

UNIVERSIDADE FEDERAL DE MINAS GERAIS
Faculdade de Ciências Econômicas
Programa de Pós-Graduação em Administração

Galba Freire Moita

**AVALIAÇÃO INTEGRATIVA DE PERFORMANCE MULTIDIMENSIONAL E
DECISÃO MULTICRITÉRIO: Um *Proxy* de Painel de Indicadores de Eficiência,
Efetividade e Qualidade para Governança de Organizações Hospitalares e Serviços de
Saúde no Brasil**

Belo Horizonte

2019

Galba Freire Moita

**AVALIAÇÃO INTEGRATIVA DE PERFORMANCE MULTIDIMENSIONAL E
DECISÃO MULTICRITÉRIO: Um *Proxy* de Painel de Indicadores de Eficiência,
Efetividade e Qualidade para Governança de Organizações Hospitalares e Serviços de
Saúde no Brasil**

Versão final

Tese apresentada ao Programa de Pós-Graduação em Administração da Universidade Federal de Minas Gerais como requisito parcial para a obtenção de título de Doutor em Administração, em acordo de cotutela com o Programa de Doutorado em Gestão – Ciência Aplicada à Decisão da Universidade de Coimbra.

Orientador: Prof. Dr. Vítor Manuel dos Reis Raposo.

Co-orientador: Prof. Dr. Allan Claudius Queiroz Barbosa.

Belo Horizonte

2019

Ficha catalográfica

M175a
2019

Moita, Galba Freire.
Avaliação integrativa de performance multidimensional e decisão multicritério [manuscrito]: um proxy de painel de indicadores de eficiência, efetividade e qualidade para governança de organizações hospitalares e serviços de saúde no Brasil

/ Galba Freire Moita . – 2019.

658 f.: il., tabs.

Orientador: Vítor Manuel dos Reis Raposo.

Coorientador: Allan Claudius Queiroz Barbosa.

Tese (doutorado) – Universidade Federal de Minas Gerais, Centro de Pós-Graduação e Pesquisas em Administração.

Inclui bibliografia (f. 343-366).

1. Indicadores de saúde – Brasil – Teses. 2. Saúde – administração – Brasil – Teses. 3. Indicadores de qualidade em assistência à saúde – Teses. I. Raposo, Vítor Manuel dos Reis. II. Barbosa, Allan Claudius Queiroz. III. Universidade de Coimbra. Faculdade de Economia. IV. Título.

CDD: 658

Elaborada pela Biblioteca da FACE/UFMG. – 138/2019

Bibliotecária Rosilene Santos (CRB- 06/2527)

Ata de aprovação da banca examinadora (folha 1/2)
Coimbra, 13 de novembro de 2019

Informe o
Original
2019/11/14



João Paulo
J. P. C.
M. L.
L. M. A. T. A.

Doutoramento em Gestão - Ciência Aplicada à Decisão, da Faculdade de Economia da Universidade de Coimbra em regime de Cotutela com a Universidade Federal de Minas Gerais - Brasil, de Galba Freire Moita

ATA DA PROVA DE DOUTORAMENTO

Aos treze dias do mês de novembro de dois mil e dezanove, na sala dos Capelos da Reitoria da Universidade de Coimbra sob a presidência de João Paulo Faria Oliveira Costa, Professor Catedrático da Faculdade de Economia da Universidade de Coimbra, por despacho de delegação de competências do Reitor da Universidade de Coimbra, número 5215/2019, publicado no Diário da República, segunda série, n.º 101, de 27 de maio, reuniu o júri da prova de doutoramento acima mencionada.

Compareceram como vogais: Allan Claudius Queiroz Barbosa, da Faculdade de Ciências Económicas da Universidade Federal de Minas Gerais - Brasil; Pedro Vasconcelos Maia do Amaral, da Faculdade de Ciências Económicas da Universidade Federal de Minas Gerais - Brasil; Jorge Manuel Trigo de Almeida Simões, da Universidade de Aveiro; Ana Maria Escoval da Silva, da Escola Nacional de Saúde Pública da Universidade Nova de Lisboa; Pedro Augusto Melo Lopes Ferreira, da Faculdade de Economia da Universidade de Coimbra; Luís Miguel Alçada Tomás de Almeida, da Faculdade de Economia da Universidade de Coimbra.

Aberta a sessão e verificando-se a existência de quórum legal, deu-se início à prova, a qual consistiu no seguinte:

Apreciação e discussão da tese, em língua portuguesa, apresentada pelo candidato intitulada: "Avaliação Integrativa de Performance Multidimensional e Decisão Multicritério: Um proxy de painel de indicadores de eficiência, efetividade e qualidade para Governação de Organizações Hospitalares e Serviços de Saúde no Brasil."

Intervieram como arguentes principais Pedro Vasconcelos Maia do Amaral e Jorge Manuel Trigo de Almeida Simões. Intervieram ainda Ana Maria Escoval da Silva, Pedro Augusto Melo Lopes Ferreira, Luís Miguel Alçada Tomás de Almeida e Allan Claudius Queiroz Barbosa.

Seguidamente, reuniu o júri para proceder à votação nos termos legais. Allan Claudius Queiroz Barbosa, Pedro Vasconcelos Maia do Amaral e Luís Miguel Alçada Tomás de Almeida votaram pela Aprovação com Distinção e Louvor atendendo à qualidade e amplitude do trabalho apresentado, assim como à maturidade científica revelada pelo candidato na forma como respondeu às questões que lhe foram colocadas durante a discussão. Jorge Manuel Trigo de Almeida Simões, Ana Maria Escoval da Silva e Pedro Augusto Melo Lopes Ferreira votaram pela Aprovação com Distinção atendendo à pertinência e atualidade do tema, à clareza da apresentação com que iniciou as provas e à forma como respondeu às questões colocadas, ainda que tenham sido apontados pelo júri alguns aspetos que poderiam ter sido melhorados, nomeadamente a organização, a dispersão e ausência de reflexão conclusiva. Tendo havido um empate, o presidente do júri João Paulo Faria Oliveira Costa votou pela Aprovação com Distinção e

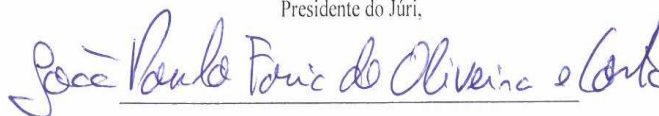
Ata de aprovação da banca examinadora (folha 2/2)
Coimbra, 13 de novembro de 2019

Louvor, exercendo o voto de qualidade ao abrigo do n.º 3 do artigo 84 do Regulamento Académico da Universidade de Coimbra. Face ao exposto, o mesmo júri deliberou, por maioria, atribuir a classificação final de Aprovado com Distinção e Louvor a Galba Freire Moita.

Nada mais havendo a tratar, foi a sessão encerrada.


Para constar se exarou a presente ata, que eu Leonor Simões dos Santos Quitério, Técnica Superior, secretariei e redigi e que depois de lida em voz alta e aprovada, vai ser assinada por todos os membros do júri presentes e por mim.

Presidente do Júri,



(João Paulo Faria Oliveira Costa)

Vogais,



(Allan Claudius Queiroz Barbosa)



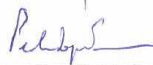
(Pedro Vasconcelos Maia do Amaral)



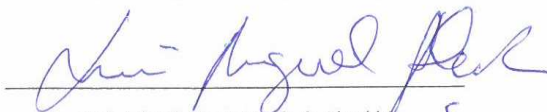
(Jorge Manuel Trigo de Almeida Simões)



(Ana Maria Escóval da Silva)

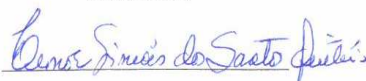


(Pedro Augusto Melo Lopes Ferreira)



(Luís Miguel Alçada Tomás de Almeida)

A Secretária,



(Leonor Simões dos Santos Quitério)

DEDICATÓRIA

A Deus, fonte de inspiração maior, que sempre tem me iluminado.

Aos meus pais (*in memoriam*) que me ensinaram valores éticos e humanos inalienáveis.

À minha esposa Valéria Neres e meus filhos Galba Júnior e Daniela Neres que abdicaram de tempos preciosos de nossa convivência para que me dedicasse a este estudo.

AGRADECIMENTOS

Aos meus orientadores Victor Manuel dos Reis Raposo e Allan Claudius Queiroz Barbosa por partilhar conhecimentos e experiências preciosas capazes de forjar aprendizados e vivências incalculáveis.

Aos gestores municipais e estaduais, aos diretores de hospitais e unidades de saúde que gentilmente autorizaram as investigações em grupos pilotos e de campo, além dos respondentes (especialistas, profissionais, gestores e usuários do SUS), sem os quais não alcançaríamos os objetivos propostos.

Ao Mardem Sousa, Ednea Miranda e Érick Darwin que me ajudaram na coordenação das pesquisas de campo e na digitação dos dados coletados.

Aos meus alunos e colegas de jornada de avaliação e gestão em saúde do Brasil que me desafiaram continuamente a buscar mais conhecimentos e construir novos caminhos...

Nada, é mais difícil, e por isso mais precioso, do que ser capaz de decidir.

Napoleão Bonaparte (s.d.)

Não existem métodos fáceis, para resolver problemas difíceis.

René Descartes (s.d.)

A verdadeira genialidade reside na capacidade de avaliação de informações incertas,
arriscadas e conflitantes.

Winston Churchill (s.d.)

RESUMO

A criação do Sistema Único de Saúde (SUS), em 1988, no Brasil, representou avanços do ponto de vista de garantias de atendimento, da organização sistêmica e da descentralização da gestão única, porém com frágil governança de desempenho. Por sua vez, a Nova Gestão Pública (NGP), ou Gerencialismo, tem exigido dos gestores públicos de saúde esforços de monitoramento de resultados, controladoria e responsabilização (*accountability*) dos gastos públicos, visto que muitos países da Europa e América Latina aplicam cerca de 9 a 10% do PIB-GDP em saúde. Este estudo analisou algumas metodologias de mensuração de resultados em saúde no cenário mundial como os projetos QIP e AHRQ (EUA), EFQM (Europa) e PATH (OMS), e projetos brasileiros como o PROADESS, PNASH, PNASS, PMAQ e as matrizes de indicadores como IDB-Ripsa e IDSUS, em busca de desenvolver um quadro teórico-conceitual (base teórico-metodológica) de avaliação de performance para organizações de saúde, aplicável às unidades do SUS. Em seguida, envolveu gestores e profissionais para a captação e validação da Percepção desses agentes chaves (*stakeholders*) quanto as dimensões e indicadores chaves para a mensuração de desempenho em organizações e serviços de saúde no contexto do SUS-Brasil de hospitais, policlínicas e outras unidades do SUS. Utilizou metodologias científicas de apoio à decisão, como *Design Research* e o método misto (qualitativo e quantitativo) de avaliação para responder à questão: há diferentes *proxies* de painel de indicadores de avaliação multidimensional para previsão e monitoramento de performance e resultados em serviços de saúde, conforme os multicritérios dos decisores chaves da rede de serviços na saúde pública do Brasil? Através de estatística descritiva, correlação quantílica e análise fatorial, foram analisadas as ponderações de itens de avaliação de performance, na perspectiva dos gestores e decisores e também de itens de análise de qualidade/satisfação de usuários. Algumas hipóteses sobre variáveis latentes e observadas desses itens de avaliação foram propostas e testadas por análise estatística, quanto às três possíveis variáveis moderadoras. Por fim, operacionalizou a validação multicêntrica, tendo construído modelos finais através de análise de cargas fatoriais, validade e confiabilidade dos itens, de *proxies* de painéis de indicadores com 453 decisores chaves e uma adaptação transcultural da escala SERVQUAL que resultou em uma escala inovadora de avaliação da qualidade e satisfação, com subescalas de intervenção, tendo sido validada por 195 especialistas e gestores e, aplicada a 2.547 usuários de 74 unidades do SUS.

Palavras-chave: Governança em saúde; Avaliação de desempenho em saúde; Indicadores de monitoramento e avaliação em saúde; Qualidade em saúde; Satisfação do paciente.

ABSTRACT

The creation of the Unified Health System (SUS) in 1988, in the Brazil, represented advances from the standpoint of service guarantees, systemic organization and decentralization of unified management but with weak governance performance. In turn, the New Public Management (NPM), or Managerialism, has required public health managers monitoring efforts results, controllership and accountability of the spending of public funds, as a lot of countries of Europe and Latin America apply about 9-10 % of health – PIB-GDP. This study analyzed some health outcomes measurement methodologies on the world stage as QIP and AHRQ projects (USA), EFQM (Europe) and PATH (WHO), and Brazilian projects as PROADESS, PNASH, PNASS, PMAQ and panel of indicators such as IDB-Ripsa and IDSUS, seeking develop a Conceitual framework (theoretical and methodological basis) performance assessment for health organizations applicable to the SUS units. Then engaged managers and professionals for the capture and validation of the perception of stakeholders as the dimensions and key indicators for performance measurement in organizations and health services in the context of SUS-Brazil hospitals, polyclinics and other SUS units. It used even scientific methods of decision support, as Design Research and mixed method (qualitative and quantitative) assessment to answer the question: there are different proxies' multidimensional evaluation indicators dashboard for forecasting and monitoring performance and outcomes in hospitals and health services, according to the multiple criteria of the key stakeholders of the care network of the public health in Brazil? Descriptive statistics, quantum correlation and factorial analysis were used to analyze the weighting of performance evaluation items from the perspective of managers and decision makers, as well as user quality/satisfaction analysis items. Some hypotheses about latent and observed variables of these evaluation items were proposed and tested by statistical analysis, regarding the three possible moderating variables. Finally, it operates the multicentric validation, having built final models through analysis of factor loads, validity and reliability of items, of proxies of indicators panels with 453 key decision makers and a cross-cultural adaptation of the SERVQUAL scale that resulted in an innovative scale of evaluation of quality and satisfaction, with intervention subscales, having been validated by 195 specialists and managers, and applied to 2,547 users of 74 SUS units.

Keywords: Health governance; Health care outcomes; in-hospital outcomes; Indicators of health monitoring and evaluation; Quality health; Patient satisfaction.

LISTA DE SIGLAS E ACRÔNIMOS

ABrES	Associação Brasileira de Economia da Saúde
ACSS	Administração Central do Sistema Nacional de Saúde (SNS) de Portugal
ACU	Análise Custo-Utilidade
AHRQ	U.S. <i>Agency for Healthcare Research and Quality</i> (Agência de Pesquisa e Qualidade nos Serviços de Saúde)
AIS	Ações Integradas de Saúde, no âmbito do Conasp
Alma-Ata	<i>Declaration. International conference on primary health care</i> , Alma-Ata, USSR, (1978).
ANS	Agência Nacional de Saúde do Brasil
BDTD	Repositório de publicações de teses das principais universidades brasileiras
BID	Banco Interamericano de Desenvolvimento
BVS	Biblioteca Virtual em Saúde, gerenciada por convênio MS-OPAS/Brasil
CAP	Caixas de Aposentadorias e Pensões
Casemix	Índice de medição da variedade das situações clínicas dos doentes tratados por cada hospital, conforme um conjunto de DRGs
CBA	Análise de Custo-Benefício
CEA	Análise de Custo-Efetividade
CEP	Comissão de Ética em Pesquisa
CHI	<i>Commission of Health Improvement</i> (Comissão de Melhoria em Saúde), no âmbito do NHS do Reino Unido.
CMB	Confederação das Misericórdias do Brasil
CNES	Cadastro Nacional de Estabelecimentos de Saúde; Módulo do Sistema de Informação do Datasus
CNS	Conferência Nacional de Saúde, realizada a cada quatro anos
CNS	Conselho Nacional de Saúde, órgão colegiado de decisão de governança sistêmica do SUS
COAP	Contrato Organizativo das Ação Pública, no âmbito do SUS
CONASP	Conselho Consultivo de Administração de Saúde Previdenciária
CONASS	Conselho Nacional de Secretários de Saúde do Brasil
CONEP	Conselho Nacional de Ética em Pesquisa do Brasil

CRS	Retornos Constantes de Escala, que ocorre quando um aumento proporcional em todos os recursos utilizados resulta num aumento proporcional (idêntico) na quantidade produzida.
DART	Europe Repositório de publicações de teses das principais universidades Europeias
Datusus	Sistema unificado de coleta e processamento de dados do Sistema Nacional de Saúde (SUS) do Brasil
DEA	<i>Data Envelopment Analysis</i> (Análise Envoltória de Dados)
DGS	Direção Geral de Saúde de Portugal
DMU	<i>Decision Making Units</i> (Unidades Tomadoras de Decisão), termo usado de forma similar para organizações, no contexto de estudos de DEA.
DRG	<i>Diagnosis Related Groups</i> (Grupos de Diagnósticos Homogêneos)
EFQM	<i>European Foundation for Quality Management</i> (Fundação Europeia para Gerenciamento da Qualidade)
EAI	Eficiência Alocativa da organização (i), valor que varia de zero a um
EEi	Eficiência Econômica Total da organização (i), valor que varia de zero a um
ENSP	Escola Nacional de Saúde Pública de Portugal
ERS	Entidade Reguladora da Saúde de Portugal que operacionaliza a regulação do sistema público e privado do SNS de Portugal
ETi	Eficiência Técnica da organização (i), valor que varia de zero a um
FBH	Federação Brasileira de Hospitais
FGV	Fundação Getúlio Vargas, destacada Universidade Privada com sedes no Sudeste do Brasil
FHEMIG	Fundação Hospitalar do Estado de Minas Gerais, que gerencia cerca de 50 unidades hospitalares e serviços públicos de saúde
FIOCRUZ	Fundação Osvaldo Cruz
GDH	Grupos de Diagnósticos Homogêneos, adaptação do conceito de DRG do Sistema Nacional de Saúde de Portugal
IAP	Institutos de Aposentadoria e Pensões
IBGE	Instituto Brasileiro de Geografia e Estatística
ICICT/Fiocruz	Instituto de Comunicação e Informação Científica em Saúde da Fiocruz
IDB	Indicadores e Dados Básicos para a Saúde no Brasil que são gerenciados pela RIPSA e o Ministério da Saúde do Brasil

IDSS	Índices de Desempenho da Saúde Suplementar, em quatro dimensões
Indicadores QUALISS	Implantados pelo Programa de Qualificação dos Prestadores de Serviços de Saúde Suplementar
IDSUS	Matriz de indicadores em 18 macrodimensões gerenciada pelo Datasus e o Ministério da Saúde do Brasil
IN	Instrução Normativa (RN) da ANS, que orientam as diretrizes das normas infralegais, obrigatórias para operadoras de planos de saúde
INAMPS	Instituto Nacional de Assistência Médica da Previdência Social
INPS	Instituto Nacional de Previdência Social
QIP	<i>Quality Indicator Project</i> (Projeto de Indicadores de Qualidade)
JCAH	<i>Joint Commission of Accreditation of Hospitals</i>
KPI	<i>Key Performance Indicators</i> (Indicadores Chaves de Performance)
MAC	Média e Alta Complexidade (hospitais e serviços especializado do SUS)
MfDR	<i>Managing for Development Results</i> , no âmbito de monitoramento e avaliação
Modelo BCR	Estima a eficiência pelo pressuposto de retornos variáveis
Modelo CCR	Estima a eficiência pelo pressuposto de retornos constantes
ML	Modelo Lógico ou Modelo Teórico-Lógico
MS	Ministério da Saúde do Brasil
NHS	<i>National Health System</i> (Serviço Nacional de Saúde do Reino Unido)
NICE	<i>National Institute for Health and Clinical Excellence</i> , que executa avaliação de tecnologias de saúde e estudos de avaliação econômica no âmbito do NHS do Reino Unido.
NGP	Nova Gestão Pública (NPM - <i>New Public Management</i>)
NPF	<i>National Performance Framework</i> (Marco Nacional de Performance), no âmbito do NHS do Reino Unido.
NSF	<i>National Service Framework</i> (Marco Nacional para Serviços), no âmbito do NHS do Reino Unido.
OCDE	Organização para a Cooperação e o Desenvolvimento Econômico, países da Europa (OECD)
OMS	Organização Mundial de Saúde (WHO), representada no Brasil, América Latina, e Caribe pela OPAS-Organização Pan-Americana de Saúde.
ONA	Organização Nacional de Acreditação, do sistema de saúde do Brasil

OPAS	Organização Pan-americana de Saúde, representante da OMS na América Latina e Caribe
PAF	<i>Performance Assessment Framework</i> do NHS (quadro teórico-conceitual de medição de desempenho do NHS).
PAS	Plano Anual de Saúde, no âmbito do SUS
PATH	<i>Performance Assessment Tool for Quality Improvement in Hospitals</i> (ferramenta de avaliação de performance para a melhoria da qualidade hospitalar)
PforR ou P4R	Programa para Resultados do Banco Mundial para o apoio ao desenvolvimento da “ <i>Accountability</i> ” de países em desenvolvimento
PGQP	Programa Gaúcho de Qualidade e Produtividade, para a certificação de empresas do Rio Grande do Sul - Brasil
PI	<i>Performance Indicator</i> (Indicador de Performance)
PIB	Produto Interno Bruto (GDP), somatório de todas as riquezas geradas em um país
PLATAFORMA_Brasil	Sistema Nacional On-Line de Controle de Investigação do Conselho Nacional de Ética do Brasil (Conep).
PNASH	Programa Nacional de Avaliação de Serviços Hospitalares, no âmbito do SUS
PNASS	Programa Nacional de Avaliação de Serviços de Saúde, no âmbito do SUS
PNQ	Programa Nacional de Qualidade, para a certificação de empresas brasileiras
POI	Programação e Orçamentação Integradas, no âmbito do Conasp
PPP	Parceria Público-Privada
PROADESS	Projeto Desenvolvimento de Metodologia de Avaliação do Desempenho do Sistema de Saúde Brasileiro
PRODEV	Programa de Estratégia e Plano de Ação para a Efetividade do Desenvolvimento do BID para os países da América Latina e Caribe
Pro-HOSP	Programa de Fortalecimento e Qualificação da Rede Hospitalar do SUS
PROM	<i>Patient Reported Outcome Measures</i> (Mensuração de Desfechos Reportados pelo Paciente)
PS	Plano de Saúde (elaboração quadrienal, nos três níveis de gestão do SUS)
PSM	<i>Problems Structuring Methods</i> (Método de Estruturação de Problemas)
QALY-Brasil	Grupo de pesquisa de mensuração de qualidade de vida no Brasil (www.qalybrasil.org.br)

RAG	Relatório Anual de Gestão (elaboração nos três níveis de gestão do SUS)
RBM	<i>Results-based Management</i> , no âmbito de monitoramento e avaliação
RIPSA	Rede Interagencial de Informações para a Saúde formatada pela OPAS/OMS, e gerenciada em conjunto com o Ministério da Saúde do Brasil
RN	Resolução Normativa (RN) da ANS, que são normas infralegais, obrigatórias para operadoras de planos de saúde
RS	Retornos de Escala, representam o grau em que um aumento proporcional em todos os recursos utilizados resulta em um aumento da quantidade produzida
SAMHPS	Sistema de Atenção Médico-Hospitalar da Previdência Social, no âmbito do Conasp
SERVQUAL	Escala <i>Service Quality</i> de avaliação de qualidade e satisfação em serviços (Parasuraman, Berry, & Zeithaml, 1991b)
SERVPERF	Escala <i>Service Performance</i> de avaliação de qualidade e satisfação em serviços (Cronin & Taylor, 1992)
SFA	<i>Stochastic Frontier Analysis</i> (Análise de Fronteira Estocástica)
SINAS@Hospitais	Sistema de Indicadores de Performance de Hospitais do Serviço Nacional de Saúde de Portugal
SIS	Sistemas de Informações em Saúde
SODA	<i>Strategic Option Development Analysis</i>
SUS	Sistema Único de Saúde, ou seja, Sistema Nacional de Saúde do Brasil
TCG	Termo de Compromisso de Gestão, no âmbito do SUS
THE KING'S FUND	Fundação independente, que desenvolve ações de melhorias no âmbito do NHS do Reino Unido, com recursos governamentais
TLCE	Termo de Consentimento Livre e Esclarecido
USAID	Agência dos Estados Unidos para o Desenvolvimento Internacional
UNISINOS	Universidade Vale dos Sinos, com sedes no Sul do Brasil.
VRS	Retornos Variáveis de Escala, que ocorre quando um aumento proporcional em todos os recursos utilizados resulta num aumento desproporcional (maior ou menor) na quantidade produzida.

LISTA DE PRINCIPAIS TERMOS SIMILARES (PORTUGUAL *VERSUS* BRASIL)

Artefacto	Artefato ou dispositivo
Aspeto	Aspecto
Conceptual	Conceitual
Controlo	Controle
Cuidados de saúde	Atenção à saúde
De seguida	Em seguida
Distrito de Saúde	Microrregional de saúde
Expetativa	Expectativa
Gestão de Topo	Gestão ou direção organizacional
Gestão intermédia	Gestão departamental/intermediária
Governação	Governança
Investigação	Pesquisa
Objectivo	Objetivo
Utente	Paciente
Perceção	Percepção
Perspetiva	Perspectiva
Planeado	Planejado
Planeamento	Planejamento
Província	Estado (unidades da federação)
Recolha de dados	Coleta de dados
Usuário	Utilizador

LISTA DE ABREVIATURAS

Graf.	Gráfico
$H1_0$	Primeira hipótese nula
$H1_1$	Primeira hipótese alternativa
Hn_0	E-ésima hipótese nula
Hn_1	E-ésima hipótese alternativa
$H11_0$	Primeira hipótese nula (sub-hipótese um)
$H11_1$	Primeira hipótese alternativa (sub-hipótese um)
Hnx_0	E-ésima hipótese nula (sub-hipótese X)
Hnx_1	E-ésima hipótese alternativa (sub-hipótese X)
me	Número de entradas ou insumos da avaliação pelo método DEA
no	Número de organizações da avaliação pelo método DEA
p.	Página
p.ex.	Por exemplo
pp.	Páginas
sp	Número de saídas ou produtos da avaliação pelo método DEA
Sig. ou P-Value	Probabilidade de significância
Tab.	Tabela
vs	<i>Versus</i>

LISTA DE QUADROS

<i>Quadro 1</i> - Gastos em saúde: percentual de despesas pública <i>versus</i> privada, países OCDE, Brasil e outros (2013). Fonte: Adaptado pelo autor – Dados disponíveis em http://data.worldbank.org/data-catalog/world-development-indicators .	37
<i>Quadro 2</i> - Síntese do enquadramento e abordagens da investigação.	55
<i>Quadro 3</i> - Quadro comparativo dos vários modelos multidimensionais analisados. Fonte: Carvalho (2007). Adaptado de Ramalho (2005: 248-249).	68
<i>Quadro 4</i> - Modelo integrativo da eficácia organizacional (Morin et al., 1994) – critérios e indicadores de cada dimensão. Fonte: Carvalho (2007). Adaptado de Morin, Savoie e Beaudin (1994).	69
<i>Quadro 5</i> - Síntese integrativa das teorias da avaliação em saúde e bases conceptuais do estudo	164
<i>Quadro 6</i> - Síntese dos procedimentos metodológicos da investigação.	180
<i>Quadro 7</i> - Síntese das análises descritivas dos níveis de conhecimento (atribuídos) dos principais domínios de avaliação (e refletidos nos macrodomínios de avaliação).	183
<i>Quadro 8</i> - Síntese das análises descritivas dos níveis de conhecimento (atribuídos) dos principais indicadores (e refletidos nas dimensões de performance).	186
<i>Quadro 9</i> - Síntese das análises descritivas dos níveis de valoração (atribuídos) aos principais constructos propostos (e refletidos nas subdimensões e macrodimensões da escala de qualidade percebida).	189
<i>Quadro 10</i> - Síntese das análises descritivas dos níveis de valoração (atribuídos) das questões validadas (e refletidos nas subdimensões e macrodimensões da escala de qualidade percebida).	192
<i>Quadro 11</i> - Visão global da fase empírica dos estudos/etapas de 1 a 3: atividades, técnicas de investigação e métodos de análises de resultados (macrodomínios/domínios de avaliação e dimensões/indicadores de performance).	198
<i>Quadro 12</i> - Visão global da fase empírica dos estudos/etapas de 4 e 5: atividades, técnicas de investigação e métodos de análises de resultados (macrodimensões/subdimensões de constructos e de questões validadas da qualidade percebida).	199
<i>Quadro 13</i> - Definições conceituais das macrodomínios e dimensões de performance.	214
<i>Quadro 14</i> - Definições conceituais das subdimensões e macrodimensões de qualidade/satisfação percebida	246
<i>Quadro 15</i> - Síntese de validações e conclusões de hipóteses sobre itens de avaliação de performance (estudos 1, 2 e 3)	305
<i>Quadro 16</i> - Síntese validações e conclusões de hipóteses sobre itens de medição de qualidade/satisfação (estudos 4 e 5)	316

Quadro 17- Síntese dos quadros conceituais (frameworks) nacionais para o desempenho do sistema de saúde (países selecionados) **Erro! Marcador não definido.**

Quadro 18- Conjunto principal dos 27 indicadores de desempenho do hospital (projeto PATH) **Erro! Marcador não definido.**

Quadro 19- Quadro de objetivos *versus* quadro teórico-conceitual (framework)**Erro! Marcador não definido.**

Quadro 20- Pesquisa integrativa de definições conceituais de satisfação de consumidores (usuários-utilizadores) – ordenamento cronológico **Erro! Marcador não definido.**

Quadro 21 - Síntese de aplicação da escala SERVQUAL em organizações hospitalares e unidades de saúde do Brasil. **Erro! Marcador não definido.**

Quadro 22- Equalização de 22 questões da escala SERVQUAL (10 dimensões originais para 5 dimensões atuais) **Erro! Marcador não definido.**

Quadro 23- Lista de 22 questões da escala SERVQUAL original **Erro! Marcador não definido.**

Quadro 24- Síntese de estudos da revisão sistemática (OPAS – AbrES - MS)**Erro! Marcador não definido.**

Quadro 25- Síntese de estudos de aplicação do método DEA no contexto da saúde brasileira **Erro! Marcador não definido.**

Quadro 26- Estudos recentes de monitoramento e avaliação desempenho aplicadas em organizações do sistema de saúde do Brasil **Erro! Marcador não definido.**

Quadro 27- Bibliometria de publicações de monitoramento e avaliação de desempenho em organizações do sistema de saúde do Brasil (1986 – 2018) **Erro! Marcador não definido.**

Quadro 28- Matriz de indicadores e fontes primárias de informação - PROADES**Erro! Marcador não definido.**

Quadro 29- Quadro-síntese das principais bases de dados e informações do SUS**Erro! Marcador não definido.**

Quadro 30- Síntese do processo de evolução do IDSS da ANS (2003 a 2009)**Erro! Marcador não definido.**

Quadro 31- Rol de indicadores IDSS e de Indicadores de mapeamento assistencial do QUALISS-ANS (2016) **Erro! Marcador não definido.**

Quadro 32- Dimensões dos indicadores do sistema SIPAGEH **Erro! Marcador não definido.**

Quadro 33- Matriz de indicadores hospitalares – SIPAGEH **Erro! Marcador não definido.**

Quadro 34- Distribuição dos indicadores identificados, segundo as instituições estudadas **Erro! Marcador não definido.**

LISTA DE FIGURAS

<i>Figura 1</i> - Evolução da Taxa Bruta de Mortalidade por mil habitantes/ano (Brasil, 1996 a 2017) - grandes regiões e estados selecionados (> média do Brasil, no último quadriênio). Fonte: Elaboração própria - Dados disponíveis em: http://datasus.saude.gov.br/informacoes-de-saude/tabnet/estatisticas-vitais .	39
<i>Figura 2</i> - Bibliometria de estudos de avaliação em saúde no Brasil (1990 - 2018).	44
<i>Figura 3</i> - Matriz de Dimensões da Avaliação de Desempenho do Sistema de Saúde. Fonte: http://www.proadess.icict.fiocruz.br/index.php?pag=matp (F Viacava et al., 2004: 115)	102
<i>Figura 4</i> - Antecedentes da satisfação mais frequentemente utilizados. Fonte: adaptado de Szymanski e Henard (2001 apud Vinagre, 2008: 26).	120
<i>Figura 5</i> - Inter-relações entre necessidades, valores e desejos <i>versus</i> expectativas. Fonte: Adaptado de (Oliver, 1997 apud Vinagre, 2008: 29).	120
<i>Figura 6</i> - Modelo completo de desconfirmação das expectativas com o desempenho percebido. Fonte: Adaptado de (Oliver, 1997 apud Vinagre, 2008: 33).	121
<i>Figura 7</i> - Princípios doutrinários e organizativos do SUS.	127
<i>Figura 8</i> - Gastos em saúde: percentual do PIB (GDP), países OCDE e outros selecionados (2016 ou ano mais próximo). Fonte: <i>OECD Health Statistics 2017</i> . <i>WHO Global Health Expenditure Database</i> . StatLink http://dx.doi.org/10.1787/888933604229 .	128
<i>Figura 9</i> - Evolução populacional por grupos etários no Brasil (2000-2060). Fonte: IBGE. Diretoria de Pesquisas. Coordenação de População e Indicadores Sociais, 2013. http://www.ibge.gov.br/home/presidencia/noticias/imprensa/ppts/00000014425608112013563329137649.pdf	129
<i>Figura 10</i> - Magnitude dos gastos por tipologia de despesas do sistema de saúde de países OCDE e outros selecionados (2015 ou ano mais próximo). Fonte: <i>OECD Health Statistics 2017</i> . <i>WHO Global Health Expenditure Database</i> . StatLink http://dx.doi.org/10.1787/888933604438 .	130
<i>Figura 11</i> - O arranjo organizacional e as inter-relações do sistema público e privado de saúde do Brasil.	134
<i>Figura 12</i> - Fluxo lógico do processo de estruturação, aplicação e validação dos instrumentos de itens de avaliação de performance e escala de qualidade/satisfação	171
<i>Figura 13</i> - Fluxo lógico de atividades para avaliação de performance e avaliação de qualidade/satisfação.	173
<i>Figura 14</i> - Fluxo lógico de metodologias para avaliação de performance e avaliação de qualidade/satisfação.	173
<i>Figura 15</i> - Fluxo lógico integrativo de aplicação da investigação (avaliação de performance – estudos 1 a 3 e avaliação de qualidade – estudos 4 e 5).	176
<i>Figura 16</i> - Fluxo lógico da aplicação da investigação e agregação de resultados (etapa 1 - macrodônios e domínios de avaliação).	181

<i>Figura 17</i> - Fluxo do processo de validação de macrodomínios e domínios de avaliação (etapa 1).	182
<i>Figura 18</i> - Fluxo lógico da aplicação da investigação e agregação de resultados (estudos/etapas 1 & 2 - dimensões e indicadores de performance).	183
<i>Figura 19</i> - Fluxo do processo de validação de dimensões e indicadores de performance (etapas 2 & 3)	185
<i>Figura 20</i> - Fluxo lógico da aplicação da investigação e agregação de resultados (estudo/etapa 4 - macrodimensões, subdimensões e constructos de qualidade percebida).	186
<i>Figura 21</i> - Fluxo do processo de validação de constructos adaptados ao SUS (qualidade percebida) (etapa 4).	188
<i>Figura 22</i> - Fluxo lógico da aplicação da investigação e agregação de resultados (estudo/etapa 5 - macrodimensões, subdimensões e questões validadas).	189
<i>Figura 23</i> - Fluxo do processo de validação de questões de qualidade percebida (etapa 5 – questões validadas).	191
<i>Figura 24</i> - Modelo gráfico das variáveis latentes e observadas de performance.	192
<i>Figura 25</i> - Modelo gráfico das variáveis latentes e observadas de qualidade percebida.	194
<i>Figura 26</i> - Síntese de fluxo de análises de itens de análise dos painéis de indicadores e dimensões de performance.	200
<i>Figura 27</i> - Esquema lógico de variáveis da investigação e análises de resultados (etapa 1 - macrodomínios e domínios de avaliação).	201
<i>Figura 28</i> - Esquema lógico de variáveis da investigação e análises de resultados (etapas 2 & 3 - dimensões e indicadores de performance).	202
<i>Figura 29</i> - Síntese dos fluxos, métodos e técnicas de análises de resultados (etapas 1 a 3 - ciclo 1 e 2 - proxies A a G).	203
<i>Figura 30</i> - Síntese de fluxo de análises de itens de avaliação de qualidade/satisfação.	204
<i>Figura 31</i> - Esquema lógico de variáveis da investigação e análises de resultados (etapa 4 - macrodimensões, subdimensões e constructos de qualidade percebida).	205
<i>Figura 32</i> - Esquema lógico de variáveis da investigação e análises de resultados (etapas 5 - macrodimensões, subdimensões e questões validadas de qualidade percebida).	206
<i>Figura 33</i> - Síntese dos fluxos, métodos e técnicas de análises de resultados (etapas 4 e 5 - ciclo 1 e 2 - proxies H a N).	207
<i>Figura 34</i> - Gráfico de barras das variáveis de caracterização selecionadas para estudo de fatores mediadores – Proxy A+B (ciclo 1) e proxy C (ciclo 2) (estudo 1) (Visão de especialistas).	216
<i>Figura 35</i> - Grau de importância/relevância de domínios e macrodomínios de avaliação (proxy A+B versus proxy C) (Visão de especialistas).	219
<i>Figura 36</i> - Gráfico de barras das variáveis de caracterização selecionadas para estudo de fatores mediadores – Proxy D a F e proxy E+F (ciclo 1, estudo 2) (Visão de especialistas, profissionais e gestores).	221

<i>Figura 37-</i> Gráfico de barras com intervalos de confiança para os indicadores e as dimensões de performance - <i>Proxy E+F</i> (ciclo 1, estudo 2) (Perspectiva indireta para Decisores).	224
<i>Figura 38-</i> Gráfico de barras das variáveis de caracterização selecionadas para estudo de fatores mediadores – <i>Proxy G</i> (ciclo 2, estudo 3) (Visão de Decisores).	227
<i>Figura 39-</i> Grau de importância/relevância de indicadores e dimensões de performance (<i>proxy G</i>).	230
<i>Figura 40-</i> Análise Fatorial Confirmatória da Performance Global Percebida – <i>Proxy C</i> (ciclo 2, estudo 1) (Visão de gestores e profissionais).	235
<i>Figura 41-</i> Análise Fatorial Confirmatória da Performance Global Percebida – <i>Proxy E+F</i> (ciclo 1, estudo 2) (Perspectiva indireta para Decisores).	239
<i>Figura 42-</i> Análise Fatorial Confirmatória da Performance Global Percebida – <i>Proxy G</i> (ciclo 2, estudo 3) (Visão de Decisores).	244
<i>Figura 43-</i> Gráfico de barras com intervalos de confiança para os constructos originais - <i>Proxy H+I</i> (ciclo 1, estudo 4) (Perspectiva indireta para Usuários e Profissionais).	251
<i>Figura 44-</i> Gráfico de barras com intervalos de confiança para as subdimensões de constructos - <i>Proxy H+I</i> (ciclo 1, estudo 4) (Perspectiva indireta para Usuários e Profissionais).	253
<i>Figura 45-</i> Gráfico de barras com intervalos de confiança para os constructos de qualidade - <i>Proxy J</i> (ciclo 2, estudo 4) (Visão de Usuários).	260
<i>Figura 46-</i> Gráfico de barras com intervalos de confiança para as subdimensões de constructos – <i>Proxy J</i> (ciclo 2, estudo 4) (Visão de Usuários).	261
<i>Figura 47 -</i> Gráfico de barras com intervalos de confiança para questões validadas - <i>Proxy L</i> (ciclo 1, estudo 5) - Perspectiva indireta para Usuários (Expectativa e Percepção).	268
<i>Figura 48-</i> Gráfico de barras com intervalos de confiança para as subdimensões de questões validadas – <i>Proxy L</i> (ciclo 1, estudo 5) - Perspectiva indireta para Usuários (Expectativa e Percepção).	270
<i>Figura 49-</i> Gráfico de barras com intervalos de confiança para questões validadas - <i>Proxy M+N</i> (ciclo 2, estudo 5) - Visão de Usuários (Expectativa e Percepção).	276
<i>Figura 50-</i> Gráfico de barras com intervalos de confiança para as subdimensões de questões validadas – <i>Proxy M+N</i> (ciclo 2, estudo 5) - Visão de Usuários (Expectativa e Percepção).	278
<i>Figura 51 - Análise Fatorial Confirmatória da Qualidade Global Percebida - Proxy J (ciclo 2, estudo 4) (Visão de Usuários).</i>	288
<i>Figura 52-</i> Análise Fatorial Confirmatória da Qualidade Global Percebida (ciclo 2: proxy M+N) – Visão de Usuários (Expectativa).	295
<i>Figura 53-</i> Análise Fatorial Confirmatória da Qualidade Global Percebida (ciclo 2: proxy M+N) – Visão de Usuários (Percepção).	296
<i>Figura 54-</i> Modelo gráfico das variáveis latentes (Macrodomínios) e observadas (Domínios) de performance	298
<i>Figura 55-</i> Modelo gráfico das variáveis latentes (Dimensões) e observadas (Indicadores) de performance	299

- Figura 56- Modelo gráfico das variáveis latentes (Macro & Subdimensões) e observadas (Questões de qualidade) de qualidade percebida 307
- Figura 57- Modelo lógico 1: Atributos ou características para avaliação em saúde **Erro! Marcador não definido.**
- Figura 58 - Modelo lógico 2: Atributos de avaliação equalizada oferta vs demanda vs resultados de serviços de saúde **Erro! Marcador não definido.**
- Figura 59- Gráfico de barras com intervalos de confiança para os macrodomínios de avaliação – Proxy A+B (ciclo 1, estudo 1) (Visão de especialistas). **Erro! Marcador não definido.**
- Figura 60- Gráfico de barras com intervalos de confiança para os macrodomínios de avaliação – Proxy C (ciclo 2, estudo 1) (Visão de gestores e profissionais). **Erro! Marcador não definido.**
- Figura 61- Gráfico de barras com intervalos de confiança para as dimensões de performance – Proxy E+F (ciclo 1, estudo 2) (Perspectiva indireta para Decisores). **Erro! Marcador não definido.**
- Figura 62 - Gráfico de barras com intervalos de confiança para as dimensões de performance (DD1 a DD7) – Proxy G (ciclo 2, estudo 3) (Visão de Decisores). **Erro! Marcador não definido.**
- Figura 63- Correlações entre domínios e macrodomínios de avaliação - Proxy C (ciclo 2, estudo 1) (Visão de gestores e profissionais). **Erro! Marcador não definido.**
- Figura 64 - Correlações entre dimensões e indicadores de performance – Proxy E+F (ciclo 1, estudo 2) (Perspectiva indireta para Decisores). **Erro! Marcador não definido.**
- Figura 65- Correlações entre dimensões e indicadores de performance – Proxy G (ciclo 2, estudo 3) (Visão de Decisores). **Erro! Marcador não definido.**
- Figura 66- Correlações entre constructos originais e subdimensões de constructos- Proxy J (ciclo 2, estudo 4) (Visão de Usuários). **Erro! Marcador não definido.**
- Figura 67- Correlações entre as subdimensões e as macrodimensões de constructos – Proxy J (ciclo 2, estudo 4) (Visão de Usuários). **Erro! Marcador não definido.**
- Figura 68- Correlações entre questões validadas e subdimensões de questões validadas – Proxy M+N (Expectativa). **Erro! Marcador não definido.**
- Figura 69- Correlações entre as subdimensões e as macrodimensões de questões validadas – Proxy M+N (Expectativa). **Erro! Marcador não definido.**
- Figura 70- Correlações entre questões validadas e subdimensões de questões validadas – Proxy M+N (Percepção). **Erro! Marcador não definido.**
- Figura 71- Correlações entre as subdimensões e as macrodimensões de questões validadas – Proxy M+N (Percepção). **Erro! Marcador não definido.**
- Figura 72- Quadro teórico-conceitual (framework) EFQM adaptado aos cuidados de saúde, conforme as dimensões do modelo PATH **Erro! Marcador não definido.**

LISTA DE TABELAS

Tabela 1 - Análise descritiva dos domínios de avaliação - Proxy A+B (ciclo 1) e proxy C (ciclo 2) (estudo 1) (Visão de especialistas).	217
Tabela 2 - Análise descritiva dos macrodomínios de avaliação - Proxy A+B (ciclo 1) e proxy C (ciclo 2) (estudo 1) (Visão de especialistas).	218
Tabela 3 - Análise descritiva dos indicadores de performance – Proxy E+F (ciclo 1, estudo 2) (Perspectiva indireta para Decisores).	222
Tabela 4- Análise descritiva das dimensões de performance – Proxy E+F (ciclo 1, estudo 2) (Perspectiva indireta para Decisores).	223
Tabela 5- Análise descritiva dos indicadores de performance (IDD01 a IDD56) – Proxy G (ciclo 2, estudo 3) (Visão de Decisores).	227
Tabela 6- Análise descritiva das dimensões de performance (DD1 a DD7) – Proxy G (ciclo 2, estudo 3) (Visão de Decisores).	229
Tabela 7- Análise Fatorial Confirmatória dos domínios de avaliação – Proxy C (ciclo 2, estudo 1) (Visão de gestores e profissionais).	232
Tabela 8- Validação por critérios de qualidade e validade dos macrodomínios de avaliação - Proxy C (ciclo 2, estudo 1) (Visão de gestores e profissionais).	233
Tabela 9- Análise Fatorial Confirmatória da Performance Global Percebida – Proxy C (ciclo 2, estudo 1) (Visão de gestores e profissionais).	233
Tabela 10 - Validação da Performance Global Percebida – Proxy C (ciclo 2, estudo 1) (Visão de gestores e profissionais).	234
Tabela 11 - Parâmetros de qualidade do modelo final da Performance Global Percebida – Proxy C (ciclo 2, estudo 1) (Visão de gestores e profissionais).	234
Tabela 12- Análise Fatorial Confirmatória dos indicadores de performance – Proxy E+F (ciclo 1, estudo 2) (Perspectiva indireta para Decisores).	236
Tabela 13- Validação por critérios de qualidade e validade das dimensões de performance – Proxy E+F (ciclo 1, estudo 2) (Perspectiva indireta para Decisores).	237
Tabela 14- Análise Fatorial Confirmatória das dimensões de performance – Proxy E+F (ciclo 1, estudo 2) (Perspectiva indireta para Decisores).	237
Tabela 15 - Validação da Performance Global Percebida – Proxy E+F (ciclo 1, estudo 2) (Perspectiva indireta para Decisores).	238
Tabela 16- Parâmetros de qualidade (ajustamento) do modelo final da Performance Global Percebida – Proxy E+F (ciclo 1, estudo 2) (Perspectiva indireta para Decisores).	238

Tabela 17- Análise Fatorial Confirmatória dos indicadores de performance – Proxy G (ciclo 2, estudo 3) (Visão de Decisores).	240
Tabela 18- Validação por critérios de qualidade e validade das dimensões de performance – Proxy G (ciclo 2, estudo 3) (Visão de Decisores).	242
Tabela 19- Análise Fatorial Confirmatória das dimensões de performance – Proxy G (ciclo 2, estudo 3) (Visão de Decisores).	242
Tabela 20- Validação da Performance Global Percebida – Proxy G (ciclo 2, estudo 3) (Visão de Decisores).	243
Tabela 21- Parâmetros de qualidade (ajustamento) do modelo final da Performance Global Percebida – Proxy G (ciclo 2, estudo 3) (Visão de Decisores).	243
Tabela 22- Análise descritiva das variáveis de caracterização selecionadas - Proxy H+I (ciclo 1, estudo 4) (Perspectiva indireta para Usuários e Profissionais).	248
Tabela 23- Análise descritiva de constructos originais - Proxy H+I (ciclo 1, estudo 4) (Perspectiva indireta para Usuários e Profissionais).	249
Tabela 24- Análise descritiva das subdimensões de constructos - Proxy H+I (ciclo 1, estudo 4) (Perspectiva indireta para Usuários e Profissionais).	252
Tabela 25- Análise descritiva das macrodimensões de qualidade - Proxy H+I (ciclo 1, estudo 4) (Perspectiva indireta para Usuários e Profissionais).	254
Tabela 26- Análise descritiva das variáveis de caracterização selecionadas - Proxy J (ciclo 2, estudo 4) (Visão de Usuários).	256
Tabela 27- Análise descritiva dos constructos de qualidade – Proxy J (ciclo 2, estudo 4) (Visão de Usuários).	258
Tabela 28- Análise descritiva das subdimensões de constructos – Proxy J (ciclo 2, estudo 4) (Visão de Usuários).	260
Tabela 29- Análise descritiva das macrodimensões de constructos – Proxy J (ciclo 2, estudo 4) (Visão de Usuários).	262
Tabela 30- Análise descritiva das variáveis de caracterização selecionadas - Proxy K+L - (ciclo 1, estudo 5) (Perspectiva indireta para Usuários).	265
Tabela 31- Análise descritiva das questões validadas – Proxy L (ciclo 1, estudo 5) - Perspectiva indireta para Usuários (Expectativa e Percepção).	266
Tabela 32- Análise descritiva das subdimensões de questões validadas – Proxy L (ciclo 1, estudo 5) - Perspectiva indireta para Usuários (Expectativa e Percepção).	268
Tabela 33- Análise descritiva das macrodimensões de qualidade – Proxy L (ciclo 1, estudo 5) - Perspectiva indireta para Usuários (Expectativa e Percepção).	270

Tabela 34- Análise descritiva das variáveis de caracterização selecionadas - Proxy M+N (ciclo 2, estudo 5) (Visão de Usuários).	274
Tabela 35- Análise descritiva das questões validadas – Proxy M+N (ciclo 2, estudo 5) - Visão de Usuários (Expectativa e Percepção).	275
Tabela 36- Análise descritiva das subdimensões de questões validadas– Proxy M+N (ciclo 2, estudo 5) - Visão de Usuários (Expectativa e Percepção).	277
Tabela 37 - Análise descritiva das macrodimensões de qualidade – Proxy M+N (ciclo 2, estudo 5) - Visão de Usuários (Expectativa e Percepção).	278
Tabela 38- Análise Fatorial Confirmatória dos constructos originais – Proxy J (ciclo 2, estudo 4) (Visão de Usuários).	282
Tabela 39- Validação das subdimensões de constructos – Proxy J (ciclo 2, estudo 4) (Visão de Usuários).	283
Tabela 40- Análise Fatorial Confirmatória das subdimensões de constructos – Proxy J (ciclo 2, estudo 4) (Visão de Usuários).	284
Tabela 41- Validação das subdimensões de constructos – Proxy J (ciclo 2, estudo 4) (Visão de Usuários).	285
Tabela 42- Análise Fatorial Confirmatória da Qualidade Global Percebida – Proxy J (ciclo 2, estudo 4) (Visão de Usuários).	285
Tabela 43 - Validação da Qualidade Global Percebida – Proxy J (ciclo 2, estudo 4) (Visão de Usuários).	285
Tabela 44- Parâmetros de qualidade (ajustamento) do modelo final da Qualidade Global Percebida – Proxy J (ciclo 2, estudo 4) (Visão de Usuários).	286
Tabela 45- Análise Fatorial Confirmatória das subdimensões de questões validadas (Expectativa)– Proxy M+N.	290
Tabela 46 - Análise Fatorial Confirmatória das subdimensões de questões validadas (Percepção) – Proxy M+N	290
Tabela 47 - Validação das subdimensões de questões validadas (Expectativa) - Proxy M+N.	291
Tabela 48 - Validação das subdimensões de questões validadas (Percepção) –ProxyM+N.	291
Tabela 49- Análise Fatorial Confirmatória das macrodimensões de questões validadas (Expectativa) – Proxy M+N.	292
Tabela 50- Análise Fatorial Confirmatória das macrodimensões de questões validadas (Percepção) – Proxy M+N.	293
Tabela 51- Validação das macrodimensões questões validadas (Expectativa) - Proxy M+N.	293
Tabela 52- Validação das macrodimensões de questões validadas (Percepção) - Proxy M+N.	293

Tabela 53- Parâmetros de qualidade do modelo final (Expectativa) – Proxy M+N.	294
Tabela 54 - Parâmetros de qualidade do modelo final (Percepção) – Proxy M+N.	294
Tabela 55- Comparação dos macrodomínios de avaliação em relação às variáveis de caracterização selecionadas (nível de atenção à saúde e de cargo/função) - Proxy A+B (ciclo 1, estudo 1) (Visão de especialistas).	Erro! Marcador não definido.
Tabela 56- Comparação dos macrodomínios de avaliação em relação às variáveis de caracterização selecionadas (nível de atenção à saúde e de cargo/função) - Proxy C (ciclo 2, estudo 1) (Visão de gestores e profissionais).	Erro! Marcador não definido.
Tabela 57- Comparação das dimensões de performance em relação aos níveis de atenção à saúde - Proxy E+F (ciclo 1, estudo 2) (Perspectiva indireta para Decisores).	Erro! Marcador não definido.
Tabela 58 - Comparação das dimensões de performance em relação aos níveis de cargo/função ocupados - Proxy E+F (ciclo 1, estudo 2) (Perspectiva indireta para Decisores).	Erro! Marcador não definido.
Tabela 59- Comparação das dimensões de performance (DD1 a DD7) em relação aos níveis de atenção à saúde – Proxy G (ciclo 2, estudo 3) (Visão de Decisores).	Erro! Marcador não definido.
Tabela 60- Comparação das dimensões de performance (DD1 a DD7) em relação aos níveis de complexidade da assistência – Proxy G (ciclo 2, estudo 3) (Visão de Decisores).	Erro! Marcador não definido.
Tabela 61- Comparação das dimensões de performance (DD1 a DD7) em relação aos níveis de cargo/função ocupados – Proxy G (ciclo 2, estudo 3) (Visão de Decisores).	Erro! Marcador não definido.
Tabela 62- Correlações entre domínios e macrodomínios de avaliação - Proxy C (ciclo 2, estudo 1) (Visão de gestores e profissionais).	Erro! Marcador não definido.
Tabela 63- Comparação da Performance Global Percebida em relação as variáveis de caracterização selecionadas (nível de atenção à saúde e de cargo/função) – Proxy C (ciclo 2, estudo 1) (Visão de gestores e profissionais).	Erro! Marcador não definido.
Tabela 64- Correlações entre dimensões e indicadores de performance – Proxy E+F (ciclo 1, estudo 2) (Perspectiva indireta para Decisores).	Erro! Marcador não definido.
Tabela 65- Comparação da Performance Global Percebida em relação às variáveis de caracterização selecionadas – Proxy E+F (ciclo 1, estudo 2) (Perspectiva indireta para Decisores).	Erro! Marcador não definido.
Tabela 66- Correlações entre dimensões e indicadores de performance – Proxy G (ciclo 2, estudo 3) (Visão de Decisores).	Erro! Marcador não definido.

Tabela 67- Comparação da Performance Global Percebida em relação às variáveis de caracterização selecionadas – Proxy G (ciclo 2, estudo 3) (Visão de Decisores). **Erro! Marcador não definido.**

Tabela 68 - Comparação das subdimensões de constructos em relação aos níveis de atenção à saúde - Proxy H+I (ciclo 1, estudo 4) (Perspectiva indireta para Usuários). **Erro! Marcador não definido.**

Tabela 69 - Comparação das subdimensões de constructos em relação aos níveis de atenção à saúde - Proxy H+I (ciclo 1, estudo 4) (Perspectiva indireta para Profissionais).**Erro! Marcador não definido.**

Tabela 70- Comparação das subdimensões de constructos em relação aos níveis de complexidade da assistência - Proxy H+I (ciclo 1, estudo 4) (Perspectiva indireta para Usuários).**Erro! Marcador não definido.**

Tabela 71- Comparação das subdimensões de constructos em relação aos níveis de complexidade da assistência - Proxy H+I (ciclo 1, estudo 4) (Perspectiva indireta para Profissionais).**Erro! Marcador não definido.**

Tabela 72- Comparação das subdimensões de constructos em relação aos níveis de cargo/função ocupados - Proxy H+I (ciclo 1, estudo 4) (Perspectiva indireta para Usuários).**Erro! Marcador não definido.**

Tabela 73- Comparação das subdimensões de constructos em relação aos níveis de cargo/função ocupados - Proxy H+I (ciclo 1, estudo 4) (Perspectiva indireta para Profissionais).**Erro! Marcador não definido.**

Tabela 74- Comparação das macrodimensões de qualidade em relação aos níveis de atenção à saúde - Proxy H+I (ciclo 1, estudo 4) (Perspectiva indireta para Usuários e Profissionais).**Erro! Marcador não definido.**

Tabela 75- Comparação das macrodimensões de qualidade em relação aos níveis de complexidade da assistência- Proxy H+I (ciclo 1, estudo 4) (Perspectiva indireta para Usuários e Profissionais). **Erro! Marcador não definido.**

Tabela 76- Comparação das macrodimensões de qualidade em relação aos níveis de cargo/função ocupados - Proxy H+I (ciclo 1, estudo 4) (Perspectiva indireta para Usuários e Profissionais). **Erro! Marcador não definido.**

Tabela 77- Comparação das subdimensões de constructos em relação aos níveis de atenção à saúde – Proxy J (ciclo 2, estudo 4) (Visão de Usuários). **Erro! Marcador não definido.**

Tabela 78- Comparação das subdimensões de constructos em relação aos níveis de complexidade da assistência – Proxy J (ciclo 2, estudo 4) (Visão de Usuários). **Erro! Marcador não definido.**

Tabela 79- Comparação das subdimensões de questões validadas em relação aos níveis de escolaridade de usuários – Proxy J (ciclo 2, estudo 4) (Visão de Usuários). **Erro! Marcador não definido.**

Tabela 80- Análise descritiva das macrodimensões de qualidade – Proxy J (ciclo 2, estudo 4) (Visão de Usuários). **Erro! Marcador não definido.**

Tabela 81- Comparação das macrodimensões de constructos em relação aos níveis de complexidade da assistência– Proxy J (ciclo 2, estudo 4) (Visão de Usuários). **Erro! Marcador não definido.**

Tabela 82- Comparação das macrodimensões de constructos em relação aos níveis de escolaridade de usuários – Proxy J (ciclo 2, estudo 4) (Visão de Usuários). **Erro! Marcador não definido.**

Tabela 83- Comparação das subdimensões de questões validadas em relação aos níveis de atenção à saúde – Proxy L (ciclo 1, estudo 5) - Perspectiva indireta para Usuários (Percepção). **Erro! Marcador não definido.**

Tabela 84- Comparação das subdimensões de questões validadas em relação aos níveis de complexidade da assistência – Proxy L (ciclo 1, estudo 5) - Perspectiva indireta para Usuários (Percepção). **Erro! Marcador não definido.**

Tabela 85- Comparação das subdimensões de questões validadas em relação aos níveis de cargo/função – Proxy L (ciclo 1, estudo 5) - Perspectiva indireta para Usuários - (Percepção). **Erro! Marcador não definido.**

Tabela 86- Comparação das macrodimensões de questões validadas em relação aos níveis de atenção à saúde – Proxy L (ciclo 1, estudo 5) - Perspectiva indireta para Usuários (Expectativa, Percepção). **Erro! Marcador não definido.**

Tabela 87- Comparação das macrodimensões de questões validadas em relação aos níveis de complexidade da assistência – *Proxy L* (ciclo 1, estudo 5) - Perspectiva indireta para Usuários (Expectativa, Percepção). **Erro! Marcador não definido.**

Tabela 88- Comparação das macrodimensões de questões validadas em relação aos níveis de cargo/função – *Proxy L* (ciclo 1, estudo 5) - Perspectiva indireta para Usuários (Expectativa, Percepção). **Erro! Marcador não definido.**

Tabela 89- Associação entre Expectativa e Percepção nas subdimensões de questões validadas – *Proxy L* (ciclo 1, estudo 5) - Perspectiva indireta para Usuários. **Erro! Marcador não definido.**

Tabela 90- Associação entre Expectativa e Percepção nas macrodimensões – *Proxy L* (ciclo 1, estudo 5) - Perspectiva indireta para Usuários. **Erro! Marcador não definido.**

Tabela 91- Comparação das subdimensões de questões validadas em relação aos níveis de atenção à saúde – *Proxy M+N* (ciclo 2, estudo 5) - Visão de Usuários (Expectativa). **Erro! Marcador não definido.**

Tabela 92- Comparação das subdimensões de questões validadas em relação aos níveis de atenção à saúde – *Proxy M+N* (ciclo 2, estudo 5) - Visão de Usuários (Percepção). **Erro! Marcador não definido.**

Tabela 93- Comparação das subdimensões de questões validadas em relação aos níveis de complexidade da assistência – *Proxy M+N* (ciclo 2, estudo 5) - Visão de Usuários (Expectativa).**Erro! Marcador não definido.**

Tabela 94- Comparação das subdimensões de questões validadas em relação aos níveis de complexidade da assistência – *Proxy M+N* (ciclo 2, estudo 5) - Visão de Usuários (Percepção).**Erro! Marcador não definido.**

Tabela 95- Comparação das subdimensões de questões validadas em relação aos níveis de escolaridade de usuários – *Proxy M+N* (ciclo 2, estudo 5) - Visão de Usuários (Expectativa).**Erro! Marcador não definido.**

Tabela 96- Comparação das subdimensões de questões validadas em relação aos níveis de escolaridade de usuários – *Proxy M+N* (ciclo 2, estudo 5) - Visão de Usuários (Percepção).**Erro! Marcador não definido.**

Tabela 97- Comparação das macrodimensões de qualidade em relação aos níveis de atenção à saúde – *Proxy M+N* (ciclo 2, estudo 5) - Visão de Usuários (Expectativa, Percepção).**Erro! Marcador não definido.**

Tabela 98- Comparação das macrodimensões de qualidade em relação aos níveis de complexidade da assistência – *Proxy M+N* (ciclo 2, estudo 5) - Visão de Usuários (Expectativa, Percepção). **Erro! Marcador não definido.**

Tabela 99- Comparação das macrodimensões de questões validadas em relação aos níveis de escolaridade de usuários – *Proxy M+N* (ciclo 2, estudo 5) - Visão de Usuários (Expectativa, Percepção). **Erro! Marcador não definido.**

Tabela 100- Associação entre Expectativa e Percepção nas subdimensões de questões validadas – *Proxy M+N* (ciclo 2, estudo 5) - Visão de Usuários. **Erro! Marcador não definido.**

Tabela 101- Associação entre Expectativa e Percepção nas macrodimensões de questões validadas – *Proxy M+N* (ciclo 2, estudo 5) - Visão de Usuários. **Erro! Marcador não definido.**

Tabela 102- Correlações entre constructos e subdimensões de constructos – *Proxy J* (ciclo 2, estudo 4) (Visão de Usuários). **Erro! Marcador não definido.**

Tabela 103- Correlações entre as subdimensões e as macrodimensões de constructos – *Proxy J* (ciclo 2, estudo 4) (Visão de Usuários). **Erro! Marcador não definido.**

Tabela 104 - Comparação da Qualidade Global Percebida em relação às variáveis de caracterização selecionadas – Proxy J (ciclo 2, estudo 4) (Visão de Usuários). Erro! Marcador não definido.

Tabela 105- Correlações entre questões validadas e subdimensões de questões validadas (Expectativa) – *Proxy M+N*. **Erro! Marcador não definido.**

Tabela 106- Correlações entre as subdimensões e as macrodimensões questões validadas (Expectativa)
– *Proxy M+N*. **Erro! Marcador não definido.**

Tabela 107- Correlações entre questões validadas e subdimensões de questões validadas (Percepção) –
Proxy M+N. **Erro! Marcador não definido.**

Tabela 108- Correlações entre as subdimensões e as macrodimensões de questões validadas (Percepção)
– *Proxy M+N*. **Erro! Marcador não definido.**

Tabela 109- Comparação da Qualidade Global Percebida em relação às variáveis de caracterização
selecionadas – *Proxy M+N* (ciclo 2, estudo 5) -Visão de Usuários (Percepção)**Erro! Marcador não
definido.**

Tabela 110- Bibliometria da pesquisa integrativa de estudos relevantes de avaliação, monitoramento e
mensuração de performance, no âmbito mundial (países selecionados)**Erro! Marcador não
definido.**

SUMÁRIO

1. INTRODUÇÃO	35
1.1. Introdução	35
1.2. Enquadramento do tema	39
1.2.1. Contexto geral	41
1.2.2. Contexto nacional (brasileiro)	44
1.2.3. O problema	46
1.3. Objetivos do estudo	49
1.4. Estrutura da tese	50
PARTE I - CONCEITUAL	56
2. REVISÃO CONCEITUAL DA TEORIA DA AVALIAÇÃO E DAS BASES DA MEDIÇÃO DE PERFORMANCE NAS ORGANIZAÇÕES DE SAÚDE	57
2.1. A estruturação do campo de avaliação em saúde e o modelo teórico-lógico	57
2.2. A delimitação do campo da avaliação de performance do estudo	62
2.3. A evolução das metodologias de medição de performance em organizações de saúde	70
2.4. As metodologias e os instrumentos de medição de performance e apoio à decisão em organizações de saúde	74
2.4.1. As metodologias de avaliação da eficiência em organizações de saúde	75
2.4.2. As metodologias de avaliação de efetividade e ajustamento de risco em saúde	76
2.4.3. O enquadramento teórico-conceitual (<i>framework</i>) dos indicadores e dimensões de performance da investigação	81
2.4.4. O enquadramento teórico-conceitual (<i>framework</i>) da avaliação e medição da qualidade percebida em saúde e a escala SERVQUAL	111
3. EVOLUÇÃO RECENTE DA AVALIAÇÃO NO CONTEXTO DE GOVERNANÇA DA SAÚDE PÚBLICA DO BRASIL	126
3.1. Breve históricos da governança e da avaliação no sistema de saúde brasileiro	126
3.2. O arcabouço legal da avaliação e monitoramento no âmbito do Sistema Nacional de Saúde (SUS) do Brasil	135
3.3. As principais fontes de dados e informações do sistema de saúde do Brasil	138
4. FUNDAMENTOS E ESTRATÉGIAS DE ENVOLVIMENTO DOS DECISORES DE ORGANIZAÇÕES DE SAÚDE	145
4.1. A utilização da avaliação no contexto da tomada de decisão em sistemas e organizações de saúde	146

4.2.	A teoria organizacional subjacente ao envolvimento dos agentes de decisão (<i>stakeholders</i>) na avaliação participativa	150
4.3.	Estratégias mistas (qualitativa e quantitativa) de coleta de informações e construção de consensos entre decisores (<i>stakeholders</i>) em organizações de saúde	151
4.3.1.	A entrevista semiestruturada ou estruturada	154
4.3.2.	Os grupos focais ou grupos de especialistas	155
4.3.3.	A técnica de grupo nominal (TGN)	155
4.3.4.	A técnica <i>Delphi</i>	156
4.4.	Síntese teórico-conceitual da investigação	157
	PARTE II – EMPÍRICA	165
	5. METODOLOGIA DO ESTUDO	166
5.1.	O processo de recolha de dados	168
5.2.	Os modelos e as variáveis utilizadas no estudo	171
5.2.1.	O fluxo de aplicação dos instrumentos do painel de indicadores e dimensões de performance.	
	181	
5.2.2.	O fluxo de aplicação dos instrumentos da escala/satisfação de qualidade percebida.	186
5.3.	As hipóteses do estudo (itens de avaliação de performance e de avaliação de qualidade/satisfação percebida)	192
5.4.	Os modelos de análise e as técnicas de validação dos itens de avaliação da investigação	196
	6. RESULTADOS E DISCUSSÃO	213
6.1.	Os resultados dos painéis de indicadores e de dimensões de performance	213
6.1.1.	Análises descritivas dos estudos 1 a 3 (<i>proxies A a G</i>)	214
6.1.1.1.	Análise descritiva do estudo 1 (ciclos 1 e 2: <i>proxies A a C</i>) (domínios <i>versus</i> macrodomínios de avaliação) (Visão de especialistas, gestores e profissionais)	215
6.1.1.2.	Análise descritiva do estudo 2 (ciclo 1: <i>proxies D a F</i>) (indicadores <i>versus</i> dimensões de performance) (Visão de profissionais e gestores)	220
6.1.1.3.	Análise descritiva do estudo 3 (ciclo 2: <i>proxy G</i>) (indicadores <i>versus</i> dimensões de performance) (Visão de decisores)	225
6.1.2.	Análises inferenciais dos estudos 1 a 3 (<i>proxy C, proxy E+F e proxy G</i>)	231
6.1.2.1	Análise inferencial do estudo 1 (domínios <i>versus</i> macrodomínios de avaliação (ciclo 2: <i>proxy C</i>) (Visão de gestores e profissionais)	232
6.1.2.2	Análise inferencial do estudo 2 (indicadores <i>versus</i> dimensões de performance) (ciclo 1: <i>proxy E+F</i>) (Visão de gestores e profissionais)	235

6.1.2.3	Análise inferencial do estudo 3 (indicadores <i>versus</i> dimensões de performance) (ciclo 2: <i>proxy</i> G) (Visão de decisores)	240
6.2.	Os resultados dos itens de medição de qualidade/satisfação (escala de qualidade percebida)	245
6.2.1.	Análises descritivas dos estudos 4 e 5 (<i>proxies</i> H+I e J, <i>proxies</i> M e M+N)	246
6.2.1.1.	Análise descritiva do estudo 4 (ciclo 1: <i>proxies</i> H e I) (constructos <i>versus</i> subdimensões e macrodimensões de qualidade) (Visão profissionais <i>versus</i> usuários)	247
6.2.1.2.	Análise descritiva do estudo 4 (ciclo 2: <i>proxy</i> J) (constructos <i>versus</i> subdimensões e macrodimensões de qualidade) (Visão profissionais <i>versus</i> usuários)	256
6.2.1.3.	Análise descritiva do estudo 5 (ciclo 1: <i>proxies</i> K e L) (questões validadas <i>versus</i> subdimensões e macrodimensões de qualidade) (perspectiva de expectativa <i>versus</i> percepção).	263
6.2.1.4.	Análise descritiva do estudo 5 (ciclo 2: <i>proxies</i> M e N) (questões validadas <i>versus</i> subdimensões e macrodimensões de qualidade) (perspectiva de expectativa <i>versus</i> percepção)	273
6.2.2.	Análises inferenciais dos estudos 4 e 5 (<i>proxies</i> J e M+N)	281
6.2.2.1.	Análise inferencial do estudo 4 (ciclo 2: <i>proxy</i> J)	281
6.2.2.2.	Análise inferencial do estudo 5 (ciclo 2: <i>proxy</i> M+N)	288
6.3.	Validações e conclusões de hipóteses sobre itens de avaliação de performance	297
6.4.	Validações e conclusões de hipóteses sobre itens de medição de qualidade/satisfação (escala de qualidade percebida)	306
7.	CONCLUSÕES	317
7.1.	Descrição da abordagem	317
7.2.	Dificuldades encontradas	320
7.3.	Limitações do trabalho empírico	322
7.4.	Investigação futura	325
7.5.	Contributos desta investigação	326
	Referências Bibliográficas	338
	Anexos/Apêndices	Erro! Marcador não definido.
9.1	Os resultados complementares da validação dos painéis de indicadores e de dimensões de performance	Erro! Marcador não definido.
9.1.1.	Análises descritivas complementares dos estudos 1 a 3 (<i>proxies</i> A a G)	Erro! Marcador não definido.

9.1.1.1. Análise descritiva complementar do estudo 1 (ciclos 1 e 2: *proxies* A a C) (domínios *versus* macrodomínios de avaliação) (Visão de especialistas, gestores e profissionais) **Erro! Marcador não definido.**

9.1.1.2. Análise descritiva complementar do estudo 2 (ciclo 1: *proxies* D a F) (indicadores *versus* dimensões de performance) (Visão de profissionais e gestores) **Erro! Marcador não definido.**

9.1.1.3. Análise descritiva complementar do estudo 3 (ciclo 2: *proxy* G) (indicadores *versus* dimensões de performance) (Visão de decisores) **Erro! Marcador não definido.**

9.1.2. Análises inferenciais complementares dos estudos 1 a 3 (*proxy* C, *proxy* E+F e *proxy* G) **Erro! Marcador não definido.**

9.1.2.1. Análise inferencial complementar do estudo 1 (domínios *versus* macrodomínios de avaliação) (ciclo 2: *proxy* C) (Visão de gestores e profissionais) **Erro! Marcador não definido.**

9.1.2.2. Análise inferencial complementar do estudo 2 (indicadores *versus* dimensões de performance) (ciclo 1: *proxy* E+F) (Visão de gestores e profissionais) **Erro! Marcador não definido.**

9.1.2.3. Análise inferencial complementar do estudo 3 (indicadores *versus* dimensões de performance) (ciclo 2: *proxy* G) (Visão de decisores) **Erro! Marcador não definido.**

9.2. Os resultados complementares da validação dos itens de medição de qualidade/satisfação (escala de qualidade percebida) **Erro! Marcador não definido.**

9.2.1. Análises descritivas complementares dos estudos 4 e 5 (*proxies* H+I e J, *proxies* M e M+N) **Erro! Marcador não definido.**

9.2.1.1. Análise descritiva complementar do estudo 4 (ciclo 1: *proxies* H e I) (constructos *versus* subdimensões e macrodimensões de qualidade) (Visão profissionais *versus* usuários) **Erro! Marcador não definido.**

9.2.1.2. Análise descritiva complementar do estudo 4 (ciclo 2: *proxy* J) (constructos *versus* subdimensões e macrodimensões de qualidade) (Visão profissionais *versus* usuários) **Erro! Marcador não definido.**

9.2.1.3. Análise descritiva complementar do estudo 5 (ciclo 1: *proxies* K e L) (questões validadas *versus* subdimensões e macrodimensões de qualidade) (perspectiva de expectativa *versus* percepção) **Erro! Marcador não definido.**

9.2.1.4. Análise descritiva complementar do estudo 5 (ciclo 2: *proxies* M e N) (questões validadas *versus* subdimensões e macrodimensões de qualidade) (perspectiva de expectativa *versus* percepção) **Erro! Marcador não definido.**

9.2.2. Análises inferenciais complementares dos estudos 4 e 5 (*proxies* J e M+N) **Erro!**
Marcador não definido.

9.2.2.1. Análise inferencial complementar do estudo 4 (ciclo 2: *proxy* J)**Erro! Marcador não definido.**

9.2.2.2. Análise inferencial complementar do estudo 5 (ciclo 2: *proxy* M)**Erro! Marcador não definido.**

1. INTRODUÇÃO

1. Enquadramento da investigação

Esta investigação destaca a evolução recente das teorias e modelos de avaliação na área de saúde, quanto à estruturação do campo da avaliação para a tomada de decisão em sistemas e organizações de saúde. Estabelece uma visão global das tendências metodológicas da medição de performance em organizações e serviços de saúde, além de caracterizar os alguns programas e projetos de reconhecidas relevâncias na avaliação de organizações de saúde desenvolvidos no Brasil, em especial, no contexto de saúde pública, em busca de construir um quadro teórico-conceitual (*framework*) para a geração de indicadores de performance aplicável aos hospitais e serviços de saúde do SUS.

A etapa empírica desse propósito, centrou-se na finalidade de elaborar uma proposta metodológica e um *proxy* de painel de indicadores multidimensionais (produtividade assistencial, eficiência, efetividade, impacto, qualidade percebida etc), devidamente validados pelos agentes de decisão chaves, em uma amostra selecionada multicêntrica.

Em outro escopo dessa investigação abordaram-se os aspectos de medição de satisfação do usuário-utilizador de serviços de saúde do Brasil, contexto que há uma histórica escassez de iniciativas. Para fazer face à escassez de instrumentos padronizados e validados de mensuração de qualidade e satisfação, elaborou-se uma proposta de estratégia de mensuração de avaliação de satisfação dos usuários, através de uma escala inovadora validada por decisores chaves e usuários do SUS.

Na etapa empírica desse propósito adicional procedeu-se a tradução e a adaptação transcultural da escala SERVQUAL em um novo instrumento (escala de medição de qualidade/satisfação de usuários do SUS) fundamentados no campo da translação de conhecimentos no qual as teorias consolidadas são alvos de transliteração à luz de experiências dos atores sociais envolvidos (Barbosa & Neto, 2017; Clavier, Sénéchal, & Potvin, 2011; Hartz & John, 2008).

1.1. Introdução

O sistema de saúde brasileiro no último século apresentou uma evolução notável, ao nível estrutural e organizacional, apesar das demandas da sociedade serem crescentes. Desde a

criação do Sistema Único de Saúde (SUS)¹ no Brasil, muitos são os avanços do ponto de vista de garantias de atendimento, da organização sistêmica e da descentralização da gestão única da saúde, no entanto, pouco se observam avanços na governança do sistema de saúde.

No ano seguinte a redemocratização do Brasil, as diretrizes aprovadas na oitava Conferência Nacional de Saúde (VIII CNS²) (Brasil, 1986b) serviram de esteio para os artigos 196-200 da atual Constituição Brasileira de 1988 (Brasil, 1988), e apontaram para a reestruturação do Sistema Nacional de Saúde e fomentaram a criação do Sistema Único de Saúde (SUS) sob a visão de um Sistema Nacional de Saúde, que garante a saúde como direito do cidadão, sendo regularizado pela Lei Orgânica de Saúde nº. 8.080/1990(Brasil, 1990b), complementada pela Lei nº. 8.142/1990(Brasil, 1990c), ambas em vigor até os dias atuais, e Normas Operacionais complementares infralegais, que recentemente tiveram seus avanços organizacionais e de atenção à saúde condensados no chamado “Pacto Pela Saúde 2006”, e adendos posteriores, para organizar a assistência à saúde brasileira.

Em paralelo a esta evolução a Nova Gestão Pública (NGP)³, ou Gerencialismo, enquanto tendência gerencial recente tem exigido dos gestores públicos mundiais, esforços de monitoramento de resultados em diversas áreas para a controladoria e responsabilização (*accountability*) dos gastos e recursos públicos.

O elevado volume de recursos aplicados na saúde, por si, exige avanços no monitoramento de resultados. Dados relativos a 2016 apontam que 35 países da OCDE⁴ gastam em torno de 9,0% do PIB⁵ (GDP) no sistema de saúde, nos Estados Unidos cerca de 17,2%, em Portugal 8,9%, e no Brasil 6,2% (OECD, 2017), Além disso, os gastos com a saúde têm crescido nos últimos anos (2003-2016) em média de 3,6% nestes países (OECD, 2017), e países como Estados Unidos e Brasil registram inflações no setor saúde muito acima da média mundial, grande parte pela pressão de incorporação de novas tecnologias em saúde, e ampliação do volume de cuidado demandado pela população.

Um estudo recente do Banco Mundial aponta, no âmbito da OCDE e para alguns países emergentes, o rápido crescimento de despesas de saúde no setor privado, sendo que no Brasil que conta com o SUS, sistema nacional de saúde brasileiro, público, gratuito e de acesso universal, já se gasta mais recursos com o setor privado (54,26%) que com o sistema público

1 SUS - Sistema Único de Saúde, Sistema Nacional de Saúde (SNS) do Brasil

2CNS - Conferência Nacional de Saúde

3 NGP – Nova Gestão Pública (NPM - New Public Management)

4OCDE - Organização para a Cooperação e o Desenvolvimento Econômico (OECD), países da Europa

5PIB – Produto Interno Bruto (GDP) somatório de todas as riquezas geradas em um país

(45,74%), enquanto a média de gastos de 25 países europeus situa-se em torno de 77% no sistema público de saúde e o Canadá aplica mais de 70%, ou seja, média muito superior de recursos públicos face aos recursos da saúde privada (quadro 1).

Países Avaliados	GDP-PIB per capita (US\$ PPP)	Valor Despesa per capita (US\$ PPP)	Despesa privada (%GDP-PIB)	Despesa pública (%GDP-PIB)	Despesa pública vs % Gasto Total Saúde	Gastos Total Saúde (% GDP-PIB)
Brasil	10.278,63	1.042,73	4,83	4,07	45,74	8,90
Argentina	10.941,96	1.473,70	3,19	4,92	60,64	8,11
China	7.417,89	432,28	2,28	2,89	55,89	5,16
Espanha	26.952,35	3.040,78	2,49	6,95	73,59	9,44
Reino Unido	32.808,79	3.321,67	1,61	7,71	82,70	9,32
Austria	34.852,66	3.691,55	2,84	6,19	68,51	9,03
França	29.820,09	4.085,48	2,70	8,92	76,74	11,63
EUR-25	28.048,60	3.259,80	2,33	7,82	77,04	10,15
Canadá	35.714,79	4.519,96	3,31	7,87	70,41	11,18

Quadro 1 - Gastos em saúde: percentual de despesas pública *versus* privada, países OCDE, Brasil e outros (2013). Fonte: Adaptado pelo autor – Dados disponíveis em <http://data.worldbank.org/data-catalog/world-development-indicators>.

Um dos maiores desafios de gestão em sistemas de saúde, em países em desenvolvimento como o Brasil, é a presença de uma tripla carga de doença (Schramm et al., 2004) pelo recrudescimento de doenças infectocontagiosas; epidemia de doenças oriundas da violência e condições de vidas urbanas; e avanço de doenças crônicas-degenerativas, em virtude da inversão da pirâmide etária, com tendência expressa pelos gráficos 3 e 4.

Outro exemplo das dificuldades de governança do SUS é que apenas recentemente o Ministério da Saúde concluiu o primeiro levantamento das filas nacionais de cirurgias eletivas que registrou uma lista de espera de 667.014 pacientes (Frasão, 2017) apesar de anualmente aplicar recursos elevados na área.

Neste contexto de elevação de gastos e conforme a visão da NGP ou Gerencialismo, que exige a responsabilização dos gestores públicos por gastos governamentais, surge a necessidade de avaliação e monitoramento de resultados do sistema de saúde. Em resposta, pode-se citar

exemplos de projetos de avaliação e monitoramento de resultados da administração pública como PRODEV⁶ e P4R (PforR)⁷, que têm sido implantados na gestão pública de países em desenvolvimento, inclusive no Brasil, através de agências internacionais de fomento em busca de otimizar os recursos e garantir qualidade às ações públicas.

Nos primeiros momentos as prioridades do SUS foram a organização da atenção à saúde, a garantia de acesso aos serviços do SUS, a organização das instâncias de gestão nas três esferas governamentais.

Na área de saúde do Brasil, os esforços de governança têm sido maioritariamente focados nas políticas de saúde e sistemas coletivos de saúde, como os programas de atenção primária, de vacinação, de controle de mortalidade infantil e materna, dentre outros fatores sistêmicos. Por outro lado, a governança dos sistemas curativos e reabilitadores, na maioria das vezes, limita-se a implantação de novas estruturas de atendimentos e quase nunca em gestão da performance destas organizações e serviços de saúde.

No entanto, a mídia brasileira tem mostrado que pouco se tem avançado na gestão de resultados em saúde, por exemplo, na evolução das taxas de mortalidade (Barreto, 2016), que apresenta drástica elevação das taxas que representam mortes a cada cem mil pessoas e são enviadas ao Ministério da Saúde por estados e municípios. Neste caso, o primeiro desafio é equalizar os dados, pois os sistemas do Datasus comportam duas séries históricas: uma que vai de 1984 a 2007; e outra após o ano de 2008, quando houve revisão geral de códigos dos registros de eventos e procedimentos do SUS, que impactam qualquer estudo que se pretende efetivar – ressalte-se que as limitações quanto à fiabilidade nos dados do Datasus da rede de atenção à saúde será objeto posterior de análise. O outro é estabelecer um sistema de monitoramento e regulação de índices voltados para a rede de atenção à saúde, que subsidie a governança do SUS, escopo deste estudo.

Ampliou-se a análise dessa tendência da mortalidade, através da análise de dados das duas últimas décadas nos estados que no último quadriênio tiveram maiores taxas de mortalidade, além da média por região do Brasil (figura 1).

6 PRODEV - Programa de Estratégia e Plano de Ação para a Efetividade do Desenvolvimento do BID para os países da América Latina e Caribe

7 PforR ou P4R - Programa para Resultados do Banco Mundial para o apoio ao desenvolvimento da “*accountability*” de países em desenvolvimento

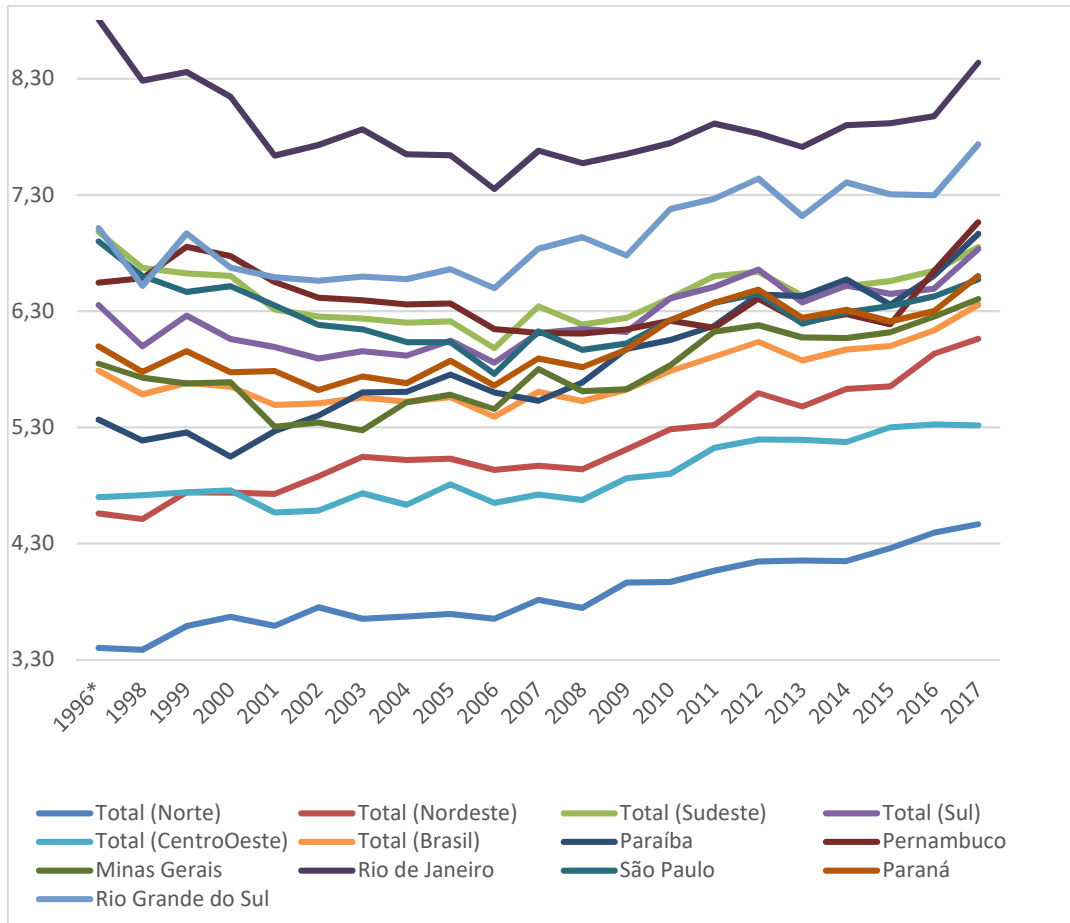


Figura 1 - Evolução da Taxa Bruta de Mortalidade por mil habitantes/ano (Brasil, 1996 a 2017) - grandes regiões e estados selecionados (> média do Brasil, no último quadriênio). Fonte: Elaboração própria - Dados disponíveis em: <http://datasus.saude.gov.br/informacoes-de-saude/tabnet/estatisticas-vitais>.

Nesse contexto, a regulação do sistema público de saúde do Brasil pauta-se basicamente na gestão da demanda e na garantia do acesso e na organização dos fluxos assistenciais, sem uma preocupação de desempenho nos pontos de atenção à saúde. Com isto, agrava-se a crise de recursos e de financiamento, não havendo transparência nos resultados de hospitais e demais serviços de saúde, quanto ao seu desempenho e resultados alcançados. Assim, há um dilema recorrente no SUS: há insuficiência de recursos ou ineficiência na gestão? Sem uma metodologia de avaliação e monitoramento de desempenho dificilmente pode-se enfrentar este dilema e avançar com as conquistas sociais do SUS.

Dessa forma, pode-se supor que há relevância no estabelecimento de mecanismos e instrumentos de mensuração de performance e monitoramento de resultados de políticas, programas e ações de saúde no âmbito do SUS, principal escopo dessa investigação.

1.2. Enquadramento do tema

A fim de delimitar o campo desta investigação e permitir a estruturação de um enquadramento teórico conceitual (*framework*), contextualiza-se o espaço funcional considerando o constructo de “serviços”, o que permitirá cingir o campo onde se processam as inter-relações, sendo base para a elaboração de modelos teórico-lógicos de avaliação em saúde, inclusive quanto aos instrumentos de avaliação da satisfação/insatisfação no âmbito dos serviços oferecidos.

Ressalta-se que as organizações e serviços de saúde, tal como se apresentam no SUS, constituem-se como serviços públicos de acesso geral, ou unidades credenciadas e, na medida em que são de acessos universais, contam com um grau elevado de utilização. Logo, o envolvimento dos usuários é grande e o seu autoconceito importante.

Na literatura não há um consenso sobre o escopo e a amplitude da avaliação e do monitoramento. Nesta investigação escolheu-se trabalhar com a visão integrada do monitoramento e avaliação, defendida pela *United Nations Development Programme* (UNDP). Nesta visão, o processo de monitoramento pode gerar questões a serem respondidas pela avaliação, que se situa como um processo de análises aprofundadas de dados e situações, pertinentes para a tomada de decisão e gestão baseada em resultados (UNDP, 2009).

Assume-se, ainda, o escopo instrumental do monitoramento de resultados definido no âmbito da OECD, que trata o monitoramento como a função contínua que usa a recolha sistemática de dados, para fornecer à gestão e aos decisores (*stakeholders*) parâmetros decisores sobre uma intervenção em curso, através de indicadores específicos sobre a extensão do progresso, do atingimento dos objetivos e dos progressos alcançados na utilização de recursos alocados (OECD, 2002; World Bank, 2009).

Esta perspectiva instrumental entende o monitoramento como o processo que fornece informações em tempo real exigidas, para a gestão baseada em resultados e, por sua vez, a avaliação fortemente baseada em dados gerados durante o monitoramento do programa ou do ciclo do projeto, incluindo, por exemplo, dados de base, informações sobre o processo e as medidas de implementação do programa ou projeto. Nesta visão, o processo de monitoramento pode gerar questões a serem respondidas por uma avaliação, que se situa como um processo de análises aprofundadas de dados e situações, pertinentes para a tomada de decisão e gestão baseada em resultados.

Por fim, o Artefato validado nesta investigação pretende subsidiar um sistema de monitoramento e avaliação de performance de hospitais e unidades de saúde, no âmbito do SUS, enquanto um sistema de Monitoramento e Avaliação (M&A) baseado em resultados que

fornece informações cruciais sobre políticas públicas ou desempenho organizacional - *results-based management* (RBM) -, como conceituam Morra-Imas e Rist (2009). Numa ampliação de propósito e utilização, nada impede que os decisores possam se apropriar dos Artefatos gerados enquanto mobilizadores da gestão que ultrapasse o alcance de resultados para priorizar o desenvolvimento institucional, ou seja, com a perspectiva ampliada: “*managing for development results* (MfDR)”(UNDP, 2009).

Nortearia este trabalho a tendência emergente na gestão de resultados em saúde quanto a avaliação, que sugere que “a mudança central do foco de atenção foi das atividades ou “produtos” das agências governamentais (output) para os resultados dessas atividades (*outcomes*)” (Kettl, 1996; Almeida, 1999 apud F Viacava et al., 2004: 712). Na parte de análise de qualidade percebida, segue a tendência metodológica de valorizar os resultados e impactos referidos pelos pacientes (PROMs - *Patient Reported Outcome Measures*, ou seja, a mensuração de desfechos reportados pelo paciente), sendo esta mais uma das bases desta investigação.

Em busca de explorar o tema para construir uma base científica de suporte à investigação procedeu-se uma ampla revisão integrativa da literatura recente (Jan/2005 a Dez/2016) (tabela 110, anexo 1). Assim, em linhas gerais explorou-se o contexto geral do campo de avaliação e mensuração de performance em sistemas e serviços de saúde, abordou-se as experiências e programas do Brasil de avaliação em saúde, além do estudo das bases de medição de qualidade e satisfação de usuários, caracterizados globalmente a seguir:

1.2.1. Contexto geral

De acordo com Travassos et al. (1999), é necessário destacar a importância do desenvolvimento de estratégias para permitir que os resultados das avaliações melhorem a efetividade dos cuidados hospitalares. Assim, o monitoramento de indicadores pode promover a efetividade da assistência, bem como da busca pela eficiência organizacional em saúde. A monitorização dos indicadores constitui também uma boa prática de governança hospitalar (Raposo, 2007) uma vez que reforçam a transparência e a prestação de contas/responsabilização (*accountability*). Entretanto, a multiplicidade de perspectivas e de complexidades envolvidas, além das diversidades de escopos e definições demandam a análise dos principais quadros teóricos e metodológicos de avaliação relevantes da literatura.

Na literatura há diversas propostas de quadro teórico-Conceitual (*framework*) para avaliação de performance e a seleção de indicadores, por exemplo, painéis de dimensões de

avaliação e indicadores de performance usados para a medição da qualidade de alguns países e organizações multilaterais (Reino Unido, Canadá, Austrália, EUA, OMS e OCDE) citados como relevantes em outro importante estudo (Arah, Klazinga, Delnoij, Asbroek, & Custers, 2003).

Ao nível mundial pode-se dizer que os primeiros esforços estruturados de medição de performance direcionados às organizações hospitalares e serviços de saúde foram iniciados na década de 1980. Após este período, surgiram relevantes sistemas de avaliação em saúde, que este estudo caracterizou enquanto as principais metodologias e sistemas mundiais de medição de resultados em saúde, como por exemplo, quadros teóricos-conceptuais (*frameworks*) de alguns projetos em destaque:

- *Quality Indicator Project (QIP)*⁸ da Agência de Pesquisa e Qualidade nos Serviços de Saúde – *U.S. Agency for Healthcare Research and Quality (AHRQ)*⁹;
- Projeto da Fundação Europeia para Gerenciamento da Qualidade - *European Foundation for Quality Management (EFQM)*¹⁰;
- Sistema de indicadores do *Project Assessment Tool of Quality Improvement in Hospitals (PATH)*¹¹ da Organização Mundial de Saúde (OMS)¹²;
- No Reino Unido, duas abordagens. A primeira baseada na definição de indicadores do Serviço Nacional de Saúde-NHS¹³, a segunda interessada na concretização de uma metodologia para estabelecer “*hospital ratings*”;
- Em Portugal, o Sinas@Hospitais¹⁴ sob a coordenação da Entidade Reguladora de Saúde (ERS)¹⁵, que classifica os hospitais avaliados (ranking) que é utilizado pela Administração Central do Sistema de Saúde (ACSS)¹⁶, entre outros que surgirem como relevantes na revisão da literatura (anexo 2).

No propósito de integrar os quadros conceituais (*frameworks*) mais destacados, investigou-se na literatura relevante (quadros 17 a 19, anexo 2) a importância de cada uma dos indicadores e dimensões de performance, bem como, buscou-se identificar quais as

8QIP – *Quality Indicator Project* (Projeto de Indicadores de Qualidade)

9AHRQ – *U.S. Agency for Healthcare Research and Quality* (Agência de Pesquisa e Qualidade nos Serviços de Saúde)

10EFQM - *European Foundation for Quality Management* (Fundação Europeia para Gerenciamento da Qualidade)

11PATH – *Performance Assessment Tool for Quality Improvement in Hospitals* (ferramenta de avaliação de performance para a melhoria da qualidade hospitalar)

12 OMS- Organização Mundial de Saúde(WHO), representada no Brasil, América Latina, e Caribe pela OPAS-Organização Pan-Americana de Saúde.

13 NHS- *National Health System* (Serviço Nacional de Saúde do Reino Unido)

14Sinas@Hospitais – Sistema de Indicadores de Performance de Hospitais do Serviço Nacional de Saúde de Portugal

15 ERS-Entidade Reguladora de Saúde que operacionaliza a regulação do sistema público e privado do SNS de Portugal

16ACSS - Administração Central do Sistema Nacional de Saúde (SNS) de Portugal

metodologias e instrumentos de mensuração de performance em saúde mais utilizadas nos contextos mundial e brasileiros.

No outro escopo desta investigação, estudou-se destaques na literatura quanto as bases de avaliação de qualidade percebida para a estruturação de escalas de medição de qualidade e satisfação de usuários de serviços de saúde.

Apesar do relativo consenso que a qualidade é vista como um dos aspectos centrais a ser considerado para a avaliação em saúde (Donabedian, 1980c; H. V. Vuori, 1982), não existe uma definição universalmente aceita de cuidado em saúde, qualidade ou qualidade de atendimento. No nível individual, Campbell et al. (2000: 1621) propõe a qualidade como sendo a capacidade de acessar o atendimento eficaz com o objetivo de maximizar os benefícios à saúde em relação à necessidade.

No escopo da saúde, uma ampla revisão da literatura mundial aponta que a satisfação do usuário é multifacetada e de difícil mensuração (M. Esperidião & Trad, 2005; M. A. Esperidião, 2009), sendo de interesse maior para os decisores de saúde (Aharony & Strasser, 1993 apud Esperidião e Trad, 2005: 304).

Por sua vez, Vinagre (2008) enfatiza a complexidade de aferição dos serviços pelos aspectos da *intangibilidade* (ausência de atributos físicos) dos serviços (Parasuraman, Berry, & Zeithaml, 1991b) e o fato dos serviços constituírem-se essencialmente enquanto processos, sendo afetados por interações e atividades (Grönroos, 1990), ou seja, dependente de aspectos relacionais entre usuários e prestadores de serviços.

A partir de uma análise integrativa realizada por Giese e Cote (2000 apud Vinagre, 2008) podem-se sintetizar os achados de alguns estudos sobre as bases da avaliação da satisfação, comparando as definições conceituais, as respostas, o enfoque e o momento da análise da satisfação (quadro 20, anexo 2).

Diante das muitas definições do constructo satisfação e dos inúmeros quadros conceituais (*frameworks*) propostos na literatura para a avaliação da satisfação, resolveu-se considerar, neste trabalho, a satisfação enquanto resposta a um processo de avaliação da experiência de consumo, considerando tanto o resultado quanto o processo como elementos importantes para a satisfação dos consumidores com enfoque nos antecedentes/preditores da satisfação, ou seja, amparado pela teoria da atribuição causal (Bitner, Booms, & Mohr, 1994).

Para além disso, elegeu-se neste estudo o escopo de experiência de serviços, visto que em serviços de saúde a experiência de serviços parece ser muito sensível aos constructos expectativas, emoções e experiências, que são constructos que estão na base do enquadramento teórico da linha experiência de serviços para mensuração da satisfação de usuários e que

resultou na escala SERVQUAL (Parasuraman, Berry, & Zeithaml, 1988; Parasuraman et al., 1991b), que será adaptada nesta investigação para o contexto do SUS.

1.2.2. Contexto nacional (brasileiro)

Uma revisão sistemática (Machado, Martins, & Martins, 2013) voltada para avaliação de qualidade hospitalar mostra que os primeiros estudos surgiram entre 1991 a 2003, tendo encontrado apenas 48 publicações (quadro 26, anexo 4). Em outro estudo sobre a avaliação em saúde no Brasil (Fernandes, Ribeiro, & Moreira, 2011) acharam apenas 44 estudos.

Neste propósito, procedeu-se uma revisão sistemática ampliada nas principais bases publicações de artigos científicos do Brasil (SciELO e Lilacs e Medline, via Biblioteca Virtual em Saúde - BVS) tendo localizado 112 publicações no período de 1990 a 2018 (figura 2). Assim, buscou-se mapear e descrever qualquer iniciativa inovadora de monitoramento e avaliação, com similaridade de objeto e escopo desta investigação que seja considerada de relevante aplicação no contexto das organizações de saúde do Brasil, a fim de avaliar sua aplicabilidade no SUS (quadro 27, anexo 4).

A incipiência de publicações em avaliação e monitoramento de sistemas e serviços de saúde no Brasil comprovada nesta bibliometria realça, ainda mais, a importância do objeto de análise desta investigação.

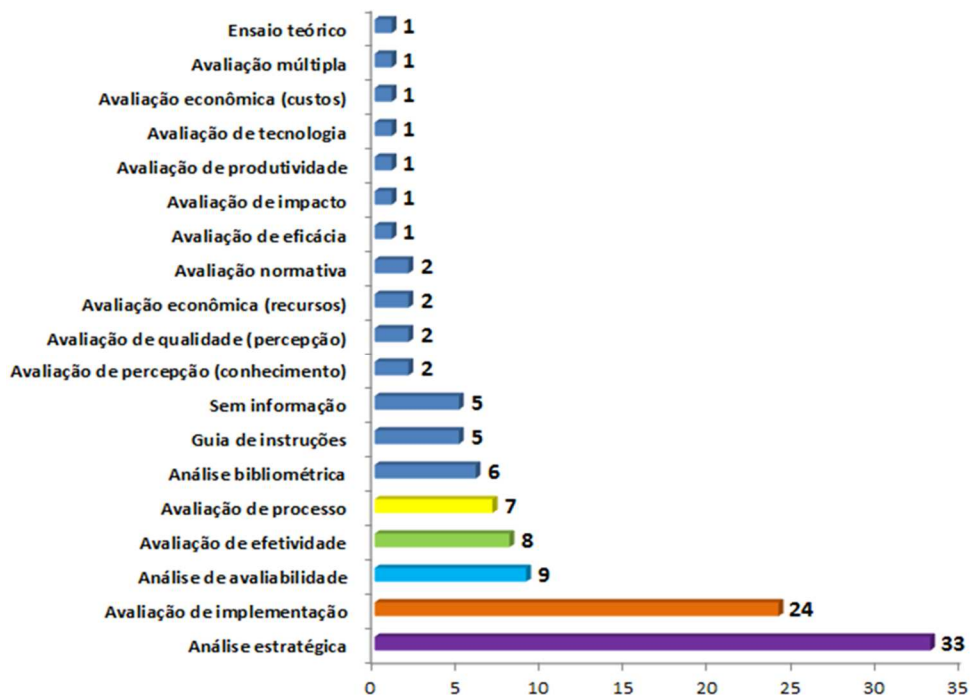


Figura 2 - Bibliometria de estudos de avaliação em saúde no Brasil (1990 - 2018).

No propósito de aproximar-se da realidade do objeto de avaliação deste estudo, caracterizam-se alguns programas e projetos de reconhecidas relevâncias na avaliação de organizações de saúde desenvolvidos no Brasil, no contexto de saúde pública e iniciativas privadas de ampla aplicação. Assim, buscou-se mapear e descrever qualquer iniciativa inovadora de monitoramento e avaliação, com similaridade de objeto e escopo desta investigação que seja considerada de relevante aplicação no contexto das organizações de saúde do Brasil, a fim de avaliar sua aplicabilidade no SUS.

De forma complementar, este trabalho examinou, no contexto da saúde brasileira, os sistemas informatizados, indicadores e instrumentos gerenciais, no campo de saúde pública, tais como: indicadores do DATASUS¹⁷, Projeto Desenvolvimento de Metodologia de Avaliação do Desempenho do Sistema de Saúde Brasileiro (PROADESS)¹⁸, indicadores do Índice de Desempenho do SUS (IDSUS)¹⁹, Indicadores e Dados Básicos da Saúde (IDB)²⁰ da Rede Interagencial de Informações para a Saúde (RIPSA)²¹ e, indicadores do novo PNASS; no âmbito da Agência Nacional de Saúde Suplementar (ANS)²², Índices de Desempenho da Saúde Suplementar (IDSS)²³ e indicadores QUALISS²⁴ do Programa de Qualificação dos Prestadores de Serviços de Saúde Suplementar (QUALISS) – em fase de estruturação e; nas organizações privadas: indicadores PROHASA/CQH²⁵ e indicadores SIPAGEH²⁶, dentre outros que surgirem na revisão da revisão integrativa da literatura (anexos 5 e 6).

No outro propósito desta investigação, identificou-se no Brasil, duas pesquisas de opinião de base nacional do Conselho Federal de Medicina (CFM) e da Confederação Nacional da Indústria (CNI), com os mais renomados institutos brasileiros de pesquisa, Datafolha e Ibope, respectivamente, com dados de 2011 a 2018, que revelam insatisfações da população que variam de 54 a 93% (CFM, 2014, 2015, 2018). E, ainda, que elevados percentuais de brasileiros avaliam a saúde pública como ruim ou péssima, que aumentou de 61%, em 2011, para 75% em 2018 (CNI, 2012, 2018).

17DATASUS - Conjunto de módulos do sistemas de informação de dados do sistema nacional de saúde do Brasil (SNS-SUS), inclusive de hospitais e serviços de saúde

18PROADESS - Projeto Desenvolvimento de Metodologia de Avaliação do Desempenho do Sistema de Saúde Brasileiro

19 IDSUS – Matriz de indicadores em 18 macrodimensões gerenciada pelo DATASUS e o Ministério da Saúde do Brasil

20 IDB - Indicadores e Dados Básicos para a Saúde no Brasil gerenciados pela RIPSA e o Ministério da Saúde do Brasil

21 RIPSA - Rede Interagencial de Informações para a Saúde

22ANS - Agência Nacional de saúde Suplementar, criada pela Lei Federal 9661/00 para regular as atividades privadas de saúde, através de suas resoluções, pautadas na Lei 9656/98 (“operadoras de saúde”)

23 IDSS - Índices de Desempenho da Saúde Suplementar

24Indicadores QUALISS, implantados pelo Programa de Qualificação dos Prestadores de Serviços de Saúde Suplementar

25PROHASA/CQH – Indicadores do Programa de Estudos Avançados em Adm.Hospitalar e de Sistemas de Saúde da FMUSP

26 SIPAGEH – Sistema de Indicadores Padronizados para Gestão Hospitalar, com adesão voluntária de 37 hospitais do Sul do Brasil

Em uma pesquisa sistemática efetivada nas principais bases de artigos científicos do Brasil (Medline, Lilacs e Scielo), além de ter obtido resultados extremamente escassos, não se identificou nenhuma metodologia robusta de avaliação de qualidade e satisfação através de instrumentos padronizados e validados e/ou de larga utilização no contexto de saúde do Brasil (figura 2). Esses achados estão alinhados com os resultados de uma revisão sistemática extensiva, entre 1990 e 2013, nas principais bases bibliográficas de publicação brasileiras em saúde pública, na qual Volpato (2014) identificou poucas publicações com características voltadas à percepção dos usuários e/ou dos profissionais referente a qualidade dos serviços públicos de saúde, também não tendo identificado instrumentos ou escalas validadas de mensuração de satisfação e/ou qualidade em saúde no Brasil.

1.2.3. O problema

O campo de avaliação em saúde, por suas complexidades, subjetividades e múltiplas perspectivas, engloba uma diversidade de termos, conceitos e métodos o que está coerente com a multiplicidade de questões consideradas como pertinentes na área da saúde, na heterogeneidade e complexidade das intervenções, sejam elas ações, serviços, programas ou políticas públicas.

Por sua vez Minayo et al. (2010) afirmam que os investimentos maciços em políticas públicas, no bojo do Estado do bem-estar social exigiram a adoção da investigação avaliativa que devem priorizar aspectos como utilidade, viabilidade, precisão técnica, ética e métodos e técnicas de análise de dados adequados.

Alkin (2004 apud Samico et al., 2010) estabelece as origens do campo da avaliação a partir de duas necessidades: a prestação de contas (*accountability*) e o controle dos programas.

Por sua vez, Chianca e Youker (2004) registram um crescente interesse no campo de avaliação de sistemas de saúde da América Latina e Caribe. No entanto, Bosi e Mercado (2006) elaboram uma extensa revisão de literatura quanto às experiências de avaliação nos países da América Latina e concluem que a maioria dos trabalhos é de índole teórica, sendo poucos os estudos empíricos ou de caso. Citam ainda que quase sempre os enfoques avaliativos usados inspiram-se em certos autores e obras eurocêntricas (Fals Borda & Mora-Osejo, 2004), do mundo anglo-saxão, em especial, provenientes das teorias e modelos emergentes da península ibérica (Bosi & Mercado, 2006).

Em outro prisma, Hartz e Vieira-da-Silva (2005), abordam a avaliação enquanto desenhos de modelos teóricos-lógicos para a compreensão das relações de lógica entre os objetivos, atividades e resultados do programa. Assim, o objeto da investigação científica é um constructo da realidade passível de modelagem, instrumentalizado em dois momentos: a idealização do objeto-modelo ou modelo conceitual e a construção da teoria do objeto-modelo ou, em outros termos, na elaboração do modelo teórico (Bunge, 1974). O modelo lógico (ML) é entendido enquanto um esquema visual que apresenta como um programa deve ser implementado e que resultados são esperados (Rowan, 2000).

Enquanto isso, Samico et al. (2010), afirmam que o ML pode incluir todos ou alguns dos seguintes elementos: objetivos e metas, insumos, atividades, produtos, público-alvo e resultados (imediatos, intermediários e de longo prazo). Estes autores propõem que a elaboração do ML deve contar com a participação ativa dos responsáveis e envolvidos no programa, e que o processo de validação do ML deve envolver as partes interessadas (*stakeholders*), como os profissionais responsáveis por sua implementação nos serviços de saúde e os usuários dos programas, ou ainda, deve consultar comitês de especialistas, como forma de identificar lacunas, sobreposições ou esclarecer dúvidas sobre o ML construído (Souza, Vieira-Da-Silva, & Hartz, 2005).

Vale salientar que Carvalho (2007) afirma que importa não só identificar os atores principais (*stakeholders*), mas, também, caracterizá-los e agrupá-los em função da sua importância considerada e do impacto que poderão ter ao nível do comportamento e dos resultados alcançados pela organização. Dessa forma, para o escopo deste trabalho, adotou-se a perspectiva da teoria dos *stakeholders* e considerou-se como decisores chaves (*key stakeholders*) os gestores dos níveis setoriais, organizacionais e governacionais, no processo de estruturação de modelos para mensuração da performance organizacional. Diante disso, esta investigação firma-se no enquadramento conceitual do campo da avaliação participativa e construtivista de 4ª. geração proposta por Guba e Lincoln (2011).

Em seguida, em uma breve análise dos aspectos filosóficos da avaliação, escolheu-se trabalhar sob as bases da ontologia idealista, na qual os modelos e representações que formam o nível empírico são contingentes e socialmente produzidos - neste caso, pelos agentes de decisão chaves (Bosi & Mercado, 2006).

Sendo assim, nos aspectos metodológicos, assume-se nesta investigação um processo integrativo e participativo da avaliação, em busca de construir consensos quanto a visão global, metodologias, macrodimensões e indicadores de mensuração de performance organizacional,

em especial, no âmbito de agentes de decisão chaves de unidades de saúde pública e/ou vinculadas ao SUS.

Por sua vez, Guba e Lincoln (2011) pressupõem a avaliação enquanto um processo de negociação entre os atores envolvidos na intervenção a ser avaliada. Diante disto, uma das primeiras etapas de aproximação ao objeto sob avaliação consta do mapeamento das partes interessadas (*stakeholders*) na avaliação, em especial, os agentes de decisão, que contribuem ativamente na construção e no desenho do modelo de avaliação da investigação.

A inclusão dos diversos atores na decisão da modelagem da avaliação se faz necessário também pela variedade de dimensões mensuráveis nos serviços de saúde, pelo que há uma necessidade de priorização dos objetivos e instrumentos desenvolvidos pelos seus futuros usuários-utilizadores. Neste quesito, Uchimura e Bosi (2002) referem autores e listas das possíveis dimensões de avaliação:

Gattinara et al. (1995) sinalizam vários fatores que determinam a qualidade dos serviços de saúde: competência profissional; [...] satisfação dos clientes; [...] acessibilidade; [...] eficácia; [...] eficiência [...].

Por outro lado, Vuori (1991; também citado por Acurcio et al., 1991; Akerman & Nadanovsky, 1992; Santos, 1995) citam outros fatores: [...] efetividade; [...] eficácia; [...] eficiência; [...]

Assim, pela complexidade do fenômeno em estudo e a subjetividade do processo avaliativo dos sistemas de saúde, cabe ao investigador, em parceria com os decisores, selecionar as dimensões que mais apresentam relevância na realidade na tomada de decisão em organizações de saúde, e assim, direcionar a construção da metodologia de avaliação de performance proposta nesta pesquisa.

Nesse contexto colaborativo, no intuito de elaborar uma proposta de quadro teórico-conceitual (*framework*) para a seleção de indicadores mais relevantes foram estudados alguns dos principais sistemas mundiais de avaliação de performance em organizações de saúde, como por exemplo, quadros conceptuais dos projetos de indicadores de performance e qualidade internacional em destaque de países da Europa, EUA e/ou Canadá. Em seguida, no propósito de aproximar-se da realidade do objeto de avaliação deste estudo, caracterizam-se alguns programas e projetos de reconhecidas relevâncias na avaliação de organizações de saúde desenvolvidos no Brasil, em especial, no contexto de saúde pública (tópico 2.4.3). Os principais fundamentos destes quadros teóricos-conceituais (*frameworks*) estruturaram as bases para a proposição do *proxy* de painel de indicadores de performance, validada por decisores chaves (tópico 6.1).

No Aspecto de medição de satisfação do usuário-utilizador de serviços de saúde do Brasil, há uma histórica escassez de iniciativas neste campo evidenciada na pesquisa bibliográfica (quadro 21, anexo 2). Para fazer face à escassez de instrumentos padronizados e validados de mensuração de qualidade e satisfação, caracterizou-se alguns dos principais fundamentos de mensuração de qualidade e satisfação aplicáveis aos usuários de saúde (tópico 2.4.4), em especial, sobre a escala SERVQUAL, enquanto base para a elaboração de uma proposta de estratégia de mensuração de avaliação de satisfação dos usuários-utilizadores, pautada em princípios científicos, validada por decisores chaves e usuários-utilizadores do SUS (tópico 6.2).

Dessa forma, a proposta de tradução e adaptação transcultural da escala SERVQUAL em um novo instrumento (escala de medição de qualidade/satisfação de usuários do SUS) fundamenta-se no campo da translação de conhecimentos no qual as teorias consolidadas são alvos de transliteração à luz de experiências dos atores sociais envolvidos (Barbosa & Neto, 2017; Clavier et al., 2011; Hartz & John, 2008), visando à melhoria de serviços em prol de preenchimentos de lacunas de saberes e instrumentos para transformar políticas e práticas (Barbosa & Neto, 2017; CIHR, 2004; WHO, 2006) para a gestão de serviços do SUS.

Diante desses argumentos e contextos explanados, surge a problematização central desta investigação: um *proxy* de modelo de monitoramento e avaliação multidimensional de performance e resultados, construído em colaboração com os decisores chaves, tem relevâncias/importâncias para os diferentes níveis de gestores, na governança das organizações de saúde do SUS, em uma amostra de hospitais e serviços de saúde pública no Brasil? No escopo adicional deste estudo surge a segunda questão central: quais os níveis de adaptação, relevância e validação de um grupo de questões em uma escala inovadora de mensuração de qualidade percebida por especialistas e usuários do SUS?

Destacam-se outras questões transversais, exploradas no percurso desta investigação: quais as experiências relevantes de sistemas de monitoramento e avaliação e/ou de gestão de indicadores de performance no contexto dos hospitais e serviços de saúde do SUS? Quais as bases normativas e de sistemas de dados capazes de fundamentar um sistema de monitoramento e avaliação no âmbito do SUS?

1.3. Objetivos do estudo

A fim de orientar esta investigação, a partir da problematização traçada definiram-se alguns objetivos. Assim, o Objetivo geral da investigação é estruturar e validar um *proxy* de um

painel de indicadores e dimensões de performance, além de uma escala de qualidade percebida em saúde, por decisão multicritério dos principais agentes decisores, aplicado ao monitoramento de resultados e à gestão estratégica das organizações de saúde, em unidades selecionadas no sistema público de saúde do Brasil. Enquanto objetivos específicos foram identificados os seguintes:

- descrever os princípios norteadores de avaliação em sistemas e serviços de saúde a fim de estabelecer um enquadramento metodológico-Conceitual sobre os aspectos relevantes da avaliação de resultados e mensuração de performance em organizações e serviços de saúde;
- investigar e sintetizar os principais modelos de avaliação, painéis de indicadores e de dimensões de performance nas organizações de saúde mundiais, a fim de priorizar a análise de sua aplicabilidade na otimização e governança (governança) das organizações de saúde brasileiras;
- identificar e caracterizar as fontes de informações, relevantes e disponíveis no Brasil, que permitam a recolha de dados confiáveis para estruturação de indicadores de performance aplicáveis aos hospitais e serviços de saúde pública do Brasil;
- identificar as categorias de atores chaves (*key stakeholders*), em especial, os principais agentes decisores (implementadores de políticas, planejadores, reguladores e gestores da saúde) no sistema público de saúde do Brasil;
- validar um *proxy* de painéis de indicadores e de dimensões de performance nas organizações de saúde, em uma amostra estratificada do universo selecionado de decisores chaves de organizações e serviços de saúde no sistema público de saúde do Brasil;
- adaptar e aplicar uma escala de mensuração da qualidade percebida dos usuários-utilizadores do SUS, pautada em princípios científicos, validada por decisores chaves, em uma amostra de especialistas e usuários de unidades de saúde selecionadas.

1.4. Estrutura da tese

Em virtude do desafio de fundamentar a avaliação das organizações de saúde este estudo destaca a evolução das teorias e modelos de avaliação na área de saúde, quanto à estruturação do campo da avaliação e a utilização da avaliação no contexto da tomada de decisão em sistemas e organizações de saúde. Em seguida, faz uma análise do percurso de estruturação da saúde pública do Brasil, quanto aos aspectos recentes de governança e avaliação do SUS.

O percurso da investigação bibliográfica foi iniciado pela análise dos escritos recentes no campo de avaliação e monitoramento de sistemas de saúde e, neste propósito, procedeu-se uma revisão da literatura dividida nos seguintes blocos ordenados de investigação: o estado da arte do desenvolvimento do campo da avaliação de resultados em saúde; uma visão global das tendências metodológicas da medição de performance em organizações e serviços de saúde e do desenvolvimento de estratégias e instrumentos de medição de performance em unidades de saúde; os aspectos metodológico e instrumental de mensuração de performance para o apoio à decisão em sistemas de saúde; um breve histórico da governança e da avaliação do sistema de saúde brasileiro; a caracterização das principais fontes de informações do sistema de saúde do Brasil, que possibilitem a geração de indicadores de performance da rede de atenção à saúde (quadro 2).

Na etapa posterior, com o propósito de identificar um quadro teórico-conceitual (*framework*), explorou-se o estado da arte de metodologias avaliativas dos sistemas de saúde, a fim de sedimentar um modelo teórico-lógico de avaliação multidimensional aplicável aos hospitais e serviços de saúde do SUS.

Por fim, a etapa empírica, centrou-se na finalidade de elaborar uma proposta metodológica e um *proxy* de painel de indicadores multidimensionais (produtividade assistencial, eficiência, efetividade, impacto, qualidade percebida etc), devidamente validados pelos agentes de decisão chaves, em uma amostra selecionada multicêntrica, enquanto instrumento de mensuração de performance para a governança de redes de organizações de saúde, no contexto de hospitais e serviços de saúde do SUS.

Nesse propósito, explorou-se a teoria das partes intervenientes (*stakeholders*) e apropriou-se de instrumentos de pesquisa de percepção dos agentes de decisão e de construção de consenso em grupo, em busca de selecionar as dimensões mais relevantes e, posteriormente, modelar uma matriz de indicadores por critérios de julgamento de valor, conforme a relevância atribuída pelos decisores chaves, envolvidos no processo de validação do modelo proposto.

O percurso metodológico desta investigação, no âmbito do doutorado em Gestão e Ciência da Decisão da Universidade de Coimbra, por sua complexidade e multidimensionalidade, envolveu diferentes prismas. A natureza da pesquisa caracterizou-se por uma metodologia avaliativa do tipo observacional (controle de possíveis influências intencionais no objeto pesquisado), transversal, com abordagem por método misto (qualitativo e quantitativo) de avaliação, durante a etapa de validação metodológica e instrumental por decisores chaves da amostra (gestores, profissionais e usuários da saúde).

A investigação apoiou-se em estudo de casos múltiplos (Yin, 2009 apud Minayo et al., 2010) selecionados por conveniência e disponibilidade de e disponibilidade de participação durante a aplicação dos instrumentos para a sua validação (ciclo 1 – agregação de todas as fases de uma das etapas do estudo voltadas para a estruturação e validação dos instrumentos) e para a pesquisa de dados de campo (ciclo 2 – agregação de todas as fases de uma das etapas do estudo voltadas para estudos de dados de campo), porém com amostragem de usuários do tipo estratificado proporcional, na aplicação de campo da escala de qualidade percebida (ciclo 2) na mensuração da satisfação da rede de saúde.

Assim, o quadro abaixo sintetizou a estrutura geral, tópicos, subtópicos e as principais abordagens do estudo.

Estrutura Geral	Principais tópicos	Descrição de objetivos
Enquadramento, abordagens e tópicos da investigação	Avaliação quali-quantitativa para medição de desempenho saúde pública brasileira	Suporte teórico-metodológico da investigação
1– INTRODUÇÃO, ENQUADRAMENTO DO TEMA E OBJETIVOS DO ESTUDO		
O enquadramento e os objetivos da investigação	1.1. Introdução 1.2. Enquadramento do tema (Contexto geral e nacional, brasileiro) 1.3. Objetivos do estudo 1.4. Estrutura da tese	Delimitar o enquadramento do tema, os aspectos de contextos (geral e nacional) e os objetivos da investigação, além de oferecer um quadro estrutural da tese.
2. REVISÃO CONCEITUAL DA TEORIA DA AVALIAÇÃO E DAS BASES DA MEDIÇÃO DE PERFORMANCE NAS ORGANIZAÇÕES DE SAÚDE		
A evolução das metodologias de medição de desempenho em organização de saúde	2.1. O campo de avaliação em saúde e o modelo teórico-lógico	Examinar a evolução teórico conceitual do campo de avaliação em saúde
	2.2. A delimitação do campo da avaliação de desempenho do estudo	Delimitar o campo da avaliação de desempenho aplicado
	2.3. A evolução das metodologias de medição de performance em saúde	Explorar as principais metodologias de medição de performance em saúde

	2.4. As metodologias e os instrumentos de medição de performance e apoio às decisões em organizações de saúde	Descrever as metodologias de medição de performance em saúde aplicada no estudo
3. EVOLUÇÃO RECENTE DA AVALIAÇÃO NO CONTEXTO DA GOVERNANÇA DA SAÚDE DO BRASIL		
O contexto de governança da saúde pública brasileira	3.1. Breve histórico da governança e da avaliação no sistema de saúde brasileiro	Descrever o contexto global da evolução da governança e da avaliação no SUS
	3.2. Arcabouço legal da avaliação e monitoramento no âmbito do Sistema Nacional de Saúde (SUS)	Conhecer as legislações e normatizações do SUS com repercussões na investigação
	3.3. As principais fontes de dados e informações do sistema de saúde do Brasil	Caracterizar as principais fontes de dados e informações do SUS
4. FUNDAMENTOS E ESTRATÉGIAS DE ENVOLVIMENTO DOS DECISORES DE ORGANIZAÇÕES DE SAÚDE		
A utilização da avaliação para a tomada de decisão	4.1. A utilização da avaliação no contexto da tomada de decisão em sistemas e organizações de saúde	Argumentar sobre o valor da utilização da avaliação no contexto da tomada de decisão
A teoria organizacional subjacente ao envolvimento das partes interessadas (<i>stakeholders</i>)	4.2. A teoria organizacional subjacente ao envolvimento dos agentes de decisão (<i>stakeholders</i>) na avaliação participativa	Examinar a teoria subjacente ao envolvimento dos agentes de decisão (<i>stakeholders</i>) para fundamentar a seleção do método colaborativo de validação dos itens de análise de todos os instrumentos
Estratégias mistas (qualitativa e quantitativa) de coleta de informações e	4.3. Estratégias mistas (qualitativa e quantitativa) de coleta de informações e construção de consensos entre	Fundamentar as metodologias e técnicas utilizadas no processo colaborativo de validação dos itens de análise de todos os instrumentos

construção de consensos	de decisores (<i>stakeholders</i>) em organizações de saúde	
5. METODOLOGIA DO ESTUDO		
Procedimentos metodológicos	5.1. Visão global do percurso metodológico da investigação	Propor um enquadramento metodológico e um delineamento da aplicação da investigação; Definir a natureza da investigação, abordagem e métodos de pesquisa; Definição de universo, população e amostra da investigação, Aspectos éticos
Fluxo dos métodos aplicados na coleta de dados investigação	5.2. O fluxo do processo de estruturação e aplicação dos instrumentos para recolha de dados da amostra	Descrever a aplicação dos métodos e estratégias de pesquisa; identificar as fases, etapas e grupos de informantes de cada ciclo da amostragem
Estabelecimento de hipóteses do estudo	5.3. As hipóteses do estudo (itens de avaliação de performance e de avaliação de qualidade/ satisfação percebida)	Definir as hipóteses do estudo quanto aos itens de avaliação de performance e aos itens de avaliação de qualidade/ satisfação.
Modelos e técnicas de análise de dados	5.4. Os modelos de análise e técnicas de validação dos itens de avaliação da investigação	Descrever os fluxos de processos e técnicas aplicadas às análises de resultados e validação de itens de avaliação da investigação.
6. TESTES DE HIPÓTESES E DISCUSSÃO DE RESULTADOS		
A análise dos painéis de indicadores e de dimensões de performance	6.1. Os resultados dos painéis de indicadores e de dimensões de performance	Analisar e discutir os resultados dos painéis de indicadores e de dimensões de performance, à luz da literatura
A análise dos itens de medição de qualidade/ satisfação (escala de qualidade percebida)	6.2. Os resultados dos itens de medição de qualidade/ satisfação (escala de qualidade percebida)	Analisar e discutir os resultados da validação das questões da escala de qualidade percebida, à luz da literatura

Os parâmetros do teste de hipóteses e da validação dos painéis de indicadores e de dimensões de performance	6.3. Os resultados da validação dos painéis de indicadores e de dimensões de performance	Testar os achados com as hipóteses para a validação dos painéis de indicadores e de dimensões de performance
Os parâmetros do teste de hipóteses e da validação das questões da escala de qualidade percebida.	6.4. Os resultados da validação das questões da escala de qualidade percebida	Testar os achados com as hipóteses para a validação das questões da escala de qualidade percebida
7. CONSIDERAÇÕES FINAIS		
Síntese dos achados e conclusões da investigação	Descrição das principais conclusões e contribuições, além das, limitações, dificuldades e possibilidades de investigações futuras	Sintetizar os resultados relativos aos itens de análise em cada estudo; Descrever a abordagem aplicada; identificar limitações e dificuldade do estudo; listar contribuições da investigação; articular sobre possíveis investigações futuras
Apêndices 2-A a 2-I	Abordagem dos aspectos éticos e de autorizações (Aprovação na Plataforma Brasil, em comitê ético - CAAE 54972816.9.3001.5040).	Descrever documentos de participação e autorizações (Investigação autorizada de 30 de junho de 2016 a 30 de maio de 2018).
Apêndices 4 a 8	Apresentação das versões de Instrumentos validados	Descrever os instrumentos validados de coletas de dados
Apêndice 9	Apresentação dos resultados complementares obtidos	Analisar e discutir os resultados complementares
Anexos 1 a 7	Apresentação de tabelas e quadros com conceitos e informações relevantes, procedentes de pesquisa bibliográficas	Descrever conceitos e informações relevantes procedentes da pesquisa bibliográficas

Quadro 2 - Síntese do enquadramento e abordagens da investigação.

PARTE I - CONCEITUAL

2. REVISÃO CONCEITUAL DA TEORIA DA AVALIAÇÃO E DAS BASES DA MEDIÇÃO DE PERFORMANCE NAS ORGANIZAÇÕES DE SAÚDE

Esta revisão Conceitual analisa o embasamento teórico do campo de avaliação e, em seguida, explora o contexto de saúde pública no Brasil através de uma breve síntese da evolução organizativa da saúde brasileira nas últimas décadas, no âmbito dos modelos de governança vigentes na saúde do Brasil, bem como quanto as tentativas de monitoramento de saúde pública.

2.1. A estruturação do campo de avaliação em saúde e o modelo teórico-lógico

Nesta investigação utiliza-se o conceito ampliado de performance como diretamente ligado aos resultados genéricos de uma organização, um serviço ou uma pessoa. Dessa forma, desempenho é pensado especialmente para as pessoas, englobando a performance e os comportamentos verificados. Porém não faz muito sentido avaliar o comportamento individual dos gestores de topo, e assim, nesse nível, os termos desempenho e performance podem ser usados indistintamente. Segue-se a ideia que o termo performance é normalmente usado em várias acepções (Alves, 2012: 2), relativamente as organizações ou referente a indivíduos como o resultado de uma actuação (traduzido em valores de indicadores). Nas organizações assume-se que a performance tem um carácter:

- i) multidimensional – diversas variáveis (ex. qualidade, tempo, custo; dimensão econômica ou social);
- ii) multicausal (ex. derivadas de recursos, de processos e de produtos) e facetas;
- iii) conotada com um tempo determinado – ex. recorre-se a indicadores avançados e atrasados; a rentabilidade obtida hoje pode ser a falência da empresa amanhã;
- iv) sempre medida relativamente aos objetivos dos seus *stakeholders* (Alves, 2012: 2-3).

Em busca dos primórdios do processo de avaliação em saúde Dos Reis *et al.* (1990) sintetizaram que a avaliação de resultados na área de saúde, talvez por sua complexidade, é uma iniciativa relativamente recente. Lembram do relatório de Flexner (1910), que sob o patrocínio da Fundação Carnegie propôs uma avaliação do ensino e sistematização da prática médica, e, do relatório Codman (Porterfield, 1976) nos quais identificam os primeiros sinais da preocupação da qualidade e da performance das organizações de saúde e a criação do embrião

da *Joint Commission of Accreditation of Hospitals (JCAH)*²⁷, em 1928, uma das principais agências mundiais de fomento à qualidade no setor de saúde.

Por seu turno, o aumento indiscriminado de custos da área médica e a necessidade do controle da qualidade da assistência médica pressionam os sistemas de saúde para uma adequada gestão dos recursos, conforme Perez Arias (1983, apud Dos Reis et al., 1990: 51). Estes autores ainda argumentam que “estas são razões suficientes para se buscar o controle dos custos e da qualidade a atenção médica, e que a tendência ao aumento dos custos do setor saúde é, na verdade, universal”.

As bases da avaliação de qualidade e dos estudos econômicos em saúde, foram lançadas antes da década de 1970, mas a maioria dos estudos apontam Avedis Donabedian como debatedor seminal. A partir da proposta inicial de Avedis Donabedian (1980a) – Avaliação de Estrutura, Processos e Resultados -, há uma busca por modelos integrativos de avaliação, em que se avaliam as relações entre estado de saúde, qualidade do cuidado e gastos de recursos. Neste propósito, pode-se citar o modelo de Brook & Lohr (1985b apud Dos Reis et. al, 1990: 54-55), que propõe a avaliação das dimensões: eficácia do cuidado, efetividade do cuidado; as variações das características populacionais e dos níveis de qualidade de cuidado.

Também pode-se pressupor que a gestão por resultados no setor de saúde pública foi potencializada, como em outras áreas governamentais, pelo surgimento da moderna tendência da Nova Gestão Pública (NPG) ou Gerencialismo por responsabilização (*accountability*) dos gastos públicos, em busca de otimização e garantia da eficiência dos gastos públicos.

Nesse mesmo propósito Silva e Formigli (1994 apud Matus, 1997) defendia a importância de “um processo de avaliação continuada dos serviços, também chamada de monitoramento, ou, num estágio mais avançado de organização dos serviços de saúde, como uma “sala de situações” para o planejamento”.

Fletcher *et al.* (1993), articulam sobre a importância de se avaliar os resultados em saúde e citam que “a determinação da eficácia de tratamentos e procedimentos é de fundamental relevância, pois os recursos econômicos são limitados...”.

Por sua vez, Minayo et al. (2010) afirmam que os investimentos maciços em políticas públicas, no bojo do Estado do bem-estar social exigiram a adoção da investigação avaliativa que devem priorizar aspectos como utilidade, viabilidade, precisão técnica, ética e métodos e técnicas de análise de dados adequados.

27 JCAH - Joint Commission of Accreditation of Hospitals

Utilidade para possibilitar a redução de incertezas e propiciar a tomada de decisão relevante; *viabilidade* do ponto de vista político, prático e de custo-benefício; *ética* que permite uma análise que respeita os valores dos interessados, mas que oferece transparência à sociedade nos resultados obtidos; *precisão técnica* que na avaliação de políticas públicas, além do emprego de *métodos e técnicas de análise de dados* adequados, que demandam a colaboração ativa dos atores envolvidos.

A partir destas necessidades e de outras perspectivas analíticas, pesquisadores da área de avaliação organizacional iniciaram a estruturação de um campo de avaliação em saúde.

Neste propósito, Hartz e Vieira-da-Silva (2005) citam que a avaliação das práticas cotidianas corresponderia a um julgamento que se faz a partir do recurso a noções oriundas do “senso comum”, as técnicas não sistemáticas de observação, e, à análise e formulação de juízo de valor dicotômicos e simplificados. Citam ainda que a pesquisa avaliativa corresponderia ao julgamento que é feito sobre as práticas sociais a partir da formulação de uma pergunta não respondida na literatura especializada, sobre características destas práticas, em geral, ou em um contexto particular, através de metodologias científicas (Hartz & Vieira-da-Silva, 2005).

Segundo Samico *et al.* (2010) o campo de avaliação em saúde, engloba uma diversidade de termo, conceitos e métodos conforme sua heterogeneidade, complexidade e subjetividade das intervenções, sejam elas ações, serviços, programas ou políticas públicas, e por isso é um dilema a seleção de um modelo ou abordagem de avaliação (Worthen, Sanders, & Fitzpatrick, 2004).

Quanto ao espectro da avaliação, Hartz e Vieira-da-Silva (2005) argumentam que

O espectro de avaliação tem sido reconhecido por alguns autores, tanto em relação ao objeto da avaliação quanto ao método e abordagem.

[...] Quanto ao objeto, alguns autores, apesar de reconhecerem a diversidade do campo da avaliação enfocam principalmente os métodos e técnicas relacionados a avaliação de programas sociais (Scriven, 1991; Rossi; Freeman, 2004; Patton, 1997).

Outro debate intenso é sobre a aplicabilidade de abordagens qualitativas ou quantitativas e suas características, sendo uma tendência o uso de abordagens mista, como nesta investigação. Por sua vez, alguns aspectos específicos da avaliação têm sido discutidos e classificados por Guba e Lincoln (2011) em três gerações de avaliação supostamente tradicionais: mensuração e comparabilidade a padrões e objetivos (primeira geração); a identificação e descrição do processo, e de como a atividade atinge os resultados almejados (segunda geração); o estabelecimento de critérios de julgamentos baseado em referenciais externos (terceira geração). Porém, Guba e Lincoln (2011) apontam graves problemas e limitações comuns a estas três gerações de avaliação tradicionais que se caracterizariam por:

- tendência à supremacia do ponto de vista gerencial nos processos avaliativos;
- incapacidade em acomodar o pluralismo de atores e projetos envolvidos em torno de qualquer programa;
- hegemonia do paradigma positivista: desconsideração do contexto, privilégio de métodos quantitativos, crença numa verdade única e absoluta caracterização como "não científico" tudo que fuja ao que foi anteriormente citado e desresponsabilização moral e ética do avaliador (já que a ciência seria livre de valores).

Em busca de superar estas limitações, Guba e Lincoln (2011) propõem a avaliação de quarta geração, pautada em um processo construtivista de negociação entre interessados, que se apresenta como inclusiva e participativa em busca de promover a participação pluralistas dos atores sociais e escutar as vozes múltiplas das partes interessadas (*stakeholders*) na avaliação.

A avaliação de quarta geração, termo sugerido por Guba e Lincoln (2011), também é denominada de Avaliação Participativa Construtivista, pois pauta-se em um processo construtivista de negociação entre interessados, que se apresenta como inclusiva e participativa em busca de promover a participação pluralista dos atores sociais e escutar as vozes múltiplas das partes interessadas (*stakeholders*) na avaliação (Guba & Lincoln, 2011), tendo sido selecionada como uma das bases desta investigação para o envolvimento das partes interessadas na avaliação.

De outro lado, a produção em torno da avaliação tem registrados estudos concentrados em três eixos: aqueles especialmente ocupados com as questões de métodos envolvidos nos processos avaliativos; aqueles voltados para os valores e julgamentos dos grupos de interesses envolvidos (*stakeholders*) com um dado serviço ou programa; e outros preocupados com a efetiva utilização dos resultados da avaliação realizada (Alkin & Christie, 2004 apud Campos e Furtado, 2011). Os dois últimos grupos de estudos citados podem, de uma forma ou de outra, contribuir quando se trabalha com o conceito de avaliação para a tomada de decisão.

Patton (1997) propôs a avaliação enquanto metodologia focada na utilização dos resultados para a tomada de decisão. Ao desenvolver uma abordagem da avaliação denominada “avaliação focada para a utilização” inclui na sua definição: a) a maneira de fazer a avaliação que corresponderia à coleta sistemática de informações sobre as atividades, características e resultados dos programas; b) o propósito da avaliação, ou seja, fazer julgamentos sobre os programas e/ou para subsidiar o processo de tomada de decisões sobre futuras programações (Vieira-Da-Silva, 2005).

O conceito de utilização da avaliação pressupõe a atribuição de valor a uma intervenção, além de priorizar a confiabilidade e informações válidas obtidas de forma sistemática, usando comparações com o propósito de entender os mecanismos causais ou princípios gerais de um fenômeno e/ou apoiar a tomada de decisão (Ovretveit, 1998).

Por sua vez, Contrandiopoulos *et al.*(1997: 31) entende que avaliação “baseia-se fundamentalmente em fazer um julgamento de valor a respeito de uma intervenção ou sobre qualquer um de seus componentes, com o Objetivo de ajudar na tomada de decisões”.

Neste prisma, avaliação consiste em fazer um julgamento de valor sobre uma intervenção empregando um dispositivo que permita fornecer informações cientificamente válidas e socialmente legítimas sobre uma intervenção, ou qualquer um de seus componentes, considerando os diferentes atores envolvidos que possam ter julgamentos diferentes, de modo a revelar a posição sobre a intervenção e, construir um julgamento (individualmente ou coletivamente) que possa ser traduzido em ações (Champagne, Contandriopoulos, Broussele, Hartz, & Denis, 2009 apud Samico et al., 2010).

Dessa forma, a avaliação é vista não só como instrumento para verificar a efetividade de programas, mas utilizada também com o intuito de melhorar a qualidade dos serviços. Sendo assim, a avaliação pode ser entendida como um processo de julgamento de valor orientado para a tomada de decisão e governança(Raposo, 2007).

No foco deste trabalho, abordam-se alguns aspectos de aplicabilidade da avaliação para a tomada de decisão em saúde no Brasil, como nos exemplos dos relatos de Tanaka e Tamaki (2012b), de Contandriopoulos (2006), de Fernandes, *et al.*(2011), de Campos e Furtado (2011) e alguns outros estudos iniciais aplicados no sistema de saúde do Brasil.

Neste contexto, escolheu-se trabalhar sob um conceito ampliado de avaliação para a tomada de decisão e considerar os aspectos de utilização de resultados, além da negociação e participação no processo avaliativo.

Uma das primeiras etapas operacionais do processo de avaliação é o desenho do modelo teórico-lógico ou modelo lógico (ML) - *logic model ou logic framework* -, que é concebido no âmbito da avaliação centrada na teoria (*theory-driven evaluation*) nascida na década de 1970 no bojo de metodologias aplicadas pela Agência dos Estados Unidos para o Desenvolvimento Internacional (USAID) e Banco Mundial, conforme afirmam Hartz e Vieira-da-Silva (2005).

Segundo Bunge (1974) o modelo teórico é um sistema hipotético-dedutivo que concerne a um objeto-modelo uma representação Conceitual, esquemática de uma coisa, Artefato ou situação real ou suposta como tal. Assim, o modelo teórico encerra os objetos-modelo, que fazem parte da teoria científica. Para Chen (1990) a teoria é um conjunto inter-relacionado de pressupostos, princípios e/ou proposições que explicam ou dirigem as ações sociais. O modelo teórico tem a função de complementar o ML, visto que inclui as especificações sobre o problema, a população alvo, as condições do contexto, o conteúdo do programa e os atributos

necessários para produzir isolada ou conjuntamente os efeitos esperados (Medina, Silva, Aquino, & Hartz, 2005).

Por outro lado, o modelo lógico (ML) no campo da avaliação, é a ideia que operacionaliza o objeto-modelo, enquanto um conjunto de elementos em um esquema inter-relacionado (McLaughlin & Jordan, 1999; Rowan, 2000).

Para o âmbito desta pesquisa, com o intuito de estruturar uma visão global de eixos temáticos de avaliação (categorias de itens de avaliação) de performance em organizações de saúde, iniciou-se pela análise da pertinência e adequabilidade dos vários itens do ML sugerido no esquema de Hartz e Vieira-da-Silva (2005) para a avaliação de saúde a fim de mapear os conhecimentos tácitos dos agentes chaves da amostra nas categorias de macrodomínios e domínios de avaliação, e, Em seguida, pela etapa de análise e consolidação de painéis de indicadores e dimensões de performance em organizações de saúde. Este modelo será objeto de análise e valoração pelos respondentes e validado ou adaptado em um modelo final pela análise estatística.

Ressalte-se que a estruturação do campo de avaliação em saúde tem uma dinâmica de atualizações constantes e nesta investigação caracterizou-se as bases seminais do campo, por pesquisa bibliográfica integrativa, mas sem a pretensão de revisão exaustiva, que foram sistematizadas (quadro 5) e sintetizadas pelos quadros-conceituais mundiais (quadros 17 a 19, anexo 2), além dos mais relevantes da saúde privada brasileira (quadros 30 a 34, anexo 6) e síntese de publicações no campo de saúde pública do Brasil (quadros 26 e 27, anexo 4).

2.2. A delimitação do campo da avaliação de performance do estudo

No tópico anterior desenvolveu-se uma revisão Conceitual das metodologias avaliativas aplicadas às organizações de saúde em busca de estabelecer a fundamentação teórica e o escopo desta investigação. Para, além disso, torna-se necessário construir um enquadramento teórico (*framework*) para delimitar o campo da presente pesquisa bem como embasar as metodologias e instrumentos aplicados.

O interesse pela avaliação de performance organizacional, possivelmente, emana do âmbito da eficiência do trabalho e do desempenho das pessoas, conforme argumentam Carvalho e Gomes (2000, 2002).

Carvalho (2007) enfatiza que, sobre a definição do constructo *eficácia organizacional*, não há consenso na literatura (Rerimann, 1982; Brunet et al., 1991 Morin et al., 1994; Rogers e Wright, 1998; Savoie e Morin, 2001), além de não haver um modelo explicativo que englobe

a complexidade e contemple a diversidade dos fatores implicados na análise do fenômeno eficácia organizacional.

Sobre o constructo *eficácia*, a despeito de não haver consenso, Navarro e Quijano (2006) e Carvalho (2007) defendem ser majoritária a aceitação da eficácia como um conceito ampliado, focado naquilo que a organização consegue relativamente ao que havia sido previsto ou estava estipulado conseguir, em concordância com o conceito basilar de Etzioni (1975) que concebe a eficácia como o grau segundo o qual a organização consegue alcançar os seus Objetivos e finalidades.

Esses objetivos podem ser considerados desde aspectos de produção (produtividade), de boas práticas de gestão (excelência e qualidade total), de implementação das ações planejadas, ou em uma visão mais generalista até objetivos de otimização de recursos (*eficiência*).

Por sua vez, Brunet (1991 apud Carvalho, 2007: 151), atendendo ao seu sentido etimológico, afirmam que *eficácia* se refere a capacidade de gerar o efeito esperado, definição predominante na literatura.

Enquanto isso, Katz e Khan (1977, Carvalho, 2007: 149) enquadram a *eficiência* no campo da otimização da organização em termos de produção de resultados, definindo-a como a capacidade de alcançar um máximo de vantagens para a organização através de meios econômicos e técnicos ou por meios políticos.

Por não ser escopo da investigação abordar os aspectos mais aprofundados dos conceitos dos constructos *eficácia*, *eficiência*, *efetividade* e, demais termos correlatos, assume-se as bases conceituais acima aludidas no âmbito deste estudo, visto que surgem como majoritárias da literatura (quadros 17 a 19, anexo 2).

Na avaliação participativa, há de se considerar que os múltiplos atores participantes ou agentes interessados (*stakeholders*) atuam com diferentes interesses e comportamentos em relação a variados entendimentos e representações da *eficácia organizacional* e, por conseguinte, com diferentes critérios de mensuração de performance organizacional, como diz Carvalho (2007):

Que os múltiplos *stakeholders*, com motivações diversas e muitas vezes opostas, têm inevitavelmente comportamentos diferenciados que assentam, muitas vezes, em diferentes representações da eficácia (Morin et al., 1994) ou da performance (Buono, 2005; Conlon et al., 2004; Driscoll & Starik, 2004; Lawton et al., 2000; Trieschmann et al., 2000).

Esta investigação pressupõe como razoável que os múltiplos atores sociais exercem influência na escolha dos critérios de avaliação da *eficácia organizacional* e tendem a valorizar

e a adoptar diferentes critérios de avaliação da *eficácia*, como articulam Carvalho e Gomes (2000, 2002).

Nesse sentido, tendem a exercer influência na escolha dos critérios de avaliação da eficácia adoptados por cada organização, bem como nos resultados alcançados por essas mesmas organizações (Morin et al., 1994) ou da performance (Buono, 2005; Conlon et al., 2004; Driscoll & Starik, 2004; Lawton et al., 2000; Trieschemann et al., 2000).

[...] Assim, a eficácia pode ser objecto de avaliações diferenciadas, dando origem a que as organizações possam constituir um espaço de confrontação de interesses, expectativas e representações.

Noutra perspectiva, Carvalho (2007) afirma que o termo *eficácia organizacional* evoluiu na literatura e assumiu a forma de *organizational performance*, segundo estudos de Hirsch e Lewin (1999) que identificam claramente o constructo *performance organizacional* como o sucessor da *eficácia organizacional*, alinhados com Glunk e Wilderom (1999) que notaram que há uma tendência na literatura de substituir o termo *eficácia organizacional* e consolidar o constructo *performance organizacional*. Esta visão é compactuada por Ramalho (2005), enquanto uma possível consolidação da literatura em torno do constructo *performance organizacional*.

Outra definição necessária diz respeito ao campo de enquadramento da avaliação organizacional desta investigação. Neste propósito, Glunk e Wilderom (1999 apud Carvalho, 2007: 152-153) avaliam alguns dos constructos mais significantes e frequentemente utilizados na literatura, tais como: *corporate performance*, *organizational performance* e *organizational effectiveness*²⁸.

Glunk e Wilderom (1999 apud Carvalho, 2007: 153) sugerem ainda que da fusão destas tradições emergiu e domina a literatura o constructo *organizational performance*, enquanto um constructo multidimensional do desempenho organizacional, que pode incluir indicadores de diversas dimensões.

Dessa forma, dado o escopo multidimensional abordado nesta investigação, que pretende matriciar aspectos de resultados económicos (*eficiência*) de efeitos/impactos (*efetividade*) e sociais (*qualidade percebida - satisfação e adequação ao uso-*), pode-se definir o constructo *organizational performance*, ou seja, *performance organizacional*, ou ainda, *desempenho organizacional* – mais usado no contexto de países de língua portuguesa-, enquanto o campo desta investigação.

²⁸Tradução livre do autor: “desempenho corporativo, desempenho organizacional, efetividade organizacional” (Carvalho, 2007: 153)

Deve-se, ainda, selecionar os métodos pertinentes para medição da *eficácia organizacional*. Carvalho (2007) caracteriza estes métodos em duas fases de evolução. Na primeira fase de concepção da eficácia enquanto unidimensional, centrada no alcance dos objetivos ou na otimização de recursos, o problema principal está na definição de critérios válidos e ponderação relativa dos mesmos (Brunet et al., 1991 apud Carvalho, 2007: 164).

Por sua vez, Cameron (1980 apud Carvalho, 2007: 165), define os estudos da *eficácia* como problemas de definição e escolhas adequadas de critérios, sendo assim, de forma complementar, envolve determinar quais são os indicadores apropriados. Isto reforça o envolvimento das partes interessadas (*stakeholders*), nas fases de validação de critérios e na seleção de indicadores, que se aplicou nesta investigação.

Na fase de estudos unidimensionais da *eficácia*, a literatura propõe três abordagens dominantes sobre a *eficácia*. A abordagem *racional dos objetivos* ou modelos de objetivos (J. Navarro & Quijano, 2006 apud Carvalho, 2007: 166), a abordagem *sistêmica e de aquisição de recursos* (K. Cameron, 1981 apud Carvalho, 2007: 169), a abordagem *de processos internos* (K. Cameron, 1981 apud Carvalho, 2007: 170). Estes modelos unidimensionais da *eficácia* são tidos apenas como precursores dos métodos multidimensionais, portanto, não se abordará detalhes desta abordagem nesta investigação.

Em uma fase de transição, surgiu a metáfora política (Carvalho, 2007), a partir da compreensão da natureza complexa e multifacetada das organizações que despertou os estudos iniciais da fase multidimensional da *eficácia*, consolidada na abordagem dos múltiplos constituintes (*stakeholders*), sendo importante discernir que há partes interessadas estratégicas, e/ou outras que são legítimas, classificadas como constituintes (*stakeholders*) relevantes e com legitimidade (Suchman, 1995 apud Carvalho, 2007: 172) ou com saliência (Mitchell, Agle, & Wood, 1997 apud Carvalho, 2007: 172), que influenciam diferentemente a *performance organizacional*. Esta influência dos múltiplos agentes relevantes (*stakeholders*) é uma das bases teóricas-conceituais para promover o envolvimento dos agentes chave interessados na parte empírica desta investigação.

Navarro e Quijano (2006) afirmam que na abordagem dos múltiplos atores interessados (*stakeholders*) é necessário identificar os diversos interesses estabelecidos e avaliar em que medida esses interesses estão a ser satisfeitos, assim propõem alguns diferentes critérios de *eficácia*, por tipo de grupos de interessados.

Assim, pode-se perceber que mais do que oferecer critérios universalmente válidos para avaliar a *eficácia* ou *performance* organizacional, é preciso, em primeira instância, analisar que critérios consideram os membros (*stakeholders*) da organização, alvo da avaliação executada,

como indicadores válidos de eficácia organizacional. Dessa forma, a *eficácia* organizacional passa a ser vista como um conceito/constructo baseado nos valores dos *stakeholders* envolvidos em sua validação (J. Navarro & Quijano, 2006 apud Carvalho, 2007).

Para além disto, há de se considerar que os constructos racionalizados pelos agentes chaves (*stakeholders*), considerados relevantes e legítimos, não são a realidade objetiva, mas trata-se de uma realidade socialmente construída pelas múltiplas visões desses *stakeholders*. Assim, esses constructos validados constituem abstrações resultantes da ação de observação e análise sobre essa mesma realidade permite atribuir-lhe um sentido (K. Cameron, 1981 apud Carvalho, 2007). O grau de aceitabilidade e legitimidade dos critérios depende do contexto organizacional na qual eles são gerados e/ou integrados (Lourença, Miguez, Gomes, & Carvalho, 2004).

Diante das limitações dos modelos de medidas unidimensionais da *eficácia* ou *performance* organizacional prosperou variados modelos operacionais *multidimensionais*, sendo os principais caracterizados por Carvalho (2007), todos com suas vantagens e limitações, com diferentes perspectivas.

Neste contexto, surgiu na literatura a fase multidimensional da avaliação da *eficácia organizacional*, na qual dada a complexidade e as multifacetadas das organizações percebeu-se a necessidade de modelos multidimensionais de *eficácia* ou *performance* organizacional.

Após extensa análise da literatura para identificar as diferentes representações multidimensionais da *eficácia organizacional*, Carvalho (2007) caracteriza ao menos 10 modelos multidimensionais, sintetizados no quadro 3.

Modelo	Componentes	Pontos fortes	Pontos fracos
Modelo da Gestão Sócio-económica de Savall (1975)	-Social -Económica	-Paridade explícita entre as componentes.	-Subordinação real do social ao económico -Validação teórica (parcial e fragmentada) e empírica.
Modelo das Seis Caixas de Weisbord (1976)	-Relações -Mecanismos Auxiliares -Recompensas -Estruturas -Estratégias/ Propósitos -Liderança	-A simplicidade da estrutura permite uma compreensão rápida do modelo.	-Reduccionismo da complexidade organizacional. -Dificuldades ligadas aos nexos causais entre os propósitos e outras caixas.
Modelo dos Valores Contrastant	-Social -Sistémica -Económica	-Fundamentação teórica	-Viés académico

es de Quinn e Rohrbaugh (1983)	-Burocrática		
Modelo ROR: Retorno-dos-Recursos De Halal (1986)	-Retorno de Recursos -Accionista -RH -Cliente -Público -Associados	-Sustentação na perspectiva dos constituintes múltiplos	-O interesse da organização é equacionado como a soma dos interesses dos seus constituintes múltiplos. -Há medidas de eficácia organizacional que não correspondem ao interesse de nenhum <i>stakeholder</i> considerado individualmente mas sim ao conjunto de <i>stakeholders</i> .
Modelo do Fluxo Organizacional de Porras (1987)	-Tecnologia -Fatores Sociais -Arranjos Organizacionais - Cenário Físico	-Considera simultaneamente processos e estruturas na avaliação da eficácia organizacional - É graficamente apelativo e de fácil interpretação. -É teoricamente integrado com outras perspectivas na mudança planejada. -Validado por alguns estudos.	-Limitação das variáveis aos quatro domínios (tecnologia, social, arranjo organizacional e físico). -Potencial desajustamento a idiosincrasias organizacionais. -Atenua a importância relativa dos <i>stakeholders</i> ou fatores de mudança organizacional. -Não permite compreender os gaps entre o que se é e o que se gostaria de ser.
Modelo de McNair, Lynch e Cross (1990)	-Visão -Mercado -Finanças -Satisfação do Cliente -Flexibilidade -Produtividade -Qualidade -Serviço -Tempo de ciclo -Desperdício	-Integração entre a componente tangível e a intangível. -Estruturação inclusiva (a eficácia organizacional é dependente da eficácia das unidades e subunidades).	- Visão racional top-down da gestão, mas esquece a influência ascendente. -Estrutura funcionalista. -Não integra a noção de antagonismo entre medidas de eficácia organizacional. -Inexistência de eixos estruturantes.
Modelo de Burke-Litwin (1992)	-Desperdício individual -Desempenho organizacional	-Integrador da literatura dominante na Mudança Organizacional e	-Complexidade dos nexos causais (tudo está ligado a tudo).

		Gestão do desempenho. -Validade reconhecida por entidades oficiais e estabelecidas no mercado.	-Importância atribuída ao meio enquanto factor de mudança organizacional. -Estrutura não validada empiricamente.
Modelo Balanced Scorecard de Kaplan e Norton (1992)	-Financeira -Cliente -Negócio -Inovação	-Tomou dianteira na contestação à concepção economicista-financeira unidimensional	-A autoridade é a fonte de validação -Sem fundamentação teórica -Modelo prescritivo
Modelo Integrativo de Morin, Savoie e Beaudin (1994)	-Psicossocial -Económica -Política -Sistémica	-Assenta nas práticas sendo balizado por uma reflexão com base na Teoria Organizacional.	- Modelo construído de forma indutiva, baseado em estudo empírico exploratório, com estruturação referida a tipologia consagrada, mas atribuição resultante de opinião da autora – Morin.
Modelo de Skandia Navigator (199)	-Financeira -Cliente - Processo -Social -Inovação	-Utilizado na Skandia -Grande aceitação na Academia e Empresas	Modelo prescritivo.

Quadro 3 - Quadro comparativo dos vários modelos multidimensionais analisados. Fonte: Carvalho (2007). Adaptado de Ramalho (2005: 248-249).

Enquanto exemplo, resolveu-se destacar apenas dois métodos multidimensionais citados, por sua relevância na literatura, a seguir apresentados.

Em um dos modelos multidimensionais de destaque na literatura, Kaplan e Norton (1992), desenvolveram o método *Balanced Scorecard* (BSC), sendo precursores do rompimento do paradigma dominante de medidas exclusivamente financeiras da avaliação da eficácia organizacional e introduziram a “gestão dos intangíveis”, para realçar a importância de articular a identificação dos objetivos com a estratégia organizacional. No BSC, um dos modelos multidimensionais mais utilizados, propõe-se sintetizar indicadores em um painel de quatro perspectivas (financeira, do cliente, do processo interno de negócio e da aprendizagem e crescimento).

Também pode-se destacar, dentre outros, o modelo *Interativo de Eficácia Organizacional*, exposto no quadro 4a seguir:

Valor dos Recursos Humanos	Eficiência Económica
----------------------------	----------------------

<p>* <u>Mobilização do Pessoal</u> - Grau de interesse que os empregados manifestam pelo seu trabalho e pela organização, bem como o esforço envidado para atingir objetivos</p> <p>* <u>Moral do Pessoal</u> - Grau em que a experiência de trabalho é avaliada positivamente pelo empregado</p> <p>* <u>Desempenho do Pessoal (Produtividade)</u>- Qualidade ou quantidade de produção por empregado ou por grupo/equipa</p> <p>* <u>Desenvolvimento do Pessoal</u> - Grau de incremento das competências dos membros da organização</p>	<p>* <u>Economia de Recursos</u> - Grau em que a organização reduz a quantidade de recursos utilizados mantendo o bom funcionamento do sistema</p> <p>* <u>Produtividade</u> - Quantidade ou qualidade de bens e serviços produzidos pela organização face à quantidade de recursos utilizados para a sua produção durante um determinado período de tempo.</p>
<p>Legitimidade da Organização face a grupos externos</p>	<p>Perenidade da Organização</p>
<p>* <u>Satisfação dos Financiadores</u> - Grau em que os financiadores estimam que os seus fundos são utilizados de forma óptima</p> <p>* <u>Satisfação dos Clientes</u>- Julgamento efectuado pelos clientes sobre a forma como a organização responde às suas necessidades</p> <p>* <u>Satisfação dos Organismos Reguladores</u> - Grau em que a organização respeita as leis e regulamentos que regem as suas actividades</p> <p>* <u>Satisfação da Comunidade</u>- Apreciação que a comunidade alargada faz das actividades e dos efeitos da organização</p>	<p>* <u>Qualidade do Produto</u> - Grau em que o produto responde às necessidades dos clientes</p> <p>* <u>Rentabilidade Financeira</u> - Grau em que certos indicadores financeiros da organização aumentam ou diminuem face aos exercícios precedentes, ou face a um Objetivo fixado</p> <p>* <u>Competitividade</u> - Grau em que certos indicadores económicos se apresentam favoráveis ou desfavoráveis, por comparação com a concorrência, ramo ou sector de actividade.</p>

Quadro 4 - Modelo integrativo da eficácia organizacional (Morin et al., 1994) – critérios e indicadores de cada dimensão. Fonte: Carvalho (2007). Adaptado de Morin, Savoie e Beaudin (1994).

Diante de tais bases teóricas-conceituais pode-se considerar a *eficácia* ou *performance* organizacional como um constructo multidimensional, subjetivo e socialmente construído, tratando-se de um juízo de valor das partes interessadas (*stakeholders*) que validam seus critérios, segundo Morin *et al.* (1994 apud Carvalho, 2007: 180). Este embasamento teórico da *eficácia* ou *performance* organizacional foi incorporado neste estudo pela mensuração da percepção das relevâncias/importâncias atribuídas aos itens de análise (macrodomínios e domínios de avaliação e dimensões e indicadores de performance) pelos grupos de interessados, envolvidos em nossa investigação.

Dessa forma, justifica-se analisar a literatura quanto as metodologias, instrumentos, dimensões e indicadores de performance disponíveis para a mensuração dos constructos propostos (itens de avaliação de performance), para posterior validação pelos agentes chaves (*key stakeholders*) envolvidos nesta investigação.

Neste propósito, de forma sintética, deve-se conceber um processo avaliativo para a tomada de decisão, através da escolha de metodologias, indicadores e de parâmetros que contemple essa diversidade de pontos de vistas, através da formulação da melhor estratégia, na seleção da abordagem, na definição de níveis e atributos, bem como na seleção de critérios, indicadores e padrões (Vieira-Da-Silva, 2005; Tanaka & Tamaki, 2012).

Assim, algumas metodologias e sistemas de indicadores de destaque mundiais (anexo 2), também no campo de saúde pública do Brasil (anexo 5) além dos mais relevantes da saúde privada brasileira (anexo 6), foram sistematizadas quanto as possíveis potencialidades, virtudes, limitações e/ou desafios de implementação, além de abordagens destacadas de quadro teórico-conceitual (*framework*) destacados na literatura de avaliação de performance em saúde (tópico 2.4.3).

A partir destas definições conceituais, enquanto fundamentos para a fase instrumental de aplicação empírica desta investigação, destacam-se a seguir alguns aspectos metodológicos e/ou de instrumentos de avaliação e mensuração de *performance organizacional*, em especial, no âmbito de unidades de saúde pública.

2.3. A evolução das metodologias de medição de performance em organizações de saúde

A evolução dos processos de medição de performance em organizações de saúde revela diversidades, complexidades e desafios a serem enfrentados. Por exemplo, apesar do relativo consenso que a qualidade é vista como um dos aspectos centrais a ser considerado para a avaliação em saúde (Donabedian, 1980c; H. V. Vuori, 1982), não existe uma definição universalmente aceita de cuidado em saúde, qualidade ou qualidade de atendimento. No nível individual, Campbell et al. (2000: 1621) propõe a qualidade como sendo a capacidade de acessar o atendimento eficaz com o Objetivo de maximizar os benefícios à saúde em relação à necessidade, no entanto, destaca que deve-se abordar também a qualidade do ponto de vista populacional:

No nível societário, a qualidade do atendimento é a capacidade de acessar o atendimento efetivo de forma eficiente e com base equitativa para a otimização do benefício/bem-estar de saúde para toda a população.

[...] No entanto, as complexidades e necessidades de pacientes individuais ou consultas só são discerníveis no nível do indivíduo e focalizando o que é necessário para um indivíduo em sua interação com um prestador de cuidados de saúde (Campbell et al., 2000: 1621-22).

No mesmo trabalho, Campbell et al. (2000) lembram do enfoque especial da qualidade e adequação do cuidado nas necessidades individuais dos pacientes (McGlynn, 1997; Stott, Kinnersley, & Elwyn, 1997), porém afirmam que as medidas das necessidades de saúde dos indivíduos devem ser equilibradas quanto a otimização dos resultados de saúde para a sociedade. Lembram que diferentes partes interessadas valorizam diferentes componentes da qualidade do atendimento para articular que estes aspectos incorporam complexidades para a adequada mensuração da performance em sistemas de saúde.

Por sua vez, Arah et al. (2006: 7) defendem que o funcionamento do sistema de prestação de cuidados de saúde é avaliado em relação aos objetivos públicos estabelecidos para o nível e a distribuição dos benefícios e os custos dos cuidados pessoais e de saúde pública.

Noutra perspectiva, Murray & Frenk (2000b) que ressaltam a avaliação para os tomadores de decisão, que precisam quantificar a variação no desempenho do sistema de saúde, de seus subcomponentes, como regiões de saúde, serviços e unidades de saúde, a fim de obter evidências para decisões através de informações significativas e comparáveis sobre o desempenho do sistema de saúde.

Outra complexidade inerente ao processo de avaliação em saúde trata-se da diversidade de termos e conceitos envolvidos. Neste propósito, Silva & Formigli (1994) afirmam que há frequentemente uma superposição entre as definições de eficácia, efetividade, eficiência e, mesmo, qualidade, com diversas evoluções em seus sentidos e aplicações:

A *qualidade* tem sido definida como “efeito” (Donabedian, 1980c); a *efetividade*, como qualidade (Freeborn & Greenlick, 1973). Tanto a *eficácia* como a *efetividade* têm sido definidas em relação ao cumprimento de objetivos propostos (Rodrigues & Torres, 1982; Rundall, 1992). A *efetividade* tem sido vista, ainda, como um atributo composto pela *eficiência* e pela *eficácia* (F. E. Campos, 1988).

[...] A *eficácia*, por exemplo, inicialmente era definida como benefício ou efeito da atividade pela OMS (1970), ou como cumprimento de metas em livros didáticos de administração sanitária (Ferrara, Acebal, & Paganini, 1976), ou ainda utilizada como equivalente a *efetividade* (Doll, 1974). Mais recentemente, verifica-se uma tendência à convergência entre algumas definições. A *eficácia* vem sendo referida mais consistentemente como o efeito potencial ou o *efeito* em determinadas condições experimentais (OTA, 1978; Vuori, 1982; Cavallo et al., 1988; Donabedian 1990). Já a *efetividade* e o *impacto* traduziriam o efeito real num sistema operacional (OTA, 1978; Chavez, 1981; Vuori 1982; Brook & Lohp 1985; Donabedian, 1990).

[...] A diferença entre *impacto* e *efetividade* tem sido buscada tanto em relação ao objeto da avaliação quanto no que concerne ao tempo. Ou seja, o efeito de

determinado serviço sobre um grupo populacional corresponderia à *efetividade*. Por outro lado, o efeito de um “sistema” sobre uma população expressaria o *impacto* (De Geyndt, 1970). Em relação ao tempo, o curto prazo diria respeito à *efetividade* e o longo prazo, ao *impacto* (Rundall, 1992; Silva & Formigli, 1994: 82).

Silva e Formigli (1994) abordam as noções de qualidade, efetividade, impacto, eficiência e satisfação do usuário pelo prisma consideradas por alguns autores, como atributos do cuidado médico. Já a cobertura, o acesso e a equidade são referidas ao “sistema de saúde” (Lane & Kelman, 1975).

Algumas iniciativas de medição do desempenho nos cuidados de saúde buscam a avaliação direta da qualidade, suficiência, eficiência e adequação do suprimento de serviços assistenciais, dentre outras características, entretanto, há outras que abordam o desempenho através de um *proxy* de medidas de performance no atendimento.

Neste contexto, segundo argumenta Kazandjian (2003) o fator de confusão predominante durante a fase de implementação de qualquer sistema de medição de performance talvez seja se as medidas do sistema refletem a qualidade do atendimento ou se são indicadores dos processos subjacentes ao atendimento. Afirmam que a adequação do cuidado deve ser decidida primeiro - só então pode-se avançar para um julgamento sobre a eficiência da produção, pois não importa quão eficiente seja, a produção de cuidados inadequados permanece inaceitável. Assinala, ainda, que a adequação compreende uma série de dimensões essenciais para abordar a interpretação do desempenho medida por meio de indicadores (V. A. Kazandjian, 2003).

Ainda segundo Kazandjian (2003), como proposto por Donabedian, a qualidade do cuidado pode ser capturada pela construção operacional dicotômica de adequação e eficiência de produção. É razoavelmente aceito pela literatura que as bases da avaliação de qualidade e dos estudos econômicos em saúde foram lançadas antes da década de 1970, além da maioria dos estudos apontarem Avedis Donabedian como debatedor seminal da mensuração de resultados em saúde, pautado em três pilares: *Estrutura, Processos e Resultados* (Donabedian, 1980b, 1988).

Fletcher *et al.* (1993), articulam sobre a importância de se avaliar os resultados em saúde e citam que a determinação da eficácia de tratamentos e procedimentos é de fundamental relevância, pois os recursos econômicos são limitados.

No entanto, as medidas contingenciais dos gastos em saúde nos sistemas públicos mundiais têm oscilado entre o controle da gestão da aplicação de recursos e na dosagem entre um sistema público e privado. Porém, há uma nova tendência emergente na gestão de resultados

em saúde que sugere que “a mudança central do foco de atenção foi das atividades ou “produtos” das agências governamentais (*output*) para os resultados (*outcomes*) dessas atividades” (Kettl, 1996; Almeida, 1999 apud F Viacava et al., 2004: 712).

Segundo alguns autores (FIOCRUZ, 2012; Veillard et al., 2005; F Viacava et al., 2004) o relatório sobre a saúde *World Health Report 2000*(WHO, 2000a) identificou três objetivos gerais de um sistema de saúde, além de enfatizar que a estrutura de prestação de serviços impactam no desempenho do sistema de saúde:

Alcançar boa saúde para a população, garantir que os serviços de saúde sejam acolhedor, respeitem os direitos e a dignidade dos usuários e, assegurar sistemas de pagamento justos. O hospital tem um papel central na consecução desses objetivos (McKee & Healy, 2002). Obviamente, a organização, configuração e prestação de serviços de saúde impactam no desempenho do sistema geral de saúde (Veillard et al., 2005: 487).

Por sua vez, Escorel *et al.*(2007) afirmam que os serviços de saúde precisam ser constantemente monitorados e avaliados, como forma de garantir a qualidade do atendimento à população. De acordo com os autores, em casos onde não há um sistema de avaliação efetivo, não seria possível controlar falhas no atendimento e no manejo profissional e, garantir qualidade aos usuários.

Nesta nova visão, deve-se avaliar a quantidade de serviços ou produtos de saúde (atendimentos, exames, cirurgias, internações etc) produzidos pelo sistema de saúde, além da mensuração multidimensional dos resultados, para monitorar a gestão eficiente dos recursos, o alcance de efetividade e qualidade dos atendimentos e, o impacto resolutivo na situação de saúde dos atendidos, da sociedade e no sistema de saúde.

No âmbito das organizações, os indicadores de performance ou desempenho podem ser entendidos como um conjunto de elementos para avaliar o grau com que os resultados pretendidos são alcançados (Barzelay, 1997; Boland & Fowler, 2000). Estas métricas podem ser usadas para melhorar a qualidade ou desempenho dos serviços, estabelecer padrões desejáveis, balizar as aferições de resultados alcançados e verificar o alcance de metas (Stevens, Stokes, & O’Mahony, 2006).

Assim, no setor público, campo deste estudo, Stevens *et al.*(2006) notaram a tendência de racionalização da quantidade dos indicadores e, ênfase para o alcance de resultados. Por sua vez, Boland e Fowler (2000) citam que uma das principais questões a resolver para a otimização dos indicadores se refere a inexistência de um padrão de quais são as dimensões de avaliação

de desempenho que devem ser consideradas, e, por conseguinte, quais os indicadores devem ser usados para avaliar e monitorar adequadamente os resultados organizacionais pretendidos.

Segundo Viacava (2004: 153) a variedade de modelos de desempenho descrita na literatura “reflete a complexidade da questão, além da diversidade de definições, das dimensões englobadas, dos determinantes do desempenho, dos indicadores utilizados e dos critérios de julgamento” (Sicotte et al., 1998).

A multiplicidade de perspectivas e de complexidades envolvidas, além das diversidades de escopos e definições demandam a análise dos principais quadros teóricos e metodológicos de avaliação relevantes da literatura, por exemplo, para estabelecer um enquadramento teórico-conceitual (*framework*) do sistema de medição da performance no âmbito de unidades de saúde pública, um dos desafios desta investigação, cujos fundamentos e potencialidades são destacados a seguir.

2.4. As metodologias e os instrumentos de medição de performance e apoio à decisão em organizações de saúde

Para início de exploração da literatura relevante, montou-se uma estratégia de revisão integrativa através de uma ampla pesquisa genérica, em publicações recentes (Jan/2005 a Dez/2016). Foram investigadas as publicações consideradas relevantes da área de governança, gestão e avaliação de saúde além de temas transversais ao campo de avaliação em saúde, tendo explorado as seguintes bases de dados: B-On; ABI-Inform/Proquest; Repositórios de teses digitais DART-Europe, RCAAP-Portugal e BDTD-Brasil. Uma limitação inicial da pesquisa bibliográfica foi a impossibilidade de explorar outros repositórios, por exemplo, Pubmed (Medline), Cochrane Library, Embase, CRD (York), Portal Capes-BR (Bireme, Scielo e Lilacs), AHRQ (EUA), etc.

Na fase exploratória publicações sobre métodos e instrumentos de avaliação, buscou-se inicialmente uma visão global de governança e decisão em organizações de saúde, tendo se pesquisado por termos como *health decison-making*, *hospitals decision-making*, *stakeholders in hospitals*. Em seguida, avançou-se para pesquisar sobre áreas mais específicas como: avaliação econômica, análise de efetividade parametrizada pelo risco e avaliação da qualidade percebida - satisfação e adequação ao uso, centrados em metodologias de avaliação multidimensional aplicadas à saúde, detalhadas nos tópicos a seguir. Após estudar as dimensões de avaliação de performance propostas nos objetivos desta investigação, sob uma visão mais instrumental, explorou-se as bases de artigos e teses pelos termos *Hospital Dashboard*, *Clinical*

Dashboard, Benchmarking, Balanced Scorecard e termos similares com foco em saúde, sendo patente a limitação de publicações nestes termos no contexto de gestão de saúde (tabela 110, anexo 1).

Para o escopo deste tópico, foram considerados apenas os sistemas de indicadores que abordassem de forma objetiva e instrumental pelo menos uma das dimensões deste estudo (produtividade assistencial, eficiência, efetividade, impacto, efeitos, qualidade percebida ou outras que surgirem na investigação), com aplicação multicêntrica em unidades de saúde, para posterior detalhamento e avaliação de sua possível adaptabilidade ao ambiente das organizações e unidades de saúde pública do Brasil, contexto deste estudo. Também identificou-se alguns sistemas de indicadores de avaliação de performance que são regularmente aplicados em hospitais e unidades de saúde, sendo selecionados para o escopo deste trabalho apenas aqueles que são multidimensionais (produtividade assistencial, eficiência, efetividade, eficácia, qualidade centrada no paciente etc), desde que se percebesse a relevância de aplicação em diferentes contextos de sistemas de saúde transnacionais, tendo os principais resultados sintetizados(anexo 2).

Neste ponto, efetivou-se as leituras cuidadosas dos principais estudos considerados relevantes, sobre metodologias de mensuração de avaliação de performance e de resultados em hospitais e serviços de saúde, para analisar a adequabilidade dos itens de avaliação de performance abordados e sua possível incorporação no quadro teórico-conceitual (*framework*) desta investigação, no âmbito do sistema público de saúde do Brasil (quadros 15 a 17, anexo 2). Diante disto, analisou-se criticamente alguns destes quadros conceituais (*frameworks*), além de seus indicadores e dimensões de avaliação de performance em saúde, a fim de selecionar os considerados mais adequados para os propósitos desta investigação (tópico 2.4.3) e uma base teórica bem como possíveis quadros conceituais (*frameworks*) (quadros 20 a 23, anexo 2) para a modelagem de uma escala de qualidade percebida (satisfação) validada para o contexto do SUS (tópico 2.4.4).

2.4.1. As metodologias de avaliação da eficiência em organizações de saúde

Nesta etapa, aprofundou-se a pesquisa avaliando as publicações na área de economia de saúde e gestão de eficiência em organizações de saúde, além de seus termos (MeSH) mais utilizados na literatura mundial, por exemplo, *Health Efficiency, Health Economics, Health Economics Evaluation, Data Envelopment Analysis and Health (DEA), Stochastic Frontier*

Analysis (SFA),²⁹ dentre outros similares, sendo sintetizados os resultados das aplicações no sistema de saúde do Brasil e comprovada a incipiência do seus usos (quadro 25, anexo 3).

A literatura aponta vários métodos de estimação da eficiência. O método DEA, do tipo não-paramétrico, tem a facilidade de avaliar apenas as interações integradas das múltiplas entradas para definir as saídas ou consequências. Enquanto primeira escolha pretende-se analisar a aplicabilidade do método de Análise Envoltória de Dados (DEA) que pode ser modelado através de programação linear, e apesar de sua simplicidade relativa disponibiliza uma vasta gama de informações para os decisores e formuladores de políticas, além de oferecer fundamentos para a medição de performance em saúde.

La Forgia e Couttolenc (2009) afirmam que no âmbito do Brasil, o uso do modelo DEA, ainda é incipiente. O modelo SFA é exceção nos estudos publicados na saúde do Brasil. Assim, quanto aos estudos no contexto brasileiro, foram tímidos os resultados obtidos sobre a aplicação de métodos de análise de eficiência durante revisão sistemática das bases de dados de artigos e teses, optou-se por complementar esta análise dos achados do estudo de revisão extensiva da literatura sobre a produção científica em economia da saúde no Brasil (2004-2012), apresentado pela OPAS³⁰, ABrES³¹ e MS³²(Brasil, 2013a), no final de 2013, com os estudos mais relevantes sintetizados no quadro 24(anexo 3).

Dessa forma, os resultados da pesquisa da literatura (tabela 110, anexo 1) quanto as aplicações de métodos científicos de medição da eficiência no âmbito da saúde do Brasil foram incipientes, realidade reforçada na fase empírica da investigação visto que os decisores chaves basicamente identificaram alguns indicadores operacionais de medidas de eficiência, sem apontar métodos de análise técnica da eficiência. Diante disso, decidiu-se limitar esta investigação à análise de relevância dos indicadores de eficiência sugeridos na fase empírica, além de definir os seus respectivos pesos (importâncias/relevâncias), conforme os critérios dos decisores, sem aprofundar a análise técnica da eficiência, e, sugerir estudos futuros sobre modelo de cálculo e análise da eficiência técnica, a fim de comparar os aspectos de eficiência entre as organizações da saúde estudadas.

2.4.2. As metodologias de avaliação de efetividade e ajustamento de risco em saúde

29 SFA - Stochastic Frontier Analysis (Análise de Fronteira Estocástica)

30 OPAS- Organização PanAmericana de Saúde, representante da OMS na América Latina e Caribe

31 ABrES – Associação Brasileira de Economia da Saúde

32 MS – Ministério da Saúde do Brasil

A avaliação da dimensão de efetividade do atendimento é afetada pelas condições clínicas dos usuários atendidos. Deve ser senso comum que as organizações de saúde atendem a diversas combinações diferentes de pacientes, a depender da procura, que podem variar, em menor ou maior grau, de acordo com idade, gênero, diagnósticos, fatores de risco, e severidade do caso ou complexidade do tratamento. Assim, pode-se perceber que estas diferenças nos perfis e nas condições clínicas dos usuários atendidos podem causar impactos nos recursos (insumos), nos processos produtivos e nos resultados (produtos) das organizações e dos serviços de saúde.

Para permitir a mensuração comparativa adequada da efetividade entre as organizações de saúde depende-se da homogeneização dos pacientes ou dos grupos de diagnósticos atendidos nas unidades de serviços. Diante disto, segundo Shaughnessy e Hittle (2002), pode-se afirmar que o principal motivo para utilização das metodologias de ajustamento pelo risco é a possibilidade de comparação entre resultados e/ou performance das organizações de saúde, por exemplo, pelo uso de *índices de casemix* ou *Diagnosis Related Groups (DRG)*.

Por sua vez, Iezzoni (2003) cita alguns dos seguintes objetivos no uso da parametrização pelo risco:

Possibilitar a comparação dos níveis de eficiência e custos entre prestadores e seguros de saúde; permitir a construção de relatórios públicos sobre a performance dos prestadores individuais; permitir a comparação interna dos resultados obtidos por doente entre médicos no sentido de melhorar a performance assistencial; subsidiar políticas de reembolsos de acordo com níveis de pagamentos por indivíduo ou para planos de seguros; incentivar os prestadores a aceitar pacientes com elevados riscos.

Por exemplo, do ponto de vista do financiamento per capita, Bonilla e Rubio (2000) argumentam que o ajustamento pelo risco pode ser usado como um processo que adequa o valor distribuído em função das características dos indivíduos (IHPS, 2005).

Assim, uma alternativa hegemônica presente na literatura (tabela 110, anexo 1) é compor um *Casemix* unidimensional pelo custo médio relativo, pela facilidade de apuração, mas há restrições neste método visto que é limitado apenas à dimensão do uso de recursos.

Outras aproximações unidimensionais de *casemix*, também medidas ao nível de unidades de serviços, são usadas para refletir a severidade dos casos ou complexidade dos tratamentos, tais como: proporção dos casos atendidos de média *versus* de alta complexidade; a relação entre o número de casos de uma determinada especialidade e a quantidade total de atendimentos realizados; a relação da quantidade de determinado tipo de exame e a quantidade de atendimentos realizados, etc. Este tipo de *casemix* reflete os atendimentos efetivamente

prestados aos pacientes, e por isto são melhores que os *casemix* focados exclusivamente em recursos, mas são limitados por ser unidimensional, e não permitir detecção de diferenças ao nível de pacientes, mas apenas ao nível de grupos de procedimentos ou organizações de saúde.

Segundo La Forgia e Couttolenc (2009) os sistemas de classificação do perfil dos pacientes e gravidade dos casos e/ou no risco relativo são os melhores *casemix*, por ter abordagem direcionada ao perfil clínico do paciente, que tem relação direta com a dimensão da severidade clínica do caso, que reúne indicadores importantes das necessidades clínicas do paciente, e, pode incluir os fatores de riscos, o histórico e as condições clínicas atuais do paciente.

Nas metodologias de parametrização de risco há pelo menos três perspectivas dominantes, quanto ao foco na utilização de recursos *versus* foco no estadiamento da doença. Segundo afirma Iezzoni (1997 apud Costa & Lopes, 2011: 9), há diferentes pontos de vistas quanto a predominância de uma perspectiva em relação à outra, com sistemas que privilegiam a *utilização de recursos*, correlacionado com a severidade da doença, como por exemplo, os *Diagnosis Related Groups (DRG)* e o *Acuity Index Method (AIM)*, outros priorizam a severidade do doente ou *estadiamento da doença*, como por exemplo, o *Acute Physiologic and Chronic Health Evaluation (APACHE)*, e o *System of Acute Physiologic Score (SAPS)*, e o *Mortality Probability Model (MPM)*, e o *Disease Staging* ou os *Medis Groups*, enquanto há outras metodologias que consubstanciam os dois princípios (*utilização de recursos e estadiamento de doenças*), como os *Patient Management Categories (PMCs)*.

Qualquer que seja a abordagem ou perspectiva, a classificação dos pacientes ou grupos de diagnósticos deve explicitar as diferenças de resultados através de instrumentos de avaliação clínica ou efeitos assistenciais, tais como: taxa de mortalidade ajustada pelo risco, taxa de complicações ajustada pelo risco, taxa de readmissões ou reinternações ajustada pelo risco, etc.

Dessa base teórica percebe-se que a ampla maioria dos estudos de avaliação de resultados em saúde, tem se pautado nos Grupos Relacionados de Diagnósticos - *Diagnosis Related Groups (DRG)* ou em outros sistemas similares de grupos homogêneos de diagnósticos, que correlacionam as severidades dos casos às intensidades de uso de recursos.

Os DRG's começaram a ser concebidos nos finais da década de 1960 na Universidade de Yale (EUA), com a finalidade de associar os diferentes tipos de pacientes tratados (*casemix*) com os respectivos custos (Fetter, Shin, Freeman, Averill, & Thompson, 1980). O *casemix* trata-se da medição da variedade das situações clínicas dos doentes tratados por cada hospital, organização de saúde ou prestador, enquanto o *índice de casemix* é o valor que expressa a diversidade dos casos tratados em cada hospital (Lichtig, 1986).

Ao considerar os sistemas oficiais de classificação de pacientes do Brasil, percebem-se grandes dificuldades a superar para implantar os DRG's, visto que o SUS tem a codificação de procedimentos pautada em um sistema SIH/AIH, que se baseia apenas nos principais procedimentos realizados, sendo que os registros de procedimentos secundários são pouco sistematizados (R. C Couto, 2012; Rodrigo Camargos Couto et al., 2012; Daibert, 2015; Gomes, 2009; Noronha et al., 1991; Serufo-Filho, 2014). A recente padronização de codificação da ANS no sistema privado e suplementar seguiu o foco nos principais procedimentos, e, portanto, há limitações de identificação de comorbidades e parametrização de riscos clínicos.

De fato, uma investigação de Veras e Martins (1994 apud La Forgia e Couttolenc, 2009: 135), apontou que apenas 19% das comorbidades tinham registros em AIH, enquanto apareciam em 42% dos prontuários.

Em dois estudos recentes, o Projeto DRG-Brasil, de um grupo de pesquisa da UFMG³³, aplicou análises de *índices de casemix* (R. C Couto, 2014; Serufo-Filho, 2014). Em uma análise de produtividade, em mais de 145 mil pacientes em 116 hospitais privados, os casos de internações desta amostra apresentavam comorbidades em 78,0% em DRG clínicos, sendo que 20,8% deles apresentam três ou mais diagnósticos, além de comorbidades em 52,6% em DRG cirúrgicos (R. C Couto, 2014; Serufo-Filho, 2014). Este grupo iniciou a ampliação destas análises para o contexto de hospitais do SUS com resultados a serem divulgados.

A metodologia DRG exige uma codificação baseada em grupos de diagnósticos para mensurar o uso relativo de recursos (custos), as categorias de doenças, e adequação dos pagamentos por riscos diferenciais. Por sua vez, ambos os sistemas de registros de procedimentos de saúde no Brasil (SIH/AIH do SUS e CIHA da ANS) são de caráter prospectivo, pela padronização e agrupamentos de serviços e insumos individuais em casos ou procedimentos mais abrangentes, sem uma preocupação com o uso de recursos, custos dos procedimentos ou gestão de resultados clínicos.

Enquanto os DRG's incluem o ajuste de risco de forma explícita, comorbidades e índices de *casemix*, o sistema SIH/AIH do SUS, na conformação atual, não permite qualquer mecanismo de ajuste de risco. As diretrizes da AIH são rígidas. Por exemplo, em caso de complicações do tratamento ou excesso de permanência na internação, na prática, os prestadores de serviços emitem uma nova guia AIH, sem vinculação a inicial, para garantir o pagamento adequado, além de frequentes falhas nos registros de comorbidades.

33 UFMG – Universidade Federal de Minas Gerais

Assim, muitas vezes, os prestadores ajustam a codificação da AIH para códigos não adequados ao diagnóstico adequado da doença. Dessa forma, agrega-se mais distorção no sistema de classificação de pacientes no sistema SIH/AIH. Isto resulta em informações inconfiáveis e imprecisas quanto à classificação dos tratamentos, comorbidades, gravidade e riscos dos pacientes atendidos no SUS.

Porém, recentemente, iniciou-se um amplo debate no Ministério da Saúde, quanto a necessidade de estratégias robustas de integração de dados, inclusive quanto a adoção do DRG como estratégia de otimização do SUS (Brasil, 2015f).

Experiências na Espanha e Portugal demonstram a adoção do DRG na origem dos sistemas de informação hospitalar, apontando maior eficiência por utilizar variáveis que diferenciam pacientes segundo a gravidade da doença. Outra característica identificada nos sistemas destes países é a composição de dados clínicos e administrativos, que originaram o Conjunto Mínimo de Dados (CMD), um “conjunto mínimo de variáveis que devem ser recolhidas em cada hospitalização e que permitem a coleta de informação assistencial ‘válida, confiável e oportuna’” (Pinto, 2010b).

Por sua vez, as tabelas (SIA/SUS e SIH/AIH) de reembolso por procedimentos do SUS, historicamente, foram formatadas de forma pulverizadas. A fim de consolidar os códigos de procedimentos do SUS, a tabela SIGTAP (Tabela de Procedimentos, Medicamentos, Órteses, Próteses e Materiais Especiais – OPM) oficializada pela portaria nº 2.848/2007 (Brasil, 2015f).

Em virtude das limitações do SIH/AIH pode-se propor a análise da viabilidade de proposição de um *índice de casemix* simplificado que discrimine a severidade dos casos e permita a comparabilidade de resultados entre unidades de saúde. Em seguida, este *índice de casemix* proposto pode ser usado, por exemplo, para testar algumas hipóteses de correlações quanto aos índices de *eficiência clínica* (resultados clínicos alcançados, tais como: taxa de mortalidade, taxa de reinternação de pacientes, taxa de complicações cirúrgicas, taxa de ocupação de leitos, etc, parametrizadas por um determinado *índice casemix*), para analisar se têm iguais relevâncias para os três níveis de decisores deste estudo.

No entanto, os resultados da pesquisa da literatura (tabela 110, anexo 1) quanto a aplicação de métodos científicos de medição da efetividade pela aplicação de *índices de casemix* ou *Diagnosis Related Groups (DRG)* no âmbito da saúde do Brasil foram incipientes, realidade reforçada na fase empírica da investigação visto que os decisores chaves basicamente identificaram alguns indicadores operacionais de medidas de efetividade, sem apontar métodos de análise técnica da efetividade, por exemplo, pelo uso de *índices de casemix* ou DRG. Diante disso, decidiu-se limitar esta investigação à análise de relevância destes indicadores de efetividade sugeridos na fase empírica, além de definir os seus respectivos pesos

(importâncias/relevâncias), conforme os critérios dos decisores, sem aprofundar a análise técnica da efetividade, e, sugerir estudos futuros sobre modelo de cálculo e análise da efetividade, a fim de propor instrumentos de medição dos aspectos de efetividade entre as organizações da saúde.

De acordo com a visão inicial de utilização dos Artefatos gerados nesta investigação, conforme a relevância atribuída pelos decisores envolvidos e as limitações técnicas nos conhecimentos dos especialistas e gestores, decidiu-se limitar a revisão Conceitual ao enquadramento teórico-conceitual (*framework*) dos indicadores e dimensões de performance e, Em seguida, às bases teórico-conceitual (*framework*) da avaliação e medição da qualidade percebida em saúde, a fim de gerar uma visão aplicada dos resultados desta investigação.

2.4.3. O enquadramento teórico-conceitual (*framework*) dos indicadores e dimensões de performance da investigação

Em busca de delimitar um quadro teórico-conceitual (*framework*) do escopo da avaliação de performance organizacional desta investigação, a partir da análise da pesquisa integrativa da literatura (tabela 110, anexo 1) matriciado pelas respostas iniciais dos painéis de especialistas e profissionais que, inicialmente, prospectou os níveis de conhecimentos de especialistas e profissionais da amostra nos serviços públicos de saúde do Brasil (tópicos 5.2.1 e 6.1.1) decidiu-se delimitar, o escopo da avaliação de performance de organizações de saúde, no âmbito deste trabalho, de forma a considerar a avaliação de performance de serviços e unidades de saúde enquanto um modelo multidimensional, mediada por indicadores e dimensões nas categorias mais relevantes da literatura desde que houvesse referências (mesmo que indiretas) dos indivíduos da amostra dos painéis de especialistas, por exemplo, quanto as dimensões: produtividade assistencial, eficiência, efetividade, impacto, qualidade percebida (satisfação) dentre outras. Em seguida, durante a fase de aplicação de instrumentos validados nos grupos pilotos de gestores e profissionais percebeu-se a necessidade de incorporar outros indicadores e dimensões de performance no instrumento de pesquisa de campo, sendo necessário a expansão do quadro teórico-conceitual (*framework*) definitivo que a seguir será apresentado.

Neste propósito, investigou-se na literatura relevante (quadros 17 a 19, anexo 2) a importância de cada uma dos indicadores e dimensões de performance, bem como, buscou-se identificar quais as metodologias e instrumentos de mensuração de performance em saúde mais utilizadas nos contextos mundial e brasileiros, como forma de considerar a possibilidade de

utilização dos mesmos na avaliação de performance em serviços de saúde para posterior validação de grupos pilotos e de campo da amostra selecionada.

A fim de garantir a possibilidade de construção e usabilidade dos indicadores selecionados e validados, estudou-se as disponibilidades de fontes de dados de informações publicizadas no contexto brasileiro, e/ou nos gabinetes de Controle das organizações de saúde. Assim, os Sistemas de Informações em Saúde (SIS) destacam-se enquanto fontes de informações para a composição de indicadores em saúde (tópico 3.3). Portanto, estabeleceu-se outro parâmetro delimitador do escopo desta investigação, a fim de garantir o acesso aos dados para a composição dos indicadores e dimensões validadas, conforme a relevância apontada pelos gestores dos níveis estratégicos de decisão do sistema de saúde.

Além disso, para conceder legitimidade da aplicabilidade destes indicadores e dimensões, procedeu-se a análise da importância/relevância e usabilidade de cada um dos indicadores e dimensões pela validação dos decisores chaves. A fim de sedimentar uma teoria de base para o envolvimento dos indivíduos na validação destes itens de avaliação da performance, estudou-se a teoria dos *stakeholders* (partes interessadas) enquanto arcabouço teórico para construir o envolvimento dos principais agentes de decisão (*key stakeholders*) das organizações e serviços de saúde deste estudo (tópico 4).

Neste contexto, cabe uma questão norteadora: qual seria o papel e a importância dos indicadores dentro do sistema de informação em saúde? Para Sheldon (1998 apud Da Silva et al., 2008: 50), indicadores de desempenho (*performance indicator; clinical indicator*) são medidas indiretas da qualidade utilizadas como instrumento de monitoramento para salientar os processos, serviços ou profissionais que podem estar apresentando problemas e que necessitam de uma avaliação mais direta. São informações expressas através de um evento, uma taxa ou uma razão.

De acordo com Travassos *et al.*(1999), é necessário destacar a importância do desenvolvimento de estratégias para permitir que os resultados das avaliações melhorem a efetividade dos cuidados hospitalares. Assim, o monitoramento de indicadores pode despontar como uma estratégia de monitoramento e promoção da efetividade da assistência, bem como da busca pela eficiência organizacional em saúde. Mas, os indicadores podem ser usados também como balizadores da eficiência alocativa de recursos e monitoramento das despesas dos serviços de saúde e, ainda, como medidas de qualidade assistencial e gerencial.

Segundo Da Silva *et al.*(2008), cabe ainda destacar que bons indicadores são os que apresentam forte associação, estabelecida com base no conhecimento científico, entre o processo de cuidado e a medida de resultado utilizada (validade causal). Entretanto, limitações

no conhecimento sobre a eficácia dos procedimentos diagnósticos e terapêuticos restringem a determinação da validade causal de diferentes indicadores de resultado (Brook & McGlynn, 1991).

Existem diversos parâmetros usados para a medição da performance em serviços de saúde. No entanto, destacam-se indicadores ou taxa de mortalidade, notadamente com um viés de mapear a quantidade de mortes evitáveis, que teoricamente podem estar relacionadas às infecções hospitalares, o uso inadequado de medicamentos e de outras tecnologias, as falhas na supervisão, os erros durante a cirurgia ou as altas inapropriadas, entre outras causas (OTA, 1988).

Noutra perspectiva, Murray & Frenk (2000b) lembram que a diversidade dos modelos propostos para medir o desempenho do sistema de saúde (Aday, 1993; AHCPR, 1999; Hoffmeyer & Kluwer, 1994; C. W. Hsiao, 1995; Knowles, Leighton, & Stinson; OECD; OECD; Organization, 1999) comprovam a importância dada a estruturação de uma base conceitual para a mensuração da performance do sistema de saúde, que podem ser influenciados por vários fatores, por exemplo, diferenças no *design* (desenho), conteúdo e gestão dos sistemas de saúde se traduzem em diferenças em uma gama de resultados socialmente valorizados, como saúde, capacidade de resposta ou justiça distributiva, dentre outros.

Um quadro teórico-conceitual (*framework*) coerente e consistente tem sua construção iniciada pela definição do escopo os aspectos que se pretende mensurar quanto a performance dos sistemas de saúde. Uma vez que os objetivos intrínsecos dos sistemas de saúde tenham sido claramente articulados, a medição desses objetivos, o conceito de desempenho e a noção distinta dos fatores chaves que influenciam o desempenho devem ser explorados (Murray & Frenk, 2000b). Quanto a seleção de dimensões e indicadores de performance, do sistema de saúde lembram que:

Frequentemente revela-se uma tensão entre os grupos focados no conteúdo técnico dos serviços de saúde (por exemplo, imunizações ou unidades de terapia intensiva) e aqueles focados nos arranjos institucionais do sistema de saúde (por exemplo, mecanismos de pagamento de provedor ou seguro social). É necessário fornecer uma estrutura para pensar sobre as dimensões dos sistemas de saúde que podem influenciar o desempenho e operacionalizar as medidas dessas dimensões. Este tópico é tratado mais extensivamente em outros lugares (Frenk; Kutzin; Londono & Frenk, 1997; Murray & Frenk, 2000b: 12).

Nesta investigação o Objetivo foi estabelecido desde seu princípio enquanto a mensuração de performance de serviços e unidades de saúde do SUS, sendo que as dimensões

e indicadores foram definidos pela importância/relevância atribuída na pesquisa de campo, a partir de um modelo básico proposto para a análise e valoração dos indivíduos da amostra.

Portanto, a fim de elaborar uma proposta de quadro teórico-conceitual (*framework*) para uma metodologia robusta de avaliação de performance e a seleção de indicadores mais relevantes e destacados na literatura foram estudados alguns dos principais sistemas mundiais de avaliação de performance em organizações de saúde, como por exemplo, quadros conceituais de projetos que englobem indicadores de performance e/ou medição da qualidade de alguns países e organizações multilaterais (Reino Unido, Canadá, Austrália, EUA, OMS e OCDE) citados como relevantes em outro estudo (Arah, Klazinga, et al., 2003):

Examinamos o sistema de saúde, estruturas de desempenho, indicadores, qualidade e documentos de gestão publicados, principalmente na Internet, sobre ou pelo Reino Unido, Canadá, Austrália, EUA, OMS e OCDE.

[...] A primeira estratégia consistia em utilizar os motores de busca e hiper ligações localizados nos sites oficiais dos departamentos ou ministérios ou agências de saúde de cada país e organização: Reino Unido (<http://www.doh.gov.uk>), Canadá (<http://www.hc-sc.gc.ca>; <http://www.cihi.ca>), Austrália (<http://www.aihw.gov.au>), EUA (<http://www.health.gov>), OMS (<http://www.who.int>) e OCDE (<http://www.oecd.org>) (Arah, Klazinga, et al., 2003, p. 378).

De forma mais instrumental, estudou-se as bases das principais metodologias e sistemas mundiais de medição de resultados em saúde, como por exemplo, quadros teóricos-conceituais (*frameworks*) de alguns projetos em destaque: *Quality Indicator Project* (QIP) da Agência de Pesquisa e Qualidade nos Serviços de Saúde – *U.S. Agency for Healthcare Research and Quality* (AHRQ), com destaque no projeto da Fundação Europeia para Gerenciamento da Qualidade - *European Foundation for Quality Management* (EFQM), além de sistema de indicadores do *Project Assessment Tool of Quality Improvement in Hospitals* (PATH) da Organização Mundial de Saúde (OMS); no Reino Unido, duas abordagens. A primeira baseada na definição de indicadores do Serviço Nacional de Saúde-NHS, a segunda interessada na concretização de uma metodologia para estabelecer “*hospital ratings*”; em Portugal, o *Sinas@Hospitais* sob a coordenação da Entidade Reguladora de Saúde (ERS), que classifica os hospitais avaliados (*ranking*) que é utilizado pela Administração Central do Sistema de Saúde (ACSS), entre outros que surgirem como relevantes na revisão integrativa da literatura (anexo 2).

No escopo delineado por esta investigação, a seguir, destacaram-se algumas experiências relevantes mundiais e do Brasil a fim de construir um quadro teórico-conceitual (*framework*) para uma proposta inovadora de um *proxy* de painel de *indicadores e dimensões* de performance no âmbito de organizações, unidades e serviços de saúde pública do Brasil.

Ao nível mundial pode-se dizer que os primeiros esforços estruturados de medição de performance direcionados às organizações hospitalares e serviços de saúde foi iniciado, em 1985, no *Center for Performance Sciences* (CPS), aplicado em 7 hospitais de *Meryland-EUA*, que hoje congrega mais de 200 hospitais e 1.950 unidades de saúde, em nove países (Áustria, Canadá, Bélgica, Alemanha, Holanda, Portugal, Singapura, Reino Unido e Taiwan), além dos EUA, no Projeto de Indicadores de Qualidade - *Quality Indicator Project* (QIP), referido por Kazandjian (1999; 2003), com 250 medidas de internação e ambulatorial, mas os indicadores mais populares referem-se ao manejo do parto, mortalidade hospitalar e perioperatória, readmissões não programadas ao centro cirúrgico, internação, unidade de terapia intensiva (UTI) e emergência. Ainda nos Estados Unidos, a Agência de Pesquisa e Qualidade nos Serviços de Saúde – *U.S. Agency for Healthcare Research and Quality* (AHRQ), tem desenvolvido grupos de indicadores oficiais de rastreamento da qualidade em saúde, elaboração de relatórios públicos e a contratualização baseada em performance, além de indicadores para serviços de saúde, em quatro áreas: atendimento preventivo, atendimentos a pacientes em tratamento, segurança do paciente e pediatria.

Na Europa, em 1988, surgiu outra tentativa relevante de avaliação de resultados em saúde, através da Fundação Europeia para Gerenciamento da Qualidade - *European Foundation for Quality Management* (EFQM) (EFQM, 1999), com um modelo não prescritivo, baseado em nove critérios: Liderança, Pessoas, Política e Estratégia, Parcerias e Recursos, Processos, Resultados nas Pessoas, Resultados para o Cliente, Resultados para a Sociedade, e Resultados Chaves do Desempenho. Cada critério do modelo EFQM inclui um número de subcritérios (total de 32) que são desdobrados em 172 itens de avaliação e medições. Os nove critérios estão agrupados nos critérios “*enabler*” e “*result*”. Alinhados com a trilogia de Donabedian (Donabedian, 1980d), os facilitadores (*enabler*) cobrem o processo, a estrutura e os meios de uma organização. As quatro dimensões do resultado (*result*) são divididas em oito subcritérios, que exigem medidas objetivas, dados e fatos, permitindo a comparação de desempenhos com outras organizações (Walburg, Klazinga, & Nabitiz, 2000). Quando adaptado para os sistemas de saúde a abordagem EFQM resultou em uma proposta abrangente e adaptável em diversos contextos:

O Modelo EFQM mostra muitos paralelos com o modelo de avaliação do Prêmio americano Malcolm Baldrige (W. L. Hart & Bogan, 1992; "Malcolm Baldrige National Quality Award, 1999 Criteria for Performance Excellence," 1999), o australiano, o Prêmio de Qualidade Sul-Africano e Prêmio Deming no Japão. [...] Segundo a abordagem EFQM fornece um quadro mais amplo e genérico do que a maioria das abordagens tradicionais de cuidados de saúde. Sendo genérica, não entra

em padrões e normas específicas para cuidados de saúde como os Sistemas Europeus de Acreditação, como o King's Fund (Fund, 1998) and PACE (Gennip, 1998) or the Joint Commission for Accreditation of Healthcare Organizations in North America (J. Commission, 1998). A Abordagem EFQM é geral e se alinha conceitualmente com as idéias formuladas por Donabedian (Walburg et al., 2000: 192).

Enquanto isso, Walburg et. al. (2000) lembram que diversos países europeus, aplicam adaptações do modelo EFQM para autoavaliação por organizações de saúde e oferecem uma breve evolução das adaptações efetivadas nos principais países europeus, outros muitos autores relatam experiências de adaptação a países específicos, por exemplo, na Alemanha (Moeller, 2001), Valejo et. al. (2006) destacam diversos estudos sobre as adequações do modelo EFQM que concluem que o EFQM é amplamente aplicável aos cuidados de saúde (Arcelay, Sanchez, & Hernandez, 1999; Holland & Fennell, 2000; Jackson, 1999; Klazinga, 2000; Mira, Lorenzo, Rodríguez-Marín, Aranaz, & Sitges, 1998; Moeller, 2001; Simón, Guix, Nualart, Surroca, & Carbonell, 2001).

A avaliação de performance dos serviços de saúde foi o objeto central do *World Health Report 2000 - WHR 2000* (WHO, 2000b). Esta proposta de indicadores unificados foi muito criticada. Em seguida, diversos países membros da OCDE (WHO, 2003) iniciaram o desenvolvimento de sistemas de monitoramento específicos, tendo a medição de resultados (*outcomes*), quase sempre como uma preocupação do ponto de vista de avaliação (F Viacava et al., 2004).

O modelo de avaliação de performance em saúde proposto pela OMS (WHO, 2000b) teve suas limitações fortemente debatidas. Os principais questionamentos se referem aos problemas relativos ao modelo teórico aplicado (Braveman, Starfield, & Geiger, 2001; V. Navarro, 2000; Williams, 2001) e outro conjunto de problemas que se refere aos aspectos metodológicos da mensuração de desempenho adotada quanto as dimensões da avaliação (Almeida et al, 2001).

Deste intenso debate, surgiu uma proposta alternativa de avaliação de desempenho dos países da OCDE, caracterizada, sinteticamente, através dos estudos de Hurst e Hughes (2001 apud F Viacava et al., 2004: 715-716):

a) incluir indicadores de resultados dos serviços de saúde (*outcomes*) como parte da avaliação de desempenho (*eficiência microeconômica*); b) incluir acesso como componente da responsividade, possibilitando avaliação da equidade; c) incluir o nível de gasto sanitário como meta dos sistemas de saúde (*eficiência macroeconômica*); d) excluir qualquer ponderação para avaliação de metas; e) referir a *avaliação de desempenho* a várias e distintas dimensões dos sistemas de saúde; e f) considerar a avaliação de desempenho restrita à assistência médica como oposta às atividades de saúde pública ou quaisquer outras mais amplas.

Viacava *et al*(2004) fizeram uma ampla revisão integrativa e destacaram que vários países membros da OCDE estão desenvolvendo quadros de referência e indicadores de avaliação de performance dos sistemas de saúde (Cozzens, 1995; Hurst & Hughes, 2001; AIHW, 2000; NHPC, 2002; CIHI, 2001; Hurst, 2002; OR, 2002; Smee, 2002; Wolfson e Alvarez, 2002). Estes sistemas específicos de avaliação de performance na saúde foram construídos sobre os princípios gerais das propostas da OMS e da OCDE.

Os sistemas de avaliação que vêm sendo implementados por esses países têm como características comuns: a) a definição de quadros de referência (objetivos e metas) e distintas dimensões de desempenho (sobretudo em relação à qualidade e eficiência); b) a preocupação com a facilidade de operacionalização dos conceitos enquanto as organizações internacionais (como OMS e a OECD) se preocupam mais com definições em nível macro; c) dar maior ênfase às dimensões de estrutura e processo como *proxy* de resultados; d) não inclusão de medidas mais gerais de eficiência.

No Reino Unido, existem basicamente duas grandes abordagens. A primeira baseada na definição de indicadores do Serviço Nacional de Saúde-NHS (Chang, Stephen, & Northcott, 2002), a segunda interessada na concretização de uma metodologia para estabelecer “*hospital ratings*” (NHS, 2002). Ambas englobam indicadores de *estrutura*, de *processo* e de *resultados* (a trilogia de *Avendis Donabedian*). Também há o trabalho do *The King’s Fund*,³⁴(Fund, 1998) uma fundação independente que recebe recursos públicos e que desenvolve projetos de análise dos sistemas de saúde, a partir de dados disponíveis nos sistemas do NHS, e executa projetos de medições de performance nos hospitais. Há outra organização independente, o *National Institute for Health and Clinical Excellence* (NICE)³⁵ que executa avaliação de tecnologias de saúde e estudos de avaliação econômica, para a governança da incorporação de novas tecnologias, inclusive com proposição de protocolos e diretrizes clínicas.

No âmbito do NHS, uma estrutura estatal foi implantada, a Comissão de Melhoria em Saúde – *Commission of Health Improvement* (CHI)³⁶ que desenvolve indicadores, coleta dados e faz a disseminação de informações de performance comparativa entre prestadores de serviços do NHS. Assim, o CHI desenvolveu um sistema de classificação da performance hospitalar em sete dimensões: gerenciamento de risco, auditoria clínica, pesquisa e educação, envolvimento do paciente, envolvimento dos funcionários, gestão da informação, treinamento e

³⁴The King’s Fund - Fundação independente, que desenvolve ações de melhorias no âmbito do NHS - Reino Unido, com recursos governamentais

³⁵ NICE - National Institute for Health and Clinical Excellence, que executa avaliação de tecnologias de saúde e estudos de avaliação econômica no âmbito do NHS do Reino Unido.

³⁶ CHI - Commission of Health Improvement(Comissão de Melhoria em Saúde), no âmbito do NHS do Reino Unido.

desenvolvimento. O CHI divulga anualmente o *ranking* dos hospitais e aqueles que não atingem nenhuma estrela (nível de desempenho mínimo) devem apresentar um plano de ação que será monitorado pelo CHI.

Recentemente, no Reino Unido, foram implantadas duas estratégias relevantes em busca da melhoria da qualidade da saúde que são partes de um acordo de performance do setor de saúde: o Marco Nacional para Serviços - *National Service Framework* (NSF)³⁷, que são padrões e estratégias desenvolvidos por especialistas sob o patrocínio do NICE, a fim de executarem desenvolvimentos e implementações em áreas específicas da saúde e estabelecer medidas de desempenho para julgar o progresso dos serviços de cuidados específicos (McLoughlin, Leatherman, Fletcher, & Owen, 2001); e o Marco Nacional de Performance - *National Performance Framework* (NPF)³⁸, que inclui indicadores de performance para mensuração e comparação de performance das regiões, hospitais e grupos de serviços básicos de saúde em seis áreas prioritárias: melhorias da saúde, acesso, atendimento resolutivo e adequado, eficiência, satisfação do paciente e, resultados em saúde (Campbell et al., 2000).

De forma objetiva, Arah *et. al.*(2003) apresentam uma comparação global das diversas abordagens de quadro teórico-conceitual (*framework*) dos sistemas de medição de performance do Reino Unido (NHS), do Canadá, da Austrália, dos EUA, do relatório *World Health Report 2000*(WHO, 2000a), além das publicações *Health at a Glance* e *OECD Health Data* da OECD (Development, 2001; Unit, 2001).

Quanto ao sistema do NHS (Arah, Klazinga, et al., 2003) caracterizam as bases conceituais e contextuais para a estruturação do quadro teórico (*framework*) dos sistemas de performance destes países, apresentados na quadro 17 (anexo 2). Em seguida, especificaram o quadro teórico de desempenho do NHS, *Performance Assessment Framework* (PAF) do NHS. Conceitualmente, a estrutura do PAF do NHS é baseada em uma abordagem de *balanced scorecard*, sendo um sistema unificado de medição, avaliação e recompensa, projetado para fornecer uma visão mais abrangente do desempenho do NHS (Health, 2001), no qual "o conjunto geral de indicadores deve fornecer uma imagem equilibrada do desempenho da organização, refletindo os principais aspectos, incluindo os resultados e a perspectiva dos usuários" (A. Commission, 2000). A iniciativa engloba um conjunto indicadores nacionais distribuídos por áreas ou dimensões de desempenho.

37NSF - *National Service Framework* (Marco Nacional para Serviços) , no âmbito do NHS do Reino Unido.

38NPF - *National Performance Framework* (Marco Nacional de Performance), no âmbito do NHS do Reino Unido.

Seis áreas de desempenho são identificadas dentro da autoridade de saúde PAF, a saber: (a) melhoria da saúde; (b) acesso suficiente; c) entrega efetiva de cuidados de saúde adequados; (d) eficiência; (e) experiência do paciente / cuidador; e (f) resultados de saúde dos cuidados do NHS (*D. o. Health, 2001b*). Ao nível do subsistema NHS Trust, o quadro tem apenas quatro áreas principais de desempenho: (a) eficácia e resultados clínicos; (b) eficiência; (c) experiência do paciente / cuidador; e (d) competência e capacidade. Dentro do PAF, um conjunto de 'Indicadores Nacionais de Desempenho do NHS' dá um resumo das atividades do NHS, que aborda uma ampla gama de questões como saúde mental, tratamento do câncer, listas de espera, acesso aos médicos GPs, saúde geral da população e pessoal (Health, 2002). Este conjunto relativamente pequeno de indicadores nacionais, publicados anualmente, será apoiado por “indicadores de *benchmarking*”, que contêm informações contextuais mais específicas, identificando assim um bom ou mau desempenho em todas as áreas do NHS (Health, 2001, 2002)(Arah, Klazinga, et al., 2003: 380).

No Canadá, o desempenho do sistema de saúde faz parte do *Canadian Health Information Roadmap Initiative Indicators Framework*, conforme informa o *Canadian Institute for Health Information* (CIHI, 2000). Segundo Arah *et. al.*(2003) essa Iniciativa de Roteiro sedimenta quatro dimensões e diversos indicadores de performance (PIs).

Essa Iniciativa do Roteiro, que se baseia no modelo de saúde da população, conceitua o seu quadro de indicadores de saúde em termos de fornecimento de informação comparativa de alta qualidade em quatro dimensões: (a) estado de saúde; (b) determinantes não médicos da saúde; c) desempenho do sistema de saúde; e (d) características da comunidade e do sistema de saúde (CIHI, 1999, 2000).

[...] Os oito domínios do desempenho do sistema de saúde neste quadro são: (a) aceitabilidade; (b) acessibilidade; (c) adequação; (d) competência; (e) continuidade; (f) eficácia; (g) eficiência; e (h) segurança (Acreditação, 1996; CIHI, 2000). Atualmente, essas dimensões, exceto aceitabilidade, competência e continuidade, parecem ser apoiadas por vários PIs. Sob as características de comunidade e saúde, os PIs são agrupados em outros que fornecem informações contextuais sobre: (a) comunidade; (b) sistema de saúde; e (c) recursos.

[...] Alguns dos indicadores empregados na estrutura canadense para capturar o domínio de desempenho da eficiência incluem: o respectivo número de novos casos de coqueluche, sarampo, HIV e clamídia relatados em um determinado ano; taxas de internação por pneumonia e influenza padronizadas por idade aos 65 anos ou mais; taxas de mortalidade padronizadas por idade devido a doenças tratáveis clinicamente, como câncer do colo do útero, pneumonia e bronquite não especificada; e taxas de hospitalização de cuidados agudos padronizados por idade para condições sensíveis à atenção ambulatorial. Alguns desses indicadores também capturam domínios de acessibilidade e adequação, como é o caso das condições sensíveis ao atendimento ambulatorial (Arah, Klazinga, et al., 2003: 384).

Segundo Arah *et. al.*(2003), na Austrália, a estrutura de desempenho nacional é baseada em um modelo de determinantes de saúde, é não hierárquica, mas relacional, pois presta atenção a outras variáveis contextuais que podem influenciar consideravelmente os insumos, processos, produtos ou resultados dos serviços de saúde. De forma objetiva, dizem que:

O quadro teórico-conceitual (*framework*) tem três níveis: (a) estado de saúde e resultados; (b) determinantes da saúde; e (c) desempenho do sistema de saúde. A

equidade é vista como uma propriedade abrangente desses três níveis (Committee, 2001). Condições de saúde, função humana, expectativa de vida e bem-estar e mortes são as quatro dimensões contidas na seção "estado de saúde e resultados" da estrutura. O nível de "determinantes da saúde" inclui fatores ambientais e socioeconômicos, capacidade da comunidade, comportamentos de saúde e dimensões relacionadas à pessoa.

[...] O desempenho do sistema de saúde possui nove dimensões, a saber: (a) efetividade; (b) adequação; (c) eficiência; (d) responsividade; (e) acessibilidade; (f) segurança; (g) continuidade; (h) capacidade; e (i) sustentabilidade. (Arah, Klazinga, et al., 2003: 386).

Quanto aos EUA, Arah *et. al.*(2003) descrevem diversas iniciativas e modelos de monitoramento e avaliação, sintetizadas a seguir:

a) uma estrutura proposta de melhoria do sistema nacional de saúde (Medicine, 2001); (b) um modelo de melhoria da saúde da população (Services, 2000); (c) o *National Health Care Quality Report* (Medicine, 2001); e (d) duas proeminentes iniciativas de relato de desempenho amplamente utilizadas, mas não nacionais - o *Consumer Assessment of Health Plans Study* (CAHPS) e o *Health Employer Data Information Set* (HEDIS[®]) (Reilly, Zema, & Crofton, 2002).

b) [...] o programa *Quality Improvement Organization* (QIO) dos Centros de Serviços *Medicare* e *Medicaid*, a iniciativa ORYX da *Joint Commission on Accreditation of Healthcare Organizations* e as medidas de qualidade da *Foundation for Accountability*.

c) [...] O *Institute of Medicine* (IOM) propôs uma estrutura para a melhoria do sistema nacional, delineado por seis domínios de desempenho: (a) segurança; (b) eficácia; (c) centrada no paciente; (d) tempestividade; (e) eficiência; e (f) patrimônio líquido. (Medicine, 2001a). Os PIs que podem preencher essa estrutura de reinvenção de sistemas estão disponíveis na *National Quality Measures Clearinghouse*™ (NQMC™), um repositório nacional de indicadores mantido pela *Agency for Healthcare Research and Quality* (AHRQ)(Arah, Klazinga, et al., 2003: 387-88).

Noutra vertente, Arah *et. al.*(2003) apresentam o quadro teórico-conceitual (*framework*) de avaliação de performance do sistema de saúde proposto pela OMS (WHO, 2000a) que pauta-se em três objetivos de um sistema de saúde: (a) melhorar a saúde; (b) aumentar sua capacidade de resposta às expectativas da população; e (c) assegurar a equidade da contribuição financeira (WHO, 2000b). Em seguida, distribui-se componentes de avaliação em cada um dos objetivos:

O Objetivo de "melhorar a saúde" é dividido em dois: (a) estado de saúde médio; e (b) reduzir as desigualdades em saúde. A "capacidade de resposta" inclui dois componentes principais: (a) respeito pelas pessoas; e (b) orientação do cliente. Para o terceiro Objetivo, a justiça na contribuição financeira, apenas sua distribuição, mas não seu nível médio, é considerado. O nível de recursos é tratado como uma variável extrínseca para avaliar o desempenho. Os níveis de obtenção de metas para melhoria da saúde e responsividade refletem a qualidade geral do sistema de saúde, enquanto as medidas de distribuição (saúde, responsividade e contribuição financeira) descrevem a equidade do sistema (Murray & Frenk, 2000a). Assim, os cinco principais componentes do desempenho do sistema de saúde segundo a OMS são: o nível geral de saúde da população, a distribuição da saúde na população, o nível geral de responsividade, a distribuição da capacidade de resposta na população e a distribuição dos sistemas de saúde. carga financeira dentro da população. O desempenho é estimado a partir da soma ponderada dos cinco componentes listados

acima, e um "indicador geral de realização" é calculado (Arah, Klazinga, et al., 2003: 389).

Ainda neste prisma, Arah *et. al.*(2003) apresentam o quadro teórico-conceitual (*framework*) de avaliação de performance do sistema de saúde proposto pela OECD (OECD, 2001) com três objetivos principais: (a) melhoria da saúde e resultados; (b) capacidade de resposta e acesso; e (c) contribuição financeira e despesas com saúde (Hurst, 2002; OECD, 2001).

Essas metas são submetidas a dois "componentes para avaliação" para explorar o desempenho do sistema, a saber: o nível médio e a distribuição de cada meta. Como resultado, seis componentes do desempenho do sistema de saúde emergem dessa matriz, onde os níveis médios das três metas mencionadas são tomados para medir a eficiência e suas respectivas distribuições indicam a equidade. Diferentemente da estrutura da OMS, onde o acesso é visto como um determinante da capacidade de resposta, aqui o "acesso" é designado como um componente essencial da "capacidade de resposta", de modo a abordar questões de equidade de acesso dentro da estrutura da OCDE. Em termos de dimensões de desempenho, o quadro da OCDE tem quatro principais, assim: (a) melhoria / resultados de saúde; (b) responsividade; (c) equidade (de resultados de saúde, acesso e financiamento, respectivamente); e (d) eficiência (macroeconômica e microeconômica) (Arah, Klazinga, et al., 2003: 389).

Muitos países membros da OECD desenvolvem medições de desempenho e gerenciamento que captam equidade, qualidade e eficiência de seus sistemas de saúde (Hurst, 2000, 2002; OECD, 2001). De forma instrumental, a OCDE compila e publica indicadores-chave nos formatos impresso (*Health at a Glance*) e eletrônico (*OECD Health Data*) (Development, 2001; Unit, 2001), sendo cerca de 10 categorias de indicadores, a saber:

Os 10 principais dados encontrados nestas bases de dados são: estado de saúde, recursos de saúde, utilização de saúde, gastos com saúde, financiamento e remuneração, proteção social, mercado farmacêutico, determinantes não médicos de saúde, referências demográficas e referências econômicas. Os dados para os muitos indicadores contidos nesses dados são provenientes de países membros, do banco de dados do *Eurostat New Cronos*, do *WHO's World Health Statistics Annual* da OMS, do *WHO-Europe Health for All* e de outros bancos de dados internacionais (Arah, Klazinga, et al., 2003: 391).

Além das diversas abordagens por quadro teórico-conceitual (*framework*) acima expostas, Arah *et. al.*(2003) apresentam de forma global os indicadores de eficácia, iniciativas de melhoria de qualidade e os mecanismos de mudança do sistema de saúde de cada um dos sistemas de avaliação de performance analisados, não sendo sua análise escopo desta investigação.

Outro modelo teórico-conceitual (*framework*) de destaque na literatura (Arah et al., 2006; Mattke, 2004; Mattke, Epstein, & Leatherman, 2006; Mattke, Kelley, & Scherer, 2004) trata-se do projeto *Health Care Quality Indicators* (HCQI) que envolveu, desde 2003, painéis de especialistas de 23 países signatários da OECD, organizações internacionais como *World Health Organization* (WHO) e *European Commission* (EC), organizações especializadas como *International Society of Quality in Healthcare* (ISQua) e *European Society for Quality in Healthcare* (ESQH) além de diversas universidades, em busca de desenvolver e implementar um quadro de indicadores de qualidade a nível internacional (Mattke et al., 2006). Os painéis de especialistas selecionaram e caracterizaram as dimensões e indicadores específicos em cinco áreas prioritárias: atendimento cardíaco (Lambie & Mattke, 2004), diabetes mellitus (Greenfield, Nicolucci, & Mattke, 2004), saúde mental (Hermann & Mattke, 2004a), segurança do paciente (Millar & Mattke, 2004), atenção primária, prevenção e promoção da saúde (Marshall, Leatherman, & Mattke, 2004). Cada uma destas cinco áreas prioritárias desenvolveu seu próprio quadro teórico-conceitual (*framework*) tendo sido consolidado uma recomendação de 86 indicadores selecionados: nove para diabetes, 12 para saúde mental, 17 para cuidados cardíacos, 21 para segurança do paciente e 27 para cuidados primários e prevenção (Mattke et al., 2006).

Em outro estudo, no âmbito do projeto *Health Care Quality Indicators* (HCQI) da OECD, Arah et al. (2006) propõem um alinhamento conceitual das principais dimensões de performance contidas nas diferentes abordagens de quadro teórico-conceitual (*framework*) dos sistemas de medição de performance do Reino Unido (NHS), do Canadá, da Austrália, dos EUA, do relatório *World Health Report 2000* (WHO, 2000a), além das publicações *Health at a Glance* e *OECD Health Data* da OECD (Development, 2001; Unit, 2001), além dos projetos *European Community Health Indicators Project* (ECHI) e do *The Commonwealth Fund's International Health Indicators Project*, por exemplo, eficácia, adequação, segurança, eficiência, continuidade, acessibilidade, equidade, responsividade, centralidade do paciente e pontualidade:

Eficácia é o grau de alcançar resultados desejáveis, dada a prestação adequada de serviços de cuidados de saúde baseados em evidências a todos os que poderiam beneficiar (Arah, Klazinga, et al., 2003; Donabedian, 1980b; Medicine, 1990, 2001; Veillard et al., 2005)

Adequação é o grau em que a assistência à saúde é relevante para as necessidades clínicas, dada a melhor evidência atual (JCAHO, 1997).

Segurança é uma dimensão em que o sistema tem as estruturas corretas, presta serviços e obtém resultados de maneiras que evitam danos ao usuário, provedor ou ambiente (JCAHO, 1997; Veillard et al., 2005).

Eficiência envolve encontrar o nível coerente de recursos para o sistema e garantir que esses recursos sejam usados para gerar o máximo de benefícios ou resultados (Arah, Custers, & Klazinga, 2003; Donabedian, 2003; Veillard et al., 2005). O quadro da OCDE inclui “despesas com saúde” ou custo como parte da eficiência: o Objetivo “macroeconômico” é encontrar o nível sustentável de gastos com saúde (especialmente de gastos públicos com saúde), e o Objetivo de eficiência “microeconômica” é minimizar gastos para qualquer nível de resultados e capacidade de resposta (Hurst, 2002).

Continuidade aborda até que ponto os cuidados de saúde para usuários específicos, ao longo do tempo, são organizados sem problemas em provedores e instituições, que pode ser medida a partir da perspectiva do paciente e pode acabar como parte da centralização no paciente. A coordenação pode então ser vista como cuidados de saúde sendo organizados de forma suave entre provedores e instituições.

Acessibilidade é a facilidade com que os serviços de saúde são alcançados. O acesso pode ser físico, financeiro ou psicológico e requer que os serviços de saúde estejam disponíveis a priori. Uma dimensão intimamente relacionada, portanto, é a *equidade*, que define o quanto um sistema lida de forma justa com todos os envolvidos. *Equidade* também lida com a distribuição do ônus do pagamento pelos serviços de saúde (Hurst, 2002; WHO, 2000a) e com a distribuição dos cuidados de saúde e seus benefícios entre as pessoas (Donabedian, 1980b, 2003).

Responsividade refere-se a como um sistema facilita as pessoas a satisfazer suas expectativas legítimas de não-saúde (WHO, 2000a) A *centralidade do paciente* capta o grau em que um sistema realmente funciona colocando o paciente/usuário no centro de sua prestação de assistência médica, sendo cada vez mais medido como experiências do paciente de cuidados de saúde com ênfase no cuidado (Arah, Custers, et al., 2003; Medicine, 2001; Veillard et al., 2005). *Responsividade e centralidade do paciente* são muitas vezes consideradas equivalentes.

Pontualidade é o grau em que a assistência médica é fornecida dentro da janela de tempo mais benéfica ou necessária (Donabedian, 2003; JCAHO, 1997). A *pontualidade* pode tornar-se parte da *acessibilidade* ou *capacidade de resposta*, de modo a refletir as experiências do paciente com a prontidão dos cuidados de saúde (WHO, 2000a). *Aceitabilidade* é a conformidade com os desejos, desejos e expectativas dos usuários de cuidados de saúde e suas famílias (Donabedian, 2003). Como tal, a *aceitabilidade* é frequentemente apresentada como parte ou substituto da *centralidade do paciente*, como visto no Canadá (Arah et al., 2006: 8-10).

Há outras diversas dimensões de performance bem como diferentes definições das acima apresentadas, sendo estes conceitos usados apenas quando os informantes da amostra externarem dúvidas, assim, decidiu-se assumir estas definições sem maiores aprofundamentos.

Por sua vez, Vallejo *et al.* (2006) fizeram uma avaliação Conceitual do modelo *European Foundation for Quality Management* (EFQM) no propósito de assinalar os alinhamentos da adaptação da proposta EFQM aos sistemas de saúde. Neste estudo Vallejo *et al.* (2006) destacam que nas experiências de aplicação do modelo EFQM ele é amplamente aceito como aplicável aos cuidados de saúde (Arcelay et al., 1999; Holland & Fennell, 2000; Jackson, 1999; Klazinga, 2000; Mira et al., 1998; Moeller, 2001; Simón et al., 2001). Lembrem, ainda, que o EFQM promove a melhoria na qualidade das organizações (Nabitz & Walburg, 2000) e a qualidade do tratamento dispensado aos pacientes (Jackson & Bircher, 2002) além de outros aspectos positivos assinalados na literatura:

O uso da autoavaliação (Nabitz & Klazinga, 1999), por ser considerada uma atividade motivadora para gestores e profissionais que nela participam (Moeller, Breinlinger, O'Reilly, & Elser, 2000; Simón et al., 2001; Stewart, 2003).

[...] Promoção da melhoria por um sistema simples de identificação de áreas para melhoria (Holland & Fennell, 2000; Jackson, 1999; Moeller, 2001).

[...] Possibilidade de fazer atividades de benchmarking (Arcelay et al., 1999; Moeller et al., 2000; Moeller & Sonntag, 1998).

[...] Validade de face (Moeller, 2001; Walburg et al., 2000).

[...] Flexibilidade de sua estrutura que permite a inclusão de práticas já existentes (Geraedts, Montenarie, & van Rijk, 2001; Jackson, 1999; Pitt, 1999; Stewart, 2003)(Vallejo et al., 2006: 327).

No entanto, Vallejo *et al.*(2006) enfatizam que o EFQM é um modelo genérico e destacam alguns desafios na aplicação do EFQM no setor de saúde.

O modelo não é específico o suficiente para abordar todas as áreas relevantes para a saúde (Moeller, 2001; Walburg et al., 2000).

[...] Mesmo quando os critérios podem ser adequados, os subcritérios devem ser adaptados aos cuidados de saúde (Moracho et al., 2001), especialmente para organizações públicas (Mira et al., 1998).

[...] Linguagem utilizada para descrever o modelo é identificada como uma de suas principais dificuldades (Mira et al., 1998; Stewart, 2003), por ser complexa, pouco clara e distante para os profissionais de saúde acostumados à terminologia clínica (Rodríguez & Ferrándiz-Santos, 2004).

[...] É difícil desenvolver indicadores operativos para avaliar os critérios de resultados em cuidados de saúde, porque os resultados esperados não são especificados (Vallejo et al., 2006: 328).

Por fim, Vallejo *et al.*(2006)destacam os esforços para a adaptação dos conteúdos do modelo, seus critérios e subcritérios, para a atenção à saúde (Holland & Fennell, 2000; Lorenzo, Arcelay, & Bacigalupe, 2001; Nabitz & Walburg, 2000; Vitoria-Gasteiz, 2000), diretrizes para o uso do modelo em organizações de saúde (Nabitz & Klazinga, 1999), estratégias para implementá-lo combinadas com outras abordagens (Brandt, Schmidt, Dziewas, & Groene, 2005), ou indicadores que são específicos para este campo ("Identification of a Set of Key Quality Indicators in the Hospital Setting Using the European Excellence Model ", 2001).

Outro sistema notável foi desenvolvido pelo escritório Europeu da OMS, que lançou, em 2003, um projeto com o Objetivo de desenvolver e disseminar uma ferramenta flexível e abrangente para a avaliação de performance e melhoria da qualidade hospitalar (projeto PATH) com os seguintes focos (Veillard et al., 2005):

(i) definição dos conceitos e identificação das dimensões-chave de desempenho do hospital, (ii) O projeto da arquitetura do PATH para melhorar a gestão baseada em evidências e de melhoria da qualidade, através da avaliação de desempenho; (iii) a seleção de um núcleo e um conjunto de indicadores de desempenho sob medida com definições operacionais detalhados, (iv) a identificação de trade-offs entre os indicadores, (v) elaboração de folhas descritivas para cada indicador para apoiar os

hospitais na interpretação de seus resultados; (vi) projeto de um painel de instrumentos equilibrado, e (vii) as estratégias para a implementação do painel PATH.

A iniciativa *Project Assessment Tool of Quality Improvement in Hospitals* (PATH) objetiva apoiar os hospitais na avaliação de sua performance pela análise de seus próprios resultados, e traduzi-los em ações de melhoria, pelo uso de ferramentas para avaliação de performance e do apoio de uma rede de parceiros dos hospitais participantes.

As etapas de concepção do projeto PATH foram descritas de forma sintética por Veillard *et al.*(2005). Um grupo de especialistas de 11 países anglo-saxônicos analisaram os 100 indicadores pré-selecionados quanto a disponibilidade de indicadores, sua relevância em diferentes contextos nacionais, seu impacto potencial na melhoria da qualidade e o ônus da coleta de dados. A lista de indicadores incluídos no modelo operacional foi restrita a um conjunto final de 24 indicadores principais de desempenho e a um conjunto personalizado de 27 indicadores(Veillard et al., 2005: 492). O conjunto principal foi projetado para permitir *benchmarking* em diferentes contextos mundiais e os personalizados devem adequar-se de forma flexível às diferentes realidades de cada país (Veillard et al., 2005: 493-95). Estes dois conjuntos de indicadores baseiam-se em evidências para uso em hospitais europeus e sugerem formas para seu uso estratégico na avaliação de desempenho hospitalar (Ibrahim, 2001). Neste contexto, sintetizou-se o quadro de indicadores, dimensões e subdimensões do PATH (quadro 18, anexo 2).

Ao final do processo, a partir de uma síntese de diferentes teorias de desempenho organizacional (K. S. Cameron & Whetten, 1983; Sicotte et al., 1998), um instrumento foi estruturado que englobou seis dimensões de performance hospitalar inter-relacionadas do modelo conceitual PATH, a saber:

Eficácia clínica, segurança, centralidade do paciente, governança responsiva, orientação da equipe e eficiência. Duas perspectivas transversais (*segurança e centralidade do paciente*) atravessam quatro dimensões do desempenho hospitalar (*eficácia clínica, eficiência, orientação do pessoal e governança responsiva*). [...] A *segurança* está relacionada à *eficácia clínica (segurança do paciente), orientação do pessoal (segurança do pessoal) e governança responsiva (segurança ambiental)*, enquanto a *centralidade do paciente* está relacionado à *governança responsiva (continuidade percebida), orientação de pessoal (aspectos interpessoais) e eficácia clínica (continuidade cuidado dentro da organização)*.(Veillard et al., 2005: 488).

Além disso, o PATH visa contribuir na melhoria dos sistemas de informação e qualidade dos dados e reforçará a credibilidade dos sistemas de medição de desempenho e a confiança

dos hospitais nos dados de que precisam para avaliar sua atuação e promover a *accountability* (V. Kazandjian, 2003; Roski & Gregory).

Por sua vez, Vallejo *et al.* (2006) propuseram uma aproximação dos oito conceitos fundamentais de excelência da EFQM (EFQM, 1999) para o campo de cuidados de saúde através de um novo enquadramento teórico-Conceitual do projeto PATH. A partir da proposta do projeto PATH, além da agregação de novos elementos, Vallejo *et al.* (2006) apresentaram um quadro teórico-conceitual (*framework*) do EFQM devidamente adaptado com especificidades para os cuidados de saúde, com oito dimensões de qualidade e duas dimensões metodológicas (figura 12, anexo 2).

Groene *et al.* (2008a), apresentaram um estudo comparativo através de uma revisão sistemática integrativa de estudos aplicados de dez projetos³⁹ consagrados de mensuração de performance em hospitais, de países europeus e de quatro países selecionados, a fim de comparar com o modelo PATH, em termos de dez critérios de avaliação dos instrumentos, quanto a aplicabilidade, vantagens e desvantagens na aplicação dos instrumentos do PATH. As principais conclusões foram que os sistemas de indicadores de dez diferentes países parecem buscar coberturas similares às consolidadas no projeto PATH, que é destacado como um método flexível e adaptável em contextos de diferentes países.

Em seguida, em uma investigação de campo, Groene *et al.* (2008b) analisaram um projeto piloto de aplicações do projeto PATH, concluídas em 37 hospitais de regiões de seis diferentes países (Bélgica, Canadá - Ontário, Dinamarca, França, Eslováquia e África do Sul - KwaZulu Natal), através de entrevistas semiestruturadas sobre o contexto de qualidade, o processo de implantação de indicadores e os possíveis benefícios percebidos. Concluem que os respondentes apontaram limitações, pontos fortes e fracos mas que em geral o projeto PATH

³⁹ACHS - Australian Council on Health care Standards, <http://www.achs.org.au>
 BQS - Bundesgesundheitsstelle Qualitätssicherung, <http://www.bqs-outcome.de>
 CIST - Clinical Indicators Support Team, NHS Quality improvement Scotland, <http://www.indicators.scot.nhs.uk>
 COMPAQH - Coordination pour la Mesure de la Performance et l' Amélioration de la Qualité Hospitalière, <http://ifr69.vjf.inserm.fr/compaqh>
 IQIP - International Quality Indicator Project, <http://www.internationalqip.com>
 JCAHO - Joint Commission Accreditation of Health Care Organization, <http://www.jointcommission.org>
 MARQuIS - Method of Assessing Response to Quality Improvement Strategies, <http://www.marquis.be>
 NIP - The National Indicator Project, <http://www.nip.dk>
 OAH - Ontario Hospitals Association, <http://www.oaha.com>
 OECD - Organization for Economic Co-operation and Development
 PATH - The Performance Assessment Tool for Quality Improvement in Hospitals, www.pathqualityproject.eu
 QIP - Quality Indicator Project, <http://www.qiproject.org>
 The Dutch project on reporting of performance in hospitals, <http://www.rivm.nl/>
 Verein Outcome, www.vereinoutcome.ch
 WHO - World Health Organization, <http://www.euro.who.int>

facilitou as atividades de melhoria de qualidade nos níveis regional/nacional e hospitalar, tendo sido altamente valorizado pelas partes interessadas (*stakeholders*) e demonstrado a aplicabilidade de indicadores genéricos para melhorar a qualidade da atenção em diferentes contextos da saúde (V. A. Kazandjian, 2003).

No âmbito do projeto EuroREACH, Papanicolas & Smith (2010) estudaram algumas abordagens de quadro teórico-conceitual (*framework*) de sistemas de saúde a fim de consolidar uma proposta alinhada com os objetivos do projeto. Neste propósito lembram que Hsiao & Sidat (2008) classificam os quadros conceituais (*frameworks*) de sistemas de saúde em três categorias: descritivo; analítico; e determinístico e preditivo. Assim, Papanicolas & Smith (2010) comparam dez abordagens destacadas na literatura de quadro teórico-conceitual (*framework*) de sistemas de saúde e classificando-as segundo estas três categorias, além de expor as limitações de cada um dos dez modelos estudados e citam que todos os modelos analisados têm similaridades, diferenças de escopo e de funcionalidades, mas assinalam (Papanicolas & Smith, 2010: 32) que a distinção mais óbvia entre estes modelos de quadro teórico-conceitual (*framework*) é a separação dos objetivos finais e intermediários, visto que os objetivos intermediários contribuem para a realização dos objetivos finais e, muitas vezes, fornecem informações valiosas sobre o desempenho do sistema de saúde. Por fim, Papanicolas & Smith (2010: 33-35) sintetizaram objetivos finais e intermediários (quadro 19, anexo 2) similares às dimensões e subdimensões de performance sistematizadas por outros estudos.

Em Portugal, destaca-se que foi implantada uma metodologia interessante, apesar das limitações, inclusive com diretrizes de medição de resultados especificadas em legislação de 2002, sob a coordenação da Entidade Reguladora de Saúde (ERS), órgão do Ministério de Saúde, que tem como uma de suas missões avaliar e divulgar indicadores de medição de resultados de hospitais e serviços de saúde, ligados ao Serviço Nacional de Saúde (SNS) de Portugal. Para tal, criou-se o sistema *Sinas@Hospitais*, que conta com diversos indicadores de resultados gerenciados nas categorias: Excelência Clínica (eficiência por especialidades médicas), Segurança do Doente (gestão de riscos), Adequação e Conforto das Instalações (análise estrutural), Focalização no “*Paciente*” (qualidade do tratamento do utilizador-utilizador) e Satisfação do “*Paciente*” (satisfação do utilizador-utilizador).

O *Sinas@Hospitais* classifica os hospitais avaliados (*ranking*) que é utilizado pela Administração Central do Sistema de Saúde (ACSS), outro órgão do Ministério da Saúde, para acompanhar a performance das unidades hospitalares e oferecer uma base para pactuação e contratualização de acordos de resultados para cada unidade de saúde, tanto ao nível gerencial quanto assistencial, numa ampla gama de hospitais de Portugal. O *Sinas@Hospitais* permite a

negociação de orçamentos de custeios anuais para cada unidade de saúde, conforme o volume de atendimento, parametrizado pelo risco e a complexidade dos pacientes atendidos, identificando ainda os resultados pactuados e alcançados por cada unidade de saúde. Em relação à perspectiva de avaliação são consideradas três vertentes: efetividade dos cuidados prestados, eficiência e desempenho financeiro das organizações de saúde.

Por fim, em virtude da diversidade de abordagens de quadro teórico-conceitual (*framework*) de avaliação de performance aplicados à avaliação de sistemas, políticas e instituições de saúde, ao nível mundial, concluiu-se esta análise, sem a pretensão de esgotar o assunto, mas sedimentar um possível quadro teórico-conceitual (*framework*) de avaliação de performance para basear as dimensões e indicadores de performance procedentes da pesquisa de campo.

No propósito de aproximar-se da realidade do objeto de avaliação deste estudo, caracterizam-se alguns programas e projetos de reconhecidas relevâncias na avaliação de organizações de saúde desenvolvidos no Brasil, no contexto de saúde pública e iniciativas privadas de ampla aplicação. Assim, buscou-se mapear e descrever qualquer iniciativa inovadora de monitoramento e avaliação, com similaridade de objeto e escopo desta investigação que seja considerada de relevante aplicação no contexto das organizações de saúde do Brasil, a fim de avaliar sua aplicabilidade no SUS.

Uma revisão sistemática (Machado et al., 2013) voltada para avaliação de qualidade hospitalar mostra que os primeiros estudos surgiram entre 1991 a 2003, quando foram publicados apenas nove documentos no período. Entre os anos de 2004 e 2011, surgiram 48 publicações (quadro 26, anexo 4). Esse estudo classifica as publicações como transversal, qualitativos, caso-controle dentre outros. Em outro estudo sobre a avaliação em saúde no Brasil (Fernandes et al., 2011) foram achados apenas 44 estudos. Apoiou-se nas proposições de Novaes (2000: 547-59) para classificar 23 artigos como “Investigação avaliativa”, 17 outros categorizados como “Investigação para a decisão” e ainda quatro artigos de “Avaliação para a gestão”, enquanto apenas dois artigos relatam a aplicação da avaliação na tomada de decisão.

Neste propósito, procedeu-se uma revisão sistemática ampliada nas principais bases publicações de artigos científicos do Brasil (SciELO e Lilacs e Medline, via Biblioteca Virtual em Saúde - BVS) tendo localizado 112 publicações no período de 1990 a 2018 (quadro 27, anexo 4), que foram classificados como: Análise bibliométrica (06), Análise estratégica (33), Avaliação de implementação (24), Análise de avaliabilidade (09); Avaliação de processo (07), Avaliação de efetividade (08), Avaliação de eficácia (01) Avaliação de impacto (01), Avaliação de produtividade (01), Avaliação de qualidade/percepção (02), dentre outros, sendo apenas 12

produções na linha de avaliação de performance em saúde e somente dois artigos no campo de avaliação de qualidade percebida que confirmou que o escopo desta investigação é uma agenda negligenciada (figura 1). A incipiência de publicações em avaliação e monitoramento de sistemas e serviços de saúde no Brasil realça ainda mais a importância do objeto de análise desta investigação. Assim, após caracterizar algumas experiências e quadros conceituais (*frameworks*) destacados mundialmente, buscou-se mapear e descrever qualquer iniciativa inovadora de monitoramento e avaliação, com similaridade de objeto e escopo desta investigação que seja considerada de relevante aplicação no contexto das organizações de saúde do Brasil, a fim de avaliar sua aplicabilidade no SUS.

De forma complementar, apesar dos processos incipientes de avaliação de performance e resultados na saúde do Brasil, caracterizaram-se alguns programas e projetos de reconhecidas relevâncias na avaliação de organizações de saúde desenvolvidos no Brasil, no contexto de saúde pública e iniciativas privadas de ampla aplicação. Os sistemas informatizados, indicadores e instrumentos gerenciais, no campo de saúde pública, tais como: indicadores do DATASUS, Projeto Desenvolvimento de Metodologia de Avaliação do Desempenho do Sistema de Saúde Brasileiro (PROADESS), indicadores do Índice de Desempenho do SUS (IDSUS), Indicadores e Dados Básicos da Saúde (IDB) da Rede Interagencial de Informações para a Saúde (RIPSA) e, indicadores do novo PNASS; no âmbito da Agência Nacional de Saúde Suplementar (ANS), Índices de Desempenho da Saúde Suplementar (IDSS) e indicadores QUALISS do Programa de Qualificação dos Prestadores de Serviços de Saúde Suplementar (QUALISS) – em fase de estruturação e; nas organizações privadas: indicadores PROHASA/CQH e indicadores SIPAGEH, dentre outros que surgirem na revisão da revisão integrativa da literatura (anexos 5 e 6).

Em meados de 2003, surgiram esforços de estruturação de um sistema de medição de performance em organizações e sistemas de saúde através do Projeto Desenvolvimento de Metodologia de Avaliação do Desempenho do Sistema de Saúde Brasileiro (PROADESS), sendo uma iniciativa do MS desenvolvida em parceria com a Fundação Osvaldo Cruz (Fiocruz)⁴⁰, no centro de estudos do Instituto de Comunicação e Informação Científica em Saúde (ICICT/Fiocruz)⁴¹, que se propunha a monitorar o desempenho do sistema de saúde nas seguintes dimensões: efetividade, acesso, eficiência, respeito aos direitos das pessoas, aceitabilidade, continuidade, adequação, segurança e equidade (F Viacava et al., 2004).

40 FIOCRUZ - Fundação Osvaldo Cruz

41 ICICT/Fiocruz - Instituto de Comunicação e Informação Científica em Saúde da Fiocruz

O primeiro relatório do PROADESS(FIOCRUZ, 2003) constou do estabelecimento de um quadro de referência teórico-Conceitual e sua adaptação à realidade do SUS. No entanto, do ponto de vista pragmático, a iniciativa original do PROADESS mostrou limitações quanto a seleção e validação de indicadores, ausência de abordagem sobre as fontes de dados e informações, além de não aplicação de cálculo e divulgação de resultados dos indicadores selecionados. Apenas em 2011, o PROADESS originou a matriz de indicadores IDSUS (Brasil, 2011b), com o Objetivo de avaliar o desempenho dos sistemas de saúde municipais, conforme as dimensões determinantes de saúde, condições de saúde da população, estrutura do sistema de saúde, e porte populacional, no entanto, mais uma vez, pouco se avançou na avaliação de hospitais e serviços de saúde (FIOCRUZ, 2011).

Apesar do PROADESS manter um razoável esforço de pesquisa e uma plataforma online no âmbito do ICIT/Fiocruz, ainda não obteve nenhum avanço na sua institucionalização na saúde do Brasil, exceto quanto a seleção e adequação de uma matriz de indicadores e suas fichas de indicadores do índice IDSUS (Brasil, 2011b). Neste contexto, a equipe do ICIT/Fiocruz procedeu uma revisão conceitual dos modelos originais propostos e a revisão das fichas de indicadores, tendo aplicado o novo modelo para a mensuração de performance para *benchmarking* dos sistemas municipais de saúde do Brasil (FIOCRUZ, 2012). No entanto, após uma única aplicação em 2012, o IDSUS também foi descontinuado pelo MS por dificuldade de negociações com demais gestores tripartites do SUS. Além disso, ao analisar o escopo do IDSUS percebeu-se ser mais focado em resultados sistêmicos do SUS, distante do escopo desta investigação, voltada para a avaliação de organizações e serviços de saúde no âmbito do SUS.

Os projetos PROADESS e IDSUS não se encontram institucionalizados no âmbito do SUS, mas em virtude de ser uma importante iniciativa de sistematização de indicadores e dimensões de performance de saúde brasileira resolveu-se caracterizá-los, a seguir, para analisar a possível adequação de parte de seus indicadores e dimensões para o escopo deste estudo.

O Projeto PROADESS(F Viacava et al., 2004; Francisco Viacava, Ugá, Porto, Laguardia, & Moreira, 2012) desenvolveu um estudo sobre as macrodimensões da avaliação em saúde, a partir da revisão feita sobre as metodologias de avaliação utilizadas no Canadá, Austrália, Reino Unido e as diretrizes propostas pela Organização Panamericana de Saúde (OPAS). Os autores destacam que o enfoque principal do PROADESS está nas dimensões de avaliação, sem a exigência de ter indicadores vinculados:

Os itens incluídos no PROADESS não são indicadores na sua maioria, mas dimensões a serem eventualmente estudadas através de indicadores a que deverão ser desenvolvidos levando em conta uma série de quesitos como validade interna e externa, confiabilidade, consistência, capacidade de discriminação, além da viabilidade e o custo da obtenção dos dados/informações necessários (F Viacava et al., 2004: 114).

Em busca de consolidar as macrodimensões desta matriz, os autores (F Viacava et al., 2004: 116-53) fazem uma extensa análise comparativa de cada macrodimensão proposta em importantes sistemas de indicadores mundiais (A. I. O. H. A. W.-. AIHW, 2000; CIHI, 2000, 2001a; Health, 2001; NHS, 1999; US/DHHS, 2000) com modelagens específicas de países selecionados: Austrália, Canadá, Reino Unido e EUA, respectivamente. Também consideraram os indicadores do Sispecto (Brasil, 2003b) e da Ripsa (Brasil, 2002a), enquanto comparativos para o contexto do SUS. Os autores (F Viacava et al., 2004: 154) destacam, ainda, a utilização de painéis de controle (*dashboard*) para monitorar o desempenho dos sistemas de saúde aplicados em países europeus, norte-americano, na Austrália e Nova Zelândia, a fim de sintetizar as dimensões estudadas.

Dos países que implementaram esta abordagem o Canadá e a Austrália vêm dependendo mais esforços para operacionalização desta estratégia. Particularmente com relação ao bloco de desempenho do sistema de saúde existem diferenças quanto as dimensões definidas. Diferentemente da Austrália o Canadá adotou as seguintes dimensões: aceitabilidade e competência. Por sua vez a Austrália definiu como dimensões responsividade, capacidade e sustentabilidade. As dimensões presentes em ambas experiências são: efetividade, eficiência, acesso, adequação, continuidade e segurança (F Viacava et al., 2004: 154).

A partir desta análise comparada, surgiu uma proposta para um sistema de monitoramento do sistema de saúde brasileiro com as seguintes dimensões do desempenho dos serviços de saúde: (1) efetividade; (2) acesso; (3) eficiência; (4) respeito ao direito das pessoas; (5) aceitabilidade; (6) continuidade; (7) adequação e; (8) segurança. Em seguida, os autores Viacava et al. (2004: 155-171) consolidaram propostas de quadros de indicadores para cada uma destas oito dimensões. Dessa forma, sintetizou-se o quadro teórico-conceitual (*framework*) no âmbito do projeto PROADESS - que também foi base para o projeto IDSUS-, através de uma matriz de macrodimensões de avaliação aplicáveis ao sistema de saúde brasileira (F Viacava et al., 2004: 115), conforme a figura 3, a seguir. Noutro estudo, Viacava e colaboradores (FIOCRUZ, 2012: 75) revisitaram esse quadro teórico-conceitual (*framework*) proposto anteriormente para avaliação da validade e da viabilidade dos indicadores e consolidaram uma lista de indicadores de desempenho do sistema de saúde distribuídos pelas dimensões propostas.



Figura 3 - Matriz de Dimensões da Avaliação de Desempenho do Sistema de Saúde. Fonte: <http://www.proadess.icict.fiocruz.br/index.php?pag=matp>(F Viacava et al., 2004: 115)

Por fim, os autores consolidaram uma lista de 113 indicadores através de uma plataforma online (FIOCRUZ, 2012: 78-80), listado no quadro 28 (anexo 5), que foram considerados nas fases finais de sistematização e expansão de conhecimentos dos gestores e decisores a fim de ampliar a análise, validação e consolidação de indicadores e dimensões de performance da fase empírica deste estudo.

[...] No final desse processo de discussão, produziu-se uma relação com 113 indicadores contemplando todas as dimensões da matriz conceitual, e 104 deles foram calculados. Em muitos casos, os indicadores já haviam sido formulados por outras instituições, como os indicadores básicos de saúde reunidos em Indicadores e Dados Básicos (IDB), que se encontram no Datasus, e aqueles normalmente calculados pela

Base de Dados do IPEA (Ipeadata) ou pelo Instituto Brasileiro de Geografia e Estatística (IBGE).

Paralelamente a esse processo, foi desenvolvida a nova página eletrônica do PROADESS (www.proadess.icict.fiocruz.br), constituída por três áreas principais: Matriz conceitual, Matriz de indicadores e Biblioteca. A matriz conceitual apresenta as mesmas dimensões propostas em 2003, com uma pequena alteração [...] na subdimensão condução (FIOCRUZ, 2012: 75).

Este mesmo grupo envidou esforços em outra tentativa de institucionalização de um modelo avaliativo de indicadores de desempenho de sistemas municipais de saúde do SUS, a partir de uma adequação das dimensões e indicadores consolidados no projeto PROADESS, Viacava e colaboradores (Brasil, 2012b) publicaram uma proposta estruturada, o Índice IDSUS, em parceria com o Departamento de Monitoramento e Avaliação do SUS (DEMAS) do MS:

Inspirada no PRO-ADESS, a avaliação de desempenho do IDSUS fixou como foco a avaliação do cuidado em saúde, por meio de indicadores capazes de mensurar o acesso potencial ou obtido e a efetividade das ações e serviços de saúde, regionalizados e realizados nos níveis da Rede de Atenção à Saúde, para os residentes de cada município brasileiro.

O modelo avaliativo do IDSUS conta com um bloco de caracterização, segundo a homogeneidade dos municípios brasileiros, que foram agrupados em relação às semelhanças socioeconômicas, do perfil de mortalidade infantil e da estrutura do sistema de saúde existente no território municipal.

[...] Para a avaliação de desempenho do SUS, focada no cuidado em saúde, foram selecionados 24 indicadores (14 de acesso potencial ou obtido e 10 de efetividade), distribuídos entre a atenção básica, as atenções ambulatorial e hospitalar e a urgência e emergência.

O índice IDSUS foi publicado apenas em 2012 (Brasil, 2012b), tendo sido descontinuado em virtude de muitas críticas dos agentes de decisão da saúde pública municipal e estadual, pois apesar de ter bases comparativas por municípios homogêneos e controle de vieses (*biases*) e outros métodos de validação e controle dos resultados, houve críticas severas ao IDSUS por efetivar a comparação e o ordenamento dos municípios por três índices de desempenho global.

Este grupo de 24 indicadores foram calculados e padronizados através de Análise de Componentes Principais para gerar três índices compostos sintéticos (Índice de Desenvolvimento Socioeconômico (IDSE), Índice de Condições de Saúde (ICS) e Índice de Estrutura do Sistema de Saúde do Município (IESSM) que foram aplicados de forma comparativa por municípios similares e por nível de atenção à saúde a fim de classificar os sistemas de saúde municipais do Brasil (Brasil, 2012b: 22).

Quanto aos programas oficiais de monitoramento e avaliação do sistema de saúde foram tímidas as iniciativas no Brasil. Algumas alternativas de programas de monitoramento de organizações e serviços de saúde foram estruturadas na década de 2000, como o Programa

Nacional de Avaliação dos Serviços Hospitalares (PNASH)⁴² que centrou-se apenas em uma avaliação normativa de serviços psiquiátricos e foi aplicado em três edições (2003/2004, 2007/2009, 2012/2014) e o Programa Nacional de Avaliação de Serviços de Saúde (PNASS)⁴³ – que foi redesenhado a partir do PNASH – e, teve apenas uma edição (2004/2006) (Monteiro, 2008; La Forgia e Couttolenc, 2009).

A partir da experiência do programa PNASH, no final de 2004, o MS estruturou-se o Programa Nacional de Avaliação de Serviços de Saúde (PNASS)⁴⁴ – que teve apenas uma edição (2004/2006) (Monteiro, 2008; La Forgia e Couttolenc, 2009). Esse programa PNASS original foi aplicado de forma censitária, nos hospitais gerais, especializados, obstétricos, unidades mistas, estabelecimento de saúde com serviços próprios de alta complexidade em oncologia e nefrologia (Monteiro, 2008).

O PNASS original contava com quatro componentes interconectados: um componente sobre aspectos estruturais e normas de licenciamento; dois componentes de pesquisa de satisfação dos usuários-utilizadores e trabalhadores, selecionados por processo aleatório; e por fim, uma avaliação de um conjunto de indicadores de qualidade e eficiência, a partir de dados dos sistemas do Datasus. No âmbito do PNASS original, a dimensão dos aspectos estruturais e normativos gerou roteiros de padrões de conformidades, dividido em 22 critérios e três blocos: gestão organizacional, apoio técnico e logístico, e gestão da atenção à saúde. Estabeleceu padrões sinalizadores de risco ou qualidade discriminados em três categorias: imprescindíveis (I), necessários (N) e recomendáveis (R), que por sua vez são avaliados por dois níveis: presença, representado pela resposta sim, ou ausência, sinalizado pela resposta não (Brasil, 2004; Monteiro, 2008). Além disso, nesta parte de avaliação normativa, propôs-se alguns critérios de julgamentos para os padrões. Os fatores imprescindíveis (I), são aqueles exigidos em norma e sua inexistência representam risco imediato à saúde, os necessários (N) são aqueles que estão normatizados, mas não impactam em riscos imediatos à saúde, e por fim os recomendáveis (R) que não estão descritos em normas, mas representam melhorias na qualidade dos serviços (Brasil, 2004; Monteiro, 2008). Ressalte-se que no contexto do PNASS o termo “normas” envolve as diversas legislações, manuais, padrões e recomendações técnicas.

O processo desse PNASS original funcionava por autoavaliação das unidades e por avaliação por inspetores indicadores pelo sistema de saúde local. Quase 40% das unidades não responderam o questionário e mais de 37% foram consideradas inadequadas, na parte avaliativa

42 PNASH - Programa Nacional de Avaliação de Serviços Hospitalares, no âmbito do SUS

43 PNASS - Programa Nacional de Avaliação de Serviços de Saúde, no âmbito do SUS

44 PNASS - Programa Nacional de Avaliação de Serviços de Saúde, no âmbito do SUS

da satisfação, em uma escala de *Likert* de 5 níveis, que considerava a nota ponderada mínima aceitável de 61 pontos (notas de 0 a 100) (La Forgia & Couttolenc, 2009: 302). Neste caso, a ideia era envolver os gestores locais do SUS no processo, mas o programa foi descontinuado talvez pela baixa aderência na aplicação do PNASS nos serviços de saúde brasileiros (La Forgia & Couttolenc, 2009: 348).

Recentemente, o MS resolveu reformular e institucionalizar novamente o programa PNASS, através de portaria ministerial (Brasil, 2015e), tendo como Objetivo geral avaliar a eficiência, eficácia e efetividade das estruturas, processos e resultados relacionados ao risco, acesso e satisfação dos cidadãos frente aos serviços de saúde SUS. O novo PNASS (Brasil, 2015f) permaneceu centrado em matrizes de indicadores normativos e estruturais, tendo alguns aspectos das relações de trabalho e satisfação de usuários, além de alguns indicadores de mortalidade, morbidade e assistenciais, ainda em fase de consolidação no Departamento de Regulação, Avaliação e Controle (DRAC) e no Departamento de Monitoramento do SUS (DEMAS) do MS. A normativa (Brasil, 2015e) estabeleceu que deve ser criado um sistema de dados e informações, o Sistema de Informações do PNASS (SIPNASS), capaz de permitir suporte a tomada de decisão através de indicadores multidimensionais (estrutura, processos de trabalho, resultados relacionados ao risco e a satisfação de usuários), além de indicar que a avaliação através do PNASS será aplicada periodicamente nos estabelecimentos de atenção especializada em saúde, ambulatorial e hospitalar, contemplados com recursos financeiros e incentivos do MS, estando os resultados em fase de análise e estruturação no DRAC e DEMAS no MS.

Uma evolução do processo de organização de indicadores dos sistemas públicos de saúde no Brasil, foi a iniciativa pela Organização Panamericana de Saúde (OPAS), em parceria com o Departamento de Informática do SUS (DATASUS) do MS, que estruturou o projeto Indicadores e Dados Básicos da Saúde (IDB) da Rede Interagencial de Informações para a Saúde (RIPSA) em face da necessidade de utilização de sistemas de informações em saúde para a adequada gestão da rede de saúde, além de fazer face ao hiato entre a implantação de serviços de saúde e a adequada gestão dos mesmos. Assim, foi iniciada em 1996, através da cooperação entre o Ministério da Saúde e da Representação da Opas no Brasil acordaram em cooperar no aperfeiçoamento de informações para a saúde, em alinhamento com a Iniciativa Regional de Dados Básicos em Saúde:

[...] A estratégia proposta baseou-se na criação da Rede Interagencial de Informações para a Saúde (Ripsa), concebida por grupo de trabalho ad hoc no qual estiveram representadas as principais estruturas do Ministério da Saúde, a Opas e instituições-

chave da política de informações em saúde no País (IBGE, Abrasco, Faculdade de Saúde Pública da USP e Ipea).

[...] Caberia à Ripsa promover esforços continuados para articular bases de dados e informações produzidas pelas instituições integrantes.

[...] Em especial, a Rede promoveria consenso sobre conceitos, métodos e critérios de utilização das bases de dados. As entidades seriam co-participantes de um processo de trabalho colaborativo, realizado com a mobilização de profissionais e de bases técnico-científicas das próprias instituições, segundo suas vocações.

[...] Formalizada em 1996, por Portaria Ministerial (Brasil, 1996b, 1997b, 2000b)⁴⁵ e por acordo de cooperação com a Opas, a Ripsa tem como propósito promover a disponibilidade adequada e oportuna de dados básicos, indicadores e análises sobre as condições de saúde e suas tendências, visando aperfeiçoar a capacidade de formulação, gestão e avaliação de políticas e ações públicas pertinentes (Brasil, 2002c: 13-14).

A construção da matriz de indicadores baseou-se nos critérios de: relevância para a compreensão da situação de saúde, bem como de suas causas e consequências, tendo considerados outros fatores como mensurabilidade, facilidade de acesso a dados dentre outros:

[...] Os indicadores estão agrupados em seis subconjuntos temáticos – demográficos, socioeconômicos, mortalidade, morbidade e fatores de risco, recursos e cobertura – e são apresentados, na matriz, segundo a sua denominação, conceituação, método de cálculo, categorias de análise e fontes de dados.

De maneira geral estão disponíveis, para cada indicador, dados desagregados por: unidade geográfica⁷; distribuição etária, segundo categorias apropriadas⁸; distribuição por sexo e situação do domicílio (urbana ou rural).

[...] A matriz serve de instrumento para a elaboração de Indicadores e Dados Básicos (IDB), produto finalístico da Ripsa periodicamente divulgado, que constitui plataforma para a realização de análises e informes sobre a situação de saúde e suas tendências (Brasil, 2002c: 17).

Sendo assim, três edições de matrizes de indicadores foram publicados (Brasil, 2002a: 22-45; 2008a:23-53), sendo em 2012 publicadas apenas em sítio eletrônico do DATASUS⁴⁶, em uma lógica de aprimoramento incrementais e validação participativa da rede RIPSA, que serviram de bases para a publicação de Boletins de Indicadores IDB, com atuais fichas de indicadores disponíveis do sítio eletrônico da RIPSA⁴⁷, sendo assim descrito:

O conjunto adotado na Rede contém cerca de 100 indicadores selecionados para compor os seis subconjuntos temáticos já referidos, dispostos segundo a sua denominação, conceituação, método de cálculo, categorias de análise e fontes de dados utilizadas. Foram adotados os seguintes critérios de seleção de indicadores: (i) relevância para a compreensão da situação de saúde, suas causas e consequências; (ii)

⁴⁵ BRASIL. Ministério da Saúde. Portaria nº 2.390/GM de 11 de dezembro de 1996, atualizada pelas Portarias nº 820/GM, de 25 de junho de 1997 e nº 443/GM, de 26 de abril de 2000.

Nota: Portarias (GM/MS) mais recentes: nº 2.004, de 17 de outubro de 2003, nº 495 de 10 de março de 2006. Está em tramitação uma nova portaria de reimplantação da Ripsa, planejada para 2020.

⁴⁶ Disponível em <http://tabnet.datasus.gov.br/cgi/idb2012/matriz.btm>

⁴⁷ Disponível em <http://fichas.ripsa.org.br/2012/>

validade para orientar decisões de política e apoiar o controle social; (iii) identidade com processos de trabalho na gestão do SUS; e (iv) possibilidade de construção regular a partir de bases de dados, sistemas de informação ou estudos nacionais (RIPSA, 2007: 12).

Em relatório publicado (RIPSA, 2007) pode-se conhecer o funcionamento das estruturas de governança e decisão da rede. Entretanto, desde 2012 o projeto encontra-se descontinuado, tendo sido oficialmente redirecionado para o Departamento DEMAS/MS que tem a missão de retomar a RIPSA em um curto intervalo de tempo. Por este Aspecto e em virtude do escopo da matriz de indicadores ser mais voltadas para aspectos de morbimortalidade, demografias, coberturas, recursos e determinantes de saúde não foi possível considerar muitos indicadores para o escopo deste trabalho, mais direcionados ao desempenho de organizações, unidades e serviços de saúde.

Noutra perspectiva, em busca de caracterizar outra iniciativa de avaliação de performance, no âmbito da atenção básica (AB) do SUS, Lima (2016: 61) articula que:

A partir de 2005, o Ministério da Saúde desenvolveu a proposta de uma Política Nacional de Monitoramento e Avaliação da Atenção Básica, para qualificar processos decisórios nos âmbitos da gestão e do cuidado, e com foco na integralidade e resolutividade das ações

Dessa forma, Lima (2016: 64) destaca o programa PMAQ-AB, do Departamento de Atenção Básica do Ministério da Saúde, enquanto estratégia institucionalizada de avaliação da atenção básica do SUS:

Atualmente, o PMAQ-AB destaca-se como a principal estratégia de avaliação da AB no Brasil, cujos resultados são o foco de análise desse trabalho.

[...] A partir de 2011, a gestão do Departamento de Atenção Básica/Ministério da Saúde teve como uma das prioridades o Programa Nacional para Melhoria do Acesso e Qualidade da Atenção Básica (PMAQ-AB). O programa foi criado mediante a Portaria GM/MS 1.654 de 19 de julho de 2011, como base para repasse do incentivo financeiro por desempenho denominado “Componente Qualidade” do Piso de Atenção Básica variável (Brasil, 2011e).

[...] O principal Objetivo do PMAQ-AB é induzir a ampliação do acesso e a melhoria da qualidade da Atenção Básica (Brasil, 2013b).

As fases de aplicação da avaliação externa do PMAQ-AB, dos dois primeiros ciclos, ocorreram entre o final de 2014 e o início de 2014 e avaliaram a ampla maioria das unidades básicas municipais de saúde:

No ciclo 1 foi realizada entre junho a novembro de 2012 e realizou um censo das UBS brasileiras, totalizando 38.812 UBS (J. G. Lima, 2016: 68), no ciclo 2 entre novembro de 2013 e abril de 2014, voltou-se para unidades básicas por adesão

voluntária e atingiu 30.562 UBS (J. G. Lima, 2016: 71). Houve ajustes nos componentes e variáveis, padrões de referências e indicadores de desempenho que se refletiram na mudança dos instrumentos de coleta e análise de dados (J. G. Lima, 2016: 227-244).

A terceira edição (3º. ciclo) do PMAQ-AB somente foi regularizada no final de 2018, e está em aplicação em unidades básicas de municípios de todo o Brasil, com publicação de resultados planejados para 2019.

Em uma análise global dos questionários, formulários e resultados publicados dos três ciclos do PMAQ do MS, pode-se afirmar que é uma avaliação focada em estrutura, processos e normas funcionais sendo incipiente a análise de resultados e/ou desempenho das unidades de saúde pelo que se descartou qualquer alinhamento com o objeto desta investigação, mas deve-se enfatizar que está em evolução contínua, em fase avançada de institucionalização pelo MS.

Em outro prisma, na saúde suplementar do Brasil, a implementação progressiva da gestão efetiva de resultados no âmbito da ANS, surgiu com o Programa de Qualificação das Operadoras através de Índice de Desempenho da Saúde Suplementar (IDSS) com início de implementação em 2004. O segundo ciclo, também com metodologia incipiente foi executado com dados de 2005 e 2006. Durante o terceiro ciclo, o conjunto de indicadores definidos (com melhorias incrementais entre os três ciclos) serviu de base até a competência de 2010 (Brasil, 2010b) (quadro 30, anexo 6). Atualmente, com uma metodologia muito mais estruturada, o conjunto de indicadores do programa IDSS foi atualizado através de um Rol de Indicadores com suas respectivas fichas de técnicas (Brasil, 2016b, 2016c)(quadro 31, anexo 6), sendo aguardados novos ajustes evolutivos.

Por sua vez, em uma iniciativa de *Benchmarking* hospitalar, segundo Gonçalo (2010), um grupo de 170 hospitais participou do SIPAGEH, que mensalmente monitorou 17 indicadores (quadros 32 e 33, anexo 6). Segundo Rotta (2004) o programa SIPAGEH iniciou em 1999 com 34 hospitais e 16 indicadores de diversas áreas, no entanto, foi descontinuado.

O programa coleta mensalmente 16 indicadores de 34 hospitais de vários Estados do Sul e Sudeste, sendo eles: satisfação do paciente SUS, satisfação do paciente particular, turnover, absenteísmo, acidentes de trabalho, horas de treinamento, taxa cesarianas, média de permanência, permanência Obstétrica, permanência Pediátrica, mortalidade, mortalidade obstétrica, mortalidade pediátrica, margem líquida, infecção cateter venoso central, infecção cirurgia limpa (Rotta, 2004: 32).

Por outro lado, no Brasil, do ponto de vista de tentativa estruturada de construção de um sistema amplo de indicadores de desempenho hospitalar, pode-se citar algumas iniciativas de *Benchmarking*(La Forgia & Couttolenc, 2009: 75), por exemplo, do Programa de Controle de

Qualidade (CQH) (APM/CREMESP, 1998), da Associação Paulista de Medicina, em conjunto com o Conselho Regional de Medicina de São Paulo, que contava com 120 hospitais públicos e privados (Malik & Schiesari, 1998), a fim de monitorar indicadores de produtividade, financeiros e de recursos humanos (Rosalem, 2013). Segundo Rotta (2004) o programa CQH iniciou em 1991 com 100 hospitais e 17 indicadores:

Uma avaliação interinstitucional agregando aproximadamente, 100 hospitais do Estado de São Paulo, reunindo 17 indicadores, a saber: mortalidade institucional, mortalidade operatória, taxa de ocupação hospitalar, tempo médio de permanência, cirurgias suspensas, taxa de cesarianas, apgar em recém-nascidos, infecção hospitalar geral e por clínica, percentual de médicos especialistas, relação enfermeiro/leito, relação funcionários enfermagem/leito, índice de rotatividade, taxa absenteísmo, acidentes de trabalho e horas de treinamento. Esses indicadores são enviados mensalmente por essas instituições. A partir daí são elaborados relatórios trimestrais que comparam os indicadores das instituições.

[...] O programa também possui instrumento (manual) que possibilita avaliar instituições hospitalares dentro de padrões pré-estabelecidos, visando a obtenção de uma certificação de qualidade (Associação & Medicina, 1998)(Rotta, 2004: 31-32).

Desde 1992, foi implantado outra iniciativa de *Benchmarking*, no âmbito do Programa de Estudos Avançados em Administração Hospitalar (PROHASA) da Fundação Getúlio Vargas (FGV)⁴⁸, em parceria com o Hospital das Clínicas de São Paulo (La Forgia & Couttolenc, 2009), que avalia, monitora indicadores de custos e desempenho de hospitais voluntários. Esses indicadores tem sido publicados no Boletim Indicadores PROAHSA bimensal, que também divulga dados dos hospitais do CQH e de outras fontes (Rosalem, 2013), mas com resultados limitados pois tinham apenas cerca de 30 hospitais participantes (La Forgia & Couttolenc, 2009).

Em outro estudo (Rosalem, 2013) tem-se um quadro integrativo (do rol de indicadores das metodologias do CQH, PROHASA, SIPAGEH e de três outras listas de indicadores de alguns órgãos e agencias internacionais (*ACHS - Australian Council on Healthcare Standards, CND - Canadá Health Indicators – 2003, HP - Health People 2010 – USA, NHS - National Health Security, RALC - Grupo para Reforma del Sector Salud en América Latina y Caribe – OPAS*) listados no quadro 34 (anexo 6).

Quanto ao campo de indicadores de *Benchmarking* entre organizações do sistema público de saúde brasileiro, as únicas iniciativas estruturadas e publicadas de aplicação subnacional estão em desenvolvimento na Secretaria de Saúde de Minas Gerais (SES/MG)⁴⁹,

48 FGV - Fundação Getúlio Vargas, importante Universidade Privada com sedes no Sudeste do Brasil

49 SES/MG - Secretaria de Saúde de Minas Gerais

no bojo do programa do governo do estado (província) de Minas Gerais denominado de “Choque de Gestão” em sua diretriz “Acordos por Resultados”. Neste programa tem sido desenvolvido contratualização por resultados e sistemas de acompanhamento de indicadores de performance em unidades de saúde e hospitais públicos e filantrópicos, chamado de Programa de Fortalecimento e Qualificação da Rede Hospitalar do SUS (Pro-Hosp)⁵⁰, além de outra iniciativa similar na Fundação Hospitalar do Estado de Minas Gerais (FHEMIG)⁵¹, que controla 20 unidades assistências de propriedade do governo estadual, sendo os resultados controlados pela SES/MG (FHEMIG, 2009; Ramos & De Oliveira, 2010), podendo serem consideradas duas experiências exitosas.

Com o avanço do programa, em sua 2ª. geração, visando a consolidação dos Acordos de Resultados na SES/MG, estabeleceu-se um arcabouço legal subnacional através do decreto nº 45.468/2010 (Minas, 2010a), inclusive para o Programa Pro-Hosp que resultou em um *proxy* de painel de indicadores de governança de organizações e serviços de saúde (Marques, Mendes, & Lima, 2010), que monitora 130 hospitais públicos e filantrópicos através de contratualização por resultados (H. d. O. Lima, Lima, Campos, & Oliveira, 2010), em três macrodimensões (acesso, resolubilidade, e regionalização da assistência) com um grupo de cinco (5) indicadores, desde 2012 (Minas, 2012): Taxa de mortalidade institucional, taxa de referências (pacientes encaminhados entre municípios), taxa de cumprimentos de pactos regionais do SUS, taxa de ocupação hospitalar, e taxa de mortalidade infantil.

Esta iniciativa foi operacionalizada através do programa Pro-Hosp, inclusive com indicadores e critérios de avaliação e monitoramento de resultados, e coube aos municípios – induzidos pelo Estado – estabelecerem contratos de gestão (com metas e resultados) com as organizações (hospitais, unidades de atenção primária, centros de especialidades) que terão repercussão em remuneração variável dos profissionais.

Nessa nova etapa, para suporte de governança do programa foi concebida a elaboração de sistema de indicadores para controle de resultados governamentais, o Gerenciador de Informações, Compromissos e Metas (GEICOM), definido em legislação subnacional (Minas, 2010a). O sistema GEICOM, contribuiu com diversas vantagens e benefícios para a governança do sistema de saúde, na percepção dos gestores (Junior, 2013):

Os principais benefícios percebidos passam por: Agilidade, facilidade e desburocratização no processo de contratualização, principalmente no que tange à

50 Pro-Hosp - Programa de Fortalecimento e Qualificação da Rede Hospitalar do SUS

51 FHEMIG - Fundação Hospitalar do Estado de Minas Gerais, que gerencia cerca de 50 unidades hospitalares e serviços públicos de saúde

assinatura e documentação necessária; rapidez no fluxo de ações e informação entre as esferas do governo; possibilidade de armazenar, acessar e recuperar informações sobre o processo de contratualização (contratos, monitoramento, avaliação, desempenhos e transferências realizadas); acessibilidade do controle interno/externo, garantindo a transparência e accountability; e segurança em todo o processo, através da certificação digital que garante o acesso restrito às ações e informações do sistema.

Pela relevância da experiência da FHEMIG, que apesar do resumido quadro de indicadores padronizados mostrou-se como um processo estruturado de controle de resultados de unidades de saúde, relatou-se alguns tópicos principais no anexo 7.

Por fim, a partir da literatura relevante, sintetizou-se os quadros teóricos-conceituais (*frameworks*) sobre os sistemas de indicadores de performance de destaque mundiais (quadros 17 a 19, anexo 2), também no campo de saúde pública do Brasil (quadros 26 e 27, anexo 4) além dos mais relevantes da saúde privada brasileira (quadros 30 a 34, anexo 6), sistematizados quanto as suas possíveis potencialidades, virtudes, limitações e/ou desafios de implementação, para compor uma matriz de dimensões e de indicadores para avaliação e validação pelos agentes de decisão da amostra desta investigação.

2.4.4.O enquadramento teórico-conceitual (*framework*) da avaliação e medição da qualidade percebida em saúde e a escala SERVQUAL

Nesta etapa explorou-se as bases de publicações sobre instrumentos e métodos de avaliação da satisfação em organizações e serviços de saúde, através de termos tais como: *Perceived Quality in Hospital*, *Perceived Satisfaction in Hospital*, *SERVQUAL Scale*, *ServPerf Scale*, dentre outros similares. Encontrou-se outro hiato de publicações, mas apesar de não funcionar como palavra chave, a metodologia da escala SERVQUAL surge como uma das mais usadas no contexto de saúde mundial.

Na realidade do Brasil, identificou-se duas pesquisas de opinião de base nacional do Conselho Federal de Medicina (CFM) e da Confederação Nacional da Indústria (CNI), com os mais renomeados institutos brasileiros de pesquisa, Datafolha e Ibope, respectivamente, com dados de 2011 a 2018, que revelam a importância da avaliação de qualidade e satisfação. No entanto, apontam insatisfações da população que variam de 54 a 93% (CFM, 2014, 2015, 2018). E, ainda, que elevados percentuais de brasileiros avaliam a saúde pública como ruim ou péssima, que aumentou de 61%, em 2011, para 75% em 2018 (CNI, 2012, 2018), que variam conforme a perspectiva de análise e abordagem dessas pesquisas, em especial quanto ao uso recente dos serviços, seja na saúde pública ou suplementar.

As pesquisas do CFM (2014, 2015, 2018) indicam que ao abordar os indivíduos que usaram os serviços nas unidades do SUS, nos últimos 24 meses, a insatisfação máxima com a qualidade dos serviços reduziu-se para 66 a 69%. Além disso, os fatores mais críticos foram tempo de espera para agendamento de atendimentos e para a realização de procedimentos especializados e cirurgias, além da dificuldade de acesso a leitos hospitalares e de UTI e aos atendimentos médicos (64 a 89%). Em seguida, surgiram os problemas nas instalações estruturais, de quantidade dos demais profissionais de saúde e de má administração das unidades. Apenas de 13 a 16% dos entrevistados consideram não haver dificuldades na gestão de recursos do SUS. Mapearam-se variações dos resultados obtidos por diversos fatores, por exemplo, uso recente dos serviços de saúde, regiões geográficas do Brasil, cidade de região metropolitana ou interior, faixa etária, gênero, escolaridade, nível de renda e posse de planos de saúde pelos entrevistados.

Por sua vez, as pesquisas da CNI (2012, 2018) indicam que ao abordar os indivíduos que usaram os serviços nas unidades do SUS, nos últimos 12 meses, a insatisfação máxima com a qualidade dos serviços reduz-se para 54 a 56%, quando analisam o sistema de saúde da cidade que habitam, ou seja, os usuários estão mais satisfeitos com a saúde de sua cidade que a do País. Destaca-se também enquanto fator relevante o porte da cidade do sistema de saúde: municípios com até 50 mil habitantes (43% de insatisfação) e com mais de 500 mil habitantes (69% de insatisfação). Além disso, os fatores mais críticos foram tempo de espera para agendamento de atendimentos e para a realização de procedimentos especializados e cirurgias (37 a 57%), além da falta de equipamentos e investimentos nas unidades (10 a 15%). Em seguida, surgem os problemas falta de médicos e de má administração das unidades. Apenas de 13 a 16% dos entrevistados consideram não haver dificuldades na gestão de recursos do SUS. Identificaram variações dos resultados por diversos fatores: uso recente dos serviços de saúde, regiões geográficas do Brasil, cidade de região metropolitana ou interior, faixa etária, gênero, escolaridade, nível de renda e posse de planos de saúde pelos entrevistados. De forma objetiva, avaliou a satisfação no último atendimento hospitalar em uma escala de 0 a 10, tendo os seguintes resultados:

A nota média dada pela população acerca da estrutura dos hospitais públicos cai em comparação com 2011 em todos os pontos avaliados, fazendo com que a média geral da avaliação recue de 5,7, em 2011, para 4,6, em 2018.

Condições de limpeza/manutenção é o fator com a maior nota média (5,6), seguido por facilidade de acesso ao hospital (5,5), por roupas, toalhas e roupas de cama (5,1) e por alimentação oferecida (5,1).

As piores notas são conferidas ao tempo de espera para atendimento/internação (3,2), tempo médio para realização de exames (3,5) e rapidez no atendimento na recepção

dos hospitais (3,9). Os três fatores citados obtiveram a pior avaliação entre os aspectos e estão diretamente ligados à dificuldade de se obter atendimento, eleita pelos brasileiros como o principal problema do sistema público de saúde de sua cidade.

[...] Entre os itens melhor avaliados, encontram-se competência e conhecimento dos médicos (6,3), frequência e presença dos enfermeiros no hospital (5,9) e competência e conhecimento dos enfermeiros (5,9).

Já entre os itens pior avaliados estão a quantidade de médicos (4,3) e a frequência e presença dos médicos no hospital (4,9), coincidindo então com o terceiro item mais citado como principal problema da saúde pública da cidade dos entrevistados, a falta de médicos.

Diante de tais resultados, sintetizou-se que quanto menor o tempo decorrido do último atendimento, quanto mais próximo for a unidade de atendimento e quanto menor o porte da cidade menores foram os índices de insatisfação registrados. Revelou-se também que cidades das regiões metropolitanas têm maiores insatisfações, achados alinhados com as insatisfações das grandes cidades, talvez pela sobrecarga de transferências de atendimentos oriundas dos menores municípios. Destacaram-se importantes fatores mediadores da satisfação nessas pesquisas de opinião sobre a saúde, tais como: faixa etária, gênero, escolaridade e nível de renda dentre outros, que sugerem possíveis variáveis mediadoras para estudos da qualidade e satisfação em saúde.

Por fim, vale ressaltar os fatores mais relevantes quanto as insatisfações dos usuários, tendo destaque o tempo de espera para agendamento de atendimentos e para a realização de procedimentos especializados e cirurgias, em ambas as pesquisas. Seguiram-se outros fatores: dificuldade de acesso a leitos hospitalares e de UTI e aos atendimentos médicos, falta de equipamentos e de médicos além dos investimentos nas unidades. Ambos os inquiridos destacam a insatisfação com aspectos gerenciais e de gestão de recursos na saúde do Brasil.

Nesse contexto, desde 2002, o MS tem operacionalizado duas estratégias de ouvidoria da opinião dos pacientes do SUS: a Carta SUS, um instrumento que consolida as informações dos atendimentos nas unidades hospitalares vinculadas ao SUS, que envia uma carta à residência do usuário para fornecer estas informações e solicitar retorno sobre sua percepção e satisfação quanto ao atendimento recebido; e o Disque Saúde - 196, que é uma linha telefônica gratuita pela qual qualquer cidadão pode encaminhar reclamações e sugestões sobre o sistema de saúde, inclusive quanto sua satisfação de atendimento no SUS. Porém, nestas duas iniciativas, pouco se avançou em modelos de avaliação de performance das unidades e/ou serviços do SUS.

Por sua vez, duas tentativas incipientes foram implantadas em 1998 e 2004, por uma parceria entre o MS e o Conselho Nacional de Secretários de Saúde (CONASS)⁵² não tendo se transformado em uma ferramenta de governança de organizações de saúde (Brasil, 2003a). Outras duas iniciativas foram a Pesquisa Mundial de Saúde (PMS) que foi aplicada em unidades de saúde de todo o Brasil, em 2003, além da PMS com foco específico na Atenção Básica (PMS-AB) aplicada somente nos estados (províncias) do Rio de Janeiro e Pernambuco, em 2005. Ambas não tiveram continuidade e nem resultados sistematizados no SUS (Gouveia et al., 2009).

No contexto do Programa Nacional da Gestão Pública e Desburocratização (Gespública), instituído pelo Decreto nº 5.378/2005 (Brasil, 2005) criou-se a Carta de Serviço ao Cidadão além da proposição de um método de mensuração da satisfação de usuários de serviços públicos (Brasil, 2010c), no âmbito do Ministério do Planejamento do Brasil, no qual se propôs a criação do Instrumento Padrão de Pesquisa e Satisfação (IPPS) tendo analisado três metodologias internacionais (*American Consumer Satisfaction Index* da Universidade de Michigan; o SERVQUAL - dos especialistas Zeithaml, Parasuraman e Berry- e também do *Common Measurement Tool* do Centro Canadense de Gestão). No IPPS, destacam-se as cinco dimensões proposta na escala SERVQUAL (confiabilidade, capacidade de resposta, segurança, empatia e tangibilidade) e as 22 questões originais do método que parecem ter sido objeto de traduções simplificadas e incorporadas diretamente na proposta de questionário vinculado ao IPPS, que segundo Batista e Paiva (2014) foi base para um software que combina elementos das principais metodologias internacionais de medição de satisfação dos usuários, na forma de um questionário de pesquisa de opinião padronizado.

Não há registros de continuidade do IPPS, mas, o governo brasileiro publicou outra normativa (Brasil, 2017c) que reforça a importância da participação no acompanhamento da prestação e na avaliação dos serviços pelos usuários que demandou ajustes quanto a proatividade das ouvidorias dos órgãos públicos, desenvolvimento de mecanismos e procedimentos de respostas às manifestações dos usuários, inclusive exigências de destaques deste tema no relatório anual de gestão do órgão público, sob análise dos órgãos externos de controle público.

Registre-se que, em 2013, o IBGE aplicou em uma extensa amostra de domicílios de todo o Brasil, a Pesquisa Nacional de Saúde (PNS)⁵³ que talvez possa subsidiar estudos de

52 CONASS - Conselho Nacional de Secretários de Saúde do Brasil

53 PNS - Pesquisa Nacional de Saúde, no Brasil

performance, qualidade e satisfação quanto ao sistema de saúde, com incipientes resultados publicados, como por exemplo no artigo de Boccolini et. al. (2016).

Na pesquisa de Gastal *et al.*(2005 apud La Forgia e Couttolenc, 2009: 339), investigou-se junto aos gestores de organizações com certificados de acreditação quais eram os métodos e ferramentas mais relevantes na dimensão de qualidade. Na percepção de 37 gestores de topo de hospitais certificados em qualidade, a relevância atribuída à medição de satisfação do usuário de saúde alcançou quase o nível máximo (4,7), de uma escala até 5,0.

Noutro contexto mais recente, o Decreto Federal(Brasil, 2017b) pautado na Lei 13.460(Brasil, 2017c), determinou a obrigatoriedade de avaliação periódica dos serviços pelas instituições públicas e estabeleceu que os órgãos e entidades federais utilizem ferramenta de pesquisa de satisfação dos usuários e usem as informações coletadas para a melhoria dos serviços. Além disso, nesse escopo, um inquérito recente da Escola de Administração Pública do Ministério do Planejamento do Brasil revelou que mais de 75% dos serviços de órgãos e instituições federais não possuem avaliação regular de qualidade e/ou satisfação dos usuários, que destaca a importância da metodologia validada para a medição da qualidade e satisfação percebida em saúde proposta nesta investigação a qual, futuramente, pode ser adequada e validada para outros serviços públicos do Brasil.

Esses contextos explanados, podem sugerir a relevância atribuída à avaliação da qualidade e satisfação pelos usuários, em especial, da saúde brasileira. Assim, resolveu-se investigar um método de mensuração da qualidade e satisfação percebida, através da participação de profissionais e especialistas da área, além da promoção do empoderamento dos usuários, inclusive na validação de constructos, questionários e na aferição efetiva da qualidade percebida, em uma amostra selecionada da rede de unidades do SUS.

Ressalte-se que na pesquisa integrativa da literatura ampliada (tabela 110, anexo 1) e na pesquisa sistemática nas principais bases de artigos científicos do Brasil (quadro 27, anexo 4), não se identificou nenhuma metodologia robusta de avaliação de qualidade e satisfação através de instrumentos padronizados e validados e/ou de larga utilização no contexto de saúde do Brasil. Esses achados estão alinhados com os resultados de uma revisão sistemática extensiva, entre 1990 e 2013, nas principais bases bibliográficas de publicação brasileiras em saúde pública (Medline, Lilacs e Scielo), na qual Volpato (2014) identificou cerca de 195 trabalhos publicados, mas apenas 30 deles estavam voltados à percepção dos usuários e/ou dos profissionais referente a qualidade dos serviços públicos de saúde, porém não se identificou nenhum instrumento ou escala validada e consolidada de mensuração de satisfação e/ou qualidade em saúde no Brasil.

Concluiu-se que, no Brasil, há uma histórica escassez de instrumentos adequados para a medição de satisfação dos usuários de saúde. Neste cenário, surge a importância de se elaborar uma metodologia cientificamente validada para a mensuração da satisfação e da qualidade percebida por usuários no contexto do SUS, sendo um dos objetivos desta investigação.

Por outro lado, no aprofundamento da revisão integrativa da literatura sobre aferição de satisfação em saúde, sintetizada no quadro 21 (anexo 2), no sistema de saúde brasileiro, surgiram três trabalhos de pequeno porte de aplicação da SERVQUAL em contextos específicos (Hercos & Berezovsky, 2004; Borges, et al., 2006; Cruz & Melleiro, 2007). Também foram identificados outro três trabalhos com sistematização da escala SERVQUAL para aplicação em ambientes hospitalares, mas com traduções e adaptações simplificadas e aplicação em amostra resumida em hospitais do Brasil (Albuquerque, 2012; Da Silva, 2014; Rosalem, 2013).

No estudo de Albuquerque (2012) a escala SERVQUAL foi aplicada no ambiente de seis hospitais gerais, vinculados ao SUS. Os resultados sugerem que a utilização da avaliação dos serviços pode apoiar aos gestores e profissionais a mensurar a qualidade de serviço através da identificação e análise de suas dimensões. Por sua vez, Rosalem (2013) aplicou a SERVQUAL para mensurar a percepção no ambiente de hospitais gerais, vinculados ao SUS, sendo 10 públicos e 10 privados. Os resultados apontaram a adequação da aplicação da escala SERVQUAL e forte diferenciação da percepção de qualidade entre hospitais públicos e privados e também entre os três grupos de entrevistados: gestores, profissionais e pacientes. Por fim, os estudos de Da Silva (2014) avaliou a percepção de qualidade e satisfação com pacientes de unidades clínicas e cirúrgicas do Hospital da Universidade de Sergipe, sendo que as cinco dimensões e os 22 atributos das questões da Escala SERVQUAL aplicadas apresentaram as diferenças entre percepção e expectativa negativas e próximas de zero, ou seja, não se verificou a associação negativa citada pela literatura de aplicação da escala SERVQUAL (E. Babakus & W. G. Mangold, 1992; Parasuraman et al., 1988; Parasuraman, Berry, & Zeithaml, 1991a; Parasuraman, Zeithaml, & Berry, 1994). Destacou ainda que a Empatia foi considerada a dimensão de maior importância, seguida da dimensão Confiabilidade, enquanto a Tangibilidade obteve a menor importância.

No escopo delineado por esta investigação, a seguir, destacaram-se algumas propostas e experiências relevantes a fim de construir um enquadramento teórico conceitual (*framework*) para construir e validar os constructos (ideias-força) e questões de avaliação de qualidade percebida e satisfação dos usuários, com distribuição por subdimensões e macrodimensões de qualidade, que podem funcionar como itens de intervenção para a gestão de organizações, unidades e serviços de saúde pública do Brasil.

No propósito de elaborar uma metodologia cientificamente validada para a mensuração da qualidade percebida e satisfação de usuários no âmbito do SUS, apesar de possíveis limitações relatadas na literatura (Chahal & Kumari, 2011, 2012; Cronin & Taylor, 1992; J. J. Ferreira, Raposo, Rodrigues, Dinis, & Paço, 2012; P. L Ferreira, 1993; Pedro, Ferreira, & Mendes, 2010), justifica-se a proposta de adaptação transcultural, com rigor metodológico, da escala *Service Quality* – SERVQUAL, elaborada por Parasuraman *et al.*(1985a), que é pautada no modelo “*gap model*” de satisfação de Oliver (1980), que teoriza ser a satisfação do cliente uma função da diferença entre a expectativa e o desempenho.

Dessa forma, em virtude da larga aplicação da escala *Service Quality* – SERVQUAL em sistemas de saúde de diversos países (quadro 20, anexo 2) e das primeiras tentativas de sistematização no Brasil (quadro 21, anexo 2), acima destacadas, resolveu-se selecionar este método como ponto de partida através de sua tradução, adaptação e validação pela colaboração de atores estratégicos (*key stakeholders*): especialistas, profissionais e usuários, com rigor metodológico e analítico, devidamente autorizado pelos autores (apêndice 3), a fim de propor uma escala inovadora de qualidade percebida validada no âmbito desta investigação.

Assim, a proposta de tradução e adaptação transcultural da escala SERVQUAL em um novo instrumento, com itens de mensuração de qualidade devidamente analisados em sua validade e confiabilidade, fundamenta-se no campo da translação de conhecimentos no qual as teorias consolidadas são alvos de transliteração à luz de experiências dos atores sociais envolvidos (Barbosa & Neto, 2017; Clavier *et al.*, 2011; Hartz & John, 2008). No presente estudo, pautou-se em um processo colaborativo a fim de estabelecer os constructos (ideais-forças) de qualidade para formação de questões de avaliação de qualidade e satisfação de usuários, que sintetizaram e aplicam conhecimentos em prol de preenchimentos de lacunas de saberes e instrumentos em busca de transformar políticas e práticas (Barbosa & Neto, 2017; CIHR, 2004; WHO, 2006) para a melhoria de serviços e da gestão do SUS. O processo sistematizado pode ser replicado para outras tipologias de serviços públicos a fim de estruturar e validar escalas de qualidade específicas.

Nesse propósito, ao estudar o campo de organizações de serviços, pode-se articular que as raízes da Conceitualização dos serviços teve início nos estudos da escola nórdica, a partir dos meados dos anos 70, por exemplo, nas pesquisas de Gummesson (1977) e de Grönroos (1978, 1982, 1988).

Quanto às organizações de saúde, estas podem ser englobadas no campo de serviços e, há relativo consenso que o sistema de saúde pública brasileira enfrenta elevadas insatisfações

pelos usuários que podem variar de 54 a 93% nas pesquisas do CFM (2014, 2015, 2018) ou de 61 a 75% nas pesquisas da CNI (2012, 2018).

No escopo da saúde, uma ampla revisão da literatura mundial aponta que a satisfação do usuário é multifacetada e de difícil mensuração (M. Esperidião & Trad, 2005; M. A. Esperidião, 2009) e, assim, em muitas pesquisas surgem resultados desconectados da realidade, ou seja, existem insatisfações dos pacientes com os serviços recebidos, porém pouco explicadas nos resultados das pesquisas, possivelmente, por falhas nas metodologias de mensuração, sendo um campo de interesse para os decisores de saúde (Aharony & Strasser, 1993 apud Esperidião e Trad, 2005: 304).

Diante dessa aparente inconsistência e da extrema necessidade de avaliar a satisfação dos usuários dos serviços de saúde, diversas metodologias têm sido desenvolvidas para estruturar uma escala de mensuração confiável de satisfação de usuários-pacientes, pelo que se caracterizam os aspectos positivos e as limitações das principais propostas que surgiram na pesquisa bibliográfica integrativa (quadro 20, anexo 2), visto que, no Brasil, ainda não se registra nenhuma metodologia consolidada e/ou usada em larga escala dentro do escopo de satisfação desta investigação (quadro 21, anexo 2).

Nesse contexto, Vinagre (2008) enfatiza a complexidade de aferição dos serviços pelos aspectos da *intangibilidade* (ausência de atributos físicos) dos serviços (Parasuraman et al., 1991b) e o fato dos serviços constituírem-se essencialmente enquanto processos, sendo afetados por interações e atividades (Grönroos, 1990), ou seja, dependente de aspectos relacionais entre usuários e prestadores de serviços. Diante disso, Grönroos (2003) sugere mensurar a qualidade de serviços sob a ótica de seis critérios de qualidade percebida.

Por sua vez, Surprenant e Solomon (1987) afirmam que os serviços são concretizados pela experiência, e salientam a sua natureza interpessoal e relacional com enfoque no que alguns autores designam por *encontro* ou *experiência de serviços*, pressupondo que a produção e o consumo implicam a interação entre o utilizador, o prestador e a organização. Noutra visão, o modelo proposto por Bitner (1992 apud Vinagre 2008), caracteriza-se o *encontro de serviços* como a relação interpessoal e/ou inter-relacional entre usuário e prestador, a organização e o ambiente envolvente, enquanto a *experiência de serviços* como o resultado de um encontro ou de uma sequência destes que já ocorreram e nos quais se produziram múltiplas interações, que inclui as relações relativas ao ambiente, regras, normas e serviços fornecidos.

A partir de uma análise integrativa realizada por Giese e Cote (2000 apud Vinagre, 2008) podem-se sintetizar os achados de alguns estudos destacados sobre as bases da avaliação

da satisfação, além de comparar as definições conceituais, as respostas, o enfoque e o momento da análise da satisfação (quadro 20, anexo 2).

Diante dos diferentes escopos possíveis para a análise da satisfação e da existência de diversas linhas diferentes de investigação da satisfação dos usuários no contexto de serviços, não sendo escopo desta investigação aprofundar suas bases teóricas-conceituais, neste estudo elegeu-se o escopo de *experiência de serviços*, visto que dadas as características do atendimento em saúde, a *experiência na saúde* pode ser sensível aos constructos *expectativas*, *emoções* e *experiências*, que são constructos que estão na base do enquadramento teórico da linha *experiência de serviços* para mensuração da satisfação.

Vinagre (2008) ressalta que no ambiente de *transações* ou *encontros de serviços* específicos, muitos dos autores (p.ex. Churchill & Surprenant, 1982; Oliver, 1980; Oliver & Desarbo, 1988) avaliam a satisfação a partir de seus preditores, e provam em seus estudos empíricos que estas variáveis preditoras influenciam, em menor ou maior grau, na *satisfação* e na *qualidade percebida* pelo usuário durante uma *experiência de consumo do serviço*.

Em uma meta-análise sobre o assunto Szymanski e Henard (2001) salientam que as investigações sobre *satisfação* de consumidores focam predominantemente os efeitos sobre a *satisfação* através das variáveis preditoras, quais sejam as *expectativas*, a *desconfirmação* (como preditor com efeito dominante) e, citam outros preditores como a *equidade* (quando analisada conjuntamente com as *expectativas* e as *emoções*), além da *percepção de desempenho*, tendo efeito mais fraco e, às vezes, até sem significância estatística.

Dessa forma, diante das muitas definições do constructo *satisfação* e dos inúmeros quadros conceituais (*framework*) propostos na literatura para a avaliação da *satisfação*, neste trabalho, resolveu-se abordar a satisfação enquanto resposta a um processo de avaliação da experiência de consumo, e considerar tanto o resultado quanto o processo como elementos importantes para a *satisfação* dos consumidores com enfoque nos *antecedentes/preditores da satisfação*, ou seja, amparado pela *teoria da atribuição causal* (Bitner et al., 1994).

Assim, a partir de um modelo geral (figura 4) os constructos *expectativas*, *desconfirmação*, *desempenho* e *qualidade de serviços* quase sempre constam dos modelos de mensuração da *satisfação* (por simplicidade metodológica utiliza-se esta linha teórica neste estudo), enquanto os constructos *equidade*, *atribuição* e *emoções* geralmente são abordados apenas em estudos mais elaborados sobre a satisfação.

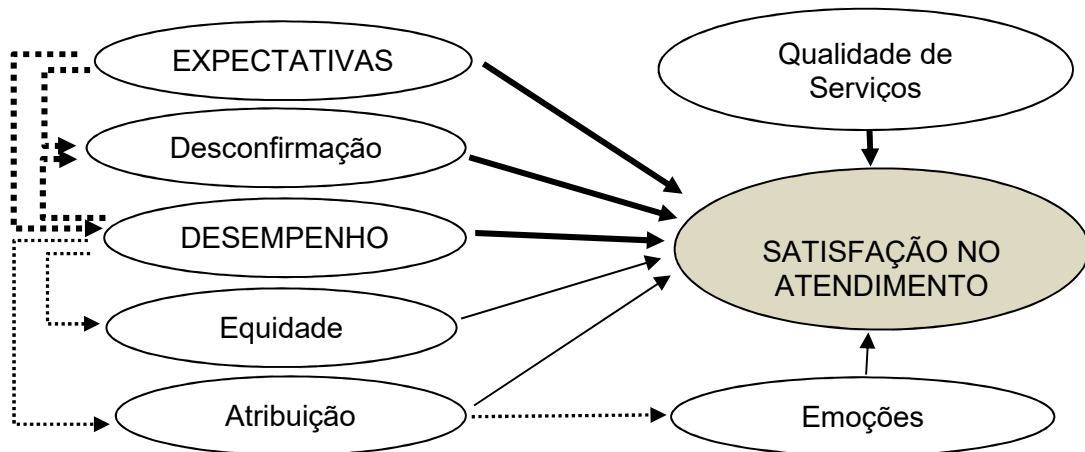


Figura 4- Antecedentes da satisfação mais frequentemente utilizados. Fonte: adaptado de Szymanski e Henard (2001 apud Vinagre, 2008: 26).

Na verdade, tem-se um enorme desafio empírico no estudo da *satisfação*, em busca de identificar os preditores que são relevantes para compor a satisfação nos *encontros de serviços*. Ao compreender a *expectativa* como um dos mais importantes preditores da *satisfação* (Szymanski & Henard, 2001) afirmam ainda que o constructo expectativa é pouco consensual, além de existirem múltiplas classes de expectativas (p.ex. Tse & Wilton, 1988; Zeithaml, Berry, & Parasuraman, 1993).

Um estudo muito relevante sobre *preditores das expectativas* de consumo foi apresentado por Oliver (1997) que desenhou um modelo (figura 5) que assume que as *necessidades* e *valores* são *preditores dos desejos* e que estes em conjunto (*necessidades, valores e desejos*) influenciam as *expectativas dos consumidores*.

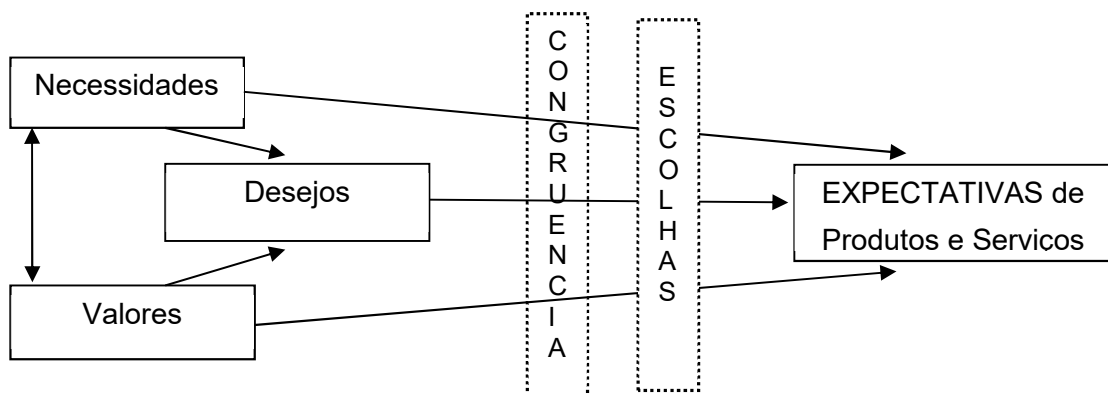


Figura 5 - Inter-relações entre necessidades, valores e desejos versus expectativas. Fonte: Adaptado de (Oliver, 1997 apud Vinagre, 2008: 29).

Assim, Oliver (1997) assume que as *necessidades* e os *valores* do consumidores são preditores diretos dos *desejos* e da *expectativa*, e a conjugação destes (*necessidades, valores e*

desejos) influenciam as *expectativas* dos consumidores. Teoriza ainda que estas inter-relações devem ser analisadas como sendo moderadas pelas congruências entre *necessidades*, *valores* e *desejos*, sempre buscada pelo consumidor, tendo ainda as escolhas de decisões de consumo como variável moderadora da formação da expectativa do consumidor.

Estes aprofundamentos de preditores são importantes, porém, para o escopo deste trabalho, resolveu-se partir da mensuração da expectativa e da percepção de desempenho e a desconfirmação objetiva (diferença algébrica) enquanto preditores da satisfação/insatisfação, baseados no figura 6, proposto por Oliver (1997).

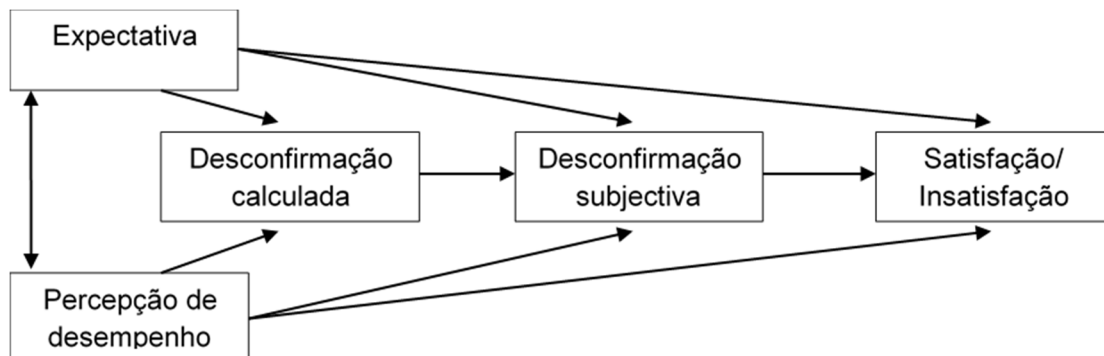


Figura 6 - Modelo completo de desconfirmação das expectativas com o desempenho percebido. Fonte: Adaptado de (Oliver, 1997 apud Vinagre, 2008: 33).

Segundo Vinagre (2008), as *expectativas* têm sido consideradas como antecipação ou como referente de comparação no contexto dos efeitos de *desconfirmação*. Enquanto antecipação da *satisfação* elas são tidas como influentes diretos da *satisfação* sem recurso comparativo com os níveis de *percepção de desempenho*(Oliver & Desarbo, 1988).

Surgem também alguns estudos empíricos (p.ex. Churchill & Surprenant, 1982; Oliver & Desarbo, 1988; Yi, 1990) que comprovam que a *desconfirmação* é um importante determinante da *satisfação*, que sugerem que a *desconfirmação* terá de ser medida de forma independente enquanto determinante específico e separado das *expectativas*, pois teorizam que as *expectativas* se alteram no pós-consumo, fase em que a *desconfirmação* é medida efetivamente.

Noutros diferentes prismas há estudos empíricos que refutam a importância da *desconfirmação* e preferem a mensuração das *expectativas* como componente integrado da *desconfirmação*(Kopalle & Lehmann, 2001; Mittal, Ross, & Baldasare, 1998). Há também estudos empíricos que suportam relação positiva entre *expectativas* e *satisfação*(Bearden & Teel, 1983; Oliver & Linda, 1981; Swan & Trawick, 1981). Por sua vez, Cronin e Taylor (1992) afirmam que as *expectativas* têm fraca saliência comparativamente com o *desempenho*

percebido. Há, ainda, estudos que utilizam a *satisfação* na condição de referente de comparação.

No contexto da *satisfação* como referente de comparação, surge o *paradigma da desconfirmação, modelo de desconfirmação* ou “*gap model*”, que sustenta que os consumidores ficam satisfeitos se o *desempenho percebido* excede (desconfirmação positiva) ou iguala as *expectativas* (confirmação) e insatisfeitos se o *desempenho observado* é inferior as *expectativas* (desconfirmação negativa) (Oliver, 1981b; Oliver & Desarbo, 1988). Assim, alguns estudos (Oliver, 1980, 1981a, 1981b; Tse & Wilton, 1988; Yi, 1990) explicam que os consumidores formariam *expectativas* pré-consumo, que quando comparadas com a *percepção de desempenho* no *encontro de serviços*, conduzem à *confirmação/desconfirmação* das *expectativas*.

Noutra perspectiva, Oliver e DeSarbo (1988) destacam algumas das principais teorias propostas para o desenvolvimento da teoria da *satisfação* ou *insatisfação* de clientes/consumidores: *Equity Theory*; *Attribution Theory*; *Performance Theory* e; *Expectancy Disconfirmation Theory (gap model)*, todas com vantagens e desvantagens.

Essas teorias foram concebidas no ambiente do marketing para estudos dos aspectos transacionais. Na *Equity Theory* relaciona-se ao princípio de que as saídas do processo de transação são proporcionais às entradas de cada uma das partes; na *Attribution Theory* pressupõe-se que a saída do processo de transação será avaliada em termos de sucesso ou de fracasso, e as causas destas saídas estão relacionadas a fatores internos, tais como as habilidades percebidas de compra ou de esforços, a fatores externos, como, por exemplo, a dificuldade relacionada ao processo de compra, ou a outros fatores, como a influência da equipe de vendas neste processo; na *Performance Theory* a satisfação do cliente está fortemente associada com a performance percebida do produto ou do serviço adquirido; e na *Expectancy Disconfirmation Theory*, talvez, a forma de mensuração de satisfação de clientes mais utilizada, parte-se do princípio de que os clientes criam um nível de *expectativa* em relação a produtos e serviços e que a *desconfirmação* (positiva ou negativa) pode gerar *satisfação* (desconfirmação positiva) ou *insatisfação* (desconfirmação negativa), definindo-se assim um “*gap*” ou hiato entre *expectativa* e *satisfação* do cliente (Oliver, 1980), denominado de modelo “*gap model*”.

Dessa forma, na linha teórica que se assumiu na presente investigação (*Expectancy Disconfirmation Theory*) abordou-se a qualidade percebida do serviço baseada no “modelo de discrepância” ou “*gap model*” adotado por diversos autores (Oliver, 1980, 1981a, 1981b; Parasuraman, Berry, & Zeithaml, 1985b; Parasuraman et al., 1988, 1991a, 1991b; Parasuraman

et al., 1994; Tse & Wilton, 1988; Yi, 1990), mais especificamente na discrepância entre *expectativa* (pré-consumo) e *desempenho* percebido (pós-consumo).

Atribui-se a Parasuraman *et al.*(1985a) a elaboração de um instrumento de mensuração de satisfação denominado de Escala *Service Quality - SERVQUAL*⁵⁴. Nessa escala, foi proposta a medição de qualidade do serviço, baseado no modelo “*gap model*” de satisfação de Oliver (1980), que afirma que a satisfação do cliente é uma função da diferença entre expectativa e desempenho. Ou seja, nesse modelo a posição da percepção do cliente sobre a qualidade de serviço percebida depende da natureza e da extensão da discrepância entre a expectativa de serviço e o desempenho percebido pelo usuário.

A perspectiva de avaliação da qualidade na metodologia do “*gap model*” é operacionalizada pela comparação algébrica entre a *expectativa* dos usuários-utilizadores e a *satisfação percebida* em organizações de serviços (Parasuraman et al., 1985a), que, Em seguida, transformou-se na Escala SERVQUAL (Parasuraman et al., 1988, 1991b), com posterior adaptação para aplicação em serviços de saúde (E. Babakus & G. Mangold, 1992), através de análises de confiabilidade e validade (por meio de Correlação e Análise Fatorial Exploratória) que concluíram que a Escala SERVQUAL é confiável e válida no ambiente hospitalar e em uma variedade de outros serviços de saúde.

Inicialmente a escala SERVQUAL foi desenvolvida com 97 itens, tomando como base 10 dimensões da qualidade (Parasuraman et al., 1985a; 1991ab) que consolidaram um instrumento com 22 questões. Por fim, essas 22 questões originais foram redistribuídas das dez (10) dimensões originais para as atuais cinco (5) dimensões da escala SERVQUAL (Berry, Parasuraman, Zeithaml, & Adsit, 1994; Parasuraman et al., 1991ab), conforme quadros 22 e 23 (anexo 2). Albuquerque (2012) caracteriza estas cinco dimensões da seguinte forma:

1. *Tangíveis visíveis*: está relacionado à atratividade das instalações, equipamentos e materiais usados por uma empresa de serviços, bem como à aparência dos funcionários do serviço.
2. *Confiabilidade/credibilidade*: significa que a empresa de serviços oferece a seus clientes serviço correto da primeira vez, sem cometer nenhum erro, e entrega o que prometeu dentro do prazo estipulado.
3. *Capacidade de resposta/prontidão*: significa que os funcionários de uma empresa de serviços estão dispostos a ajudar os clientes e atender aos seus requisitos, bem como informá-los, quando o serviço será prestado e então executá-lo com presteza.
4. *Segurança/domínio*: significa que o comportamento dos empregados transmitirá aos clientes confiança na empresa e que esta faz com que se sintam seguros. E também que os empregados são sempre corteses e têm o conhecimento necessário para responder às perguntas dos clientes.

54 Servqual – Escala Service Quality de avaliação de qualidade e satisfação em serviços (Parasuraman e Zeithaml, 1988; 1991)

5. *Empatia*: Significa que a empresa entende os problemas dos clientes e executa o serviço, tendo em vista seus melhores interesses, bem como lhes dá atenção pessoal, individual, e trabalha em horários convenientes.

Nesta investigação a partir das 10 dimensões e das 22 questões originais da escala SERVQUAL, através de painéis de especialistas e grupos pilotos de profissionais, gestores e usuários, no âmbito de hospitais, policlínicas, unidades especializadas da rede de saúde, esses atores-chave estabeleceram constructos de qualidade e questões validadas classificadas em cinco macrodimensões similares às da escala SERVQUAL (*Tangibilidade, Confiabilidade, Segurança, Presteza e Empatia*), além de sinalizar a importância de uma nova macrodimensão denominada de Acessibilidade com aspectos de garantia de acesso oportuno, efetivo e eficaz ao atendimento em saúde.

Neste processo, há de se considerar algumas limitações, restrições e cautelas na aplicação do instrumento e análise de resultados obtidos. Um exemplo de restrição analítica é que a qualidade de serviços observável ou perceptível é conceituada como um julgamento global, enquanto a satisfação é relacionada a um fato ou transação específica (Parasuraman et al., 1985a).

Noutro prisma, ao comparar a escala SERVQUAL com outras escalas, por exemplo, a escala *SERVPERF*⁵⁵ (Cronin & Taylor, 1992), a escala SERVPERVAL (Chahal & Kumari, 2011, 2012) e com o Instrumento de Avaliação de Qualidade Hospitalar (IAQH) (P. L. Ferreira, 1993; P. L. Ferreira et al., 2001; P. L. Ferreira & Mendes, 1997; P. L. Ferreira & Raposo, 2006; P. L. Ferreira, Raposo, & Pisco, 2017; Pedro et al., 2010), pode-se constatar algumas debilidades da escala SERVQUAL, que exigem cautelas em sua aplicação e análises de resultados.

Na verdade, qualquer que seja o método selecionado, enfrenta-se pelo menos três desafios na medição da expectativa dos clientes-consumidores. O primeiro deles reflete as dificuldades para investigar a fidedignidade e a validade de escalas usadas para medições diretas de constructos como a satisfação, que é um constructo complexo e multifacetado, e, reflete atributos de produtos ou de serviços, mas depende da relação preço/valor e imagens mentais do cliente (Dillon, 1997).

Outro desafio ocorre por usuários diferentes dar importâncias relativas diferentes para as diversas dimensões contidas na satisfação (produto, serviço, comunicação, etc), e assim, para se calcular o índice de satisfação geral dos clientes-consumidores, há de se considerar as

55 ServPerf – Escala Service Performance de avaliação de qualidade e satisfação em serviços (Cronin e Taylor, 1992)

respostas a todas as variáveis apresentadas no modelo, bem como suas diferentes importâncias relativas.

Há, ainda, o desafio da necessidade de definição e inclusão de variáveis de “fechamento” de cada bloco/dimensão e da satisfação geral que visam estabelecer uma situação de dependência das variáveis para o emprego da técnica de análise estatística. Sem estas variáveis seria impossível estimar-se a importância de cada variável na satisfação com a análise de uma dimensão ou com a satisfação geral.

Além de medir a satisfação dos clientes-consumidores com cada atributo, um desafio adicional pode ser definir a importância relativa entre os atributos (Fontnote, Henke, Carson, & Carson, 2007), que pode ser obtida de duas formas: *Declarada* ou *Derivada*. A *Importância Declarada* é aquela na qual se solicita ao próprio entrevistado que avalie a importância dos atributos, para tal, geralmente se recorre a determinadas escalas, como a de *diferencial semântico*, de *ranking* e de *soma constante*; enquanto a *importância Derivada* é obtida a partir da associação por métodos estatísticos entre a avaliação de cada atributo e a avaliação geral.

Quanto a seleção do método estatístico, Hanson (1992) sugere quatro formas a fim de analisar a *Importância Derivada*, ou seja, a mensuração de associação avaliação de cada atributo com a avaliação geral do usuário: pelo coeficiente de correlação de Pearson (r); pelo coeficiente da regressão linear (β); pela multiplicação do coeficiente da regressão linear (β) e o coeficiente de correlação de Pearson (r); e, pelo coeficiente de determinação parcial.

Dessa forma, nesta investigação mediu-se a *Importância Declarada* por uma escala de *Likert* de 5 níveis (*diferencial semântico*) além dos cálculos da *Importância Derivada* através de técnicas de estatística descritiva e inferencial para avaliar os itens da escala de qualidade percebida concebida (tópicos 6.2 e 6.4).

3. EVOLUÇÃO RECENTE DA AVALIAÇÃO NO CONTEXTO DE GOVERNANÇA DA SAÚDE PÚBLICA DO BRASIL

A fim de alinhar esta investigação com o contexto da saúde pública brasileira, a seguir, procedeu-se um breve relato da evolução recente de governança, mapeou-se o arcabouço legal para o campo de avaliação e monitoramento no SUS e caracterizou-se as fontes de dados e informações disponíveis e relevantes de saúde pública.

3.1 Breve históricos da governança e da avaliação no sistema de saúde brasileiro

Segundo a maioria dos historiadores, a saúde pública brasileira teve o início de sua história em 1808 (Brasil, 2011c), com a chegada da família real de Portugal. Desta época até os anos iniciais de 1900, a saúde pública limitava-se a vertente campanhista em busca de combater doenças e garantir o potencial exportador do País.

O modelo assistencial campanhista-sanitarista mostrou suas debilidades. Não havia sido estruturado um sistema de saúde. As péssimas condições de saúde da ampla maioria da população e a crescente insatisfação dos trabalhadores dos setores exportadores levaram a aprovação da Lei Elói Chaves, em 1923, que inaugurou o início da responsabilização do Estado pela saúde e a transição para o modelo assistencial médico sanitarista. Nas décadas posteriores o estado brasileiro buscou estruturar diversos modelos de governança, com sucessivas crises e ajustes no sistema de saúde.

Em 1953, criou-se o Ministério da Saúde do Brasil (MS) para coordenar as ações campanhistas de combate às doenças, principalmente no interior do Brasil, mas a assistência hospitalar estava concentrada em Instituições Filantrópicas e sem fins lucrativos. O governo incentiva a expansão da rede de assistência à saúde liberação de recursos financeiros governamentais, para o setor privado construir e equipar hospitais, com foco na incorporação tecnológica e ampliação da rede assistencial.

No início da década de 1960, a cobertura assistencial é limitada e a sociedade exige a ampliação dos direitos sociais. Nesse ambiente de pressão social, no contexto do golpe militar de 1964, o governo avança no processo de unificação de todos os IAP's existentes, e em 1966, todos são unificados pela criação do Instituto Nacional de Previdência Social (INPS)⁵⁶.

56 INPS - Instituto Nacional de Previdência Social

Outro marco da história da saúde brasileira, no âmbito ministerial, ocorreu em 1963, com a realização da 3ª. Conferência Nacional de Saúde (CNS)⁵⁷, que propôs a reordenação dos serviços de assistência médico-sanitária, através da Reforma Administrativa Federal. Assim, em 1967, ficou estabelecido que o MS assumiria a coordenação da Política Nacional de Saúde, que até então não havia saído do papel, com as seguintes áreas de competência: política nacional de saúde; atividades médicas e paramédicas; ação preventiva em geral, vigilância sanitária de fronteiras e de portos marítimos, fluviais e aéreos; controle de drogas, medicamentos e alimentos e pesquisa médico-sanitária.

Logo após a redemocratização do Brasil, ocorrida em 1985, o movimento de reforma da saúde teve seu ápice na VIII CNS – Conferência Nacional de Saúde, em 1986 (Brasil, 1986b), cujas diretrizes aprovadas serviram de esteio para os artigos 196-200 da atual Constituição Brasileira de 1988 (Brasil, 1988) e apontaram para a reestruturação do Sistema Nacional de Saúde e fomentaram a criação do SUS (Brasil, 2011c, 2011f).

Há de se reconhecer que a garantia da eficiência, da efetividade, da eficácia e da qualidade da atenção apresentam-se como os grandes desafios atuais do SUS, à luz dos princípios constitucionais (Brasil, 1988 Arts. 196-200) de integralidade, universalidade, equidade e participação social, além dos princípios organizativos do SUS (Brasil, 2011c, 2011f), conforme figura 7, abaixo:

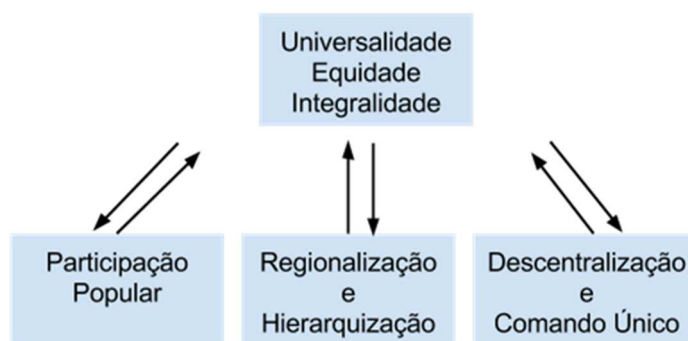


Figura 7 - Princípios doutrinários e organizativos do SUS.

Enquanto isso, o elevado volume de recursos aplicados na saúde, por si, exige avanços no monitoramento de resultados. Dados relativos a 2016 apontam que 35 países da OCDE⁵⁸ gastam em torno de 9,0% do PIB⁵⁹ (GDP) no sistema de saúde, nos Estados Unidos cerca de 17,2%, em Portugal 8,9%, e no Brasil 6,2% (OECD, 2017), Além disso, os gastos com a saúde

57 CNS - Conferência Nacional de Saúde, realizada a cada quadriênio com representação da sociedade, dos profissionais e dos governos de todo o Brasil.

58OCDE - Organização para a Cooperação e o Desenvolvimento Econômico (OECD), países da Europa

59PIB – Produto Interno Bruto (GDP) somatório de todas as riquezas geradas em um país

têm crescido nos últimos anos (2003-2016) em média de 3,6% nestes países (OECD, 2017) (figura 8).

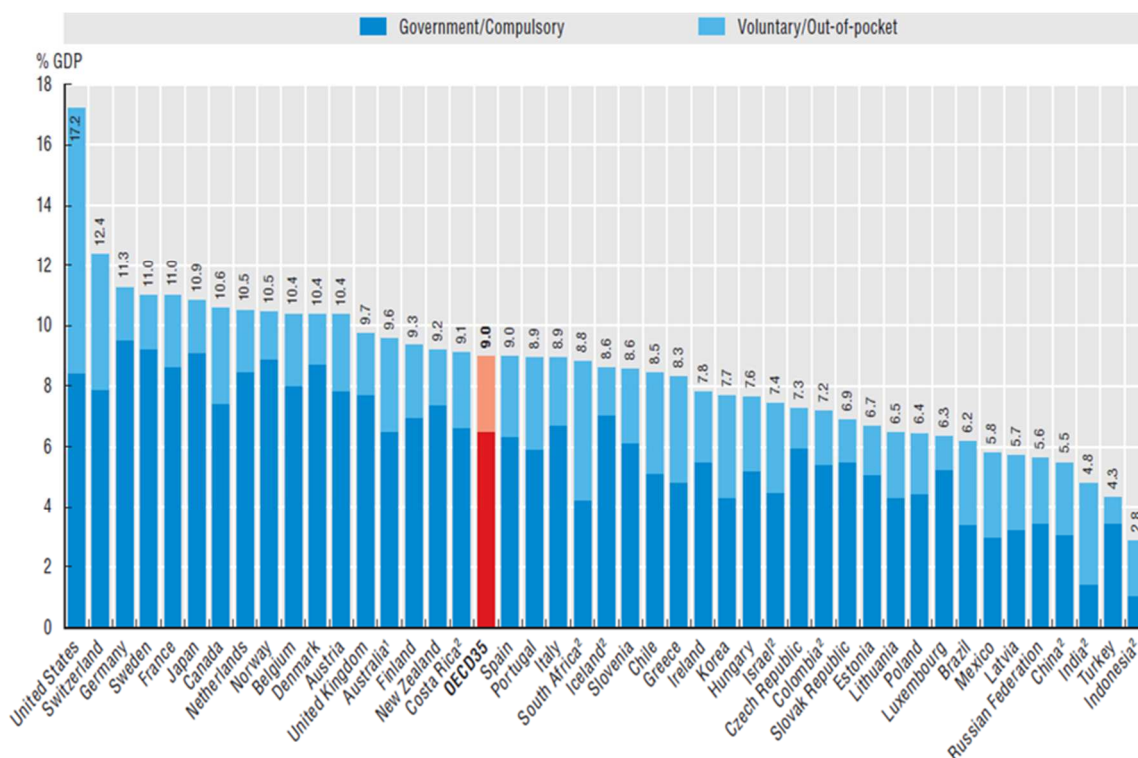


Figura 8 - Gastos em saúde: percentual do PIB (GDP), países OCDE e outros selecionados (2016 ou ano mais próximo). Fonte: *OECD Health Statistics 2017. WHO Global Health Expenditure Database.*
StatLink <http://dx.doi.org/10.1787/888933604229>.

Outro fator a considerar é a tendência mundial de envelhecimento da população, que é mais forte em países da Europa e América do Norte, mas o instituto oficial de estatística do governo brasileiro (IBGE)⁶⁰ divulgou informações sobre a inversão na pirâmide etária quanto a evolução populacional por grupos etários no Brasil (2000-2060), por indicar uma forte tendência de maior participação de idosos na população do país, conforme a realidade de muitos países (figura 9).

60 IBGE - Instituto Brasileiro de Geografia e Estatística

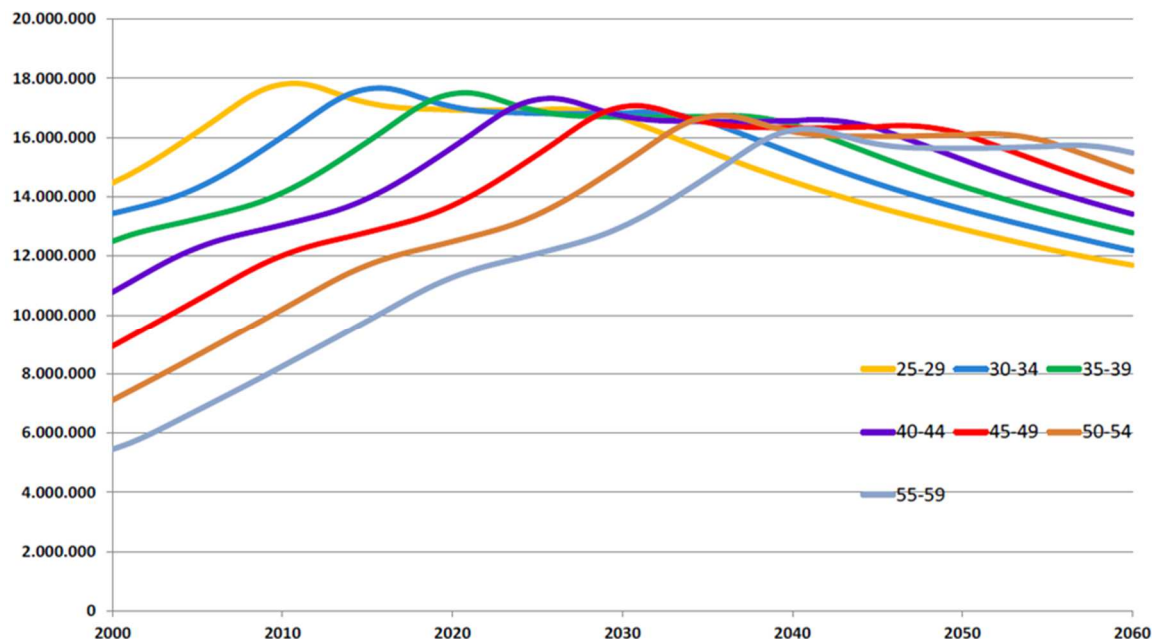


Figura 9 - Evolução populacional por grupos etários no Brasil (2000-2060). Fonte: IBGE. Diretoria de Pesquisas. Coordenação de População e Indicadores Sociais, 2013.

<http://www.ibge.gov.br/home/presidencia/noticias/imprensa/ppts/00000014425608112013563329137649.pdf>

É relativamente óbvio que os gastos com saúde se elevam ao longo da vida das pessoas, ou seja, quanto maior a idade média da população, maiores serão os gastos médios com a saúde das pessoas.

Noutra análise quanto a magnitude dos gastos por tipologia de despesas do sistema de saúde em 32 países da OCDE, EUA e outros países selecionados, percebe-se a média dos principais gastos distribuídos pelas categorias de provedores de serviços e produtos de saúde (OECD, 2017), ou seja, hospitais (agudos e psiquiátricos) (38%), cuidados residenciais de longa duração (LTC) (8%), cuidados ambulatoriais (consultórios, especialistas, centros de saúde, *home care*) (26%), varejistas (por exemplo, farmácias) (18%) e outros (por exemplo, serviços de laboratórios e institutos de saúde pública) (9%) (figura 10).

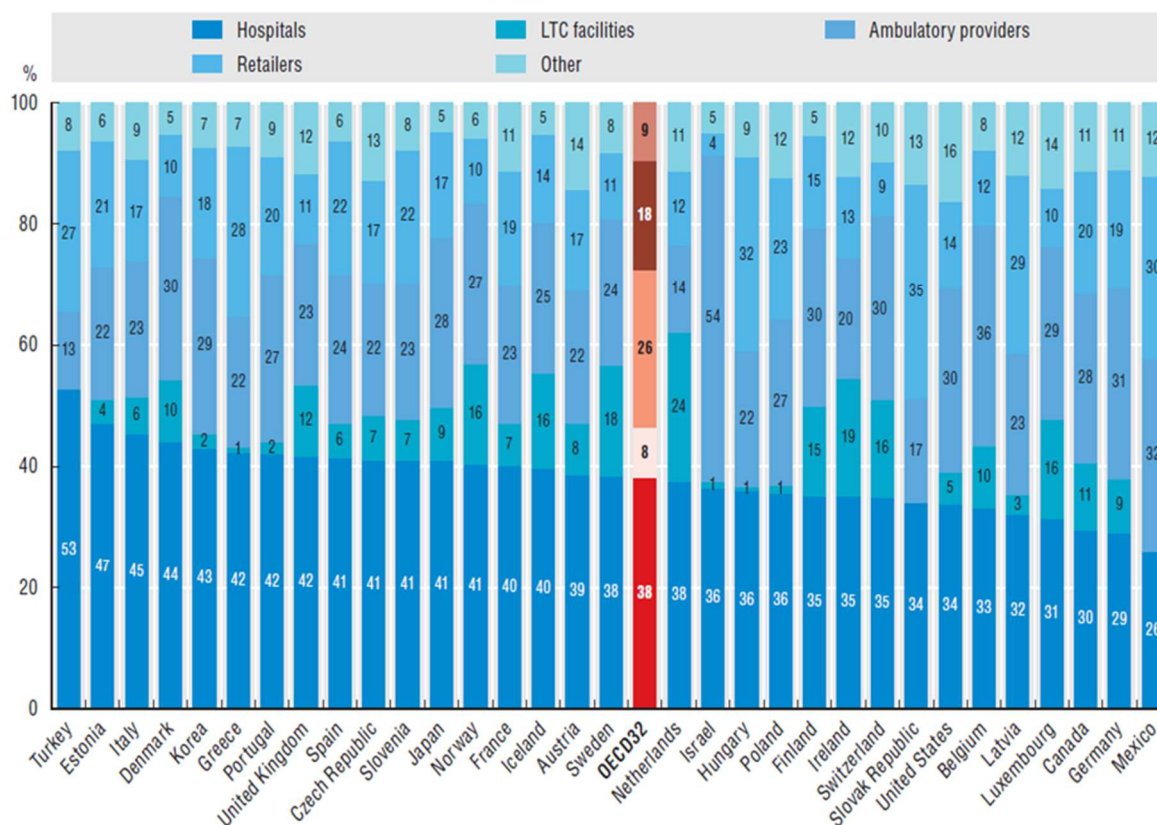


Figura 10 - Magnitude dos gastos por tipologia de despesas do sistema de saúde de países OCDE e outros selecionados (2015 ou ano mais próximo). Fonte: *OECD Health Statistics 2017*. WHO *Global Health Expenditure Database*. StatLink <http://dx.doi.org/10.1787/888933604438>.

Enquanto isso, dados de La Forgia e Counttolenc (2009), informam que o gasto total com atenção hospitalar no Brasil foi estimado em R\$ 47,3 bilhões (US\$ 16,1 bilhões) e representou 67% dos gastos totais de saúde do SUS (R\$ 70, 4 bilhões), em 2002, e, quando adicionados os hospitais privados, tem-se 58,1% de recursos do SUS, 33,4% de planos privados e 8,5% de pagamentos diretos por pacientes. No entanto, estas estimativas globais são de difícil contabilização, visto que os sistemas informatizados são de alimentação descentralizada por cada nível governamental e pode haver imprecisões na classificação contábil.

Noutro prisma, dados de La Forgia e Couttolenc (2009), mostram que no Brasil, em 2002, cerca de 25,9% foram aplicados em atenção básica (saúde primária), 50,9% em média complexidade (atenção secundária) e 23,2% em alta complexidade (atenção terciária), sendo portanto foco deste estudo os hospitais e serviços de saúde que atuam em atuação secundária e terciária, e portanto consomem mais de 74% dos recursos da atenção à saúde, sendo um forte indicador da importância de investigação de resultados no subsetor da saúde.

Nos últimos tempos a população mundial tem experimentado o fenômeno denominado de transição demográfica, que se explica pelo envelhecimento da população, visto que as taxas

de fecundidade têm caído vertiginosamente e os avanços da medicina têm ampliado a expectativas de vida das pessoas. Logicamente que o envelhecimento da população trás consigo um recrudescimento das doenças crônicas e amplia o desafio de se fornecer ações de saúde a esta população. Outro fator é que a transição epidemiológica, aponta para “as mudanças ocorridas, na frequência, na magnitude e na distribuição das condições de saúde e que se expressam nos padrões de morte, morbidade e invalidez que, em geral, acontecem, juntamente com outras transformações demográficas, sociais e econômicas” (Schramm et al., 2004).

Um dos maiores desafios de gestão em sistemas de saúde, em países em desenvolvimento como o Brasil, é a presença de uma tripla carga de doença (Schramm et al., 2004) pelo recrudescimento de doenças infectocontagiosas; epidemia de doenças oriundas da violência e condições de vidas urbanas; e avanço de doenças crônicas-degenerativas, em virtude da inversão da pirâmide etária, com tendência expressa pelos gráficos 3 e 4.

[...] doenças infecciosas, parasitárias e desnutrição, 14,7%; causas externas, 10,2%; condições maternas e perinatais, 8,8%; e doenças crônicas, 66,3%.

[...] envolve, ao mesmo tempo: uma agenda não concluída de infecções, desnutrição e problemas de saúde reprodutiva; o desafio das doenças crônicas e de seus fatores de riscos, como tabagismo, sobrepeso ou obesidade, inatividade física, uso excessivo de álcool e outras drogas, alimentação inadequada e outros; e o forte crescimento das causas externas”.

As doenças básicas foram praticamente erradicadas na maioria dos países, principalmente pelo avanço das vacinas, mas as doenças crônicas degenerativas, naturais das pessoas de mais idade, transformaram-se no maior desafio na gestão da saúde das populações mundiais e do Brasil. Segundo argumenta Mendes (2011) a aceleração da urbanização e de seus hábitos maléficos tem pressionados os sistemas de saúde.

No entanto, a realidade é que há uma crise dos sistemas de saúde que atualmente estão mais focados nas condições de doenças agudas (emergências). Assim, deixam a desejar no que tange a estruturação de redes de atenção dirigidas para doenças crônicas, em grande parte, oriundas do envelhecimento da população e da vida estressante das grandes cidades, principalmente pela pouca atenção as ações de promoção e prevenção e pela fragmentação do atendimento, que ameaça a garantia da integralidade de atenção a saúde, um dos preceitos do SUS (Brasil, 1990a). Neste contexto, os gestores são desafiados a construir e gerenciar redes integradas de serviços de atendimento a saúde, evitando a fragmentação da assistência atualmente existente.

Diante da situação epidemiológica mundial, cada vez mais se exige dos sistemas de saúde e dos serviços hospitalares respostas adequadas no sentido de promover a saúde, tomar

medidas preventivas e implantar sistemas eficientes, eficazes e efetivos na promoção, prevenção e recuperação da saúde da população. Assim, pode-se sugerir que há uma necessidade de sistemas robustos de avaliação em saúde que sejam capazes de monitorar a qualidade das redes de saúde e que possam orientar os gestores para tomar decisões custo-efetivas quanto aos resultados almejados pelos cidadãos-usuários das redes de saúde.

De acordo com a visão da NGP ou Gerencialismo, pode-se citar exemplos de projetos de avaliação e monitoramento de resultados da administração pública como PRODEV⁶¹ e P4R (*PforR*)⁶², que têm sido implantados na gestão pública de países em desenvolvimento, inclusive no Brasil, através de agências internacionais de fomento em busca de otimizar os recursos e garantir a qualidade às ações de órgãos públicos.

Desde a criação do SUS, muitos são os avanços do ponto de vista de garantias de atendimento, da organização sistêmica e da descentralização da gestão única, e organização sistêmica, através de normas organizativas infralegais de Ministério da Saúde a exemplo do atual “Pacto de Saúde 2006”, a fim de organizar a assistência à saúde nas bases de descentralização da atenção, garantia de resolutividade, redes regionalizadas de saúde e hierarquização do atendimento à população, além de garantir a governabilidade do sistema de saúde, porém, com tímidos avanços na governança do sistema de saúde.

Na área de saúde do Brasil, os esforços de governança têm sido majoritariamente focados nas políticas de saúde e sistemas coletivos de saúde, como os programas de atenção primária, de vacinação, de controle de mortalidade infantil e materna, dentre outras iniciativas sistêmicas. Por outro lado, a governança dos sistemas curativos e reabilitadores, no âmbito do SUS, prioriza a implantação de novas estruturas de atendimentos e quase nunca em gestão da performance destas organizações e serviços de saúde ou monitoramento de resultados das políticas, programas e ações de saúde.

Para fazer face ao desafio de se avaliar e monitorar desempenho e resultados nos sistemas e serviços de saúde vale lembrar, inicialmente, os argumentos de Santos-Filho (2007) que diz que, observa-se um alargamento da concepção de avaliação de intervenções em saúde, mesmo reconhecendo os diversos limites conceituais e de sua operacionalização no âmbito dos serviços.

Uma estratégia recente de governança tem a ver como os modelos gerenciais em uso nos hospitais e serviços de saúde. Atualmente no Brasil, há pelo menos sete modelos diferentes

61 PRODEV - Programa de Estratégia e Plano de Ação para a Efetividade do Desenvolvimento do BID para os países da América Latina e Caribe

62 PforR ou P4R - Programa para Resultados do Banco Mundial para o apoio ao desenvolvimento da “accountability” de países em desenvolvimento

de gestão de unidades de saúde e hospitais: administração pública direta, administração pública indireta (autarquias, etc), consórcios públicos de saúde, fundações públicas e privadas, contratualização com organizações sem fins lucrativos (entidades filantrópicas e organizações sociais), e com empresas privadas especializadas no ramo de saúde (Brasil, 2015c), e ainda, a tendência atual de parceria público e privada (PPP).⁶³

Sintetizou-se o contexto atual de governança da saúde brasileira, na figura 11, que apresenta o arranjo organizacional de governança dos sistemas subsistema público (SUS) correlacionando o subsistema suplementar e privado (operadoras de saúde) do Brasil e suas inter-relações interorganizações e unidades de saúde no âmbito do SUS.

Nessa conformação sistêmica, o subsistema público é composto pelo SUS, que será o contexto desta investigação, tem os níveis de governança organizacional federal, estadual e municipal, que se inter-relacionam de forma a compor um sistema único, mas com gestão descentralizada no nível da federação, dos estados (províncias) e dos municípios brasileiros. Por sua vez, qualquer intervenção externa entre as esferas de articulação interfederativa (Federal, Estadual, Regional e Municipal) dependente de pactuação voluntária entre estes entes, através de espaços de negociações permanentes nos níveis nacional, estadual e por regionais de saúde, sendo múltiplos os interessados chaves envolvidos nas decisões, que trabalham em mútua cooperação (Brasil, 2015b).

63 PPP - Parceria Público e Privada

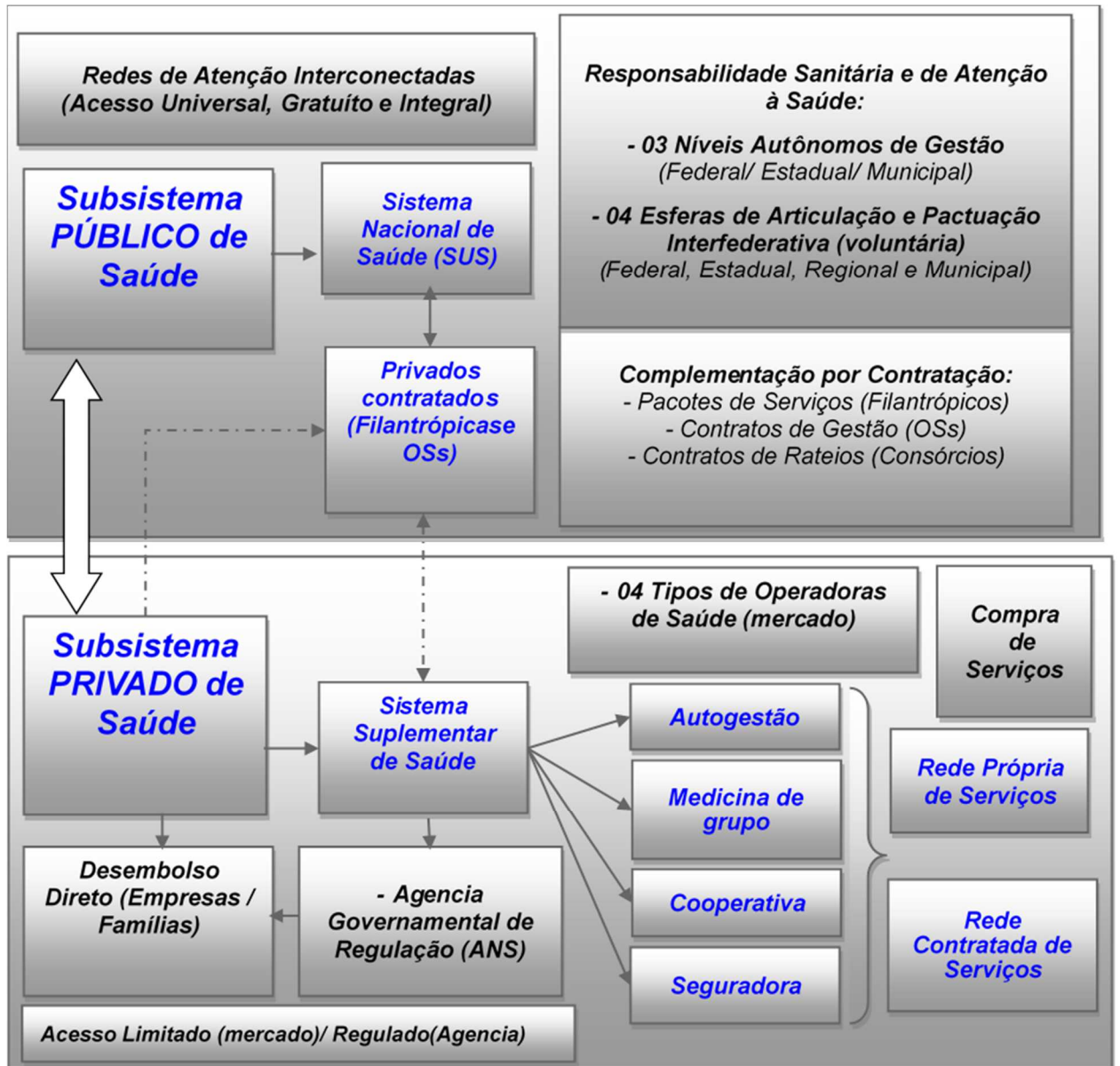


Figura 11 - O arranjo organizacional e as inter-relações do sistema público e privado de saúde do Brasil.

As leis orgânicas da saúde (nº. 8.080/1990 e nº. 8.142/1990) garantem a unidade da política de saúde, o repasse automático mensal de recursos, através de orçamentos globais provenientes do orçamento da previdência estatal, complementados pelos recursos obrigatórios de arrecadação de impostos estaduais (12% da receita própria) e municipais (15% da receita própria) e exigem a garantia de universalidade de acesso, equidade de assistência, integralidade de assistência, descentralização de ações, comando único em cada nível de gestão e rede hierarquizada de saúde, além de participação efetiva da sociedade na governança da saúde através de conselhos comunitários permanentes em todos os níveis de gestão (Brasil, 2011c, 2011f).

Em seguida, identifica-se o arcabouço legal e da avaliação e monitoramento da saúde do Brasil e, posteriormente, caracteriza-se os parâmetros dos principais sistemas de dados publicizados, no âmbito do SUS.

3.2 O arcabouço legal da avaliação e monitoramento no âmbito do Sistema Nacional de Saúde (SUS) do Brasil

Pode-se supor que os fatores econômicos, de necessidade de acesso e de exigência de qualidade de atenção à saúde reafirmam a importância da avaliação e monitoramento, em busca de eficiência, efetividade e satisfação de usuários-utilizadores das organizações de saúde.

Apesar das iniciativas de avaliação e monitoramento do sistema de saúde brasileiro serem incipientes, este campo é preceituado em diversas legislações brasileiras. Para início, os atuais pilares constitucionais do SUS assentam-se nos artigos 196 a 200 da Constituição Federal do Brasil (Brasil, 1988). Quanto aos aspectos genéricos de fiscalização e controle do sistema de saúde, a ordenação jurídica encontra base legal no artigo 197 da CF/1988 (Brasil, 1988) que, além disso, reafirma a relevância pública destas iniciativas. Estes dispositivos constitucionais foram regulamentados pela Lei Orgânica de Saúde nº. 8.080/1990(Brasil, 1990b), complementada pela Lei nº. 8.142/1990(Brasil, 1990c).

No propósito deste trabalho, destacam-se os aspectos de fiscalização e controle que desaguaram em mecanismos de planejamento e avaliação oriundos da Lei nº. 8.080/1990(Brasil, 1990b), em especial em seus artigos 15, 16 e 18:

[...] Art. 15. A União, os Estados, o Distrito Federal e os Municípios exercerão, em seu âmbito administrativo, as seguintes atribuições:

[...] Inciso VIII - elaboração e atualização periódica do plano de saúde (PS);

[...] Art. 16. Caberá à União:

[...] Inciso XVIII - elaboração do planejamento estratégico nacional no âmbito do SUS, em cooperação com os estados, os municípios e o Distrito Federal;

[...] Art.18. A direção municipal do Sistema Único de Saúde (SUS) compete:

[...] Inciso I – Planejar, organizar, controlar e avaliar as ações e os serviços de saúde e gerir e executar os serviços públicos de saúde (Grifo nosso).

No contexto das questões de governança do SUS destacam-se a implantação da política conhecida como PlanejaSUS através das Portarias nº. 3.085/2006(Brasil, 2006a) e nº. 3.332/2006(Brasil, 2006b), - com incentivos financeiros para implementação do Ministério da Saúde - Portarias nº. 376/2007(Brasil, 2007) e nº. 1.885/2008(Brasil, 2008b) - segundo Rocha et al. (2012), que iniciaram um movimento de normas orientadoras de planejamento, controle e avaliação do SUS, infelizmente revogadas e substituídas em um escopo reducionista,

posteriormente, através de outra norma infralegal, a Portaria nº. 2.135/2013(Brasil, 2013c), tendo consolidado apenas as linhas gerais para dois instrumentos estruturantes do PlanejaSUS, a exigência de Plano de Saúde (PS), e dos Relatórios Anuais de Gestão (RAG), com três sessões ou dimensões: análise situacional; definição das diretrizes, objetivos, metas e indicadores; e o processo de monitoramento e avaliação, além do desdobramento anual do PS em Programações Anuais de Saúde (PAS).

Art. 3º § 3º A elaboração do Plano de Saúde será orientada pelas necessidades de saúde da população, considerando: I - análise situacional, II - definição das diretrizes, objetivos, metas e indicadores; e III - o processo de monitoramento e avaliação.

Esta mesma portaria (Brasil, 2013c), apesar de não assegurar mecanismos de garantias operacionais, orienta sobre as dimensões mínimas que devem ser contempladas no Relatório Anual de Gestão (RAG):

Art. 6º O Relatório de Gestão é o instrumento de gestão com elaboração anual que permite ao gestor apresentar os resultados alcançados com a execução da PAS e orienta eventuais redirecionamentos que se fizerem necessários no Plano de Saúde.
 § 1º O Relatório de Gestão contemplará os seguintes itens: I - as diretrizes, objetivos e indicadores do Plano de Saúde; II - as metas da PAS previstas e executadas; III - a análise da execução orçamentária; e IV- as recomendações necessárias, incluindo eventuais redirecionamentos do Plano de Saúde.

Mais recentemente, também se registra evidência de avanços em aspectos pontuais de Gerencialismo e *accountability* em espaços de pactuação cooperativa, por exemplo, pela portaria nº. 4.279/2010(Brasil, 2010a) que instituiu diretrizes indutoras para organização das Redes de Atenção à Saúde (RAS) do SUS, além de duas recentes resoluções nº. 23/2017(Brasil, 2017e) e nº. 37/2018(Brasil, 2018) induzidas pelo Ministério da Saúde (MS) e emanadas da Comissão Intergestores Tripartite (CIT) com indícios de Gerencialismo e *accountability*(Hood, 1990, 1991; Simonet, 2011) ao definir alvos para redimensionamento da regionalização de saúde, inclusive apontando uma tendência de exigências quanto ao escopo, cobertura e governança das RAS do SUS, em todos os territórios de saúde.

A despeito deste arcabouço legal o planejamento, o monitoramento e a avaliação do SUS sempre dependeram fortemente da ação da participação social, por meio dos conselhos municipais, estaduais e nacional de saúde, que detêm o poder legal de avaliar o PS e o RAG de cada município, estado e da união. Via de regra estes conselheiros têm qualificação insuficiente para exigir metodologias adequadas de avaliação e monitoramento de resultados do SUS, sendo uma importante limitação do sistema de *accountability*.

Neste propósito, destaca-se ainda o decreto presidencial nº. 7.508/2011(Brasil, 2011a), ainda em vigor – apesar da reduzida implementação, que dispõe sobre a organização do SUS, o planejamento da saúde, a assistência à saúde e a articulação interfederativa, tendo estabelecido diretrizes nacionais para a fiscalização, o controle, a avaliação e o monitoramento de resultados na rede de saúde regionalizada, através de dispositivos expressos no Contrato Organizativo de Ação Pública de Saúde (COAP), conforme recorte abaixo explicitado:

Art. 35. [...] § 2º O desempenho aferido a partir dos indicadores nacionais de garantia de acesso servirá como parâmetro para avaliação do desempenho da prestação das ações e dos serviços definidos no Contrato Organizativo de Ação Pública de Saúde em todas as Regiões de Saúde...

[...] Art. 36. [...] Inciso IV - indicadores e metas de saúde;

[...] Inciso VI - critérios de avaliação dos resultados e forma de monitoramento permanente...

Por sua vez, o MS lançou o programa PMAQ(Brasil, 2012a), que avalia aspectos estruturais e de processo de atendimentos das unidades básicas de saúde, já foi aplicado em dois ciclos (2011/2012 e 2013/2014). O 3º. ciclo do PMAQ (2015/2017), ainda em fase de tabulação, pela primeira vez contará com um quadro de indicadores de resultados das unidades básicas de saúde (Brasil, 2017a, 2017d), mas de forma incipiente.

Mais recentemente, no âmbito da avaliação de performance de serviços públicos de saúde, o MS resolveu reinstitucionalizar o programa novo PNASS, através de portaria ministerial, onde se destacam alguns aspectos legais(Brasil, 2015e, artigos 2º., 3º. e 4º.). Nesta recente implantação do novo PNASS, estabeleceu-se um alvo de aplicação em 2.589 hospitais e organizações especializadas de saúde, que engloba as unidades que receberam aportes significativos de recursos de média e alta complexidade (MAC) do SUS, nos últimos anos(Brasil, 2015e). Os resultados do novo PNASS estão em fase de tabulação para posterior divulgação pelo MS.

Para além dessa legislação federal, registra-se ainda um arcabouço de legislações subnacionais que regulam o processo de avaliação e monitoramento, que fogem ao escopo deste trabalho. No entanto, neste escopo de legislação e programa subnacional, deve-se destacar a implementação do programa governamental “Choque de Gestão” na Secretaria de Saúde do Estado (província) de Minas Gerais (SES/MG), enquanto desdobramento de um planejamento estratégico governamental centrado em resultados (Guimarães & Tavares-De-Almeida, 2006).

Enquanto avanço do programa, segundo afirma Junior (2009), a partir de 2007, o governo de Minas Gerais definiu como uma de suas macrodiretrizes a implantação do programa “Choque de Gestão de 2ª geração”. A equipe de implantação detalhou todas as ações de

implementação desse programa governamental no bojo da SES/MG, que englobou as diversas áreas da governança do sistema de saúde de Minas Gerais (Minas, 2010b).

Ainda ao nível subnacional, a título de ilustração, pode-se citar a Lei nº 15.356/2013, na qual o Governo do Estado (província) do Ceará estabeleceu regulamentação específica de governança dos Contratos de Gestão de Organizações Sociais do terceiro setor, dentre elas as que operacionalizam hospitais e serviços de saúde, na qual pode-se exemplificar o disposto no item I, do artigo 6º. que diz “*O Contrato de Gestão [...] deve conter cláusulas estabelecendo [...] metas, prazo de execução e critérios objetivos de avaliação de desempenho, mediante indicadores de eficiência e eficácia*”.

No entanto, são incipientes os registros de contratualização e monitoramento por resultados da rede de serviços de saúde na ampla maioria dos estados (províncias) e dos municípios brasileiros – nem mesmo ao nível federal –, apesar da tendência crescente de parcerias público-privadas para operacionalização de hospitais e unidades de saúde, sendo excepcionais os relatos de experiências de avaliação e monitoramento de metas e resultados das organizações públicas e da rede de saúde (anexos 5 e 6).

Diante de tais legislações pode-se articular que há um arcabouço de normas legais sobre avaliação e monitoramento de serviços de saúde que estão na agenda nacional, com avanços e desafios relevantes, além de se perceber a incipiência da gestão por resultados.

3.3 As principais fontes de dados e informações do sistema de saúde do Brasil

Na perspectiva de Dudziak (2001), alguns dos objetivos da competência informacional consistem em formar indivíduos que:

[...] saibam determinar a natureza e a extensão de sua necessidade de informação como suporte ao processo inteligente de decisão;
 [...] sejam capazes de identificar fontes de informações de forma efetiva e eficaz;
 [...] avaliem criticamente a informação segundo critérios de relevância, objetividade, pertinência, lógica, ética, incorporando as informações selecionados ao seu próprio sistema de valores e conhecimento...

Pode-se articular que, no contexto de avaliação e monitoramento em saúde, a competência informacional e a capacidade de obter dados e informações relevantes além de pertinentes, quanto ao objeto de estudo, torna-se base para a inteligência decisória.

Neste contexto, os Sistemas de Informações em Saúde (SIS)⁶⁴ são desenvolvidos e implantados com o Objetivo de facilitar a formulação e avaliação das políticas, planos,

64 SIS- Sistemas de Informações em Saúde

programas, além da governança de unidades e serviços de saúde, subsidiando a tomada de decisões e contribui para melhoria da situação de saúde individual e coletiva. Na perspectiva de Marin (2010 apud Brasil, 2015b):

Sistemas de informação em saúde (SIS) podem ser definidos como um conjunto de componentes inter-relacionados que coletam, processam armazenam e distribuem a informação para apoiar o processo de tomada de decisão e auxiliar no controle das organizações de saúde.

[...] congregam um conjunto de dados, informações e conhecimento utilizados na área de saúde para sustentar o planejamento, o aperfeiçoamento e o processo decisório dos múltiplos profissionais da área da saúde envolvidos no atendimento aos pacientes e usuários do sistema de saúde

Os SIS suportam as funções de manipulação e tratamento de dados: coleta, aquisição, registro, armazenamento, processamento, recuperação, análise e difusão de dados, além da geração de informações. Sendo assim, subsidiam algumas funções governacionais: planejamento, coordenação, supervisão dos resultados dos processos de trabalhos, além da análise situacional da saúde e tomada de decisão em sistemas e unidades de saúde.

Tradicionalmente os SIS do Brasil, no âmbito do SUS, são fragmentados (Brasil, 1994), visto que, historicamente, foram conformados de acordo com necessidades específicas e iniciativas isoladas de diferentes áreas, e não por meio de uma estratégia mais ampla que considerasse o setor saúde como um todo (Brasil, 2009a, 2015f), estando em contínua evolução.

A institucionalização dos SIS, no âmbito do SUS, iniciou-se em 1971, quando foi criado o Núcleo de Informática do Ministério da Saúde, transformado, em 1991, no Departamento de Informática do SUS (Datusus), vigente até os dias atuais. Um marco da iniciativa organizativa foi a 1ª. reunião nacional de sistemas de informações em saúde, durante a Conferência Nacional de Saúde, em 1975, sendo que as bases dos principais SIS do SUS foram formatadas entre as décadas de 70 e 80 (Brasil, 2009a), estando em contínua evolução.

Esses sistemas de informações do SUS foram consolidados ao longo das três últimas décadas, tendo os primeiros módulos hospitalares (SIH – Sistema de Informações Hospitalar e SIA – Sistema de Informações Ambulatoriais) instituídos através da resolução nº 258/1991 do antigo Inamps, com posterior integração com os outros sistemas do Datusus até a conformação atual estabelecida na Política Nacional de Informação e Informática em Saúde (PNIIS) aprovada na Conferência Nacional de Saúde em 2004 (Brasil, 2009a), a fim de consolidar dados e informações para a gestão da saúde no Brasil.

Enquanto isso, o Datusus do Ministério da Saúde universalizou a coleta de dados e informações de saúde (portaria nº 221/1999), a partir de 2000, ancorado na lei de regulação de

saúde complementar nº 9.656 de 1998 (Brasil, 1998), quanto a exigência a todos os estabelecimentos de saúde para o fornecimento de informações dos pacientes cuja atenção é custeada por planos e seguros privados de assistência à saúde. Neste objetivo, estruturou-se o sistema CIH, que por uma década, recolhia dados apenas das internações de hospitais privados. Assim, recentemente (portaria nº 1.171/2011)(Brasil, 2011d), ampliou-se a recolha de dados, pela estruturação do sistema CIHA, a fim de controlar os registros de ações e serviços de hospitalares e ambulatoriais executados por quaisquer pessoas naturais ou jurídicas, de direito público ou privado, financiadas com recursos de fontes externas ao SUS.

Apenas nos anos 2000, o MS iniciou a unificação do processo de cadastramento de estabelecimentos de saúde, através da criação do Sistema de Cadastro Nacional de Estabelecimentos de Saúde (SCNES), que teve a 1ª. fase concluída em 2003 (Brasil, 2015f). Após 2006, o cadastro no SCNES torna-se obrigatório para qualquer estabelecimento de saúde do Brasil, tendo fechado o ano de 2015 com mais de 300 mil unidades cadastradas (Brasil, 2015f).

Em 2002, o MS implantou o Sistema Nacional de Regulação (SISREG) para apoiar a organização e gestão dos processos de regulação, concretizar os fluxos assistenciais e apoiar a gestão de acesso a unidades de saúde, por estados e municípios, sendo integrado, após 2008, ao SCNES e ao SIGTAP.

As tabelas (SIA/SUS e SIH/AIH) de reembolso por procedimentos do SUS, historicamente, foram formatadas de forma pulverizadas. A fim de consolidar os códigos de procedimentos do SUS, a tabela SIGTAP (Tabela de Procedimentos, Medicamentos, Órteses, Próteses e Materiais Especiais – OPM) oficializada pela portaria nº 2.848/2007(Brasil, 2015f), implementada a partir de janeiro de 2008, operacionalizou uma profunda reforma nos códigos dos mas dificultou a equalização de informações de produção antes e depois da SIGTAP. Houve um período de adequação da coleta de dados, sendo os dados mais confiáveis após o ano de 2010.

Ao longo deste período estruturou-se uma rede de coleta de informações em saúde, descentralizada para estados e municípios, com processamento de dados centralizado através dos sistemas gerenciados pelo Datasus, sendo os mais relevantes descritos, sinteticamente, no quadro 29 (anexo 5) (Brasil, 2004, 2009a, 2015f). Para obter uma visão global dos principais sistemas de informática do SUS pode-se consultar o relatório da Experiência Brasileira em Sistema e Informações em Saúde (Brasil, 2009a), que retrata marcos legais, desenvolvimento e uma análise crítica dos principais SIS institucionais do SUS. Além disto, apresenta a estruturação das ferramentas TabNet e TabWin, após a década de 90, a fim de facilitar o acesso

às informações geradas nos diversos SIS do SUS, em especial, para os níveis descentralizados, e permitir aos cidadãos consultas públicas, via WEB, de dados validados e consolidados da saúde pública e privada do Brasil.

Historicamente, o Brasil teve dificuldade de coletar dados e produzir informações validadas para o sistema de saúde. A fragmentação dos SIS, viés de coleta com enfoque no faturamento e dificuldades de registros automatizados, subnotificações, além de outros problemas são recorrentes na literatura (Brasil, 2009a, 2015f; De Moraes, 2014; J. C. Noronha, 2001). Possivelmente para fazer face a esses desafios, em 2003, a 12^a Conferência Nacional de Saúde (CNS) lançou as bases para a elaboração da Política Nacional de Informação e Informática em Saúde (PNIIS). Entre as ações de estruturação e organização destaca-se (De Moraes, 2014) a criação de um Comitê de Informação e Informática em Saúde (CIINFO/MS), em pleno funcionamento dentro do Ministério da Saúde, que enfrenta muitos desafios na missão de integração e interoperabilidade dos SIS do SUS.

Pode-se ter uma ideia do alcance e dos desafios da disponibilidade de informações válidas e relevantes para a construção de indicadores nos SIS no âmbito do SUS no estudo de Noronha (2001) que avalia alguns aspectos de produção e disponibilização de informações sobre internações. Destacam-se ao menos as seguintes limitações sintetizadas por Escrivão (2004):

A lista de procedimentos é incompleta, não diferenciando alguns procedimentos cirúrgicos segundo os portes das cirurgias; a classificação também não considera outras variáveis que diferenciem os pacientes quanto a sua gravidade clínica e quanto ao perfil de consumo de recursos como, por exemplo, a idade. Essa falha na classificação faz com que pacientes heterogêneos clinicamente sejam classificados num mesmo grupo, inviabilizando a realização de comparações. Por outro lado, o SIH pode ser útil para subsidiar a gestão do sistema de saúde, nos planejamentos financeiro, de insumos humanos e materiais, na avaliação dos cuidados prestados e para estudos epidemiológicos (Escrivão, 2004: 35-36).

Por sua vez, Felix et al. (2012) e Macente & Zandonade (2010), em estudos de óbitos e da completude dos dados pela declaração de óbito (DO) registrados no Sistema de Informações de Mortalidade (SIM) do Datasus, destacam assimetrias entre os preenchimentos de variáveis do sistema:

No que se refere a qualidade dos dados do SIM, várias pesquisas vêm mostrando que, para todos os tipos de óbitos, o preenchimento dos atestados de óbito não é satisfatório [...] no que diz respeito a melhora na DO no quesito causa base da mortalidade, pois houve uma redução significativa das “causas mal definidas” e por “diagnósticos incompletos”, mas ainda há um longo caminho a percorrer, pois existem questões epidemiológicas que necessitam ser melhoradas...

Um estudo de Pinto (2010a) de análise comparativa dos SIS do Brasil, Espanha e Portugal, que foram consolidados até os anos de 1990, reconhece a relevância da informação para subsidiar os gestores de saúde no processo decisório e na avaliação dos serviços e ações de saúde. Entre as muitas comparações dos SIS, inclusive quanto às possíveis limitações, os autores concluem sobre o SIH-SUS do Brasil:

A desvinculação do pagamento de serviços hospitalares da remuneração de produção, com o estabelecimento de orçamentação, incentiva a não apresentação da produção pelas unidades hospitalares. A subnotificação oriunda dos hospitais públicos federais orçados é notória.

[...] Não foram encontrados em Portugal e Espanha informações sobre subnotificação, devido, talvez, ao fato desses países estabelecerem os orçamentos hospitalares com base na produção apresentada pelos hospitais e no cumprimento de objetivos acordados.

A qualidade da informação clínica é outro grande problema no SIH-SUS, e as principais questões dizem respeito ao subregistro, ao número reduzido de variáveis e a codificação efetuada por profissional pouco qualificado na utilização da CID (Pinto, 2010a: 134).

Ao analisar diversos estudos publicados sobre os SIS do SUS, Bittencourt, Camacho e Leal (2006) destacam algumas potencialidades e desafios dos sistemas SIM; Sinan, Sinasc, por exemplo:

Predominaram os estudos que tiveram como Objetivo validar as informações geradas pelo SIM, Sinan e Sinasc. Nessa linha, em relação ao Sinasc e ao SIM, alguns estudos apontaram que o SIH pode ser uma boa estratégia para a captação de nascidos vivos e fornecer indícios para a investigação de morte materna em áreas de baixa cobertura. Em relação ao Sinan, foram identificados problemas no SIH-SUS para monitorar febre amarela, sarampo em adultos, meningite tuberculosa, hepatites A e B, para fornecer dados para preenchimento do formulário de vigilância epidemiológica das infecções hospitalares. Todavia, os resultados apontaram grande potencialidade do SIH-SUS para monitorar tuberculose pulmonar, hanseníase, meningite meningocócica, dengue, leishmanioses visceral e cutânea, doenças imunopreveníveis, cólera, febre tifoide e leptospirose (Brasil, 2009a: 59).

[...] Outro uso frequente feito a partir das informações aportadas pelo SIH é o que se refere à avaliação de desempenho de serviços, nas dimensões efetividade, adequação, acesso e eficiência.

[...] “por ser bem definido e de diagnóstico sem equívoco”, o óbito hospitalar tem sido a variável mais empregada em indicadores para medir efetividade (Brasil, 2009a: 61).

Reconhece-se (Brasil, 2009a: 62), ainda, que “a ausência de importantes variáveis clínicas e de detalhamentos de procedimentos médicos no SIH limita sua utilização para ajuizamento de adequação do tratamento dispensados aos pacientes

Por fim, destaca-se outro trabalho recente de Rocha et al. (2017) que em colaboração com o Ministério da Saúde comparou os dados do Cadastro Nacional de Estabelecimento de

Saúde (CNES) com observações *in loco* em 2.777 hospitais da amostra selecionada. Os autores articulam sobre uma razoável confiabilidade dos dados registrados no CNES, mas apontam algumas preocupações, notadamente, por se tratar de atributos físicos e estruturais que por questões lógicas deveriam ter muito mais precisão:

O status de funcionamento estava atualizado em 89% dos casos, o número de leitos em 44%, 82% mantinham o quantitativo de equipamentos correto, e 63% apresentavam coordenadas geográficas precisas.

Esta investigação não pretende explorar em profundidade os aspectos dos SIS do SUS, mas certamente deve-se considerar as limitações citadas acima, pois impactam em possíveis limitações e dificuldades para validar indicadores do SUS.

Em uma visão mais operacional, no contexto de tomada de decisão, a coleta de dados deve ser racional e objetiva, visando a construção de indicadores epidemiológicos ou operacionais que atendam aos objetivos de cada programa ou instituição, evitando-se descrédito do sistema e desperdício de tempo e recursos.

A precisão das informações disponíveis e confiáveis, por sua vez, depende, em grande parte, da qualidade dos dados gerados nos sistemas de informação de saúde, o que pode estar influenciado por múltiplos fatores técnicos e operacionais. O uso regular desses dados, por serviços e instituições acadêmicas da área de saúde, contribui decisivamente para o progressivo aprimoramento dos sistemas e bases de dados e, conseqüentemente, para a consistência das análises realizadas.

No entanto, do ponto de vista do processo de tomada de decisão, outro fator preponderante está na exigência de validação destas informações através de um processo sistematizado, que envolve o estudo do problema a partir de um levantamento de dados, da produção de informação, do estabelecimento de propostas de soluções, da escolha da decisão, da viabilização e implementação da decisão e da análise dos resultados obtidos.

Os principais SIS em funcionamento no Datasus foram amplamente caracterizados pela literatura (Brasil, 2009a, 2015b; De Moraes, 2014). Mais recentemente, o Ministério da Saúde iniciou um esforço para garantir novos avanços (Brasil, 2015f) como o Registro Eletrônico em Saúde para o SUS (RES-SUS), um ambiente que integre todas as informações assistenciais, o Conjunto Mínimo de Dados (CMD) e o estudo de um sistema de DRG, no âmbito de uma estratégia inovadora de informação em saúde (e-Saúde), que se encontra em fase de estruturação de sistemas e de novos instrumentos de coletas de dados, validação e construção de plataformas para disseminação de informações do SUS.

No contexto do presente estudo, destaca-se alguns dos principais SIS em funcionamento no Datasus enquanto fontes de informações para a composição de indicadores em saúde (quadro 29, anexo 5). Quanto às matrizes de indicadores, as iniciativas existentes no Brasil ainda são extremamente tímidas, sendo caracterizado neste estudo o rol de indicadores de saúde pública do Brasil mais relevante (PROADESS) (quadro 28, anexo 5), além dos mais relevantes da saúde privada brasileira (quadro 30 a 34, anexo 6).

4. FUNDAMENTOS E ESTRATÉGIAS DE ENVOLVIMENTO DOS DECISORES DE ORGANIZAÇÕES DE SAÚDE

Diante disto, percebe-se a complexidade do fenômeno em estudo e a subjetividade do processo avaliativo em organizações de saúde, no âmbito do SUS. Ciente deste desafio resolveu-se considerar as ações e os propósitos dos gestores dos níveis setoriais, organizacionais e governacionais, firmados no enquadramento conceitual do campo da avaliação participativa e construtivista de 4ª. geração proposta por Guba e Lincoln (2011).

Quanto aos aspectos *metodológicos*, no prisma do *paradigma construtivista*, propôs-se um processo integrativo e participativo da avaliação, em busca de construir consensos quanto a visão global, metodologias, macrodimensões e indicadores de medição de performance organizacional, no âmbito de unidades de saúde pública e/ou vinculadas ao SUS.

Diante dessa escolha, no presente tópico, parte-se da exploração da necessidade dos decisores de utilização da avaliação no contexto da tomada de decisão, para consolidar os fundamentos quanto a ação, os propósitos e o envolvimento dos principais decisores (*stakeholders*) em organizações de saúde e, caracterizar estratégias por método misto (qualitativo e quantitativo) de avaliação de construção de consensos a fim de estruturar o modelo teórico-lógico, validar as dimensões e indicadores e, por fim, operacionalizar a validação do Artefato construído para a mensuração de performance em saúde.

Uma das primeiras etapas operacionais e de aproximação ao objeto sob avaliação consta do mapeamento das partes interessadas (*stakeholders*) na avaliação, em especial, os agentes de decisão, que contribuem ativamente na construção e no desenho do modelo de avaliação.

Em seguida, deve-se construir um entendimento sobre a avaliação da performance organizacional e seus conceitos mais prevalentes na literatura vigente, a fim de propor parâmetros balizadores na condução do processo interativo que se pretende junto aos *key stakeholders* (decisores-chaves).

Para tal apropria-se de estratégias e instrumentos de abordagem mista (qualitativa e quantitativa) para captar as percepções e relevâncias atribuídas pelos agentes de decisão dos três níveis considerados nesta investigação.

A última etapa do envolvimento dos agentes de decisão é a validação de preferências de decisores quanto ao Artefato construído, e as provas de conceito, valor e usabilidade do *proxy* de indicadores para avaliação de resultados em organização de saúde.

4.1. A utilização da avaliação no contexto da tomada de decisão em sistemas e organizações de saúde

Nos dias atuais, os sistemas de saúde experimentam uma constante ampliação das disponibilidades de terapias, tecnologias e recursos aplicados à melhoria da atenção à saúde. Em paralelo a isto os gestores têm sido pressionados para a otimização e responsabilização de recursos, principalmente pela linha gerencial denominada *accountability* no bojo da Nova Gestão Pública aplicada aos sistemas públicos de saúde.

No contexto de exigências de garantias de provisão de serviços adequados e suficientes pelos usuários e de responsabilização de gastos pelo gerencialismo dominante na gestão pública, a avaliação para a tomada de decisão parece encontrar um campo fértil de crescimento e aplicação. No entanto, Contandriopoulos (2006: 709) lembra da complexidade desse processo de avaliação.

Ainda que todos estejam de acordo quanto à necessidade de aumentar a eficiência do sistema de saúde, existem grandes divergências sobre a maneira de fazê-lo e, em particular, sobre a pertinência da avaliação. Esta falta de consenso decorre da complexidade do sistema de saúde, que envolve grandes grupos de atores (os profissionais de saúde, os gestores, os planejadores, o mercado e a política), que intervêm segundo diferentes lógicas, níveis de atuação, cronogramas e valores.

As dificuldades de avaliação de programas e serviços em saúde coletiva foram descritas para o contexto Anglo-Saxão (Green & Kreuter, 1999; Rootman et al., 2004 apud Campos e Furtado, 2011). Estas dificuldades podem ser prazos relativamente restritos para a constatação de mudanças e a consideração de variáveis que se interpenetram, dentre outros.

Contandriopoulos (2006) cita que há necessidades de informação distintas dos atores da saúde que dificulta o consenso quanto ao método de avaliação a ser implantado. Cita ainda que os conceitos avaliativos são diferentes entre os atores envolvidos (governos, gestores, profissionais, etc).

Os diferentes atores atribuem ainda sentidos diversos à noção de eficiência.
[...] O que interessa ao governo é eficiência na alocação de recursos, ou seja, uma distribuição que permita atender da melhor forma possível ao bem-estar coletivo. [...] Para gestores e planejadores, distribuição de recursos eficiente é a que permite maximizar os benefícios para a saúde em relação a um determinado gasto.
[...] Na lógica econômica, somente é considerada eficiente uma distribuição de recursos que resulte no livre funcionamento do mercado.
[...] Para os profissionais de saúde, atuar de maneira eficiente significa fornecer a todos os serviços considerados necessários do ponto de vista clínico (Contandriopoulos, 2006: 709).

Assim, Contandriopoulos (2006: 710) afirma que um grande desafio para a avaliação é “incorporar nas suas estratégias os pontos de vista de atores em diferentes posições, a fim de fornecer às instâncias de decisão as informações de que precisam para fazer um julgamento o mais amplo possível”.

A despeito da capacidade de avaliação contribuir para o aperfeiçoamento do processo de decisão na saúde, isto se confronta com a complexidade relacional entre as diversas variáveis envolvidas no ambiente de serviços em saúde e os diferentes pontos de vistas dos atores interessados, além dos múltiplos fatores que condicionam a saúde e a doença. Talvez por isso, há evidências de que a utilização de avaliação é incipiente na gestão de serviços de saúde, a despeito da avaliação se constituir em um instrumento essencial de apoio à gestão pela sua capacidade de melhorar a qualidade da tomada de decisão (Tanaka & Tamaki, 2012b).

Em outro prisma, ao desenvolver uma abordagem da avaliação denominada *avaliação focada para a utilização*, Patton (1997) inclui na sua definição a maneira ou o método de fazer a avaliação que corresponderia à coleta sistemática de informações sobre as atividades, características e resultados dos programas, além do Objetivo ou propósito da avaliação, ou seja, fazer julgamentos sobre os programas e/ou para subsidiar o processo de tomada de decisões sobre futuras programações (Vieira-Da-Silva, 2005).

Assim, pode-se articular que os processos avaliativos se configuram como uma importante ferramenta de análise dos tomadores de decisão em busca de embasamento de suas decisões. Neste propósito, quanto a complexidade do processos decisórios, segundo Tanaka e Tamaki (2012b) o gestor utiliza sua competência formal (referências técnicas, políticas, institucionais, sociais, culturais entre outras), além das informações obtidas no processo avaliativo e da percepção que tem do problema para embasar suas decisões, haja visto que a tomada de decisão na gestão em saúde é complexa e permeada de subjetividade e incertezas (Paim & Teixeira, 2006).

Em outro prisma, Tanaka e Tamaki (2012b) esclarecem, no contexto de avaliação de serviços de saúde, que a avaliação para a tomada de decisão deve basear-se na utilização de métodos e técnicas de pesquisa na sua concepção, formulação e implementação, devendo ser estruturada sob seis atributos fundamentais: *utilidade, oportunidade, factibilidade, confiabilidade, objetividade e direcionalidade*.

A avaliação para a gestão de serviços de saúde é um processo técnico-administrativo e político de julgamento do valor ou mérito de algo, para subsidiar a tomada de decisão no cotidiano, o que significa produzir informações capazes de apoiar uma intervenção de forma

oportuna, e baseada na utilização de métodos e técnicas de pesquisa na sua concepção, formulação e implementação (Tanaka & Melo, 2008).

Quanto aos aspectos de aplicação de resultados da avaliação, uma das fases operacionais trata-se da seleção dos indicadores, que devem ser escolhidos pelo avaliador, envolvendo os principais atores interessados, de tal forma que possam modelar o objeto em avaliação o mais próximo possível da realidade. Neste propósito a Organização Mundial da Saúde (WHO, 1981) propõe que a escolha de indicadores deve considerar alguns atributos: *validade*, *fiabilidade*, *sensibilidade*, e a *especificidade*. Mas, há outros atributos desejáveis, por exemplo, *mensurabilidade*, *relevância* e, *custo-efetividade* (Brasil, 2002a).

A *validade* (medir efetivamente o que pretende medir), a *fiabilidade* (apresentar o mesmo resultado mesmo que seja utilizado por pessoas ou em circunstâncias diferentes), a *sensibilidade* (capacidade de captar mudanças na situação ou no objeto estudado), e a *especificidade* (refletir mudanças apenas no objeto que está sendo estudado) (WHO, 1981) (Grifos nossos).

[...] A depender da aplicação dos indicadores outros atributos devem ser incorporados aos indicadores: *mensurabilidade* (basear-se em dados disponíveis ou fáceis de conseguir); *relevância* (responder a prioridades de saúde); e *custo-efetividade* (os resultados justificam o investimento de tempo e recursos) (Brasil, 2002a) (Grifos nossos).

A etapa operacional seguinte do processo avaliativo constitui-se no julgamento de valor para subsídio à tomada da decisão. No entanto, a avaliação dos resultados obtidos dos indicadores para a emissão de juízo de valor e a precisão na tomada de decisão dependem dos parâmetros e dos padrões escolhidos. Tanaka e Tamaki (2012b) afirmam que a utilização de padrões como parâmetros trazem vieses na avaliação dos serviços de saúde, visto que as condições ideais em serviços de saúde são raras de acontecer e explicitam as diferenças entre estes conceitos:

Os parâmetros são índices quantitativos ou qualitativos estabelecidos como referência para serem alcançados com os esforços empreendidos para esse fim. Eles devem ser estabelecidos nas condições reais em que as ações são desenvolvidas. Um parâmetro se diferencia, portanto, de um padrão pelo fato de que este se refere a um índice estabelecido em condições ideais (Tanaka & Tamaki, 2012b: 826).

Diante disso, o estabelecimento dos parâmetros é uma atividade crítica na avaliação para a tomada de decisão, pois podem interferir diretamente nos resultados almejados, e sendo assim, Tanaka e Tamaki (2012b) propõem alguns atributos para estabelecimento adequado dos parâmetros:

Entende-se por adequado aquele parâmetro que: tenha correspondência com o indicador escolhido; considere os recursos disponíveis e/ou passíveis de serem mobilizados; seja aceito pelos gestores, técnicos e por outros interessados na avaliação; e seja factível de ser atingido com os esforços empreendidos para esse fim no contexto real em que a ação é desenvolvida (Tanaka & Tamaki, 2012b: 826).

O processo de tomada de decisão, propriamente dito, se inicia na fase de julgamento de valor ou mérito, que perpassa pela análise dos resultados obtidos pelos indicadores e do estabelecimento de parâmetros, e desagua na implementação de ações de mudanças da realidade, e por fim, pode institucionalizar métodos de monitoramento para garantir os resultados esperados (Contandriopoulos et al., 1997; Tamaki et al., 2012).

Sendo assim, faz parte do arsenal de atividades do avaliador/investigador conceber um processo avaliativo para a tomada de decisão, através da escolha de metodologias, indicadores e de parâmetros que contemple essa diversidade de pontos de vistas, através da formulação da melhor estratégia, na seleção da abordagem, na definição de níveis e atributos, bem como na seleção de critérios, indicadores e padrões (Vieira-Da-Silva, 2005 apud Tanaka e Tamaki, 2012), tendo em conta a validade, a credibilidade e aplicabilidade dos resultados.

Os resultados da avaliação têm que responder à questão que o gestor se coloca. Eles não se traduzem automaticamente em uma decisão, mas espera-se que as informações produzidas contribuam para o julgamento de uma determinada situação com maior validade, influenciando positivamente as decisões, para isso ela tem que ser pertinente, ter fundamentação teórica e credibilidade (Tanaka e Tamaki, 2012; Contandriopoulos, 2006).

Os resultados da avaliação devem sedimentar conhecimentos quanto ao julgamento de valor ou mérito e servir para descortinar caminhos para a ação e para a transformação da realidade avaliada. Por isso, é preciso “que os seus resultados permitam aos diferentes atores envolvidos, que podem ter campos de julgamento diferentes, se posicionarem e construir (individual ou coletivamente) um julgamento capaz de se traduzir em ação” (Contandriopoulos, 2006).

No âmbito do sistema de saúde brasileiro, pode-se articular que o campo da avaliação ainda representa um grande desafio para a gestão da saúde, visto que há escassez de publicações nesta área. Percebe-se que a investigação avaliativa ainda é incipiente, em especial, quanto a avaliação e aos processos de monitoramento de resultados através de instrumentos que possam subsidiar a tomada de decisão, campo no qual parece haver um fosso entre a necessidade e o volume de estudos de aplicação da avaliação, conforme pesquisa sistemática (quadro 27, anexo 4).

4.2. A teoria organizacional subjacente ao envolvimento dos agentes de decisão (*stakeholders*) na avaliação participativa

No enquadramento dos estudos de governança organizacional, deve-se selecionar uma teoria de embasamento. Entre as diversas teorias de governança pode-se destacar: a *teoria neoclássica*; a *teoria da agência*; a *teoria dos custos de transações*, a *teoria do direito a propriedade*, dentre outras.

Há diversas teorias no contexto de governança organizacional descritas por Alves (2012), mas que fogem do escopo deste trabalho, no entanto, há outras teorias que se destacam no âmbito da gestão pública, que não se distanciam desta investigação, por exemplo: a *teoria da escolha pública (public choice)* e a *teoria da provedoria (stewardship)* (Alves, 2012).

Todas estas teorias apresentam vantagens e desvantagens sintetizadas por Alves (2012) que destaca a *teoria dos stakeholders* com diversas vantagens para a gestão dos sistemas produtivos de bens e serviços públicos. Nesta teoria a visão é ampliada e prioriza a integração dos objetivos de todos os *stakeholders* (partes interessadas) da organização que possuem algum interesse, ou direito (moral ou legal), ou propriedade, ou seja, todos aqueles que têm direitos legítimos de ser beneficiários da performance e dos resultados da organização.

Ressalte-se que pesquisadores da *teoria dos stakeholders* sustentam que o conhecimento acerca da natureza dos *stakeholders* organizacionais constitui informação relevante para a previsão do comportamento organizacional (Mellahi & Wood, 2003; Déniz-Déniz e Zárraga-Oberty, 2004).

Segundo Freeman (1984 apud Carvalho, 2007: 60), um *stakeholder* é qualquer indivíduo ou grupo de agentes que pode afetar ou ser afetado pelo alcance dos objetivos de uma organização.

Do ponto de vista teórico há pelo menos quatro principais perspectivas com diferentes abordagens da ação dos *stakeholders* nas organizações: a visão *instrumental*, *descritiva*, *normativa* e a de *responsabilidade social*, conforme Carvalho (2007).

A perspectiva *instrumental* postula que os resultados previsíveis são contingentes a um determinado tipo de comportamento (Caldeira & Faia-Correia, 2002).

[...] A perspectiva *descritiva* focaliza-se na caracterização das interações que as organizações estabelecem com os diferentes grupos de interesse (Mellahi & Wood, 2003).

[...] A perspectiva *normativa* envolve a especificação do que as organizações devem ou não fazer, de um ponto de vista ético e moral, relativamente ao tratamento dos interesses das várias partes envolvidas (dos *shareholders* e dos *stakeholders*) (Kujala, 2001).

[...] Na perspectiva de *responsabilidade social* a prioridade é proteger os diversos interesses presentes daqueles (*stakeholders*) que têm de fazer com que a organização funcione (Caldeira & Faia-Correia, 2002).

Carvalho (2007), por sua vez, destaca o papel dos gestores para a gestão da performance organizacional ao afirmar que

Os gestores emergem, eles próprios, como um *stakeholder* com responsabilidades ao nível da tomada de decisão e da negociação, cuja importância e contributo para uma gestão eficaz do comportamento e dos resultados das organizações são incontestáveis.

Assim, parece lógico, diante dos objetivos desta investigação, embasar-se na *teoria dos stakeholders*, em especial, nas suas perspectivas *instrumental* e *descritiva* destacadas pela literatura (Bryson, 2004; S. Hart & Sharma, 2004 apud Carvalho, 2007: 54) nomeadamente quanto aos grupos de *stakeholders* exercerem um impacto considerável sobre os resultados organizacionais e sobre o seu desempenho e competitividade. Assim, mantém-se alinhado no foco desta investigação que objetiva mapear a visão das partes interessadas (*stakeholders*) quanto aos desempenhos e resultados das organizações e serviços de saúde.

Dessa forma, para o escopo deste trabalho, adotou-se a perspectiva da *teoria dos stakeholders* com duas diferentes aplicações. Envolveu-se como decisores chaves (*key stakeholders*) gestores de três níveis setoriais, organizacionais e governacionais, no processo de estruturação e validação das dimensões e indicadores para mensuração da performance organizacional, além de se abordar especialistas, profissionais e usuários-utilizadores enquanto principais partes interessadas (*key stakeholders*) na análise e validação de constructos, dimensões e questões do instrumento de avaliação de qualidade percebida.

Estabelecidos os intervenientes (partes interessadas) e o enquadramento Conceitual de avaliação de performance, então a identificação específica de decisores chaves (*key stakeholders*) e dos usuários-utilizadores, além dos propósitos dos decisores, pode deixar mais claro aspectos como situações de decisão, aspectos de coordenação dos intervenientes, resolução de divergências entre eles, indicação de objetivos e captação de preferências.

4.3. Estratégias mistas (qualitativa e quantitativa) de coleta de informações e construção de consensos entre decisores (*stakeholders*) em organizações de saúde

Este trabalho, inicialmente, mapeou as preferências dos decisores quanto aos diversos aspectos da avaliação e monitoramento de performance em organização de saúde, para Em

seguida, se apropriar destas preferências e estruturar um modelo teórico-lógico de avaliação da performance das organizações de saúde. Neste propósito inicial, fez-se um trabalho de campo para o envolvimento e captação das preferências dos decisores, que passa por uma metodologia de matiz qualitativa.

A investigação qualitativa, segundo Boggan e Biklen (1994), surgiu no final do século XIX e início do século XX, atingindo o seu apogeu nas décadas de 1960 e 1970 por via de novos estudos e sua divulgação.

Utiliza-se uma metodologia qualitativa quando se procura compreender e descrever fenómenos na sua totalidade e a partir do contexto em que ocorrem, através da recolha e analisa em profundidade dados de difícil mensuração (sentimentos, sensações, percepções, pensamentos, intenções, comportamentos passados, entendimento de razões, significados e motivações) de um determinado grupo de indivíduos em relação ao fenómeno em estudo. Entre eles estão sentimentos, sensações e motivações que podem explicar determinados comportamentos, apreendidos com o foco no significado que adquirem para os indivíduos. É um método indutivo que procura entender por que o indivíduo age como age ou pensa como pensa. (Denzin & Lincoln, 2006:105).

Segundo Denzin e Lincoln (2006) a investigação qualitativa pode ser conduzida em diferentes estratégias ou planos de investigação, como por exemplo: estudo de projeto; estudo de caso; etnografia; observação participante; fenomenologia; etno metodologia; teoria ancorada (*grounded theory*); método biográfico; método histórico; investigação aplicada e ação (pesquisa-ação); investigação clínica; análise assistida por computador; e análise textual.

Por sua vez, Savin-Baden & Major (2013) identificam e descrevem oito tipos de planos de investigação ou estratégias de natureza qualitativa (*Estudo de caso; Teoria ancorada ou fundamentada - grounded theory-; Etnografia ; Fenomenologia; Investigação narrativa; Investigação-ação; Avaliação de programa ou objeto; e Investigação artística*), pelo que se destaca a *Teoria ancorada ou fundamentada (grounded theory)*, que visa desenvolver teoria a partir da análise sistemática dos dados recolhidos em entrevistas conforme as percepções dos fatos, eventos, atos e experiências dos participantes, sendo adequada para o envolvimento e captação das preferências dos decisores desta investigação.

De acordo com Pardal e Lopes (2011) o método de investigação social ajusta-se ao objeto de estudo e possibilita a progressão do conhecimento sobre esse objeto.

Strauss & Corbin (1990) desenvolveram a estratégia ou plano de investigação de pesquisa qualitativa, a *Grounded Theory* (Teoria Ancorada ou Teoria Fundamentada em Dados) que pode ser considerada como uma forma de analisar dados etnográficos de matiz qualitativa.

Trata-se de uma estratégia ou plano de investigação de pesquisa qualitativa sistematizada que se apoia no construtivismo social que privilegia o local (estudo a realizar) em detrimento do geral (o universal) (Strauss & Corbin, 1990). Consiste na descoberta e no desenvolvimento de uma teoria a partir de informações obtidas e analisadas sistematicamente e comparativamente em entrevistas que buscam conhecer as percepções de indivíduos sobre suas experiências (fatos, atos, pessoas, etc)

A metodologia da *Grounded Theory* (Teoria Ancorada), selecionada para esta investigação, consiste na construção da teoria a partir de dados baseados na realidade obtidos pela captação das percepções dos participantes de entrevistas. O pesquisador forma uma versão teórica da realidade e esta formulação teórica ou teoria substantiva não é apenas usada para explicar a realidade, mas também permite obter um esquema de referência para a ação.

As questões de projetos de pesquisas que demandam uma abordagem *Grounded Theory* (Teoria Ancorada) são geralmente dos seguintes tipos: Como é o processo, conceitos e vivências experienciadas por gestores no processo de decisão em organizações de saúde? Como têm sido as experiências, na perspectiva dos gestores de topo, vivenciando o processo de governança em organização de saúde?

Definidos os critérios de inclusão e exclusão de indivíduos no estudo, seleciona-se a população a ser considerada no estudo. A inclusão das pessoas nas amostras deve considerar as similaridades das experiências na primeira fase de entrevistas, e em seguida para entrevistas de segunda fase deve-se selecionar indivíduos conforme os conceitos surgidos na primeira fase que estejam relacionados aos aspectos do objeto de pesquisa, sendo isto parte do método *Grounded Theory* (Strauss & Corbin, 1990). O tamanho da amostra é determinado pela saturação teórica, ou seja, nos grupos selecionados encerra-se o processo de amostragem quando não mais surgir novos conceitos e temas nas entrevistas.

A *Grounded Theory* (Teoria Ancorada) é um método no qual primeiramente os dados coletados são recortados de acordo com temas, expressões ou partes do discurso. Em seguida, faz-se uma categorização dos conteúdos e, por fim, procede-se uma análise de acordo com a percepção, sensibilidade e conhecimento do assunto do pesquisador.

Enquanto isso, Denzin e Lincoln (2006) descrevem os seguintes métodos de recolha de informação: entrevista, observação, Artefatos, documentos e gravações, métodos visuais, métodos de experiências pessoais, métodos de gestão de dados, dentre outros, por sua vez, Pardal e Lopes (2011) identificam sete (7) técnicas de recolha de dados para investigação: observação, questionário, entrevista (estruturada, não-estruturada, semiestruturada), escalas de atitudes e opiniões, análise de conteúdo, análise documental e, semântica diferencial.

Para o propósito desta investigação e conforme o objeto (estruturação do modelo teórico-lógico de avaliação de performance em saúde) a investigar na etapa de captação da percepção dos *stakeholders* quanto as dimensões e indicadores relevantes de performance em organizações de saúde, preferiu-se aplicar uma análise qualitativa pautada na entrevista semiestruturada com fases sucessivas de adequação incremental por grupos pilotos, conforme as percepções dos fatos, eventos, atos e experiências dos participantes, que é a base da Teoria Ancorada. Neste caso, através de uma análise descritiva qualitativa comparativa, a análise de conteúdo - especificamente em seu Aspecto instrumental de categorização dos conteúdos relevantes para a posterior sistematização das relevâncias atribuídas às categorias de análises que surgiram das entrevistas qualitativas.

Assim, a fim de efetivar a captação das percepções dos agentes de decisão quanto as bases da estruturação do modelo teórico-lógico de avaliação de performance em saúde deste estudo, analisou-se um rol de possibilidades da literatura e aplicou-se cinco estratégias: a entrevista aberta para a captação qualitativa dos conhecimentos prévios em painéis de especialistas; a entrevista semiestruturada ou estruturada; os grupos focais ou grupos de especialistas; a técnica de grupo nominal (TGN) e a técnica *Delphi*, caracterizadas a seguir:

4.3.1. A entrevista semiestruturada ou estruturada

Com intuito de caracterizar as entrevistas realizadas como semiestruturadas, as mesmas podem se basear em questionamentos, perguntas norteadoras, e reflexões em teorias formuladas antes e durante o processo de investigação. Afirma Minayo et al. (2010) que é importante a elaboração de um roteiro na entrevista semiestruturada, o qual deve desdobrar os indicadores considerados essenciais e suficientes em tópicos que contemplem a abrangência das informações esperadas.

Na fase final de aplicação as entrevistas estruturadas foram elaboradas mediante questionário totalmente estruturado, sendo as perguntas previamente formuladas, com cautela de não fugir a elas. O principal motivo deste cuidado é a possibilidade de se comparar o mesmo conjunto de perguntas e que as diferenças devem refletir diferenças entre os respondentes e não diferença nas perguntas (Lakatos & Marconi, 2007).

A utilização de questionário possibilita algumas vantagens, como: nem sempre é necessária a presença do pesquisador para que o informante responda as questões. Além disso, o instrumento a ser aplicado consegue atingir várias pessoas ao mesmo tempo obtendo um grande número de dados, podendo abranger uma área geográfica mais ampla se for o Objetivo

da pesquisa. Garantindo ainda uma maior liberdade das respostas em razão do anonimato, evitando vieses do entrevistador. Geralmente, através do questionário, obtêm-se respostas rápidas e precisas.

Na presente investigação inicialmente construiu-se modelos semiestruturados desenvolvidos a partir da análise da literatura e ponderação de ideias, conceitos e constructos norteadores formulados antes e durante o processo de investigação, devidamente registrados nos apêndices 4 a 8, em fases sucessivas de adequação por grupos de especialistas e informantes selecionados. Cada entrevistado tinha a possibilidade de ponderar os itens propostos no questionário semiestruturado consolidado pelo grupo na fase anterior, além da condição de propor a agregação de novos itens. No caso de agregação de itens relevantes por algum dos entrevistados outra versão do instrumento semiestruturado foi proposto e aplicado nas fases posteriores.

4.3.2. Os grupos focais ou grupos de especialistas

Os grupos focais são grupos de discussão que dialogam acerca de um tema em particular, ao receberem estímulos apropriados para o debate, permitindo, pois, que o processo de interação grupal se desenvolva, o que favorece trocas, descobertas, e a formação de ideias novas e originais (Ressel et al., 2008).

Para a realização dos grupos focais, foram previstos os seguintes momentos, conforme estabelece Dall’Agnol e Trench (1999) abertura da sessão; 2) apresentação dos participantes entre si; 3) esclarecimento sobre a dinâmica de discussões; 4) estabelecimento do setting (contrato ético da pesquisa); 5) debate; 6) síntese; 7) encerramento da sessão.

Cabe ainda destacar que cada grupo possuía uma equipe composta por um coordenador e/ou observador. Consoante Dall’Agnol e Trench (1999), aqueles desempenham papéis específicos na condução de uma dinâmica grupal, devendo pensar juntos e intercambiar impressões sobre o desenvolvimento grupal e os fatos mais relevantes de cada encontro grupal, realizam, assim, momentos distintos de trabalho como o pré-grupo, o grupo propriamente dito e o pós-grupo.

4.3.3. A técnica de grupo nominal (TGN)

A TGN trata-se de um instrumento de pesquisa qualitativa grupal, também chamada de Painel de Especialistas (Furtado e Onocko-Campos, 2005; Onocko-Campos e Furtado, 2011),

que reúne grupos de indivíduos, sem priorizar o debate entre eles. Visa aumentar a produtividade do grupo, facilitar a decisão grupal, estimular a geração de ideias críticas e servir como instrumento de agrupamento dessas ideias (Cassiani & Rodrigues, 1996). A TGN permite reunir especialistas em um assunto para selecionar ideias, fazer julgamentos de opções e fomentar a criatividade de sugestões para a resolução de problemas e/ou priorizar alternativas. Além disto permite a geração de ideias coletivas pelo incentivo à participação, interação entre os diversos saberes do grupo de especialistas, em busca da geração de consensos.

Na 1ª. etapa de aplicação do método TGN pode-se reunir um grupo de 9 a 12 especialistas em determinado assunto, explicando de forma global o assunto que se pretende pesquisar e obter a opinião do grupo de especialistas. Após uma análise genérica do assunto, o grupo deve contribuir de forma individualizada com uma ideia (que julgue ser mais importante ou relevante). O facilitador que está à frente do grupo registra todas as sugestões que parecem ser semelhantes, de forma agrupada, em lugar de fácil visualização para o grupo (por exemplo, lousa ou *datashow*). Se necessário, ele conduz um debate grupal, somente para esclarecer o significado de cada ideia e os agrupamentos sugeridos pelo grupo, sem nenhum juízo de valor inicial.

Na etapa seguinte, executa-se a 1ª. rodada de análise em grupo. Solicita-se a todos os indivíduos do grupo que emitam pesos (importância ou relevância) para cada uma das ideias apresentadas e registrada de forma agrupada, que são ponderadas isoladamente por cada participante do grupo, para posterior tabulação e apresentação ao grupo na forma de uma classificação (*ranking*) destas ideias.

Na etapa final, executa-se a 2ª. rodada de análise em grupo, apresenta-se a classificação geral (*ranking global*) das ideias que deve ser discutida pelo grupo genericamente, e em seguida, geradas novas ponderações individuais. Uma nova classificação geral (*ranking global*) definitiva é formada pela tabulação das ideias que é apresentado ao grupo como consenso grupal (Jones & Hunter, 1995).

4.3.4. A técnica *Delphi*

Utiliza-se a técnica *Delphi* para a construção de consenso, no âmbito da pesquisa qualitativa grupal. A técnica *Delphi* tem como Objetivo a busca de opiniões coletivas qualificadas (R. T. O. Campos & Furtado, 2011), consensos de opiniões de um grupo de especialistas a respeito de eventos futuros, sendo os consensos entendidos como a consolidação do julgamento intuitivo do grupo de especialistas (Wright & Giovinazzo, 2000).

A operação da Técnica Delphi inicia-se pela construção de um questionário (para cada uma das questões deve existir uma síntese bem elaborada das principais informações sobre o assunto) que circula repetidas vezes por um grupo de especialistas no assunto em análise, escolhidos previamente e, sem que haja encontros pessoais entre os mesmos. Neste primeiro questionário as respostas requeridas são quantitativas (notas, conceitos, ponderações), com espaços para possíveis justificativas (1ª. rodada).

Após a coleta dos dados estrutura-se um segundo questionário tabulando as respostas com os resultados obtidos (estatística simples e intuitiva), associando todas as argumentações de cada questão da primeira rodada, mas mantendo o anonimato dos respondentes. Na 2ª. rodada os participantes são motivados a analisar os dados tabulados e justificativas, construindo novas respostas quantitativas e justificativas conforme os resultados da rodada anterior. Este processo é repetido até que as respostas obtidas em uma rodada possam ser consideradas como o consenso do grupo (Wright & Giovinazzo, 2000).

Há relatos de uso da técnica *Delphi* pela internet, em todas as etapas, mantendo o mesmo percurso metodológico, com resultados positivos (Giovinazzo & Fischmann, 2001). Atualmente o método tem sido ampliado, inclusive para estratégias e métodos para a tomada e decisão. As bases do método são: o anonimato dos respondentes, a apresentação dos dados tabelas e de estatísticas simples, e o *feedback* das respostas para o grupo executar a reavaliação em rodadas subsequentes e, assim rever suas ponderações às questões e emitir novo posicionamento (notas, conceitos, ponderações).

Na presente pesquisa, será aplicada a técnica *Delphi* para a construção de consenso de opiniões coletivas qualificadas, na fase de validação de questionários de itens de avaliação de performance e dos itens da escala de qualidade percebida, junto à gestores, profissionais e decisores do SUS através de grupos focais sucessivos.

Após a aplicação destas técnicas, na fase final de validação dos painéis de indicadores, usou-se da ferramenta pública e gratuita utilizada no âmbito do SUS (FormSUS), enquanto plataforma de internet para a coleta das opiniões (percepções) dos decisores da amostra selecionada.

4.4. Síntese teórico-conceitual da investigação

Antes de avançar para descrever a metodologia utilizada nesta investigação, sintetizou-se a teoria, no quadro 5, onde se identificam três grandes dimensões do arcabouço teórico-conceitual (*framework*) do estudo: Macro e Inter-Organizacional, Intra-organizacional e

Individual, subdivididas em dimensões de análise, sendo a base teórico-conceitual desta investigação, apresentada nos tópicos 2, 3 e 4 (teoria da avaliação, bases da medição de performance e fundamentos e estratégias de envolvimento dos decisores).

Nível	Dimensões de Análise	Autores
MACRO E INTER-ORGANIZACIONAIS	Ambientes geral e setorial: marco regulatório; lei orgânica do SUS; ambiente político; princípios doutrinários e organizativos do SUS; arranjo organizacional do SNS-SUS; elevação de gastos da saúde.	(Brasil, 1986b)(Brasil, 1986a, 1988); (Brasil, 1988);(Brasil, 1990b); Mendes (2011); Santos-Filho (2007);(Brasil, 1990c); (Brasil, 2011c); (Brasil, 2015a); (OECD, 2013); (OECD, 2015a); (Schramm et al., 2004); (Frasão, 2017).
	Ambiente institucional: três níveis de governança organizacional federal, estadual e municipal; comando único em cada nível de gestão; participação efetiva da sociedade na governança da saúde; financiamento tripartite (união, estados e municípios) definido em leis federais; arranjo organizacional com inter-relações do subsistema público (SUS) e do subsistema suplementar e privado (operadoras de saúde) do Brasil.	(Brasil, 2011c); (Brasil, 2011c, 2011f); (Brasil, 2015c); (Brasil, 2015a).
INTRA-ORGANIZACIONAIS	Fatores contingenciais: PlanejaSUS, Plano de Saúde (PS), Relatórios Anuais de Gestão (RAG); instrumentos de controle e avaliação; sistema nacional de informação (SIS) do SUS; sete modelos de gestão de unidades de saúde e hospitais distribuídos em administração pública direta e indireta (autarquias).	(Brasil, 2006a); (Brasil, 2006b); (Brasil, 2006c, 2013c); (Brasil, 2013c); (Brasil, 2015d);(Brasil, 2011a); (Brasil, 2012a);(Brasil, 2015e); (Brasil, 2015f);

Nível	Dimensões de Análise	Autores
	<p>Orientação estratégica para inovação: Modelo seminiais de avaliação em saúde: relatório Flexner (1910), relatório Codman (1916) e Donabedian: estrutura, processo e resultados; origens do campo da avaliação: a prestação de contas (accountability) e o controle dos programas; estratégia de gestão baseada em resultados - results-based management (RBM); gestão de resultados focada nos usuários-utilizadores; mudança do foco das atividades ou “produtos” (outputs) para os resultados (outcomes).</p>	<p>(Barreto, 2016); Alkin (2004 apud Samico et al., 2010); (UNDP, 2009) Avendis Donabedian (1980a); Flexner (1910); Dos Reis et al. (1990); (Alves, 2012: 2, 3); (Porterfield, 1976); Perez Arias (1983, apud Dos Reis et al., 1990: 51); Donabedian (1980a); Silva formigli (1994 apud Matus, 1997); Fletcher et al. (1993); (Minayo et al., 2010); (Lakatos & Marconi, 2007); (Ressel et al., 2008); Dall’Agnol e Trench (1999); (Furtado e Onocko-Campos, 2005; Onocko-Campos e Furtado, 2011); (Cassiani & Rodrigues, 1996); (Jones & Hunter, 1995); (R. T. O. Campos & Furtado, 2011); (Wright & Giovinazzo, 2000); (Giovinazzo & Fischmann, 2001); Gil (2010); (Austen, 2012); Samico et al. (2010); (Worthen et al., 2004); Hartz e Vieira-da-Silva (2005); Patton (1997); Morra-Imas e Rist (2009); (Kettl, 1996; Almeida, 1999 apud F Viacava et al., 2004: 712).</p>
	<p>Configuração estrutural: modelo teórico-lógico de avaliação; multidimensões e modelos de avaliação; crescente interesse no campo de avaliação de sistemas de saúde; predomínio de estudos teóricos-conceituais; delimitação do campo da avaliação de performance; definição do constructo eficácia organizacional; eficácia enquanto capacidade de gerar o efeito esperado ou atingir o Objetivo; eficiência enquanto otimização meios e recursos da organização em termos de produção de resultados; efetividade enquanto qualidade-satisfação percebida; abordagens</p>	<p>Hartz e Vieira-da-Silva (2005); Bunge (1974); (Medina et al., 2005); (Mclaughlin & Jordan, 1999; Rowan, 2000). Chen (1990); Samico et al. (2010); Gattinara et al. (1995); Vuori (1991; também citado por Acurcio et al., 1991; Akerman & Nadanovsky, 1992; Santos, 1995); Brook & Lohr (1985 apud Dos Reis et al., 1990: 54-55); (Barzelay, 1997; Boland & Fowler, 2000; Stevens et al., 2006); Chianca e Youker (2004); Bosi e Mercado (2006); (Fals Borda & Mora-Osejo, 2004) Carvalho (2007); Uchimura e Bosi (2002); (Brasil, 2003a); (Alkin & Christie, 2004 apud R. T. O. Campos & Furtado, 2011); (OECD, 2002; World Bank, 2009); (WHO, 2000b); (Braveman et al., 2001; V. Navarro, 2000; Williams, 2001); (Almeida et al, 2001); Carvalho e Gomes (2000); Carvalho (2007); (Rerimann, 1982; Brunet et al., 1991 Morin et al., 1994; Rogers e Wright, 1998; Savoie e Morin, 2001); Navarro e Quijano (2006); Etzioni (1975); Brunet (1991 apud Carvalho, 2007: 151); Katz e Khan (1977 apud Carvalho, 2007: 149); (Morin et al., 1994); Santos (1995); (Navarro & Quijano, 2006 apud Carvalho, 2007: 166); (K. Cameron, 1981 apud Carvalho 2007: 169); (Cameron 1981 apud Carvalho, 2007: 170); Glunk e Wilderom (1999 apud Carvalho, 2007: 153); (Buono, 2005; Conlon et al., 2004; Driscoll & Starik, 2004; Lawton et al., 2000; Trieschemann et al., 2000); Hirsch e Lewin (1999); Glunk e Wilderom (1999); Ramalho</p>

Nível	Dimensões de Análise	Autores
	unidimensionais da eficácia; racional dos objetivos ou modelos de objetivos, sistêmica e de aquisição de recursos e a de processos internos; constructo multidimensional organizacional performance ou desempenho organizacional;	(2005); Glunk e Wilderom (1999 apud Carvalho, 2007: 152-153); Glunk e Wilderom (1999 apud Carvalho, 2007: 153); (Brunet et al., 1991 apud Carvalho, 2007: 164); Cameron (1980 apud Carvalho, 2007: 165); (J. Navarro & Quijano, 2006 apud Carvalho, 2007: 166); (K. Cameron, 1981 apud Carvalho, 2007: 169); (K. Cameron, 1981 apud Carvalho, 2007: 170); (Carvalho, 2007); (Suchman, 1995 apud Carvalho, 2007: 172); (Mitchell et al., 1997 apud Carvalho, 2007: 172); Navarro e Quijano (2006); (J. Navarro & Quijano, 2006 apud Carvalho, 2007); (K. Cameron, 1981 apud Carvalho, 2007); Kaplan e Norton (1992); Savoie e Beaudin (1994); Morin et al. (1994 apud Carvalho, 2007: 180).
	Enquadramento teórico-conceitual (<i>framework</i>) (avaliação de performance): o escopo do campo de avaliação em saúde; avaliação da eficiência em organizações de saúde; ajustamento pelo risco e índices de casemix ou Diagnosis Related Groups (DRG); modelos estruturados de avaliação de performance dos serviços de saúde; World Health Report 2000; projetos mundiais QIP e AHRQ (EUA), EFQM (Europa) e PATH (OMS).	(Hutubessy, 2003: 24); Farrel (1957); Banker, Charnes e Cooper (1984); Coelli, Rao e Battese (1999 apud La Forgia & Couttolenc, 2009: 128); (Cooper, Seiford, & Tone, 2002 apud La Forgia & Couttolenc, 2009: 128); Amado e Dyson (2008); Amado e Santos (2009: 44); Podinovski (2004); La Forgia e Couttolenc (2009); (Brasil, 2013a); Shaughnessy e Hittle (2002); Iezzoni (2003); Bonilla e Rubio (2000); (IHPS, 2005); Iezzoni (1997 apud Costa & Lopes, 2011: 9); (Fetter et al., 1980); (Lichtig, 1986); Sheldon (1998 apud Da Silva et al., 2008: 50); (Campbell et al., 2000); (Campbell et al., 2000: 1621-22); (McGlynn, 1997; Stott et al., 1997); (Arah et al., 2006); (Donabedian, 1980c; H. V. Vuori, 1982); (Donabedian, 1980c) Silva & Formigli (1994); (Freeborn & Greenlick, 1973); (F. E. Campos, 1988); OMS (1970); (Ferrara et al., 1976); (Lane & Kelman, 1975); (Rodrigues & Torres, 1982; Rundall, 1992); (Doll, 1974); (OTA, 1978; Vuori, 1982; Cavallo et al., 1988; Donabedian 1990); (OTA, 1978; Chavez, 1981; Vuori 1982; Brook & Lohp 1985; Donabedian, 1990); (De Geyndt, 1970); (Rundall, 1992; Silva & Formigli, 1994: 82); Kazandjian (2003); (Donabedian, 1980b, 1988); (Kettl, 1996; Almeida, 1999 apud F Viacava et al., 2004: 712); (FIOCRUZ, 2012; Veillard et al., 2005; F Viacava et al., 2004); (WHO, 2000a); (McKee & Healy, 2002); (Veillard et al., 2005); (Veillard, 2012); Escorel et al. (2007); (Barzelay, 1997; Boland & Fowler, 2000); (Stevens et al., 2006); Boland e Fowler (2000); Sheldon (1998 apud Da Silva et al., 2008: 50); (Brook & McGlynn, 1991); (OTA, 1988); Murray & Frenk (2000b); (Aday, 1993; AHCPR, 1999; Hoffmeyer & Kluwer, 1994; C. W. Hsiao, 1995; Knowles et al.; OECD; OECD; Organization, 1999); (Frenk; Kutzin; Londono & Frenk, 1997; Murray & Frenk, 2000b: 12); (Monteiro,

Nível	Dimensões de Análise	Autores
		<p>2008; La Forgia e Couttolenc, 2009) Kazandjian (1999; 2003); (EFQM, 1999); (Donabedian, 1980d), (Walburg et al., 2000: 192-93); (W. L. Hart & Bogan, 1992; "Malcolm Balridge National Quality Award, 1999 Criteria for Performance Excellence," 1999); (Gennip, 1998); (J. Commission, 1998); (Moeller, 2001: 45-46), Valejo et. al. (2006: 327-28); (WHO, 2000b); (Arcelay et al., 1999; Holland & Fennell, 2000; Jackson, 1999; Klazinga, 2000; Mira et al., 1998; Moeller, 2001; Simón et al., 2001); (WHO, 2003); (Braveman et al., 2001; V. Navarro, 2000; Williams, 2001); (Almeida et al, 2001); Hurst e Hughes (2001 apud F Viacava et al., 2004: 715-716); (Cozzens, 1995; Hurst & Hughes, 2001; AIHW, 2000; NHPC, 2002; CIHI, 2001; Hurst, 2002; OR, 2002; Smee, 2002; Wolfson e Alvarez, 2002); (Chang et al., 2002); (NHS, 2002); (Fund, 1998); (McLoughlin et al., 2001); (Campbell et al., 2000); (A. Commission, 2000); (Health, 2001, 2002); (CIHI, 1999, 2000); (Committee, 2001); (Medicine, 2001); (Services, 2000); (Medicine, 2001); (Reilly et al., 2002); (Murray & Frenk, 2000a); (OECD, 2001); Arah et. al. (2003); (Hurst, 2000, 2002; OECD, 2001); (Arah et al., 2006; Mattke, 2004; Mattke et al., 2006; Mattke et al., 2004); (Greenfield, Nicolucci, & Mattke; Greenfield et al., 2004; Hermann & Mattke, 2004b; Lambie & Mattke, 2004; Marshall, Leatherman, & Mattke; Marshall et al., 2004; Millar & Mattke, 2004); (Development, 2001; Unit, 2001); Vallejo et al. (2006); (Arcelay et al., 1999; Holland & Fennell, 2000; Jackson, 1999; Klazinga, 2000; Mira et al., 1998; Moeller, 2001; Simón et al., 2001); (Nabitz & Walburg, 2000); (Jackson & Bircher, 2002); (Nabitz & Klazinga, 1999); (Moeller et al., 2000; Simón et al., 2001; Stewart, 2003); (Arcelay et al., 1999; Moeller et al., 2000; Moeller & Sonntag, 1998); (Moeller, 2001; Walburg et al., 2000); (Geraedts et al., 2001; Jackson, 1999; Pitt, 1999; Stewart, 2003); (Mira et al., 1998); (Mira et al., 1998; Stewart, 2003); (Rodríguez & Ferrándiz-Santos, 2004); (Holland & Fennell, 2000; Lorenzo et al., 2001; Nabitz & Walburg, 2000; Vitoria-Gasteiz, 2000); (Brandt et al., 2005); ("Identification of a Set of Key Quality Indicators in the Hospital Setting Using the European Excellence Model ", 2001); (Veillard et al., 2005); (K. S. Cameron & Whetten, 1983; Sicotte et al., 1998); (V. Kazandjian, 2003; Roski & Gregory). (Ibrahim, 2001); Groene et al. (2008a); Groene et al. (2008b); Papanicolas & Smith (2010); Hsiao & Sidat (2008).</p>

Nível	Dimensões de Análise	Autores
	<p>Enquadramento teórico-conceitual (<i>framework</i>) (avaliação de qualidade): modelos estruturados de avaliação de qualidade e serviços de saúde; satisfação do usuário multifacetada, de aferição complexa por aspectos da intangibilidade; principais teorias da satisfação ou insatisfação de clientes; diagrama de preditores e antecedentes da satisfação; Modelo “gap model” de mensuração de satisfação; escala Service Quality – SERVQUAL; escala SERVPERF; escala SERVPERVAL; Instrumento de Avaliação de Qualidade Hospitalar (IAQH).</p>	<p>(M. Esperidião & Trad, 2005; M. A. Esperidião, 2009); (Aharony & Strasser, 1993 apud Esperidião e Trad, 2005: 304); Vinagre (2008); (Parasuraman et al., 1991b)(Grönroos, 1990); Grönroos (1978, 1982, 1988); Grönroos (2003); Oliver e DeSarbo (1988); (Szymanski & Henard, 2001); Oliver (1980); (Berry et al., 1994; Cronin & Taylor, 1992; Parasuraman et al., 1991a)(Chahal & Kumari, 2011, 2012; Cronin & Taylor, 1992; J. J. Ferreira et al., 2012; P. L. Ferreira, 1993; Pedro et al., 2010). Oliver (1980); Gummesson (1977); Surprenant e Solomon (1987); Bitner (1992 apud Vinagre 2008); Giese e Cote (2000 apud Vinagre, 2008); (Bitner et al., 1994); (p.ex. Churchill & Surprenant, 1982; Oliver, 1980; Oliver & Desarbo, 1988); Oliver (1997); (Oliver & Desarbo, 1988). (Bearden & Teel, 1983; Oliver & Linda, 1981; Swan & Trawick, 1981); (Oliver, 1981b; Oliver & Desarbo, 1988); (Oliver, 1980, 1981a, 1981b; Tse & Wilton, 1988; Yi, 1990); Szymanski e Henard (2001); Szymanski e Henard (2001, apud Vinagre, 2008: 26); (p.ex. Tse & Wilton, 1988; Zeithaml et al., 1993); (Kopalle & Lehmann, 2001; Mittal et al., 1998); Cronin e Taylor (1992); (Parasuraman et al., 1985a, 1991a); Parasuraman et al. (1985a); (Parasuraman et al., 1988, 1991b); (Parasuraman et al., 1988, 1991a); Lemke (2005); Atkinson (1988); Bentler e Chou (1987); Green et al. (1997); (Marôco, 2014); (E. Babakus & G. Mangold, 1992); (Chahal & Kumari, 2011, 2012); (P. L. Ferreira, 1993; P. L. Ferreira et al., 2001; P. L. Ferreira & Mendes, 1997; P. L. Ferreira & Raposo, 2006; P. L. Ferreira et al., 2017; Pedro et al., 2010); (Berry et al., 1994; Parasuraman et al., 1991a); (Dillon, 1997); (Fontnote et al., 2007); Hanson (1992).</p>
	<p>Interações internas e externas: envolvimento de decisores chaves (<i>key stakeholders</i>); avaliação participativa e construtivista – ontologia idealista; priorização dos objetivos e instrumentos desenvolvidos pelos seus futuros usuários-utilizadores; julgamento e atribuição de valor; utilização e aplicação dos resultados da avaliação para decisão; seleção de critérios, indicadores e</p>	<p>Guba e Lincoln (2011); (Yin, 2009 apud Minayo et al., 2010); Deslandes e Gomes (2004); (Becker, 1993; Minayo, 2006; Triviños, 1987; Yin, 2001); (Strauss & Corbin, 1990); Sampieri et al. (2006); Alves (2012); (Scriven, 1991; Rossi; Freeman, 2004; Patton, 1997); Patton (1997); (Paim & Teixeira, 2006); (Worthen et al., 2004 apud Campos e Furtado, 2011); (Tanaka & Melo, 2008); (WHO, 1981); (Ovretveit, 1998); Contrandriopoulos et al.(1997); (Champagne et al., 2009 apud Samico et al., 2010); Tanaka e Tamaki (2012b); Carvalho e Gomes (2000, 2002); Carvalho (2007); Carvalho (2007)(Lourença et al., 2004); Tanaka e Tamaki (2012a); (Contandriopoulos et al., 1997 apud Tanaka e Tamaki, 2012) Contandriopoulos (2006); (Green & Kreuter, 1999; Rootman et al., 2004 apud Campos e Furtado, 2011); Fernandes, et al.(2011);</p>

Nível	Dimensões de Análise	Autores
	padrões; avaliar os serviços de saúde para garantir a qualidade do atendimento à população; teoria dos <i>stakeholders</i> e a previsão do comportamento organizacional; perspectivas descritiva e instrumental da teoria dos <i>stakeholders</i> .	Campos e Furtado (2011); (Vieira-Da-Silva, 2005; Tanaka & Tamaki, 2012); Escorel et al.(2007); (1984 apud Carvalho, 2007: 60); Novaes (2000); (Mellahi & Wood, 2003; Déniz-Déniz & Zárrega-Oberty, 2004); Freeman (1984 apud Carvalho, 2007: 60); (Caldeira & Faia-Correia, 2002); (Kujala, 2001); Carvalho (2007); (Bryson, 2004; S. Hart & Sharma, 2004 apud Carvalho, 2007); Boggan e Biklen (1994); (Denzin & Lincoln, 2006); Savin-Baden & Major (2013); Pardal e Lopes (2011).
	Agentes e abordagens de mudança internas: modelos operacionais multidimensionais de eficácia organizacional: <i>Balanced Scorecard</i> (BSC); modelo multidimensional, subjetivo pelo juízo de valor e validação dos <i>stakeholders</i> ; modelo Interativo de Eficácia Organizacional etc;	Kaplan e Norton (1992), Morin et al.(1994 apud Carvalho, 2007: 180); Carvalho (2007); Morin, Savoie e Beaudin (1994).
INDIVIDUAIS	Iniciativas e aplicações no âmbito do SNS-Brasil: projetos brasileiros de avaliação de desempenho em saúde (PROADESS, PNASH, PNASS, PMAQ e as matrizes de indicadores IDB-Ripsa e IDSUS); projeto DRG-Brasil; pesquisa de satisfação de pacientes no Brasil; estudo da escala SERVQUAL em hospitais e unidades de saúde do Brasil; programa choque de gestão em MG-Brasil; evolução dos sistemas de informações em saúde do DATASUS; Índice de Desempenho da Saúde Suplementar (IDSS); programa QUALISS;	(F Viacava et al., 2004; Francisco Viacava et al., 2012); (AIHW, 2008; CIHI, 2000, 2001; Health, 2001; NHS, 1999; US/DHHS, 2000 apud Viacava et al., 2004: 116-153); (Brasil, 2003b); (Brasil, 2004; La Forgia & Couttolenc, 2009: 8; Monteiro, 2008); (Brasil, 2015e); (FIOCRUZ, 2011, 2012), (Brasil, 2011b); (Brasil, 2003a); Veras e Martins (1994 apud La Forgia & Couttolenc, 2009: 135); (Brasil, 2013a); (Brasil, 2017a, 2017d); (R. C Couto, 2014; Serufo-Filho, 2014); (CFM, 2014, 2015, 2018)(CNI, 2012, 2018); (Brasil, 2003a);(Brasil, 2009b); (Albuquerque, 2012; Rosalem, 2013); (Borges et al., 2006; Cruz & Melleiro, 2007; Hercos & Berezovsky, 2004); (Guimarães & Tavares-De-Almeida, 2006); Dudziak (2001); (Brasil, 2015f); (Pinto, 2010b); Veras e Martins (1994 apud La Forgia e Couttolenc, 2009: 135); Travassos et al (1999); Da Silva et al. (2008); (Brasil, 2015e); (Brasil, 2004; Monteiro, 2008); (FIOCRUZ, 2003); (Brasil, 2002a); (Brasil, 2010b); (F Viacava et al., 2004; Francisco Viacava et al., 2012); (Gouveia et al., 2009); (Brasil, 2005); Batista e Paiva (2014); (Brasil, 2010c); (Brasil, 2014); (Brasil, 2017c); Gastal et al. (2005 apud La Forgia e Couttolenc, 2009, p. 339); Boccolini et. al. (2016); Volpato (2014); (Albuquerque, 2012; Da Silva, 2014; Rosalem, 2013);

Nível	Dimensões de Análise	Autores
	Programa PM-QUALISS (ANS).	Albuquerque (2012); (Barbosa & Neto, 2017; Clavier et al., 2011; Hartz & John, 2008); (Barbosa & Neto, 2017; CIHR, 2004; WHO, 2006); (Vinagre, 2008); (M. Esperidião & Trad, 2005; M. A. Esperidião, 2009); (Brasil, 1994), (Brasil, 2004; 2009a); Marin (2010 apud Brasil, 2015f); (Brasil, 2009a, 2015f); (Brasil, 2009a); (Brasil, 1998);(Brasil, 2011d); (Brasil, 2015f); (Brasil, 2004, 2009a, 2015f); (Brasil, 2009a, 2015f; De Moraes, 2014; J. C. Noronha, 2001); (De Moraes, 2014); Noronha (2001); Escrivão (2004); Felix et al. (2012); Macente & Zandonade (2010); Pinto (2010a); Camacho e Leal (2006); (Brasil, 2009a: 59); (Brasil, 2009a: 61); Rocha (2017); (Brasil, 2009a, 2015b; De Moraes, 2014); (Monteiro, 2008; La Forgia e Couttolenc, 2009); (Brasil, 1996a, 1997a, 2000a); (Brasil, 2002b: 13-14); (Brasil, 2002a: 22-45; 2008a:23-53); (RIPSA, 2007); Lima (2016: 64-71); (J. G. Lima, 2016: 227-244); (Brasil, 2011e); (Brasil, 2013b); (Brasil, 2010b); (Brasil, 2016d, 2016e); (Brasil, 2016b, 2016c); Gonçalo (2010); (APM/CREMESP, 1998); (Malik & Schiesari, 1998); Rotta (2004:31-32); (Associação & Medicina, 1998); (FHEMIG, 2009; Ramos & De Oliveira, 2010); (Minas, 2010a); (Marques et al., 2010); (H. d. O. Lima et al., 2010);(Minas, 2012); (Minas, 2010a); (Junior, 2013).

Quadro 5 - Síntese integrativa das teorias da avaliação em saúde e bases conceptuais do estudo

PARTE II – EMPÍRICA

5. METODOLOGIA DO ESTUDO

Em linhas gerais, esta investigação estruturou um enquadramento conceitual (*framework*) alinhado com a percepção dos principais agentes de decisão (*key stakeholders*) voltado à gestão de hospitais e unidades de saúde do SUS. Estruturou questionários semiestruturados para capturar os pontos de vistas dos principais interessados e intervenientes na avaliação da performance do sistema de saúde. Procedeu a adequação e validação de um instrumento para a mensuração da qualidade percebida. Propôs métodos e instrumentos de análise e apoio à decisão disponíveis com adequação à realidade e utilidade percebida pelos decisores chaves e usuários, potenciais utilizadores e/ou clientes, do painel de indicadores e do instrumento de avaliação da qualidade percebida. Por fim, aplicou o processo de validação dos indicadores, provas de valor e usabilidade do *proxy* de painel de indicadores, a fim de envolver assegurar a usabilidade do instrumento final, pelos decisores chaves (*decision making*), além de ter aplicado a escala de qualidade/satisfação em uma amostra de usuários do SUS.

A natureza global da metodologia do presente estudo situou-se no campo da *Design Research* visto que se estudou um Artefato existente (a percepção sobre a gestão de organizações e serviços de saúde, estruturados e em funcionamento) e propôs-se uma solução de melhoria para o sistema através da concepção de uma metodologia de intervenção, enquanto instrumento gerencial da performance da organização ou serviço de saúde.

Os processos metodológicos deste estudo, por sua complexidade e multidimensionalidade, envolveram diferentes prismas. O estudo tratou-se de uma *pesquisa avaliativa*. Esse tipo de estudo corresponde ao julgamento das práticas sociais, principalmente das resultantes da ação social planejada, tais como políticas, programas e serviços de saúde (Hartz & Vieira-da-Silva, 2005).

A natureza da pesquisa caracterizou-se por uma metodologia avaliativa do tipo *observacional* (controle de possíveis influências intencionais no objeto pesquisado), *transversal*, com abordagem por método misto (qualitativo e quantitativo) de avaliação durante a etapa de validação metodológica e instrumental por decisores chaves da amostra (gestores, profissionais e usuários da saúde).

A parte *qualitativa e descritiva*, do tipo *exploratória* e *retrospectiva*, aplicou-se na revisão da literatura para caracterizar o problema com riquezas de detalhes, pelo estudo do estado da arte e do enquadramento metodológico da avaliação aplicada à tomada de decisão em saúde.

A investigação apoiou-se em *estudo de casos múltiplos* (Yin, 2009 apud Minayo et al., 2010). Saliente-se que para Deslandes e Gomes (2004), o estudo de caso é um dos desenhos mais utilizados na pesquisa desenvolvida na avaliação dos serviços de saúde, traduzindo de forma emblemática a lógica científica desta abordagem. Assim, em espaços e tempos determinados, os estudos de casos investigam um fenômeno contemporâneo dentro de seu contexto de vida real (Becker, 1993; Minayo, 2006; Triviños, 1987; Yin, 2001).

A metodologia aplicada ao propósito inicial foi uma *revisão integrativa bibliográfica* (anexos 1 e 2), através de uma pesquisa descritiva e exploratória em artigos e obras publicadas quanto ao percurso da estruturação da rede hospitalar do Brasil e, ainda ao quanto aos modelos de avaliação de destaque em sistema de saúde que tivesse similaridades de enquadramento teórico-conceitual com o SUS, além de ter investigado as metodologias e os sistemas de medição e monitoramento de resultados em hospitais como forma de subsidio às proposições iniciais para os painéis de especialistas quanto às metodologias de aferição de performance de organizações de saúde incorporadas neste trabalho, na medida do possível.

No segundo propósito, ainda por revisão sistemática integrativa da bibliografia, *descritiva e exploratória*, caracterizou-se de forma pragmática os sistemas de informações da saúde públicos e/ou publicizados disponíveis no Brasil, bem como, as experiências de construção de indicadores de organizações de saúde no contexto brasileiro, públicos e privados (anexos 5 e 6), além de quadros teórico-conceitual (*framework*) de outros países, em especial, quanto às suas possíveis limitações e potencialidades.

Na parte empírica, para a validação das metodologias e instrumentos propostos, utilizou-se uma metodologia avaliativa de painel de especialista (técnica *Delphi*), através da participação dos principais decisores da saúde, em especial, dos agentes das instâncias decisórias superiores de gestão estratégica e de departamentos de organizações e de serviços de saúde e, procedeu-se as provas do instrumento construído para a aferição de adequação do modelo proposto pela consultas sucessivas aos diferentes grupos de decisores chaves selecionados.

O tipo de amostragem adotado foi o de *acessibilidade*, que de acordo com Gil (2010), o pesquisador seleciona os elementos a que tem acesso, admitindo que estes possam, de alguma forma, representar o universo. Dadas as dificuldades operacionais do processo de amostragem probabilística para escolha indivíduos desta investigação, optou-se pelo levantamento (inquérito ou *survey*) de uma amostra não probabilística, por conveniência ou intencional, obtida pelo critério de acessibilidade. No caso da amostragem de usuários para aplicação de campo de instrumentos de mensuração da qualidade percebida, buscou-se uma amostra

equilibrada ao selecionar em torno de 25 a 30 usuários-utilizadores em cada unidade avaliada, porém, sem controle de outras variáveis como gênero, faixa etária, escolaridade etc.

Por fim, utilizou-se a metodologia quantitativa através de recursos da estatística descritiva, correlação quantílica e análise fatorial para mensurar os resultados obtidos e validar (ou refutar) as hipóteses iniciais propostas do estudo.

5.1. O processo de recolha de dados

A investigação apoiou-se em estudo de casos múltiplos selecionados por conveniência e disponibilidade de participação durante a aplicação dos instrumentos para a sua validação (ciclo 1 – agregação de todas as fases de uma das etapas do estudo voltadas para a estruturação e validação dos instrumentos) e para a pesquisa de dados de campo (ciclo 2 – agregação de todas as fases de uma das etapas do estudo voltadas para estudos de dados de campo), porém, com amostragem de usuários do tipo estratificado proporcional, na aplicação de campo da escala de qualidade percebida (ciclo 2) na mensuração da satisfação da rede de saúde.

Na etapa inicial da fase empírica da investigação, aplicou-se entrevistas qualitativas para a prospeção de informações quanto aos conhecimentos tácitos procedentes dos agentes chaves envolvidos nos painéis de especialistas e nos grupos pilotos iniciais para sistematizar estas informações em uma visão global por categorias analíticas de conhecimentos prévios (no estudo 1, macrodomínios e domínios de avaliação - apêndice 4 e, no estudo 4, constructos de qualidade – apêndice 7-A), através de instrumentos de entrevistas pré-estruturados.

Na exploração inicial de conhecimentos prévios dos agentes chaves, seguiu-se os princípios de Sampieri et al. (2006) que sugerem a amostragem por conveniência quando a investigação está centrada em indivíduos ou grupos portadores de um ou vários atributos que possam contribuir para o desenvolvimento de uma teoria. Citam a limitação de generalização de resultados, mas a boa captação e compreensão dos significados em contextos específicos. Para tal, propôs-se duas metodologias: uma descritiva qualitativa comparativa, a análise de conteúdo - especificamente em seu aspecto instrumental de categorização dos conteúdos relevantes pautada na teoria *Grounded Theory* (Strauss & Corbin, 1990), complementada pela análise de pesos atribuídos (apêndice 5 e 7-B, respectivamente) para a sistematização das importâncias/relevâncias atribuídas às categorias de análises que surgiram das entrevistas. Neste caso, o tamanho da amostra é determinado pela saturação teórica, ou seja, até que não haja surgimento de novos conceitos e temas nas entrevistas, sendo este o critério usado para

limitar os grupos pilotos em fases sucessivas de aprimoramento dos instrumentos de coletas de dados.

Os três estudos iniciais (1 a 3) aplicados foram do tipo *quantitativo, analítico, prospetivo e transversal*, nos quais aplicou-se instrumentos semiestruturados quanto a percepção dos agentes de decisão, sobre dimensões e indicadores de performance em saúde, através de ajustes e validações incrementais, com grupos pilotos de especialistas e profissionais (apêndices 5, 6-A e 6-B) e, por fim, com grupo decisores de saúde (apêndice 6-C) e, Em seguida, desenvolveu-se uma análise de estatística descritiva e/ou inferencial em buscar de confirmar (ou refutar) as hipóteses de diferentes painéis de dimensões e indicadores de performance.

Os dois estudos finais (4 e 5) aplicados foram do tipo *quantitativo, analítico, prospetivo e transversal*, nos quais aplicou-se instrumentos semiestruturados para a adaptação da escala de mensuração de qualidade percebida, através de ajustes e validações incrementais, com grupos pilotos de especialistas e de usuários (apêndices 7-B.1 e 7-B.2), além de grupos de consolidação de gestores e profissionais (apêndices 8-A e 8-B) e, por fim, aplicação de campo em usuários das unidades participantes (apêndices 8-C) e, Em seguida, desenvolveu-se uma análise de estatística descritiva e/ou inferencial, e caracterizar os níveis de satisfação dos usuários-utilizadores e/ou estabelecer possíveis correlações com a performance entre unidades.

O universo da pesquisa envolveu especialistas, profissionais, gestores e decisores da rede de saúde pública do Brasil e na aplicação da escala validada de qualidade percebida englobou adicionalmente usuários-utilizadores do SUS. A população da pesquisa contemplou um grupo de especialistas, profissionais, gestores e decisores da rede de serviços de atenção básica, secundária e terciária do SUS, que de forma voluntária aceitaram participar da pesquisa. Na aplicação de campo da escala de qualidade a população limitou-se à rede de saúde do Ceará.

Através de visita às organizações ou via correio eletrônico, a cada um dos hospitais e/ou serviços de saúde pesquisados foram entregues cartas de apresentação da pesquisa e solicitado um Termo de Anuência institucional. Em seguida, foi entregue um Termo de Consentimento Livre e Esclarecido (TLCE-decisores), a cada um dos gestores das instituições participantes, conforme as normas da resolução Conselho Nacional de Saúde (CNS)⁶⁵ n.º 466/12, sobre os princípios éticos apresentados, e cumpridos os requisitos de inscrição e controle ético da Plataforma_Brasil⁶⁶, do Conselho Nacional de Ética em Pesquisa do Brasil (Conep)⁶⁷, inclusive

65 CNS - Conselho Nacional de Saúde, órgão colegiado de decisão de governança(nança) sistêmica do SUS

66Plataforma_Brasil – Sistema Nacional on-Line de Controle de Investigação e Comitês de Ética do Brasil

67 CONEP - Conselho Nacional de Ética em Pesquisa do Brasil

quanto aos riscos potenciais⁶⁸ (constrangimentos, transtornos, danos, etc) e benefícios⁶⁹ esperados para os participantes, tendo sido facultado a desistência de participantes em qualquer fase da investigação, sem qualquer prejuízo ou dano. O projeto foi submetido à Comissão de Ética em Pesquisa (CEP)⁷⁰, tendo sido aprovado pelo nº CAAE: 54972816.9.0000.5051.

Foram critérios de seleção prévia para os gestores participantes: a assinatura do termo de anuência de participação na pesquisa pelo gestor municipal e/ou de um dos diretores da organização, a concordância com o TCLE-decisores, ter experiência na função de governança sistêmica, direção organizacional e/ou gerência departamental de hospitais e/ou serviços de saúde, ou ainda, ser profissional de saúde de nível médio ou superior de escolaridade.

Quanto aos gestores e profissionais informantes, dos grupos pilotos e de campo, foram incluídos aqueles que trabalham em unidades de saúde pública, tendo sido fatores de exclusão a não vinculação às unidades do SUS ou a recusa ao termo de participação (TLCE-decisores)⁷¹.

No caso da avaliação da percepção da qualidade e satisfação, foram critérios de seleção prévia para os usuários participantes: a assinatura do termo de anuência de participação na pesquisa pelo gestor municipal e/ou um dos diretores da organização, a concordância com o TCLE-usuários, e ter cursado, no mínimo, metade das séries do nível fundamental de escolaridade, para garantir a compreensão e respostas adequadas das questões do instrumento escala de qualidade percebida pelo cliente-utilizador entrevistado.

Na seleção de usuários-utilizadores foram incluídos aqueles abordados em atendimentos recentes nas unidades do SUS, sendo excluídos os menores de 18 anos e os que se recusaram o termo de participação (TCLE-usuários)⁷². Essas organizações e unidades de saúde foram caracterizados e selecionados conforme os dados disponíveis no Cadastro Nacional de Estabelecimentos de Saúde (CNES)⁷³.

Dessa forma, a visão global do fluxo do processo de estruturação e aplicação dos instrumentos e validação das matrizes avaliativas foi sintetizada na figura 12 abaixo.

68 Vide TCLE's (apêndices 2 e 3)

69 Como benefício direto para os participantes, nos comprometemos a fazer a devolutiva dos dados aos serviços participantes, quando assim solicitado, e disponibilização de um *proxy* de um Painel de Indicadores de eficiência, efetividade e qualidade (satisfação) a fim de monitorar os resultados de hospitais e serviços especializados do SUS, para subsidiar a gestão estratégica e o fomento de políticas de melhorias de qualidade e resolutividade destas organizações de saúde do Brasil.

70 CEP- Comissão de Ética em Pesquisa

71 TLCE - Termo de Consentimento Livre e Esclarecido – versão decisores (apêndice 2-A)

72 TLCE - Termo de Consentimento Livre e Esclarecido – versão usuários (apêndice 2-B)

73 CNES - Cadastro Nacional de Estabelecimentos de Saúde; Módulo do Sistema de Informação do Datasus

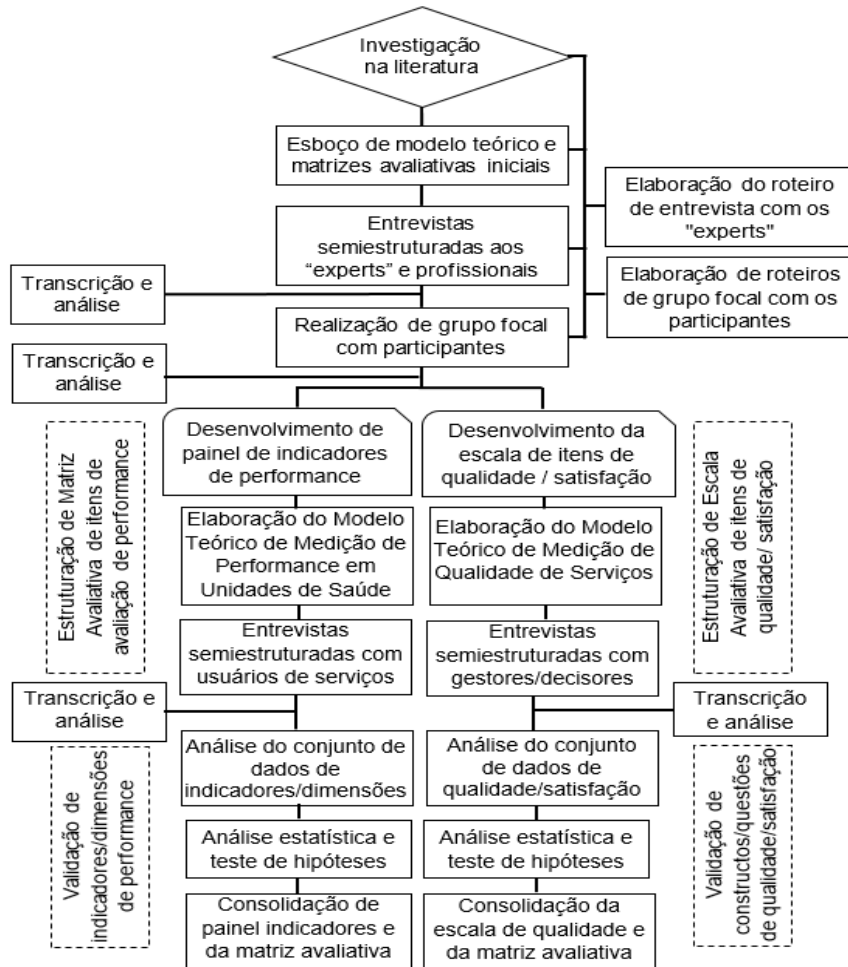


Figura 12 - Fluxo lógico do processo de estruturação, aplicação e validação dos instrumentos de itens de avaliação de performance e escala de qualidade/satisfação

No propósito de efetivar a coleta de dados em sistemas de informações disponíveis das organizações pesquisadas, foi usada uma técnica exploratória, de dados secundários, em especial, para o acesso no sistema nacional de informação do TabWin-Datasus e no TabNet-ANS. Para tal, foi firmado um Termo de Fiel Depositário através de um gestor do MS (apêndice 2).

Os dados coletados foram sistematizados e analisados para a validação dos itens de análises propostos em cada instrumento, através de uma triangulação de métodos (qualitativo e quantitativo), em um fluxo de validação, que envolveu métodos e técnicas de estatística descritiva e inferencial, conforme detalhado no tópico 6.

5.2. Os modelos e as variáveis utilizadas no estudo

As etapas e fases operacionais de pesquisa, todas de matiz *prospectivas e transversais*, envolveram grupos focais (painéis de especialistas), grupos pilotos (gestores e profissionais) e, após a adequação dos instrumentos semiestruturados em fases sucessivas e incrementais, a aplicação de pesquisa de campo, inclusive com uma fase de pesquisa à distância (*online*) de captação das preferências dos decisores chaves e, ainda, das dos usuários (no caso da avaliação da qualidade percebida).

A teoria dos *stakeholders*, parece aplicar-se bem na organização pública, e em especial na área de saúde, pois considera a satisfação dos objetivos dos múltiplos atores, mesmo com interesses contraditórios (empregados, gestores, fornecedores, clientes, consumidores, população, contribuintes, e apesar de ampliar algumas das limitações da *teoria da agência* (Alves, 2012). Além disso, as diferentes categorias de atores chaves (*key stakeholders*) têm diferentes pontos de vistas sobre como e onde devem ser aplicados os recursos humanos, financeiros e organizacionais, além de que fatores baseados em recursos e fatores cognitivos influenciam como a organização acomoda as pressões dos diferentes grupos (Austen, 2012). Diante das diversas perspectivas possíveis dos diferentes grupos de *stakeholders*, uma questão central no desenvolvimento de instrumentos de medição de desempenho é a validação da metodologia pelas principais partes interessadas (*key stakeholders*) e, em especial, pelos tomadores de decisão. Primeiro, para se saber se os itens de avaliação propostos respondem as questões práticas dos processos de decisão, depois para testar a facilidade de aplicabilidade do instrumento no contexto de todos os envolvidos no processo de avaliação e decisão.

Diante disso, na presente investigação, fez-se um esforço especial para estabelecer um método adequado de validação do instrumento de recolha de dados, do *proxy* do painel de indicadores e dimensões de performance e da escala de qualidade/satisfação percebida, através da metodologia colaborativa aplicada neste trabalho como descrito a seguir.

Para validar os diferentes itens de avaliação de cada um dos instrumentos, usou-se de uma triangulação de métodos (qualitativo e quantitativo), através da participação dos decisores, planejadores e gestores da saúde, notadamente nas instâncias decisórias superiores de governança das organizações de saúde, através de um fluxo de validação detalhado a seguir.

A parte empírica desta investigação foi desenvolvida em diversas fases subsequentes e incrementais de estruturação e aprimoramento dos instrumentos e questionários (ciclo 1 – agregação de todas as fases para a estruturação e validação dos instrumentos) e, por fim, aplicada em uma pesquisa de campo em fases únicas (ciclo 2), conforme as atividades e metodologias distribuídas em cada um dos estudos/etapas da investigação, sintetizadas nas duas figuras 13 e 14, a seguir.

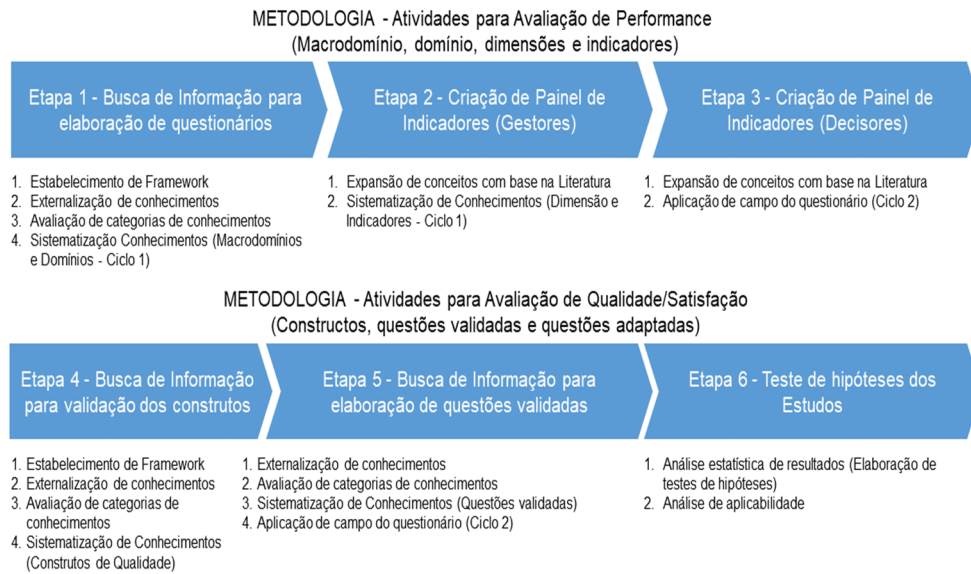


Figura 13 - Fluxo lógico de atividades para avaliação de performance e avaliação de qualidade/satisfação.

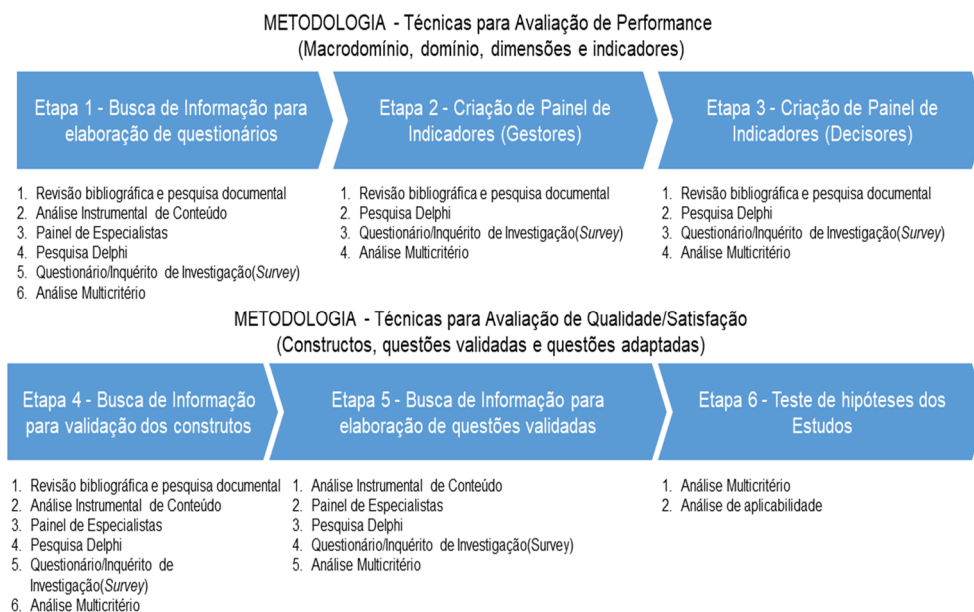


Figura 14 - Fluxo lógico de metodologias para avaliação de performance e avaliação de qualidade/satisfação.

Após a definição inicial de categorias e subcategorias analíticas pelos painéis de especialistas, definiu-se as variáveis de cada um dos estudos a fim de estruturar o processo analítico das respostas. No estudo/etapa 1 estabeleceu-se como variáveis dependentes os domínios e macrodomínios de avaliação. Nos estudos/etapas 2 e 3 as variáveis dependentes foram os indicadores e dimensões de performance. No estudo/etapa 4 definiu-se como variáveis dependentes os constructos, as subdimensões e macrodimensões de constructos. No estudo/etapa 5 as variáveis dependentes foram as questões, as subdimensões e macrodimensões

de qualidade percebida. Por sua vez, nos estudos 1 a 3, a variável latente construída foi a Performance Global Percebida e, nos estudos 4 e 5, foi a Qualidade Global Percebida. Por fim, as possíveis variáveis moderadoras testadas, em todos os estudos, foram o nível de assistência à saúde, o nível de complexidade assistencial e o nível de cargo/função dos decisores, sendo que, no caso dos usuários, a terceira variável moderadora foi o nível de escolaridade.

Nas primeiras fases empíricas dos estudos 1 e 2, em consulta aos especialistas e profissionais, indagou-se sobre as preferências e conhecimentos prévios dos decisores chaves em busca de externalizar e sistematizar conhecimentos prévio dos mesmos. Para a estruturação do problema e ordenação das categorias e subcategorias de avaliação (domínios e macrodomínios de avaliação e indicadores e dimensões de performance do painel de indicadores de performance), utilizou-se a técnica *Delphi* de entrevistas, que consiste em buscar o consenso por sucessivas consultas ao grupo de decisores de forma escrita e estruturada (Minayo et al., 2010). Este mesmo procedimento ocorreu nas primeiras fases empíricas dos estudos 4 e 5, em consulta aos especialistas e profissionais, em busca de uma tradução transcultural da escala SERVQUAL quanto aos constructos (ideias-forças) de cada uma das 22 questões para estabelecimento de categorias, subcategorias e macrocategorias de qualidade (constructos, questões, subdimensões e macrodimensões de qualidade).

Nas fases subsequentes os participantes foram indagados de forma simplificada sobre o nível ou grau de relevância/importância quanto aos itens de avaliação de performance (estudos 1 a 3) e aos itens de mensuração de qualidade (estudos 4 e 5). Em todas as etapas e fases de ponderações (valorações) dos itens de categorias e subcategorias analíticas propostos, através de questionários semiestruturados, utilizou-se uma escala de *Likert* de cinco níveis (1 = Nada importante/ relevante até 5 = Extremamente importante/ relevante).

Na primeira rodada de validação, a partir dos conhecimentos prévios e das preferências dos decisores chaves, sistematizados na fase anterior, coletou-se as valorações atribuídas pelos informantes, em fases sucessivas e incrementais, que resultou em instrumentos validados com um grupo de macrodomínios e domínios de avaliação (apêndice 5). Na próxima etapa (estudo 2) estes itens de avaliação validados foram devidamente alinhados com os quadros conceituais (*framework*) e resultaram em um grupo de indicadores e dimensões de performance – o primeiro *proxy* de painel de indicadores e dimensões de performance (apêndice 6), que foram as bases para uma segunda rodada de validação (estudo 3). Após a análise destes itens de avaliação por outro grupo de gestores, com realinhamento com os quadros conceituais (*framework*) agregou-se uma importante expansão de novos indicadores, sem alterações das dimensões de performance, que resultou no *proxy* de painel de indicadores e dimensões de performance

definitivo (apêndice 6), sendo concluído o ciclo 1. Por fim, no ciclo 2 de aplicação de campo, conduziu-se uma aplicação pragmática dos instrumentos consolidados no ciclo 1 para executar uma avaliação de resultados de forma analítica dos indicadores e dimensões de performance, em busca de aferir a adequação do modelo proposto na visão de uma amostra de decisores.

Da mesma forma, na primeira rodada de validação, partiu-se das 22 questões da escala SERVQUAL original e dos constructos (ideias-forças) propostos pelos painéis de especialistas iniciais, sistematizados na fase anterior, e coletou-se as valorações atribuídas pelos informantes, em fases sucessivas e incrementais, que resultou em instrumentos validados com um grupo de constructos, distribuídos em subdimensões e macrodimensões de constructos (apêndice 7). Na próxima etapa de validação (estudo 5) estes itens de mensuração de qualidade validados foram devidamente analisados por um novo painel de especialistas e profissionais do SUS que compuseram questões de qualidade, distribuídos em novas subdimensões e macrodimensões de qualidade em uma inovadora escala de qualidade/ satisfação (apêndice 8), sendo concluído o ciclo 1 pela validação, em uma amostra de usuários, dos constructos, subdimensões e macrodimensões de constructos. Por fim, no ciclo 2 de aplicação de campo, conduziu-se uma aplicação pragmática dos instrumentos consolidados no ciclo 1, para executar uma avaliação de resultados de forma analítica das questões, subdimensões e macrodimensões da escala de qualidade/ satisfação, em busca de aferir a adequação do modelo proposto na visão de uma amostra ampliada de usuários do SUS.

Na figura 15, estruturou-se uma visão global do fluxo lógico da fase empírica da investigação e suas respectivas lógicas de agregação de resultados, em seus ciclos de análises (ciclos 1 e 2) e suas etapas de investigação (estudos 1 a 5).

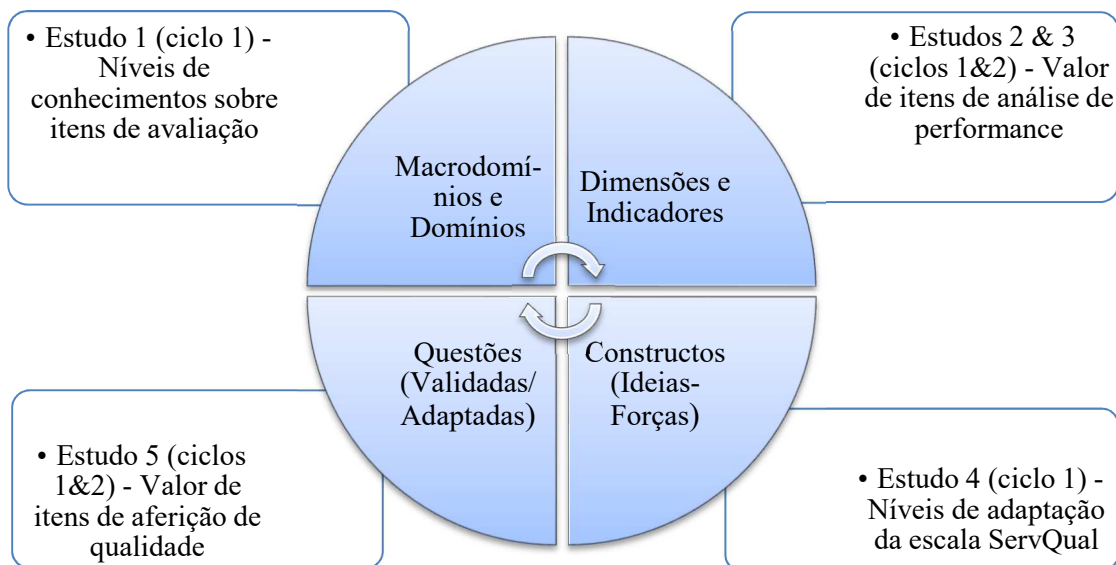


Figura 15 - Fluxo lógico integrativo de aplicação da investigação (avaliação de performance – estudos 1 a 3 e avaliação de qualidade – estudos 4 e 5).

A aplicação empírica de instrumentos e questionários envolveu diferentes amostras e grupos de respondentes nas diversas fases sucessivas e incrementais.

Nas fases de validação incremental (ciclo 1) dos instrumentos de indicadores e dimensões de performance, a amostra foi não-aleatória por conveniência conforme aceite de 220 participantes convidados (especialistas, profissionais e gestores). Assim, por facilidade de autorização de acesso, resolveu-se aplicar os questionários em visitas pré-agendadas ou por mensagem eletrônica (e-mail institucional) junto aos gestores de hospitais e serviços de saúde da amostra. Por sua vez, nas fases de validação incremental (ciclo 1) dos instrumentos da escala de qualidade/satisfação percebida, a amostra foi não-aleatória por conveniência conforme aceite de 195 participantes convidados (especialistas, profissionais de ouvidoria e gestores de qualidade) distribuídos em 10 estados/províncias do Brasil (quadro 6).

Na fase de aplicação de campo do instrumento validado (ciclo 2) de painéis de indicadores e dimensões de performance, após a consolidação inicial com alguns dos participantes do ciclo anterior convidou-se 5.165 dos 5.595 secretários de saúde do Brasil que tinham cadastros na base oficial do Conass e Conasems, por mensagem eletrônica *online* (e-mail institucional), com devolução imediata de 1.045 por problemas no cadastro ou pela alta rotatividade de gestores do SUS e, Em seguida, com uma amostra não-aleatória por conveniência, tendo recebido o aceite e respostas de 233 participantes convidados (decisores), cerca de 5,6 % dos convidados, após três rodadas de envios de mensagens, em meses diferentes.

Na fase de aplicação de campo da escala de qualidade/satisfação percebida (ciclo 2), foram incluídas na amostra 2.547 usuários em atendimento nas unidades de saúde, que aceitaram participar do inquérito, desde que os gestores de topo tivessem firmado o termo autorização da investigação, sendo excluídas as unidades que não cumpriram estes requisitos, sendo 25 a 40 respondentes por cada extrato (por estabelecimento ou por setor da unidade de saúde), em busca da aproximação da amostra para a distribuição normal, com cautelas quanto a alocação não-intencional (aleatória) de indivíduos, a fim de permitir a aplicação da teoria de limite central.

Durante as diversas fases do ciclo de validação de instrumentos, os decisores-chaves poderiam propor novos itens de avaliação de performance e/ou adequar suas descrições, não indicadas em fases anteriores da validação, desde que indicassem também uma matriz avaliativa (grau de preferência, o grau de usabilidade por nível decisório e o grau de conhecimento pessoal, etc), além dos dados de identificação do gestor participante. Em cada uma das fases posteriores incrementais os instrumentos foram ajustados para incorporar os novos itens de avaliação sugeridos, conforme as preferências indicadas pelos gestores desde que o nível mínimo de valoração do item incluído fosse acima da média (superiorou cerca de 3,0). Esta mesma lógica foi usada em todas as fases do ciclo de validação dos itens de análise da escala de qualidade/satisfação percebida, distribuídos por subdimensões e macrodimensões de qualidade percebida.

A partir da avaliação multicritério e do grau de relevância atribuído aos itens de avaliação da performance, foi proposto um *proxy* de painel de indicadores e dimensões de performance em saúde, devidamente validado (tópicos 6.1.1 e 6.1.2). Da mesma forma, a partir da avaliação multicritério e do grau de relevância atribuído aos itens de avaliação da qualidade/satisfação, propôs-se um *proxy* de questionário de escala de qualidade/satisfação percebida pelo usuário do SUS, devidamente validada (tópicos 6.2.1 e 6.2.2).

No quadro abaixo, sintetizou-se o fluxo e o percurso da investigação empírica, com os métodos e técnicas utilizadas, além da identificação de cada etapa, fases e grupos de respondentes abordados.

Procedimentos metodológicos	Dimensões de Avaliação	Indicadores de Desempenho	Constructos de Qualidade	Avaliação de Qualidade e Satisfação
Abordagem de pesquisa	Quantitativa, analítica, prospectiva e transversal			

Método	Estudos de multicaseos (Minayo et al., 2010; Yin, 2009)			
Coleta de dados – Levantamento documental	Bases teóricas e evolução da avaliação; quadro teórico-conceitual (<i>framework</i>) de avaliação em saúde; modelo teórico-lógico das dimensões de avaliação em serviços de saúde.	Modelo teórico-lógico das dimensões <i>versus</i> Indicadores de avaliação em serviços de saúde; modelos de painéis de indicadores de saúde (mundial e Brasil).	Bases teóricas de avaliação de qualidade e satisfação no campo de “serviços”; teorias e modelos de mensuração da satisfação; escala SERVQUAL.	Bases teóricas de avaliação de qualidade e satisfação no campo de “serviços”; teorias e modelos de mensuração da satisfação; escala SERVQUAL.
Definição de amostra	Amostragem por conveniência e saturação teórica (painel de especialista e grupo piloto)	Amostragem por conveniência (painel de especialista e grupo piloto e de campo)	Amostragem por conveniência (grupo piloto) e estratificada proporcional (campo)	Amostragem por conveniência (grupo piloto) e estratificada proporcional (campo)
Técnicas de pesquisa	Entrevista semiestruturada (painéis de especialistas e grupos focais)	Entrevista semiestruturada (painéis de especialistas, grupos focais e Delphi); (Pesq. de campo)	Entrevista semiestruturada (grupos focais e Delphi); Entrevista estruturada (grupo pilotos)	Entrevista semiestruturada (grupos focais e Delphi); Entrevista estruturada (Pesq. de campo)

<p>- Casos de estudos (natureza e categorias)</p> <p>- Sujeitos de pesquisa (quantidade de grupos / número de Informantes)</p> <p>- Abrangência de casos abordados (Siglas UF – estados / províncias)</p>	<p>- Rede pública e privada (Gestor e Profissional)</p> <p>- Fases 1&2: 05 painéis, 53 informantes (CE - RN - PI - MA – BA - PA – DF-RO)</p> <p>- Fase 3: 03 grupos pilotos, 55 informantes (CE - MA)</p> <p>- Fase 4: 03 grupos pilotos, 112 informantes (BA - PA - DF - GO - MT - PR)</p>	<p>- Rede pública e privada (Gestor e Profissional)</p> <p>- Fase 1: 02 painéis, 15 informantes (CE - MA - PA - AM)</p> <p>- Fase 2: 01 painéis, 31 informantes (BA - PA - DF - GO)</p> <p>- Fase 3: 03 grupos pilotos, 81 informantes (MT - PR)</p> <p>- Fase 4: 233 Pesq. Campo (CE, MA, PI, BA, SE, PE, PB, RN, DF, GO, MT, MS, PR, RS, SC, RJ, MG, SP, RO, PA, AP, TO)</p>	<p>- Rede pública (Gestor-Profiss./usuário)</p> <p>- Fase 1&2: 02 painéis, 25 informantes (CE - RN - PI - MA - PA - RO)</p> <p>- Fase 3: 07 grupos pilotos, 170 informantes (CE - MA - BA - DF - GO - MT)</p> <p>- Fase 4: 15 grupos de campo, 506 informantes - usuários (CE)</p>	<p>- Rede pública (Gestor-Profiss./usuário)</p> <p>- Fase 1&2: 04 painéis, 78 informantes (CE - RN - PI - MA - PA - DF - RO)</p> <p>- Fase 3: 03 grupos pilotos, 62 informantes (CE- MA - BA - PA - DF - GO - MT)</p> <p>- Fase 4: 15 grupos de campo, 506 informantes - usuários (CE)</p>
<p>Coleta de dados – (Roteiro de entrevista/ Questionário, com ponderação - <i>Likert</i> de 5 níveis)</p>	<p>- Fases 1&2: Versão 5A&5B (Apend.)</p> <p>- Fase 3: Versão 5C&5D (Apend.)</p> <p>- Fase 4: Versão 5E (Apend.)</p>	<p>- Fase 1&2: Versão 6A (Apend.)</p> <p>- Fase 3: Versão 6B (Apend.)</p> <p>- Fase 4: Versão 6C (Apend.)</p>	<p>- Fase 1&2: Versão 7A&7B (Apend.)</p> <p>- Fase 3: Versão 7C&7D (Apend.)</p> <p>- Fase 4: Versão 7E (Apend.)</p>	<p>- Fase 1&2: Versão 8A&8B (Apend.)</p> <p>- Fase 3: Versão 8C&8D (Apend.)</p> <p>- Fase 4: Versão 8D&8E (Apend.)</p>

Período da coleta de dados	Março a maio de 2017(Fases 1 a 4)	- Março a maio de 2017(Fases 1 a 3) - Abril a junho de 2018 (Fase 4)	Maio a agosto de 2017 (Fases 1 a 4)	Julho a setembro de 2017 (Fases 1 a 4)
Tratamento dos dados	Itens de performance analisados - Cinco macrodomínios com 24 domínios; Sete dimensões com 20 indicadores de gestores e com 56 indicadores de decisores.		Itens de qualidade analisados -: 40 constructos e 20 questões validadas, distribuídas por 13 subdimensões (e pelas cinco macrodimensões originais da SERVQUAL, tendo macrodimensão adicional (Acesso/Acessibilidade).	
Análise dos dados	Análise descritiva qualitativa comparativa; Análise estatística quantílica e factorial; teste de hipóteses.		Análise descritiva qualitativa comparativa.	Análise estatística quantílica e factorial; teste de hipóteses.

Quadro 6 - Síntese dos procedimentos metodológicos da investigação.

Destacou-se a amplitude de representatividade das amostras. Na fase de validação dos instrumentos (ciclo 1) dos itens de avaliação de performance e dos itens da escala de qualidade/satisfação percebida, alcançou-se mais de 1/3 dos estados/províncias do Brasil (Ceará - CE, Rio Grande do Norte - RN, Piauí - PI, Maranhão - MA, Bahia - BA, Pará - PA, Distrito Federal - DF, Rondônia - RO, Mato Grosso - MT e Paraná PR). Na fase de coleta de campo (ciclo 2) dos painéis de indicadores e dimensões de performance, por facilidade de aplicação *online* (uso do FormSUS), envolveu decisores de quase todas as unidades federadas do Brasil, de todas as regiões geográficas: Norte (Rondônia - RO, Pará - PA, Amapá - AP e Tocantins - TO), Nordeste (Ceará - CE, Maranhão - MA, Piauí - PI, Bahia - BA, Sergipe - SE, Pernambuco - PE, Paraíba - PB e Rio Grande do Norte - RN), Centro-Oeste (Distrito Federal - DF, Goiás - GO, Mato Grosso - MT e Mato Grosso do Sul - MS), Sudeste (Rio de Janeiro - RJ, Minas Gerais - MG e São Paulo - SP) e Sul (Paraná - PR, Rio Grande do Sul - RS e Santa Catarina - SC). Na fase de coleta de campo (ciclo 2), da aplicação da escala de qualidade/satisfação percebida de usuários, dada a dificuldade de aplicação manual dos questionários, a amostra limitou-se à quatro municípios do Ceará, mas alcançou sete grupos pilotos de usuários em 74 unidades assistenciais do SUS das diversas tipologias e níveis de atenção/cuidados de saúde.

5.2.1.O fluxo de aplicação dos instrumentos do painel de indicadores e dimensões de performance.

No estudo/etapa 1 voltou-se para a perspectiva de domínios e macrodomínios de avaliação e nos estudos/etapas 2 e 3 abordou-se a perspectiva de indicadores e dimensões de performance, com seus fluxos e etapas descritos a seguir.

A etapa 1 da investigação foi dividida em quatro fases sucessivas e incrementais de adaptação, sintetizadas nas figuras 16 e 17, a seguir.

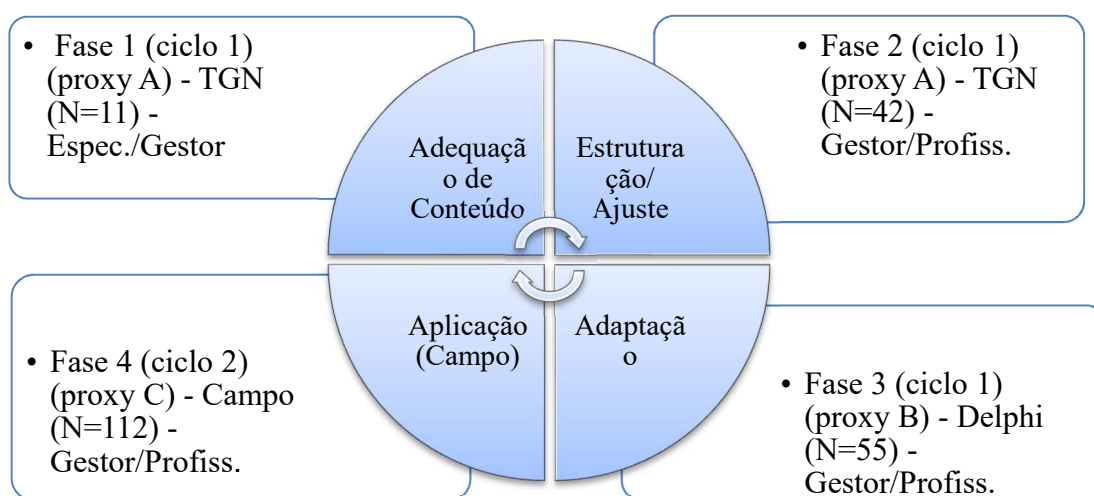


Figura 16 - Fluxo lógico da aplicação da investigação e agregação de resultados (etapa 1 - macrodomínios e domínios de avaliação).

Na primeira fase – *qualitativa e descritiva* -, através de grupos focais e técnica de grupo nominal (TGN) aplicou-se dois questionários semiestruturados (prospecção em duas fases -1 e 2 - e análise conjunta na *proxy A*) quanto as metodologias e indicadores chaves para a tomada de decisão em saúde, através de questões norteadoras com respostas abertas além da ponderação de alguns itens propostos, provenientes do ML proposto por Hartz e Vieira-Da-Silva (2005) (figura37). Estruturou-se uma visão global das preferências e relevâncias, atribuídas pelas partes interessadas, quanto aos níveis de conhecimentos globais de *macrodomínios* e *domínios* de avaliação de mensuração de performance (desempenho) em organizações ou serviços de saúde.

Na terceira fase, da etapa 1 – *quantitativa e analítica* -, através de grupos *Delphi*, partiu-se dos resultados construídos pelos dois grupos focais da fase inicial e aplicou-se um novo instrumento semiestruturado (prospecção da fase 3 e análise na *proxy B*). Estruturou-se uma visão adaptada das preferências e valorações quanto as importâncias (relevâncias) atribuídas

pelos agentes de decisão, quanto aos níveis de conhecimentos globais de macrodomínios e domínios de avaliação, que resultou em uma versão ampliada para a fase posterior.

Por fim, a fase final da etapa 1, seguiu as mesmas técnicas da fase 3 e aplicou-se um instrumento semiestruturado ampliado (prospecção da fase 5 e análise na *proxy C*). Estruturou-se uma visão adaptada das preferências e valorações quanto as importâncias (relevâncias) atribuídas pelos agentes de decisão, quanto aos níveis de conhecimentos globais de macrodomínios e domínios de avaliação, que resultou nas bases de dimensões e indicadores de performance em saúde a ser validados na etapa 2 da investigação operacional.

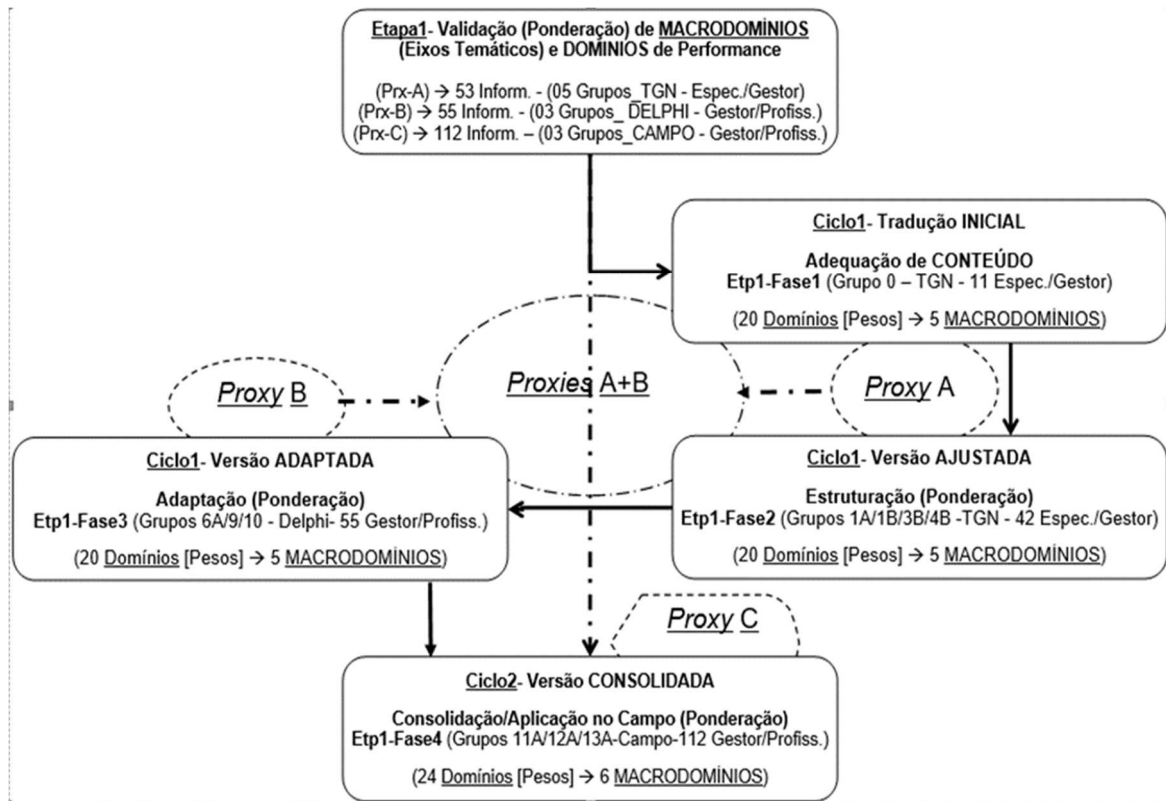


Figura 17 - Fluxo do processo de validação de macrodomínios e domínios de avaliação (etapa 1).

Estes fluxos têm suas análises complementadas pelo esquema lógico de análise de macrodomínios e domínios de avaliação. Na primeira etapa, no âmbito do estudo 1, procedeu-se a análise estatística de dados das *proxies* da valoração dos domínios de avaliação, além de seus reflexos nos macrodomínios avaliação, descritos a seguir:

Análise Cód.	Etapas/Estudo	Fases	Proxy	Informantes (N) / Categorias
A1) Análises de respostas atribuídas através de 4 <i>proxies</i> (pesos atribuídos para os <u>domínios</u>)				
A1.a)-Inicial	Etp1 – Ciclo1	1+2	A	53 (11+42) Espec./Gestor
A1.b)-Ajuste	Etp1 – Ciclo1	3+4	B	55 (11+42) Gestor/Profiss.

A1.c)-Síntese	Etp1 – Ciclo1	1+2+3+4	A+B	108 (53+55) Gestor/Profiss.
A1.d)-Campo	Etp1 – Ciclo2	5	C	112 Gestor/Profiss.
A.2) Análises de respostas calculadas através de 4 <i>proxies</i> (pesos refletidos para os <u>macrodomínios</u>):				
A2.a)-Inicial	Etp1 – Ciclo1	1+2	A	53 (11+42) Espec./Gestor
A2.b)-Ajuste	Etp1 – Ciclo1	3+4	B	55 (11+42) Gestor/Profiss.
A2.c)-Síntese	Etp1 – Ciclo1	1+2+3+4	A+B	108 (53+55) Gestor/Profiss.
A2.d)-Campo	Etp1 – Ciclo2	5	C	112 Gestor/Profiss.
A3) Análises de respostas calculadas através de um <i>proxy</i> (pesos refletidos para os <u>macrodomínios</u> e para a <u>Performance Global Percebida</u>)				
A3.a)-Campo	Etp1 – Ciclo2	5	C	112 Gestor/Profiss.

Quadro 7 - Síntese das análises descritivas dos níveis de conhecimento (atribuídos) dos principais domínios de avaliação (e refletidos nos macrodomínios de avaliação).

Por sua vez, a etapa 2 da investigação foi dividida em três fases sucessivas e incrementais de adaptação, sintetizadas nas figuras 18 e 19, a seguir. Por fim, a quarta fase foi constituída pela aplicação de campo (*online* – investigação a distância) que conformou a etapa 3 deste estudo.



Figura 18 - Fluxo lógico da aplicação da investigação e agregação de resultados (estudos/etapas 1 & 2 - dimensões e indicadores de performance).

Na primeira fase – *quantitativa e analítica* -, através de grupo focal e técnica de grupo nominal (TGN) aplicou-se um instrumento semiestruturado (prospecção da fase 1 e análise na *proxy D*) quanto as principais dimensões e indicadores de performance em organizações e unidades de saúde, desta vez, voltado para ponderações de cada item proposto, mas com possibilidade de acréscimos de novos itens, construído a partir das bases consolidadas na etapa 1 da pesquisa e do ML (figura 37). Estruturou-se uma visão inicial (adaptada) das percepções, preferências e relevâncias atribuídas pelos agentes de decisão às dimensões e indicadores de performance (desempenho) em organizações ou serviços de saúde.

Ainda na etapa 2, nas fases 2 e 3 – *quantitativa e analítica* -, através de grupos *Delphi*, partiu-se dos resultados construídos pelos grupo focal da fase inicial e aplicou-se dois questionários semiestruturados através da análise e ponderação de itens propostos em uma listagem de dimensões e indicadores chaves de performance (desempenho) para a tomada de decisão em saúde (prospecção em duas fases - 2 e 3 - e análise conjunta na *proxy E*), a partir das bases da versão construída na fase anterior da pesquisa. Estruturou-se uma versão ajustada (fase 2), e outra validada (fase 3) das percepções, preferências e relevâncias atribuídas pelos agentes de decisão às dimensões e indicadores de performance (desempenho) em organizações ou serviços de saúde.

Na etapa 3, surgiram as fases 6 e 7 – *quantitativa e analítica* -, através de grupos *Delphi*, partiu-se dos resultados construídos pelos dois grupos pilotos anteriores e adicionou-se novos indicadores apontados pela literatura mundial sobre indicadores de performance, mas que os gestores e profissionais não haviam apontados. Assim, aplicou-se este novo instrumento semiestruturado ampliado. (Na fase 6 consolidou-se com 10 decisores respondentes das fases anteriores e na fase 7 a adição da aplicação de campo *online* com 223 decisores do SUS, tendo a análise conjunta na *proxy G*). Estruturou-se uma visão validada e aplicada no campo das percepções, preferências e relevâncias atribuídas pelos agentes de decisão às dimensões e indicadores de performance (desempenho) em organizações ou serviços de saúde, numa amostra selecionada de gestores da rede de atenção do SUS.

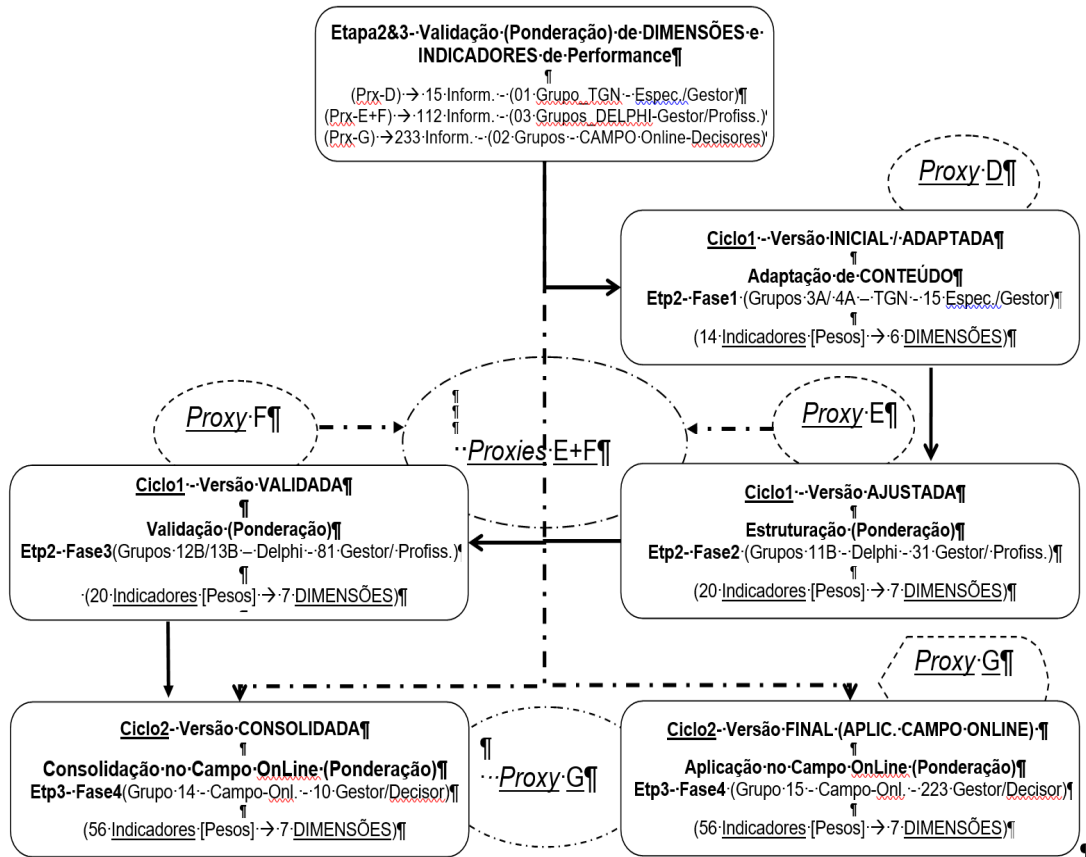


Figura 19 - Fluxo do processo de validação de dimensões e indicadores de performance (etapas 2 & 3)

Estes fluxos têm suas análises complementadas pelo esquema lógico de análise de dimensões e indicadores de performance. Na primeira etapa, no âmbito dos estudos 2 e 3, procedeu-se a análise estatística de dados das *proxies* da valoração dos indicadores de performance, além de seus reflexos nas dimensões de performance, descritos a seguir:

Análise Cód.	Etapas/Estudo	Fases	Proxy	Informantes (N) / Categorias
A1) Análises de respostas atribuídas através de 4 <i>proxies</i> (pesos atribuídos para os indicadores)				
A1.a)-Inicial	Etp2 – Ciclo1	2	E	31 Espec./Gestor
A1.b)-Ajuste	Etp2 – Ciclo1	3	F	81 Espec./Gestor
A1.c)-Síntese	Etp2 – Ciclo1	2+3	E+F	112 Espec./Gestor
A1.d)-Campo	Etp3– Ciclo2	6+7	G	233 Decisores/Gestores
A2 e A3) Análises de respostas calculadas através de 4 <i>proxies</i> (pesos refletidos para as dimensões):				
A2.a)-Inicial	Etp2 – Ciclo1	2	E	31 Espec./Gestor
A2.b)-Ajuste	Etp2 – Ciclo1	3	F	81 Espec./Gestor
A2.c)-Síntese	Etp2 – Ciclo1	2+3	E+F	112 Espec./Gestor
A3.a)-Campo	Etp3– Ciclo2	6+7	G	233 Decisores/Gestores
A4) Análises de respostas calculadas através de um <i>proxy</i> (pesos refletidos nas dimensões e para a Performance Global Percebida)				
A4.a)-Campo	Etp2 – Ciclo2	5	C	112 Gestores/Profissionais

A4.b)-Campo	Etp2 – Ciclo2	6+7	C	233 Gestores/Profissionais
A5.a) Análises de respostas calculadas através de um <i>proxy</i> (pesos simultâneos refletidos dos indicadores nas dimensões de Performance)				

Quadro 8 - Síntese das análises descritivas dos níveis de conhecimento (atribuídos) dos principais indicadores (e refletidos nas dimensões de performance).

Dessa forma, descreveu-se os fluxos de etapas e fases da pesquisa empírica quanto aos instrumentos dos painéis de itens de avaliação de performance além das atividades, métodos e técnicas de análises aplicadas aos resultados obtidos.

5.2.2.O fluxo de aplicação dos instrumentos da escala/satisfação de qualidade percebida.

No estudo/etapa 4 voltou-se para a perspectiva de constructos de qualidade percebida e no estudo/etapa 5 abordou-se a perspectiva de questões validadas de qualidade percebida, com seus fluxos e etapas descritos a seguir.

Para iniciar os estudos para a estruturação da escala de qualidade percebida, a etapa 4 da pesquisa foi dividida em quatro fases sucessivas e incrementais de adaptação, sintetizada nas figuras 20 e 21, a seguir.

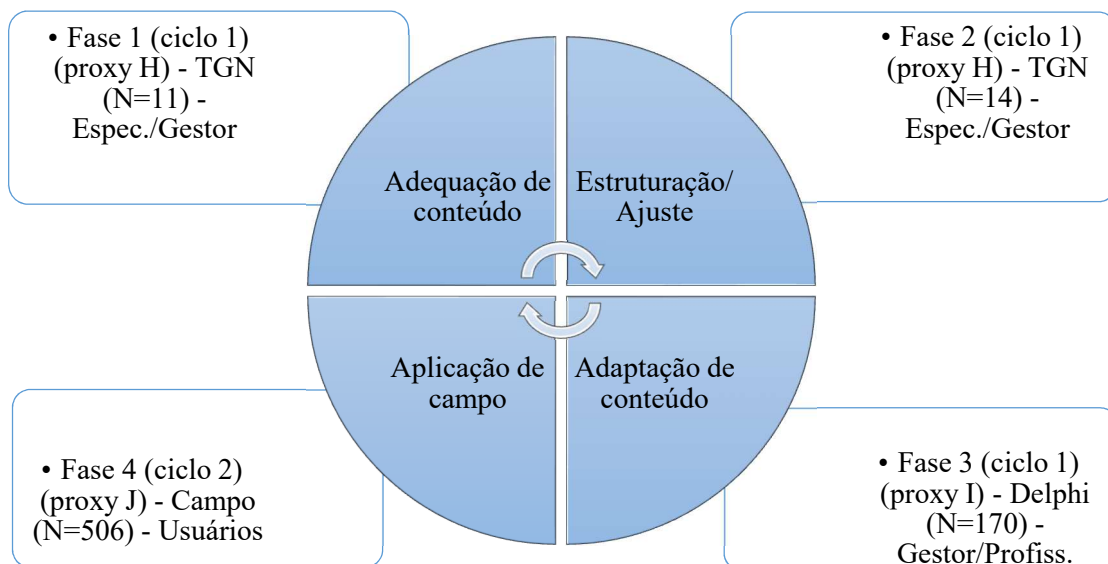


Figura 20 - Fluxo lógico da aplicação da investigação e agregação de resultados (estudo/etapa 4 - macrodimensões, subdimensões e constructos de qualidade percebida).

Na primeira fase – *qualitativa e descritiva* -, através de grupos focais e técnica de grupo nominal (TGN) aplicou-se dois questionários semiestruturados (prospecção em duas fases -1 e 2 - e análise conjunta na *proxy H*) quanto aos constructos (ideias-força ou conceitos que

relacionam a observação e o enquadramento teórico) relevantes para a mensuração de qualidade percebida, com respostas abertas além da ponderação de alguns itens propostos, provenientes da tradução transcultural e a análise da adaptabilidade do instrumento original escala SERVQUAL (Parasuraman et al., 1988, 1991a) no contexto do SUS. Estruturou-se uma visão adaptada das preferências e relevâncias atribuídas pelos grupos de especialistas e de partes interessadas (profissionais e usuários-utilizadores), aos constructos relevantes (importantes) para a mensuração de qualidade percebida (satisfação e adequação ao uso) em organizações ou serviços de saúde.

Na terceira fase, da etapa 4 – *quantitativa e analítica* -, através de grupos *Delphi*, partiu-se dos resultados construídos pelos dois grupos focais da fase inicial e aplicou-se um novo instrumento semiestruturado (prospecção da fase 3 e análise na *proxy I*). Estruturou-se uma visão validada das preferências e valorações quanto as importâncias (relevâncias) atribuídas pelos grupos pilotos (visão direta de profissionais *versus* visão indireta dos usuários) aos constructos relevantes (importantes) para a mensuração de qualidade percebida (satisfação e adequação ao uso) em organizações ou serviços de saúde, que resultou em uma versão ampliada para a fase posterior.

Por fim, a fase final da etapa 4, seguiu as mesmas técnicas da fase 3 e aplicou-se um instrumento semiestruturado ampliado (prospecção da fase 4 e análise na *proxy J*). Estruturou-se uma visão consolidada e aplicada no campo das percepções, preferências e relevâncias atribuídas pelos grupos de usuários aos constructos relevantes (importantes) para a mensuração de qualidade percebida (satisfação e adequação ao uso) em uma amostra de usuários-utilizadores selecionados em uma rede assistencial do SUS. Estes constructos validados e consolidados, quanto aos níveis de relevância (importância) atribuídos pelos grupos de especialistas e de partes interessadas (profissionais e usuários-utilizadores) resultou numa base adicional de conhecimento para a posterior adequação de conteúdo das questões relevantes para a mensuração de qualidade percebida e, portanto, subsidiou a estruturação de dois instrumentos inovadores a serem validados nas etapas 5 e 6 da investigação operacional.

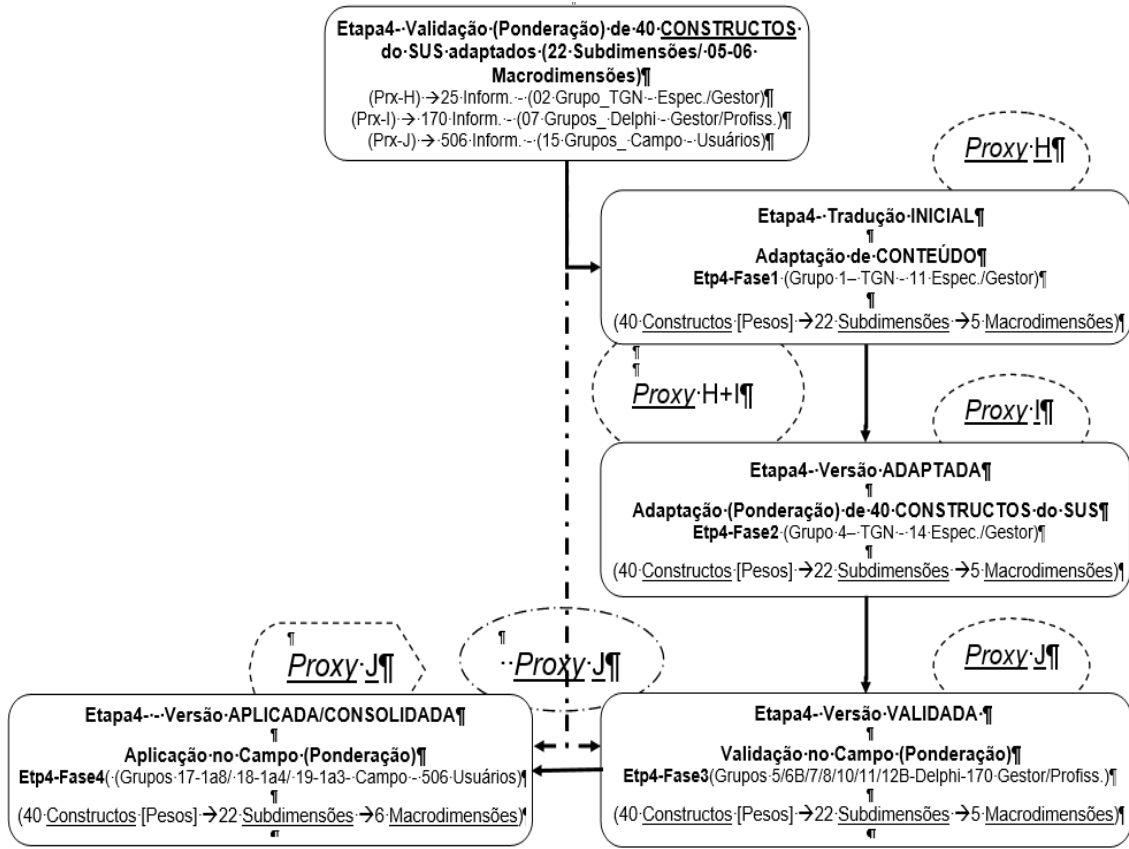


Figura 21 - Fluxo do processo de validação de constructos adaptados ao SUS (qualidade percebida) (etapa 4).

Estes fluxos têm suas análises complementadas pelo esquema lógico de análise de constructos de qualidade e das macrodimensões e subdimensões de qualidade. Na primeira etapa, no âmbito do estudo 4, procedeu-se a análise estatística de dados das proxies da avaliação das subdimensões de qualidade, além de seus reflexos nas macrodimensões de qualidade, descritos a seguir:

Análise Cód.	Etapas/Estudo	Fases	Proxy	Informantes (N) / Categorias
A1) Análises de respostas atribuídas através de 4 <i>proxies</i> (pesos atribuídos para os constructos)				
A1.a)-Inicial	Etp4 – Ciclo1	1+2	H	25 Espec./Gestor
A1.b)-Ajuste	Etp4 – Ciclo1	3	I	170 Espec./Gestor
A1.c)-Síntese	Etp4 – Ciclo1	1+2+3	H + I	195 Espec./Gestor
A1.d)-Campo	Etp4 – Ciclo2	4	J	506 Usuários
A2) Análises de respostas calculadas através de 4 <i>proxies</i> (pesos refletidos para as subdimensões):				
A2.a)-Inicial	Etp4 – Ciclo1	1+2	H	25 Espec./Gestor
A2.b)-Ajuste	Etp4 – Ciclo1	3	I	170 Espec./Gestor
A2.c)-Síntese	Etp4 – Ciclo1	1+2+3	H + I	195 Espec./Gestor
A2.d)-Campo	Etp4 – Ciclo2	4	J	506 Usuários

A3) Análises de respostas calculadas através de um <i>proxy</i> (pesos refletidos para os macrodómnios e para a Performance Global Percebida)				
A3.a)-Campo	Etp4 – Ciclo2	4	J	506 Usuários

Quadro 9 - Síntese das análises descritivas dos níveis de valoração (atribuídos) aos principais constructos propostos (e refletidos nas subdimensões e macrodimensões da escala de qualidade percebida).

Em seguida, a etapas 5 da pesquisa, foi sintetizada nas figuras 22 e 23, sendo dividida em quatro fases sucessivas e incrementais de adaptação.



Figura 22 - Fluxo lógico da aplicação da investigação e agregação de resultados (estudo/etapa 5 - macrodimensões, subdimensões e questões validadas).

Na primeira fase – *qualitativa e descritiva* -, através de grupos focais e técnica de grupo nominal (TGN) aplicou-se um questionário semiestruturado (prospecção etapa 5 – questões validadas) e análise na *proxy* K, quanto aos conteúdos das questões relevantes para a mensuração de qualidade percebida (um questionário com 20 questões voltadas para a satisfação), provenientes da tradução transcultural e a análise da adaptabilidade do instrumento original escala SERVQUAL (Parasuraman et al., 1988, 1991a) no contexto do SUS e adicionalmente dos constructos consolidados na etapa 4 deste estudo. Estruturou-se uma visão adaptada das preferências e relevâncias atribuídas pelos grupos de especialistas e partes interessadas, aos constructos relevantes (importantes) para a mensuração de qualidade percebida (satisfação e adequação ao uso) em organizações ou serviços de saúde.

Na segunda fase, da etapa 5 – *quantitativa e analítica* -, através de grupos *Delphi*, partiu-se dos resultados construídos pelos grupos focais de especialistas da fase inicial e aplicou-se um novo instrumento semiestruturado e análise na *proxy* L. Estruturou-se uma visão validada

das preferências e valorações quanto as importâncias (relevâncias) atribuídas pelos grupos pilotos (visão direta de profissionais *versus* visão indireta dos usuários) quanto aos conteúdos das *questões* relevantes para a mensuração de qualidade percebida (um questionário com 20 questões voltadas para a satisfação), que resultou em versão validada do questionário para a fase posterior.

A terceira fase, da etapa 5 – *quantitativa e analítica* -, através de grupos *Delphi*, partiu-se dos resultados construídos pelos grupos pilotos da fase anterior e aplicou-se um novo instrumento semiestruturado e análise na *proxy* M. Estruturou-se uma visão consolidada das preferências e valorações quanto as importâncias (relevâncias) atribuídas pelos grupos pilotos (visão direta de profissionais *versus* visão indireta dos usuários) quanto aos conteúdos das *questões* relevantes para a mensuração de qualidade percebida (um questionário com 20 questões voltadas para a satisfação), que resultou em versão consolidada do questionário para a fase posterior.

Por fim, a fase final da etapa 5, seguiu as mesmas técnicas e aplicou-se um instrumento semiestruturado consolidado e análise na *proxy* M+N. Estruturou-se uma visão consolidada e aplicada no campo das percepções, preferências e relevâncias atribuídas pelos grupos de usuários aos conteúdos das *questões* relevantes (importantes) para a mensuração de qualidade/satisfação percebida em uma amostra de usuários selecionados em sete grupos pilotos e 74 unidades assistenciais do SUS.

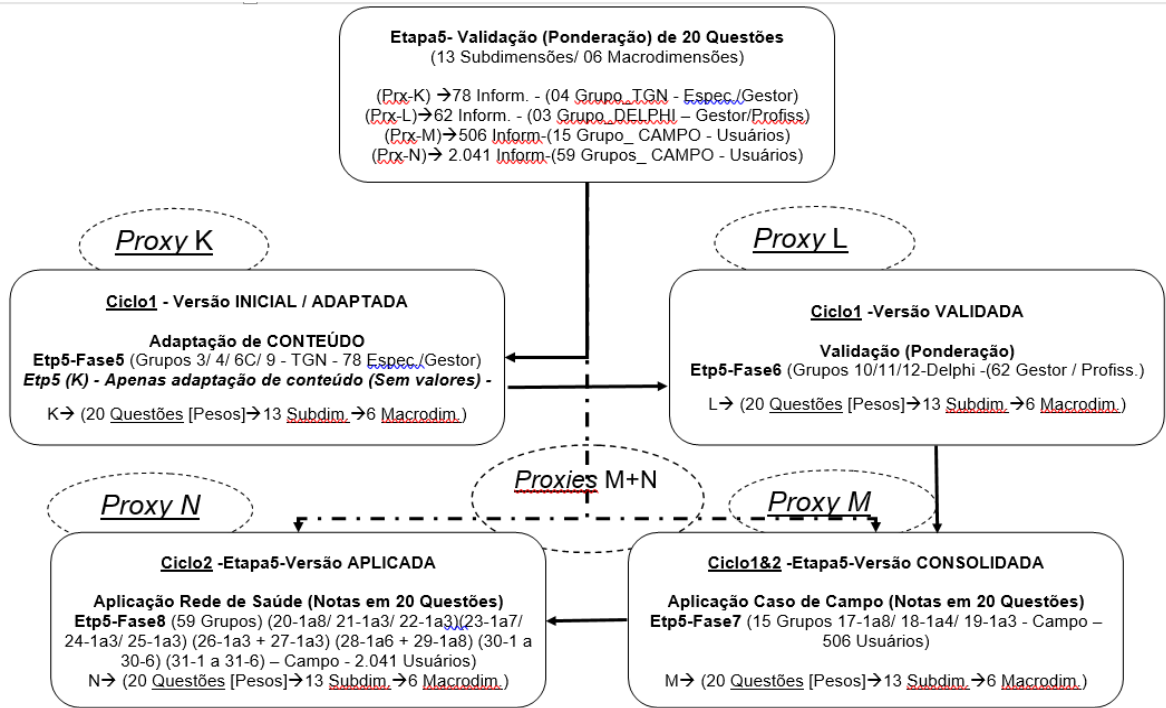


Figura 23 - Fluxo do processo de validação de questões de qualidade percebida (etapa 5 – questões validadas).

Estes fluxos têm suas análises complementadas pelo esquema lógico de análise de macrodimensões, subdimensões e questões validadas de qualidade percebida. Nas duas últimas fases, no âmbito do estudo 5, procedeu-se a análise estatística de dados das *proxies* da valoração das questões validadas além de seus reflexos nas macrodimensões e nas subdimensões da escala qualidade percebida, descritos a seguir:

Análise Cód.	Etapas/Estudo	Fases	Proxy	Informantes (N) / Categorias
A1) Análises de respostas atribuídas através de 4 proxies (pesos atribuídos para as questões validadas e para as questões adaptadas)				
A1.a)-Inicial	Etp5– Ciclo1		K	78 Espec./Gestor (sem notas)
A1.b)-Ajuste	Etp5– Ciclo1		L	62 Espec./Gestor
A1.c)-Campo/ Síntese	Etp5– Ciclo1/2		M	506 Usuários
A1.d) – Campo/ Integração	Etp5– Ciclo2		M+N	2.547 Usuários
A2) Análises de respostas calculadas através de 4 proxies (pesos refletidos para as subdimensões):				
A2.a)-Inicial	Etp5– Ciclo1		K	78 Espec./Gestor (sem notas)
A2.b)-Ajuste	Etp5– Ciclo1		L	62 Espec./Gestor
A1.c)-Campo/ Síntese	Etp5– Ciclo1/2		M	506 Usuários

A1.d) – Campo/ Integração	Etp5– Ciclo2		M+N	2.547 Usuários
A3) Análises de respostas calculadas através de um <i>proxy</i> (pesos refletidos para os macrodomínios e para a Performance Global Percebida)				
A3.a)-Campo/ Síntese	Etp5– Ciclo2		M	506 Usuários
A3.b) – Campo/ Integração	Etp5– Ciclo2		M+N	2.547 Usuários

Quadro 10 - Síntese das análises descritivas dos níveis de valoração (atribuídos) das questões validadas (e refletidos nas subdimensões e macrodimensões da escala de qualidade percebida).

Dessa forma, descreveu-se os fluxos de etapas e fases da pesquisa empírica quanto aos instrumentos da escala de qualidade/satisfação percebida além das atividades, métodos e técnicas de análises aplicadas aos resultados obtidos.

5.3. As hipóteses do estudo (itens de avaliação de performance e de avaliação de qualidade/satisfação percebida)

Inicialmente, propôs-se um diagrama gráfico (*path diagram*), pressupondo algumas relações entre as variáveis latentes (preditas), os itens de avaliação de performance, por exemplo, macrodomínios de avaliação ou dimensões de performance selecionadas e as variáveis observadas (preditoras), as categorias de avaliação, por exemplo, domínios de avaliação ou indicadores de performance (figura 24).

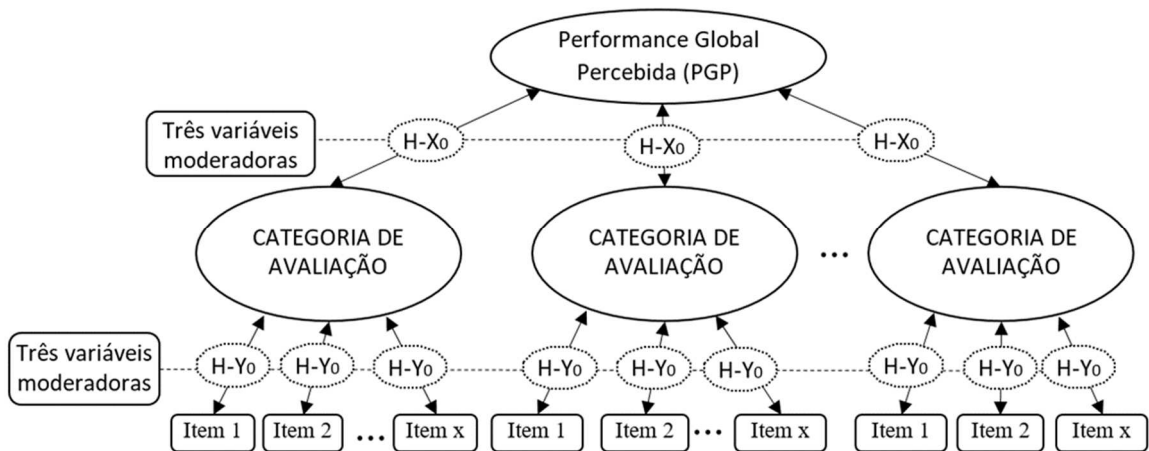


Figura 24 - Modelo gráfico das variáveis latentes e observadas de performance.

Através de recursos da estatística descritiva, correlação quantílica e análise fatorial, foram analisadas as ponderações dos domínios e macrodomínios de avaliação e as mensurações das relevâncias/importâncias das dimensões e indicadores de performance, na perspectiva dos gestores e decisores.

De forma complementar, propôs-se algumas hipóteses sobre variáveis latentes e observadas de performance (figura 24) que foram objeto de análise estatística nesta investigação. Assim, testou-se as hipóteses nulas listadas a seguir, comparativamente às três possíveis variáveis moderadoras selecionadas (nível de atenção/cuidado de saúde, nível de complexidade da assistência e nível de cargo/função dos decisores).

- H1₀ : Os níveis de conhecimentos atribuídos aos domínios de avaliação e seus reflexos sobre os macrodomínios são maioritariamente superiores às notas médias;
- H2₀ : Os níveis de conhecimentos atribuídos aos macrodomínios de avaliação são discriminados pelas três variáveis moderadoras;
- H3₀ : Os níveis de relevâncias/importâncias atribuídos aos indicadores de performance e seus reflexos sobre as dimensões são maioritariamente superiores às notas médias;
- H4₀ : Os níveis de relevâncias/importâncias atribuídas às dimensões são discriminados pelas três variáveis moderadoras;
- H5₀ : Os indicadores de performance apresentam maioritariamente carga fatorial relevante sobre suas respectivas dimensões de performance;
- H6₀ : As dimensões de performance apresentam maioritariamente qualidade e validação comprovadas;
- H7₀ : As dimensões de performance apresentam carga fatorial relevante sobre a Performance Global Percebida;
- H8₀ : Os indicadores de performance analisados apresentam maioritariamente maiores correlações com suas respectivas dimensões de performance;
- H9₀ : A validação e adequação do modelo da Performance Global Percebida foram comprovadas.

Para validar estas hipóteses nulas (H1₀ a H9₀), ou refutá-las em contraposição às suas respectivas hipóteses alternativas, apoiou-se na teoria dominante que emergiu da análise bibliográfica sobre domínios e macrodomínios de avaliação em saúde, sobre indicadores e dimensões de performance em saúde, através das análises de resultados obtidos nesta investigação (tópico 6.1) e, ainda, das análises destas hipóteses propostas, sintetizadas em um quadro de validações e conclusões a partir dos testes de hipóteses aplicados (tópico 6.3).

Em seguida, propôs-se a tradução transcultural e validação da escala de qualidade percebida selecionada (SERVQUAL) efetivadas em dois estudos/etapas consecutivas e incrementais pelos grupos de especialistas: tradução da escala original em constructos, subdimensões e macrodimensões de qualidade percebida, estruturação de um instrumento com questões validadas, com revisões das subdimensões e macrodimensões propostas.

Dessa forma, na mesma lógica, propôs-se um diagrama gráfico (*path diagram*), pressupondo algumas relações entre as variáveis latentes (preditas), as categorias de análise de qualidade, por exemplo, as macrodimensões de qualidade com outras variáveis latentes (preditas), as subcategorias de análise de qualidade, por exemplo, as subdimensões de qualidade e as suas respectivas variáveis observadas (preditoras), os itens de análise de qualidade percebida, por exemplo, os constructos e as questões validadas (figura 25).

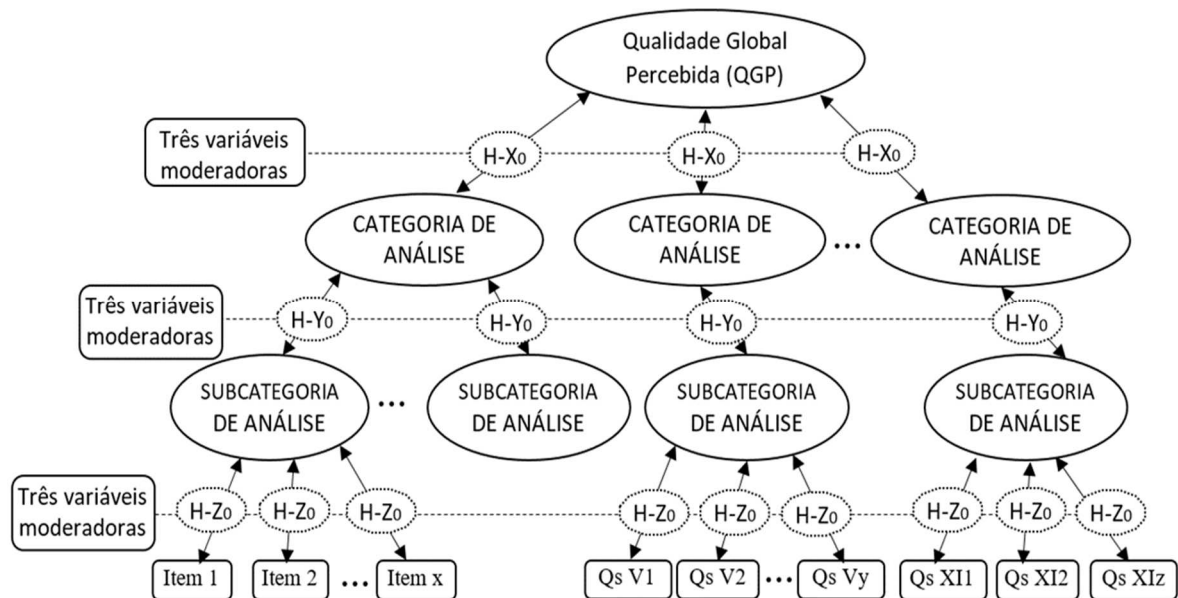


Figura 25 - Modelo gráfico das variáveis latentes e observadas de qualidade percebida.

Através de recursos da estatística descritiva, correlação quantílica e análise fatorial, foram analisadas as ponderações dos constructos e suas subdimensões e macrodimensões de constructos e as mensurações das relevâncias/importâncias das questões validadas e suas subdimensões e macrodimensões de qualidade percebida, na perspectiva dos gestores e decisores e complementarmente de usuários do SUS.

De forma complementar, propôs-se algumas hipóteses sobre variáveis latentes e observadas de qualidade percebida (figura 25) que foram objeto de análise estatística nesta investigação. Assim, testou-se as hipóteses nulas listadas a seguir, comparativamente às três possíveis variáveis moderadoras selecionadas (nível de atenção/cuidado de saúde, nível de

complexidade da assistência e nível de cargo/função dos decisores ou nível de escolaridade do usuário).

- H10₀ : Os níveis de relevâncias/importâncias atribuídos às questões validadas de qualidade percebida e seus reflexos sobre as subdimensões de questões validadas e suas respectivas macrodimensões são majoritariamente superiores às notas médias;
- H11₀ : Os níveis de relevâncias/importâncias atribuídos às subdimensões de questões validadas e suas respectivas macrodimensões são discriminados pelas três variáveis moderadoras;
- H12₀ : Há associação negativa entre as subdimensões de questões validadas e suas respectivas macrodimensões, nas perspectivas Expectativa e Percepção;
- H13₀ : As questões validadas apresentam carga fatorial relevante sobre suas respectivas subdimensões;
- H14₀ : As subdimensões de questões validadas apresentam qualidade e validação comprovadas;
- H15₀ : As subdimensões de questões validadas apresentam carga fatorial relevante sobre suas respectivas macrodimensões;
- H16₀ : As macrodimensões de questões validadas apresentam qualidade e validação comprovadas;
- H17₀ : As questões validadas apresentam maiores correlações com suas respectivas subdimensões;
- H18₀ : As subdimensões de questões validadas apresentam maiores correlações com suas respectivas macrodimensões;
- H19₀ : A validação e adequação do modelo da Qualidade Global Percebida, de questões validadas, foram comprovadas;

Para validar estas hipóteses nulas (H10₀ a H19₀), ou refutá-las em contraposição às suas respectivas hipóteses alternativas, apoiou-se nomeadamente na teoria dominante que emerge da análise bibliográfica sobre escalas de mensuração de qualidade em saúde, a fim de selecionar uma escala de qualidade/satisfação de usuários aplicável ao contexto de saúde brasileira, através das análises de resultados obtidos nesta investigação (tópico 6.2) e, ainda, das análises destas hipóteses propostas, sintetizadas em quadro de validações e conclusões a partir dos testes de hipóteses aplicados (tópico 6.4).

5.4. Os modelos de análise e as técnicas de validação dos itens de avaliação da investigação

A discussão de resultados foi composta principalmente pela etapa de análise dos Artefatos projetados (instrumento com itens de avaliação de performance das unidades de saúde e escala com itens de medição de qualidade/satisfação do usuário), através das provas de conceito, de valor e de utilidade próprias do método da abordagem *Design Research*, bem como avaliar a validade interna dos resultados obtidos, além de analisar a amplitude de possível generalização do instrumento de indicadores e escala de qualidade/satisfação validados.

Neste sentido considerou-se a análise da validação e aceitação pelo decisores envolvidos e, ainda, a disponibilidade de acesso aos dados, bem como aspectos sobre a facilidade de aplicação dos instrumentos propostos, e também sobre outros fatores relevantes que surgiram ao longo da pesquisa da literatura e da estruturação da metodologia proposta.

Após a aplicação dos questionários de recolha de dados nas diversas etapas e fases dos ciclos de validação dos instrumentos (ciclo 1) e nas várias etapas e fases das pesquisas de campo (ciclo 2) estruturou-se uma proposta de modelos e técnicas de análise, como rigor metodológico apropriado, devidamente alinhada com os objetivos e hipóteses traçados no início da investigação, nomeadamente quanto aos resultados de validação dos Artefatos produzidos pelos agentes de decisão, conforme explicitado a seguir. Assim, a análise estatística aplicada aos dados coletados dividiu-se em dois campos: análise descritiva (para todas as etapas e fases, nos ciclos 1 e 2) e análise inferencial (apenas para as fases finais constantes do ciclo 2 de cada estudo).

Neste propósito, apropriou-se de uma técnica de Modelagem de Equação Estrutural (*SEM-Structural Equation Model*), pelo uso das funcionalidades do *Software R*, em busca de explicar estatisticamente as relações entre as variáveis do modelo proposto, através de uma Análise Fatorial Confirmatória (AFC), a fim de validar as magnitudes e significâncias estatísticas da associação entre as variáveis observadas (itens) e latentes (fatores ou constructos), através das estimativas dos parâmetros, a partir dos dados amostrais coletados e da análise de suas variâncias. Também foram usados índices de ajustamento para otimizar o modelo resultante da análise.

Ressalta-se que a análise tipo SEM, geralmente, é aplicada com variáveis dos tipos *intervalar* e de *razão*, porém Lemke (2005) cita os estudos de Atkinson (1988), Bentler e Chou (1987) e Green et al. (1997) para defender que a análise SEM pode ser aplicada também para *variáveis ordinais* (que violam os pressupostos de continuidade e normalidade da variável,

exigidos na análise SEM original), além de que, e muitos estudos, tem sido aplicada em variável do tipo *Escala de Likert*, desde que tenha pelo menos 5 níveis/classes e que a distribuição de frequência da variável analisada se aproxime da normalidade. Ressalte-se que Green et al. (1997) demonstraram que o aumento do número de classes ou categorias de respostas ao item ampliaram as possibilidades de usar escalas intervalares neste tipo de estudo. No caso de não validação do pressuposto de normalidade multivariada das amostras estudadas, sugere-se utilizar o teste de Satorra-Bentler para correção dos testes de qui-quadrado (Marôco, 2014), por exemplo, através do uso do pacote Lavaan do software R.

Após a aplicação da investigação e coleta de dados os resultados obtidos, foram sistematizados e propostos os métodos e técnicas de análises dos dados das amostras. Os resultados obtidos na parte empírica foram sistematizados através de planilhas Excel (2016) sendo a base de análise para a construção de versões evolutivas dos instrumentos de coletas de dados ao longo de cada etapa do estudo e, por fim, sistematizou-se um banco de dados para as análises estatísticas nas quais se utilizou os pacotes do software R versão 3.2.2, de acesso livre.

Quanto aos modelos de análises aplicados sobre os itens de avaliação de performance e suas vinculações aos resultados obtidos desenhou-se os fluxos lógicos de técnicas e atividades usadas, de agregação de resultados e técnicas analíticas aplicadas, conforme seus ciclos de análise (ciclos 1 e 2), etapas de investigação (estudos 1 a 3), além de suas fases sucessivas e incrementais, que envolveram vários grupos e painéis de informantes (quadro 11).

Etapas do Estudo	Técnicas	Revisão bibliográfica e pesquisa documental	Análise Instrumental de Conteúdo (Bardin)	Painel de Especialistas	Pesquisa Delphi	Questionários / inquérito (Surveys)	Análise Multicritério (Análise Fatorial Confirmatória de qualidade e validade do modelo)
	Atividades						
Etapa 1 - Busca de Informação para elaboração de questionários	Estabelecimento de Framework	Relevância e Quadro conceitual de Itens de Avaliação de Performance					
	Externalização de conhecimentos		Categorias de Conhecimento Tácitos sobre avaliação de performance				
	Avaliação de categorias de conhecimentos			Estruturação de Macrodinâmios e Domínios de Avaliação			
	Sistematização de Conhecimentos (Macrodinâmios e Domínios - Ciclo 1)				Validação de Macrodinâmios e Domínios de Avaliação (Proxy A, B e C - Questionário)	Respostas obtidas sobre Macrodinâmios e Domínios (Tópicos 6.1.1.1)	Sem Análise de Modelo Inferencial
Etapa 2 - Criação de Painel de Indicadores (Gestores)	Expansão de conceitos com base na Literatura	Alinhamento e expansão de conceitos com base no Quadro Conceitual					
	Sistematização de Conhecimentos (Dimensões e Indicadores - Ciclo 1)				Validação de Dimensões e Indicadores de Performance - (proxy D, E e F - Questionário)	Respostas obtidas sobre indicadores de performance (Tópicos 6.1.1.2 e 6.1.1.3)	Modelo Inferencial de Dimensões e Indicadores de Performance (Tópicos 6.1.2.1 e 6.1.2.2)
Etapa 3 - Criação de Painel de Indicadores (Decisores)	Expansão de conceitos com base na Literatura	Alinhamento e expansão de conceitos com base no Quadro Conceitual					
	Aplicação de campo do questionário (Ciclo 2)				Validação de Dimensões e Indicadores de Performance - (proxy G - Questionário)	Respostas obtidas sobre indicadores de performance (Tópico 6.1.1.3)	Modelo Inferencial de Dimensões e Indicadores de Performance (Tópico 6.1.2.3)
Etapa 6 - Teste de hipóteses dos Estudos	Análise estatística de resultados (Elaboração de testes de hipóteses)						Hipóteses (Performance) refutadas ou validadas e Correlações/ Associações (Tópico 6.3)
	Análise de aplicabilidade	Relatório de Conclusões sobre os Painéis de Indicadores de Performance e Escala de Qualidade/ Satisfação					

Quadro 11 - Visão global da fase empírica dos estudos/etapas de 1 a 3: atividades, técnicas de investigação e métodos de análises de resultados (macrodinâmios/domínios de avaliação e dimensões/indicadores de performance).

Quanto aos modelos de análises aplicados sobre os itens da escala de mensuração de qualidade/ satisfação percebida e suas vinculações aos resultados obtidos desenhou-se os fluxos lógicos de técnicas e atividades usadas, de agregação de resultados e técnicas analíticas aplicadas, conforme seus ciclos de análise (ciclos 1 e 2), etapas de investigação (estudos 1 a 5), além de suas fases sucessivas e incrementais, que envolveram vários grupos e painéis de informantes (quadro 12).

Etapas do Estudo	Técnicas	Revisão bibliográfica e pesquisa documental	Análise Instrumental de Conteúdo (Bardin)	Painel de Especialistas	Pesquisa Delphi	Questionários / inquérito (Surveys)	Análise Multicritério (Análise Fatorial Confirmatória de qualidade e validade do modelo)
	Atividades						
Etapa 4 - Busca de Informação para validação dos construtos	Estabelecimento de Framework	Relevância e Quadro conceitual de Itens de qualidade / satisfação					
	Externalização de conhecimentos		Categorias de Conhecimento Tácitos sobre percepção de qualidade / satisfação				
	Avaliação de categorias de conhecimentos			Estruturação de Construtos de qualidade / satisfação			
	Sistematização de Conhecimentos (Construtos de Qualidade)				Validação de Construtos, Macro & Subdimensões de Qualidade - (proxy H, I e J - 40 construtos)	Respostas obtidas sobre Construtos, Macro & Subdimensões de Qualidade (Tópicos 6.2.1.1 e 6.2.1.2)	Modelo Inferencial de Construtos, Macro & Subdimensões de Qualidade (Tópico 6.2.2.1)
Etapa 5 - Busca de Informação para elaboração de questões validadas	Externalização de conhecimentos		Expansão de conceitos com base no Quadro Conceitual e Construtos validados				
	Avaliação de categorias de conhecimentos			Estruturação de Questões validadas a partir dos 40 construtos validados			
	Sistematização de Conhecimentos (Construtos de Qualidade)				Validação de Questões, Macro & Subdimensões de Qualidade - (proxy K e L - 20 Questões Validadas)	Respostas obtidas sobre Construtos, Macro & Subdimensões de Qualidade (Tópicos 6.2.1.3)	Sem Análise de Modelo Inferencial
	Aplicação de campo do questionário (Ciclo 2)				Validação de Questões, Macro & Subdimensões de Qualidade - (proxy M e N - 20 Questões Validadas)	Respostas obtidas sobre Questões, Macro & Subdimensões de Qualidade (Tópicos 6.2.1.4)	Modelo Inferencial de Questões, Macro & Subdimensões de Qualidade (Tópico 6.2.2.2)
Etapa 6 - Teste de hipóteses dos Estudos	Análise estatística de resultados (Elaboração de testes de hipóteses)						Hipóteses (Qualidade) refutadas ou validadas e Correlações / Associações (Tópico 6.4)
	Análise de aplicabilidade	Relatório de Conclusões sobre os Painéis de Indicadores de Performance e Escala de Qualidade / Satisfação					

Quadro 12 - Visão global da fase empírica dos estudos/etapas de 4 e 5: atividades, técnicas de investigação e métodos de análises de resultados (macrodimensões/subdimensões de constructos e de questões validadas da qualidade percebida).

A parte empírica desta investigação foi desenvolvida em diversas fases subsequentes de estruturação e aprimoramento incremental dos instrumentos de coleta de dados (ciclo 1) e, por fim, aplicada em uma pesquisa de campo em uma única fase (ciclo 2). A fase empírica foi iniciada pelo mapeamento dos conhecimentos tácitos dos agentes chaves da amostra nas categorias de macrodomínios e domínios de avaliação, seguidos pela etapa de análise e consolidação de painéis de indicadores e dimensões de performance em organizações de saúde.

Para a primeira categoria de análise, em busca de um índice sintético proveniente dos itens de análise de performance, analisou-se os reflexos sobre uma *variável latente* Performance Global Percebida pelo que se mensurou os efeitos refletidos das valorações dos indicadores e dimensões de performance, medidos por suas magnitudes (cargas fatoriais), significâncias estatísticas e suas correlações em um modelo estrutural a ser analisado, especificado e validado. Assim, aplicou-se os métodos de análise estatística descritiva e inferencial para a geração de escores de cada indicador e de cada dimensão de performance a fim de validar a composição de painéis de Controle (*dashboard* ou *tableau de bordo*) de indicadores chaves (*Key Performance Indicators – KPI*). Dessa forma, construiu-se alguns *proxies* do modelo de painéis de indicadores e performance, a partir dos indicadores e dimensões de performance selecionados e dados obtidos da pesquisa de campo com gestores e decisores (figura 26).

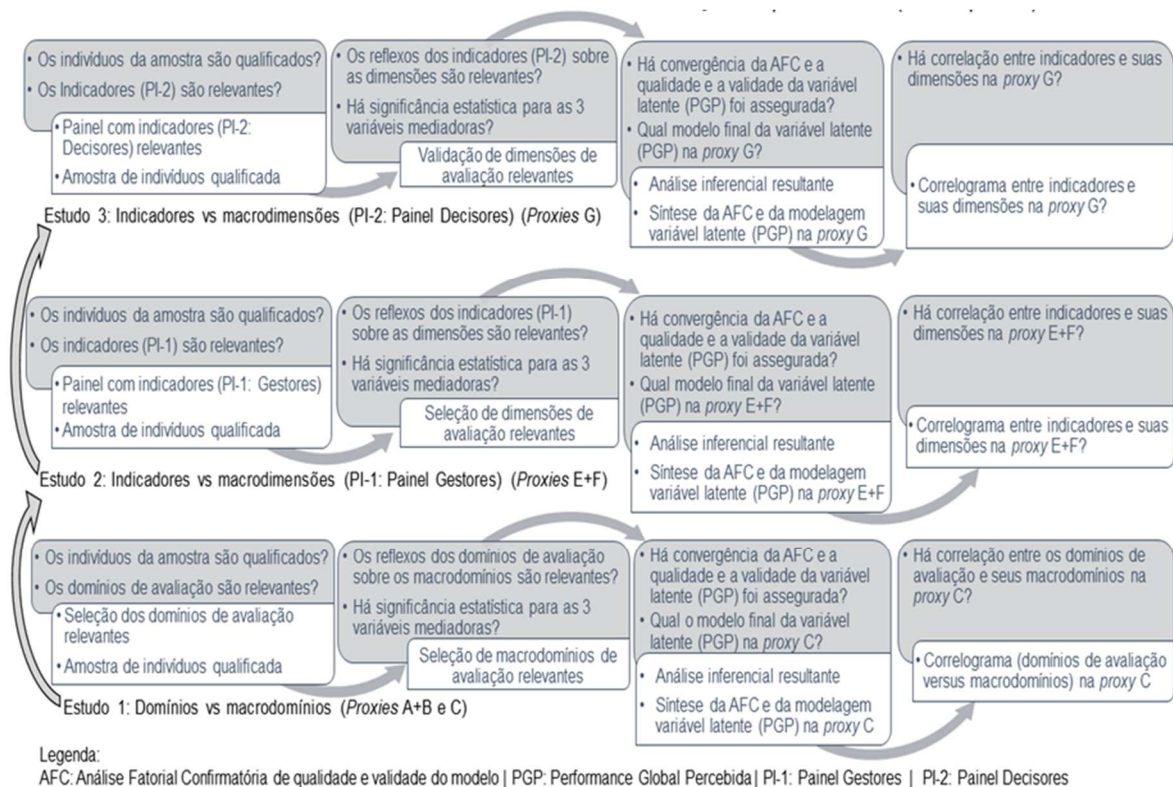


Figura 26 - Síntese de fluxo de análises de itens de análise dos painéis de indicadores e dimensões de performance.

Quanto ao painel de domínios e macrodínios de avaliação, para o conjunto de análises descritivas aplicou-se os seguintes métodos: medidas descritivas e gráficos (barplot e boxplot) para as médias (e intervalos de confiança) dos pesos atribuídos para os domínios e, de forma complementar, para os pesos refletidos dos domínios sobre os macrodínios de avaliação. Aplicou-se ainda as comparações das médias (e intervalos de confiança) entre os

domínios sobre os macrodomínios de avaliação, inclusive quanto às possíveis variáveis moderadoras (nível de atenção, nível de complexidade e nível de cargo/função).

Para o conjunto de análises inferenciais aplicou-se os seguintes métodos: análise das cargas fatoriais e contribuições percentuais de cada um dos domínios sobre os macrodomínios de avaliação e, de forma complementar, dos macrodomínios de avaliação sobre a Performance Global Percebida; verificação dos níveis de correlações de cada um dos domínios sobre os macrodomínios de avaliação; verificação da dimensionalidade, validade convergente e a confiabilidade de cada macrodomínio de avaliação e análise dos critérios de qualidade e validade, além dos índices de qualidade (ajustamento) do Modelo Final.

Por fim, aplicou-se uma análise descritiva quanto a variável latente construída e validada (Performance Global Percebida) a fim de efetivar as análises dos reflexos dos macrodomínios de avaliação sobre a Performance Global Percebida, inclusive as comparações quanto às possíveis variáveis moderadoras (nível de atenção, nível de complexidade e nível de cargo/função). De forma esquemática estas análises foram sintetizadas abaixo (figura 27):

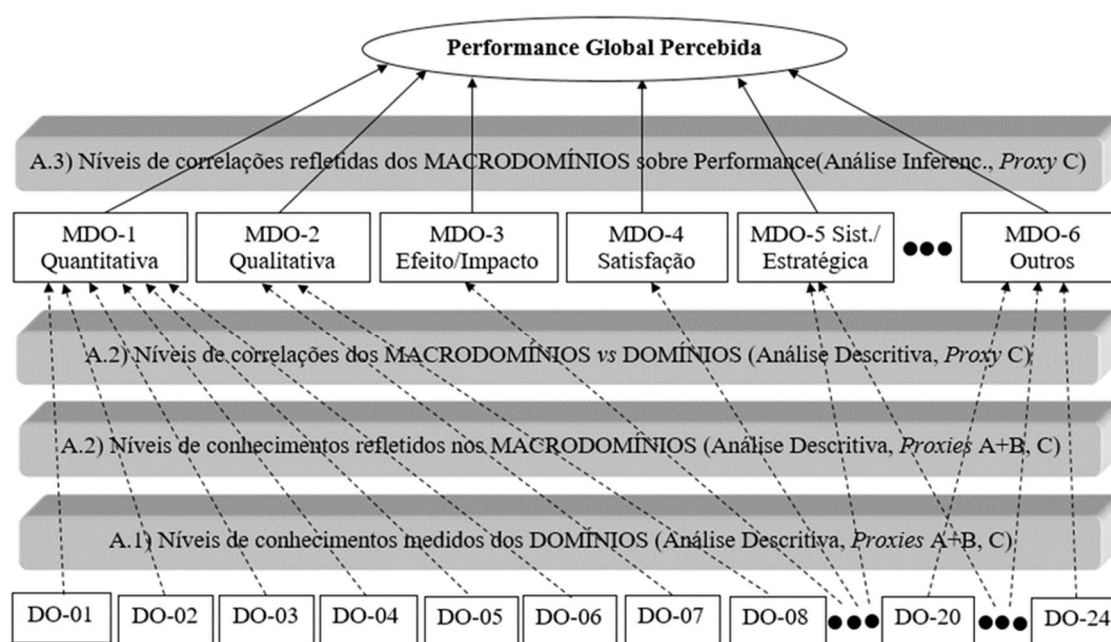


Figura 27 - Esquema lógico de variáveis da investigação e análises de resultados (etapa 1 - macrodomínios e domínios de avaliação).

Quanto ao painel de indicadores e dimensões de performance, para o conjunto de análises descritivas aplicou-se os seguintes métodos: medidas descritivas e gráficos (barplot e boxplot) para as médias (e intervalos de confiança) dos pesos atribuídos para os indicadores e, de forma complementar, para os pesos refletidos dos indicadores sobre as dimensões de performance. Aplicou-se ainda as comparações das médias (e intervalos de confiança) entre os

indicadores sobre as dimensões de performance, inclusive quanto às possíveis variáveis moderadoras (nível de atenção, nível de complexidade e nível de cargo/função).

Para o conjunto de análises inferenciais aplicou-se os seguintes métodos: análise das cargas fatoriais e contribuições percentuais de cada um dos indicadores sobre as dimensões de performance e, de forma complementar, das dimensões de performance sobre a Performance Global Percebida; verificação dos níveis de correlações de cada um dos indicadores sobre as dimensões de performance; verificação da dimensionalidade, validade convergente e a confiabilidade de cada dimensão de performance e análise dos critérios de qualidade e validade, além dos índices de qualidade (ajustamento) do Modelo Final.

Por fim, aplicou-se uma análise descritiva quanto a variável latente construída e validada (Performance Global Percebida) a fim de efetivar as análises dos reflexos das dimensões de performance sobre a Performance Global Percebida, inclusive as comparações quanto às possíveis variáveis moderadoras (nível de atenção, nível de complexidade e nível de cargo/função). De forma esquemática estas análises foram sintetizadas abaixo (figura 28):

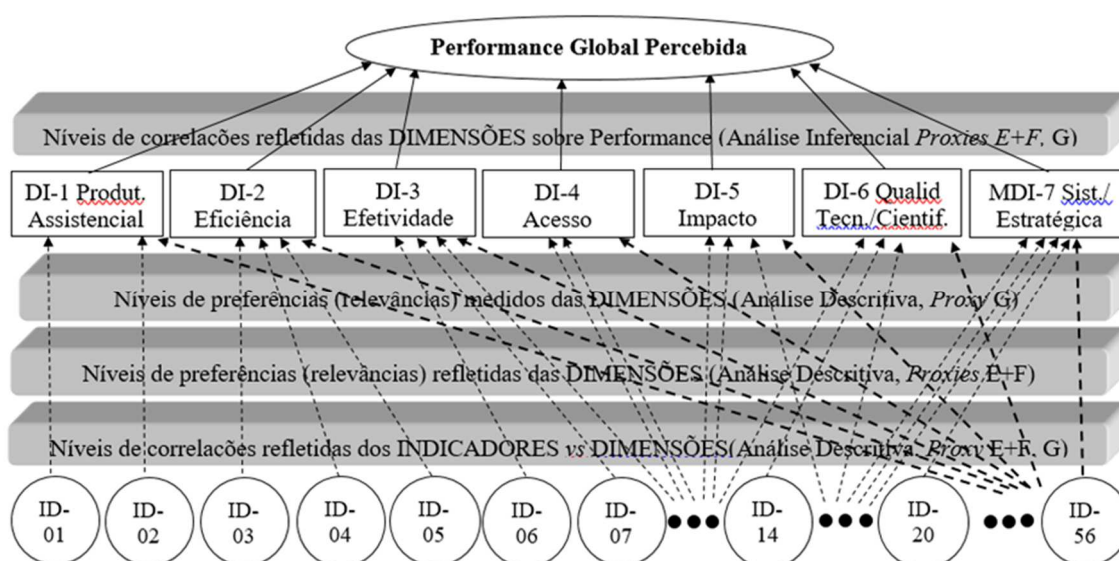


Figura 28 - Esquema lógico de variáveis da investigação e análises de resultados (etapas 2 & 3 - dimensões e indicadores de performance).

As técnicas analíticas para a validação dos itens de avaliação dos painéis de domínios e macrodomínios de avaliação e de indicadores e dimensões de performance obtiveram resultados das análises estatísticas descritivas (tópico 6.1.1) e das análises estatísticas inferenciais (tópico 6.1.2).

A figura 29 sintetizou as etapas, os métodos e as técnicas de análise usados para a validação dos painéis de indicadores e de dimensões de performance.

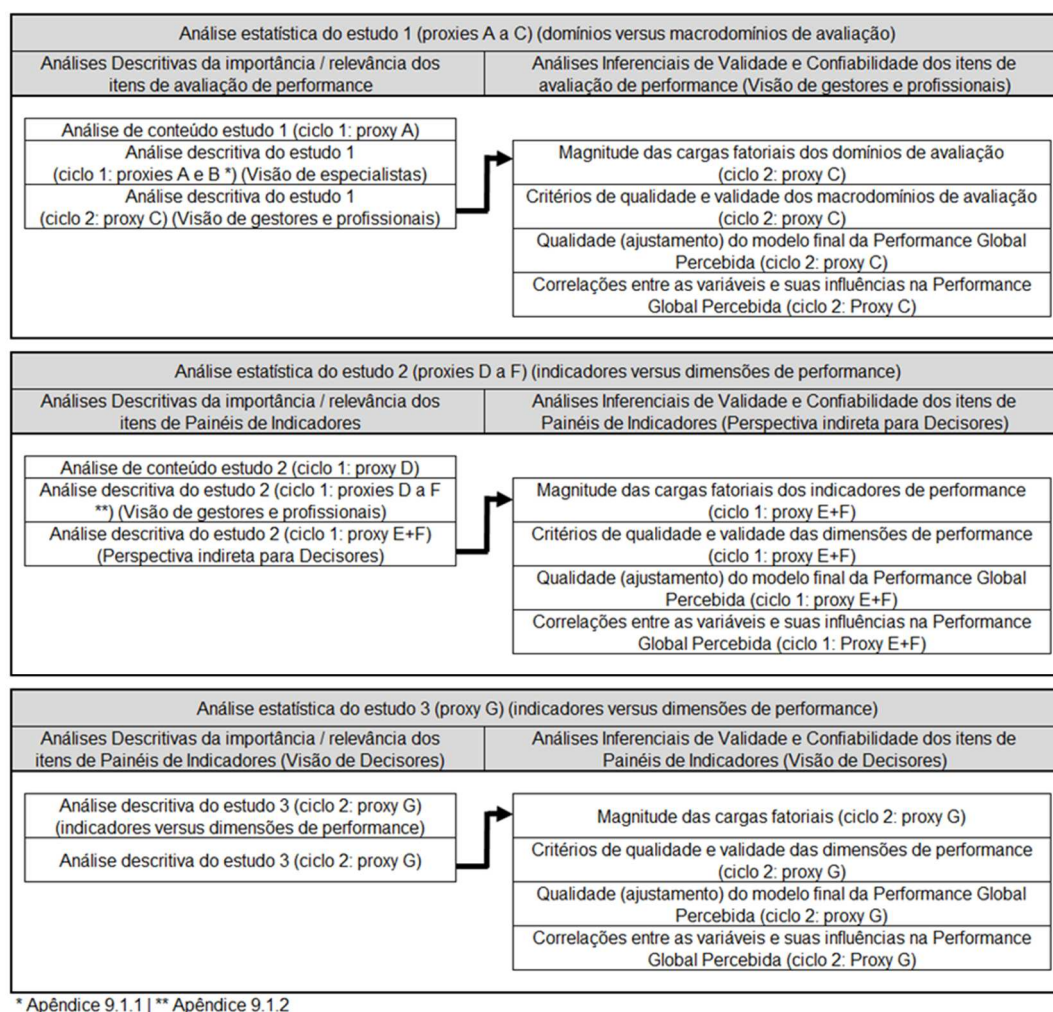


Figura 29 - Síntese dos fluxos, métodos e técnicas de análises de resultados (etapas 1 a 3 - ciclo 1 e 2 - proxies A a G).

Noutro escopo da fase empírica inicial, abordou-se a adequação e validação de itens da escala internacional SERVQUAL de qualidade e satisfação em busca de mapear os constructos de qualidade (ideias-força) relevantes, na visão de agentes chaves para avaliação de qualidade/satisfação em serviços de saúde, posteriormente, consolidados por adaptações em etapas e fases incrementais de validação de questões validadas, além de subdimensões e macrodimensões de qualidade e satisfação para usuários do SUS.

Para a segunda categoria de análise, em busca de um índice sintético proveniente dos itens de avaliação da qualidade percebida, analisou-se os reflexos sobre uma *variável latente* Qualidade Global Percebida pelo que se mensurou os efeitos refletidos das valorações dos itens de análise da escala de qualidade adaptada (constructos e questões validadas) bem como suas subdimensões e macrodimensões de qualidade, medidos por suas magnitudes (cargas fatoriais), significâncias estatísticas e suas correlações em um modelo estrutural a ser analisado,

especificado e validado. Assim, aplicou-se os métodos de análise estatística descritiva e inferencial para a geração de escores de cada item de análise (questões validadas e suas respectivas subdimensões e macrodimensões de qualidade percebida) a fim de validar a composição de uma escala de qualidade e satisfação de usuários consolidada em um questionário aplicado ao contexto de unidades e serviços de saúde do SUS. Dessa forma, construiu-se alguns *proxies* do modelo de escala de questões validadas e suas respectivas subdimensões e macrodimensões de qualidade mais relevantes para compor a proposição de um modelo inovador de escala de medição de qualidade e satisfação de usuários do SUS, a partir dos itens de análises propostos e dados obtidos da pesquisa de campo com gestores, decisores e usuários (figura 30).

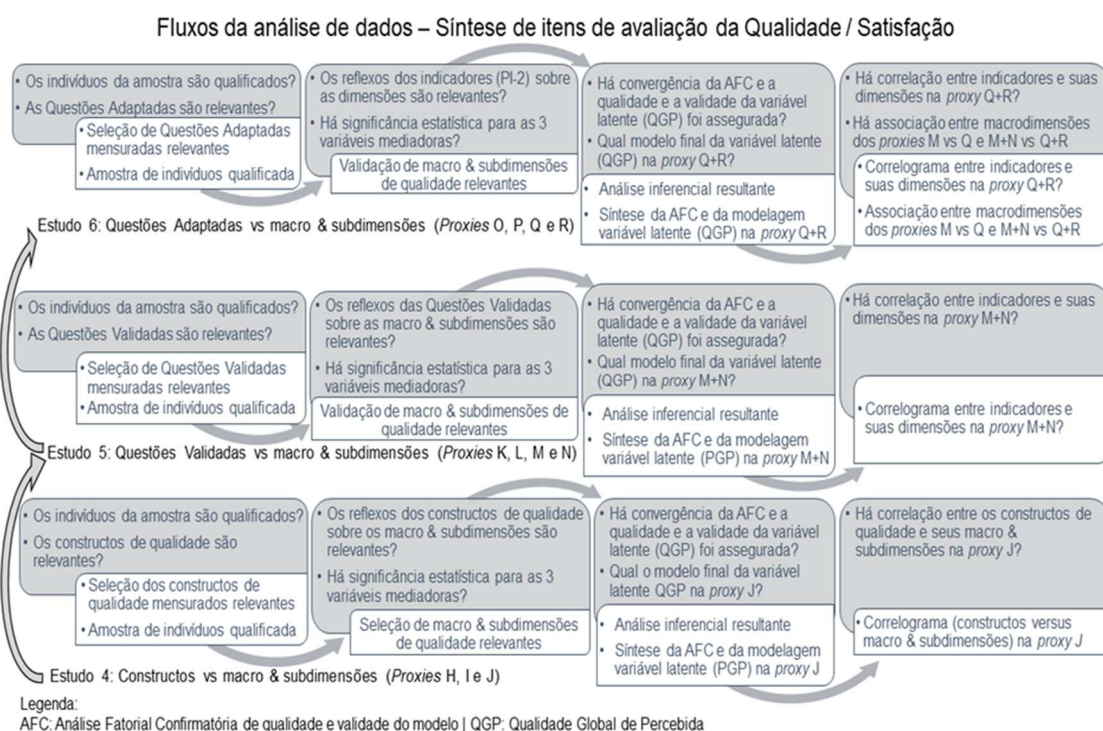


Figura 30 - Síntese de fluxo de análises de itens de avaliação de qualidade/satisfação.

Quanto a escala de constructos de qualidade de constructos, subdimensões e macrodimensões de qualidade, para o conjunto de análises descritivas aplicou-se os seguintes métodos: medidas descritivas e gráficos (barplot e boxplot) para as médias (e intervalos de confiança) dos pesos atribuídos para os constructos de qualidade e, de forma complementar, para os pesos refletidos nas subdimensões e macrodimensões de qualidade. Aplicou-se ainda as comparações das médias (e intervalos de confiança) entre as macrodimensões e subdimensões de qualidade, inclusive quanto às possíveis variáveis moderadoras (nível de

atenção, nível de complexidade e nível de cargo/função, para gestores ou nível de escolaridade, para usuários).

Para o conjunto de análises inferenciais aplicou-se os seguintes métodos: análise das cargas fatoriais e contribuições percentuais de cada uma das macrodimensões sobre as subdimensões de qualidade e, de forma complementar, das macrodimensões de qualidade; verificação dos níveis de correlações de cada um das macrodimensões de qualidade; verificação da dimensionalidade, validade convergente e a confiabilidade de cada macrodimensão de qualidade e análise dos critérios de qualidade e validade, além dos índices de qualidade (ajustamento) do Modelo Final.

Por fim, aplicou-se uma análise descritiva quanto a variável latente construída e validada (Qualidade Global Percebida) a fim de efetivar as análises dos reflexos das macrodimensões de qualidade sobre a Qualidade Global Percebida, inclusive as comparações quanto às possíveis variáveis moderadoras (nível de atenção, nível de complexidade e nível de cargo/função, para gestores ou nível de escolaridade, para usuários). De forma esquemática estas análises foram sintetizadas abaixo (figura 31):

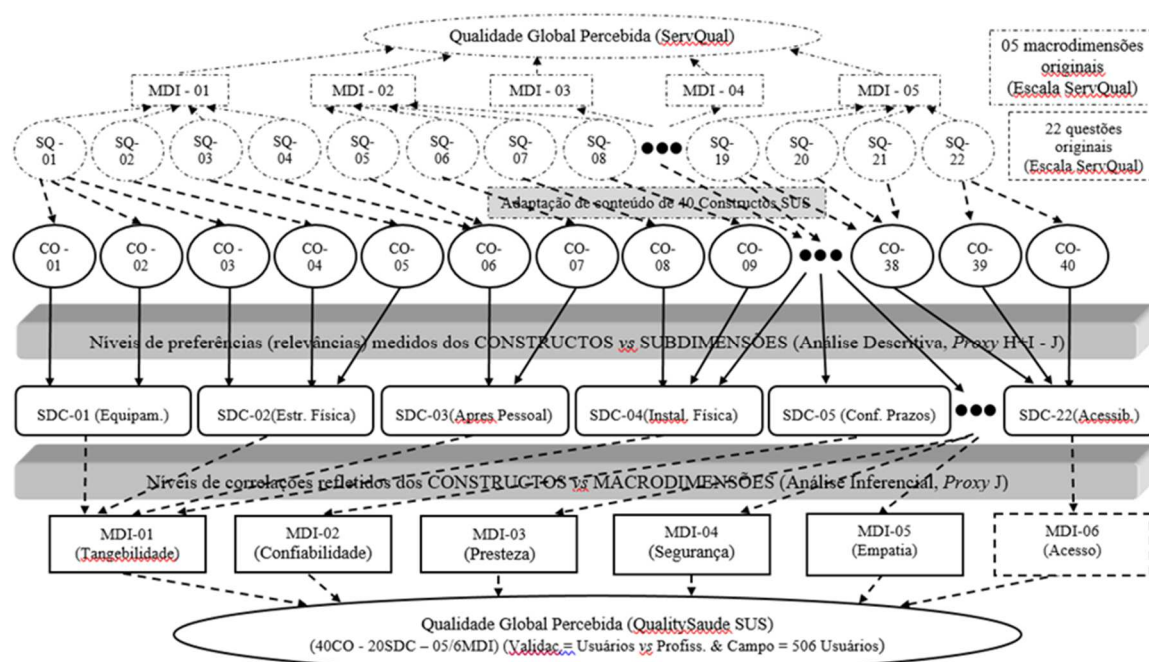


Figura 31 - Esquema lógico de variáveis da investigação e análises de resultados (etapa 4 - macrodimensões, subdimensões e constructos de qualidade percebida).

Quanto a escala de questões de qualidade de questões, subdimensões e macrodimensões de qualidade, para o conjunto de análises descritivas aplicou-se os seguintes métodos: medidas descritivas e gráficos (barplot e boxplot) para as médias (e intervalos de confiança) dos pesos

atribuídos para as questões de qualidade e, de forma complementar, para os pesos refletidos nas subdimensões e macrodimensões de qualidade. Aplicou-se ainda as comparações das médias (e intervalos de confiança) entre as macrodimensões e subdimensões de qualidade, inclusive quanto às possíveis variáveis moderadoras (nível de atenção, nível de complexidade e nível de cargo/função, para gestores ou nível de escolaridade, para usuários).

Para o conjunto de análises inferenciais aplicou-se os seguintes métodos: análise das cargas fatoriais e contribuições percentuais de cada uma das macrodimensões sobre as subdimensões de qualidade e, de forma complementar, das macrodimensões de qualidade; verificação dos níveis de correlações de cada um das macrodimensões de qualidade; verificação da dimensionalidade, validade convergente e a confiabilidade de cada macrodimensão de qualidade e análise dos critérios de qualidade e validade, além dos índices de qualidade (ajustamento) do Modelo Final.

Por fim, aplicou-se uma análise descritiva quanto a variável latente construída e validada (Qualidade Global Percebida) a fim de efetivar as análises dos reflexos das macrodimensões de qualidade sobre a Qualidade Global Percebida, inclusive as comparações quanto às possíveis variáveis moderadoras (nível de atenção, nível de complexidade e nível de cargo/função, para gestores ou nível de escolaridade, para usuários). De forma esquemática estas análises foram sintetizadas abaixo (figura 32):

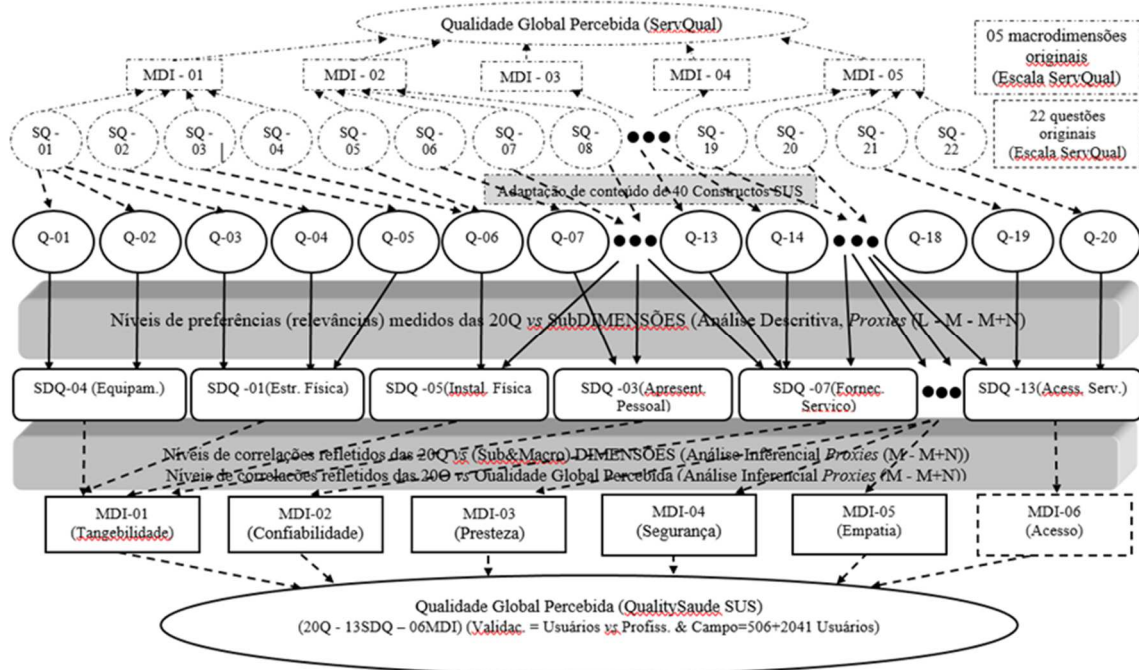


Figura 32 -Esquema lógico de variáveis da investigação e análises de resultados (etapas 5 - macrodimensões, subdimensões e questões validadas de qualidade percebida).

As técnicas analíticas para a validação dos itens de avaliação dos painéis de domínios e macrodomínios de avaliação e de indicadores e dimensões de performance obtiveram resultados das análises estatísticas descritivas (tópico 6.1.1) e das análises estatísticas inferenciais (tópico 6.1.2).

A figura 33 sintetizou as etapas, os métodos e as técnicas de análise usados para a validação dos itens de medição de qualidade/satisfação (escala de qualidade percebida).

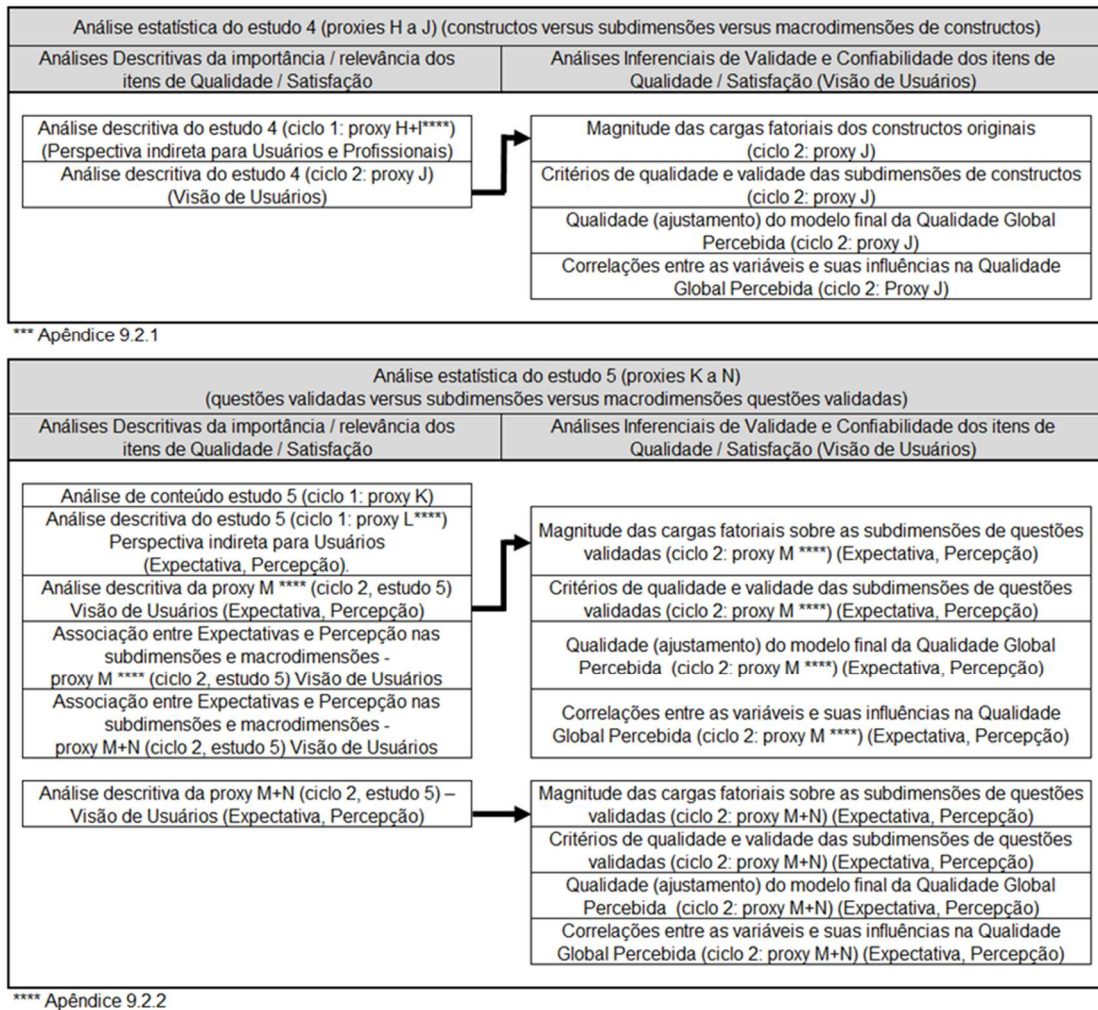


Figura 33 - Síntese dos fluxos, métodos e técnicas de análises de resultados (etapas 4 e 5 - ciclo 1 e 2 - proxies H a N).

Este processo empírico e de análise dos dados obtidos foi dividido em 6 estudos. Na primeira categoria de Artefato quanto aos resultados da validação dos itens de análise dos painéis de mensuração de performance (no estudo 1, domínios e macrodomínios de avaliação e nos estudos 2 e 3 indicadores e dimensões de performance). Na segunda categoria de artefato quanto aos resultados da validação dos itens de análise da escala de qualidade percebida (no

estudo 4, *constructos*, *subdimensões* e *macrodimensões* de qualidade percebida, no estudo 5, *questões validadas*, *subdimensões* e *macrodimensões* de qualidade percebida).

Por sua vez, cada estudo realizado (estudos/etapas 1 a 3 e estudos/etapas 4 e 5) constou de fases subsequentes e incrementais de ajuste dos instrumentos gerados e validados por diferentes grupos de informantes, que gerou diversos *proxies* com itens de análise de performance (A a G) e *proxies* com itens de avaliação de qualidade (H a N) enquanto modelos de aprimoramento incrementais dos instrumentos estruturados e validados pelos agentes de decisão. Por facilidade de análise, agregou-se alguns destes *proxies* para o escopo de itens de análise de performance (A+B, E+F) e também para o escopo de itens de avaliação de qualidade (H+I, M+N), agrupados por similaridades de categorias analíticas, visto que os itens analisados tinham ampla semelhança entre nestes instrumentos agregados em cada fase da pesquisa e seus respectivos *proxies* originais.

As análises estatísticas foram estruturadas a partir dos dados coletados em seis estudos específicos (três diferentes etapas (1 a 3) para a validação do painel de dimensões e indicadores de performance e dois diferentes etapas (4 e 5) para a adaptação e validação transcultural da escala SERVQUAL, através de fases sequenciais de análises (*proxies*), nas quais os modelos de instrumentos foram estruturados, adaptados (aprimoramentos incrementais) e/ou validados pelos entrevistados da amostra.

Nos estudos 1 a 3 abordou-se diferentes perspectivas de análises das respostas obtidas. No estudo 1 diferenciou-se as respostas pela perspectiva indireta das respostas obtidas dos especialistas informantes, através de notas atribuídas para a visão de gestores e na fase de campo apenas para a visão direta de gestores. Enquanto isso, no estudo 2 diferenciou-se as respostas pela perspectiva indireta das respostas obtidas dos gestores e profissionais informantes, através de notas atribuídas para a visão de gestores e na fase de campo apenas para a visão direta de gestores. No estudo 3, analisou-se a resposta de campo apenas para a visão direta de gestores quanto à perspectiva de Decisores da saúde.

Nos estudos 4 e 5 abordou-se diferentes perspectivas de análises das respostas obtidas. No estudo 4 diferenciou-se as respostas pela perspectiva indireta das respostas obtidas dos especialistas e gestores informantes, através de notas atribuídas para a visão de Usuários e para a visão dos Profissionais. Enquanto isso, no estudo 5 diferenciou-se as respostas pela perspectiva direta dos Usuários informantes quanto a observação da Expectativa (relativo ao atendimento recebido nos últimos 12 meses) e da percepção (relativo ao atendimento recebido naquele momento), sobre a qualidade/ satisfação percebida pelo Usuário do SUS.

Todas as análises descritivas (estudos 1 a 5, ciclo 1 - validação de instrumentos e ciclo 2 - aplicação de campo) foram iniciadas pela caracterização da amostra dos respondentes e os posteriores cálculos das valorações médias atribuída por grupos de entrevistados a cada uma das variáveis investigadas em cada estudo. De forma complementar, através de métodos de análises descritivas, foram estudadas as influências de três possíveis fatores mediadores nas respostas obtidas de especialistas, profissionais e gestores (nível de atenção, nível de complexidade e nível de cargo/função) e para respostas obtidas de Usuários (nível de atenção, nível de complexidade e nível de escolaridade).

Na análise descritiva das variáveis de caracterização selecionadas foram utilizadas as frequências absolutas e relativas. Três destas variáveis relevantes (nível de atenção, nível de complexidade e nível de cargo/função, para gestores ou nível de escolaridade, para usuários) foram analisadas quanto às possíveis variáveis moderadoras das respostas obtidas para os itens em análise, nos estudos 1 a 5. Já na descrição dos itens em análise (macrodomínios e domínios de avaliação do estudo 1; dimensões e indicadores de performance dos estudos 2 e 3; constructos, subdimensões e macrodimensões de qualidade percebida do estudo 4; questões validadas, subdimensões e macrodimensões de questões validadas do estudo 5) foram utilizadas medidas de posição, tendência central e dispersão, sendo uma das medidas utilizadas o intervalo percentílico bootstrap com 95% de confiança. O método *bootstrap* (Efron & Tibshirani, 1993) é muito utilizado na realização de inferências quando não se conhece a distribuição de probabilidade da variável de interesse. Cabe ressaltar que todos os itens analisados estavam dispostos numa de escala ordinal, discreta e finita de *Likert* de importância (relevância) com níveis de 1 (Nada importante) a 5 (Extremamente importante).

Para análise da valoração média tomou-se a média por indivíduo, das valorações atribuídas. Por definição, o conjunto de dados não apresenta distribuição normal, uma vez que todas as variáveis analisadas são limitadas em uma escala ordinal, discreta e finita (*Likert* de 1 a 5). Dessa forma, para comparar os itens de análise em relação as variáveis de caracterização foram utilizados o Teste de *Kruskal-Wallis* e o Teste de *Mann-Whitney* (Hollander & Wolfe, 1999). Além disso, quando o teste de *Kruskal-Wallis* evidenciou diferença significativa entre os níveis das variáveis de caracterização, foi utilizado o teste de *Nemenyi* (Hollander & Wolfe, 1999) para as comparações múltiplas.

Quanto a seleção destes testes para utilização na análise de todas variáveis dos estudos, foi por serem robustos a *outliers* e desvios de normalidade, tendo as seguintes características de uso:

- *Teste U deMann-Whitney*: teste não paramétrico para comparação das médias dos postos de duas amostras independentes. Correspondente ao teste da razão “t” para verificação de igualdade entre as médias de duas populações independentes. Utiliza “distribuição livre”.

- *ANOVA deKruskal-Wallis*: teste não paramétrico para comparação das médias dos postos de três ou mais amostras independentes. Correspondente à ANOVA (*one-way*) paramétrica para verificação de igualdade entre as médias de três ou mais populações independentes. Utiliza “distribuição livre”.

Em seguida, aplicou-se métodos de análise inferencial (estudos 2 a 5, ciclo 2 - aplicação de campo) a fim de verificar a validade das hipóteses propostas e consolidar os modelos propostos desenhados nesta investigação, em quatro categorias de análises inferenciais: estudo das cargas fatoriais entre variáveis; validação por critérios de qualidade e validade; análise da qualidade (ajustamento) do modelo final resultante e análise das correlações entre as variáveis e suas influências na variável latente (Performance Global, estudos 1 a 3 e Qualidade Global, estudos 4 e 5).

Para a análise inferencial das respostas obtidas em cada estudo desta investigação foram feitas algumas Análises Fatoriais Confirmatórias (AFC) (Hair, William, Babin, & Anderson, 2009). A Análise Fatorial Confirmatória (AFC) pressupõe que as variáveis latentes apresentaram distribuição normal. Por definição, as variáveis em estudo não apresentaram distribuição normal, uma vez que estavam dispostas em uma escala ordinal, discreta e finita (*Likert* de 1 a 5). Logo, foram utilizados estimadores robustos à desvios de normalidade para estrutura de covariância na Análise Fatorial Confirmatória (AFC) com a estatística de teste reescalada pelo método de Satorra e Bentler (1994).

Na análise do modelo de mensuração, são verificadas a validade convergente, a validade discriminante e a confiabilidade dos construtos analisados em cada estudo. A validade convergente garante que os indicadores de um construto estão correlacionados o suficiente para medir o conceito latente. A validade discriminante verifica se os construtos medem efetivamente diferentes aspectos do fenômeno de interesse. A confiabilidade revela a consistência das medidas usadas para mensurar o conceito pretendido. A fim de testar a validade convergente dos construtos, foi utilizado o critério proposto por Fornell e Larcker (1981) que propuseram que pelo menos 50% da variabilidade em cada item em análise deveria ser explicada pelo fator subjacente, Ele garante tal validade, caso a Variância Média Extraída (AVE), que indica o percentual médio de variância compartilhada entre o construto latente e seus indicadores e varia de 0% a 100% (Hair et al., 2009), seja superior a 50% (Henseler, Ringle, & Sinkovics, 2009) ou a 40%, no caso de pesquisas exploratórias ($AVE > 0,40$)

(Nunnally & Bernstein, 1994). Para validade discriminante, foi utilizado novamente o critério de Fornell e Larcker (1981), que garantem a validade discriminante quando a variância média extraída (AVE) de um constructo não for menor que a variância compartilhada desse constructo com os demais. Para mensurar a confiabilidade dos constructos, foi utilizado o Alfa de Cronbach (A.C.) e a Confiabilidade Composta (C.C.). De acordo com Tenenhaus et al. (2005), os índices A.C. e C.C. devem ser maiores que 0,70 para uma indicação de confiabilidade do constructo ou maiores que 0,60 no caso de pesquisas exploratórias. Para verificar a dimensionalidade foi utilizado o critério de autovalor (também chamado de *eigenvalue*) maior do que 1,0 ou critério de Guttman-Kaiser (Guttman, 1954; Kaiser, 1960 apud Yeomans & Golder, 1982). O autovalor corresponde à quantidade da variância explicada por um componente, sendo que um autovalor igual a 1,0 representa a totalidade de porcentagem da variância explicada por uma única variável. A soma da quantidade de autovalores corresponde ao número de variáveis analisadas. O critério Guttman-Kaiser é baseado na consideração de que um fator precisa explicar pelo menos a quantidade de variância que é explicada por uma única variável. Para um bom modelo de mensuração, esperam-se cargas fatoriais acima de 0,70 ou comunalidades acima de 0,40, porém itens com cargas fatoriais menores que 0,50 devem ser eliminados (Hair et al., 2009), pois, ao não contribuir de forma relevante para a formação da variável latente, prejudicam o alcance das suposições básicas para validade e qualidade dos indicadores criados para representar o conceito de interesse.

Há uma diversidade de parâmetros e índices de ajustamento e qualidade dos modelos de equações estruturais (Bollen & Long, 1993). Decidiu-se por usar um grupo de índices amplamente aplicados na literatura relevante ($X^2/G.L.$ – teste de qui-quadrado por grau de liberdade, CFI – índice de ajustamento comparativo, TLI – índice de ajustamento não normalizado ou índice Tucker-Lewis e o índice RMSEA – erro quadrático médio de aproximação). Estes parâmetros selecionados de qualidade (ajustamento) do modelo preceituam adequação quando o valor-p e o RMSEA forem estatisticamente menor que 0,050. Além disso, espera-se que $X^2/G.L.$ seja menor que 3 (Arbuckle, 2008; Wheaton, 1987), CFI seja maior que 0,80 (Bentler, 1990), TLI seja maior que 0,80 (Bentler & Bonnet, 1980) e que o RMSEA (Steiger, Shapiro, & Browne, 1985) seja menor que 0,10, sendo o ideal abaixo de 0,05.

Para avaliar as correlações entre os componentes de cada estudo foi utilizada a matriz de Correlação de *Pearson* e o correlograma.

No estudo 3, que teve as sete dimensões da Performance Global Percebida mensuradas, utilizou-se modelos de Regressão Linear Univariados (MRL) para avaliar os fatores de influência (itens mensurados) na valoração média atribuída a cada uma das dimensões e

posteriormente à Performance Global Percebida, pois todas as variáveis explicativas (itens mensurados, possíveis fatores de influência) foram muito correlacionados entre si, que não permitiu as inclusões conjuntas destas variáveis em um modelo de regressão multivariado, tendo-se decidido manter o modelo de regressão univariado (MRL) e apresentar apenas o percentual (%) de explicação individual de cada item mensurado sobre a variável latente estudada. Da mesma forma, no estudo 5 em que a Qualidade Global Percebida foi medida, utilizou-se modelos de Regressão Linear Univariados (MRL) para avaliar os fatores de influência (itens mensurados) na valoração média atribuída a cada uma das dimensões da Qualidade Global Percebida. Em estudos com modelos univariados (MRL) não há possibilidade de retiradas de variáveis explicativas visto que as análises envolvem apenas uma variável por vez.

Para o propósito de validação de resultados obtidos e de análise das hipóteses propostas, a estatística de teste encontra-se associada a uma probabilidade de significância (*sig.* ou *p-value*), que se compara diretamente com um nível de significância (alfa - α) estipulado. A probabilidade de significância (*sig.* ou *p-value*) é a probabilidade de obter o valor específico de estatística teste, caso se verifique a hipótese nula, ou seja, caso não exista efeito no modelo (igualdade). Deste modo, se *sig.* (*p-value*) for inferior ao valor (alfa - α) estipulado, neste caso, estipulou-se a probabilidade de significância (*sig.* ou *p-value*) igual a 0,05, rejeita-se a hipótese nula e não rejeitamos a hipótese alternativa, caso contrário não existe evidência estatística para rejeitar a hipótese nula (Martinez & Ferreira, 2008). Ao nível de probabilidade dos resultados destes estudos, estipulou-se um intervalo de confiança (IC) de 95% para os testes aplicados, portanto, pode-se afirmar que se tem 95% de probabilidade de que os resultados ou valores obtidos retratem a realidade, dentro da margem de erro estipulada, neste caso 2%.

O software utilizado nas análises foi o R (versão 3.3.2) por ser livre de licença e contemplar todas as análises aplicadas.

6. RESULTADOS E DISCUSSÃO

A seguir apresenta-se detalhes dos processos e resultados obtidos nas análises dos dados coletados nos diversos *proxies* (A a G, para estudos 1 a 3 e H a R, para estudos 4 e 5) enquanto modelos de aprimoramento incrementais dos instrumentos estruturados e validados pelos agentes de decisão para a produção de um relatório com os resultados comentados das análises alinhadas com os objetivos e hipóteses da investigação, tendo como prioridades a descrição analíticas dos resultados obtidos e dos métodos de análise aplicados. Assim, apresenta-se o detalhamento de cada etapa de aplicação da investigação com os fluxos, métodos e técnicas utilizadas nas análises de resultados de cada etapa, fase e grupos de respondentes abordados conforme as características dos dados coletados e sistematizados.

Diante do exposto, aplicou-se as análises de regressão quantílica e análise fatorial nos dados obtidos para identificar os indicadores e dimensões de performance mais relevantes para compor a proposição de modelos de painéis de indicadores de performance multidimensional e multicritério (tópico 6.1). Da mesma forma, aplicou-se as análises de regressão quantílica e análise fatorial nos dados obtidos para identificar os constructos (ideais-forças) de qualidade, questões validadas e questões adaptadas qualidade, além de suas subdimensões e macrodimensões de qualidade percebida mais relevantes para compor a proposição de um modelo inovador de escala de medição de qualidade e satisfação de usuários do SUS (tópico 6.2). Em seguida, efetivou-se os respectivos testes das hipóteses iniciais propostas (tópico 3), com desdobramentos de hipóteses desagregadas conforme as prioridades identificadas no processo de validação dos gestores, decisores e usuários, além de articular sobre a possibilidade de generalização de resultados desta investigação (tópico 6.3).

6.1. Os resultados dos painéis de indicadores e de dimensões de performance

Os resultados foram agregados em subcategorias (domínios de avaliação e indicadores de Performance) e categoriais (*macrodomínios* de avaliação e dimensões de performance) a fim de construir diferentes painéis de análise da performance, através de um processo empírico colaborativo de painéis de especialistas e grupos pilotos de gestores e de campo de decisores que livremente expressaram seus conhecimentos prévios e atribuíram importância/relevância para a tomada de decisão de cada um dos itens apresentados a partir de conceitos globais e genéricos apresentados, não sendo objeto desta investigação discutir conceitos teóricos, mas a

título de suporte de análise quando necessário foram usados conceitos propostos anteriormente caracterizados (tópicos 2.3 e 2.4.3), tendo sido estruturados modelos inovadores de mensuração de performance com 24 domínios e seis macrodomínios de avaliação (*proxy* C), outro com 20 indicadores e sete dimensões de performance (*proxy* E+F) além de outro com 56 indicadores e sete dimensões de performance (*proxy* G), sintetizados no quadro 13 abaixo:

<u>Macrodomínios</u>	<u>Definições conceituais</u>
I. Quantitativos	Seis <u>domínios</u> referentes a aspectos de produtividade, estrutura assistencial, cobertura e de finanças/economia (D1A_Produtiv, D1B_EstrutAssist; D1C_CobertAssit, D1D_OtimizCusto, D1E_EficienEcon e D1F_SustEconFin).
II. Qualitativos	Sete <u>domínios</u> referentes a aspectos de qualidade, humanização e integralidade assistencial, acessibilidade e entrega de serviços e equidade (D2H_QualidAssist, D2I_RiscoAssist, D2J_AtendHumaniz, D2K_AcessibilServ, D2L_AcessoServ, D2M_IntegralAtenc e D2N_Equidade).
III. Efeitos	Cinco <u>domínios</u> referentes a aspectos de eficácia, efetividade, eficiência, resolutividade e impacto (D3P_EficaciaProj, D3Q_EfetivClinica, D3R_EficienAssist, D3S_ResolubAssist e D3T_Impacto).
IV. Satisfação	Dois <u>domínios</u> referentes a aspectos de qualidade e satisfação (D4U_SatisfUsuario e D4V_SatisfProfiss).
V. Sistêmico/ Estratégico	Cinco <u>domínios</u> referentes a aspectos de gestão estratégica do sistema de saúde (D5W_IndICSAP, D5X_AtendVincReg, D5Y_AtendDesVincReg e D5V_ExtrapTetoOrç).
<u>Dimensões</u>	<u>Definições conceituais</u>
1. Produtividade Assistencial	Dois <u>indicadores</u> na <i>proxy</i> E+F (ID01 a ID02) e 14 indicadores na <i>proxy</i> G (IDD01 a IDD14).
2. Eficiência	Três <u>indicadores</u> na <i>proxy</i> E+F (ID03 a ID05) e seis indicadores na <i>proxy</i> G (IDD15 a IDD20).
3. Efetividade	Três <u>indicadores</u> na <i>proxy</i> E+F (ID06 a ID08) e sete indicadores na <i>proxy</i> G (IDD21 a IDD27).
4. Acesso	Dois <u>indicadores</u> na <i>proxy</i> E+F (ID09 a ID10) e sete indicadores na <i>proxy</i> G (IDD28 a IDD34).
5. Impacto	Dois <u>indicadores</u> na <i>proxy</i> E+F (ID11 a ID12) e 13 indicadores na <i>proxy</i> G (IDD35 a IDD47).
6. Qualidade Técnica/ Científica	Dois <u>indicadores</u> na <i>proxy</i> E+F (ID13 a ID14) e dois indicadores na <i>proxy</i> G (IDD48 a IDD49).
7. Sistêmico/ Estratégico	Seis <u>indicadores</u> na <i>proxy</i> E+F (ID15 a ID21) e sete indicadores na <i>proxy</i> G (IDD50 a IDD56).

Quadro 13 - Definições conceituais das macrodomínios e dimensões de performance.

6.1.1. Análises descritivas dos estudos 1 a 3 (*proxies* A a G)

Os resultados das análises descritivas dos domínios de avaliação integrada das *proxies* A e B (108 informantes), no ciclo 1 de validação do instrumento, além da *proxy* C (112

informantes) de aplicação de campo (ciclo 2), conforme métodos de estatística descritiva detalhados no início do tópico 6, constam do subtópico 6.1.1.1 e de validação dos indicadores e dimensões de performance das *proxies* D a F, com especificidades para a *proxy* E+F (220 informantes) constam do subtópico 6.1.1.2., e para a *proxy* G (233 informantes) constam do subtópico 6.1.1.3.

6.1.1.1. Análise descritiva do estudo 1 (ciclos 1 e 2: *proxies* A a C) (domínios *versus* macrodomínios de avaliação) (Visão de especialistas, gestores e profissionais)

Inicialmente, abordou-se uma análise integrada de fases incrementais (fase 1 a 3, *proxies* A e B) que se iniciou pela validação de face (fase 1) e aplicou dois instrumentos diferentes (fases 2 e 3), através de uma amostra de 108 informantes especialistas, gestores e profissionais de saúde, assim distribuídos: *proxy* A - 53 informantes, dividido em duas fases, *proxy* B - 55 informantes. Efetivou-se uma análise integrada das respostas obtidas nestes dois instrumentos (*proxy* A+B) pois apresentaram muitas similaridades (cinco macrodomínios e 20 domínios de avaliação). Em fase posterior (fase 4, *proxy* C), abordou-se a análise da *proxy* C - 112 informantes (seis macrodomínios e 24 domínios de avaliação).

A figura 34 abaixo apresenta uma análise descritiva das variáveis de caracterização selecionadas na amostra, na análise integrada das *proxies* A e B (108 informantes) que aplicou dois instrumentos com muitas similaridades, no ciclo 1 de validação do instrumento, comparativamente com a *proxy* C (112 informantes), no ciclo 2 de aplicação de campo, na visão de gestores e decisores.

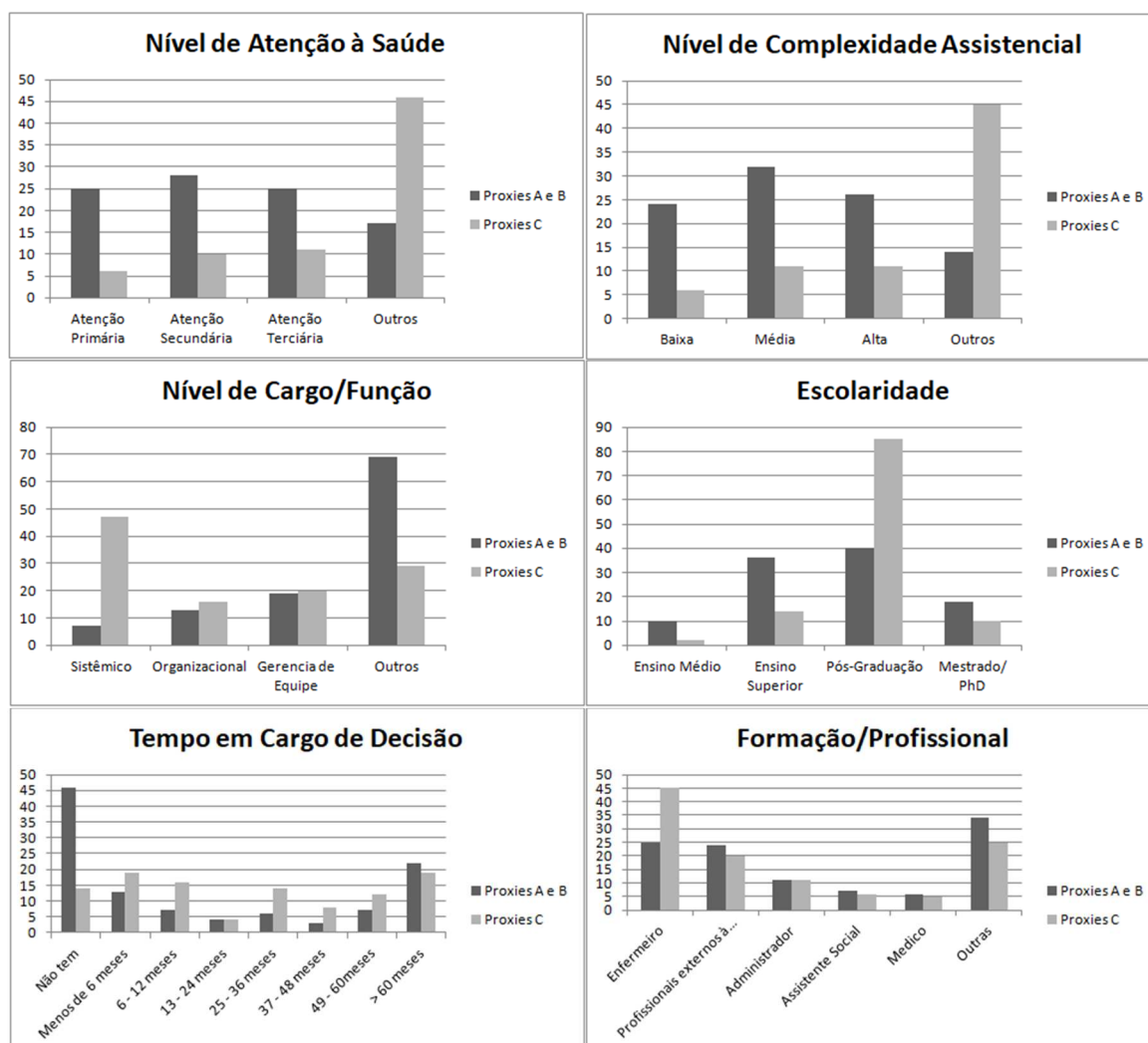


Figura 34 - Gráfico de barras das variáveis de caracterização selecionadas para estudo de fatores mediadores – Proxy A+B (ciclo 1) e proxy C (ciclo 2) (estudo 1) (Visão de especialistas).

Nas fases 1 a 3 (*proxy A+B*), destaca-se que a amostra se mostrou balanceada (figura 34) no nível de atenção à saúde que os indivíduos trabalhavam: níveis Secundário (29,47%), Primário (26,32%) e Terciário (26,32%) de atenção à saúde. Grande parte dos indivíduos (63,89%) exercia o nível de cargo/função Outros (por exemplo, assessores de gestão, profissionais em geral etc). Considerou-se a amostra (n=108) razoavelmente qualificada visto que a maior parte dos indivíduos (38,46%) possuía pós-graduação e ainda 34,62% possuía nível superior, a maioria dos indivíduos (49,07%) possuía mais de 60 meses de experiência e que uma boa parte dos indivíduos (20,37%) trabalhava a mais de 60 meses em cargo/função de decisão enquanto apenas 42,59% não possuía experiência em cargo/função de decisão. Na fase 4 (*proxy C*), destaca-se que a maior parte dos indivíduos da amostra (63,01%) trabalhava no nível de atenção à saúde Outros (estruturas não assistenciais, por exemplo, apenas de gestão ou governança) e que a maior parte dos indivíduos na amostra (61,64%) trabalhava no nível de

complexidade Outros (estruturas não assistenciais, por exemplo, apenas de gestão ou governança). Considerou-se esta amostra (n = 112) mais qualificada pois a ampla maioria dos indivíduos (76,58%) possuía pós-graduação e ainda 12,61% possuía nível superior; a ampla maioria dos indivíduos (70,27%) possuía experiência maior que 60 meses enquanto a parte que trabalhava em cargo/função de decisão acima de 6 meses somou 69,87% dos indivíduos da amostra.

A Tabela 1 apresenta uma análise descritiva relativa à valoração média atribuída aos domínios de avaliação, na análise integrada das *proxies* A e B (108 informantes), no ciclo 1 de validação do instrumento, comparativamente com a *proxy* C (112 informantes), no ciclo 2 de aplicação de campo, na visão de especialistas, profissionais e gestores.

Ressalte-se que os respondentes acrescentaram quatro novos domínios de avaliação (*D5W_IndICSAP*, *D5X_AtendVincReg*, *D5Y_AtendDesVincReg*, *D5V_ExtrapTetoOrç*), adicionados em um novo macrodomínio de avaliação (Sistêmico/Estratégico), que não tiveram respostas individualizadas nas análises iniciais (fases 1 a 3), mas foram incorporados nos instrumentos (*proxy* C) da nova fase de investigação (fase 4).

Tabela 1 - Análise descritiva dos domínios de avaliação - *Proxy* A+B (ciclo 1) e *proxy* C (ciclo 2) (estudo 1) (Visão de especialistas).

Macrodo mí-nios	Domínios	<i>Proxy</i> A+B (fases 1 a 3)				<i>Proxy</i> C (fase 4)			
		N	Médi a	D.P.	I.C-95%	N	Média	D.P.	I.C-95% ¹
Quantitati- -vos	D1A_Produtiv	108	3,85	0,90	[3,69; 4,03]	111	4,28	0,86	[4,12; 4,42]
	D1B EstrutAssist	107	4,00	0,90	[3,84; 4,16]	111	4,40	0,79	[4,25; 4,54]
	D1C CobertAssit	95	4,23	0,80	[4,06; 4,38]	102	4,36	0,81	[4,20; 4,51]
	D1D OtimizCust	108	4,23	0,80	[4,07; 4,36]	111	4,14	0,82	[3,99; 4,28]
	D1E EficienEcon	108	4,30	0,90	[4,15; 4,45]	111	4,27	0,82	[4,13; 4,42]
	D1F SustEconFin	81	4,24	0,80	[4,06; 4,40]	111	4,15	0,88	[4,00; 4,31]
	D1G Outros	11	4,18	0,87	[3,64; 4,64]	6	4,50	0,84	[3,83; 5,00]
Qualitati- -vos	D2H QualidAssis	108	4,46	0,79	[4,30; 4,61]	111	4,38	0,78	[4,23; 4,51]
	D2I RiscoAssist	107	4,38	0,75	[4,24; 4,52]	110	4,40	0,68	[4,27; 4,53]
	D2J AtendHuman	108	4,58	0,60	[4,46; 4,70]	109	4,35	0,77	[4,20; 4,49]
	D2K AcessibilSer	88	4,25	0,73	[4,10; 4,41]	112	4,12	0,81	[3,97; 4,27]
	D2L AcessoServ	106	4,36	0,72	[4,21; 4,50]	112	4,26	0,83	[4,10; 4,42]
	D2M IntegralAte	107	4,24	0,80	[4,09; 4,39]	110	4,04	0,92	[3,85; 4,20]
	D2N Equidade	96	4,20	0,85	[4,02; 4,35]	55	3,84	1,05	[3,56; 4,11]
D2O_Outros	6	3,83	0,98	[3,17; 4,50]	3	4,33	0,58	[4,00; 5,00]	
Efeitos	D3P_EficaciaProj	106	3,90	0,85	[3,73; 4,08]	112	3,89	1,00	[3,71; 4,08]
	D3Q EfetivClinic	107	4,33	0,70	[4,20; 4,45]	112	4,48	0,74	[4,35; 4,61]
	D3R EficienAssis	105	4,36	0,70	[4,24; 4,50]	112	4,29	0,81	[4,13; 4,44]

	D3S ResolubAssi	94	4,48	0,60	[4,35; 4,61]	110	4,50	0,75	[4,35; 4,64]
	D3T_Impacto	94	4,30	0,80	[4,14; 4,45]	111	4,38	0,74	[4,24; 4,52]
Satisfação	D4U SatisfUsuari	95	4,54	0,70	[4,40; 4,67]	111	4,26	0,84	[4,11; 4,41]
	D4V_SatisfProfiss	37	3,78	1,60	[3,24; 4,24]	112	4,19	0,89	[4,02; 4,35]
	D5W_IndICSAP					111	3,87	1,14	[3,67; 4,07]
Sistêmico / Estraté- gico	D5X AtendVincR					112	3,76	0,87	[3,59; 3,92]
	D5Y AtendDesVincReg					111	3,26	1,01	[3,08; 3,45]
	D5V_ExtrapTeto Orç					111	3,92	1,00	[3,73; 4,11]
Outros	D6X Outros	5	4,40	0,90	[3,60; 5,00]	1	4,00	-	[1,50; 3,75]
	D6Z_Outros	1	4,00	-	-	1	5,00	-	[1,80; 4,20]

Nota.¹ Intervalo Bootstrap

A análise global permite afirmar que a ampla maioria dos 20 domínios de avaliação nas fases 1 a 3, e dos 24 domínios de avaliação na fase 4, obtiveram valoração média elevada (superior ou cerca de 4, muito importante/relevante), sendo considerados relevantes/importantes para a estruturação de novos instrumentos com domínios e macrodomínios de avaliação devidamente alinhados pela literatura relevante de dimensões e indicadores de performance em fases posteriores desta investigação.

Por sua vez, a análise dos macrodomínios de avaliação, estruturou-se a partir da valoração média atribuída aos domínios de avaliação pelos indivíduos e, em seguida, calculadas para os macrodomínios de avaliação.

A Tabela 2 apresenta uma análise descritiva relativa à valoração média obtida para os macrodomínios de avaliação, na análise integrada das *proxies* A e B (108 informantes), no ciclo 1 de validação do instrumento, comparativamente com a *proxy* C (112 informantes), no ciclo 2 de aplicação de campo, na visão de especialistas, profissionais e gestores.

Tabela 2 - Análise descritiva dos macrodomínios de avaliação - *Proxy* A+B (ciclo 1) e *proxy* C (ciclo 2) (estudo 1) (Visão de especialistas).

Macrodomínios	<i>Proxy</i> A+B (fases 1 a 3)				<i>Proxy</i> C (fase 4)			
	N	Média	D.P.	I.C-95% ¹	N	Média	D.P.	I.C-95% ¹
Quantitativos	108	4,12	0,54	[4,01; 4,22]	111	4,27	0,53	[4,17; 4,37]
Qualitativos	108	4,36	0,47	[4,27; 4,44]	112	4,23	0,55	[4,13; 4,33]
Efeitos	107	4,27	0,50	[4,18; 4,36]	112	4,31	0,58	[4,20; 4,41]
Satisfação	96	4,40	0,83	[4,23; 4,56]	112	4,22	0,77	[4,09; 4,37]
Outros	5	4,40	0,89	[3,60; 5,00]	112	3,70	0,74	[3,56; 3,83]
					1	4,50	-	[4,50; 4,50]

Nota.¹ Intervalo Bootstrap

A análise global permite afirmar que todos os quatro macrodomínios de avaliação apresentaram valoração média elevada (superior ou cerca de 4, muito importante/relevante),

nas *proxies* A e B (n=108) e na *proxy* C (n = 112), além do quinto macrodomínio (Sistêmico/Estratégico) adicional, composto por agregações de quatro novos domínios de avaliação (D5W_IndICSAP, D5X_AtendVincReg, D5Y_AtendDesVincReg, D5V_ExtrapTetoOrç).

As análises de significância estatística quanto as possíveis influências de fatores mediadores das notas obtidas pelos macrodomínios constam do apêndice 9 (tópico 9.1.1.1). Em síntese, a variável nível de atenção à saúde, pode ser considerada como fator mediador apenas dos macrodomínios de avaliação Quantitativos e Efeitos e que as variáveis nível de cargo/função e nível de complexidade da assistência não puderam ser consideradas como fator moderador de nenhum dos cinco macrodomínios de avaliação (Quantitativos, Qualitativos, Efeitos, Satisfação e Sistêmico/Estratégico), na *proxy* C (fase 4).

A figura 35, a seguir, mostra o grau de importância comparada entre os domínios e os macrodomínios atribuídos nas fases 1 a 3 (*proxy* A+B) e fase 4 (*proxy* C).

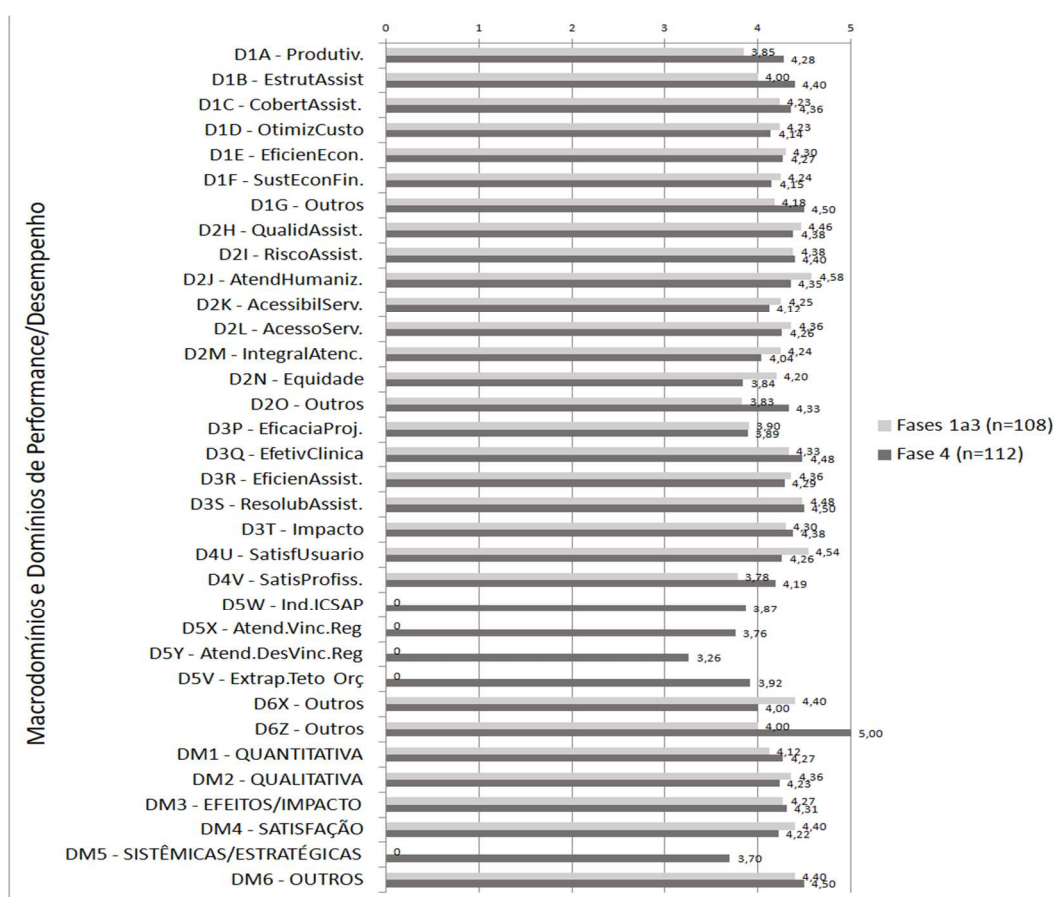


Figura 35 - Grau de importância/relevância de domínios e macrodomínios de avaliação (*proxy* A+B versus *proxy* C) (Visão de especialistas).

Desta análise descritiva sintetizou-se que os especialistas e profissionais atribuíram elevados graus de importâncias/relevâncias das subcategorias (20 e 24 domínios de avaliação)

e das categorias (cinco macrodomínios de avaliação) dos itens de análise, sendo considerados relevantes/importantes para a estruturação de um instrumento de avaliação de performance. Além disso, em média, as valorações da ampla maioria destes itens de análise foram razoavelmente dentro das mesmas margens de valores, uma vez que a maioria dos intervalos de confiança se sobrepuseram.

Quanto aos resultados do estudo das possíveis influências de três variáveis moderadoras, das análises das tabelas 55 e 56 (apêndice 9, tópico 9.1.1.1), sintetizou-se que, respectivamente, na *proxy* A+B e *proxy* C, não houve tendência de mediação de valorações médias obtidas na ampla maioria dos macrodomínios de avaliação quanto aos níveis de atenção à saúde, de complexidade da assistência e de cargo/função dos especialistas e profissionais, com raras exceções.

Estes resultados subsidiaram a construção de um novo instrumento, devidamente alinhado pela literatura relevante de indicadores e dimensões de performance, aplicado na fase posterior da investigação (*proxy* D, estudo 2).

6.1.1.2. Análise descritiva do estudo 2 (ciclo 1: *proxies* D a F) (indicadores *versus* dimensões de performance) (Visão de profissionais e gestores)

O segundo estudo (etapa 2) da investigação foi desenvolvido a partir dos conhecimentos e modelos previamente validados no estudo 1, através de uma reorganização das respostas obtidas e validadas no estudo inicial (domínios e macrodomínios de avaliação), devidamente reagrupados e alinhados com a literatura relevante das dimensões e indicadores de performance.

Inicialmente, na *proxy* D (fase 1), aplicou-se uma análise dos conhecimentos genéricos de 15 especialistas, enquanto validação de face dos conteúdos propostos e consolidados nesta fase. Os respondentes, além de atribuir valoração nos itens de avaliação instrumento da *proxy* D (6 dimensões, 14 indicadores), propuseram uma importante ampliação incremental de seis novos indicadores e a adição de uma macrodimensão de performance (Sistêmico/Estratégico) para estruturar as *proxies* E e F (7 dimensões, 20 indicadores).

Na transição do instrumento inicial da *proxy* D para o instrumento aprimorado da *proxy* E+F, observou-se que não houve eliminações de dimensões e de indicadores de performance e que grande parte dos seis novos indicadores tinham sido classificados na dimensão Outros ou tinham nomenclaturas similares em diferentes respostas de informantes, nas respostas da *proxy* D. Assim, efetivou-se um reagrupamento e realinhamento inicial com a literatura relevante das dimensões e indicadores de performance e preferiu-se fazer uma análise integrada das respostas

obtidas em uma amostra de 112 profissionais e gestores de saúde, a fim de consolidar os níveis de conhecimentos específicos iniciais sobre as dimensões e indicadores de performance, visto que estas respostas têm maior nível de comparabilidade pelas similaridades dos dois instrumentos (*proxy* E+F).

Em seguida, no ciclo 1 de validação do instrumento, abordou-se uma análise integrada de duas fases incrementais (fases 2 e 3, *proxies* E e F) que aplicou dois instrumentos com ajustes, mas com muitas similaridades (7 dimensões, 20 indicadores), na visão de profissionais e gestores, através de uma amostra de 112 informantes, assim distribuídos: *proxy* E - 31 informantes e *proxy* F - 81 informantes.

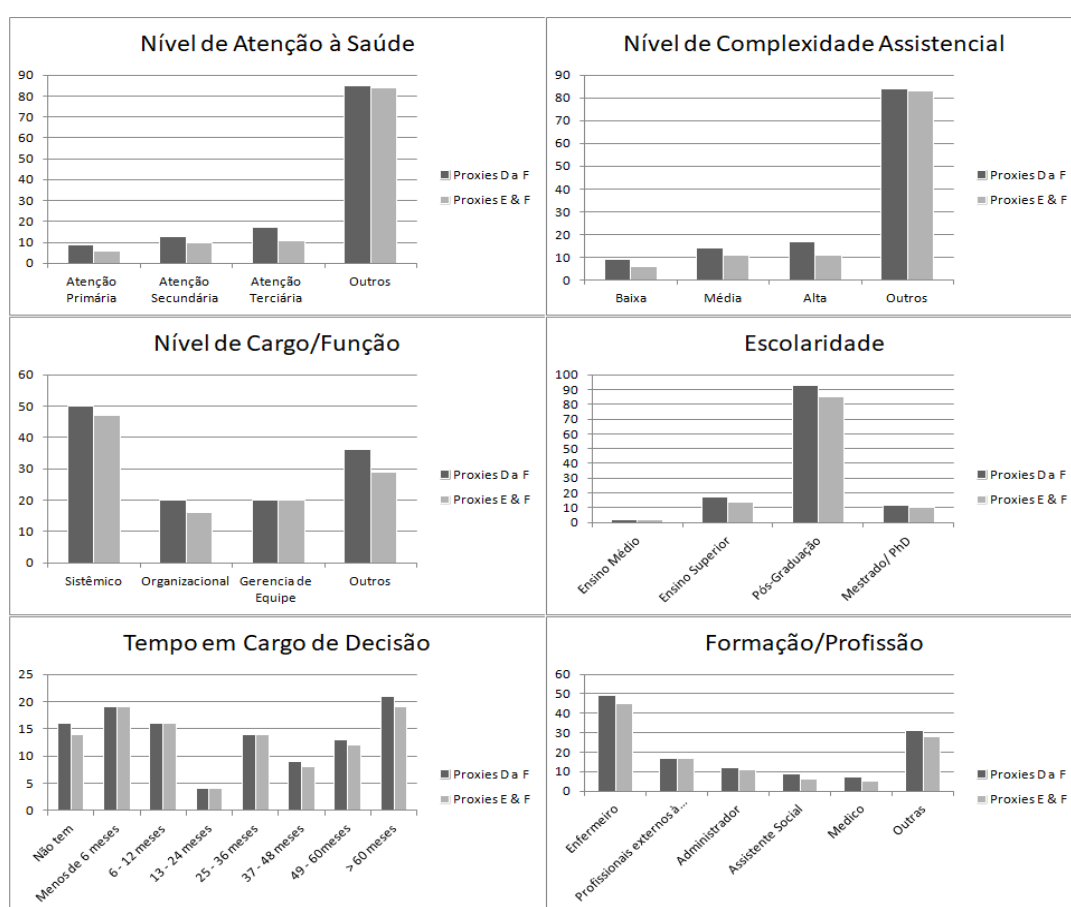


Figura 36 - Gráfico de barras das variáveis de caracterização selecionadas para estudo de fatores mediadores – *Proxy* D a F e *proxy* E+F (ciclo 1, estudo 2) (Visão de especialistas, profissionais e gestores).

A figura 36 acima apresenta uma análise descritiva das variáveis de caracterização selecionadas na amostra, na análise integrada das *proxies* D a F (127 informantes), comparativamente com a *proxy* E+F (112 informantes), na visão de especialistas, profissionais e gestores.

A partir deste ponto, preferiu-se apresentar apenas a análise integrada das respostas obtidas na *proxy* E+F, em virtude do painel inicial de 15 especialistas ter validado o conteúdo de indicadores e dimensões de performance (*proxy* D), mas ampliado o escopo de conteúdos dos instrumentos, com muitas similaridades nas *proxies* E e F.

Nas fases 2 e 3 (*proxy* E+F) a amostra mostrou-se bastante relevante (Gráfico 7). A maior parte dos indivíduos trabalhava no nível de atenção à saúde (75,68%) e nível de complexidade (74,77%) da categoria Outros (estruturas não assistenciais, por exemplo, apenas de gestão ou governança). A maioria dos indivíduos (41,96%) exercia nível de cargo/função Sistêmico/Estratégico, seguidos de boa parte (25,89%) que exercia nível de cargo/função Outros (por exemplo, assessores de gestão, profissionais em geral etc). A ampla maioria dos indivíduos (76,58%) possuía pós-graduação e ainda 12,61% possuía nível superior. A ampla maioria dos indivíduos (70,27%) possuía experiência maior que 60 meses. Somente 13,21% não possuía experiência em cargo/função de decisão.

A Tabela 3 apresenta uma análise descritiva relativa à valoração média atribuída aos indicadores de performance, na *proxy* E+F (112 informantes), no ciclo 1 de validação do instrumento, na visão de profissionais e gestores de saúde.

Tabela 3 - Análise descritiva dos indicadores de performance – *Proxy* E+F (ciclo 1, estudo 2) (Perspectiva indireta para Decisores).

Dimensões	Indicadores	N	Média	D.P.	I.C-95% ¹
Produtividade Assistencial	ID01_TaxaMedPerm	110	3,36	0,70	[3,24; 3,47]
	ID02_TaxaMedProced	111	4,77	0,57	[4,65; 4,87]
Eficiência	ID03_TaxaRotatLeitos	111	4,07	0,42	[3,99; 4,14]
	ID04_CustoSetAtend	78	3,92	0,83	[3,73; 4,10]
	ID05_NoEnfer.leitos	74	3,88	0,99	[3,65; 4,11]
Efetividade	ID06_PropCirurgRescind	80	4,16	0,91	[3,95; 4,35]
	ID07_TaxaReintern	80	4,11	0,83	[3,94; 4,30]
	ID08_TaxaInfecCirurg	80	4,51	0,89	[4,31; 4,68]
Acesso	ID09_TaxaPacienEspec	80	4,33	0,82	[4,14; 4,50]
	ID10_PropPacienEmerg	79	4,49	0,71	[4,33; 4,65]
Impacto	ID11_TaxaMortInst	80	4,73	0,60	[4,59; 4,84]
	ID12_PropNascVivos	78	4,28	0,82	[4,10; 4,46]
Qualidade Técnica/Científica	ID13_PropMedMest	82	3,37	0,98	[3,15; 3,59]
	ID14_PropProfMest	81	3,26	0,91	[3,06; 3,46]
Sistêmico/Estratégico	ID15_IndiceSatisfUsua	83	4,43	0,75	[4,28; 4,58]
	ID16_IndiceSatisfProf	81	4,03	0,87	[3,84; 4,21]
	ID17_IndiceInternICSAP	80	4,14	0,78	[3,95; 4,30]
	ID18_IndiceAtendUsuaComVinc	80	3,80	0,83	[3,63; 4,00]
	ID20_IndiceAtendUsuaSemVinc	78	3,44	0,85	[3,26; 3,63]
	ID21_IndiceExtrapTetosOrçProj	79	4,01	0,84	[3,82; 4,19]
Outros	ID22_Outros	21	4,63	0,57	[4,35; 4,85]

Nota.¹ Intervalo Bootstrap. ² O Indicador ID19 não teve resultados apurados, pois foi desagregado em ID18 e ID20, ao longo da coleta de dados.

A análise global permite afirmar que todos os 20 indicadores de performance obtiveram valoração média superior ao nível 3 (importante/relevante), na análise da *proxy* E+F, sendo considerados para a estruturação de um instrumento final de investigação com dimensões e indicadores de performance, devidamente reagrupados e ampliados (realinhados) pela literatura relevante, limitados pelas bases de dados nacionais do SUS (*proxy* G, ciclo 2 do estudo 3).

Por sua vez, a análise das dimensões de performance, estruturou-se a partir da valoração média atribuída aos indicadores de performance pelos indivíduos e calculadas para as dimensões.

A Tabela 4 apresenta uma análise descritiva das valorações médias obtidas para as dimensões de performance na *proxy* E+F (112 informantes), na visão de profissionais e de saúde.

Tabela 4- Análise descritiva das dimensões de performance – *Proxy* E+F (ciclo 1, estudo 2) (Perspectiva indireta para Decisores).

Dimensões	N	Média	D.P.	I.C-95% ¹
Produtividade Assistencial	111	4,06	0,42	[3,99; 4,14]
Eficiência	111	3,98	0,46	[3,89; 4,07]
Efetividade	82	4,26	0,75	[4,10; 4,43]
Acesso	83	4,42	0,70	[4,26; 4,57]
Impacto	81	4,52	0,59	[4,39; 4,64]
Qualidade Técnica/Científica	82	3,32	0,91	[3,13; 3,52]
Sistêmico/Estratégico	83	3,99	0,52	[3,88; 4,10]
Outros	21	4,63	0,57	[4,38; 4,84]

Nota.¹ Intervalo Bootstrap

A análise global permite afirmar que todas as sete dimensões de performance obtiveram valoração média superior ao nível 3 (importante/relevante), sendo consideradas no instrumento aplicado na fase posterior da investigação (*proxy* G, estudo 3).

A figura 37, a seguir, mostra o grau de importância comparada entre os indicadores e as dimensões de performance atribuídos nas fases 2 e 3 (*proxy* E+F).

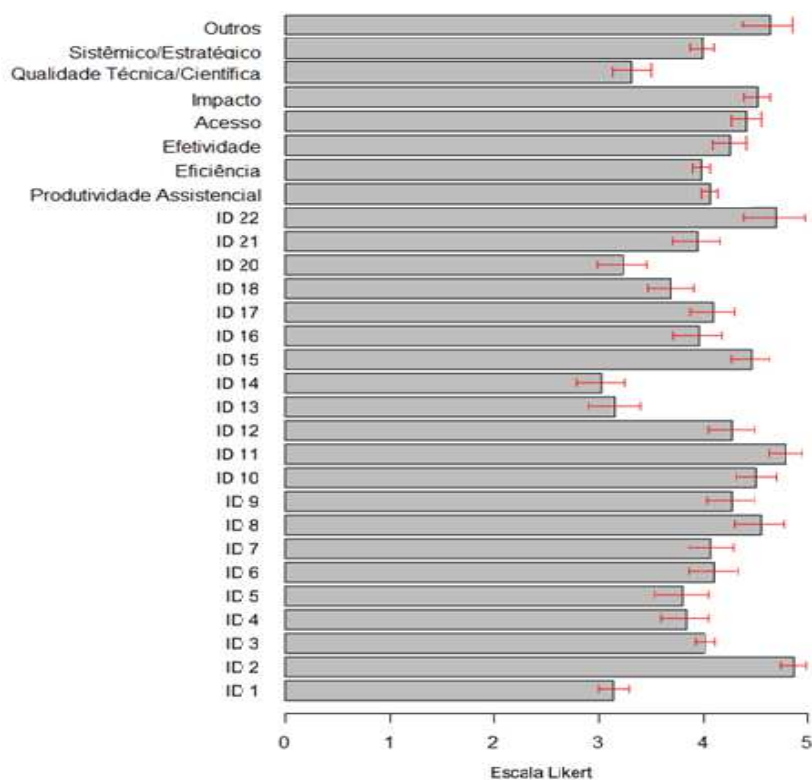


Figura 37- Gráfico de barras com intervalos de confiança para os indicadores e as dimensões de performance - Proxy E+F (ciclo 1, estudo 2) (Perspectiva indireta para Decisores).

Na ampla maioria dos 20 indicadores e das sete dimensões de performance, não houve diferença significativa em relação aos demais, exceto em quatro indicadores (*ID01*, *ID13*, *ID14* e *ID20* - este apenas marginalmente) e na dimensão Qualidade Técnica/Científica, que formaram um grupo com valorações significativamente menor em relação aos demais, uma vez que todos os demais tiveram intervalos de confiança que se sobrepuseram. Dessa forma, em média, os indivíduos tenderam a dar menor importância/relevância para quatro indicadores de (*ID01_TaxaMedPerm*, *ID13_PropMedMest*, *ID14_PropProfMest* e *ID20_IndiceAtendUsuaSemVincReg*) e também na dimensão Qualidade Técnica/Científica, em relação aos demais indicadores e dimensões, mas tenderam a dar valorações (importância/relevância) dentro das mesmas margens de valores para todos os demais indicadores e dimensões de performance, na *proxy* E+F.

Desta análise descritiva sintetizou-se que os profissionais e gestores atribuíram elevados graus de importâncias/relevâncias das subcategorias (20 indicadores de performance) e das categorias (sete dimensões de performance) dos itens de análise, sendo consideradas relevantes/importantes para a estruturação de um instrumento de avaliação de performance. Além disso, em média, as valorações da ampla maioria destes itens de análise foram

razoavelmente dentro das mesmas margens de valores, uma vez que a maioria dos intervalos de confiança se sobrepuseram, com algumas exceções: *ID01_TaxaMedPerm*, *ID13_PropMedMest*, *ID14_PropProfMest* e *ID20_IndiceAtendUsuaSemVincReg* e, ainda, na dimensão Qualidade Técnica/Científica, que se propõem aprofundar investigações.

Quanto aos resultados do estudo das possíveis influências de três variáveis moderadoras, das análises das tabelas 57 e 58 (apêndice 9, tópico 9.1.1.2), sintetizou-se que, na *proxy* E+F, não houve tendência de mediação de valorações médias obtidas na ampla maioria das dimensões de performance quanto aos níveis de atenção à saúde, de complexidade da assistência e de cargo/função dos especialistas e profissionais, com algumas exceções: quanto aos níveis de atenção à saúde (exceto nas dimensões Produtividade Assistencial e Efetividade) e de cargo/função (exceto na dimensão Efetividade).

Estes resultados subsidiaram a construção de um novo instrumento, devidamente alinhado pela literatura relevante de indicadores e dimensões de performance, aplicado na fase posterior da investigação (*proxy* G, estudo 3).

6.1.1.3. Análise descritiva do estudo 3 (ciclo 2: *proxy* G) (indicadores *versus* dimensões de performance) (Visão de decisores)

A parte empírica do estudo 3 foi desenvolvida a partir do instrumento consolidado nas diversas fases subsequentes procedentes no ciclo 1 (*proxy* E+F), inicialmente com estruturação e validação do novo instrumento ampliado (*proxy* G), com 7 dimensões e 56 indicadores, através de um grupo piloto de respondentes do ciclo 1 (*proxy* E+F) e, por fim, aplicado em uma pesquisa de campo em fase única (ciclo 2).

Inicialmente, na *proxy* G, ciclo 2 do estudo 3, aplicou-se uma análise da validação em um grupo piloto de 15 informantes chaves (perspectiva direta de gestores da saúde para a visão de Decisores), que participaram da fase anterior (*proxy* E+F), tendo atribuído valorações médias elevadas (superior a 4, muito relevante/importante) para todos os itens do novo instrumento proposto (*proxy* G). Observou-se que no grupo piloto não houve acréscimos e nem eliminações de dimensões e de indicadores de performance. Também não se efetivou novos reagrupamentos e/ou realinhamentos com a literatura relevante e preferiu-se fazer, uma análise integrada da *proxy* G (233 informantes no ciclo 2 de aplicação de campo, na visão de Decisores).

Em seguida, executou-se uma análise integrada das 233 respostas obtidas na *proxy G* (7 dimensões e 56 indicadores) a fim de consolidar os níveis de relevância/importância atribuídas para as dimensões e indicadores de performance em serviços de saúde, na visão de Decisores.

A figura 38 abaixo apresenta uma análise descritiva das variáveis de caracterização selecionadas na amostra, na análise da *proxy G* (233 informantes), no ciclo 2 de aplicação de campo, na visão de Decisores.

Na fase 4, de consolidação dos indicadores e dimensões de performance, fase única do estudo 3 (*proxy G*) a amostra mostrou-se bastante relevante (Gráfico 9). A maior parte dos indivíduos trabalhava no nível de atenção à saúde (48,50%) trabalhava no nível de atenção à saúde Primária, seguido por parte (23,61%) que trabalhavam em Outros níveis (estruturas não assistenciais, por exemplo, apenas de gestão ou governança); a maior parte dos indivíduos na amostra (47,21%) trabalhava no nível de Baixa complexidade seguidos pelos indivíduos que trabalhavam nos níveis da Média complexidade (34,33%). A maior parte dos indivíduos na amostra (39,68%) exercia nível de cargo/função Outros (por exemplo, assessores de gestão, profissionais em geral etc), enquanto que boa parte (23,61%) exercia cargos Sistêmicos e de indivíduos (20,60%) de cargos de Gerência/Chefia Setorial. A maioria dos indivíduos na amostra (56,22%) possuía pós-graduação, 25,75% possuía nível superior de escolaridade e 10,73% mestrado ou PhD. A ampla maioria dos indivíduos (73,39%) possuía experiência maior que 60 meses. Menos de 1 % possuía experiência em cargo/função de decisão menor que 6 meses.

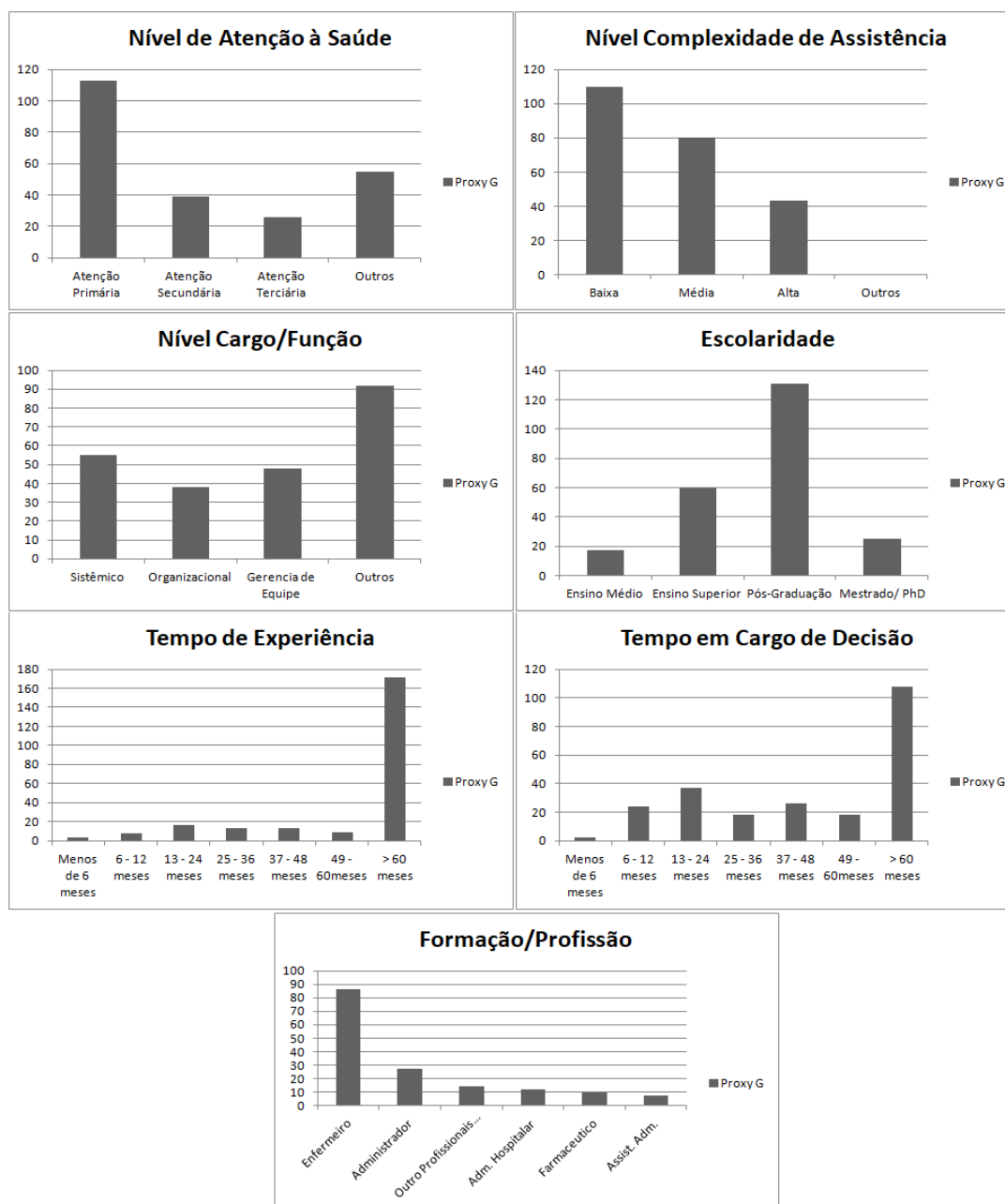


Figura 38- Gráfico de barras das variáveis de caracterização selecionadas para estudo de fatores mediadores – Proxy G (ciclo 2, estudo 3) (Visão de Decisores).

A Tabela 5 apresenta uma análise descritiva relativa às valorações médias atribuídas aos 56 indicadores de performance, na análise da *proxy G* (233 informantes no ciclo 2 de aplicação de campo, na visão de Decisores).

Tabela 5- Análise descritiva dos indicadores de performance (IDD01 a IDD56) – Proxy G (ciclo 2, estudo 3) (Visão de Decisores).

Dimensão	Indicador	N	Média	D.P.	I.C-95% ¹
Produtividade	IDD01_TaxaConsulBas	169	4,40	0,92	[4,27; 4,53]
Assistencial	IDD02_TaxaConsultUrgenc	168	4,30	1,08	[4,14; 4,45]

	IDD03_TaxaConsulMedBasGOPed	167	4,35	1,02	[4,18; 4,49]
	IDD04_TaxaConsulMedAmbEspec	165	4,46	0,75	[4,34; 4,57]
	IDD05_TaxaPacienObsHospPronAtend	168	4,33	1,04	[4,16; 4,48]
	IDD06_TaxaProcCirurg	168	4,23	1,16	[4,06; 4,39]
	IDD07_TaxaInterHosp	165	4,42	0,88	[4,28; 4,55]
	IDD08_TaxaExamCitop	168	4,39	1,12	[4,22; 4,56]
	IDD09_TaxaExamMamogra	168	4,35	1,15	[4,17; 4,52]
	IDD10_TaxaExamRadiol	167	4,28	1,14	[4,11; 4,44]
	IDD11_IndicSessQuimioSist	167	4,22	1,22	[4,02; 4,39]
	IDD12_IndicSessRadiotSist	168	4,18	1,24	[3,99; 4,36]
	IDD13_TaxaSessHemodCron	167	4,18	1,24	[3,99; 4,36]
	IDD14_TaxaProcOdontEspec	167	4,25	1,13	[4,07; 4,41]
Eficiência	IDD15_TaxaPermanInter	154	4,50	0,80	[4,37; 4,62]
	IDD16_TaxaProcConsMed	155	4,49	0,80	[4,35; 4,62]
	IDD17_TaxaRotativLeitos	154	4,47	0,94	[4,33; 4,62]
	IDD18_PropEnfLeitos	154	4,44	0,94	[4,28; 4,57]
	IDD19_PercentPacienSaidosInternSIGTAP	152	4,36	1,05	[4,18; 4,51]
	IDD20_PercentExcesPermLimiarMaxSIGTAP	154	4,36	1,08	[4,17; 4,53]
Efetividade	IDD21_PropCirurgRescind	148	4,11	1,35	[3,89; 4,33]
	IDD22_TaxaReintern	148	4,26	1,12	[4,07; 4,44]
	IDD23_PropPastosNormSUS	148	4,32	1,13	[4,13; 4,49]
	IDD24_TaxaDensIncidInfecCirurg	146	4,42	1,08	[4,23; 4,60]
	IDD25_TaxaDensIncidInfecCorrentUTI	146	4,30	1,20	[4,08; 4,47]
	IDD26_TaxaSepticPosOper	147	4,39	1,14	[4,20; 4,57]
	IDD27_PercAltasHospUTI	146	4,23	1,32	[3,99; 4,43]
Acesso	IDD28_TaxaPacienListaEsperAmb	146	4,50	0,87	[4,33; 4,64]
	IDD29_TaxaPacienListaEspeEletiva	146	4,50	0,87	[4,35; 4,64]
	IDD30_TempMedListaEsperaAgendAmb	146	4,43	0,86	[4,29; 4,56]
	IDD31_TempMedListaEsperaAgendCirurgEletiva	145	4,42	0,86	[4,28; 4,56]
	IDD31_TempMedListaEsperaLeitosTransf	144	4,38	1,08	[4,19; 4,55]
	IDD33_TempMedAtendUsuaGravesCrit	145	4,57	0,87	[4,42; 4,70]
	IDD34_PropTransfOutrosEstab	144	4,48	0,82	[4,34; 4,60]
Impacto	IDD35_TaxaMortInst	146	4,39	0,87	[4,25; 4,52]
	IDD36_TaxaMortUTI	146	4,12	1,09	[3,95; 4,30]
	IDD37_TaxaMortIAM	146	4,13	1,10	[3,95; 4,30]
	IDD38_TaxaMortAVC	146	4,23	1,00	[4,06; 4,38]
	IDD39_TaxaMortPacienTransfPosCirurg	145	4,12	1,09	[3,95; 4,30]
	IDD40_TaxaMortPacienOncol	144	3,99	1,20	[3,79; 4,17]
	IDD41_TaxaMortPacienCronHemod	144	4,04	1,19	[3,83; 4,23]
	IDD42_PropEncamTranspCronicHemod	144	4,06	1,17	[3,85; 4,25]
	IDD43_TaxaMortMaterna	146	4,27	1,11	[4,10; 4,43]
	IDD44_TaxaMortPerinatalGest	146	4,31	1,10	[4,12; 4,48]
	IDD45_PropMortNeonatalBaixoPeso	146	4,32	1,10	[4,14; 4,49]
	IDD46_PropMortInfantil	146	4,32	1,14	[4,12; 4,49]
	IDD47_PropNascVivos	146	4,27	1,11	[4,08; 4,45]

Qualidade Tecnológica/Científica	IDD48_PropMedEspec	146	4,07	1,01	[3,90; 4,23]
	IDD49_PropProfEspec	146	4,08	1,02	[3,92; 4,23]
Sistêmico/Estratégico	IDD50_IndiceSatisfUsua	147	4,20	0,93	[4,04; 4,34]
	IDD51_IndiceSatisfProf	147	4,20	0,88	[4,06; 4,35]
	IDD52_IndiceInternICSAP	147	4,12	1,06	[3,94; 4,29]
	IDD53_IndiceAtendUsuaComVincReg	146	3,99	1,02	[3,82; 4,16]
	IDD54_IndiceAtendUsuaSemVincReg	147	3,92	1,04	[3,74; 4,08]
	IDD55_VariaPercentCustoTotal	146	4,14	0,90	[4,01; 4,30]
	IDD56_IndiceExtrapTetosOrçProj	144	4,13	1,00	[3,97; 4,27]
Outros indicadores		137	4,04	1,01	[3,86; 4,19]

Nota.¹ Intervalo Bootstrap

A análise global permite afirmar que todos os 56 indicadores de performance avaliados obtiveram valoração média elevada (superior ou cerca de 4, muito importante/relevante), sendo todos considerados importantes/relevantes para a estruturação do painel definitivo com indicadores e dimensões de performance desta investigação (*proxy* G, ciclo 2 do estudo 3).

Por sua vez, a análise das dimensões de performance, estruturou-se a partir da valoração média atribuída diretamente às dimensões de performance pelos indivíduos, não sendo necessários cálculos indiretos.

A Tabela 6 apresenta uma análise descritiva das valorações médias atribuídas diretamente às dimensões de performance na *proxy* G (233 informantes), na visão de Decisores da saúde.

Tabela 6- Análise descritiva das dimensões de performance (DD1 a DD7) – *Proxy* G (ciclo 2, estudo 3) (Visão de Decisores).

Dimensão	N	Média	D.P.	I.C-95% ¹
Produtividade Assistencial	169	4,31	0,74	[4,20; 4,42]
Eficiência	155	4,44	0,70	[4,33; 4,54]
Efetividade	148	4,29	0,98	[4,12; 4,44]
Acesso	146	4,47	0,69	[4,36; 4,57]
Impacto	146	4,20	0,96	[4,04; 4,34]
Qualidade Tecnológica/Científica	146	4,07	1,01	[3,91; 4,23]
Sistêmico/Estratégico	147	4,10	0,84	[3,95; 4,23]

Nota.¹ Intervalo Bootstrap

A análise global permite afirmar que todas as sete dimensões de performance obtiveram valoração média elevada (superior a 4, muito importante/relevante), sendo todas consideradas importantes/relevantes para a estruturação do painel definitivo com indicadores e dimensões de performance desta investigação (*proxy* G).

A figura 39, a seguir, mostra o grau de importância comparada entre os indicadores e as dimensões de performance atribuídos na fase 4 (*proxy* G).

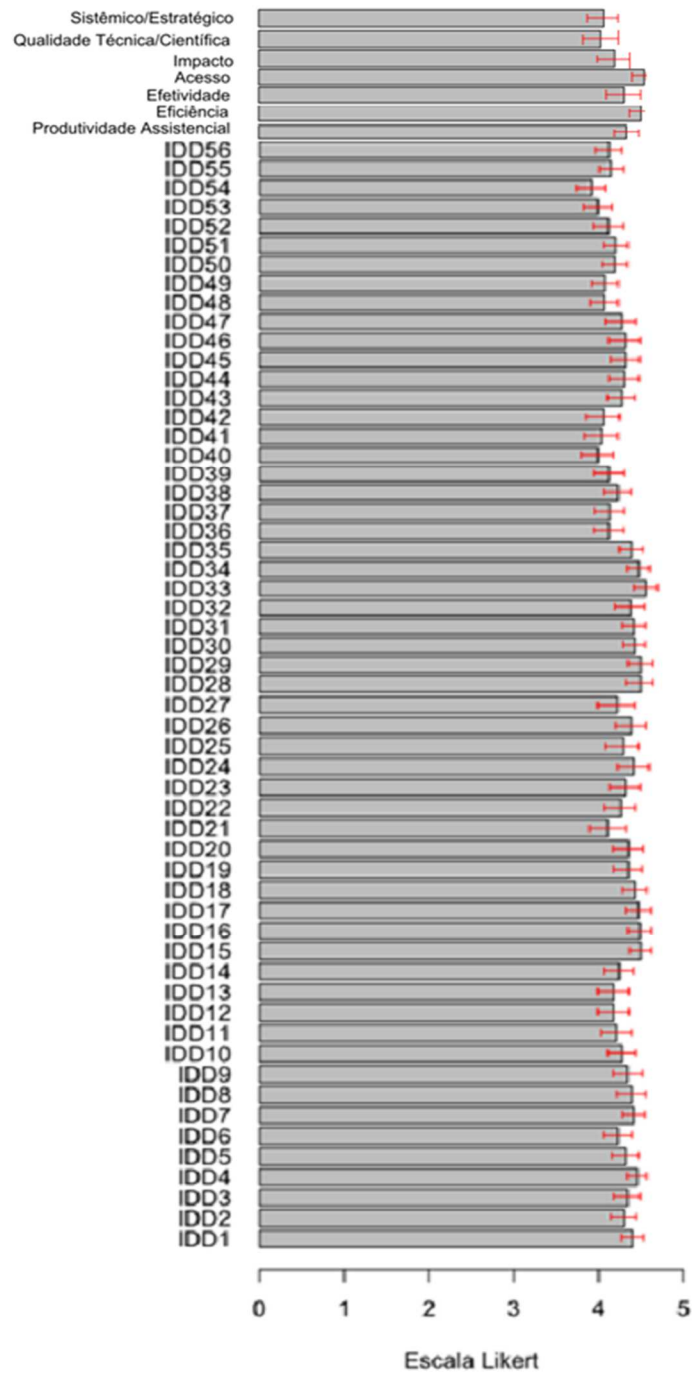


Figura 39- Grau de importância/relevância de indicadores e dimensões de performance (*proxy* G).

Em todos os 56 indicadores e nas sete dimensões de performance, não houve diferença significativa em relação aos demais, uma vez que todos tiveram intervalos de confiança que se sobrepuseram. Dessa forma, em média, os indivíduos tenderam a dar valorações (importância/relevância) dentro das mesmas margens de valores para todos demais indicadores e dimensões de performance, na *proxy* G.

Desta análise descritiva sintetizou-se que os decisores atribuíram elevados graus de importâncias/relevâncias das subcategorias (56 indicadores de performance) e das categorias (sete dimensões de performance) dos itens de análise, sendo consideradas relevantes/importantes para a estruturação de um instrumento de avaliação de performance. Além disso, em média, as valorações de todos estes itens de análise foram razoavelmente dentro das mesmas margens de valores, uma vez que todos os intervalos de confiança dos 56 indicadores e das sete dimensões de performance se sobrepuseram.

Quanto aos resultados do estudo das possíveis influências de três variáveis moderadoras, das análises das tabelas 59 a 61 (apêndice 9, tópico 9.1.1.3), sintetizou-se que, na *proxy* G, não houve tendência de mediação de valorações médias obtidas na ampla maioria das dimensões de performance quanto aos níveis de atenção à saúde, de complexidade da assistência e de cargo/função dos especialistas e profissionais, exceto na dimensão Produtividade Assistencial.

Estes resultados subsidiaram as próximas fases de validação por estatística inferencial destes 56 indicadores de performance e sete dimensões de performance propostos nesta investigação (*proxy* G) a fim de propor um painel de indicadores e performance para unidades de saúde do SUS.

6.1.2. Análises inferenciais dos estudos 1 a 3 (*proxy* C, *proxy* E+F e *proxy* G)

Após a conclusão da análise estatística descritiva, aplicou-se métodos de análise inferencial, a fim de verificar a validade das hipóteses propostas e consolidar os modelos desenhados nesta investigação para os indicadores e dimensões de performance (estudos 1 a 3), conforme métodos de análise inferencial detalhados no início do tópico 6, que dividiram-se em quatro análises de resultados: estudo das cargas fatoriais dos domínios *versus* macrodomínios de avaliação estudados (*proxy* C) e indicadores *versus* dimensões de performance (*proxy* E+F e *proxy* G), validação por critérios de qualidade e validade dos macrodomínios de avaliação (*proxy* C) e de dimensões de performance (*proxy* E+F e *proxy* G), análise da qualidade (ajustamento) do modelo final resultante, e, análise das correlações entre as variáveis e suas influências na Performance Global Percebida nas três *proxies* estudadas (*proxy* C, *proxy* E+F e *proxy* G).

6.1.2.1 Análise inferencial do estudo 1 (domínios *versus* macrodomínios de avaliação (ciclo 2: *proxy* C) (Visão de gestores e profissionais)

Em busca de validar os cinco macrodomínios de avaliação, além de estruturar e validar a variável latente Performance Global Percebida, procedentes da *proxy* C (112 informantes), aplicou-se métodos de análise inferencial a fim de consolidar os modelos desenhados nesta investigação, em três categorias iniciais de análises inferenciais: estudo das cargas fatoriais entre variáveis, validação por critérios de qualidade e validade além de análise da qualidade (ajustamento) do modelo final resultante, que se apresentam a seguir, sendo complementadas pela análise das correlações entre as variáveis e suas influências na Performance Global Percebida (apêndice 9, tópico 9.1.2.1).

Para validar os macrodomínios de avaliação estudados (Quantitativos, Qualitativos, Efeitos, Satisfação e Sistêmico/Estratégico) foi feita uma Análise Fatorial Confirmatória (AFC) (Hair et al., 2009) para os macrodomínios de avaliação, na *proxy* C (n = 112). Logo, pode-se destacar que, no modelo final, após análise e exclusão de sete domínios de avaliação (*D1A_Produtiv*, *D1B_EstrutAssist*, *D1C_CobertAssit*, *D2H_QualidAssist*, *D2M_IntegralAtenc*, *D2N_Equidade* e *D3P_EficaciaProj*), todos os demais 17 domínios de avaliação que permaneceram na análise, apresentaram carga fatorial superior a 0,50 (ou suas permanências não impediram a validação convergente da AFC), na *proxy* C, com as cargas fatoriais mostradas na tabela 7.

Tabela 7- Análise Fatorial Confirmatória dos domínios de avaliação – *Proxy* C (ciclo 2, estudo 1) (Visão de gestores e profissionais).

Macrodomínios	Domínios	Modelo Inicial		Modelo Final	
		Peso ¹	C.F. ²	Peso ¹	C.F. ²
Quantitativos	D1A_Produtiv	15%	0,48		
	D1B_EstrutAssist	11%	0,39		
	D1C_CobertAssit	10%	0,37		
	D1D_OtimizCusto	21%	0,71	29%	0,63
	D1E_EficienEcon	20%	0,66	37%	0,82
	D1F_SustEconFin	21%	0,65	34%	0,71
Qualitativos	D2H_QualidAssist	12%	0,47		
	D2I_RiscoAssist	13%	0,60	23%	0,66
	D2J_AtendHumaniz	15%	0,61	29%	0,75
	D2K_AcessibilServ	14%	0,54	22%	0,54
	D2L_AcessoServ	16%	0,62	26%	0,60
	D2M_IntegralAtenc	19%	0,65		
Efeitos	D2N_Equidade	12%	0,53		
	D3P_EficaciaProj	20%	0,51		
	D3Q_EfetivClinica	17%	0,58	21%	0,58
	D3R_EficienAssist	26%	0,80	32%	0,79

	D3S_ResolubAssist	19%	0,63	23%	0,61
	D3T_Impacto	18%	0,61	24%	0,64
Satisfação	D4U_SatisfUsuario	45%	0,71	44%	0,70
	D4V_SatisfProfiss	55%	0,81	56%	0,83
Sistêmico/Estratégia	D5W_IndICSAP	22%	0,50	22%	0,50
	D5X_AtendVincReg	24%	0,72	24%	0,72
	D5Y_AtendDesVincReg	31%	0,81	31%	0,81
	D5V_ExtrapTetoOrç	23%	0,59	23%	0,59

Nota.¹ Peso relativo; ² Carga Fatorial.

Na Tabela 8, pode-se observar que a confiabilidade e a validade dos cinco macrodomínios de avaliação (Quantitativos, Qualitativos, Efeitos, Satisfação e Sistêmico/Estratégico) foram asseguradas, uma vez que todos apresentaram validação convergente ($AVE > 0,40$), confiabilidade adequada ($A.C. > 0,60$ ou $C.C. > 0,60$), unidimensionalidade e validação discriminante ($VCM < AVE$), na *proxy C* ($n = 112$).

Tabela 8- Validação por critérios de qualidade e validade dos macrodomínios de avaliação - *Proxy C* (ciclo 2, estudo 1) (Visão de gestores e profissionais).

Macrodomínios	Domínios	A.C. ¹	C.C. ²	Dim. ³	AVE ⁴	VCM ⁵
Quantitativos	3	0,76	0,76	1	0,52	0,13
Qualitativos	4	0,71	0,73	1	0,41	0,24
Efeitos	5	0,75	0,75	1	0,44	0,24
Satisfação	2	0,73	0,74	1	0,59	0,24
Sistêmico/Estratégia	4	0,72	0,75	1	0,44	0,14

Nota.¹ Alfa de Cronbach; ² Confiabilidade Composta; ³ Dimensionalidade; ⁴ Variância Média Extraída;

⁵ Variância Máxima Compartilhada.

Para validar a variável latente Performance Global Percebida foi feita uma Análise Fatorial Confirmatória (AFC) (Hair et al., 2009) da valoração média obtida para os macrodomínios de avaliação, na *proxy C* ($n = 112$). Logo, pode-se destacar que, no modelo final, todos os cinco macrodomínios (Quantitativos, Qualitativos, Efeitos, Satisfação e Sistêmico/Estratégico) permaneceram na análise e apresentaram carga fatorial superior a 0,50 (ou suas permanências não impediram a validação convergente da AFC), na *proxy C*, com as cargas fatoriais mostradas na tabela 9.

Tabela 9- Análise Fatorial Confirmatória da Performance Global Percebida – *Proxy C* (ciclo 2, estudo 1) (Visão de gestores e profissionais).

Macrodomínios	Modelo Inicial		Modelo Final	
	Peso ¹	C.F. ²	Peso ¹	C.F. ²
Quantitativa	15%	0,63	15%	0,48
Qualitativa	19%	0,90	21%	0,82
Efeitos	24%	0,82	19%	0,76
Satisfacao	25%	0,74	27%	0,78

Sistêmico	17%	0,53	19%	0,56
-----------	-----	------	-----	------

Nota.¹ Peso relativo; ² Carga Fatorial.

Na Tabela 10, pode-se observar que a confiabilidade e a validade da variável latente Performance Global Percebida foram asseguradas, uma vez que apresentou validação convergente ($AVE > 0,40$), confiabilidade adequada ($A.C. > 0,60$ ou $C.C. > 0,60$), unidimensionalidade e validação discriminante ($VCM < AVE$), na *proxy C* ($n = 112$).

Tabela 10 - Validação da Performance Global Percebida – *Proxy C* (ciclo 2, estudo 1) (Visão de gestores e profissionais).

Macrodomínios	A.C. ¹	C.C. ²	Dim. ³	AVE ⁴	VCM ⁵
5	0,73	0,82	1	0,48	-

Nota.¹ Alfa de Cronbach; ² Confiabilidade Composta; ³ Dimensionalidade; ⁴ Variância Média Extraída; ⁵ Variância Máxima Compartilhada.

Também aplicaram-se testes para medir a qualidade (ajustamento) do modelo construído para os macrodomínios e domínios de avaliação da *proxy C* (112 informantes), através de alguns índices selecionados na literatura relevante ($X^2/G.L.$, CFI, TLI e RMSEA). Para um bom ajuste espera-se que $X^2/G.L.$ seja menor que 3 (Arbuckle, 2008; Wheaton, 1987), CFI seja maior que 0,80 (Bentler, 1990), TLI seja maior que 0,80 (Bentler & Bonnet, 1980) e que o RMSEA (Steiger et al., 1985) seja menor que 0,10, sendo o ideal abaixo de 0,05.

Na Tabela 11, pode-se observar que os parâmetros e índices de ajustamento e qualidade dos modelos de equações estruturais indicaram bom ajuste, já que a razão entre a estatística qui-quadrado e os graus de liberdade foi menor que 3, as estatísticas TLI e o CFI foram maiores que 0,80 e o RMSEA do foi menor que o limite máximo de 0,10.

Tabela 11 - Parâmetros de qualidade do modelo final da Performance Global Percebida – *Proxy C* (ciclo 2, estudo 1) (Visão de gestores e profissionais).

Medidas de Qualidade do Ajuste	Estatística de Teste	Estatística de Teste Reescalada (Satorra e Bentler 1994)
χ^2	166,61	150,32
G.L.	114	114
$\chi^2/G.L.$	1,46	1,32
TLI	0,888	0,888
CFI	0,906	0,906
RMSEA	0,064	0,053
	[0,042; 0,084]	[0,028; 0,074]
valor-p (RMSEA \leq 0,05)	0,137	0,389

Diante destes resultados, a figura 40 sintetizou o ajuste da Análise Fatorial Confirmatória (AFC) e a modelagem para a variável latente Performance Global Percebida, na *proxy C* (112 informantes, visão direta de gestores e profissionais). Devido ao elevado número de domínios

de avaliação optou-se por omitir suas nomenclaturas, que ficaram identificados como D1A a D5Y.

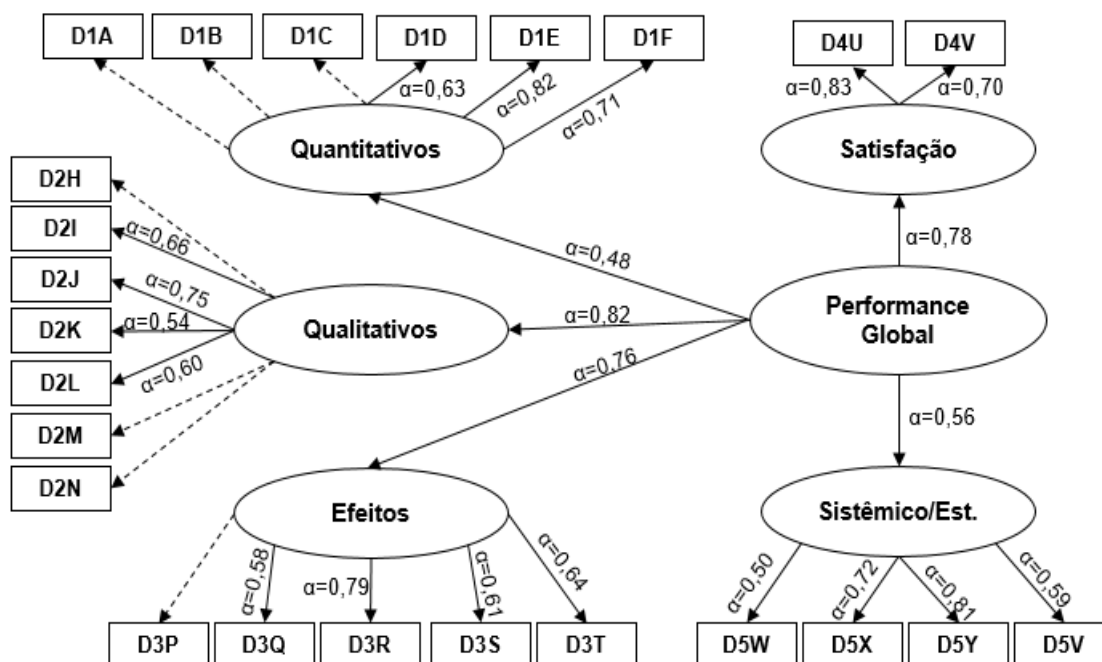


Figura 40- Análise Fatorial Confirmatória da Performance Global Percebida – Proxy C (ciclo 2, estudo 1) (Visão de gestores e profissionais).

Para concluir a análise inferencial, a figura 63 sintetizou os resultados apresentados na Tabela 62 (apêndice 9, tópico 9.1.2.1), nos quais foi possível observar que os domínios de avaliação apresentaram maiores correlações com seus respectivos macrodomínios, na proxy C (n = 112).

Por sua vez, os resultados obtidos mostraram que a Performance Global Percebida média foi de 4,16 e a mediana de 4,18 e que 75% das respostas atribuídas resultaram em valoração superior a 3,88 (cerca de 4, muito importante/relevante). Pode-se observar uma variabilidade na valoração média da variável latente Performance Global Percebida. Porém, não houve diferença significativa (valor-p > 0,050) para considerar as variáveis nível de atenção à saúde, de complexidade da assistência ou de cargo/função enquanto fator mediador da Performance Global Percebida, na proxy C (n = 112) (Tabela 63, apêndice 9, tópico 9.1.2.1).

6.1.2.2 Análise inferencial do estudo 2 (indicadores *versus* dimensões de performance) (ciclo 1: proxy E+F) (Visão de gestores e profissionais)

Em busca de validar as sete dimensões de performance, além de estruturar e validar a variável latente Performance Global Percebida, procedentes da *proxy* E+F (112 informantes), aplicou-se métodos de análise inferencial a fim de consolidar os modelos desenhados nesta investigação, em três categorias iniciais de análises inferenciais: estudo das cargas fatoriais entre variáveis, validação por critérios de qualidade e validade além de análise da qualidade (ajustamento) do modelo final resultante, que se apresentam a seguir, sendo complementadas pela análise das correlações entre as variáveis e suas influências na Performance Global Percebida (apêndice 9, tópico 9.1.2.2).

Para validar as sete dimensões de performance estudadas (Produtividade Assistencial, Eficiência, Efetividade, Acesso, Impacto, Qualidade Técnica/Científica, Sistêmico/Estratégico) foi feita uma Análise Fatorial Confirmatória (AFC) (Hair et al., 2009) para as dimensões de performance, na *proxy* E+F (n = 112). Logo, pode-se destacar que, no modelo final, após análise e exclusão de seis indicadores de performance (*ID02_TaxaMedProced*, *ID03_TaxaRotatLeitos*, *ID04_CustoSetAtend*, *ID12_PropNascVivos*, *ID17_IndiceInternICSAP* e *ID21_IndiceExtrapTetosOrç*), todos os demais 14 indicadores que permaneceram na análise, apresentaram carga fatorial superior a 0,50 (ou suas permanências não impediram a validação convergente da AFC), na *proxy* E+F, com as cargas fatoriais mostradas na tabela 12.

Tabela 12- Análise Fatorial Confirmatória dos indicadores de performance – *Proxy* E+F (ciclo 1, estudo 2)
(Perspectiva indireta para Decisores).

Dimensões	Indicadores	Modelo Inicial		Modelo Final	
		Peso ¹	C.F. ²	Peso ¹	C.F. ²
Produtividade Assistencial	ID01_TaxaMedPerm	100%	1,00	100%	1,00
	ID02_TaxaMedProced				
Eficiência	ID03_TaxaRotatLeitos	19%	0,42		
	ID04_CustoSetAtend	36%	0,49		
	ID05_NoEnfer.leitos	44%	0,51	100%	1,00
Efetividade	ID06_PropCirurgRescind	36%	0,82	37%	0,82
	ID07_TaxaReintern	33%	0,82	33%	0,81
	ID08_TaxaInfecCirurg	31%	0,71	29%	0,68
Acesso	ID09_TaxaPacienEspec	53%	0,82	54%	0,83
	ID10_PropPacienEmerg	47%	0,83	46%	0,82
Impacto	ID11_TaxaMortInst	55%	0,79	100%	1,00
	ID12_PropNascVivos	45%	0,47		
Qualidade Técnica/Científica	ID13_PropMedMest	54%	0,98	53%	0,95
	ID14_PropProfMest	46%	0,90	47%	0,92
Sistêmico/ Estratégico	ID15_IndiceSatisfUsua	20%	0,70	25%	0,68
	ID16_IndiceSatisfProf	20%	0,63	28%	0,68
	ID17_IndiceInternICSAP	10%	0,34		
	ID18_IndiceAtendUsuaCom	19%	0,61	24%	0,61

ID20_IndiceAtendUsuaSem	19%	0,60	24%	0,61
ID21_IndiceExtrapTetosOrç	12%	0,39		

Nota.¹ Peso relativo; ² Carga Fatorial.

Na Tabela 13, pode-se observar que a confiabilidade e a validade das sete dimensões de performance estudadas (Produtividade Assistencial, Eficiência, Efetividade, Acesso, Impacto, Qualidade Técnica/Científica, Sistemico/Estratégico) foram asseguradas, uma vez que todas apresentaram validação convergente ($AVE > 0,40$), confiabilidade adequada ($A.C. > 0,60$ ou $C.C. > 0,60$), unidimensionalidade e validação discriminante ($VCM < AVE$), na *proxy* E+F ($n = 112$). Três dimensões (Produtividade Assistencial, Eficiência e Impacto) eram formadas por um único indicador de performance e apresentaram validação por definição.

Tabela 13- Validação por critérios de qualidade e validade das dimensões de performance – *Proxy* E+F (ciclo 1, estudo 2) (Perspectiva indireta para Decisores).

Dimensões	Indicadores	A.C. ¹	C.C. ²	Dim. ³	AVE ⁴	VCM ⁵
Produtividade Assistencial	1	-	-	1	-	-
Eficiência	1	-	-	1	-	-
Efetividade	3	0,82	0,63	1	0,60	0,23
Acesso	2	0,81	0,64	1	0,68	0,19
Impacto	1	-	-	1	-	-
Qualidade Técnica/Científica	2	0,93	0,87	1	0,88	0,15
Sistemico/Estratégico	4	0,68	0,46	1	0,42	0,19

Nota.¹ Alfa de Cronbach; ² Confiabilidade Composta; ³ Dimensionalidade; ⁴ Variância Média Extraída; ⁵ Variância Máxima Compartilhada.

Para validar a variável latente Performance Global Percebida foi feita uma Análise Fatorial Confirmatória (AFC) (Hair et al., 2009) da valoração média obtida para as dimensões de performance, na *proxy* E+F ($n = 112$). Logo, pode-se destacar que, no modelo final, após análise e exclusão de uma dimensão de performance (Produtividade Assistencial) que apresentou carga fatorial inferior a 0,50, as seis dimensões de performance que permaneceram na análise (Eficiência, Efetividade, Acesso, Impacto, Qualidade Técnica/Científica e Sistemico/Estratégico) apresentaram carga fatorial superior ou cerca de 0,50 (ou suas permanências não impediram a validação convergente da AFC), na *proxy* E+F, com as cargas fatoriais mostradas na tabela 14.

Tabela 14- Análise Fatorial Confirmatória das dimensões de performance – *Proxy* E+F (ciclo 1, estudo 2) (Perspectiva indireta para Decisores).

Dimensões	Modelo Inicial		Modelo Final	
	Peso ¹	C.F. ²	Peso ¹	C.F. ²
Produtividade Assistencial	4%	0,12	-	-
Eficiência	8%	0,98	17%	0,48
Efetividade	17%	0,56	18%	0,63
Acesso	19%	0,70	17%	0,65

Impacto	13%	0,70	11%	0,49
Qualidade Técnica/Científica	20%	0,53	18%	0,51
Sistêmico/Estratégico	19%	0,90	19%	0,96

Nota.¹ Peso relativo; ² Carga Fatorial.

Na Tabela 15, pode-se observar que a confiabilidade e a validade da variável latente Performance Global Percebida foram asseguradas, uma vez que apresentou validação convergente ($AVE > 0,40$), confiabilidade adequada ($A.C. > 0,60$ ou $C.C. > 0,60$), unidimensionalidade e validação discriminante ($VCM < AVE$), na *proxy* E+F ($n = 112$).

Tabela 15 - Validação da Performance Global Percebida – *Proxy* E+F (ciclo 1, estudo 2) (Perspectiva indireta para Decisores).

Dimensões	A.C. ¹	C.C. ²	Dim. ³	AVE ⁴	VCM ⁵
6	0,72	0,57	1	0,41	-

Nota.¹ Alfa de Cronbach; ² Confiabilidade Composta; ³ Dimensionalidade; ⁴ Variância Média Extraída; ⁵ Variância Máxima Compartilhada.

Também se aplicaram testes para medir a qualidade (ajustamento) do modelo construído para as dimensões e indicadores de performance da *proxy* E+F (112 informantes), através de alguns índices selecionados na literatura relevante ($X^2/G.L.$, CFI, TLI e RMSEA). Para um bom ajuste espera-se que $X^2/G.L.$ seja menor que 3 (Arbuckle, 2008; Wheaton, 1987), CFI seja maior que 0,80 (Bentler, 1990), TLI seja maior que 0,80 (Bentler & Bonnet, 1980) e que o RMSEA (Steiger et al., 1985) seja menor que 0,10, sendo o ideal abaixo de 0,05.

Na Tabela 16, pode-se observar que os parâmetros e índices de ajustamento e qualidade dos modelos de equações estruturais indicaram bom ajuste, já que a razão entre a estatística qui-quadrado e os graus de liberdade foi menor que 3, as estatísticas TLI e o CFI foram maiores que 0,80 e o RMSEA do foi menor que o limite máximo de 0,10.

Tabela 16- Parâmetros de qualidade (ajustamento) do modelo final da Performance Global Percebida – *Proxy* E+F (ciclo 1, estudo 2) (Perspectiva indireta para Decisores).

Medidas de Qualidade do Ajuste	Estatística de Teste	
	Estatística de Teste	Reescalada (Satorra e Bentler 1994)
χ^2	142,18	101,97
G.L.	60,00	60,00
$\chi^2/G.L.$	2,37	1,70
TLI	0,840	0,749
CFI	0,877	0,807
RMSEA	0,111 [0,087;	0,079 [0,056; 0,101]
valor-p (RMSEA \leq 0,05)	0,134]	
	0,000	0,021

Diante destes resultados, a figura 41 sintetizou o ajuste da Análise Fatorial Confirmatória (AFC) e a modelagem para a variável latente Performance Global Percebida, na *proxy* E+F (112 informantes, perspectiva indireta para a visão dos Decisores). Devido ao elevado número de indicadores de performance optou-se por omitir suas nomenclaturas na figura 2, que ficaram identificados como ID01 a ID21.

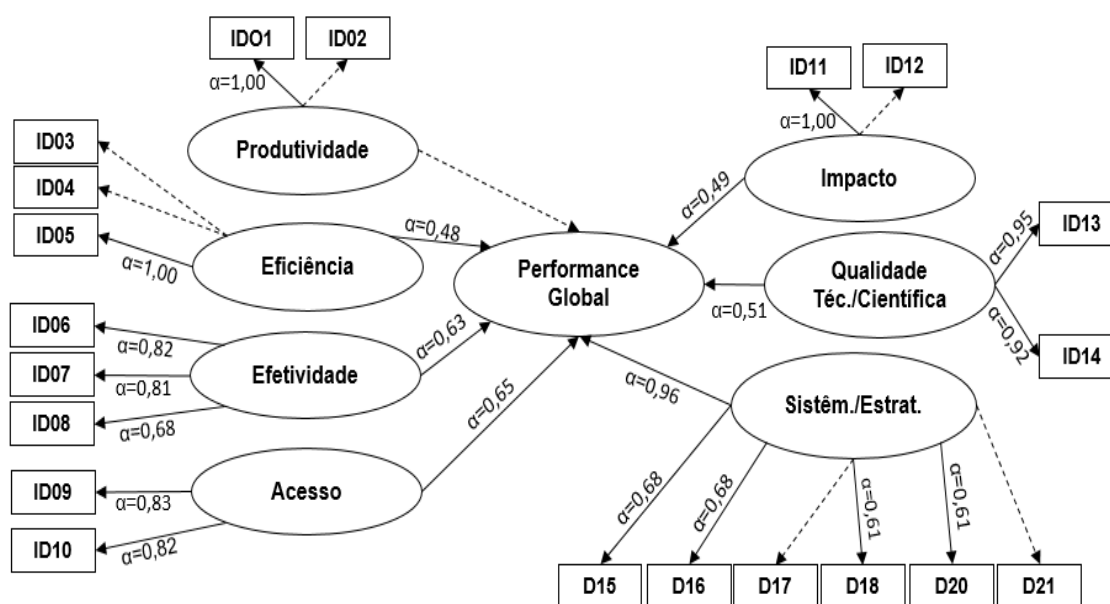


Figura 41- Análise Fatorial Confirmatória da Performance Global Percebida – *Proxy* E+F (ciclo 1, estudo 2) (Perspectiva indireta para Decisores).

Para concluir a análise inferencial, a figura 64 sintetizou os resultados apresentados na Tabela 64 (apêndice 9, tópico 9.1.2.2), nos quais foi possível observar que os indicadores de performance apresentaram maiores correlações com suas respectivas dimensões, na *proxy* E+F ($n = 112$).

Por sua vez, os resultados obtidos, mostraram que a Performance Global Percebida média foi de 4,04 e a mediana de 4,04 e que 75% das respostas atribuídas resultaram em valoração superior a 3,96 (cerca de 4, muito importante/relevante). Pode-se observar uma variabilidade na valoração média da variável latente Performance Global Percebida. Porém, não houve diferença significativa (valor- $p > 0,050$) para considerar as variáveis nível de atenção à saúde, de complexidade da assistência enquanto fator mediador da Performance Global Percebida. No entanto, houve diferença significativa (valor- $p = 0,019$) quanto a influência do nível de cargo/função na valoração da Performance Global Percebida. Ao fazerem-se as comparações múltiplas, houve diferença significativa (valor- $p = 0,015$) entre os de cargo/função

Sistêmicos e Organizacionais, sendo que os indivíduos de cargo/função Sistêmicos apresentaram Performance Global Percebida significativamente maior, na *proxy* E+F (n = 112) (Tabela 65, apêndice 9, tópico 9.1.2.2).

6.1.2.3 Análise inferencial do estudo 3 (indicadores *versus* dimensões de performance) (ciclo 2: *proxy* G) (Visão de decisores)

Em busca de validar as sete dimensões de performance, além de estruturar e validar a variável latente Performance Global Percebida, procedentes da *proxy* E+F (112 informantes), aplicou-se métodos de análise inferencial a fim de consolidar os modelos desenhados nesta investigação, em três categorias iniciais de análises inferenciais: estudo das cargas fatoriais entre variáveis, validação por critérios de qualidade e validade além de análise da qualidade (ajustamento) do modelo final resultante, que se apresentam a seguir, sendo complementadas pela análise das correlações entre as variáveis e suas influências na Performance Global Percebida (apêndice 9, tópico 9.1.2.3).

Para validar as sete dimensões de performance estudadas (Produtividade Assistencial, Eficiência, Efetividade, Acesso, Impacto, Qualidade Técnica/Científica, Sistêmico/Estratégico) foi feita uma Análise Fatorial Confirmatória (AFC) (Hair et al., 2009) para as dimensões de performance, na *proxy* G (n = 233). Logo, pode-se destacar que, no modelo final, após análise e exclusão de sete indicadores de performance (*IDD01_TaxaConsulBas*, *IDD02_TaxaConsultUrgenc*, *IDD03_TaxaConsulMedBasGOPed*, *IDD04_TaxaConsulMedAmbEspec*, *IDD05_TaxaPacienObsHospPronAtend*, *IDD15_TaxaPermanInter* e *IDD16_TaxaProcConsMed*), todos os demais 49 indicadores que permaneceram na análise, apresentaram carga fatorial superior a 0,50 (ou suas permanências não impediram a validação convergente da AFC), com as cargas fatoriais mostradas na tabela 17.

Tabela 17- Análise Fatorial Confirmatória dos indicadores de performance – *Proxy* G (ciclo 2, estudo 3) (Visão de Decisores).

Dimensões	Indicadores	Modelo Inicial		Modelo Final	
		Peso ¹	C.F. ²	Peso ¹	C.F. ²
Produtividade de Assistencia l	<i>IDD01_TaxaConsulBas</i>	4%	0,43		
	<i>IDD02_TaxaConsultUrgenc</i>	5%	0,46		
	<i>IDD03_TaxaConsulMedBasGOPed</i>	4%	0,38		
	<i>IDD04_TaxaConsulMedAmbEspec</i>	1%	0,17		
	<i>IDD05_TaxaPacienObsHospPronAtend</i>	4%	0,39		
	<i>IDD06_TaxaProcCirurg</i>	7%	0,56	9%	0,55

	IDD07_TaxaInterHosp	5%	0,50	6%	0,49
	IDD08_TaxaExamCitop	8%	0,69	10%	0,67
	IDD09_TaxaExamMamogra	9%	0,74	11%	0,72
	IDD10_TaxaExamRadiol	8%	0,64	10%	0,63
	IDD11_IndicSessQuimioSist	13%	0,98	16%	0,98
	IDD12_IndicSessRadiotSist	13%	0,97	17%	0,98
	IDD13_TaxaSessHemodCron	10%	0,73	12%	0,73
	IDD14_TaxaProcOdontEspec	7%	0,55	8%	0,54
Eficiência	IDD15_TaxaPermanInter	7%	0,32		
	IDD16_TaxaProcConsMed	6%	0,30		
	IDD17_TaxaRotativLeitos	13%	0,53	15%	0,52
	IDD18_PropEnfLeitos	18%	0,70	21%	0,70
	IDD19_PercentPacienSaidosInternSIGTAP	28%	0,98	32%	0,99
	IDD20_PercentExcesPermLimiarMaxSIGTAP	28%	0,94	32%	0,94
Efetividade	IDD21_PropCirurgRescind	16%	0,75	16%	0,75
	IDD22_TaxaReintern	10%	0,59	10%	0,59
	IDD23_PropPastosNormSUS	11%	0,62	11%	0,62
	IDD24_TaxaDensIncidInfecCirurg	14%	0,88	14%	0,88
	IDD25_TaxaDensIncidInfecCorrentUTI	16%	0,89	16%	0,89
	IDD26_TaxaSepticPosOper	16%	0,89	16%	0,90
	IDD27_PercAltasHospUTI	17%	0,85	17%	0,85
Acesso	IDD28_TaxaPacienListaEsperAmb	10%	0,52	10%	0,48
	IDD29_TaxaPacienListaEspeEletiva	19%	0,96	14%	0,70
	IDD30_TempMedListaEsperaAgendAmb	12%	0,62	11%	0,52
	IDD31_TempMedListaEsperaAgendCirurgEletiva	19%	0,95	13%	0,68
	IDD32_TempMedListaEsperaLeitosTransf	14%	0,57	20%	0,83
	IDD33_TempMedAtendUsuaGravesCrit	12%	0,61	18%	0,84
	IDD34_PropTransfOutrosEstab	12%	0,61	15%	0,75
Impacto	IDD35_TaxaMortInst	4%	0,58	4%	0,58
	IDD36_TaxaMortUTI	5%	0,55	5%	0,54
	IDD37_TaxaMortIAM	7%	0,79	7%	0,77
	IDD38_TaxaMortAVC	5%	0,63	5%	0,66
	IDD39_TaxaMortPacienTransfPosCirurg	7%	0,72	7%	0,73
	IDD40_TaxaMortPacienOncol	9%	0,84	8%	0,82
	IDD41_TaxaMortPacienCronHemod	8%	0,84	8%	0,82
	IDD42_PropEncamTranspCronicHemod	8%	0,83	8%	0,81
	IDD43_TaxaMortMaterna	9%	0,97	9%	0,97
	IDD44_TaxaMortPerinatalGest	9%	0,98	9%	0,99
	IDD45_PropMortNeonatalBaixoPeso	9%	0,98	9%	0,99
	IDD46_PropMortInfantil	9%	0,93	9%	0,93
	IDD47_PropNascVivos	9%	0,95	9%	0,95
Qualidade Tecnológica/Científica	IDD48_PropMedEspec	50%	1,00	50%	1,00
	IDD49_PropProfEspec	50%	0,97	50%	0,97

Sistêmico/ Estratégico	IDD50_IndiceSatisfUsua	14%	0,84	13%	0,77
	IDD51_IndiceSatisfProf	13%	0,83	12%	0,77
	IDD52_IndiceInternICSAP	16%	0,87	17%	0,88
	IDD53_IndiceAtendUsuaComVincReg	16%	0,87	16%	0,90
	IDD54_IndiceAtendUsuaSemVincReg	15%	0,84	16%	0,86
	IDD55_VariaPercentCustoTotal	12%	0,77	12%	0,74
	IDD56_IndiceExtrapTetosOrçProj	14%	0,82	15%	0,82

Nota. ¹ Peso relativo; ² Carga Fatorial;

Na Tabela 18, pode-se observar que a confiabilidade e a validade das sete dimensões de performance estudadas (Produtividade Assistencial, Eficiência, Efetividade, Acesso, Impacto, Qualidade Técnica/Científica, Sistêmico/Estratégico) foram asseguradas, uma vez que todas apresentaram validação convergente ($AVE > 0,40$), confiabilidade adequada ($A.C. > 0,60$ ou $C.C. > 0,60$), unidimensionalidade e validação discriminante ($VCM < AVE$), na *proxy* G ($n = 233$).

Tabela 18- Validação por critérios de qualidade e validade das dimensões de performance – *Proxy* G (ciclo 2, estudo 3) (Visão de Decisores).

Dimensões	Indicadores	A.C. ¹	C.C. ²	Dim. ³	AVE ⁴	VCM ⁵
Produtividade Assistencial	9	0,91	0,90	1	0,52	0,23
Eficiência	4	0,88	0,88	1	0,65	0,24
Efetividade	7	0,92	0,92	1	0,63	0,30
Acesso	7	0,89	0,86	1	0,51	0,37
Impacto	13	0,97	0,96	1	0,69	0,37
Qualidade Tecnológica/Científica	2	0,98	0,98	1	0,97	0,34
Sistêmico/Estratégico	7	0,94	0,94	1	0,69	0,34

Nota. ¹ Alfa de Cronbach; ² Confiabilidade Composta; ³ Dimensionalidade; ⁴ Variância Média Extraída; ⁵ Variância Máxima Compartilhada.

Para validar a variável latente Performance Global Percebida foi feita uma Análise Fatorial Confirmatória (AFC) (Hair et al., 2009) da valoração média obtida para as dimensões de performance, na *proxy* G ($n = 233$). Logo, pode-se destacar que, no modelo final, após análise e exclusão de uma dimensão de performance (Eficiência) que apresentou carga fatorial inferior a 0,50, as seis dimensões de performance que permaneceram na análise (Produtividade Assistencial, Efetividade, Acesso, Impacto, Qualidade Técnica/Científica e Sistêmico/Estratégico) apresentaram carga fatorial superior ou cerca de 0,50 (ou suas permanências não impediram a validação convergente da AFC), na *proxy* G, com as cargas fatoriais mostradas na tabela 19.

Tabela 19- Análise Fatorial Confirmatória das dimensões de performance – *Proxy* G (ciclo 2, estudo 3) (Visão de Decisores).

Dimensões	Modelo Inicial		Modelo Final	
	Peso ¹	C.F. ²	Peso ¹	C.F. ²
Produtividade Assistencial	9%	0,53	14%	0,54

Eficiência	3%	0,33	23%	0,57
Efetividade	21%	0,50	13%	0,80
Acesso	11%	0,62	15%	0,75
Impacto	16%	0,82	19%	0,49
Qualidade Tecnológica/Científica	20%	0,50	16%	0,58
Sistêmico/Estratégico	19%	0,61		

Nota.¹ Peso relativo; ² Carga Fatorial;

Na Tabela 20, pode-se observar que a confiabilidade e a validade da variável latente Performance Global Percebida foram asseguradas, uma vez que apresentou validação convergente ($AVE > 0,40$), confiabilidade adequada ($A.C. > 0,60$ ou $C.C. > 0,60$), unidimensionalidade e validação discriminante ($VCM < AVE$), na *proxy* G ($n = 233$).

Tabela 20- Validação da Performance Global Percebida – *Proxy* G (ciclo 2, estudo 3) (Visão de Decisores).

Dimensões	A.C. ¹	C.C. ²	Dim. ³	AVE ⁴	VCM ⁵
6	0,78	0,79	1	0,40	-

Nota.¹ Alfa de Cronbach; ² Confiabilidade Composta; ³ Dimensionalidade; ⁴ Variância Média Extraída; ⁵ Variância Máxima Compartilhada.

Também aplicaram-se testes para medir a qualidade (ajustamento) do modelo construído para as dimensões e indicadores de performance da *proxy* G (233 informantes), através de alguns índices selecionados na literatura relevante ($X^2/G.L.$, CFI, TLI e RMSEA). Para um bom ajuste espera-se que $X^2/G.L.$ seja menor que 3 (Arbuckle, 2008; Wheaton, 1987), CFI seja maior que 0,80 (Bentler, 1990), TLI seja maior que 0,80 (Bentler & Bonnet, 1980) e que o RMSEA (Steiger et al., 1985) seja menor que 0,10, sendo o ideal abaixo de 0,05.

Na Tabela 21, pode-se observar que os parâmetros e índices de ajustamento e qualidade dos modelos de equações estruturais indicaram bom ajuste, já que a razão entre a estatística qui-quadrado e os graus de liberdade foi menor que 3, as estatísticas TLI e o CFI foram maiores que 0,80 e o RMSEA do foi menor que o limite máximo de 0,10.

Tabela 21- Parâmetros de qualidade (ajustamento) do modelo final da Performance Global Percebida – *Proxy* G (ciclo 2, estudo 3) (Visão de Decisores).

Medidas de Qualidade do Ajuste	Estatística de Teste	Estatística de Teste Reescalada (Satorra e Bentler 1994)
χ^2	5335,16	1726,37
G.L.	1109,00	1109,00
$\chi^2/G.L.$	4,81	1,56
TLI	0,714	0,793
CFI	0,730	0,804
RMSEA	0,128 [0,124; 0,131]	0,049 [0,046; 0,051]
valor-p (RMSEA \leq 0,05)	0,000	0,765

Diante destes resultados, a figura 42 sintetizou o ajuste da Análise Fatorial Confirmatória (AFC) e a modelagem para a variável latente Performance Global Percebida, na *proxy* G (233 informantes, visão de Decisores). Devido ao elevado número de indicadores de performance optou-se por omitir suas nomenclaturas na figura 42, que ficaram identificados como IDD01 a IDD56.

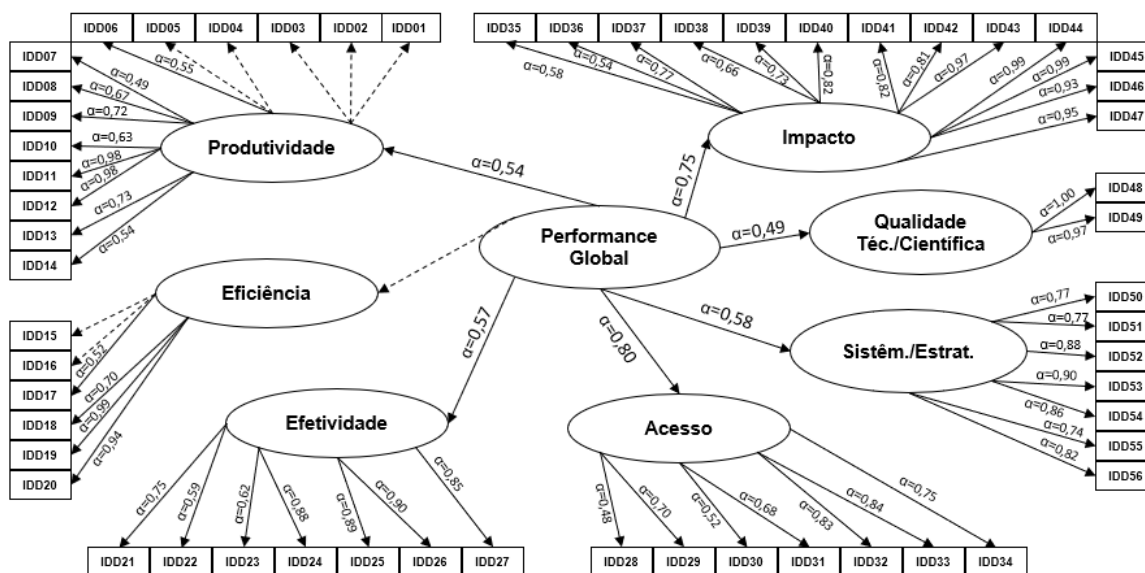


Figura 42- Análise Fatorial Confirmatória da Performance Global Percebida – *Proxy* G (ciclo 2, estudo 3) (Visão de Decisores).

Para concluir a análise inferencial, a figura 65 sintetizou os resultados apresentados na Tabela 66 (apêndice 9, tópico 9.1.2.3), nos quais foi possível observar que os indicadores de performance apresentaram maiores correlações com suas respectivas macrodimensões, na *proxy* G ($n = 233$).

Por sua vez, os resultados obtidos mostraram que a Performance Global Percebida média foi de 4,22 e a mediana de 4,22 e que 75% das respostas atribuídas resultaram em valoração superior a 4,45 (acima de 4, muito importante/relevante). Pode-se observar uma variabilidade na valoração média da variável latente Performance Global Percebida. Porém, não houve diferença significativa (valor- $p > 0,050$) para considerar as variáveis nível de atenção à saúde, de complexidade da assistência e de cargo/função ocupados enquanto fator mediador da Performance Global Percebida, na *proxy* G ($n = 233$) (Tabela 67, apêndice 9, tópico 9.1.2.3).

Dessa forma, concluiu-se o mapeamento dos conhecimentos tácitos dos agentes chaves da amostra nas categorias de macrodomínios e domínios de avaliação, seguidos pela etapa de análise e consolidação de painéis de indicadores e dimensões de performance, além dos reflexos sobre uma *variável latente* Performance Global Percebida, medidos por suas magnitudes

(cargas fatoriais), significâncias estatísticas e suas correlações em um modelo estrutural validado.

6.2. Os resultados dos itens de medição de qualidade/satisfação (escala de qualidade percebida)

Os resultados foram agregados em subcategorias (constructos e questões) e categoriais (subdimensões) e macrocategorias (macrodimensões) de análise da a fim de estruturar uma escala de medição de qualidade/satisfação percebida pelo usuário, a partir dos conceitos consolidados da escala SERVQUAL (Albuquerque, 2012; E. Babakus & W. G. Mangold, 1992; Parasuraman et al., 1985b, 1988, 1991a; Parasuraman et al., 1994; Rosalem, 2013) anteriormente caracterizados (tópico 2.4.4) através de um processo empírico colaborativo de painéis de especialistas e grupos de tradução transcultural das 22 questões e das cinco macrodimensões da escala SERVQUAL, tendo sido estruturado um modelo inovador de escala de medição de qualidade/satisfação no SUS (*QualitySaude*), com 20 questões validadas agregadas em 13 subdimensões de qualidade percebida distribuídas em seis macrodimensões qualidade percebida, sintetizadas no quadro 14 abaixo:

<u>Macrodimensões</u>	Definições conceituais
I. Tangibilidade	Está relacionado à atratividade das instalações, equipamentos e materiais usados por uma empresa de serviços, bem como à aparência dos funcionários do serviço.
II. Confiabilidade	A empresa de serviços oferece a seus clientes serviço correto da primeira vez, sem cometer nenhum erro, e entrega o que prometeu dentro do prazo estipulado.
III. Segurança	O comportamento dos funcionários transmitirá aos clientes confiança na empresa e que esta faz com que se sintam seguros. E também que os empregados são sempre corteses e têm o conhecimento necessário para responder às perguntas dos clientes.
IV. Presteza	Os funcionários estão dispostos a ajudar os clientes e atender aos seus requisitos, bem como informá-los, quando o serviço será prestado e então executá-lo com presteza.
V. Empatia	A empresa entende os problemas dos clientes e executa o serviço, tendo em vista seus melhores interesses, bem como lhes dá atenção pessoal, individual, e trabalha em horários convenientes.
VI. Acessibilidade	Garantia de acesso oportuno, efetivo e eficaz ao atendimento em saúde.
VII. Satisfação Global	Qualidade global percebida e satisfação relatada pelo cliente.
<u>Subdimensões</u>	Definições conceituais

1. Estrutura física	Aspectos físicos e estruturais da unidade de saúde
2. Apresentação Pessoal	Aspectos da aparência pessoal, asseio e higiene dos profissionais
3. Instalações físicas	Funcionamento das instalações físicas
4. Equipamentos	Funcionalidades dos equipamentos utilizados
5. Confiabilidade organizacional	Confiança nos compromissos de entrega dos serviços com qualidade prometida e tempo oportuno
6. Fornecimento de serviços	Entrega dos serviços com qualidade prometida e tempo oportuno
7. Confiança na Competência Profissional.	Confiança nas competências técnicas dos profissionais
8. Segurança na Resolução de Problemas	Confiança no comportamento dos empregados que transmitem segurança na solução de seus problemas, e também que os empregados têm o conhecimento necessário para responder às perguntas dos clientes.
9. Atenção e Disposição para Atendimento	Percepção que os funcionários estão dispostos a ajudar os clientes e atender aos seus requisitos, quando o serviço será prestado e então executá-lo com presteza.
10. Acesso à Informação	Garantia de acesso as informações e disposição de funcionários para informar aos clientes.
11. Humanização do Atendimento	Aspectos de acolhimentos, relações interpessoais adequadas e que os empregados são sempre corteses
12. Prioridade nos Interesses do cliente	Percepção que os funcionários estão dispostos a priorizar os interesses dos clientes e suas necessidades.
13. Acessibilidade aos serviços	Garantia de acesso oportuno e eficaz ao atendimento em saúde

Quadro 14- Definições conceituais das subdimensões e macrodimensões de qualidade/satisfação percebida

6.2.1. Análises descritivas dos estudos 4 e 5 (*proxies* H+I e J, *proxies* M e M+N)

Os resultados das análises descritivas dos itens de avaliação de qualidade percebida integrada das *proxies* H e I (195 informantes), no ciclo 1 de validação do instrumento, além da *proxy* J (506 informantes) de aplicação de campo (ciclo 2), conforme métodos de estatística descritiva detalhados no início do tópico 6, constam dos subtópicos 6.2.1.1 e 6.2.1.2, respectivamente. Os resultados das análises descritivas dos itens de avaliação de qualidade percebida pela análise integrada da *proxy* K+L (140 informantes), no ciclo 1 de validação do instrumento, além da *proxy* M (506 informantes) e da análise integrada da *proxy* M+N (2.547 informantes) de aplicação de campo (ciclo 2), constam dos subtópicos 6.2.1.3 e 6.2.1.4, respectivamente.

6.2.1.1. Análise descritiva do estudo 4 (ciclo 1: *proxies* H e I) (constructos *versus* subdimensões e macrodimensões de qualidade) (Visão profissionais *versus* usuários)

A partir dos conhecimentos prévios de especialistas e profissionais de saúde, estruturou-se o processo colaborativo de tradução transcultural e validação de uma escala inovadora de medição de qualidade/satisfação de usuários de saúde, que manteve as cinco macrodimensões originais da escala SERVQUAL e para cada uma de suas 22 questões originais estruturou uma subdimensão de constructos que foram desagregadas em 40 constructos de qualidade a fim de garantir proximidade à realidade do SUS e entendimento adequada dos itens de medição de qualidade/satisfação propostos nesta fase da investigação.

A parte empírica do estudo 4 foi desenvolvida a partir do instrumento consolidado nas diversas fases subsequentes procedentes no ciclo 1 (*proxy* H+I), através de um grupo piloto de respondentes (n = 195) e, posteriormente, com estruturação e validação de um novo instrumento ampliado (*proxy* J), aplicado em uma pesquisa de campo em única fase incremental, no ciclo 2 (n = 506).

Inicialmente, abordou-se uma análise integrada de fases incrementais (fases 1 a 3, *proxies* H e I) que se iniciou pela validação de face (fase 1) e aplicou dois instrumentos diferentes (fases 2 e 3), através de uma amostra de 195 informantes especialistas, gestores e profissionais de saúde, assim distribuídos: *proxy* H - 25 informantes e *proxy* I - 170 informantes. Efetivou-se uma análise integrada das respostas obtidas nestes dois instrumentos (*proxy* H+I) pois apresentaram muitas similaridades (40 constructos, 22 subdimensões de constructos e cinco macrodimensões de qualidade percebida). Em fase posterior (fase 4), abordou-se a análise da *proxy* J - 506 informantes. (40 constructos, 22 subdimensões de constructos e cinco macrodimensões de qualidade percebida).

A Tabela 22 abaixo apresenta uma análise descritiva das variáveis de caracterização selecionadas na amostra, na análise integrada das *proxies* H e I (195 informantes) que aplicou dois instrumentos com muitas similaridades, no ciclo 1 de validação do instrumento, na visão de especialistas e profissionais.

Tabela 22- Análise descritiva das variáveis de caracterização selecionadas - Proxy H+I (ciclo 1, estudo 4) (Perspectiva indireta para Usuários e Profissionais.

Variáveis		N	%
Nível de Atenção à Saúde	Atenção Primária	28	14,36%
	Atenção Secundária	68	34,87%
	Atenção Terciária	56	28,72%
	Outros	43	22,05%
Nível de complexidade da assistência	Baixa	34	17,44%
	Média	86	44,10%
	Alta	33	16,92%
	Outros	42	21,54%
Nível de cargo/função	Sistêmico	19	10,11%
	Organizacional	51	27,13%
	Gerencia de Equipe	36	19,15%
	Outros	82	43,62%
Escolaridade	Analfabeto	0	0,00%
	Ensino Básico	0	0,00%
	Ensino Fundamental	0	0,00%
	Ensino Médio	13	6,70%
	Ensino Superior	56	28,87%
	Pós-Graduação	98	50,52%
	Mestrado/ PhD	27	13,92%
Tempo de experiência	Não tem	1	0,52%
	Menos de 6 meses	23	11,92%
	6 - 12 meses	35	18,13%
	13 - 24 meses	14	7,25%
	25 - 36 meses	9	4,66%
	37 - 48 meses	8	4,15%
	49 - 60meses	10	5,18%
	> 60 meses	93	48,19%
Tempo em cargo de decisão	Não tem	8	4,40%
	Menos de 6 meses	72	39,56%
	6 - 12 meses	25	13,74%
	13 - 24 meses	10	5,49%
	25 - 36 meses	17	9,34%
	37 - 48 meses	6	3,30%
	49 - 60meses	13	7,14%
	> 60 meses	31	17,03%
Formação/Profissão	Enfermeiro	36	18,46%
	Engenheiro	12	6,15%
	Administrador	12	6,15%
	Medico	10	5,13%
	Assistente Social	10	5,13%
	Outras	115	58,97%

Nas fases 1 a 3 (proxy H+I), destaca-se que a amostra se mostrou balanceada (Tabela 22) no nível de atenção à saúde que os indivíduos trabalhavam: níveis Secundário (34,87%),

Primário (14,36%) e Terciário (28,72%) de atenção à saúde, além de 22,05% da categoria Outros (estruturas não assistenciais, por exemplo, apenas de gestão ou governança). Grande parte dos indivíduos (43,62%) exercia o nível de cargo/função Outros (por exemplo, assessores de gestão, profissionais em geral etc) e boa parte (27,13%) exercia o nível de cargo/função Organizacional. Considerou-se a amostra (n = 195) bastante qualificada visto que a maior parte dos indivíduos (50,52%) possuía pós-graduação e ainda 93,30% possuía nível superior, a grande parte dos indivíduos (48,19%) possuía mais de 60 meses de experiência e que uma boa parte dos indivíduos (17,03%) trabalhava a mais de 60 meses em cargo/função de decisão enquanto apenas 4,40% não possuía experiência em cargo/função de decisão.

A Tabela 23 apresenta uma análise descritiva relativa à valoração média atribuída aos constructos de qualidade, na análise integrada das *proxies* H e I (195 informantes), no ciclo 1 de validação do instrumento, perspectiva indireta com atribuição de valorações por especialistas e profissionais, com observações para a perspectiva de Usuários e a dos Profissionais.

Tabela 23- Análise descritiva de constructos originais - *Proxy* H+I (ciclo 1, estudo 4) (Perspectiva indireta para Usuários e Profissionais).

Subdimensões de constructos	Constructos originais	Usuário			Profissional		
		Média	D.P.	I.C-95% ¹	Média	D.P.	I.C-95% ¹
Equipamentos	CO1_Disponib	4,55	0,74	[4,44; 4,65]	4,57	0,77	[4,47; 4,68]
	CO2_Acesso	4,52	0,76	[4,42; 4,62]	4,37	0,79	[4,27; 4,49]
	CO3_Resolut	4,36	0,82	[4,25; 4,47]	4,60	0,80	[4,48; 4,70]
	CO4_Conserv	4,02	0,92	[3,89; 4,15]	4,43	0,83	[4,32; 4,54]
	CO5_Usabilid	3,65	1,12	[3,50; 3,80]	4,40	0,81	[4,28; 4,50]
Estrutura Física	CO6_Conforto	4,24	0,97	[4,11; 4,37]	4,15	0,89	[4,04; 4,28]
	CO7_Limpeza	4,46	0,84	[4,35; 4,57]	4,63	0,74	[4,52; 4,73]
	CO8_Bem	4,32	0,89	[4,20; 4,44]	4,39	0,83	[4,28; 4,50]
Apresentação Pessoal	CO9_Vestim	3,89	1,01	[3,75; 4,03]	4,41	0,82	[4,29; 4,52]
	CO10_Visual	4,23	0,97	[4,08; 4,37]	4,51	0,82	[4,39; 4,63]
Instalações Físicas	CO11_Conserv	4,24	0,89	[4,11; 4,35]	4,39	0,80	[4,27; 4,50]
	CO12_Funcion	4,20	0,80	[4,09; 4,3]	4,55	0,76	[4,44; 4,64]
Confiança em Prazos	CO13_Cumprim.Prazos	4,42	0,91	[4,29; 4,54]	4,15	0,93	[4,02; 4,28]
Suporte ao usuário-cliente	CO14_Solidaried	4,20	0,78	[4,09; 4,31]	4,09	0,83	[3,97; 4,20]
	CO15_Apoio_U	4,46	0,73	[4,35; 4,55]	4,32	0,79	[4,21; 4,43]
	CO16_Acolhim	4,61	0,75	[4,50; 4,71]	4,15	0,86	[4,02; 4,28]
Confiabilidade Organizacional	CO17_Confianca.Perc	4,72	0,57	[4,63; 4,79]	4,60	0,70	[4,50; 4,70]
Fornecimento de Serviços	CO18_Prazo.Resolut	4,32	0,78	[4,21; 4,43]	4,33	0,77	[4,23; 4,43]
	CO19_Prazo.Atendim	4,49	0,85	[4,36; 4,60]	4,21	0,81	[4,09; 4,32]

Garantia dos Registros	CO20_ Inform.Disp.Co ntr	3,55	1,08	[3,40; 3,71]	4,51	0,79	[4,41; 4,61]
Acesso Informação	CO21_ Esclarecer	4,23	1,11	[4,08; 4,38]	3,66	1,15	[3,50; 3,81]
	CO22_ Infor	4,31	1,07	[4,16; 4,45]	3,83	1,21	[3,64; 3,99]
Pronto Atendimento	CO23_ Atendim	4,40	0,87	[4,29; 4,52]	3,95	0,89	[3,84; 4,07]
	CO24_ Disponib	4,31	0,88	[4,19; 4,44]	3,71	0,93	[3,58; 3,84]
Disposição p/ atendimento	CO25_ Dispos.Assisten c	4,29	1,08	[4,14; 4,44]	3,91	1,08	[3,76; 4,07]
Ocupação profissional	CO26_ Profiss.Suficien te	4,19	1,22	[4,01; 4,34]	3,91	1,23	[3,74; 4,08]
Competência profissional	CO27_ Conf.Compet.Pr ofiss	4,72	0,56	[4,64; 4,79]	4,49	0,64	[4,41; 4,57]
Credibilidade e Profissional	CO28_ Confianca	4,61	0,64	[4,51; 4,69]	4,56	0,67	[4,47; 4,65]
	CO29_ Credib	4,69	0,68	[4,58; 4,78]	4,64	0,76	[4,52; 4,75]
Cortesia nas Relações	CO30_ Cortesia	4,57	0,70	[4,47; 4,67]	4,54	0,73	[4,43; 4,64]
	CO31_ Respeito	4,54	0,72	[4,43; 4,63]	4,57	0,67	[4,47; 4,66]
	CO32_ Atitude	4,54	0,67	[4,44; 4,63]	4,44	0,74	[4,34; 4,55]
Suprimento de Recursos	CO33_ Garantia	3,88	1,31	[3,68; 4,05]	4,33	1,07	[4,18; 4,47]
	CO34_ Suprim	3,89	1,27	[3,70; 4,07]	4,31	1,10	[4,15; 4,45]
Comunicação e Atenção	CO35_ Atendim.Indiv	4,30	1,11	[4,14; 4,45]	4,26	1,10	[4,10; 4,41]
Humanização no Atendimento	CO36_ Dedicção	3,78	1,18	[3,61; 3,95]	4,03	1,17	[3,84; 4,18]
	CO37_ Tratam	4,26	1,17	[4,08; 4,42]	4,11	1,14	[3,95; 4,26]
Compreensão do Cliente	CO38_ Conhec.Situac. Usuario	4,31	1,08	[4,16; 4,47]	4,31	1,12	[4,15; 4,47]
Prioridade nos interesses	CO39_ Objet.Centrado. Usuario	4,28	1,08	[4,12; 4,43]	4,22	1,12	[4,06; 4,37]
Acessibilidade dos clientes	CO40_ Facil.Acesso.Us uario	4,21	1,20	[4,04; 4,38]	3,41	1,29	[3,23; 3,59]

Nota! Intervalo Bootstrap

A análise global permite afirmar que a ampla maioria dos 40 constructos originais obtiveram valoração média elevada (superior ou cerca de 4, muito importante/relevante), para a perspectiva de Usuários e a dos Profissionais, na análise da *proxy* H+I, sendo os 40 constructos considerados relevantes/importantes para as próximas fases de investigação, na estruturação do instrumento de pesquisa de Usuários (*proxy* J, ciclo 2).

A figura 43, que sintetizou os dados da Tabela 23, mostra que na ampla maioria dos 40 constructos de qualidade (CO1 a CO40) houve variabilidade na valoração atribuída pelos especialistas e profissionais de saúde, de forma indireta, para a perspectiva de Usuários e a dos Profissionais, sendo que na maioria dos constructos analisados os intervalos de confiança não se sobrepuseram. Dessa forma, em média, os indivíduos tenderam a dar valorações

(importância/relevância) com diferenças significativas na ampla maioria dos 40 constructos para a perspectiva de Usuários e a dos Profissionais, na *proxy* H+I.

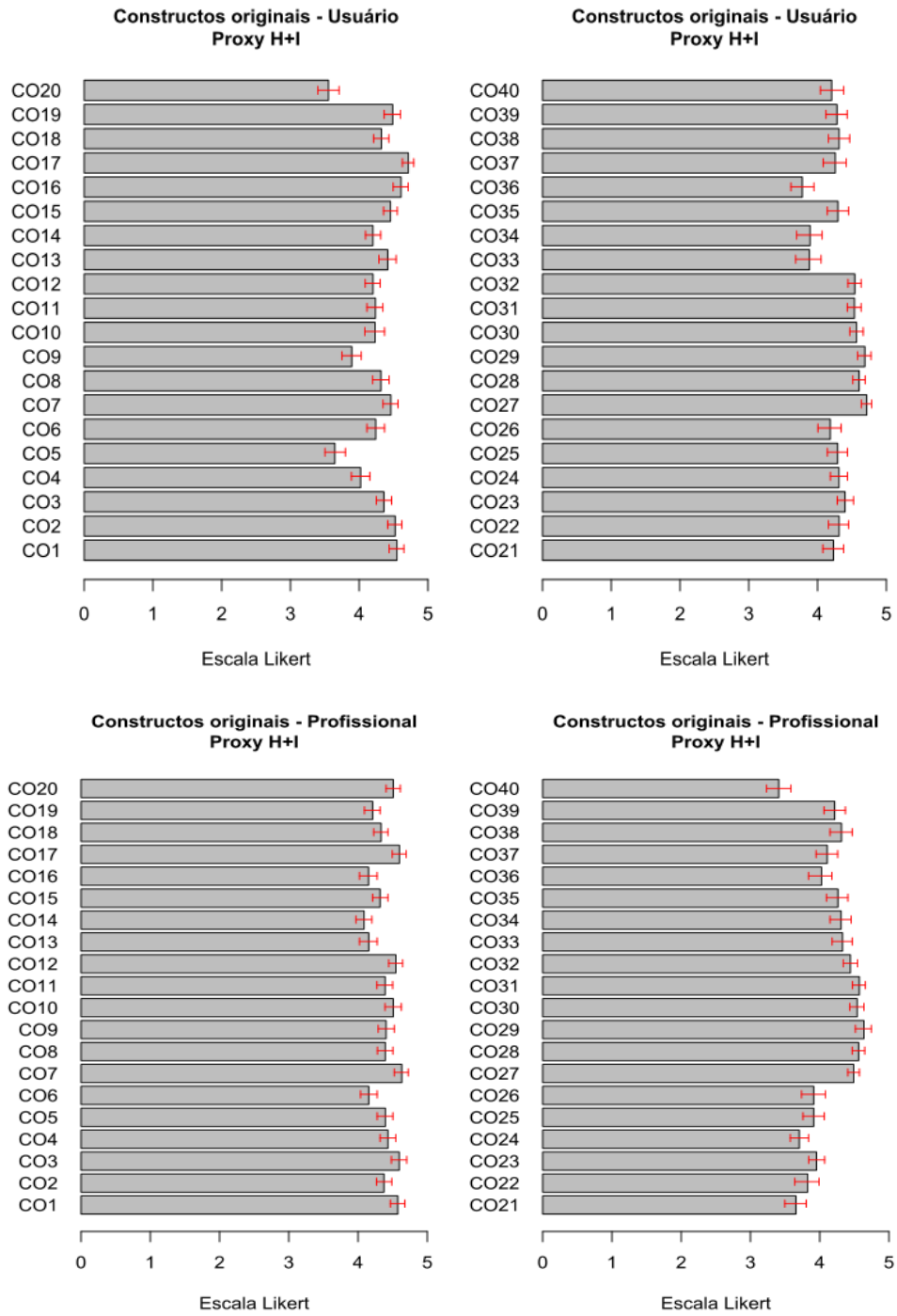


Figura 43- Gráfico de barras com intervalos de confiança para os constructos originais - *Proxy* H+I (ciclo 1, estudo 4) (Perspectiva indireta para Usuários e Profissionais).

Por sua vez, a análise das subdimensões de constructos de qualidade, estruturou-se a partir da valoração média atribuída aos constructos pelos indivíduos e calculadas para as subdimensões propostas pelos informantes.

A Tabela 24 apresenta uma análise descritiva das valorações médias obtidas pelas subdimensões de constructos, na *proxy* H+I (195 informantes no ciclo 1 de validação do instrumento, perspectiva indireta com atribuição de valorações por especialistas e profissionais, com observações para a perspectiva de Usuários e a dos Profissionais).

Tabela 24- Análise descritiva das subdimensões de constructos - *Proxy* H+I (ciclo 1, estudo 4) (Perspectiva indireta para Usuários e Profissionais).

Subdimensões de constructos	Usuário			Profissional		
	Média	D.P.	I.C-95%	Média	D.P.	I.C-95%
Equipamentos	4,22	0,59	[4,13; 4,30]	4,47	0,60	[4,38; 4,56]
Estrutura Física	4,34	0,78	[4,22; 4,44]	4,39	0,70	[4,29; 4,48]
Apresentação Pessoal	4,06	0,90	[3,94; 4,19]	4,46	0,74	[4,35; 4,56]
Instalações Físicas	4,22	0,71	[4,13; 4,31]	4,47	0,71	[4,38; 4,57]
Confiança em Prazos	4,42	0,91	[4,28; 4,53]	4,15	0,93	[4,03; 4,28]
Suporte ao usuário-cliente	4,42	0,62	[4,34; 4,52]	4,19	0,69	[4,09; 4,27]
Confiabilidade Organizacional	4,72	0,57	[4,63; 4,79]	4,60	0,70	[4,50; 4,69]
Fornecimento de Serviços	4,41	0,74	[4,30; 4,51]	4,27	0,73	[4,17; 4,37]
Garantia dos Registros	3,55	1,08	[3,40; 3,71]	4,51	0,79	[4,40; 4,61]
Acesso Informação	4,27	1,02	[4,13; 4,41]	3,74	1,10	[3,57; 3,89]
Pronto Atendimento	4,35	0,81	[4,24; 4,46]	3,83	0,81	[3,72; 3,94]
Disposição p/ atendimento	4,29	1,08	[4,15; 4,44]	3,91	1,08	[3,75; 4,06]
Ocupação profissional	4,19	1,22	[4,02; 4,37]	3,91	1,23	[3,74; 4,09]
Competência profissional	4,72	0,56	[4,63; 4,80]	4,49	0,64	[4,40; 4,59]
Credibilidade Profissional	4,64	0,57	[4,56; 4,71]	4,59	0,60	[4,51; 4,67]
Cortesia nas Relações	4,55	0,58	[4,47; 4,62]	4,52	0,57	[4,43; 4,60]
Suprimento de Recursos	3,89	1,24	[3,70; 4,07]	4,32	1,04	[4,17; 4,47]
Comunicação e Atenção	4,30	1,11	[4,13; 4,45]	4,26	1,10	[4,10; 4,40]
Humanização no Atendimento	4,02	1,08	[3,86; 4,17]	4,07	1,07	[3,91; 4,22]
Compreensão do Cliente	4,31	1,08	[4,16; 4,46]	4,31	1,12	[4,15; 4,47]
Prioridade nos interesses	4,28	1,08	[4,12; 4,43]	4,22	1,12	[4,05; 4,38]
Acessibilidade dos clientes	4,21	1,20	[4,04; 4,36]	3,41	1,29	[3,22; 3,58]

Nota.¹ Intervalo Bootstrap

A análise global permite afirmar que a ampla maioria das 22 subdimensões de constructos obtiveram valoração média elevada (superior ou cerca de 4, muito importante/relevante), para as perspectivas de Usuários e a dos Profissionais, na análise do *proxy* H+I, sendo as 22 subdimensões consideradas relevantes/importantes para as próximas fases de investigação, na estruturação do instrumento de pesquisa de Usuários (*proxy* J, ciclo 2).

A figura 44, que sintetizou os dados da Tabela 24, mostra que:

- Na perspectiva de Usuários, houve diferença significativa da valoração média obtida somente em três subdimensões (Confiabilidade Organizacional, Competência profissional e Credibilidade Profissional), que formaram um grupo com valorações significativamente maior em relação às demais e também na subdimensão Garantia dos Registros com valoração significativamente menor em relação às demais, uma vez que todas as demais subdimensões tiveram intervalos de confiança que se sobrepuseram. Dessa forma, em média, os indivíduos tenderam a dar maiores valorações (importância/relevância) em três subdimensões (Confiabilidade Organizacional, Competência profissional e Credibilidade Profissional) e menor valoração (importância) para a subdimensão Garantia dos Registros, em relação às demais 18 subdimensões, nas quais os intervalos de confiança se sobrepuseram, indicando que não houve diferenças significativas nas valorações entre estas subdimensões para a perspectiva de Usuários.
- Na perspectiva dos Profissionais, não houve diferença significativa da valoração média obtida em todas as 22 subdimensões, uma vez que os seus intervalos de confiança se sobrepuseram. Dessa forma, em média, os indivíduos tenderam a dar valorações (importância/relevância) dentro das mesmas margens de valores na ampla maioria das 22 subdimensões de constructos para a perspectiva dos Profissionais, na *proxy* H+I.

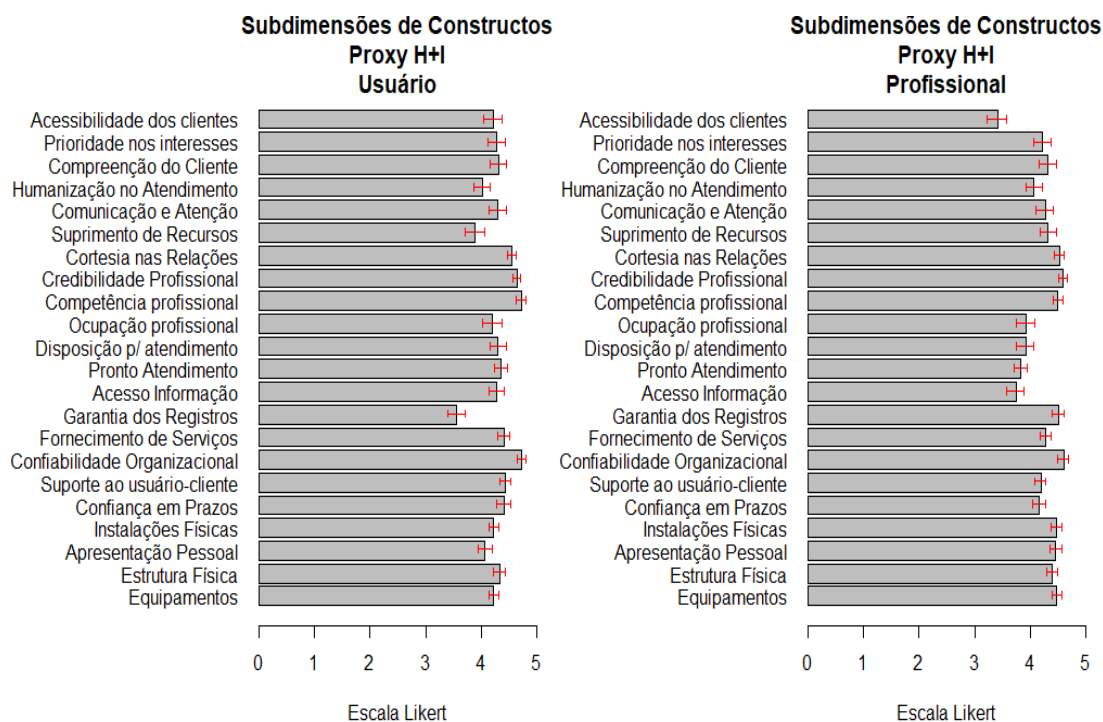


Figura 44- Gráfico de barras com intervalos de confiança para as subdimensões de constructos - *Proxy* H+I (ciclo 1, estudo 4) (Perspectiva indireta para Usuários e Profissionais).

Por sua vez, a análise das macrodimensões de qualidade, estruturou-se a partir da valoração média atribuída aos constructos pelos indivíduos e calculadas para as macrodimensões propostas pelos informantes.

A Tabela 25 apresenta uma análise descritiva das valorações médias obtidas pelas macrodimensões de qualidade, na *proxy* H+I (195 informantes no ciclo 1 de validação do instrumento, perspectiva indireta com atribuição de valorações por especialistas e profissionais, com observações para a perspectiva de Usuários e a dos Profissionais).

Tabela 25- Análise descritiva das macrodimensões de qualidade - *Proxy* H+I (ciclo 1, estudo 4) (Perspectiva indireta para Usuários e Profissionais).

Macrodimensões	Usuário			Profissional		
	Média	D.P.	I.C-95%	Média	D.P.	I.C-95%
Tangibilidade	4,22	0,55	[4,15; 4,30]	4,45	0,52	[4,38; 4,52]
Confiabilidade	4,35	0,54	[4,27; 4,42]	4,30	0,54	[4,22; 4,36]
Presteza	4,29	0,73	[4,20; 4,39]	3,83	0,71	[3,73; 3,93]
Segurança	4,43	0,48	[4,36; 4,50]	4,49	0,42	[4,43; 4,54]
Empatia	4,19	0,98	[4,06; 4,32]	4,06	0,95	[3,92; 4,19]

Nota.¹ Intervalo Bootstrap

A análise global permite afirmar que a ampla maioria das cinco subdimensões de qualidade (Tangibilidade, Confiabilidade, Presteza, Segurança e Empatia) obtiveram valoração média elevada (superior a 4, muito importante/relevante), na análise do *proxy* H+I, sendo todas as cinco macrodimensões consideradas relevantes/importantes para análise nas próximas fases de investigação, na estruturação do instrumento de questões validadas de Usuários (*proxy* J, ciclo 2). Pode ainda se destacar que:

- Na perspectiva de Usuários, não houve diferenças significativas nas valorações médias obtidas em todas as cinco macrodimensões de qualidade, uma vez que todos os seus intervalos de confiança se sobrepuseram. Dessa forma, em média, os indivíduos tenderam a dar valoração (importância/relevância) dentro das mesmas margens de valores em todas as cinco macrodimensões de constructos para a perspectiva de Usuários, na *proxy* H+I.
- Na perspectiva dos Profissionais, houve diferença significativa da valoração média obtida somente na macrodimensão Presteza, com valoração significativamente menor em relação às demais, uma vez que todas as demais quatro macrodimensões de qualidade tiveram intervalos de confiança que se sobrepuseram. Dessa forma, em média, os indivíduos tenderam a dar menor valoração (importância/relevância) para a macrodimensão Presteza, em relação às demais quatro macrodimensões (Tangibilidade, Confiabilidade, Segurança e Empatia), indicando que não houve diferenças

significativas nas valorações entre estas quatro macrodimensões de constructos para a perspectiva de Profissionais, na *proxy* H+I.

Desta análise descritiva sintetizou-se que os especialistas e profissionais atribuíram valoração média elevada (superior a 4, muito importante/relevante) para as subcategorias (40 constructos), as categorias (22 subdimensões de qualidade) e as macrocategorias (cinco macrodimensões de qualidade) dos itens de análise, sendo considerados muito relevantes/importantes para as próximas fases de investigação, na estruturação do instrumento de pesquisa de Usuários (*proxy* J, ciclo 2).

Quanto aos resultados do estudo das possíveis influências de três variáveis moderadoras, das análises das tabelas 68 a 73 (apêndice 9, tópico 9.2.1.1), sintetizou-se que houve tendência de influência da variável mediadora nível de atenção à saúde na maioria (≥ 12) das 22 subdimensões de constructos, para as perspectivas de Usuários (12 subdimensões) e dos Profissionais (13 subdimensões), que não houve tendência de influência da variável mediadora nível de complexidade da assistência na ampla maioria (≥ 14) das 22 subdimensões de constructos, para as perspectivas de Usuários (17 subdimensões) e dos Profissionais (14 subdimensões), e, que não houve tendência de influência da variável mediadora nível de cargo/função na ampla maioria (≥ 17) das 22 subdimensões de constructos nas perspectivas de Usuários (17 subdimensões) e dos Profissionais (20 subdimensões), na *proxy* H+I, que se propõe aprofundar investigações.

Por sua vez, das análises das tabelas 74 a 76 (apêndice 9, tópico 9.2.1.1), sintetizou-se que houve uma tendência de influência da variável mediadora nível de atenção à saúde, em todas as cinco macrodimensões de constructos (Tangibilidade, Confiabilidade, Presteza, Segurança e Empatia) para as perspectivas de Usuários e dos Profissionais. Por sua vez, não houve tendência de influência da variável mediadora nível de complexidade da assistência, na ampla maioria das cinco macrodimensões de constructos para a perspectiva de Usuários e a dos Profissionais, sendo exceções duas macrodimensões (Confiabilidade e Segurança), na perspectiva de Usuários. Também não houve tendência de influência da variável mediadora nível de cargo/função, na ampla maioria das cinco macrodimensões de constructos para a perspectiva de Usuários e a dos Profissionais, sendo exceções duas macrodimensões (Confiabilidade e Presteza), na perspectiva de Usuários, na *proxy* H+I, que se propõe aprofundar investigações.

Diante destes resultados, a comprovação da tendência das variáveis nível de atenção à saúde, funcionar enquanto fatores mediadores da ampla maioria das 22 subdimensões de

constructos propostas, além de todas as cinco macrodimensões de qualidade originais da escala SERVQUAL, trata-se de um achado inovador desta investigação. Os indícios dos dois fatores mediadores nível de complexidade da assistência e nível de escolaridade cargo/função funcionarem enquanto fatores mediadores da ampla maioria das 22 subdimensões de constructos propostas foram estudados e validados nas etapas posteriores (tópico 6.2.1.2).

Ressalte-se, ainda, que a validação das cinco macrodimensões de qualidade (macrocategorias de análise), na perspectiva de especialistas e profissionais, na *proxy* H+I, estão alinhadas com a validação destas mesmas macrodimensões da escala SERVQUAL, nos estudos de Babakus e Mangold (1992) no contexto de saúde, mas a validação da ampla maioria das 22 subdimensões de constructos propostas (categorias de análise) trata-se de um achado inovador desta investigação que serão objeto de aprofundamento da validação nas etapas posteriores.

Estes resultados subsidiaram a construção de um novo instrumento de questões validadas, para a validação de uma escala inovadora de medição de qualidade/satisfação de usuários de saúde, aplicado na fase posterior da investigação (*proxy* J, estudo 4).

6.2.1.2. Análise descritiva do estudo 4 (ciclo 2: *proxy* J) (constructos *versus* subdimensões e macrodimensões de qualidade) (Visão profissionais *versus* usuários)

A partir do instrumento consolidado nas fases 1 a 3 procedentes no ciclo 1 (*proxy* H+I), na fase 4, ciclo 2 do estudo 4, através de um novo instrumento (*proxy* J), aplicou-se uma análise da validação em um grupo piloto de 506 informantes (perspectiva direta de usuários da saúde). Assim, na fase 4, abordou-se a análise da *proxy* J (n = 506), na qual o instrumento manteve similaridades com o instrumento da fase anterior (*proxy* H+I) (40 constructos, 22 subdimensões e cinco macrodimensões de qualidade percebida). Assim, procedeu-se a validação definitiva de um novo instrumento (*proxy* J), por um grupo de 506 usuários entrevistados em 14 unidades de saúde da atenção primária e secundária do SUS.

A Tabela 26 abaixo apresenta uma análise descritiva das variáveis de caracterização selecionadas na amostra, na análise integrada da *proxy* J (506 informantes), que aplicou um único instrumento, no ciclo 2 de aplicação de campo, na visão de usuários de saúde.

Tabela 26- Análise descritiva das variáveis de caracterização selecionadas - *Proxy* J (ciclo 2, estudo 4) (Visão de Usuários).

Variáveis	N	%
Nível de Atenção à Saúde		
Atenção Primária	241	47,63%
Atenção Secundária	140	27,67%

	Atenção Terciária	125	24,70%
	Outros	0	0,00%
Nível Complexidade da assistência	Baixa	241	47,63%
	Média	265	52,37%
	Alta	0	0,00%
	Outros	0	0,00%
Escolaridade	Analfabeto	22	4,35%
	Ensino Básico	224	44,27%
	Ensino Fundamental	68	13,44%
	Ensino Médio	183	36,17%
	Ensino Superior	8	1,58%
	Pós-Graduação	1	0,20%
	Mestrado/ PhD	0	0,00%
Sexo	Masculino	134	26,53%
	Feminino	370	73,27%
	Outros	1	0,20%
Idade	Até 19 anos	4	0,79%
	20 a 29 anos	111	21,94%
	30 a 39 anos	121	23,91%
	40 a 49 anos	76	15,02%
	50 a 59 anos	72	14,23%
	60 a 69 anos	61	12,06%
	60 a 79 anos	36	7,11%
	80 anos ou mais	25	4,94%

Na fase 4 (*proxy J*), destaca-se que a amostra se mostrou diversificada (Tabela 26) no nível de atenção à saúde que os indivíduos estavam em atendimento: níveis Primário (47,63%), Secundário (27,67%) e Terciário (24,70%) de atenção à saúde. A maior parte dos indivíduos na amostra (52,37%) estava em atendimento em unidades de Média complexidade, enquanto 47,63% estava em atendimento em unidades de Baixa complexidade. Não houve entrevista em unidades de Alta complexidade da assistência, por exemplo, em hospitais de grande porte. Considerou-se a amostra ($n = 506$) razoavelmente representativa da realidade dos usuários do SUS, visto que a maior parte dos indivíduos (48,62%) possuía até o ensino basilar (analfabeto e ensino básico), seguidos por boa parte (39,01%) que possuía ensino médio e por 13,44% de indivíduos que possuíam ensino fundamental, e, ainda, a maioria dos indivíduos (73,27%) era do sexo Feminino - desequilibrada por histórica demanda reduzidas de homens-, a faixa de idade mais frequente entre os indivíduos foi a de 30 a 39 anos (23,91%), seguida por 20 a 29 anos (21,94%) e 40 a 49 anos (15,02%).

A Tabela 27 apresenta uma análise descritiva relativa à valoração média atribuída aos constructos de qualidade, na análise da *proxy J* (506 informantes), no ciclo 2 de aplicação de campo, na visão de usuários de saúde.

Tabela 27- Análise descritiva dos constructos de qualidade – *Proxy J* (ciclo 2, estudo 4) (Visão de Usuários).

Subdimensões de constructos	Constructos originais	Média	D.P.	I.C-95% ¹
Equipamentos	CO1_Disponib	3,32	0,59	[3,27; 3,37]
	CO2_Acesso	3,33	0,59	[3,28; 3,38]
	CO3_Resolut	3,33	0,60	[3,27; 3,37]
	CO4_Conserv	3,23	0,66	[3,17; 3,28]
	CO5_Usabilid	3,30	0,59	[3,25; 3,35]
Estrutura Física	CO6_Conforto	3,34	0,66	[3,27; 3,39]
	CO7_Limpeza	3,52	0,55	[3,47; 3,56]
	CO8_Bem	3,33	0,67	[3,27; 3,39]
Apresentação Pessoal	CO9_Vestim	3,38	0,65	[3,32; 3,44]
	CO10_Visual	3,41	0,62	[3,36; 3,46]
Instalações Físicas	CO11_Conserv	3,35	0,61	[3,30; 3,40]
	CO12_Funcion	3,38	0,58	[3,33; 3,44]
Confiança em Prazos	CO13_Cumprim.Prazos	2,90	0,56	[2,85; 2,95]
Suporte ao usuário-cliente	CO14_Solidaried	3,16	0,37	[3,13; 3,19]
	CO15_Apoio_U	3,16	0,37	[3,13; 3,20]
	CO16_Acolhim	3,17	0,37	[3,14; 3,20]
Confiabilidade Organizacional	CO17_Confianca.Percep	3,52	0,55	[3,47; 3,56]
Fornecimento de Serviços	CO18_Prazo.Resolut	3,37	0,53	[3,32; 3,41]
	CO19_Prazo.Atendim	2,90	0,59	[2,85; 2,95]
Garantia dos Registros	CO20_Inform.Disp.Contr	3,11	0,38	[3,08; 3,15]
Acesso Informação	CO21_Esclarecer	3,14	0,35	[3,11; 3,17]
	CO22_Infor	3,15	0,36	[3,12; 3,18]
Pronto Atendimento	CO23_Atendim	3,00	0,46	[2,96; 3,04]
	CO24_Disponib	3,06	0,38	[3,03; 3,09]
Disposição atendimento	p/ CO25_Dispos.Assistenc	3,13	0,34	[3,10; 3,16]
Ocupação profissional	CO26_Profiss.Suficiente	2,95	0,29	[2,92; 2,97]
Competência profissional	CO27_Conf.Compet.Profiss	3,28	0,48	[3,24; 3,33]
Credibilidade Profissional	CO28_Confianca	3,37	0,51	[3,32; 3,41]
	CO29_Credib	3,34	0,51	[3,30; 3,39]
Cortesia nas Relações	CO30_Cortesia	3,23	0,43	[3,20; 3,27]
	CO31_Respeito	3,23	0,43	[3,20; 3,27]
	CO32_Atitude	3,25	0,44	[3,21; 3,29]
Suprimento de Recursos	CO33_Garantia	2,91	0,33	[2,88; 2,93]
	CO34_Suprim	2,91	0,34	[2,88; 2,93]
Comunicação Atensão	e CO35_Atendim.Indiv	3,11	0,32	[3,08; 3,14]
Humanização Atendimento	CO36_Dedicação	3,16	0,36	[3,12; 3,19]
	CO37_Tratam	3,15	0,36	[3,12; 3,18]

Compreensão Cliente	do	CO38_Conhec.Situac.Usuari o	3,13	0,36	[3,10; 3,16]
Prioridade interesses	nos	CO39_Objct.Centrado.Usuar io	3,07	0,26	[3,05; 3,09]
Acessibilidade clientes	dos	CO40_Facil.Acesso.Usuario	2,99	0,34	[2,96; 3,02]

Nota.¹ Intervalo Bootstrap

A análise global permite afirmar que a ampla maioria dos 40 constructos de qualidade obtiveram valoração média elevada (superior ou cerca de 3, importante/relevante), na análise da *proxy* J, sendo os 40 constructos considerados relevantes/importantes para as próximas fases de investigação, na estruturação do instrumento de pesquisa de Usuários (*proxy* K, ciclo 1 do estudo 5).

A figura 45, que sintetizou os dados da Tabela 27, mostra que na ampla maioria dos 40 constructos de qualidade (CO1 a CO40) houve diferença significativa da valoração média obtida somente em seis constructos (*CO13_Cumprim.Prazos*, *CO19_Prazo.Atendim*, *CO26_Profiss.Suficiente*, *CO33_Garantia Condic.*, *CO34_Suprim Recursos* e *CO40_Facil.Acesso.Usuario*), que formaram um grupo com valorações significativamente menor em relação aos demais, como também em dois constructos (*CO7_Limpeza* e *CO17_Confianca.Percep*) que formaram um grupo com valorações significativamente maior em relação aos demais, uma vez que todos os demais 32 constructos tiveram intervalos de confiança que se sobrepuseram, na *proxy* J (506 informantes, visão de Usuários).

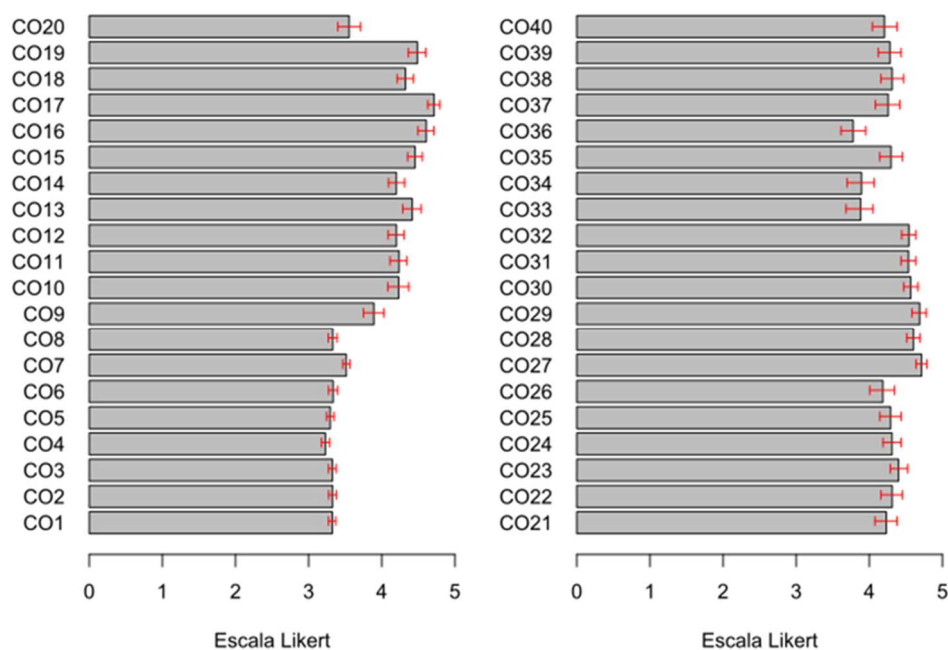


Figura 45- Gráfico de barras com intervalos de confiança para os constructos de qualidade - Proxy J (ciclo 2, estudo 4) (Visão de Usuários).

Por sua vez, a análise das subdimensões de constructos de qualidade, estruturou-se a partir da valoração média atribuída aos constructos pelos indivíduos e calculadas para as subdimensões propostas pelos informantes.

A Tabela 28 apresenta uma análise descritiva das valorações médias obtidas pelas subdimensões de constructos, na análise da proxy J (506 informantes), no ciclo 2 de aplicação de campo, na visão de usuários de saúde.

Tabela 28- Análise descritiva das subdimensões de constructos – Proxy J (ciclo 2, estudo 4) (Visão de Usuários).

Subdimensões de constructos	Média	D.P.	I.C-95% ¹
Equipamentos	3,30	0,58	[3,25; 3,35]
Estrutura Física	3,39	0,58	[3,34; 3,44]
Apresentação Pessoal	3,40	0,63	[3,34; 3,45]
Instalações Físicas	3,37	0,59	[3,32; 3,42]
Confiança em Prazos	2,90	0,56	[2,85; 2,95]
Suporte ao usuário-cliente	3,16	0,37	[3,13; 3,20]
Confiabilidade Organizacional	3,52	0,55	[3,47; 3,56]
Fornecimento de Serviços	3,14	0,42	[3,10; 3,18]
Garantia dos Registros	3,11	0,38	[3,08; 3,15]
Acesso Informação	3,15	0,35	[3,12; 3,18]
Pronto Atendimento	3,03	0,40	[3,00; 3,06]
Disposição p/ atendimento	3,13	0,34	[3,10; 3,16]
Ocupação profissional	2,95	0,29	[2,92; 2,97]
Competência profissional	3,28	0,48	[3,24; 3,33]
Credibilidade Profissional	3,34	0,51	[3,30; 3,39]
Cortesia nas Relações	3,24	0,43	[3,20; 3,28]

Suprimento de Recursos	2,91	0,33	[2,88; 2,93]
Comunicação e Atenção	3,11	0,32	[3,08; 3,14]
Humanização no Atendimento	3,15	0,36	[3,12; 3,18]
Compreensão do Cliente	3,13	0,36	[3,10; 3,16]
Prioridade nos interesses	3,07	0,26	[3,05; 3,10]
Acessibilidade dos clientes	2,99	0,34	[2,96; 3,02]

Nota.¹ Intervalo Bootstrap

A análise global permite afirmar que a ampla maioria (18) das 22 subdimensões de constructos obtiveram valoração média elevada (superior a 3, importante/relevante), na análise da *proxy* J, sendo as 22 subdimensões consideradas relevantes/importantes para as próximas fases de investigação, na estruturação do instrumento de pesquisa de Usuários (*proxy* K, ciclo 1 do estudo 5).

A figura 46, que sintetizou os dados da Tabela 28, mostra que na ampla maioria das 22 subdimensões de constructos avaliadas para a visão de Usuários, houve diferença significativa da valoração média obtida somente em três subdimensões (Suprimentos de Recursos, Ocupação Profissional e Confiança em Prazos), que formaram um grupo com valorações significativamente menor em relação às demais, como também na subdimensão Confiabilidade Organizacional com valoração significativamente maior em relação às demais, uma vez que todas as demais 18 subdimensões tiveram intervalos de confiança que se sobrepuseram.

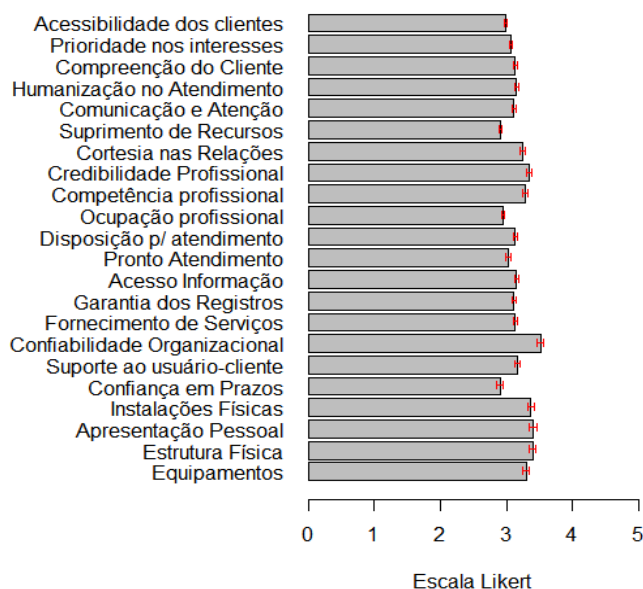


Figura 46- Gráfico de barras com intervalos de confiança para as subdimensões de constructos – *Proxy* J (ciclo 2, estudo 4) (Visão de Usuários).

Por sua vez, a análise das macrodimensões de qualidade, estruturou-se a partir da valoração média atribuída aos constructos pelos indivíduos e calculadas para as macrodimensões propostas pelos informantes.

A Tabela 29 apresenta uma análise descritiva das valorações médias obtidas pelas macrodimensões de qualidade, na análise da *proxy J* (506 informantes), no ciclo 2 de aplicação de campo, na visão de usuários de saúde.

Tabela 29- Análise descritiva das macrodimensões de constructos – *Proxy J* (ciclo 2, estudo 4) (Visão de Usuários).

Macrodimensões	Média	D.P.	I.C-95% ¹
Tangibilidade	3,35	0,50	[3,31; 3,39]
Confiabilidade	3,16	0,28	[3,14; 3,19]
Presteza	3,07	0,24	[3,05; 3,09]
Segurança	3,19	0,31	[3,16; 3,21]
Empatia	3,10	0,24	[3,08; 3,12]

Nota.¹ Intervalo Bootstrap

A análise global permite afirmar que todas as cinco macrodimensões de constructos (Tangibilidade, Confiabilidade, Presteza, Segurança e Empatia) obtiveram valoração média elevada (superior a 3, importante/relevante), na análise da *proxy J*, e, foram consideradas relevantes/importantes para análise nas próximas fases de investigação, na estruturação do instrumento de questões validadas de Usuários (*proxy K*, ciclo 1 do estudo 5). Pode-se ainda se destacar que houve diferença significativa da valoração média obtida nas macrodimensões Tangibilidade (com valoração significativamente maior) e Presteza – marginalmente (com valoração significativamente menor), em relação às demais macrodimensões de constructos, uma vez que as demais três macrodimensões tiveram intervalos de confiança que se sobrepuseram. Dessa forma, em média, os indivíduos tenderam a dar maior valoração (importância/relevância) para a macrodimensão Tangibilidade e menor valoração para Presteza, em relação às demais macrodimensões de constructos (Confiabilidade, Segurança e Empatia), na *proxy J* (506 informantes, visão de Usuários).

Desta análise descritiva sintetizou-se que os usuários de saúde atribuíram elevados graus de importâncias/relevâncias (média superior ou próxima de 3) para as subcategorias (40 constructos), as categorias (22 subdimensões de qualidade) e as macrocategorias (cinco macrodimensões de qualidade) dos itens de análise, sendo considerados relevantes/importantes para as próximas fases de investigação, na estruturação de um instrumento com questões de qualidade (*proxy K*, ciclo 1 do estudo 5).

Quanto aos resultados do estudo das possíveis influências de três variáveis moderadoras, das análises das tabelas 77 a 79 (apêndice 9, tópico 9.2.1.2), sintetizou-se que, na ampla maioria das 22 subdimensões de constructos, houve uma tendência de influência das variáveis mediadoras nível de atenção à saúde (18 subdimensões), nível de complexidade da assistência

(15 subdimensões) e nível de escolaridade de Usuários (14 subdimensões), na *proxy* J (506 informantes, visão de Usuários).

Por sua vez, das análises das tabelas 80 a 82 (apêndice 9, tópico 9.2.1.2), sintetizou-se que houve diferença significativa quanto as variáveis nível de atenção à saúde, nível de complexidade da assistência e nível de escolaridade de Usuários, para se considerar enquanto fatores mediadores em todas as macrodimensões de constructos (Tangibilidade, Confiabilidade, Presteza, Segurança e Empatia), para a perspectiva de Usuários, na *proxy* J. No entanto, quanto ao nível de complexidade esta tendência deve ser melhor investigada, por exemplo, com agregação de entrevistas em unidades de Alta complexidade.

Diante destes resultados, a comprovação da tendência das variáveis nível de atenção à saúde, nível de complexidade da assistência e nível de escolaridade cargo/função funcionarem enquanto fatores mediadores da ampla maioria das 22 subdimensões de constructos propostas, além de todas as cinco macrodimensões de qualidade originais da escala SERVQUAL, trata-se de um achado inovador desta investigação.

Ressalte-se, ainda, que a validação das cinco macrodimensões de qualidade (macrocategorias de análise), na perspectivas de usuários, na *proxy* J, estão alinhadas com a validação destas mesmas macrodimensões da escala SERVQUAL, nos estudos de Babakus e Mangold (1992) no contexto de saúde, mas a validação da ampla maioria das 22 subdimensões de constructos propostas (categorias de análise) trata-se de um achado inovador desta investigação que serão objeto de aprofundamento da validação pela análise inferencial (tópico 6.2.2.1).

6.2.1.3. Análise descritiva do estudo 5 (ciclo 1: *proxies* K e L) (questões validadas *versus* subdimensões e macrodimensões de qualidade) (perspectiva de expectativa *versus* percepção).

A partir dos conhecimentos prévios de especialistas e profissionais de saúde, e dos resultados obtidos no estudo 4 (*proxies* H+I e J) desta investigação, com cinco macrodimensões originais da escala SERVQUAL e para cada uma de suas 22 questões originais uma subdimensão de constructos que foram desagregadas em 40 constructos de qualidade, estruturou-se o processo colaborativo de melhoria incremental da tradução transcultural e validação de uma escala inovadora de medição de qualidade/satisfação de usuários de saúde a fim de garantir proximidade à realidade do SUS e entendimento adequada dos itens de medição

de qualidade/satisfação propostos nesta nova fase da investigação. Por sua vez, no estudo 5, manteve-se as cinco macrodimensões originais da escala SERVQUAL e agregou-se outra inovadora (Acessibilidade) que se revelou de muita importância no SUS. Foram estruturadas 13 subdimensões de qualidade por similaridades entre as 22 subdimensões de constructos do estudo 4. A partir dos 40 constructos procedentes do estudo 4, estruturou-se 20 questões validadas da escala inovadora de medição de qualidade/satisfação, com posterior aplicação.

A parte empírica do estudo 5 foi desenvolvida a partir do instrumento consolidado nas diversas fases subsequentes procedentes no ciclo 1 (*proxy* K+L), através de um grupo piloto de respondentes (n = 140), posteriormente, com estruturação e validação de um novo instrumento ampliado (*proxy* M), aplicado em uma pesquisa de campo (ciclo 2) em única fase incremental (n = 506), e, por fim, com o instrumento (*proxy* M) validado por usuários, aplicou-se uma pesquisa de campo complementar (*proxy* N), com foco na medição da qualidade/satisfação de uma rede de atendimento do SUS (n= 2.041), com análise final integrada das respostas obtidas (*proxies* M e N).

Inicialmente, na *proxy* K (fase 1), aplicou-se uma análise dos conhecimentos genéricos de 78 especialistas, enquanto validação de face dos conteúdos propostos e consolidados nesta fase. Os respondentes, além de atribuir valoração nos itens de avaliação instrumento da *proxy* K (20 questões validadas, 13 subdimensões e seis macrodimensões de qualidade percebida), propuseram uma importante adequação incremental, somente de conteúdos das questões validadas, na transição do instrumento inicial da *proxy* K para o instrumento aprimorado da *proxy* L, mas não houve eliminações de questões validadas, subdimensões ou macrodimensões de qualidade percebida para estruturar a *proxy* L (20 questões validadas, 13 subdimensões e seis macrodimensões de qualidade percebida).

Em seguida, efetivou-se uma análise das respostas obtidas em uma amostra de 62 especialistas, profissionais e gestores de saúde, a fim de consolidar os níveis de conhecimentos específicos iniciais sobre as questões validadas, subdimensões e macrodimensões de qualidade percebida, visto que estas respostas tinham melhor avaliabilidade no novo instrumento (*proxy* L). Dessa forma, no ciclo 1 de validação do instrumento, abordou-se uma análise das respostas obtidas na fase 2 (*proxy* L) (20 questões validadas, 13 subdimensões e seis macrodimensões de qualidade percebida).

A Tabela 30 abaixo apresenta uma análise descritiva das variáveis de caracterização selecionadas na amostra, na análise das *proxies* K e L (140 informantes) que aplicou dois instrumentos com muitas similaridades, no ciclo 1 de validação do instrumento, na visão de especialistas e profissionais.

Tabela 30- Análise descritiva das variáveis de caracterização selecionadas - Proxy K+L - (ciclo 1, estudo 5)
(Perspectiva indireta para Usuários).

Variáveis		Proxy K (N)	Proxy L (N)	Proxy K+L (N)	Proxy K+L (%)
Nível de Atenção à Saúde	Atenção Primária	23	8	31	24,60%
	Atenção Secundária	20	10	30	23,81%
	Atenção Terciária	11	10	21	16,67%
	Outros	11	33	44	34,92%
Nível Complexidade da assistência	Baixa	18	8	26	20,47%
	Média	26	10	36	28,35%
	Alta	15	11	26	20,47%
	Outros	7	32	39	30,71%
Nível Cargo/Função	Sistêmico	4	18	22	15,83%
	Organizacional	24	8	32	23,02%
	Gerencia de Equipe	5	10	15	10,79%
	Outros	45	25	70	50,36%
Escolaridade	Analfabeto	0	0	0	0,00%
	Ensino Básico	0	0	0	0,00%
	Ensino Fundamental	0	0	0	0,00%
	Ensino Médio	7	1	8	5,84%
	Ensino Superior	50	7	57	41,61%
	Pós-Graduação	6	44	50	36,50%
	Mestrado/ PhD	12	10	22	16,06%
Tempo de experiência	Não tem	29	0	29	16,76%
	Menos de 6 meses	15	3	18	10,40%
	6 - 12 meses	14	2	16	9,25%
	13 - 24 meses	1	4	5	2,89%
	25 - 36 meses	2	6	8	4,62%
	37 - 48 meses	1	5	6	3,47%
	49 - 60meses	16	6	22	12,72%
	> 60 meses	34	35	69	39,88%
Tempo em cargo de decisão	Não tem	0	14	14	13,59%
	Menos de 6 meses	26	13	39	37,86%
	6 - 12 meses	2	8	10	9,71%
	13 - 24 meses	1	2	3	2,91%
	25 - 36 meses	2	8	10	9,71%
	37 - 48 meses	2	3	5	4,85%
	49 - 60meses	3	3	6	5,83%
	> 60 meses	8	8	16	15,53%
Formação/Profissão	Assistente Social	3		3	2,21%
	Enfo Assistencial	6		6	4,51%
	Assessor Qualidade	11		11	8,66%
	Analista Sistema	17		17	14,66%

Assist. Adm.	22	3	25	25,25%
Chefia Farm.		19	19	25,68%
Administrador		7	7	12,73%
Analista Sistema		4	4	8,33%
Assessor		3	3	6,82%
Outros	19	22	41	100,00%

Nas fases 1 e 2 (*proxy* K+L), destaca-se que a amostra se mostrou balanceada (Tabela 30) no nível de atenção à saúde que os indivíduos trabalhavam nos seguintes níveis: Outros (34,92%) (atuação em estruturas não assistenciais, por exemplo, em níveis de governança), Primário (24,60%), Secundário (23,81%) e Terciário (16,67%) de atenção à saúde. Grande parte dos indivíduos (50,36%) exercia o nível de cargo/função Outros (por exemplo, assessores de gestão, profissionais em geral etc) e boa parte (23,02%) exercia o nível de cargo/função Organizacional, seguidos pelos indivíduos do cargo/função Sistêmico (15,83%) e dos que exerciam cargo/função Gerencia de Equipe (10,79%). Considerou-se a amostra (n = 140) bastante qualificada visto que a maior parte dos indivíduos (36,50%) possuía pós-graduação e (16,06%) possuía Mestrado/PhD e ainda 94,16% possuía nível superior, a maior parte dos indivíduos (39,88%) tinha mais de 60 meses de experiência, enquanto que os indivíduos com menos de 6 meses de experiência somaram 27,16% da amostra. Uma boa parte dos indivíduos (15,53%) trabalhava a mais de 60 meses em cargo/função de decisão enquanto apenas 13,59% não possuía experiência em cargo/função de decisão.

Na *proxy* K (20 questões validadas, 13 subdimensões e seis macrodimensões de qualidade percebida), os respondentes propuseram uma importante adequação incremental, somente de conteúdos das questões validadas, tendo-se considerado esta fase como validação de face, sem análise dos valores atribuídos.

A Tabela 31 apresenta uma análise descritiva relativa à valoração média atribuída às questões validadas de qualidade, na análise da *proxy* L (62 informantes), no ciclo 1 de validação do instrumento, perspectiva indireta com atribuição de valorações por especialistas e profissionais, com observações para a perspectiva de Expectativa e Percepção do Usuário.

Tabela 31- Análise descritiva das questões validadas – *Proxy* L (ciclo 1, estudo 5) - Perspectiva indireta para Usuários (Expectativa e Percepção).

Subdimensões de questões validadas	Questões validadas	Expectativa			Percepção		
		Média	D.P.	I.C-95% ¹	Média	D.P.	I.C-95% ¹
Estrutura Física	SDQ01CV1	3,52	0,82	[3,29; 3,71]	2,98	0,98	[2,74; 3,23]
	SDQ01CV2	3,84	0,92	[3,59; 4,07]	2,88	0,96	[2,65; 3,13]
Apresentação Pessoal	SDQ03CV3	3,71	0,73	[3,52; 3,87]	3,03	0,77	[2,84; 3,21]
Equipamentos	SDQ04CV4	3,68	0,87	[3,45; 3,90]	3,00	0,90	[2,78; 3,23]

Instalações Físicas	SDQ05CV5	3,65	0,73	[3,47; 3,83]	2,98	0,93	[2,77; 3,23]
Confiabilidade Organizacional	SDQ06CV6	3,32	0,74	[3,13; 3,50]	3,25	0,79	[3,05; 3,44]
Fornecimento de Serviços	SDQ07CV7	2,76	1,00	[2,52; 3,02]	2,55	1,10	[2,29; 2,84]
	SDQ07CV8	3,16	0,99	[2,93; 3,39]	2,83	0,97	[2,59; 3,07]
Confiança na Competência Profissional	SDQ09CV9	3,53	0,91	[3,33; 3,75]	3,26	0,93	[3,03; 3,49]
	SDQ09CV10	3,44	0,85	[3,25; 3,64]	3,25	0,83	[3,05; 3,44]
Segurança na resolução do problema	SDQ17CV17	3,59	0,88	[3,38; 3,80]	3,10	0,89	[2,89; 3,33]
Atenção e Disposição p/ Atendimento	SDQ11CV11	3,28	0,93	[3,05; 3,49]	2,97	0,93	[2,74; 3,20]
	SDQ11CV12	3,16	0,78	[2,97; 3,35]	2,85	0,85	[2,66; 3,07]
Acesso a Informações	SDQ15CV15	3,18	0,96	[2,95; 3,41]	2,75	0,93	[2,53; 2,98]
	SDQ15CV16	3,38	0,80	[3,20; 3,57]	2,92	0,86	[2,71; 3,12]
Humanização no Atendimento	SDQ13CV13	3,13	0,89	[2,90; 3,36]	2,95	1,01	[2,71; 3,21]
	SDQ13CV14	3,25	1,06	[2,98; 3,51]	2,92	0,97	[2,69; 3,16]
Prioridade nos interesses do cliente	SDQ18CV18	3,14	0,90	[2,93; 3,37]	2,95	0,96	[2,71; 3,19]
Acessibilidade aos Serviços	SDQ19CV19	-	-	-	2,43	0,90	[2,19; 2,69]
Satisfação Global	SDQ20CV20	-	-	-	2,86	0,78	[2,66; 3,06]

Nota.¹ Intervalo Bootstrap

A análise global permite afirmar que, na perspectiva da Expectativa, a ampla maioria das 20 questões validadas analisadas (SDQ07CV19 - Acessibilidade aos Serviços e SDQ07CV20 - Satisfação Global não foram mensuradas quanto a Expectativa) que, na perspectiva da percepção, obtiveram valoração média mediana (superior a 3, importante/relevante). Dessa forma, pelas elevadas valorações médias atribuídas, todas as 20 questões validadas foram consideradas relevantes na estruturação do instrumento de usuários (*proxy* M, ciclo 1).

A figura 47, que sintetizou os dados da Tabela 31, mostra que na ampla maioria das 20 questões validadas analisadas (SDQ07CV19 - Acessibilidade aos Serviços e SDQ07CV20 - Satisfação Global não foram mensuradas quanto a Expectativa) houve variabilidade na atribuição de valorações de Expectativa e de Percepção, sendo que os seus intervalos de confiança se sobrepuseram, indicando que não houve diferenças significativas nas valorações atribuídas para as questões dentro da perspectiva de Expectativa e para as questões dentro da perspectiva de percepção. Dessa forma, em média, os indivíduos tenderam a dar valorações (importância/relevância) dentro das mesmas margens de valores na ampla maioria das 20 questões validadas nas perspectivas de Expectativa e de Percepção, na *proxy* L (62 informantes, perspectiva indireta para a visão de Usuários).

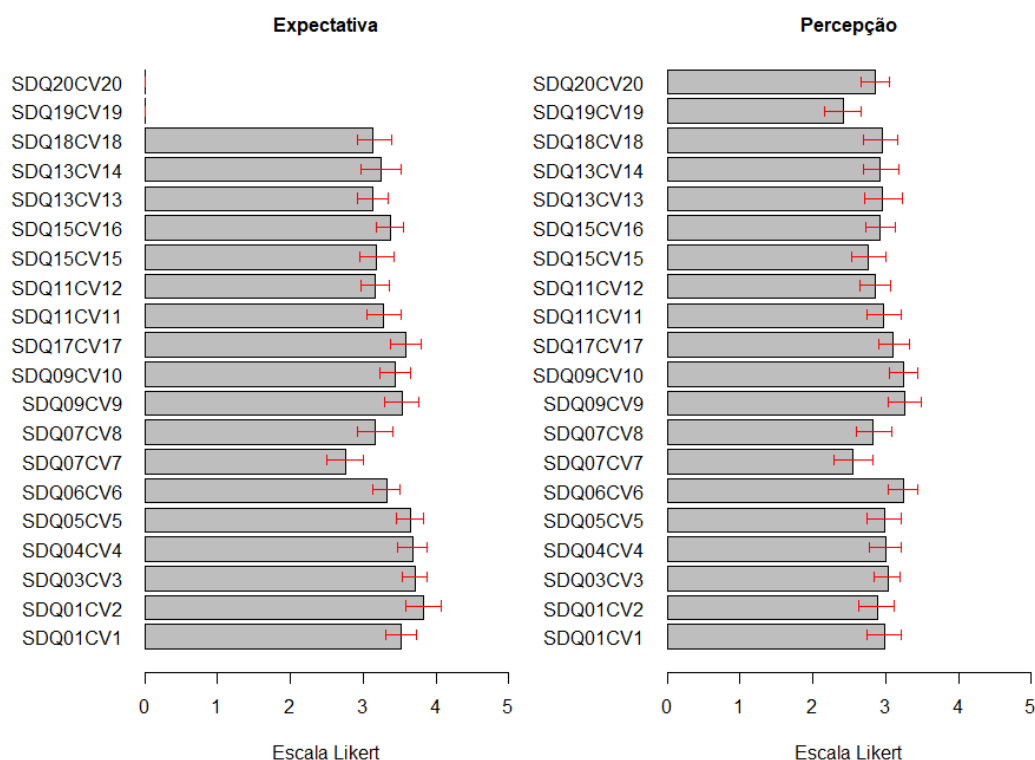


Figura 47 - Gráfico de barras com intervalos de confiança para questões validadas - *Proxy L* (ciclo 1, estudo 5) - Perspectiva indireta para Usuários (Expectativa e Percepção).

Por sua vez, a análise das subdimensões de questões validadas de qualidade, estruturou-se a partir da valoração média atribuída às questões validadas pelos indivíduos e calculadas para as subdimensões propostas pelos informantes.

A Tabela 32 apresenta uma análise descritiva das valorações médias obtidas pelas subdimensões de questões validadas, na *proxy L* (62 informantes no ciclo 1 de validação do instrumento, perspectiva indireta com atribuição de valorações por especialistas e profissionais, com observações para as perspectivas de Expectativa e de Percepção de Usuários.

Tabela 32- Análise descritiva das subdimensões de questões validadas – *Proxy L* (ciclo 1, estudo 5) - Perspectiva indireta para Usuários (Expectativa e Percepção).

Subdimensões de questões validadas	Expectativa			Percepção		
	Média	D.P.	I.C-95% ¹	Média	D.P.	I.C-95% ¹
Estrutura Física	3,67	0,79	[3,47; 3,86]	2,94	0,90	[2,72; 3,16]
Apresentação Pessoal	3,71	0,73	[3,53; 3,87]	3,03	0,77	[2,84; 3,23]
Equipamentos	3,68	0,87	[3,47; 3,92]	3,00	0,90	[2,78; 3,23]
Instalações Físicas	3,65	0,73	[3,47; 3,83]	2,98	0,93	[2,75; 3,23]
Confiabilidade Organizacional	3,32	0,74	[3,15; 3,50]	3,25	0,79	[3,05; 3,44]
Fornecimento de Serviços	2,96	0,50	[2,84; 3,09]	2,69	0,94	[2,48; 2,94]
Confiança na Competência Profissional	3,48	0,83	[3,29; 3,69]	3,25	0,83	[3,05; 3,44]
Segurança na resolução do problema	3,59	0,88	[3,38; 3,80]	3,10	0,89	[2,87; 3,31]
Atenção e Disposição p/ Atendimento	3,22	0,80	[3,03; 3,43]	2,91	0,84	[2,71; 3,12]
Acesso a Informações	3,28	0,82	[3,07; 3,49]	2,84	0,82	[2,64; 3,05]

Humanização no Atendimento	3,19	0,91	[2,97; 3,39]	2,93	0,94	[2,70; 3,16]
Prioridade nos interesses do cliente	3,14	0,90	[2,92; 3,37]	2,95	0,96	[2,71; 3,17]
Acessibilidade aos Serviços	-	-	-	2,43	0,90	[2,19; 2,67]
Satisfação Global	-	-	-	2,86	0,78	[2,66; 3,06]

Nota.¹ Intervalo Bootstrap

A análise global permite afirmar que ampla maioria das 13 subdimensões de questões validadas analisadas (Acessibilidade aos Serviços e Satisfação Global não foram mensuradas quanto a Expectativa), na perspectiva da Expectativa, obtiveram valoração média elevada (superior a 3, importante/relevante) e, na perspectiva da percepção, obtiveram valoração média mediana (próximo de 3). Dessa forma, pelas elevadas valorações médias obtidas, todas as 13 subdimensões de questões validadas foram consideradas relevantes, na estruturação do instrumento de pesquisa de Usuários (*proxy* M, ciclo 2).

A figura 48, que sintetizou os dados da Tabela 32, mostra que na ampla maioria das 13 subdimensões de questões validadas analisadas (Acessibilidade aos Serviços e Satisfação Global não foram mensuradas quanto a Expectativa), houve variabilidade nas valorações obtidas para a Expectativa e para a Percepção, sendo que os seus intervalos de confiança se sobrepuseram, indicando que não houve diferenças significativas nas valorações obtidas para as subdimensões dentro da perspectiva de Expectativa e para as subdimensões dentro da perspectiva de Percepção. Dessa forma, em média, os indivíduos tenderam a dar valorações (importância/relevância) dentro das mesmas margens de valores em todas as 13 subdimensões de questões validadas nas perspectivas de Expectativa e de Percepção, na *proxy* L (62 informantes, perspectiva indireta para a visão de Usuários).

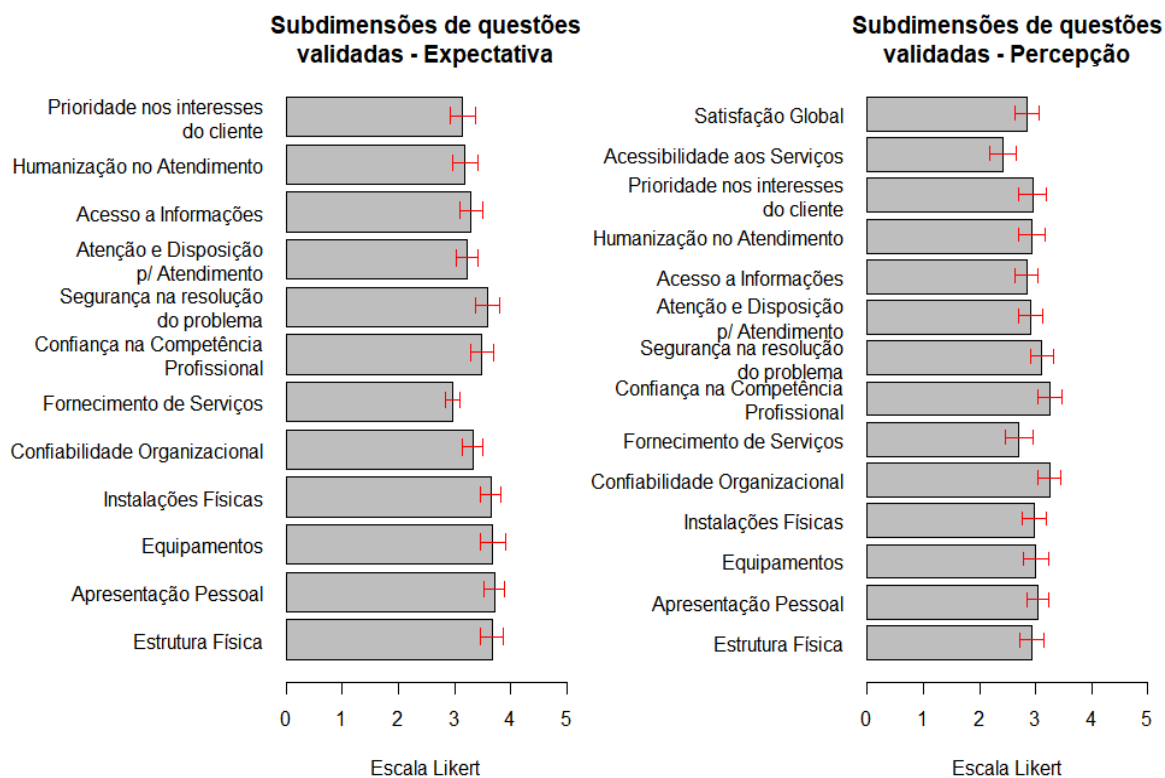


Figura 48- Gráfico de barras com intervalos de confiança para as subdimensões de questões validadas – Proxy L (ciclo 1, estudo 5) - Perspectiva indireta para Usuários (Expectativa e Percepção).

Por sua vez, a análise das macrodimensões de qualidade, estruturou-se a partir da valoração média atribuída às questões validadas pelos indivíduos e calculadas para as macrodimensões propostas pelos informantes.

A Tabela 33 apresenta uma análise descritiva das valorações médias obtidas pelas macrodimensões de qualidade, na *proxy L* (62 informantes no ciclo 1 de validação do instrumento, perspectiva indireta com atribuição de valorações por especialistas e profissionais, com observações para as perspectivas de Expectativa e de Percepção de Usuários.

Tabela 33- Análise descritiva das macrodimensões de qualidade – Proxy L (ciclo 1, estudo 5) - Perspectiva indireta para Usuários (Expectativa e Percepção).

Macrodimensões	Expectativa			Percepção		
	Média	D.P.	I.C-95%	Média	D.P.	I.C-95%
Tangibilidade	3,68	0,65	[3,52; 3,84]	2,97	0,79	[2,77; 3,16]
Confiabilidade	3,08	0,47	[2,96; 3,19]	2,87	0,84	[2,67; 3,08]
Segurança	3,52	0,79	[3,33; 3,71]	3,20	0,80	[3,02; 3,42]
Presteza	3,25	0,77	[3,07; 3,43]	2,87	0,79	[2,69; 3,07]
Empatia	3,19	0,87	[2,97; 3,41]	2,95	0,91	[2,73; 3,17]
Acessibilidade	-	-	-	2,65	0,71	[2,46; 2,83]

Nota.¹ Intervalo Bootstrap

A análise global permite afirmar que na perspectiva da Expectativa, as cinco macrodimensões de questões validadas analisadas (Acessibilidade não foi mensurada quanto a

Expectativa) obtiveram valoração média elevada (superior a 3, importante/relevante) e que na perspectiva da Percepção, as seis macrodimensões de questões validadas (Tangibilidade, Confiabilidade, Presteza, Segurança, Empatia e Acessibilidade) obtiveram valoração média mediana (próximo de 3, importante/relevante). Dessa forma, pelas elevadas valorações médias atribuídas, todas as macrodimensões de questões validadas foram consideradas relevantes, na estruturação do instrumento de pesquisa de Usuários (*proxy M*, ciclo 2).

Noutra análise, em todas as seis macrodimensões de questões validadas analisadas houve variabilidade nas valorações obtidas para a Expectativa e para a Percepção, sendo que a ampla maioria dos seus intervalos de confiança se sobrepuseram. Em relação a Expectativa, em média, os indivíduos tenderam a dar maior valoração (importância/relevância) para a macrodimensão Tangibilidade e menor valoração para Confiabilidade em relação aos demais, uma vez que os intervalos de confiança não se sobrepuseram. Por sua vez, em relação a Percepção, as valorações médias obtida para todas as seis macrodimensões de questões validadas foram razoavelmente dentro das mesmas margens de valores, uma vez que os intervalos de confiança se sobrepuseram.

Desta análise descritiva sintetizou-se que os especialistas e profissionais atribuíram graus de importâncias/relevâncias de medianos a elevados (superior ou próximo a 3, importante/relevante) para as subcategorias (20 questões validadas), as categorias (13 subdimensões de questões validadas) e as macrocategorias (seis macrodimensões de qualidade) dos itens de análise, sendo considerados relevantes/importantes para as próximas fases de investigação, na estruturação de um instrumento com questões de qualidade (*proxy M*, ciclo 2). Além disso, em média, as valorações da ampla maioria destes itens de análise foram razoavelmente dentro das mesmas margens de valores, uma vez que os intervalos de confiança se sobrepuseram.

Quanto aos resultados do estudo das possíveis influências de três variáveis moderadoras, das análises das tabelas 83 a 85 (apêndice 9, tópico 9.2.1.3), sintetizou-se que, na *proxy L*, para a Expectativa, não houve uma tendência de influência das variáveis mediadoras nível de atenção à saúde, de complexidade da assistência e de cargo/função, na ampla maioria das 12 subdimensões de questões validadas (Acessibilidade aos Serviços e Satisfação Global não foram mensuradas quanto a Expectativa), com raras exceções, que se propõe aprofundar análises. Por sua vez, para a Percepção, houve tendência de influência das variáveis mediadoras nível de atenção à saúde, de complexidade da assistência e de cargo/função, na ampla maioria (12) das 13 subdimensões de questões validadas.

Por sua vez, das análises das tabelas 86 a 88 (apêndice 9, tópico 9.2.1.3), sintetizou-se que, na *proxy* L, não houve diferenças significativas (valor- $p > 0,050$) nas valorações obtidas na ampla maioria das seis macrodimensões de questões validadas analisadas (Tangibilidade, Confiabilidade, Presteza, Segurança, Empatia e Acessibilidade – esta última não foi mensurada quanto a Expectativa), que não foram mediadas pelo nível de atenção, sendo exceções duas macrodimensões (Confiabilidade e Presteza), e quanto ao nível de complexidade da assistência uma macrodimensão (Confiabilidade) mas sem nenhuma exceção quanto ao nível de cargo/função. Por outro lado, houve significância estatística (valor- $p \leq 0,050$) nas valorações obtidas na ampla maioria das seis macrodimensões de questões validadas analisadas, que foram mediadas pelo nível de atenção, pelo nível de complexidade da assistência, sendo exceções a macrodimensão Acessibilidade quanto ao nível de atenção e quanto ao nível de complexidade da assistência e sem exceções quanto ao nível de cargo/função.

Diante destes resultados, a comprovação da tendência das variáveis nível de atenção à saúde e nível de complexidade da assistência funcionarem enquanto fatores mediadores da ampla maioria das 13 subdimensões de questões validadas propostas, além de todas as cinco macrodimensões de qualidade originais da escala SERVQUAL, além da macrodimensão Acessibilidade, apesar da necessidade de análises mais adequadas da perspectiva de Expectativa, trata-se de um achado inovador desta investigação.

Ressalte-se, ainda, que a validação das cinco macrodimensões de qualidade (macrocategorias de análise), na perspectivas de especialistas e profissionais, na *proxy* L, estão alinhadas com a validação destas mesmas macrodimensões da escala SERVQUAL, nos estudos de Babakus e Mangold (1992) no contexto de saúde, mas a validação da ampla maioria das 13 subdimensões de questões validadas (categorias de análise) trata-se de um achado inovador desta investigação que serão objeto de aprofundamento da validação nas etapas posteriores.

Por fim, diante dos resultados das análises das tabelas 89 e 90 (apêndice 9, tópico 9.2.1.3), houve uma tendência que quanto maior foram os valores da Expectativa maior tenderam a ser os valores da Percepção obtidos nas subdimensões de questões validadas e macrodimensões de qualidade, em contraposição à associação negativa citada pela literatura de aplicação da escala SERVQUAL (E. Babakus & W. G. Mangold, 1992; Parasuraman et al., 1988, 1991a; Parasuraman et al., 1994), mas alinhada com outros estudos empíricos que suportam relação positiva entre *expectativas* e *satisfação* (Oliver & Linda, 1981; Swan & Trawick, 1981; Bearden e Teel, 1983), pelo que se sugere aprofundar as análises.

Estes resultados subsidiaram a construção de um novo instrumento de questões validadas, para a validação de uma escala inovadora de medição de qualidade/satisfação de usuários de saúde, aplicado na fase posterior da investigação (*proxy M*, ciclo 2).

6.2.1.4. Análise descritiva do estudo 5 (ciclo 2: *proxies M e N*) (questões validadas *versus* subdimensões e macrodimensões de qualidade) (perspectiva de expectativa *versus* percepção)

A partir do instrumento consolidado nas fases 1 e 2 procedentes no ciclo 1 (*proxy L*), na fase 3, ciclo 2 do estudo 5, através de um novo instrumento (*proxy M*), aplicou-se uma análise da validação em um grupo piloto de 506 informantes (perspectiva direta de usuários da saúde). Assim, na fase 3, abordou-se a análise da *proxy M* (n = 506), na qual o instrumento manteve similaridades com o instrumento da fase anterior (*proxy L*) (20 questões validadas, 13 subdimensões e seis macrodimensões de qualidade percebida).

A fase 4, ciclo 2 do estudo 5 (*proxy N*), da pesquisa de campo foi desenvolvida pelo uso do instrumento consolidado e validado na fase anterior (*proxy M*), visto que não houve nenhuma mudança proposta pelo grupo piloto, aplicado em uma fase incremental complementar de pesquisa de campo (ciclo 2), voltado à análise das respostas de um grupo piloto de 2.041 usuários da saúde.

A análise dos resultados obtidos nas fases 3 e 4 foram fortemente similares, com raríssimas exceções. Dessa forma, a reprodução destes resultados em separado teria baixo significado comparadamente aos resultados da amostra ampliada, tendo-se optado por apresentar uma análise descritiva integrada da valoração média atribuída às questões validadas de qualidade da *proxy M+N* (n = 2.547), bem como dos valores obtidos de seus reflexos sobre as subdimensões e macrodimensões de qualidade, com observações para a Expectativa e a Percepção, na visão direta de usuários da saúde, a fim de consolidar os níveis de relevância/importância atribuídas para os itens de avaliação da escala qualidade percebida (20 questões validadas, 13 subdimensões e seis macrodimensões de qualidade percebida). Caso necessário pode-se fazer destaques específicos de resultados em separado.

A Tabela 34 abaixo apresenta uma análise descritiva das variáveis de caracterização selecionadas na amostra, na análise das *proxies M e N* (2.547 informantes) que aplicou dois instrumentos com muitas similaridades, no ciclo 2 de aplicação de campo, na visão de usuários da saúde.

Tabela 34- Análise descritiva das variáveis de caracterização selecionadas - Proxy M+N (ciclo 2, estudo 5)
(Visão de Usuários).

Variáveis		Proxy M (n=506)		Proxy N (n=2.041)		Proxy M+N (n=2.547)	
		N	%	N	%	N	%
Nível de Atenção à Saúde	Atenção Primária	241	47,63%	394	19,30%	635	24,93%
	Atenção Secundária	140	27,67%	777	38,07%	917	36,00%
	Atenção Terciária	125	24,70%	870	42,63%	995	39,07%
	Outros	0	0,00%	0	0,00%	0	0,00%
Nível Complexidade da assistência	Baixa	241	47,63%	394	19,30%	635	24,93%
	Média	265	52,37%	1647	80,70%	1912	75,07%
	Alta	0	0,00%	0	0,00%	0	0,00%
	Outros	0	0,00%	0	0,00%	0	0,00%
Escolaridade	Analfabeto	47	9,29%	122	6,53%	169	7,12%
	Ensino Básico	176	34,78%	500	26,77%	676	28,48%
	Ensino Fundamental	79	15,61%	377	20,18%	456	19,21%
	Ensino Médio	195	38,54%	731	39,13%	926	39,01%
	Ensino Superior	8	1,58%	126	6,75%	134	5,64%
	Pós-Graduação	1	0,20%	12	0,64%	13	0,55%
	Mestrado/ PhD	0	0,00%	0	0,00%	0	0,00%
Sexo	Masculino	144	28,51%	609	29,85%	753	29,59%
	Feminino	361	71,49%	1430	70,10%	1791	70,37%
	Outros	0	0,00%	1	0,05%	1	0,04%
Idade	Até 19 anos	4	0,79%	70	3,57%	74	3,00%
	20 - 29 anos	111	21,94%	416	21,20%	527	21,35%
	30 - 39 anos	121	23,91%	419	21,36%	540	21,88%
	40 - 49 anos	76	15,02%	431	21,97%	507	20,54%
	50 - 59 anos	72	14,23%	315	16,06%	387	15,68%
	60 - 69 anos	61	12,06%	191	9,73%	252	10,21%
	60 - 79 anos	36	7,11%	88	4,49%	124	5,02%
	80 anos ou mais	25	4,94%	32	1,63%	57	2,31%

Nas fases 3 e 4 (*proxy M+N*), destaca-se que a amostra se mostrou balanceada (Tabela 34) no nível de atenção à saúde que os indivíduos estavam em atendimento: Terciário (39,07%), Secundário (36,00%) e Primário (24,93%) de atenção à saúde. A maior parte dos indivíduos na amostra (75,07%) estava em atendimento em unidades de Média complexidade, enquanto 24,93% estava em atendimento em unidades de Baixa complexidade. Não houve entrevista em unidades de Alta complexidade da assistência, por exemplo, em hospitais de grande porte. Considerou-se a amostra (n = 2.547) razoavelmente representativa da realidade dos usuários do

SUS, visto que a maior parte dos indivíduos (39,01%) possuía ensino médio, seguidos por boa parte (35,60%) que possuía até o ensino basilar (analfabeto e ensino básico) e por 19,21% de indivíduos que possuíam ensino fundamental, enquanto apenas 6,19% da amostra possuíam ensino superior (graduação e pós-graduações), e, ainda, a maioria dos indivíduos (73,27%) era do sexo Feminino - desequilibrada por histórica demanda reduzidas de homens-, a faixa etária mais frequente (21,88%) foi a de 30 - 39 anos, seguido por 20 - 29 anos (21,35%) e 40 - 49 anos (20,54%).

A Tabela 35 apresenta uma análise descritiva relativa à valoração média atribuída às questões validadas de qualidade, na análise da *proxy* M+N (2.547 informantes), no ciclo 2 de aplicação de campo, na visão de usuários de saúde.

Tabela 35- Análise descritiva das questões validadas – *Proxy* M+N (ciclo 2, estudo 5) - Visão de Usuários (Expectativa e Percepção).

Subdimensões de questões validadas	Questões validadas	Expectativa			Percepção		
		Média	D.P.	I.C-95% ¹	Média	D.P.	I.C-95% ¹
Estrutura Física	SDQ01CV1	3,71	0,72	[3,69; 3,74]	3,19	0,76	[3,16; 3,21]
	SDQ01CV2	3,72	0,64	[3,69; 3,74]	3,24	0,72	[3,21; 3,26]
Apresentação Pessoal	SDQ03CV3	3,69	0,62	[3,67; 3,71]	3,24	0,63	[3,22; 3,26]
Equipamentos	SDQ04CV4	3,59	0,61	[3,57; 3,62]	3,21	0,69	[3,19; 3,24]
Instalações Físicas	SDQ05CV5	3,59	0,62	[3,57; 3,62]	3,20	0,69	[3,17; 3,23]
Confiabilidade Organizacional	SDQ06CV6	3,51	0,75	[3,49; 3,54]	3,29	0,64	[3,27; 3,32]
Fornecimento de Serviços	SDQ07CV7	2,58	0,87	[2,55; 2,62]	2,90	0,78	[2,87; 2,93]
	SDQ07CV8	3,09	0,52	[3,07; 3,11]	3,19	0,64	[3,17; 3,22]
Confiança na Competência Profissional	SDQ09CV9	3,33	0,57	[3,31; 3,35]	3,24	0,62	[3,22; 3,27]
	SDQ09CV10	3,48	0,71	[3,45; 3,50]	3,29	0,63	[3,26; 3,31]
Segurança na resolução do p.	SDQ17CV17	3,55	0,77	[3,52; 3,58]	3,25	0,65	[3,22; 3,28]
Atenção e Disposição p/ Atendimento	SDQ11CV11	3,34	0,66	[3,31; 3,36]	3,28	0,68	[3,26; 3,31]
	SDQ11CV12	3,23	0,59	[3,21; 3,25]	3,25	0,71	[3,22; 3,28]
Acesso a Informações	SDQ15CV15	3,34	0,74	[3,31; 3,37]	3,16	0,65	[3,13; 3,18]
	SDQ15CV16	3,27	0,63	[3,25; 3,30]	3,14	0,64	[3,12; 3,17]
Humanização no Atendimento	SDQ13CV13	3,17	0,53	[3,15; 3,20]	3,29	0,67	[3,26; 3,32]
	SDQ13CV14	3,32	0,63	[3,30; 3,35]	3,25	0,65	[3,22; 3,27]
Prioridade nos interesses	SDQ18CV18	3,11	0,53	[3,09; 3,13]	3,19	0,61	[3,17; 3,21]
Acessibilidade aos Serviços	SDQ19CV19	-	-	-	2,87	0,77	[2,84; 2,90]
Satisfação Global	SDQ20CV20	-	-	-	3,35	0,79	[3,32; 3,38]

Nota.¹ Intervalo Bootstrap

A análise global permite afirmar que a ampla maioria das 20 questões validadas analisadas (SDQ07CV19 - Acessibilidade aos Serviços e SDQ07CV20 - Satisfação Global não foram mensuradas quanto a Expectativa) obtiveram valoração média elevada (superior a 3,

importante/relevante), na análise da *proxy* M+N, sendo as 20 questões validadas consideradas relevantes/importantes na aplicação da escala de qualidade/satisfação percebida pelos usuários de saúde.

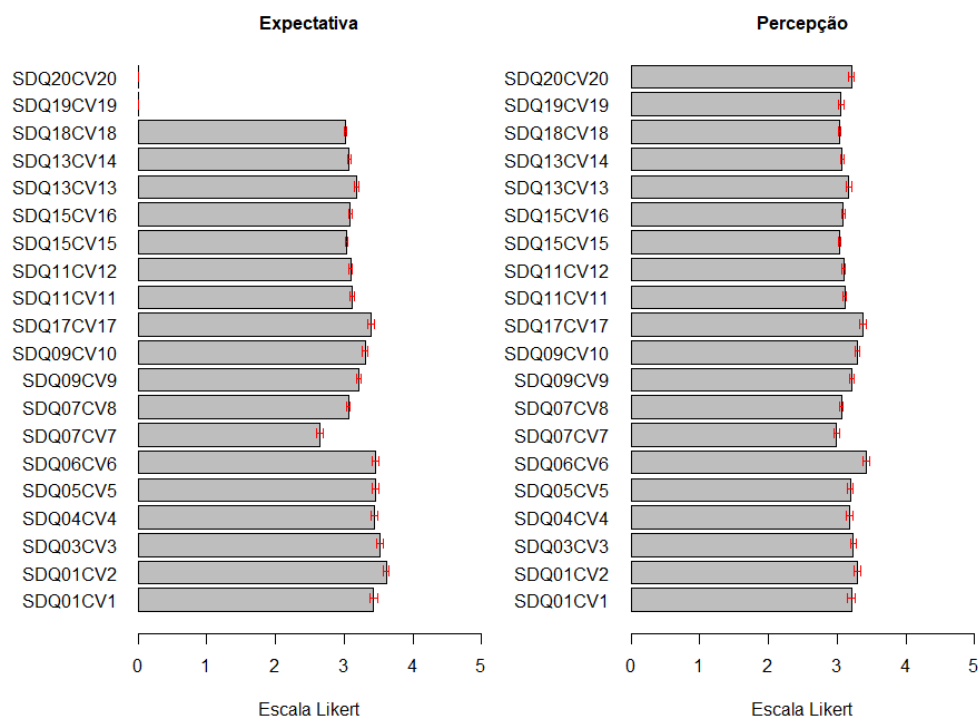


Figura 49- Gráfico de barras com intervalos de confiança para questões validadas - *Proxy* M+N (ciclo 2, estudo 5) - Visão de Usuários (Expectativa e Percepção).

A figura 49, que sintetizou os dados da Tabela 35, mostra que na ampla maioria (17) das 20 questões validadas analisadas (SDQ07CV19 - Acessibilidade aos Serviços e SDQ07CV20 - Satisfação Global não foram mensuradas quanto a Expectativa) houve variabilidade na atribuição de valorações de Expectativa e de Percepção, sendo que os seus intervalos de confiança se sobrepuseram, indicando que não houve diferenças significativas na maioria das valorações atribuídas para as questões dentro da perspectiva de Expectativa e para as questões dentro da perspectiva de Percepção, mas algumas questões validadas na perspectiva Expectativa (SDQ06CV7, SDQ06CV8, SDQ06CV18) e na perspectiva Percepção (SDQ06CV7, SDQ06CV19, SDQ06CV20) obtiveram valorações médias significativamente menores, na *proxy* M+N (2.547 informantes, visão de Usuários).

Por sua vez, a análise das subdimensões de questões validadas, estruturou-se a partir da valoração média atribuída às questões validadas pelos indivíduos e calculadas para as subdimensões propostas pelos informantes.

A Tabela 36 apresenta uma análise descritiva das valorações médias obtidas pelas subdimensões de constructos, na análise da *proxy* M+N (2.547 informantes), no ciclo 2 de aplicação de campo, na visão de usuários de saúde.

Tabela 36- Análise descritiva das subdimensões de questões validadas– *Proxy* M+N (ciclo 2, estudo 5) - Visão de Usuários (Expectativa e Percepção).

Subdimensões de questões validadas	Expectativa			Percepção		
	Média	D.P.	I.C-95%	Média	D.P.	I.C-95%
Estrutura Física	3,72	0,62	[3,69; 3,74]	3,21	0,68	[3,18; 3,24]
Apresentação Pessoal	3,69	0,62	[3,67; 3,72]	3,24	0,63	[3,21; 3,26]
Equipamentos	3,59	0,61	[3,57; 3,62]	3,21	0,69	[3,19; 3,24]
Instalações Físicas	3,59	0,62	[3,57; 3,62]	3,20	0,69	[3,17; 3,23]
Confiabilidade Organizacional	3,51	0,75	[3,49; 3,55]	3,29	0,64	[3,27; 3,32]
Fornecimento de Serviços	2,84	0,52	[2,82; 2,86]	3,05	0,61	[3,02; 3,07]
Confiança na Competência P.	3,40	0,60	[3,38; 3,42]	3,26	0,58	[3,24; 3,29]
Segurança na resolução do p.	3,55	0,77	[3,52; 3,58]	3,25	0,65	[3,22; 3,28]
Atenção e Disposição p/ At.	3,28	0,56	[3,26; 3,30]	3,27	0,66	[3,25; 3,30]
Acesso a Informações	3,31	0,63	[3,28; 3,33]	3,15	0,60	[3,13; 3,17]
Humanização no Atendimento	3,25	0,51	[3,23; 3,27]	3,27	0,62	[3,24; 3,29]
Prioridade nos interesses do c.	3,11	0,53	[3,09; 3,13]	3,19	0,61	[3,17; 3,22]
Acessibilidade aos Serviços	-	-	-	2,87	0,77	[2,84; 2,90]
Satisfação Global	-	-	-	3,35	0,79	[3,32; 3,38]

Nota.¹ Intervalo Bootstrap

A análise global permite afirmar que, nas perspectivas de Expectativa e de Percepção, a ampla maioria das 13 subdimensões de questões validadas (Acessibilidade aos Serviços e Satisfação Global não foram mensuradas quanto a Expectativa) obtiveram valoração média elevada (superior a 3, importante/relevante), na análise da *proxy* M+N, sendo as 13 subdimensões de questões validadas consideradas relevantes/importantes na aplicação da escala de qualidade/satisfação percebida pelos usuários de saúde.

A figura 50, que sintetizou os dados da Tabela 36, mostra que na ampla maioria das 13 subdimensões de questões validadas houve variabilidade na atribuição de valorações de Expectativa e de Percepção, sendo que os seus intervalos de confiança se sobrepuseram. Em relação a Expectativa, houve diferença significativa somente em duas subdimensões (Fornecimento de Serviços e Prioridade nos interesses do cliente - que formaram um grupo com valorações significativamente menor), uma vez que as demais onze subdimensões de questões validadas tiveram intervalos de confiança que se sobrepuseram; Em relação a Percepção, houve diferença significativa somente em duas subdimensões (Fornecimento de Serviços e

Acessibilidade aos Serviços - que formaram um grupo com valorações significativamente menor), uma vez que as demais onze subdimensões de questões validadas tiveram intervalos de confiança que se sobrepuseram, e, ainda na subdimensão Satisfação Global com valoração significativamente maior que todas as demais subdimensões;

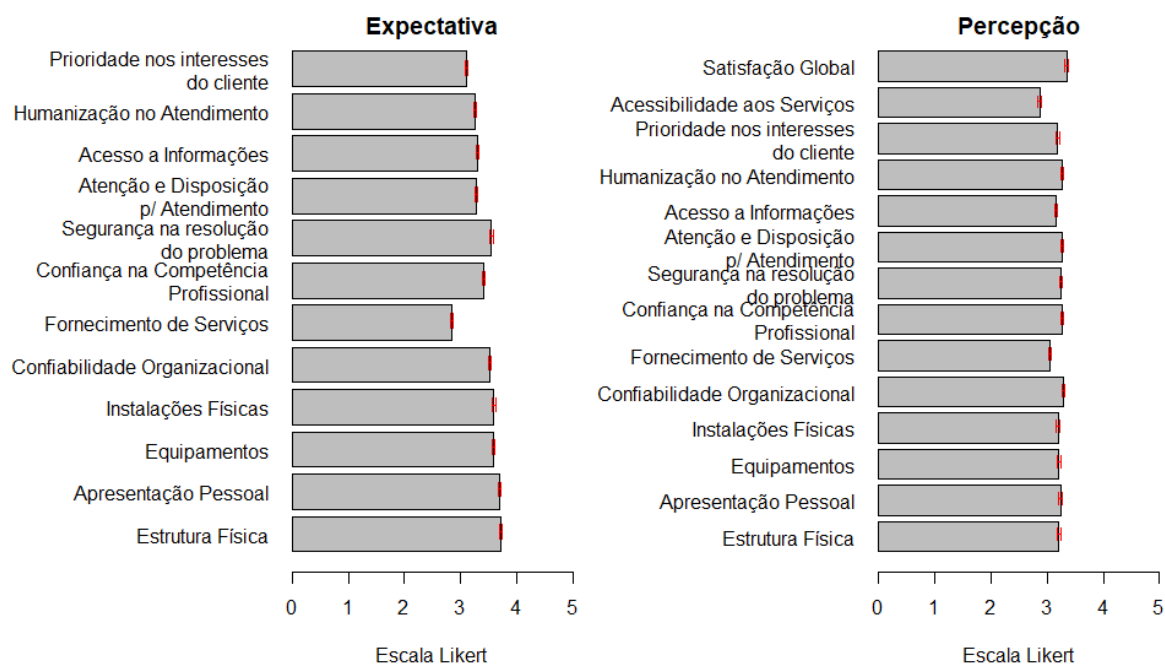


Figura 50- Gráfico de barras com intervalos de confiança para as subdimensões de questões validadas – Proxy M+N (ciclo 2, estudo 5) - Visão de Usuários (Expectativa e Percepção).

Por sua vez, a análise das macrodimensões de questões validadas, estruturou-se a partir da valoração média atribuída às questões validadas pelos indivíduos e calculadas para as macrodimensões propostas pelos informantes.

A Tabela 37 apresenta uma análise descritiva das valorações médias obtidas pelas subdimensões de constructos, na análise da *proxy* M+N (2.547 informantes), no ciclo 2 de aplicação de campo, na visão de usuários de saúde.

Tabela 37 - Análise descritiva das macrodimensões de qualidade – Proxy M+N (ciclo 2, estudo 5) - Visão de Usuários (Expectativa e Percepção).

Macrodimensões	Expectativa			Percepção		
	Média	D.P.	I.C-95% ¹	Média	D.P.	I.C-95% ¹
Tangibilidade	3,66	0,54	[3,64; 3,68]	3,22	0,59	[3,19; 3,24]
Confiabilidade	3,06	0,47	[3,04; 3,08]	3,13	0,56	[3,11; 3,15]
Segurança	3,45	0,61	[3,43; 3,47]	3,26	0,55	[3,24; 3,28]
Presteza	3,29	0,55	[3,27; 3,32]	3,21	0,58	[3,19; 3,24]
Empatia	3,20	0,46	[3,18; 3,22]	3,24	0,57	[3,22; 3,27]
Acessibilidade	-	-	-	3,13	0,69	[3,11; 3,16]

Nota.¹ Intervalo Bootstrap

A análise global permite afirmar que nas perspectivas da Expectativa e da Percepção todas as seis subdimensões de questões validadas analisadas (Tangibilidade, Confiabilidade, Presteza, Segurança, Empatia e Acessibilidade – esta última não foi mensurada na perspectiva da Expectativa) obtiveram valoração média elevada (superior a 3, importante/relevante), na análise da *proxy* M+N, sendo as seis macrodimensões de qualidade consideradas relevantes/importantes na aplicação da escala de qualidade/satisfação percebida pelos usuários de saúde.

Noutra análise, em todas as seis macrodimensões de questões validadas analisadas, houve variabilidade nas valorações obtidas para a Expectativa e para a Percepção, sendo que a ampla maioria dos seus intervalos de confiança se sobrepuseram. Em relação a Expectativa, em média, os indivíduos tenderam a dar maior valoração (importância/relevância) para a macrodimensão Tangibilidade e menor valoração para Confiabilidade em relação aos demais, uma vez que os intervalos de confiança não se sobrepuseram. Por sua vez, em relação a Percepção, as valorações médias obtida para todas as macrodimensões de questões validadas foram razoavelmente dentro das mesmas margens de valores, sendo que as maiores valorações médias foram para Segurança e Empatia.

Desta análise descritiva sintetizou-se que os usuários de saúde atribuíram elevados graus de importâncias/relevâncias das subcategorias (20 questões validadas), das categorias (13 subdimensões de questões validadas) e das macrocategorias (seis macrodimensões de qualidade) dos itens de análise, sendo considerados relevantes/importantes na aplicação da escala de qualidade/satisfação percebida pelos usuários de saúde. Além disso, em média, as valorações da ampla maioria destes itens de análise foram razoavelmente dentro das mesmas margens de valores, uma vez que os intervalos de confiança se sobrepuseram.

Quanto aos resultados do estudo das possíveis influências de três variáveis moderadoras, das análises das tabelas 91 a 96 (apêndice 9, tópico 9.2.1.4), sintetizou-se que, na *proxy* M+N, para a Expectativa, não houve uma tendência de influência das variáveis mediadoras nível de atenção à saúde, de complexidade da assistência e de cargo/função, na ampla maioria das 12 subdimensões de questões validadas (Acessibilidade aos Serviços e Satisfação Global não foram mensuradas quanto a Expectativa), com raras exceções, que se propõe aprofundar análises. Por sua vez, para a Percepção, houve tendência de influência das variáveis mediadoras nível de atenção à saúde, de complexidade da assistência e de escolaridade de usuários, na ampla maioria (≥ 11) das 13 subdimensões de questões validadas.

Por sua vez, das análises das tabelas 97 a 99 (apêndice 9, tópico 9.2.1.4), sintetizou-se que, na *proxy* M+N, houve diferenças significativas (valor- $p > 0,050$) nas valorações obtidas na ampla maioria das seis macrodimensões de questões validadas analisadas (Tangibilidade, Confiabilidade, Presteza, Segurança, Empatia e Acessibilidade – esta última não foi mensurada quanto a Expectativa), que foram mediadas pelo nível de atenção e de complexidade da assistência, sendo exceção apenas uma macrodimensão (Confiabilidade), para a perspectiva Expectativa, quanto ao nível de complexidade da assistência. Por sua vez, para a Percepção, houve tendência de influência das variáveis mediadoras nível de atenção à saúde, de complexidade da assistência e de escolaridade de usuários, na ampla maioria (≥ 05) das seis macrodimensões de qualidade.

Diante destes resultados, a comprovação da tendência das variáveis nível de atenção à saúde, nível de complexidade da assistência e nível de escolaridade de usuários funcionarem enquanto fatores mediadores da ampla maioria das 13 subdimensões de questões validadas propostas, além de todas as cinco macrodimensões de qualidade originais da escala SERVQUAL, além da macrodimensão Acessibilidade, trata-se de um achado inovador desta investigação.

Ressalte-se, ainda, que a validação das cinco macrodimensões de qualidade (macrocategorias de análise), na perspectivas de usuários, na *proxy* M+N, estão alinhadas com a validação destas mesmas macrodimensões da escala SERVQUAL, nos estudos de Babakus e Mangold (1992) no contexto de saúde, mas a validação da ampla maioria das 13 subdimensões de questões validadas (categorias de análise) trata-se de um achado inovador desta investigação que serão objeto de aprofundamento da validação pela análise inferencial (tópico 6.2.2.3).

Por fim, diante dos resultados das análises das tabelas 100 e 101 (apêndice 9, tópico 9.2.1.4), houve uma tendência que quanto maior foram os valores da Expectativa maior tenderam a ser os valores da Percepção obtidas nas subdimensões de questões validadas e macrodimensões de qualidade, em contraposição à associação negativa citada pela literatura de aplicação da escala SERVQUAL (E. Babakus & W. G. Mangold, 1992; Parasuraman et al., 1988, 1991a; Parasuraman et al., 1994), mas alinhada com outros estudos empíricos que suportam relação positiva entre *expectativas* e *satisfação* (Oliver & Linda, 1981; Swan & Trawick, 1981; Bearden e Teel, 1983), pelo que se sugere aprofundar as análises.

Estes resultados subsidiaram a análise de validação de uma escala inovadora de medição de qualidade/satisfação de usuários de saúde.

6.2.2. Análises inferenciais dos estudos 4 e 5 (*proxies* J e M+N)

Após a conclusão da análise estatística descritiva, aplicou-se métodos de análise inferencial a fim de verificar a validade das hipóteses propostas e consolidar os modelos desenhados nesta investigação para os itens da escala de avaliação de qualidade percebida (estudos 4 e 5, ciclo 2 - aplicação de campo), conforme métodos de análise inferencial especificados no início do tópico 6, que dividiram-se em quatro análises de resultados: estudo das cargas fatoriais dos constructos *versus* subdimensões de constructos e macrodimensões de qualidade (*proxy* J) e questões validadas *versus* subdimensões e macrodimensões de qualidade (*proxies* M e N), validação por critérios de qualidade e validade das macrodimensões de qualidade (*proxy* J) e das macrodimensões de questões validadas de qualidade (*proxies* M e N), análise da qualidade (ajustamento) do modelo final resultante, e, análise das correlações entre as variáveis e suas influências na Qualidade Global Percebida nas três *proxies* estudadas (*proxy* J e *proxies* M e N).

6.2.2.1. Análise inferencial do estudo 4 (ciclo 2: *proxy* J)

Nas fases 1 a 3, no ciclo 1 do estudo 4, consolidou-se um instrumento (*proxy* H+I) apenas por análise descritiva (tópico 6.2.1.1). Na fase 4, abordou-se a análise da *proxy* J (n = 506), na qual o instrumento manteve similaridades com o instrumento da fase anterior (*proxy* H+I) (40 constructos, 22 subdimensões de constructos e cinco macrodimensões de qualidade percebida) (tópico 6.2.1.2).

Em busca de validar as 22 subdimensões de constructos e as cinco macrodimensões de qualidade percebida, além de estruturar e validar a variável latente Qualidade Global Percebida, procedentes da *proxy* J (506 informantes), aplicou-se métodos de análise inferencial a fim de consolidar os modelos desenhados nesta investigação, em três categorias iniciais de análises inferenciais: estudo das cargas fatoriais entre variáveis, validação por critérios de qualidade e validade além de análise da qualidade (ajustamento) do modelo final resultante, que se apresentam a seguir, sendo complementadas pela análise das correlações entre as variáveis e suas influências na Qualidade Global Percebida (apêndice 9, tópico 9.2.2.1).

Para validar as 22 subdimensões de constructos analisadas (Equipamentos, Estrutura Física, Apresentação Pessoal, Instalações Físicas, Confiança em Prazos etc) foi feita uma Análise Fatorial Confirmatória (AFC) (Hair et al., 2009) para as subdimensões de constructos,

na *proxy J* (n = 506). Logo, pode-se destacar que, no modelo final, após análise e exclusão de dois constructos (*CO15_Apoio_U e CO19_Prazo.Atendim*), todos os demais 38 constructos que permaneceram na análise apresentaram carga fatorial superior a 0,50 (ou suas permanências não impediram a validação convergente da AFC), na *proxy J*, com as cargas fatoriais mostradas na tabela 38.

Tabela 38- Análise Fatorial Confirmatória dos constructos originais – *Proxy J* (ciclo 2, estudo 4) (Visão de Usuários).

Subdimensões de constructo	Constructos originais	Modelo Inicial		Modelo Final	
		Peso ¹	C.F. ²	Peso ¹	C.F.
Equipamentos	CO1_Disponib	20%	0,97	20%	0,97
	CO2_Acesso	20%	0,99	20%	0,99
	CO3_Resolut	20%	0,98	20%	0,98
	CO4_Conserv	20%	0,87	20%	0,87
	CO5_Usabilid	19%	0,95	19%	0,95
Estrutura Física	CO6_Conforto	39%	0,99	39%	0,99
	CO7_Limpeza	22%	0,67	22%	0,67
	CO8_Bem	39%	0,98	39%	0,98
Apresentação Pessoal	CO9_Vestim	50%	0,94	50%	0,94
	CO10_Visual	50%	1,00	50%	1,00
Instalações Físicas	CO11_Conserv	50%	0,95	50%	0,95
	CO12_Funcion	50%	0,99	50%	0,99
Confiança em Prazos	CO13_Cumprim.Prazos	100%	1,00	100%	1,00
Suporte ao usuário-cliente	CO14_Solidaried	50%	1,00	50%	0,99
	CO15_Apoio_U	-	-	-	-
	CO16_Acolhim	50%	0,99	50%	0,99
Confiabilidade Organizacional	CO17_Confianca.Percep	100%	1,00	100%	1,00
Fornecimento de Serviços	CO18_Prazo.Resolut	57%	0,57	100%	1,00
	CO19_Prazo.Atendim	43%	0,39	-	-
Garantia dos Registros	CO20_Inform.Disp.Contr	100%	1,00	100%	1,00
Acesso a Informação	CO21_Esclarecer	54%	1,11	48%	0,93
	CO22_Infor	46%	0,92	52%	0,99
Pronto Atendimento	CO23_Atendim	53%	0,86	53%	0,86
	CO24_Disponib	47%	0,91	47%	0,92
Disposição atendimento p/	CO25_Dispos.Assistenc	100%	1,00	100%	1,00
Ocupação profissional	CO26_Profiss.Suficiente	100%	1,00	100%	1,00
Competência profissional	CO27_Conf.Compet.Profiss	100%	1,00	100%	1,00
Credibilidade Profissional	CO28_Confianca	49%	0,97	49%	0,98
	CO29_Credib	51%	1,01	51%	1,00
Cortesia nas Relações	CO30_Cortesia	33%	0,98	33%	0,98
	CO31_Respeito	34%	0,99	33%	0,98

		CO32 Atitude	33%	0,94	33%	0,94
Suprimento	de	CO33_Garantia	48%	0,95	49%	0,97
Recursos		CO34_Suprim	52%	1,03	51%	1,00
Comunicação	e	CO35_Atendim.Indiv	100%	1,00	100%	1,00
Atenção						
Humanização	no	CO36_Dedicação	51%	0,99	51%	0,99
Atendimento		CO37_Tratam	49%	0,97	49%	0,97
Compreensão	do	CO38_Conhec.Situac	100%	1,00	100%	1,00
Cliente						
Prioridade	nos	CO39_Objct.Centrado	100%	1,00	100%	1,00
interesses						
Acessibilidade	dos	CO40_Facil.Acesso	100%	1,00	100%	1,00
clientes						

Nota.¹ Peso relativo; ² Carga Fatorial.

Na Tabela 39, pode-se observar que a confiabilidade e a validade das 22 subdimensões de constructos analisadas (Equipamentos, Estrutura Física, Apresentação Pessoal, Instalações Físicas, Confiança em Prazos etc) foram asseguradas, uma vez que todas apresentaram validação convergente ($AVE > 0,40$), confiabilidade adequada ($A.C. > 0,60$ ou $C.C. > 0,60$), unidimensionalidade e validação discriminante ($VCM < AVE$), na *proxy J* ($n = 506$).

Tabela 39- Validação das subdimensões de constructos – *Proxy J* (ciclo 2, estudo 4) (Visão de Usuários).

Subdimensões de constructo	Constructos originais	A.C. ¹	C.C. ²	Dim. ³	AVE ⁴	VCM ⁵
Equipamentos	5	0,98	0,91	1	0,91	0,61
Estrutura Física	3	0,91	0,80	1	0,80	0,46
Apresentação Pessoal	2	0,97	0,94	1	0,94	0,25
Instalações Físicas	2	0,97	0,94	1	0,94	0,61
Confiança em Prazos	1	-	-	1	-	-
Suporte ao usuário-cliente	2	0,99	0,99	1	0,99	0,50
Confiabilidade Organizacional	1	-	-	1	-	-
Fornecimento de Serviços	1	-	-	1	-	-
Garantia dos Registros	1	-	-	1	-	-
Acesso Informação	2	0,95	0,91	1	0,91	0,35
Pronto Atendimento	2	0,87	0,78	1	0,79	0,18
Disposição p/ atendimento	1	-	-	1	-	-
Ocupação profissional	1	-	-	1	-	-
Competência profissional	1	-	-	1	-	-
Credibilidade Profissional	2	0,99	0,98	1	0,98	0,12
Cortesia nas Relações	3	0,98	0,94	1	0,94	0,44
Suprimento de Recursos	2	0,99	0,97	1	0,97	0,06
Comunicação e Atenção	1	-	-	1	-	-
Humanização no Atendimento	2	0,98	0,96	1	0,96	0,50
Compreensão do Cliente	1	-	-	1	-	-
Prioridade nos interesses	1	-	-	1	-	-
Acessibilidade dos clientes	1	-	-	1	-	-

Nota.¹ Alfa de Cronbach; ² Confiabilidade Composta; ³ Dimensionalidade; ⁴ Variância Média Extraída; ⁵ Variância Máxima Compartilhada.

Em seguida, para validar as cinco macrodimensões de qualidade percebida analisadas (Tangibilidade, Confiabilidade, Presteza, Segurança e Empatia) foi feita uma Análise Fatorial Confirmatória (AFC) (Hair et al., 2009) para as macrodimensões de qualidade, na *proxy* J (n = 506). Logo, pode-se destacar que, no modelo final, após análise e exclusão de sete subdimensões de constructos (Confiança em Prazos, Confiabilidade Organizacional, Fornecimento de Serviços, Ocupação profissional, Suprimento de Recursos, Compreensão do Cliente e Acessibilidade dos clientes) todas as 15 subdimensões de constructos que permaneceram na análise apresentaram carga fatorial superior a 0,50 (ou suas permanências não impediram a validação convergente da AFC)., na *proxy* J, com as cargas fatoriais mostradas na tabela 40.

Tabela 40- Análise Fatorial Confirmatória das subdimensões de constructos – *Proxy* J (ciclo 2, estudo 4)
(Visão de Usuários).

Macrodimensões	Subdimensões de constructo	Modelo Inicial		Modelo Final	
		Peso ¹	C.F. ²	Peso ¹	C.F. ²
Tangibilidade	Equipamentos	28%	0,87	27%	0,87
	Estrutura Física	26%	0,73	26%	0,73
	Apresentação Pessoal	18%	0,53	18%	0,53
	Instalações Físicas	28%	0,89	28%	0,90
Confiabilidade	Confiança em Prazos	8%	0,15		
	Suporte ao usuário-cliente	30%	0,83	54%	0,86
	Confiabilidade Organizacional	20%	0,37		
	Fornecimento de Serviços	17%	0,57		
	Garantia dos Registros	25%	0,67	46%	0,71
Presteza	Acesso Informação	33%	0,71	31%	0,67
	Pronto Atendimento	22%	0,46	27%	0,47
	Disposição p/ atendimento	33%	0,84	42%	0,86
	Ocupação profissional	10%	-0,30		
Segurança	Competência profissional	33%	0,68	42%	0,93
	Credibilidade Profissional	27%	0,55	37%	0,79
	Cortesia nas Relações	29%	0,78	21%	0,51
	Suprimento de Recursos	6%	0,19		
Empatia	Comunicação e Atenção	27%	0,81	33%	0,80
	Humanização no Atendimento	35%	0,94	44%	0,96
	Compreensão do Cliente	14%	0,37		
	Prioridade nos interesses	19%	0,71	23%	0,70
	Acessibilidade dos clientes	6%	-0,17		

Nota.¹ Peso relativo; ² Carga Fatorial.

Na Tabela 41, pode-se observar que a confiabilidade e a validade das cinco macrodimensões de qualidade percebida analisadas (Tangibilidade, Confiabilidade, Presteza, Segurança e Empatia) foram asseguradas, uma vez que todas apresentaram validação

convergente ($AVE > 0,40$), confiabilidade adequada ($A.C. > 0,60$ ou $C.C. > 0,60$), unidimensionalidade e validação discriminante ($VCM < AVE$), na *proxy J* ($n = 506$).

Tabela 41- Validação das subdimensões de constructos – *Proxy J* (ciclo 2, estudo 4) (Visão de Usuários).

Macrodimensões	Subdimensões de constructos	A.C. ¹	C.C. ²	Dim. ³	AVE ⁴	VCM ⁵
Tangibilidade	5	0,84	0,60	1	0,60	0,02
Confiabilidade	2	0,76	0,63	1	0,62	0,52
Presteza	3	0,71	0,47	1	0,47	0,44
Segurança	3	0,76	0,46	1	0,58	0,35
Empatia	3	0,87	0,69	1	0,68	0,50

Nota.¹ Alfa de Cronbach; ² Confiabilidade Composta; ³ Dimensionalidade; ⁴ Variância Média Extraída;

⁵ Variância Máxima Compartilhada.

Para validar a variável latente Qualidade Global Percebida foi feita uma Análise Fatorial Confirmatória (AFC) (Hair et al., 2009) da valoração média obtida para as macrodimensões de qualidade, na *proxy J* ($n = 506$). Logo, pode-se destacar que, no modelo final, após análise e exclusão de uma macrodimensão (Tangibilidade), todas as demais quatro macrodimensões de qualidade que permaneceram na análise apresentaram carga fatorial superior a 0,50 (ou suas permanências não impediram a validação convergente da AFC), na *proxy J*, com as cargas fatoriais mostradas na tabela 42.

Tabela 42- Análise Fatorial Confirmatória da Qualidade Global Percebida – *Proxy J* (ciclo 2, estudo 4) (Visão de Usuários).

Macrodimensões	Modelo Inicial		Modelo Final	
	Peso ¹	C.F. ²	Peso ¹	C.F. ²
Tangibilidade	9%	0,17		
Confiabilidade	9%	0,95	29%	0,93
Presteza	28%	1,05	21%	1,03
Segurança	30%	0,91	28%	0,90
Empatia	24%	0,90	22%	0,90

Nota.¹ Peso relativo; ² Carga Fatorial

Na Tabela 43, pode-se observar que a confiabilidade e a validade da variável latente Qualidade Global Percebida foram asseguradas, uma vez que apresentou validação convergente ($AVE > 0,40$), confiabilidade adequada ($A.C. > 0,60$ ou $C.C. > 0,60$), unidimensionalidade e validação discriminante ($VCM < AVE$), na *proxy J* ($n = 506$).

Tabela 43 - Validação da Qualidade Global Percebida – *Proxy J* (ciclo 2, estudo 4) (Visão de Usuários).

Qualidade Global Percebida	Macrodimensões	A.C. ¹	C.C. ²	Dim. ³	AVE ⁴	VCM ⁵
Qualidade	4	0,85	0,97	1	0,88	-

Nota.¹ Alfa de Cronbach; ² Confiabilidade Composta; ³ Dimensionalidade; ⁴ Variância Média Extraída; ⁵

Variância Máxima Compartilhada.

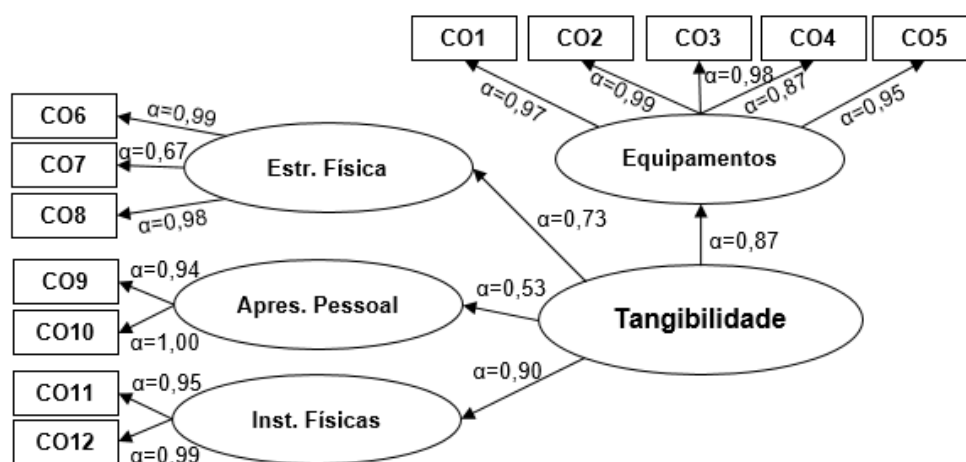
Também se aplicaram testes para medir a qualidade (ajustamento) do modelo construído para as macrodimensões de qualidade percebida da *proxy* J (506 informantes), através de alguns índices selecionados na literatura relevante ($X^2/G.L.$, CFI, TLI e RMSEA). Para um bom ajuste espera-se que $X^2/G.L.$ seja menor que 3 (Arbuckle, 2008; Wheaton, 1987), CFI seja maior que 0,80 (Bentler, 1990), TLI seja maior que 0,80 (Bentler & Bonnet, 1980) e que o RMSEA (Steiger et al., 1985) seja menor que 0,10, sendo o ideal abaixo de 0,05.

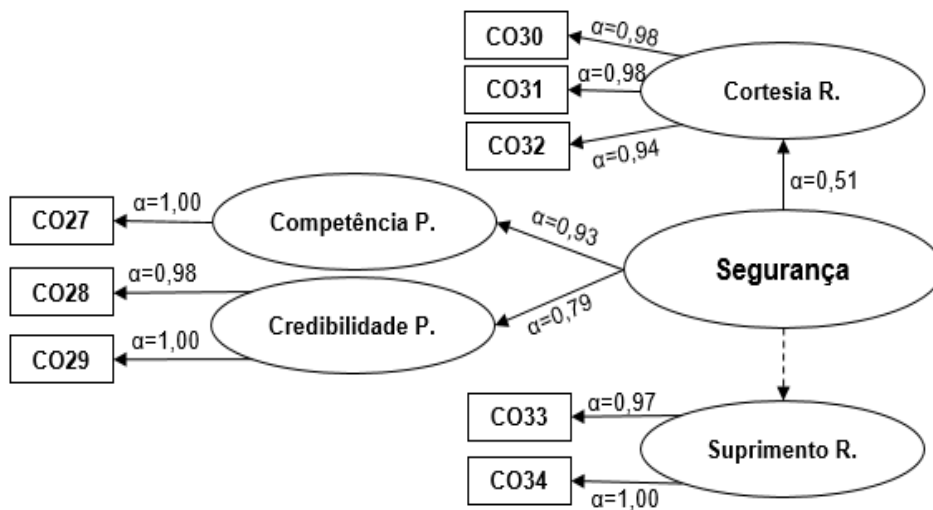
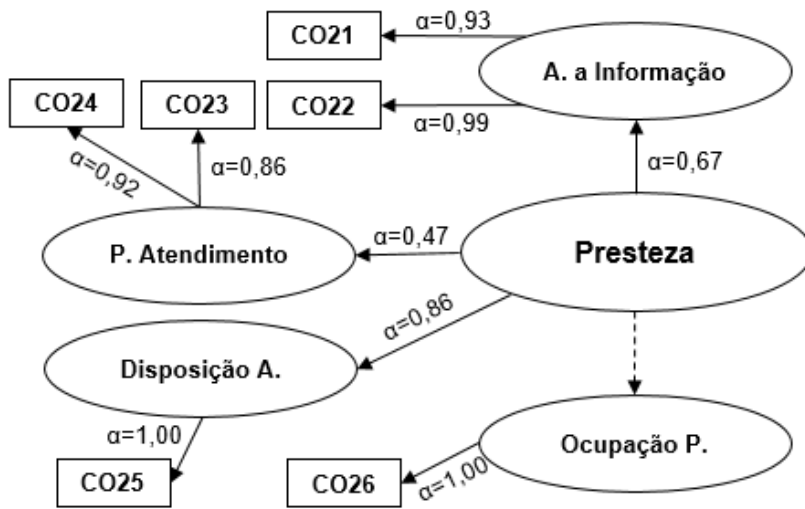
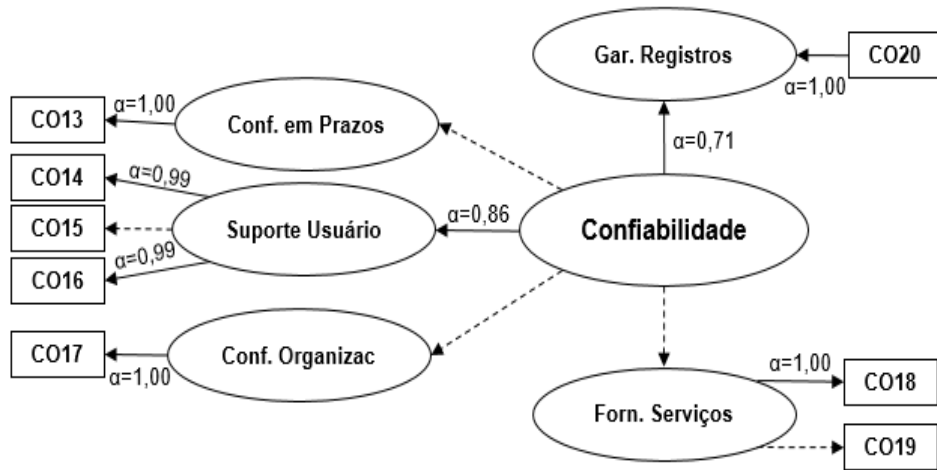
Na Tabela 44, pode-se observar que os parâmetros e índices de ajustamento e qualidade dos modelos de equações estruturais indicaram bom ajuste, já que a razão entre a estatística qui-quadrado e os graus de liberdade foi menor que 3, as estatísticas TLI e o CFI foram maiores que 0,80 e o RMSEA do foi menor que o limite máximo de 0,10.

Tabela 44- Parâmetros de qualidade (ajustamento) do modelo final da Qualidade Global Percebida – *Proxy* J (ciclo 2, estudo 4) (Visão de Usuários).

Medidas de Qualidade do Ajuste	Estatística de Teste	Estatística de Teste Reescalada (Satorra e Bentler 1994)
χ^2	2536,73	984,52
G.L.	439,00	439,00
$\chi^2/G.L.$	5,78	2,24
TLI	0,903	0,956
CFI	0,914	0,961
RMSEA	0,097	0,050
valor-p (RMSEA \leq 0,05)	[0,094; 0,101]	[0,047; 0,052]
	0,000	0,608

Diante destes resultados, a figura 51 sintetizou o ajuste da Análise Fatorial Confirmatória (AFC) e a modelagem para a variável latente Qualidade Global Percebida, na *proxy* J (506 informantes, visão de usuários). Devido ao elevado número de constructos optou-se por omitir suas nomenclaturas, que ficaram identificados como CO1 a CO40.





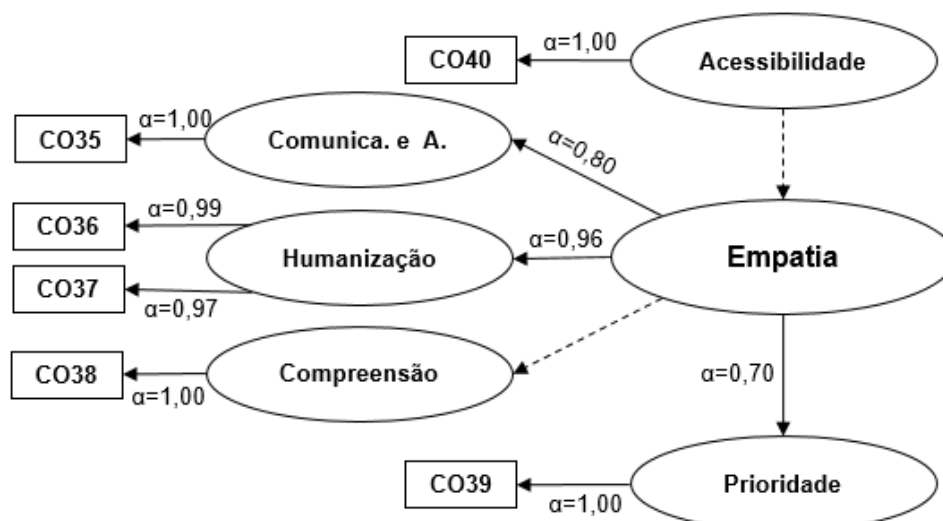


Figura 51 - Análise Fatorial Confirmatória da Qualidade Global Percebida - Proxy J (ciclo 2, estudo 4) (Visão de Usuários).

Para concluir a análise inferencial, as figuras 66 e 67 sintetizaram os resultados apresentados nas Tabelas 102 e 103, respectivamente (apêndice 9, tópico 9.2.2.1), nos quais foi possível observar que os constructos apresentaram maiores correlações com suas subdimensões de constructos e as subdimensões apresentaram maiores correlações com suas respectivas macrodimensões de qualidade percebida, na *proxy J* ($n = 506$).

Por sua vez, os resultados obtidos mostraram que a Qualidade Global Percebida média foi de 3,17 e a mediana foi de 3,01 e que 75% das respostas atribuídas resultaram em valoração superior a 3,00 (cerca de 3, importante/relevante). Pode-se observar uma variabilidade na valoração média da variável latente Qualidade Global Percebida. Houve diferença significativa (valor- $p \leq 0,050$) para considerar as variáveis nível de atenção à saúde e de escolaridade do usuário enquanto fator mediador da Qualidade Global Percebida, no entanto, não houve diferença significativa (valor- $p = 0,104$) quanto ao nível de complexidade na *proxy J* ($n = 506$) (Tabela 104, apêndice 9, tópico 9.2.2.1).

6.2.2.2. Análise inferencial do estudo 5 (ciclo 2: *proxy M+N*)

Nas fases 1 e 2, no ciclo 1 do estudo 5, consolidou-se um instrumento (*proxy L*) apenas por análise descritiva. Na fase 3, ciclo 2 do estudo 5, abordou-se a análise da *proxy M* ($n = 506$), na qual o instrumento manteve similaridades com o instrumento da fase anterior (*proxy L*) (20 questões validadas, 13 subdimensões e seis macrodimensões de qualidade percebida). Por sua vez, na fase 4, ciclo 2 do estudo 5 (*proxy N*), da pesquisa de campo foi desenvolvida pelo uso

do instrumento consolidado e validado na fase anterior (*proxy M*), visto que não houve nenhuma mudança proposta pelo grupo piloto, aplicado em uma fase incremental complementar de pesquisa de campo (ciclo 2), voltado à análise das respostas de um grupo piloto de 2.041 usuários da saúde.

A análise dos resultados obtidos nas fases 3 e 4 foram fortemente similares, com raríssimas exceções, apesar dos parâmetros da análise inferencial terem apresentado algumas diferenças que não impactaram nos modelos finais validados. Dessa forma, a reprodução destes resultados em separado teria baixo significado comparadamente aos resultados da amostra ampliada, tendo-se optado por apresentar uma análise descritiva integrada da valoração média atribuída às questões validadas de qualidade da *proxy M+N* ($n = 2.547$), bem como dos valores obtidos de seus reflexos sobre as subdimensões e macrodimensões de qualidade, com observações para a Expectativa e a Percepção, na visão direta de usuários da saúde, a fim de consolidar os níveis de relevância/importância atribuídas para os itens de avaliação da escala qualidade percebida (20 questões validadas, 13 subdimensões e seis macrodimensões de qualidade percebida). Caso necessário pode-se fazer destaques específicos de resultados em separado.

Em busca de validar as 13 subdimensões de questões validadas e as seis macrodimensões de qualidade percebida (Acessibilidade aos Serviços e Satisfação Global não foram mensuradas para a Expectativa que anulou a macrodimensão Acessibilidade), além de estruturar e validar a variável latente Qualidade Global Percebida, procedentes da *proxy M+N* (2.547 informantes), aplicou-se métodos de análise inferencial a fim de consolidar os modelos desenhados nesta investigação, em três categorias iniciais de análises inferenciais: estudo das cargas fatoriais entre variáveis, validação por critérios de qualidade e validade além de análise da qualidade (ajustamento) do modelo final resultante, que se apresentam a seguir, sendo complementadas pela análise das correlações entre as variáveis e suas influências na Qualidade Global Percebida (apêndice 9, tópico 9.2.2.2).

Para validar as 13 subdimensões de questões validadas (Equipamentos, Estrutura Física, Apresentação Pessoal, Instalações Físicas, Confiabilidade Organizacional etc) (Acessibilidade aos Serviços e Satisfação Global não foram mensuradas para a Expectativa) foi feita uma Análise Fatorial Confirmatória (AFC) (Hair et al., 2009) para as subdimensões de questões validadas, na *proxy M+N* ($n = 2.547$). Logo, pode-se destacar que, no modelo final, após análise e exclusão de apenas uma questão validada (*SDQ07CV7_Tempo*) para a perspectiva da Expectativa, as demais 17 questões (*SDQ19CV19_Prazo.Espera.Aten* não foi mensurada para

Expectativa) e todas as 19 questões para a Percepção permaneceram na análise pois apresentaram carga fatorial superior a 0,50 (ou suas permanências não impediram a validação convergente da AFC), na *proxy* M+N, com as cargas fatoriais mostradas nas tabelas 45 (Expectativa) e 46 (Percepção), a seguir:

Tabela 45- Análise Fatorial Confirmatória das subdimensões de questões validadas (Expectativa)– *Proxy* M+N.

Subdimensão de questões validadas	Questões validadas	Modelo Inicial		Modelo Final	
		Peso ¹	C.F. ²	Peso ¹	C.F. ²
Estrutura Física	SDQ01CV1_Conforto	53%	0,80	52%	0,80
	SDQ01CV2_Limpeza	47%	0,81	48%	0,82
Apresentação Pessoal	SDQ03CV3_AparenciaPes s	100%	1,00	100%	1,00
Equipamentos	SDQ04CV4_FuncEquipam	100%	1,00	100%	1,00
Instalações Físicas	SDQ05CV5_FuncInstalac	100%	1,00	100%	1,00
Confiabilidade Organizacional	SDQ06CV6_Credib.Confia nca	100%	1,00	100%	1,00
Fornecimento de Serviços	SDQ07CV7_Tempo	18%	0,08		
	SDQ07CV8_Disponib	82%	0,58	100%	1,00
Confiança na Competência Profissional	SDQ09CV9_Conf	44%	0,85	44%	0,84
	SDQ09CV10_Credib	56%	0,87	56%	0,88
Segurança na resolução do problema	SDQ17CV17_Segur.Reso. Probl.	100%	1,00	100%	1,00
Atenção e Disposição p/ Atendimento	SDQ11CV11_Atendim	55%	0,82	55%	0,82
	SDQ11CV12_Pront	45%	0,76	45%	0,76
Acesso a Informações	SDQ15CV15_Facilid	55%	0,86	55%	0,86
	SDQ15CV16_Esclarec	45%	0,82	45%	0,82
Humanização no Atendimento	SDQ13CV13_Gentileza	39%	0,65	38%	0,65
	SDQ13CV14_Tratam	61%	0,85	62%	0,85
Prioridade nos interesses do cliente	SDQ18CV18_Compreens.I nteresse	100%	1,00	100%	1,00

Nota.¹ Peso relativo; ² Carga Fatorial.

Tabela 46 - Análise Fatorial Confirmatória das subdimensões de questões validadas (Percepção) – *Proxy* M+N

Subdimensão de questões validadas	Questões validadas	Modelo Inicial	
		Peso ¹	C.F. ²
Estrutura Física	SDQ01CV1_Conforto	50%	0,81
	SDQ01CV2_Limpeza	50%	0,85
Apresentação Pessoal	SDQ03CV3_AparenciaPess	100%	1,00
Equipamentos	SDQ04CV4_FuncEquipam	100%	1,00
Instalações Físicas	SDQ05CV5_FuncInstalac	100%	1,00
Confiabilidade Organizacional	SDQ06CV6_Credib.Confianca	100%	1,00
Fornecimento de Serviços	SDQ07CV7_Tempo	44%	0,54
	SDQ07CV8_Disponib	56%	0,84

Confiança na Competência Profissional	SDQ09CV9_Conf	52%	0,89
	SDQ09CV10_Credib	48%	0,80
Segurança na resolução do problema	SDQ17CV17_Segur.Reso.Probl.	100%	1,00
Atenção e Disposição p/ Atendimento	SDQ11CV11_Atendim	48%	0,84
	SDQ11CV12_Pront	52%	0,88
Acesso a Informações	SDQ15CV15_Facilid	51%	0,84
	SDQ15CV16_Esclarec	49%	0,82
Humanização no Atendimento	SDQ13CV13_Gentileza	50%	0,86
	SDQ13CV14_Tratam	50%	0,87
Prioridade nos interesses do cliente	SDQ18CV18_Compreens.Interes	100%	1,00
Acessibilidade aos Serviços	SDQ19CV19_Prazo.Espera.Aten	100%	1,00

Nota.¹ Peso relativo; ² Carga Fatorial.

Nas Tabelas 47 (Expectativa) e 48 (Percepção), a seguir, pode-se observar que a confiabilidade e a validade de todas as 13 subdimensões de questões validadas (Equipamentos, Estrutura Física, Apresentação Pessoal, Instalações Físicas, Confiabilidade Organizacional etc) (Acessibilidade aos Serviços e Satisfação Global não foram mensuradas para a Expectativa) foram asseguradas, uma vez que todas apresentaram validação convergente ($AVE > 0,40$), confiabilidade adequada ($A.C. > 0,60$ ou $C.C. > 0,60$), unidimensionalidade e validação discriminante ($VCM < AVE$), na *proxy* M+N ($n = 2.547$).

Tabela 47 - Validação das subdimensões de questões validadas (Expectativa) - *Proxy* M+N.

Subdimensão de questões validadas	Questões adaptadas	A.C. ¹	C.C. ²	Dim. ³	AVE ₄	VCM ₅
Estrutura Física	2	0,79	0,79	1	0,65	0,48
Apresentação Pessoal	1	-	-	1	-	-
Equipamentos	1	-	-	1	-	-
Instalações Físicas	1	-	-	1	-	-
Confiabilidade Organizacional	1	-	-	1	-	-
Fornecimento de Serviços	1	-	-	1	-	-
Confiança na Competência Profissional	2	0,84	0,91	1	0,74	0,53
Segurança na resolução do problema	1	-	-	1	-	-
Atenção e Disposição p/ Atendimento	2	0,76	0,77	1	0,62	0,53
Acesso a Informações	2	0,82	0,83	1	0,70	0,51
Humanização no Atendimento	2	0,70	0,72	1	0,57	0,52
Prioridade nos interesses do cliente	1	-	-	1	-	-

Nota.¹ Alfa de Cronbach; ² Confiabilidade Composta; ³ Dimensionalidade; ⁴ Variância Média Extraída; ⁵ Variância Máxima Compartilhada.

Tabela 48 - Validação das subdimensões de questões validadas (Percepção) - *Proxy* M+N.

Subdimensão de questões validadas	Questões adaptadas	A.C. ¹	C.C. ²	Dim. ³	AVE ₄	VCM ₅
-----------------------------------	--------------------	-------------------	-------------------	-------------------	------------------	------------------

Estrutura Física	2	0,81	0,82	1	0,68	0,50
Apresentação Pessoal	1	-	-	1	-	-
Equipamentos	1	-	-	1	-	-
Instalações Físicas	1	-	-	1	-	-
Confiabilidade Organizacional	1	-	-	1	-	-
Fornecimento de Serviços	2	0,62	0,66	1	0,50	0,47
Confiança na Competência Profissional	2	0,83	0,84	1	0,72	0,47
Segurança na resolução do problema	1	-	-	1	-	-
Atenção e Disposição p/ Atendimento	2	0,85	0,85	1	0,74	0,58
Acesso a Informações	2	0,81	0,81	1	0,69	0,56
Humanização no Atendimento	2	0,86	0,86	1	0,75	0,58
Prioridade nos interesses do cliente	1	-	-	1	-	-
Acessibilidade aos Serviços	1	-	-	1	-	-

Nota.¹ Alfa de Cronbach; ² Confiabilidade Composta; ³ Dimensionalidade; ⁴ Variância Média Extraída; ⁵ Variância Máxima Compartilhada.

Em seguida, para validar as seis macrodimensões de qualidade percebida analisadas (Tangibilidade, Confiabilidade, Presteza, Segurança, Empatia e Acessibilidade – esta não foi mensurada para a Expectativa) foi feita uma Análise Fatorial Confirmatória (AFC) (Hair et al., 2009) para as macrodimensões de qualidade, na *proxy* M+N (n = 2.547). Logo, pode-se destacar que, no modelo final, após análise e exclusão de apenas uma subdimensão de questão validada (*Fornecimento de Serviços*) para a perspectiva da Expectativa, todas as 13 subdimensões de questões validadas para a Percepção e 12 subdimensões para a Expectativa (que teve exclusão de uma subdimensão), pois apresentaram carga fatorial superior a 0,50 (ou suas permanências não impediram a validação convergente da AFC), na *proxy* M+N, com as cargas fatoriais mostradas nas tabelas 49 (Expectativa) e 50 (Percepção), a seguir:

Tabela 49- Análise Fatorial Confirmatória das macrodimensões de questões validadas (Expectativa) – *Proxy* M+N.

Macrodimensões	Subdimensão de questões validadas	Modelo Inicial		Modelo Final	
		Peso ¹	C.F. ²	Peso ¹	C.F. ²
Tangibilidade	Estrutura Física	26%	0,92	28%	0,97
	Apresentação Pessoal	23%	0,77	25%	0,80
	Equipamentos	25%	0,85	24%	0,77
	Instalações Físicas	26%	0,85	24%	0,78
Confiabilidade	Confiabilidade Organizacional	90%	0,69	100%	1,00
	Fornecimento de Serviços	10%	0,86		
Segurança	Confiança na Competência Profissional	43%	0,94	41%	0,92
	Segurança na resolução do problema	57%	0,80	59%	0,82
Presteza	Atenção e Disposição p/ Atendimento	47%	0,96	47%	0,95
	Acesso a Informações	53%	0,91	53%	0,92
Empatia	Humanização no Atendimento	51%	0,97	49%	1,05
	Prioridade nos interesses do cliente	49%	0,61	51%	0,70

Nota.¹ Peso relativo; ² Carga Fatorial.

Tabela 50- Análise Fatorial Confirmatória das macrodimensões de questões validadas (Percepção) – Proxy M+N.

Macrodimensões	Subdimensão de questões validadas	Modelo inicial	
		Peso ¹	C.F. ²
Tangibilidade	Estrutura Física	25%	0,90
	Apresentação Pessoal	23%	0,80
	Equipamentos	26%	0,85
	Instalações Físicas	26%	0,84
Confiabilidade	Confiabilidade Organizacional	55%	0,77
	Fornecimento de Serviços	45%	0,95
Segurança	Confiança na Competência Profissional	51%	0,91
	Segurança na resolução do problema	49%	0,75
Presteza	Atenção e Disposição p/ Atendimento	49%	0,88
	Acesso a Informações	51%	0,96
Empatia	Humanização no Atendimento	52%	0,90
	Prioridade nos interesses do cliente	48%	0,79
Acessibilidade	Acessibilidade aos Serviços	100%	1,00

Nota.¹ Peso relativo; ² Carga Fatorial.

Nas Tabelas 51 (Expectativa) e 52 (Percepção), a seguir, pode-se observar que a confiabilidade e a validade de todas as seis macrodimensões de qualidade percebida analisadas (Tangibilidade, Confiabilidade, Presteza, Segurança, Empatia e Acessibilidade – esta não foi mensurada para a Expectativa) foram asseguradas, uma vez que todas apresentaram validação convergente ($AVE > 0,40$), confiabilidade adequada ($A.C. > 0,60$ ou $C.C. > 0,60$), unidimensionalidade e validação discriminante ($VCM < AVE$), na proxy M+N ($n = 2.547$).

Tabela 51- Validação das macrodimensões questões validadas (Expectativa) - Proxy M+N.

Macrodimensões	Subdimensões validadas	A.C. ¹	C.C. ²	Dim. ³	AVE ⁴	VCM ⁵
Tangibilidade	4	0,89	0,90	1	0,70	0,31
Confiabilidade	1	-	-	1	-	-
Segurança	2	0,81	0,86	1	0,76	0,64
Presteza	2	0,82	0,93	1	0,88	0,64
Empatia	2	0,69	0,88	1	0,79	0,58

Nota.¹ Alfa de Cronbach; ² Confiabilidade Composta; ³ Dimensionalidade; ⁴ Variância Média Extraída; ⁵ Variância Máxima Compartilhada.

Tabela 52- Validação das macrodimensões de questões validadas (Percepção) - Proxy M+N.

Macrodimensões	Subdimensões validadas	A.C. ¹	C.C. ²	Dim. ³	AVE ⁴	VCM ⁵
Tangibilidade	4	0,89	0,91	1	0,72	0,60

Confiabilidade	2	0,75	0,86	1	0,75	0,61
Segurança	2	0,77	0,82	1	0,70	0,61
Presteza	2	0,82	0,91	1	0,84	0,72
Empatia	2	0,79	0,83	1	0,72	0,71
Acessibilidade	1	-	-	1	-	-

Nota.¹ Alfa de Cronbach; ² Confiabilidade Composta; ³ Dimensionalidade; ⁴ Variância Média Extraída; ⁵ Variância Máxima Compartilhada.

Também aplicaram-se testes para medir a qualidade (ajustamento) do modelo construído para as macrodimensões de qualidade percebida da *proxy* M+N (2.547 informantes), através de alguns índices selecionados na literatura relevante ($X^2/G.L.$, CFI, TLI e RMSEA). Para um bom ajuste espera-se que $X^2/G.L.$ seja menor que 3 (Arbuckle, 2008; Wheaton, 1987), CFI seja maior que 0,80 (Bentler, 1990), TLI seja maior que 0,80 (Bentler & Bonnet, 1980) e que o RMSEA (Steiger et al., 1985) seja menor que 0,10, sendo o ideal abaixo de 0,05.

Nas tabelas 53 (Expectativa) e 54 (Percepção), a seguir, pode-se observar que os parâmetros e índices de ajustamento e qualidade dos modelos de equações estruturais indicaram bom ajuste, já que a razão entre a estatística qui-quadrado e os graus de liberdade foi menor que 3, as estatísticas TLI e o CFI foram maiores que 0,80 e o RMSEA do foi menor que o limite máximo de 0,10.

Tabela 53- Parâmetros de qualidade do modelo final (Expectativa) – Proxy M+N.

Medidas de Qualidade do Ajuste	Estatística de Teste	Estatística de Teste Reescalada (Satorra e Bentler 1994)
χ^2	1292,07	801,06
G.L.	87,00	87,00
$\chi^2/G.L.$	14,85	9,21
tli	0,937	0,927
cfi	0,954	0,947
RMSEA	0,074 [0,070; 0,077]	0,057 [0,054; 0,060]
valor-p (RMSEA \leq 0,05)	0,000	0,000

Tabela 54 - Parâmetros de qualidade do modelo final (Percepção) – Proxy M+N.

Medidas de Qualidade do Ajuste	Estatística de Teste	Estatística de Teste Reescalada (Satorra e Bentler 1994)
χ^2	1630,50	912,72
G.L.	119,00	119,00
$\chi^2/G.L.$	13,70	7,67
tli	0,944	0,928
cfi	0,956	0,944
RMSEA	0,071 [0,068; 0,074]	0,051 [0,049; 0,054]

valor-p (RMSEA \leq 0,05) 0,000 0,199

Diante destes resultados, as figuras 52(Expectativa) e 53(Percepção) sintetizaram os dois ajustes da Análise Fatorial Confirmatória (AFC) e a modelagem para a variável latente Qualidade Global Percebida, na *proxy* M+N (n = 2.547 informantes, visão direta de gestores e profissionais). Devido ao elevado número de questões validadas optou-se por omitir suas nomenclaturas, que ficaram identificados como SDQ01V1 a SDQ19V19.

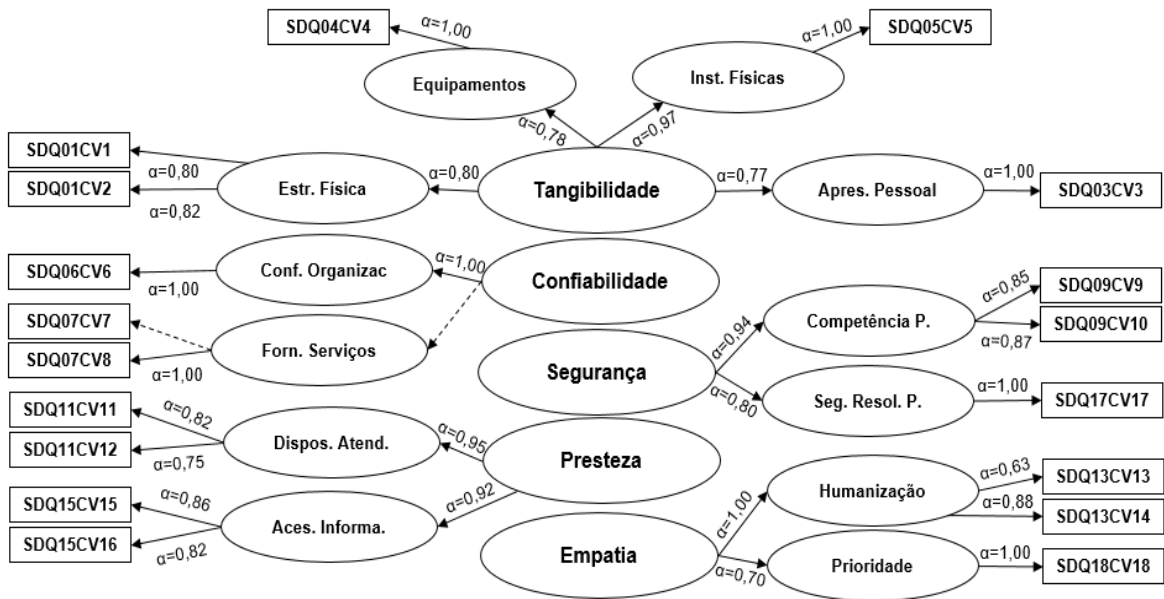


Figura 52- Análise Fatorial Confirmatória da Qualidade Global Percebida (ciclo 2: *proxy* M+N) – Visão de Usuários (Expectativa).

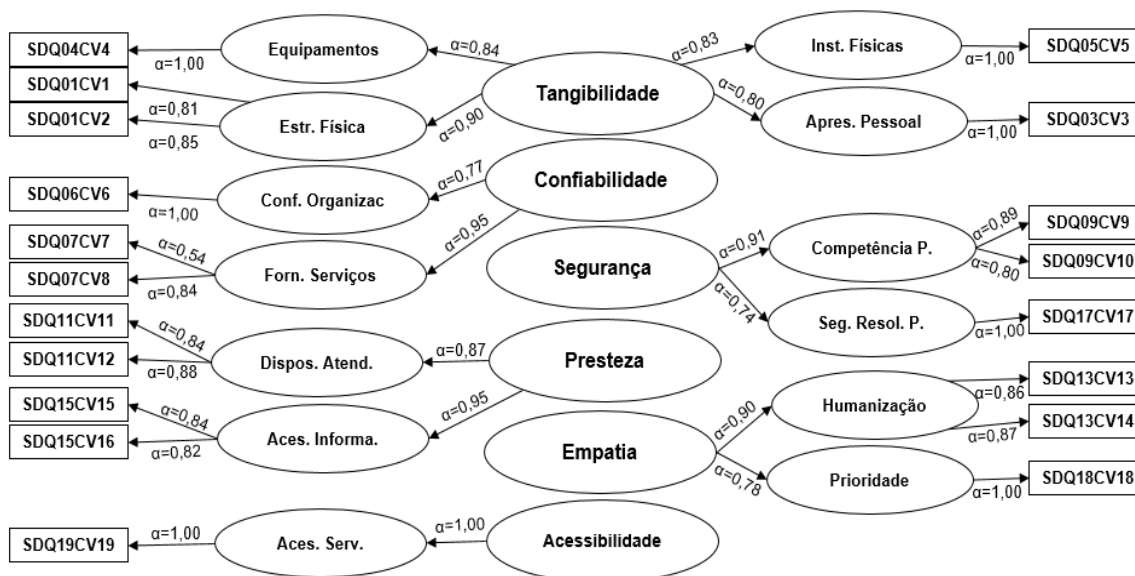


Figura 53- Análise Fatorial Confirmatória da Qualidade Global Percebida (ciclo 2: proxy M+N) – Visão de Usuários (Percepção).

Para concluir a análise inferencial, as figuras 68 e 69 sintetizaram os resultados das Tabelas 105 e 106 (Expectativa) e as figuras 70 e 71 os resultados das Tabelas 107 e 108 (Percepção)(apêndice 9, tópico 9.2.2.2), nos quais foi possível observar que as questões validadas apresentaram maiores correlações com suas respectivas subdimensões de questões validadas que por sua vez apresentaram maiores correlações com suas respectivas macrodimensões de qualidade percebida, na *proxy* M+N ($n = 2.547$).

Por sua vez, os resultados obtidos mostraram que a Qualidade Global Percebida média foi de 3,35 e a mediana de 3,00 e que 75% das respostas atribuídas resultaram em valoração superior a 3 (cerca de 4, muito importante/relevante). Pode-se observar uma variabilidade na valoração média da variável latente Qualidade Global Percebida. Houve diferença significativa (valor- $p \leq 0,050$) para considerar as variáveis nível de atenção à saúde, de complexidade da assistência e de escolaridade de usuários enquanto fator mediador da Qualidade Global Percebida, para a Percepção, na *proxy* M+N ($n = 2.547$) (Tabela 109, apêndice 9, tópico 9.2.2.2).

Dessa forma, concluiu-se a análise da adequação e validação de itens da escala internacional SERVQUAL de qualidade e satisfação em busca de mapear os constructos de qualidade (ideias-força) relevantes, na visão de agentes chaves para avaliação de qualidade/satisfação em serviços de saúde, posteriormente, consolidados por adaptações em etapas e fases incrementais de validação de questões validadas, além de subdimensões e macrodimensões de qualidade e satisfação para usuários do SUS, além dos reflexos sobre uma

variável latente Qualidade Global Percebida, medidos por suas magnitudes (cargas fatoriais), significâncias estatísticas e suas correlações em um modelo estrutural validado.

6.3. Validações e conclusões de hipóteses sobre itens de avaliação de performance

Nos tópicos 6.1. e 9.1. (apêndice 9), procedeu-se a análise dos conhecimentos tácitos dos agentes chaves da amostra nas categorias de macrodomínios e domínios de avaliação, seguidos pela etapa de análise e consolidação de painéis de indicadores e dimensões de performance, e, em busca de um índice sintético proveniente dos itens de análise de performance, analisou-se os reflexos sobre uma *variável latente* Performance Global Percebida pelo que se mensurou os efeitos refletidos das valorações dos indicadores e dimensões de performance, medidos por suas magnitudes (cargas fatoriais), significâncias estatísticas e suas correlações em um modelo estrutural validado, a fim de validar a composição de painéis de controle (*dashboard* ou *tableau de bordo*) de indicadores chaves (*Key Performance Indicators – KPI*).

A partir destes resultados relatados (tópicos 6.1 e 9.1) propôs-se a análise, a seguir, quanto a validação e conclusões de hipóteses sobre itens de avaliação de performance. Para o propósito de análise das hipóteses propostas, ao nível de probabilidade dos resultados, estipulou-se um intervalo de confiança (IC) de 95% para os testes aplicados, portanto, pode-se afirmar que se tem 95% de probabilidade de que os resultados ou valores obtidos retratem a realidade, dentro da margem de erro estipulada, neste caso 2%.

Dessa forma, testou-se as hipóteses iniciais propostas (tópico 5.3), com desdobramentos de hipóteses desagregadas conforme as prioridades identificadas no processo de validação dos gestores, decisores e usuários, articuladas com as provas de conceitos, valor e usabilidade do *proxy* construído de painel de indicadores e de dimensões de performance em serviços de saúde, como forma de subsidiar os tomadores de decisão em saúde de maneira simples e intuitiva para balizar a governança de hospitais e serviços de saúde.

Inicialmente havia sido proposto um diagrama gráfico (*path diagram*), pressupondo algumas relações entre as variáveis latentes (preditas), os itens de avaliação de performance, por exemplo, macrodomínios de avaliação ou dimensões de performance selecionadas e as variáveis observadas (preditoras), as categorias de avaliação, por exemplo, domínios de avaliação ou indicadores de performance (figura 24, tópico 5.3).

Propôs-se algumas hipóteses específicas sobre variáveis latentes e observadas de performance quanto aos resultados obtidos pelos domínios e macrodomínios de avaliação propostos (figura 54).

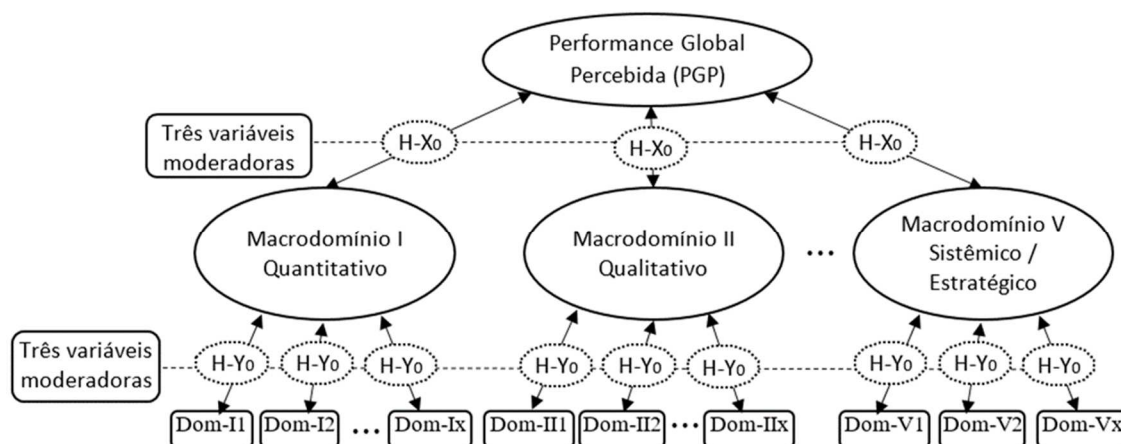


Figura 54- Modelo gráfico das variáveis latentes (Macrodomínios) e observadas (Domínios) de performance

No escopo de domínios e macrodomínios de avaliação, testou-se as hipóteses nulas listadas a seguir, comparativamente às três possíveis variáveis moderadoras selecionadas (nível de atenção/cuidado de saúde, nível de complexidade da assistência e nível de cargo/função de gestores e decisores). Para validar estas hipóteses nulas propostas ($H1_0$ a $H2_0$) – e seus desdobramentos de hipóteses desagregadas ($H1^1_0$ a $H1^2_0$ e $H2^1_0$ a $H2^3_0$), ou refutá-las em contraposição às suas respectivas hipóteses alternativas, apropriou-se das análises de resultados obtidos nesta investigação (tópicos 6.1.1.1 e 6.1.2.1., além dos tópicos 9.1.1.1 e 9.1.2.1. - apêndice 9) e análises destas hipóteses propostas.

- $H1_0$: Os níveis de conhecimentos atribuídos aos domínios de avaliação e seus reflexos sobre os macrodomínios são maioritariamente superiores às notas médias
 - ✓ $H1^1_0$: Os níveis de conhecimentos atribuídos aos domínios de avaliação são maioritariamente superiores às notas médias;
 - ✓ $H1^2_0$: Os níveis de conhecimentos obtidos nos macrodomínios de avaliação são maioritariamente superiores às notas médias.
- $H2_0$: Os níveis de conhecimentos atribuídos aos macrodomínios de avaliação são discriminados pelas três variáveis moderadoras
 - ✓ $H2^1_0$: Os níveis de conhecimentos obtidos nos macrodomínios de avaliação são discriminados pelo nível de atenção (cuidados) à saúde;
 - ✓ $H2^2_0$: Os níveis de conhecimentos obtidos nos macrodomínios de avaliação são discriminados pelo nível de complexidade da assistência;

- ✓ $H2^3_0$: Os níveis de conhecimentos obtidos nos macrodomínios de avaliação são discriminados pelo nível de cargo/função.

Propôs-se algumas hipóteses específicas sobre variáveis latentes e observadas de performance quanto aos resultados obtidos pelos indicadores e dimensões de performance propostos (figura 55).

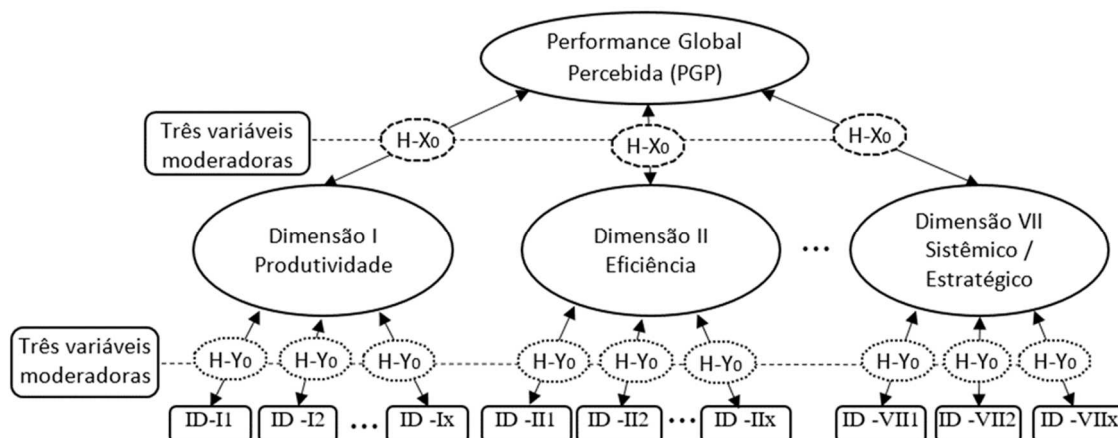


Figura 55- Modelo gráfico das variáveis latentes (Dimensões) e observadas (Indicadores) de performance

No escopo de indicadores e dimensões de performance, testou-se as hipóteses nulas listadas a seguir, comparativamente às três possíveis variáveis moderadoras selecionadas (nível de atenção/cuidado de saúde, nível de complexidade da assistência e nível de cargo/função de gestores e decisores). Para validar estas hipóteses nulas propostas ($H3_0$ a $H9_0$) – e seus desdobramentos de hipóteses desagregadas ($H3^1_0$ a $H3^2_0$ e $H4^1_0$ a $H4^3_0$), ou refutá-las em contraposição às suas respectivas hipóteses alternativas, apropriou-se das análises de resultados obtidos nesta investigação (tópicos 6.1.1.2, 6.1.1.3, 6.1.2.2. e 6.1.2.3., além dos tópicos 9.1.1.2, 9.1.1.3, 9.1.2.2. e 9.1.2.3. - apêndice 9) e análises destas hipóteses propostas.

- $H3_0$: Os níveis de relevâncias/importâncias atribuídos aos indicadores de performance e seus reflexos sobre as dimensões são maioritariamente superiores às notas médias;
 - ✓ $H3^1_0$: Os níveis de relevâncias/importâncias atribuídos aos indicadores de performance são maioritariamente superiores às notas médias;
 - ✓ $H3^2_0$: Os níveis de relevâncias/importâncias atribuídos às dimensões de performance são maioritariamente superiores às notas médias;
- $H4_0$: Os níveis de relevâncias/importâncias atribuídas às dimensões são discriminados pelas três variáveis moderadoras;

- ✓ H4¹₀ : Os níveis de relevâncias/importâncias obtidos nas dimensões são discriminados pelo nível de atenção (cuidados) à saúde;
- ✓ H4²₀ : Os níveis de relevâncias/importâncias obtidos nas dimensões são discriminados pelo nível de complexidade da assistência;
- ✓ H4³₀ : Os níveis de relevâncias/importâncias obtidos nas dimensões são discriminados pelo nível de cargo/função.
- H5₀ : Os indicadores de performance apresentam maioritariamente carga fatorial relevante sobre suas respectivas dimensões de performance;
- H6₀ : As dimensões de performance apresentam maioritariamente qualidade e validação comprovadas;
- H7₀ : As dimensões de performance apresentam carga fatorial relevante sobre a Performance Global Percebida;
- H8₀ : Os indicadores de performance analisados apresentam maioritariamente maiores correlações com suas respectivas dimensões de performance;
- H9₀ : A validação e adequação do modelo da Performance Global Percebida foram comprovadas.

Nos estudos 1 a 3, os resultados foram agregados em subcategorias (domínios de avaliação e indicadores de Performance) e categorias (macrodomínios de avaliação e dimensões de performance) e aplicadas análises descritivas (tópico 6.1.1.) em modelos inovadores de mensuração de performance com 24 domínios e seis macrodomínios de avaliação (*proxy* C), outro com 20 indicadores e sete dimensões de performance (*proxy* E+F) além de outro com 56 indicadores e sete dimensões de performance (*proxy* G). Algumas das hipóteses propostas quanto aos resultados descritivos foram verificadas. Em seguida, aplicou-se métodos de análise inferencial (tópico 6.1.2.), a fim de verificar a validade das demais hipóteses propostas e consolidar os modelos desenhados nesta investigação para os itens de avaliação de performance (domínios e macrodomínios de avaliação e indicadores e dimensões de performance) (estudos 1 a 3).

A figura 6 (*proxy* C) sintetizou o ajuste da Análise Fatorial Confirmatória (AFC) e a modelagem para a variável latente Performance Global Percebida, Dessa forma, a ampla maioria das hipóteses iniciais foram validadas e confirmaram a validação dos itens de mensuração de performance testados (24 domínios e cinco macrodomínios de avaliação), na *proxy* C. Por sua vez, a figura 7 (*proxy* E+F) e a figura 8 (*proxy* G) sintetizaram os ajustes da Análise Fatorial Confirmatória (AFC) e a modelagem para a variável latente Performance

Global Percebida. Dessa forma, a ampla maioria das hipóteses iniciais foram validadas e confirmaram a validação dos itens de mensuração de performance testados (14 indicadores e seis dimensões de performance, na *proxy* E+F, e, 49 indicadores e seis dimensões de performance, na *proxy* G).

Dessa forma, optou-se por apresentar uma síntese de validações (ou refutações) e conclusões das hipóteses analisadas com vínculos aos resultados apurados quanto aos itens de avaliação de performance (tópicos 6.1 e 9.1 – apêndice 9), conforme o quadro 15 abaixo:

Hipóteses formuladas (Itens de avaliação de performance)	Referências (<i>Proxy</i> , Tab./Graf.)	Resultados das hipóteses
- $H1_0$: Os <u>níveis de conhecimentos</u> atribuídos aos <u>domínios</u> de avaliação e seus reflexos sobre os <u>macrodomínios</u> são maioritariamente superiores às <u>notas médias</u> .		
<p>- $H1^1_0$: Os <u>níveis de conhecimentos</u> atribuídos aos <u>domínios</u> de avaliação são maioritariamente superiores às <u>notas médias</u>;</p> <p>- $H1^2_0$: Os <u>níveis de conhecimentos</u> obtidos nos <u>macrodomínios</u> de avaliação são maioritariamente superiores às <u>notas médias</u>.</p>	<p><u>Proxy C</u> (Profiss. e Gestores)</p> <p>- Domínios (Tab. 1, <i>proxy</i> C).</p> <p>- Macrodomínios (Tab. 2, <i>proxy</i> C).</p>	<p>- Para os profissionais e gestores, <u>todos os 24 domínios</u> de avaliação obtiveram <u>valorações acima das notas médias</u> (superior ou cerca de 4), na <i>proxy</i> C;</p> <p>- Para os profissionais e gestores, <u>todos os cinco macrodomínios</u> de avaliação obtiveram <u>valorações acima das notas médias</u> (superior ou cerca de 4), na <i>proxy</i> C.</p>
	<p><i>Em síntese, todos os 24 domínios de avaliação e todos os cinco macrodomínios de avaliação obtiveram valorações acima das notas médias (superior ou cerca de 4), na proxy C;</i></p>	
- $H2_0$: Os <u>níveis de conhecimentos</u> atribuídos aos <u>macrodomínios</u> de avaliação são discriminados pelas <u>três variáveis moderadoras</u> .		
- $H2^1_0$: Os <u>níveis de conhecimentos</u> obtidos nos <u>macrodomínios</u> de avaliação são discriminados pelo <u>nível</u>	<p><u>Proxy C</u> (Profiss. e Gestores)</p>	- O <u>nível de atenção à saúde não pode ser considerado como fator mediador</u> (valor- $p > 0,050$) de <u>nenhum dos cinco macrodomínios</u> de avaliação, na <i>proxy</i> C;

<p>de <u>atenção</u> (cuidados) à saúde;</p> <p>- $H2^2_0$: Os <u>níveis de conhecimentos</u> obtidos nos <u>macrodomínios</u> de avaliação são discriminados pelo <u>nível de complexidade</u> da assistência;</p> <p>- $H2^3_0$: Os <u>níveis de conhecimentos</u> obtidos nos <u>macrodomínios</u> de avaliação são discriminados pelo <u>nível de cargo/função</u>.</p>	<p>- Macrodomínios (Tab. 56, <i>proxy</i> C).</p>	<p>- O <u>nível de complexidade</u> da assistência <u>não pode ser considerado como fator mediador</u> (valor-$p > 0,050$) de <u>nenhum dos cinco macrodomínios</u> de avaliação, na <i>proxy</i> C;</p> <p>- O <u>nível de cargo/função</u> não pode ser considerado como fator mediador (valor-$p > 0,050$), de <u>três macrodomínios</u> (Qualitativos, Satisfação e Sistêmico/Estratégico), exceto nos macrodomínios de avaliação Quantitativos e Efeitos, na <i>proxy</i> C.</p> <p><i>Em síntese, os níveis de atenção à saúde e de complexidade da assistência não tiveram influência como fator moderador (valor-$p > 0,050$) em nenhum dos cinco macrodomínios de avaliação, enquanto o nível de cargo/função não teve influência como fator moderador (valor-$p > 0,050$), em três macrodomínios (Qualitativos, Satisfação e Sistêmico/Estratégico), mas teve influência mediadora em dois macrodomínios (Quantitativos e Efeitos), na <i>proxy</i> C.</i></p>
<p>- $H3_0$: Os <u>níveis de relevâncias/importâncias</u> atribuídos aos <u>indicadores</u> de performance e seus reflexos sobre as <u>dimensões</u> são maioritariamente superiores às <u>notas médias</u>.</p>		
<p>- $H3^1_0$: Os <u>níveis de relevâncias/importâncias</u> atribuídos aos <u>indicadores</u> de performance são maioritariamente superiores às <u>notas médias</u>;</p> <p>- $H3^2_0$: Os <u>níveis de relevâncias/importâncias</u> atribuídos às <u>dimensões</u> de performance são</p>	<p><u>Proxy E+F e Proxy G</u> (Gestores vs Decisores)</p> <p>- Indicadores (Tab. 3 e Graf. 8, <i>proxy</i> E+F; Tab. 5, Graf. 10, <i>proxy</i> G).</p> <p>- Dimensões (Tab. 4 e Graf. 8, <i>proxy</i> E+F;</p>	<p>- Para os <u>gestores</u>, <u>todos os 20 indicadores</u> de performance obtiveram <u>valorações acima das notas médias</u> (superior ou cerca de 3), na <i>proxy</i> E+F, e, para os <u>decisores</u> todos os <u>56 indicadores</u> de performance obtiveram <u>valorações acima das notas médias</u> (superior ou cerca de 4), na <i>proxy</i> G.</p> <p>- Para os <u>gestores</u>, <u>todas as sete dimensões</u> de performance obtiveram <u>valorações acima das notas médias</u> (superior a 3), na <i>proxy</i> E+F, e, para os <u>decisores</u> as <u>sete dimensões</u> de</p>

<p>maioritariamente superiores às <u>notas médias</u>.</p>	<p>Tab. 6, Graf. 10, <i>proxy</i> G).</p>	<p>performance obtiveram <u>valorações acima das notas médias</u> (superior a 4), na <i>proxy</i> G.</p>
<p><i>Em síntese, todos os 20 indicadores de performance, na proxy E+F, e, todos os 56 indicadores de performance, na proxy G, obtiveram valorações acima das notas médias (superior a 3 ou cerca de 4).</i></p>		
<p>- H4₀ : Os <u>níveis de relevâncias/importâncias</u> atribuídas as <u>dimensões</u> são discriminados pelas <u>três variáveis moderadoras</u></p>		
<p>- H4¹₀ : Os <u>níveis de relevâncias/importâncias</u> obtidos nas <u>dimensões</u> são discriminados pelo <u>nível de atenção</u> (cuidados) à saúde;</p> <p>- H4²₀ : Os <u>níveis de relevâncias/importâncias</u> obtidos nas <u>dimensões</u> são discriminados pelo <u>nível de complexidade</u> da assistência;</p> <p>- H4³₀ : Os <u>níveis de relevâncias/importâncias</u> obtidos nas <u>dimensões</u> são discriminados pelo <u>nível de cargo/função</u>.</p>	<p><u>Proxy E+F e Proxy G</u> (Gestores vs Decisores)</p> <p>- Atenção (Tab. 57, <i>proxy</i> E+F; Tab. 59, <i>proxy</i> G).</p> <p>- Complexidade (Tab. 60, <i>proxy</i> G).</p> <p>- Cargo/função (Tab. 58, <i>proxy</i> E+F; Tab. 61, <i>proxy</i> G).</p>	<p>- Para os <u>gestores</u>, o <u>nível de atenção</u> à saúde <u>não pode ser considerado como fator mediador</u> (valor-p>0,050) na <u>ampla maioria</u> das <u>sete dimensões</u> de performance, exceto nas <u>dimensões Produtividade Assistencial e Efetividade</u>, na <i>proxy</i> E+F, e, para os decisores o <u>nível de atenção</u> à saúde <u>não pode ser considerado como fator mediador</u> de <u>nenhuma</u> das <u>sete dimensões</u> de performance, na <i>proxy</i> G.</p> <p>- Para os <u>gestores</u>, o <u>nível de complexidade</u> da assistência <u>não pode ser considerado como fator mediador</u> (valor-p>0,050) em <u>nenhuma</u> das <u>sete dimensões</u> de performance, na <i>proxy</i> E+F, e, para os decisores o <u>nível de complexidade</u> da assistência <u>não pode ser considerado como fator mediador</u> de <u>nenhuma</u> das <u>sete dimensões</u> de performance, na <i>proxy</i> G;</p> <p>- Para os <u>gestores</u>, o <u>nível de cargo/função</u> <u>não pode ser considerado como fator mediador</u> (valor-p>0,050) na <u>ampla maioria</u> das <u>sete dimensões</u> de performance, exceto na <u>dimensão Efetividade</u>, na <i>proxy</i> E+F, e, para os decisores o <u>nível de cargo/função</u> <u>não</u></p>

		<p>pode ser considerado como fator <u>mediador</u> na <u>ampla maioria</u> das <u>sete dimensões</u> de performance, exceto na <u>dimensão Produtividade Assistencial</u>, na <i>proxy G</i></p> <p><i>Em síntese, o nível de complexidade da assistência não teve influência como fator moderador (valor-p>0,050) de nenhuma das sete dimensões de performance nas proxies E+F e G, enquanto os níveis de atenção à saúde e de carga/funçãotiveram influência (valor-p≤0,050) apenas em duas dimensões de performance (Produtividade Assistencial e Efetividade).</i></p>
<p>- H5₀ : Os <u>indicadores de performance</u> apresentam maioritariamente <u>carga fatorial</u> relevante sobre suas respectivas <u>dimensões de performance</u>.</p>	<p><u>Proxy E+F</u> <u>eproxy G</u> (Gestores vs Decisores)</p> <p>- Indicadores (Tab. 12, <i>Proxy E+F</i> e Tab. 17, <i>Proxy G</i>)</p>	<p>- A <u>ampla maioria</u> (14) dos 20 <u>indicadores de performance</u> (<i>proxy E+F</i>, gestores), e a <u>ampla maioria</u> (49) dos 56 <u>indicadores de performance</u> (<i>proxy G</i>, decisores), apresentaram <u>carga fatorial superior a 0,50</u> (ou suas permanências não impediram a validação convergente da AFC).</p>
<p>- H6₀ : As <u>dimensões de performance</u> apresentam maioritariamente <u>qualidade e validação</u> comprovadas.</p>	<p><u>Proxy E+F</u> <u>eproxy G</u> (Gestores vs Decisores)</p> <p>- Dimensões (Tab. 13, <i>Proxy E+F</i> e Tab. 18, <i>Proxy G</i>)</p>	<p>- <u>Todas as sete dimensões de performance</u> (<i>proxy E+F</i>, gestores e <i>proxy G</i>, decisores) <u>tiveram qualidade e validação asseguradas</u>, com validação convergente (AVE > 0,40), confiabilidade adequada (A.C. > 0,60 ou C.C. > 0,60), unidimensionalidade e validação discriminante (VCM < AVE).</p>
<p>- H7₀ : As <u>dimensões de performance</u> apresentam <u>carga fatorial</u> relevante sobre a <u>Performance Global Percebida</u>.</p>	<p><u>Proxy E+F</u> <u>eproxy G</u> (Gestores vs Decisores)</p> <p>- Dimensões (Tab. 14,</p>	<p>- A <u>ampla maioria</u> (6) das <u>sete dimensões de performance</u> (após a <u>retirada da dimensão Produtividade Assistencial</u>), na <i>proxy E+F</i>, e também na <i>proxy G</i> (após a <u>retirada da dimensão Eficiência</u>), apresentaram <u>carga fatorial superior a 0,50</u> (ou suas permanências não</p>

	<i>Proxy E+F e Tab. 19, Proxy G)</i>	impediram a validação convergente da AFC).
- H_{80} : Os <u>indicadores de performance</u> analisados apresentam maioritariamente maiores <u>correlações</u> com suas respectivas <u>dimensões de performance</u>	<u><i>Proxy E+F e proxy G</i></u> (Gestores vs Decisores) - Correlações (Tab. 64 e Gráf. 24, <i>Proxy E+F e Tab. 66 e Gráf. 25, Proxy G)</i>	- <u>A ampla maioria dos 20 indicadores de performance, na <i>proxy E+F</i> e dos 56 indicadores de performance, na <i>proxy G</i>, apresentaram maiores correlações com suas respectivas sete dimensões de performance.</u>
- H_{90} : A <u>validação e adequação do modelo da Performance Global Percebida</u> foram comprovadas.	<u><i>Proxy E+F e proxy G</i></u> (Gestores vs Decisores) - Performance Global Percebida (Tab. 15 e 16 e Fig.7, <i>Proxy E+F e Tab. 20 e 21 e Fig.8, Proxy G)</i>	- <u>A Performance Global Percebida (mantidas as seis dimensões de performance, após exclusão de uma dimensão, nas <i>proxies E+F e G</i>) teve qualidade e validação asseguradas, com validação convergente (AVE > 0,40), confiabilidade adequada (A.C. > 0,60 ou C.C. > 0,60), unidimensionalidade e validação discriminante (VCM < AVE);</u> - O ajuste da AFC foi considerado adequado, já que a razão entre a estatística qui-quadrado e os graus de liberdade foi menor que 3, as estatísticas TLI e o CFI foram maiores ou próximos a 0,80 e o RMSEA do foi menor que o limite máximo de 0,10, nas <i>proxies E+F e G</i>
	<u><i>Em síntese, a Performance Global Percebida teve qualidade e validação asseguradas e os parâmetros de qualidade indicaram boa adequação, sem necessidade de propor ajustes no modelo final com indicadores e dimensões (figuras 7 e 8).</i></u>	

Quadro 15- Síntese de validações e conclusões de hipóteses sobre itens de avaliação de performance (estudos 1, 2 e 3)

6.4. Validações e conclusões de hipóteses sobre itens de medição de qualidade/satisfação (escala de qualidade percebida)

Nos tópicos 6.2. e 9.2. (apêndice 9), procedeu-se a análise da adequação e validação de itens da escala internacional SERVQUAL de qualidade e satisfação para o mapeamento dos constructos de qualidade (ideias-força) relevantes, na visão de agentes chaves para avaliação de qualidade/satisfação em serviços de saúde, em busca de um índice sintético proveniente dos itens de avaliação da qualidade percebida, analisou-se os reflexos sobre uma *variável latente* Qualidade Global Percebida pelo que se mensurou os efeitos refletidos das valorações dos itens de análise da escala de qualidade adaptada (constructos e questões validadas) bem como suas subdimensões e macrodimensões de qualidade, medidos por suas magnitudes (cargas fatoriais), significâncias estatísticas e suas correlações em um modelo estrutural validado, a fim de validar a composição de uma escala de qualidade e satisfação de usuários consolidada em um questionário aplicado ao contexto de unidades e serviços de saúde do SUS.

A partir destes resultados relatados (tópicos 6.2 e 9.2) propôs-se a análise, a seguir, quanto a validação e conclusões de hipóteses sobre itens de medição de qualidade/satisfação (escala de qualidade percebida). Para o propósito de análise das hipóteses propostas, ao nível de probabilidade dos resultados, estipulou-se um intervalo de confiança (IC) de 95% para os testes aplicados, portanto, pode-se afirmar que se tem 95% de probabilidade de que os resultados ou valores obtidos retratem a realidade, dentro da margem de erro estipulada, neste caso 2%.

Dessa forma, testou-se as hipóteses iniciais propostas (tópico 5.3), com desdobramentos de hipóteses desagregadas conforme os conhecimentos prévios dos especialistas e profissionais que propuseram as valorações dos itens de análise da escala de qualidade adaptada (constructos e questões validadas) bem como suas subdimensões e macrodimensões de qualidade.

Inicialmente havia sido proposto um diagrama gráfico (*path diagram*), pressupondo algumas relações entre as variáveis latentes (preditas), as categorias de análise de qualidade, por exemplo, as macrodimensões de qualidade com outras variáveis latentes (preditas), as subcategorias de análise de qualidade, por exemplo, as subdimensões de qualidade e as suas respectivas variáveis observadas (preditoras), os itens de análise de qualidade percebida, por exemplo, os constructos, as questões validadas e questões adaptadas (figura 25, tópico 5.3).

Propôs-se algumas hipóteses específicas sobre variáveis latentes e observadas de qualidade percebida quanto aos resultados obtidos pelas questões validadas e suas respectivas subdimensões e macrodimensões (figura 56).

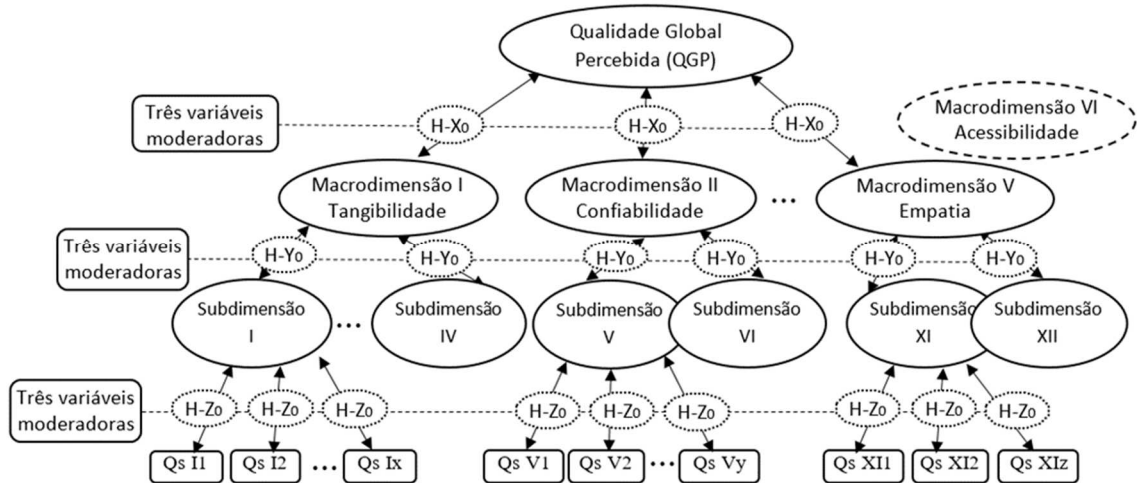


Figura 56- Modelo gráfico das variáveis latentes (Macro & Subdimensões) e observadas (Questões de qualidade) de qualidade percebida

No escopo de questões validadas e suas respectivas, subdimensões e macrodimensões de questões validadas, testou-se as hipóteses nulas listadas a seguir, comparativamente às três possíveis variáveis moderadoras selecionadas (nível de atenção/cuidado de saúde, nível de complexidade da assistência e nível de cargo/função de gestores e decisores – nível de escolaridade de usuários, no caso de pesquisa de campo). Para validar estas hipóteses nulas propostas (H10₀ a H29₀) – e seus desdobramentos de hipóteses desagregadas (H10¹₀ a H10³₀; H11¹₀ a H11⁶₀ e H12¹₀ a H12²₀), ou refutá-las em contraposição às suas respectivas hipóteses alternativas, apropriou-se das análises de resultados obtidos nesta investigação (tópicos 6.2.1.2, 6.2.1.4, 6.2.2.1. e 6.2.2.2., além dos tópicos 9.2.1.2, 9.2.1.4, 9.2.2.1. e 9.2.2.2. - apêndice 9) e análises destas hipóteses propostas.

- H10₀ : Os níveis de relevâncias/importâncias atribuídos às questões validadas de qualidade percebida e seus reflexos sobre as subdimensões de questões validadas e suas respectivas macrodimensões são maioritariamente superiores às notas médias;
 - ✓ H10¹₀ : Os níveis de relevâncias/importâncias atribuídos às questões validadas são maioritariamente superiores às notas médias;
 - ✓ H10²₀ : Os níveis de relevâncias/importâncias obtidas nas subdimensões de questões validadas são maioritariamente superiores às notas médias;
 - ✓ H10³₀ : Os níveis de relevâncias/importâncias obtidas nas macrodimensões de questões validadas são maioritariamente superiores às notas médias.

- H11₀ : Os níveis de relevâncias/importâncias atribuídos às subdimensões de questões validadas e suas respectivas macrodimensões são discriminados pelas três variáveis moderadoras;
 - ✓ H11¹₀ : Os níveis de relevâncias/importâncias obtidos nas subdimensões de questões validadas são discriminados pelo nível de atenção (cuidados) à saúde;
 - ✓ H11²₀ : Os níveis de relevâncias/importâncias obtidos nas subdimensões de questões validadas são discriminados pelo nível de complexidade da assistência;
 - ✓ H11³₀ : Os níveis de relevâncias/importâncias obtidos nas subdimensões de questões validadas são discriminados pelo nível de escolaridade do usuário;
 - ✓ H11⁴₀ : Os níveis de relevâncias/importâncias obtidos nas macrodimensões de questões validadas são discriminados pelo nível de atenção (cuidados) à saúde;
 - ✓ H11⁵₀ : Os níveis de relevâncias/importâncias obtidos nas macrodimensões de questões validadas são discriminados pelo nível de complexidade da assistência;
 - ✓ H11⁶₀ : Os níveis de relevâncias/importâncias obtidos nas macrodimensões de questões validadas são discriminados pelo nível de escolaridade do usuário.
- H12₀ : Há associação negativa entre as subdimensões de questões validadas e suas respectivas macrodimensões, nas perspectivas Expectativa e Percepção;
 - ✓ H12¹₀ : Há associação negativa entre as subdimensões de questões validadas nas perspectivas Expectativa e Percepção;
 - ✓ H12²₀ : Há associação negativa entre as macrodimensões de questões validadas nas perspectivas Expectativa e Percepção;
- H13₀ : As questões validadas apresentam carga fatorial relevante sobre suas respectivas subdimensões;
- H14₀ : As subdimensões de questões validadas apresentam qualidade e validação comprovadas;
- H15₀ : As subdimensões de questões validadas apresentam carga fatorial relevante sobre suas respectivas macrodimensões;
- H16₀ : As macrodimensões de questões validadas apresentam qualidade e validação comprovadas;
- H17₀ : As questões validadas de questões validadas apresentam maiores correlações com suas respectivas subdimensões;
- H18₀ : As subdimensões de questões validadas apresentam maiores correlações com suas respectivas macrodimensões;

- H19₀ : A validação e adequação do modelo da Qualidade Global Percebida, de questões validadas, foram comprovadas.

Nos estudos 4 e 5, os resultados foram respectivamente agregados em subcategorias (constructos e questões validadas) e categorias (subdimensões de constructos e subdimensões de questões validadas) e macrocategorais (macrodimensões de qualidade) de mensuração de qualidade/satisfação percebida através de um processo empírico colaborativo de painéis de especialistas e grupos pilotos de gestores e de campo de usuários que livremente expressaram seus conhecimentos prévios e atribuíram importância/relevância para o julgamento de cada um dos itens apresentados a partir de conceitos globais e genéricos apresentados, tendo sido estruturados e aplicadas análises descritivas (tópico 6.2.1.) em modelos inovadores de mensuração de qualidade/satisfação percebida com 40 domínios, 22 subdimensões de domínios e cinco macrodimensões de qualidade (*proxy* J) e outro com 20 questões validadas, 13 subdimensões de questões validadas e seis macrodimensões de qualidade (*proxy* M+N). Algumas das hipóteses propostas quanto aos resultados descritivos foram verificadas. Em seguida, aplicou-se métodos de análise inferencial (tópico 6.2.2.), a fim de verificar a validade das demais hipóteses propostas e consolidar os modelos desenhados nesta investigação dos itens de mensuração de qualidade percebida (constructos, questões validadas, subdimensões e macrodimensões de qualidade) (estudos 4 e 5).

A figura 9 (*proxy* J) sintetizou o ajuste da Análise Fatorial Confirmatória (AFC) e a modelagem para a variável latente Qualidade Global Percebida. Dessa forma, a ampla maioria das hipóteses iniciais foram validadas e confirmaram a validação dos itens de mensuração de qualidade testados (38 constructos, 15 subdimensões de constructos e quatro macrodimensões de qualidade, na *proxy* J).

Dessa forma, optou-se por apresentar uma síntese de validações (ou refutações) e conclusões das hipóteses analisadas com vínculos aos resultados apurados quanto aos itens de medição de qualidade/satisfação percebida (tópicos 6.2 e 9.2 – apêndice 9), conforme o quadro 16 abaixo:

Hipóteses formuladas (Itens de medição de qualidade/satisfação)	Referências (<i>Proxy</i> , Tab./Graf.)	Resultados das hipóteses
- H10 ₀ : Os <u>níveis de relevâncias/importâncias</u> atribuídos às <u>questões validadas</u> de qualidade percebida e seus reflexos sobre as <u>subdimensões</u> de questões validadas e suas respectivas <u>macrodimensões</u> são maioritariamente superiores às <u>notas médias</u> .		

<p>- H10¹₀ : Os <u>níveis de relevâncias/importâncias</u> atribuídos às <u>questões validadas</u> são maioritariamente superiores às <u>notas médias</u>;</p> <p>- H10²₀ : Os <u>níveis de relevâncias/importâncias</u> obtidas nas <u>subdimensões</u> de questões validadas são maioritariamente superiores às <u>notas médias</u>;</p> <p>- H10³₀ : Os <u>níveis de relevâncias/importâncias</u> obtidas nas <u>macrodimensões</u> de questões validadas são maioritariamente superiores às <u>notas médias</u>.</p>	<p><u>Proxy M+N</u></p> <p>- Expectativa e Percepção (Tab. 35, <i>proxy</i> M+N).</p> <p>- Expectativa e Percepção (Tab. 36, <i>proxy</i> M+N).</p> <p>- Expectativa e Percepção (Tab. 37, <i>proxy</i> M+N).</p>	<p>- Nas perspectivas de Expectativa e de Percepção, <u>todas as 20 questões validadas</u> (SDQ07CV19 - Acessibilidade aos Serviços e SDQ07CV20 - Satisfação Global não foram mensuradas quanto a Expectativa) obtiveram <u>valorações acima das notas médias</u> (superior ou cerca de 3), na <i>proxy</i> M+N;</p> <p>- Nas perspectivas de Expectativa e de Percepção a <u>ampla maioria das 13 subdimensões de questões validadas</u> (Acessibilidade aos Serviços e Satisfação Global não foram mensuradas quanto a Expectativa) obtiveram <u>valorações acima das notas médias</u> (superior ou cerca de 3), na <i>proxy</i> M+N;</p> <p>- Nas perspectivas de Expectativa e de Percepção a <u>todas as seis macrodimensões de questões validadas</u> (Acessibilidade não foi mensurada quanto a Expectativa) obtiveram <u>valorações acima das notas médias</u> (superior a 3), na <i>proxy</i> M+N.</p> <p><i>Em síntese, todas as 20 questões validadas, a ampla maioria das 13 subdimensões e as seis macrodimensões de questões validadas obtiveram valorações maioritariamente acima das notas médias (superior ou cerca de 3), na proxy M+N.</i></p>
<p>- H11₀ : Os <u>níveis de relevâncias/importâncias</u> atribuídos às <u>subdimensões</u> de questões validadas e suas respectivas <u>macrodimensões</u> são discriminados pelas <u>três variáveis moderadoras</u>.</p>		

<p>- H11¹₀ : Os <u>níveis de relevâncias/importâncias</u> obtidos nas <u>subdimensões</u> de questões validadas são discriminados pelo <u>nível de atenção</u> (cuidados) à saúde;</p> <p>- H11²₀ : Os <u>níveis de relevâncias/importâncias</u> obtidos nas <u>subdimensões</u> de questões validadas são discriminados pelo <u>nível de complexidade</u> da assistência;</p> <p>- H11³₀ : Os <u>níveis de relevâncias/importâncias</u> obtidos nas <u>subdimensões</u> de questões validadas são discriminados pelo <u>nível de escolaridade</u> do usuário.</p>	<p><u>Proxy M+N</u></p> <p>- <u>Atenção</u></p> <p>- Expectativa (Tab. 91 <i>proxy</i> M+N).</p> <p>- Percepção (Tab. 92 <i>proxy</i> M+N).</p> <p>- <u>Complexidade</u></p> <p>- Expectativa (e Tab. 93 <i>proxy</i> M+N).</p> <p>- Percepção (Tab. 94 <i>proxy</i> M+N).</p> <p>- <u>Escolaridade</u></p> <p>- Expectativa (Tab. 95 <i>proxy</i> M+N).</p> <p>- Percepção (Tab. 96 <i>proxy</i> M+N).</p>	<p>- O <u>nível de atenção à saúde</u> pode ser considerado como <u>fator mediador</u> (valor-$p \leq 0,050$) em <u>todas as 12 subdimensões</u> (Expectativa) e <u>13 subdimensões</u> (Percepção) (além da subdimensão Satisfação Global), na <i>proxy</i> M+N.</p> <p>- O <u>nível de complexidade</u> da assistência pode ser considerado como <u>fator mediador</u> (valor-$p \leq 0,050$) na <u>ampla maioria</u> das <u>12 subdimensões</u> (Expectativa) e das <u>13 subdimensões</u> (Percepção) (além da subdimensão Satisfação Global), na <i>proxy</i> M+N. Para a Expectativa, houve <u>duas exceções nas subdimensões de questões validadas</u> (Estrutura Física e Fornecimento de Serviços), na <i>proxy</i> M+N.</p> <p>- O <u>nível de escolaridade</u> do usuário pode ser considerado como <u>fator mediador</u> (valor-$p \leq 0,050$) na <u>ampla maioria</u> das <u>13 subdimensões</u> (Percepção), na <i>proxy</i> M+N, sendo três exceções para a Percepção (Instalações Físicas, Fornecimento de Serviços, Atenção e Disposição p/ Atendimento). Porém, <u>não pode ser considerado como fator mediador</u> (valor-$p > 0,050$) na <u>ampla maioria</u> das <u>12 subdimensões</u> (Expectativa) (Acessibilidade aos Serviços e Satisfação Global não foram mensuradas quanto a Expectativa), com raras exceções (<u>três subdimensões</u>).</p>
<p><i>Em síntese, os <u>níveis de atenção à saúde</u> e <u>de complexidade da assistência</u> tiveram <u>influências como fatores moderadores</u> (valor-$p \leq 0,050$) na <u>ampla maioria das 12 subdimensões</u> (Expectativa) e também nas <u>13</u></i></p>		

	<p><u>subdimensões (Percepção) de questões validadas, na proxy M+N, além da subdimensão Satisfação Global, sendo exceções até duas subdimensões de questões validadas</u> (para Expectativa - Prioridade nos interesses do cliente e Confiança na Competência Profissional). <u>Quanto ao nível de escolaridade do usuário pode ser considerado como fator mediador</u> (valor-$p \leq 0,050$) na <u>ampla maioria das 13 subdimensões (Percepção)</u> (além da subdimensão Satisfação Global), na <u>proxy M+N</u>, com raras exceções (três subdimensões).. Entretanto, <u>o nível de escolaridade não pode ser considerado como fator moderador</u> (valor-$p > 0,050$) na <u>ampla maioria das 12 subdimensões (Expectativa)</u>, com raras exceções (três subdimensões).</p>	
<p>- H11⁴₀ : Os <u>níveis de relevâncias/importâncias</u> obtidos nas <u>macrodimensões</u> de questões validadas são discriminados pelo <u>nível de atenção</u> (cuidados) à saúde;</p> <p>- H11⁵₀ : Os <u>níveis de relevâncias/importâncias</u> obtidos nas <u>macrodimensões</u> de questões validadas são discriminados pelo <u>nível de complexidade</u> da assistência;</p> <p>- H11⁶₀ : Os <u>níveis de relevâncias/importâncias</u> obtidos nas <u>macrodimensões</u> de questões validadas são discriminados pelo <u>nível de escolaridade do usuário</u>.</p>	<p><u>Proxy M+N</u></p> <p>- <u>Atenção</u></p> <p>- Expectativa (Tab. 97, <u>proxy M+N</u>).</p> <p>- Percepção (Tab. 97, <u>proxy M+N</u>).</p> <p>- <u>Complexidade</u></p> <p>- Expectativa (Tab. 98, <u>proxy M+N</u>).</p> <p>- Percepção (Tab. 98, <u>proxy M+N</u>).</p> <p>- <u>Escolaridade</u></p> <p>- Expectativa (Tab. 99, <u>proxy M+N</u>).</p>	<p>- O <u>nível de atenção à saúde</u> pode ser considerado como fator mediador (valor-$p \leq 0,050$) em <u>todas as cinco macrodimensões</u> de questões validadas, na <u>proxy M+N</u> (Expectativa e Percepção), além da <u>macrodimensão Confiabilidade</u> adicional para a Percepção.</p> <p>- O <u>nível de complexidade</u> da assistência pode ser considerado como <u>fator mediador</u> (valor-$p \leq 0,050$) em <u>todas as cinco macrodimensões</u> de questões validadas, na <u>proxy M+N</u> (Expectativa e Percepção), além da <u>macrodimensão Acessibilidade</u> para a Percepção, sendo exceção a <u>macrodimensão Confiabilidade</u> para a Expectativa.</p> <p>- O <u>nível de escolaridade do usuário</u> pode ser considerado como <u>fator mediador</u> (valor-$p \leq 0,050$) na <u>ampla maioria das cinco macrodimensões</u> de questões validadas, na <u>proxy M+N</u> (Percepção), com exceção da <u>macrodimensão Confiabilidade</u>, além da <u>macrodimensão adicional Acessibilidade</u> para a Percepção. Porém, <u>não pode ser considerado como fator mediador</u> (valor-</p>

	- Percepção (Tab. 99, proxy M+N).	p>0,050) na <u>ampla maioria das cinco macrodimensões</u> (Expectativa) (Acessibilidade não foi mensurada quanto a Expectativa), com uma exceção positiva (Empatia), na proxy M+N.
	<i>Em síntese, os <u>níveis de atenção à saúde, de complexidade da assistência tiveram influências como fatores moderadores</u> (valor-p≤0,050) na <u>ampla maioria das cinco macrodimensões</u> (Expectativa e Percepção) e, também, o <u>nível de escolaridade do usuário teve influência moderadora na ampla maioria das cinco macrodimensões</u> (Percepção), além da <u>macrodimensão adicional Acessibilidade</u>, sendo exceção a <u>macrodimensão Confiabilidade</u>. Porém, o <u>nível de escolaridade não pode ser considerado como fator mediador</u> (valor-p>0,050) na <u>ampla maioria das cinco macrodimensões</u> (Expectativa) (Acessibilidade não foi mensurada quanto a Expectativa), com exceção da <u>macrodimensão Empatia</u>, na proxy M+N.</i>	
- H12 ₀ : Há <u>associação negativa</u> entre as <u>subdimensões</u> de questões validadas e suas respectivas <u>macrodimensões</u> , nas perspectivas <u>Expectativa e Percepção</u> .		
<p>- H12¹₀ : Há <u>associação negativa</u> entre as <u>subdimensões</u> de <u>questões validadas</u>, nas perspectivas de <u>Expectativa</u> e de <u>Percepção</u>;</p> <p>- H12²₀ : Há <u>associação negativa</u> as <u>macrodimensões</u> de <u>questões validadas</u>, nas perspectivas de <u>Expectativa</u> e de <u>Percepção</u>;</p>	<p><u>Proxy M+N</u></p> <p>- <u>Subdimensão</u></p> <p>- Expectativas Percepção (Tab. 100, proxy M+N).</p> <p>- <u>Macrodimensão</u></p> <p>o</p> <p>- Expectativas Percepção (Tab. 101, proxy M+N).</p>	<p>- Houve <u>correlação significativa</u> (valor-p≤0,050) e <u>positiva</u> em <u>todas as 12 subdimensões</u> de <u>questões validadas</u> (Expectativa e Percepção), na proxy M+N, indicando que <u>quanto maior é a Expectativa maior tende a ser a Percepção</u>.</p> <p>- Houve <u>correlação significativa</u> (valor-p≤0,050) e <u>positiva</u> em <u>todas as cinco macrodimensões</u> de <u>questões validadas</u> (Expectativa e Percepção), na proxy M+N, indicando que <u>quanto maior é a Expectativa maior tende a ser a Percepção</u> obtidas em <u>todas as macrodimensões</u>.</p> <p>Obs: Achados em <u>contraposição à associação negativa</u> citada pela literatura de aplicação da escala Servqual (Parasuraman et al., 1988; 1991 1994; Babakus & Mangold, 1992).</p>

<p>- H13₀ : As <u>questões validadas</u> apresentam <u>carga fatorial</u> relevante sobre suas respectivas <u>subdimensões</u>.</p>	<p><u>Proxy M+N</u></p> <p>- Expectativa (Tab. 45, <i>proxy</i> M+N).</p> <p>- Percepção (Tab. 46, <i>proxy</i> M+N)</p>	<p>- <u>Todas as questões validadas</u> para a <u>Percepção</u> (19 questões) e a <u>ampla maioria</u> para a <u>Expectativa</u> (17 questões), (<i>SDQ19CV19_Prazo.Espera.Aten</i> não foi mensurada para Expectativa), após análise e exclusão de uma questão validada (<i>SDQ07CV7_Tempo</i>) para a Expectativa, na <i>proxy</i> M+N, apresentaram <u>carga fatorial superior a 0,50</u> (ou suas permanências não impediram a validação convergente da AFC);</p>
<p>- H14₀ : As <u>subdimensões</u> de questões validadas apresentam <u>qualidade e validação</u> comprovadas.</p>	<p><u>Proxy M+N</u></p> <p>- Expectativa (Tab. 47, <i>proxy</i> M+N).</p> <p>- Percepção (Tab. 48, <i>proxy</i> M+N)</p>	<p>- <u>Todas as 12 subdimensões</u> de questões <u>validadas</u> (Expectativa e Percepção), além da <u>subdimensão Acessibilidade adicional</u> (Percepção), na <i>proxy</i> M+N, tiveram <u>qualidade e validação asseguradas</u>, com validação convergente (AVE > 0,40), confiabilidade adequada (A.C. > 0,60 ou C.C. > 0,60), unidimensionalidade e validação discriminante (VCM < AVE).</p>
<p>- H15₀ : As <u>subdimensões</u> de questões validadas apresentam <u>carga fatorial</u> relevante sobre suas respectivas <u>macrodimensões</u>.</p>	<p><u>Proxy M+N</u></p> <p>- Expectativa (Tab. 49, <i>proxy</i> M+N).</p> <p>- Percepção (Tab. 50, <i>proxy</i> M+N)</p>	<p>- A <u>ampla maioria</u> (11) das 12 <u>subdimensões</u> de questões validadas, para Expectativa (após a <u>retirada da subdimensão Fornecimento de Serviços</u>), além de <u>todas as 13 questões validadas</u>, para Percepção, na <i>proxy</i> M+N, apresentaram <u>carga fatorial superior a 0,50</u> (ou suas permanências não impediram a validação convergente da AFC).</p>
<p>- H16₀ : As <u>macrodimensões</u> de questões validadas apresentam <u>qualidade e validação</u> comprovadas.</p>	<p><u>Proxy M+N</u></p> <p>- Expectativa (Tab. 51, <i>proxy</i> M+N).</p>	<p>- <u>Todas as cinco macrodimensões</u> de <u>questões validadas</u> (Expectativa e Percepção), além da <u>macrodimensão Acessibilidade adicional</u> (Percepção), na <i>proxy</i> M+N, tiveram <u>qualidade e validação asseguradas</u>, com validação convergente (AVE > 0,40),</p>

	- Percepção (Tab. 52, <i>proxy</i> M+N)	confiabilidade adequada (A.C. > 0,60 ou C.C. > 0,60), unidimensionalidade e validação discriminante (VCM < AVE).
- H17 ₀ : As <u>questões validadas</u> apresentam maiores <u>correlações</u> com suas respectivas <u>subdimensões</u> .	<u>Proxy M+N</u> - Expectativa (Tab. 105, Graf. 28, <i>proxy</i> M+N). - Percepção (Tab. 107, Graf. 30, <i>proxy</i> M+N).	- A <u>ampla maioria</u> das 18 <u>questões validadas</u> apresentaram <u>maiores correlações</u> com suas respectivas 12 <u>subdimensões</u> de questões validadas (Expectativa e Percepção), além da <u>subdimensão Acessibilidade</u> adicional (Percepção), na <i>proxy</i> M+N.
- H18 ₀ : As <u>subdimensões</u> de questões validadas apresentam <u>maiores correlações</u> com suas respectivas <u>macrodimensões</u> .	<u>Proxy M+N</u> - Expectativa (Tab. 106, Graf. 29, <i>proxy</i> M+N). - Percepção (Tab. 108, Graf. 31, <i>proxy</i> M+N).	- A <u>ampla maioria</u> das 12 <u>subdimensões</u> <u>questões validadas</u> apresentaram <u>maiores correlações</u> com suas respectivas cinco <u>macrodimensões</u> (Expectativa e Percepção), além da <u>macrodimensão Acessibilidade</u> adicional (Percepção), na <i>proxy</i> M+N.
- H19 ₀ : A <u>validação e adequação do modelo</u> da <u>Qualidade Global Percebida</u> , de <u>questões validadas</u> , foram comprovadas.	<u>Proxy M+N</u> - <u>Qualidade Global Percebida</u> - Expectativa (Tab. 53 e Fig. 10, <i>Proxy</i> M+N) - Percepção	- A <u>Qualidade Global Percebida</u> (mantidas as <u>cinco macrodimensões</u> de qualidade, na <i>proxy</i> M+N) teve <u>qualidade e validação asseguradas</u> , para a <u>Expectativa</u> e <u>Percepção</u> , além da <u>macrodimensão</u> adicional <u>Acessibilidade</u> (Percepção), com validação convergente (AVE > 0,40), confiabilidade adequada (A.C. > 0,60 ou C.C. > 0,60), unidimensionalidade e validação discriminante (VCM < AVE); - O ajuste da AFC foi considerado adequado, já que a razão entre a

	(Tab. 54 e Fig. 11, <i>Proxy</i> M+N)	estatística qui-quadrado e os graus de liberdade foi menor que 3, as estatísticas TLI e o CFI foram maiores ou próximos a 0,80 e o RMSEA do foi menor que o limite máximo de 0,10, na <i>proxy</i> M+N.
<i>Em síntese, a <u>Qualidade Global Percebida</u> teve <u>qualidade e validação asseguradas</u> e os <u>parâmetros de qualidade indicaram boa adequação</u>, sem necessidade de propor ajustes no modelo final com questões validadas, subdimensões e macrodimensões para a escala de qualidade e satisfação do SUS (figura 10 -Expectativa e figura 11 – Percepção).</i>		

Quadro 16- Síntese validações e conclusões de hipóteses sobre itens de medição de qualidade/satisfação (estudos 4 e 5)

7. CONCLUSÕES

A Lei Helói Chaves de 1923 pode ser considerada a primeira tentativa de organizar a saúde pública no Brasil. Sucederam-se melhorias incrementais de organização sanitária nas décadas de 1930 até meados de 1970, quando a crise da saúde pública brasileira impulsionou o movimento de reforma sanitária, que resultou na criação do SUS, na Constituição Federal de 1988. Este Sistema Nacional de Saúde (SUS) alcançou avanços notáveis nas garantias de atendimento, na organização sistêmica, na descentralização da gestão única, na participação social e em outros campos. Entretanto, passados 30 anos desse percurso estruturante do SUS, apesar de algumas iniciativas governamentais, ainda, permanecem incipientes as questões de governança sistêmica, de gestão de resultados e de avaliação e monitoramento de desempenho na saúde brasileira.

Por sua vez, a NGP ou Gerencialismo, tem exigido dos gestores públicos de saúde esforços de monitoramento de resultados, controladoria e responsabilização (*accountability*) dos gastos públicos, visto que muitos países da Europa e América Latina aplicam cerca de 9 a 10% do PIB-GDP em saúde (OECD, 2015a, 2015b).

No entanto, a regulação do sistema público de saúde do Brasil pauta-se basicamente na gestão da demanda e na garantia do acesso e na organização dos fluxos assistenciais, sem uma preocupação de desempenho nos pontos de atenção à saúde. Com isto, agrava-se a crise de recursos e de financiamento, não havendo transparência nos resultados de hospitais e demais serviços de saúde, quanto ao seu desempenho e resultados alcançados.

Nesse contexto, a presente investigação considerou relevante estudar o estabelecimento de mecanismos e instrumentos de mensuração de performance e monitoramento de resultados em saúde, como por exemplo, um *proxy* de um painel de indicadores e dimensões de performance, além de estruturar e validar uma escala inovadora de mensuração de qualidade e satisfação de usuários do SUS, que foi proposta para apuração do indicador de qualidade/satisfação de usuários do painel de indicadores desta investigação.

7.1. Descrição da abordagem

No primeiro escopo, o estudo destacou a evolução recente das teorias e modelos de avaliação na área de saúde, quanto à estruturação do campo da avaliação para a tomada de decisão e a medição de performance em organizações e serviços de saúde, em busca de construir

um quadro teórico-conceitual (*framework*) para a geração de indicadores de performance aplicável aos hospitais e serviços de saúde do SUS. Também analisou algumas metodologias de mensuração de resultados em saúde no cenário mundial como os projetos QIP e AHRQ (EUA), EFQM (Europa) e PATH (OMS), e projetos brasileiros como o PROADESS, PNASH, PNASS, PMAQ e as matrizes de indicadores como IDB-Ripsa e IDSUS (Brasil, 2011b), em busca de desenvolver um quadro teórico-conceitual (base teórico-metodológica) de avaliação de performance para organizações de saúde, aplicável às unidades do SUS.

Em seguida, na etapa empírica, após a construção da base teórico-lógico (*framework*), estruturou-se um processo colaborativo com gestores e profissionais para a captação de conhecimentos prévios e validação da percepção e valores atribuídos desses agentes chaves (*key stakeholders*), na perspectiva de relevância/importância (usabilidade para a tomada de decisão), para a construção de consensos quanto a um grupo de indicadores e dimensões de mensuração de desempenho em organizações e serviços de saúde de hospitais, policlínicas e outras unidades do SUS, tendo elaborado uma proposta de um *proxy* de painel de indicadores e dimensões multidimensionais, devidamente validado pelos agentes de decisão chaves, em uma amostra selecionada multicêntrica.

Em outro escopo dessa investigação, face à escassez de instrumentos padronizados e validados de mensuração de qualidade e satisfação no SUS, estudou-se algumas teorias de destaques na literatura quanto as bases de avaliação de qualidade percebida para a estruturação de escalas de medição de qualidade e satisfação de usuários de serviços de saúde em busca de estruturar uma base teórico-conceitual (*framework*) de mensuração de qualidade e satisfação em serviços de saúde e, elaborou-se uma proposta de estratégia de mensuração de avaliação de satisfação dos usuários. Para tal, elegeu-se o escopo de *experiência de serviços*, que está na base do enquadramento teórico da linha experiência de serviços para mensuração da satisfação de usuários e que resultou na escala SERVQUAL (Parasuraman et al., 1988, 1991b), modelo de escala, que apesar de importantes limitações elencadas neste estudo, foi adaptada nesta investigação para o contexto do SUS.

Ressalte-se que a escolha da SERVQUAL pautou-se na possível facilidade de incorporação à gestão pública brasileira, visto que na única experiência governamental registrada pelo Ministério do Planejamento do Brasil, destacaram-se as cinco dimensões proposta na escala SERVQUAL (confiabilidade, capacidade de resposta, segurança, empatia e tangibilidade) e as 22 questões originais do método que parecem ter sido objeto de traduções simplificadas e incorporadas diretamente na proposta de questionário vinculado ao IPPS, que

segundo Batista e Paiva (2014) foi base para um software que combina elementos das principais metodologias internacionais de medição de satisfação dos usuários.

Na etapa empírica desse escopo adicional procedeu-se a tradução e a adaptação transcultural da escala SERVQUAL em um novo instrumento (escala de medição de qualidade/satisfação de usuários do SUS) fundamentados no campo da translação de conhecimentos no qual as teorias consolidadas são alvos de transliteração à luz de experiências dos atores sociais envolvidos (Barbosa & Neto, 2017; Clavier et al., 2011; Hartz & John, 2008). Neste propósito, estruturou-se um processo colaborativo com especialistas e profissionais para a captação de conhecimentos prévios e validação da Percepção e valores atribuídos desses agentes chaves (*key stakeholders*), para constructos (ideias-forças), questões validadas, subdimensões e macrodimensões de qualidade percebida, com posterior validação de usuários, tendo estruturado e validado uma escala inovadora de mensuração de qualidade e satisfação de usuários do SUS.

Utilizou metodologias científicas de apoio à decisão, como *Design Research* e o método misto (qualitativo e quantitativo) para a abordagem analítica no envolvimento de agentes chaves (*key stakeholders*) no processo de captação e análise das respostas obtidas. Através de recursos da estatística descritiva, correlação quantílica e análise fatorial, foram analisadas as ponderações de itens de avaliação de performance, na perspectiva dos gestores e decisores e, também, de itens de análise de qualidade/ satisfação de usuários. De forma complementar, foram propostas algumas hipóteses sobre variáveis latentes e observadas destes itens que foram objeto de análise estatística, para testar as hipóteses, comparativamente às três possíveis variáveis moderadoras selecionadas (nível de atenção/cuidado de saúde, de complexidade da assistência e de cargo/função dos decisores ou de escolaridade, para usuários), processo inédito na literatura de painéis de indicadores e também quanto a escala de qualidade e satisfação de usuários da saúde.

A natureza da pesquisa caracterizou-se por uma metodologia avaliativa do tipo *observacional* (controle de possíveis influências intencionais no objeto pesquisado), *transversal*, com abordagem por método misto (qualitativo e quantitativo) de avaliação durante a etapa de validação metodológica e instrumental por decisores chaves da amostra (gestores, profissionais e usuários da saúde).

As etapas e fases operacionais de pesquisa, todas de matiz *prospectivas* e *transversais*, envolveram grupos focais (painéis de especialistas), grupos pilotos (gestores e profissionais) e, após a adequação dos instrumentos semiestruturados em fases sucessivas e incrementais, a

aplicação de pesquisa de campo, inclusive com uma fase de pesquisa à distância (*online*) de captação das preferências dos decisores chaves e, ainda, das dos usuários (no caso da avaliação da qualidade percebida).

7.2. Dificuldades encontradas

A primeira dificuldade foi a necessidade de estudar diversos modelos de avaliação de performance organizacionais (quadro 3, tópico 2.2.) pelo fato das bases teóricas-conceituais estudadas considerar a *eficácia* ou *performance* organizacional como um constructo multidimensional, subjetivo e socialmente construído, tratando-se de um juízo de valor das partes interessadas (*stakeholders*) que validam seus critérios, segundo Morin *et al.* (1994 apud Carvalho, 2007: 180).

Além disso, surgiu o desafio de um estudo ampliado de diversas metodologias e sistemas de indicadores de destaque mundiais (anexo 2), também no campo de saúde pública do Brasil (anexo 5) além dos mais relevantes da saúde privada brasileira (anexo 6), que foram sistematizadas quanto as possíveis potencialidades, virtudes, limitações e/ou desafios de implementação (tópico 2.4.3) visto que a multiplicidade de perspectivas e de complexidades envolvidas, além das diversidades de escopos e definições demandaram a análise dos principais quadros teóricos e metodológicos de avaliação relevantes da literatura. Segundo Viacava (2004: 153), na área de avaliação de resultados em saúde, a variedade de modelos de desempenho descrita na literatura “reflete a complexidade da questão, além da diversidade de definições, das dimensões englobadas, dos determinantes do desempenho, dos indicadores utilizados e dos critérios de julgamento” (Sicotte *et al.*, 1998). Por sua vez, Murray & Frenk (2000b) lembram que há diversidade dos modelos propostos para medir o desempenho do sistema de saúde (Aday, 1993; AHCPR, 1999; Hoffmeyer & Kluwer, 1994; C. W. Hsiao, 1995; Knowles *et al.*; OECD; OECD; Organization, 1999).

Outra dificuldade destacada foi a incipiência de publicações em avaliação e monitoramento de sistemas e serviços de saúde no Brasil, comprovada na revisão sistemática ampliada nas principais bases publicações de artigos científicos do Brasil (Scielo e Lilacs e Medline, via Biblioteca Virtual em Saúde - BVS) que localizou apenas 112 publicações no período de 1986 a 2018 (quadro 27, anexo 4). Quanto ao contexto da avaliação de organizações de saúde brasileira foi identificado a incipiência de modelos estruturados de avaliação de desempenho limitando-se a algumas experiências isoladas, por exemplo, PROADESS, IDB, RIPSAs, PNASS, IDSS, PROHASA/CQH e SIPAGEH (anexos 5 e 6). Estas incipiências de

publicações e programas oficiais resultaram em dificuldades de comparação dos resultados obtidos nesta investigação.

No outro escopo desta investigação, a primeira dificuldade enfrentada foi a multiplicidade de abordagens teórico-metodológicas para a mensuração de satisfação na área de serviços, considerado um constructo multifacetado e afetados por interações na entrega dos serviços, tendo sido necessário estudar algumas das diversas metodologias que têm sido desenvolvidas para estruturar uma escala de mensuração confiável de satisfação de usuários-utilizadores (quadro 20, anexo 2). Uma ampla revisão da literatura mundial aponta que a satisfação do usuário é multifacetada e de difícil mensuração (M. Esperidião & Trad, 2005; M. A. Esperidião, 2009). Por sua vez, Vinagre (2008) enfatiza a complexidade de aferição dos serviços pelos aspectos da *intangibilidade* (ausência de atributos físicos) dos serviços (Parasuraman et al., 1991b) e o fato dos serviços constituírem-se essencialmente enquanto processos, sendo afetados por interações e atividades (Grönroos, 1990).

Diante das muitas definições do constructo *satisfação* e dos inúmeros quadros conceptuais (*framework*) propostos na literatura para a avaliação da *satisfação*, neste trabalho, resolveu-se abordar a satisfação enquanto resposta a um processo de avaliação da experiência de consumo, e considerar tanto o resultado quanto o processo como elementos importantes para a satisfação dos consumidores com enfoque nos antecedentes/preditores da satisfação, ou seja, amparado pela teoria da atribuição causal (Bitner et al., 1994). Noutro prisma de dificuldades deste estudo, não foi possível abordar as possíveis limitações da escala SERVQUAL selecionada, *a priori*, apesar de ter registrado possíveis limitações relatadas na literatura (Chahal & Kumari, 2011, 2012; Cronin & Taylor, 1992) além da concepção de outros instrumentos (P. L. Ferreira, 1993; P. L. Ferreira et al., 2001; P. L. Ferreira & Mendes, 1997; P. L. Ferreira & Raposo, 2006; P. L. Ferreira et al., 2017; Pedro et al., 2010).

Outra dificuldade encontrada foi a ausência de estudos comparativos dos resultados obtidos com outros trabalhos na área de aplicação de instrumento ou escala validada e consolidada de mensuração de satisfação e/ou qualidade em saúde visto que, no Brasil, ainda não se registra nenhuma metodologia consolidada e/ou usada em larga escala dentro do escopo de satisfação desta investigação (quadro 21, anexo 2). Na pesquisa integrativa da literatura ampliada (tabela 110, anexo 1) e na pesquisa sistemática nas principais bases de artigos científicos do Brasil (quadro 27, anexo 4), não se identificou nenhuma metodologia robusta de avaliação de qualidade e satisfação através de instrumentos padronizados e validados e/ou de

larga utilização no contexto de saúde do Brasil. Achados alinhados com os resultados de uma revisão sistemática extensiva Volpato (2014), entre 1990 e 2013.

No entanto, por esta investigação envolver um processo colaborativo com especialistas, profissionais e gestores da saúde, uma das maiores dificuldades foi que os conhecimentos prévios destes atores chaves mostraram-se dependentes de informações captadas na literatura, adicionadas em cada etapa de consulta, em ambos os escopos do estudo. Quanto aos itens de avaliação de performance, o processo incremental de melhoria dos instrumentos necessitou de diversos aportes de conceitos de indicadores e dimensões de performance, visto que desde as primeiras fases houve limitações de conhecimentos referidos pelos atores chaves. Quanto aos itens dos instrumentos de mensuração de qualidade havia uma polissemia de termos e conceitos difusos, observados nas primeiras abordagens nos painéis de especialistas, possivelmente, pela escassez histórica de modelos de aferição de qualidade e satisfação no SUS, pelo que foi necessário selecionar de um instrumento estruturado (escala SERVQUAL) para estabelecer um parâmetro basal de análise e construção da escala de mensuração de qualidade e satisfação no contexto do SUS.

Destacaram-se, ainda, as dificuldades de acesso aos especialistas, profissionais e gestores da saúde para responder as entrevistas semiestruturadas, que foi superada pela parceria com algumas instituições de formação de gestores (painéis de especialistas e grupos *Dephi*), além de gestores de topo de secretarias estaduais e municipais (aplicação de campo) e, por fim, com as associações de gestores (Conass e Conasems) na investigação à distância para a validação final dos indicadores e dimensões de performance. Outra dificuldade foi a reduzida disponibilidade de usuários da saúde na abordagem de campo de 2.547 usuários em atendimento em 74 unidades do SUS que demandou uma mobilização especial através dos gestores das unidades de saúde e a capacitação de uma equipe de campo para aplicação em unidades de saúde em diferentes cidades.

7.3. Limitações do trabalho empírico

Uma das limitações iniciais do trabalho empírico foram as fragilidades de conhecimentos prévios identificadas juntos aos atores chaves (especialistas, profissionais e gestores da saúde) que se mostraram incipientes em ambos os escopos do estudo. Esta realidade demandou a seleção, *a priori*, de um modelo teórico-conceitual (*framework*) para condução das pesquisas iniciais dos painéis de indicadores e dimensões de performance, visto que os respondentes revelaram muitas limitações de conhecimentos nos painéis de especialistas. Pela

mesma realidade, no segundo escopo do estudo enfrentou-se limitações de construir um instrumento com um grupo de itens de mensuração de qualidade e satisfação a partir dos conhecimentos prévios de especialistas e profissionais da saúde, tendo inclusive limitações de comparação crítica prévia com outras proposições teórico-metodológicas da literatura, pelo que se decidiu iniciar o estudo pela seleção de um instrumento estruturado (escala SERVQUAL) para estabelecer um parâmetro basal de análise e construção da escala de mensuração de qualidade e satisfação do SUS. Em ambos os casos, estas limitações foram atenuadas pela liberdade plena dos respondentes efetivarem agregações e/ou mudanças durante o processo de estruturação e qualificação dos instrumentos de coleta de dados, que tiveram adequações incrementais em diversas fases sucessivas.

Outra limitação das etapas empíricas da investigação foi a ausência de um processo de amostragem aleatória, tendo sido necessária a seleção de participantes por disponibilidade de participação ou acessibilidade (não probabilística, com limitações de generalização de resultados), apesar de algumas cautelas na seleção dos respondentes para evitar vieses, e no caso de aplicação de campoescala de qualidade percebida, tendo implementado a amostragem de usuários do tipo estratificado proporcional, com uma amostra equilibrada ao selecionar em torno de 25 a 30 usuários-utilizadores em cada unidade avaliada, porém, sem controle de outras variáveis como gênero, faixa etária, escolaridade.

Outra fragilidade do estudo empírico foram as limitações dos Sistemas de Informações em Saúde (SIS), enquanto fontes de informações para a composição de indicadores em saúde, visto que por suas limitações no âmbito do SUS funcionaram como um delimitador do escopo de indicadores avaliados, na última etapa de validação do *proxy* de indicadores e dimensões de performance, pois a fim garantir a usabilidade não se agregou nenhum indicador que não tivesse garantia de existência de dados nos SIS do SUS.

Ressalte-se, ainda, as limitações naturais do processo empírico de medição das expectativas de clientes-consumidores relatadas na literatura. Independente do método selecionado, enfrenta-se pelo menos três desafios. O primeiro deles reflete as dificuldades para investigar a fidedignidade e a validade de escalas usadas para medições diretas de constructos como a satisfação, que é um constructo complexo e multifacetado, pois depende da relação preço/valor e imagens mentais do cliente (Dillon, 1997). Outro desafio ocorre por usuários diferentes dar importâncias relativas diferentes para as diversas dimensões contidas na satisfação (produto, serviço, comunicação, etc), e assim, para se calcular o índice de satisfação geral dos clientes-consumidores, há de se considerar as respostas a todas as variáveis

apresentadas no modelo, bem como suas diferentes importâncias relativas. Há, ainda, a necessidade de definição e inclusão de variáveis de “fechamento” de cada bloco/dimensão e da satisfação geral que visam estabelecer uma situação de dependência das variáveis para o emprego da técnica de análise estatística. Sem estas variáveis seria impossível estimar-se a importância de cada variável na satisfação com a análise de uma dimensão ou com a satisfação geral.

Destaque-se também as limitações naturais do processo de coleta de dados por instrumentos semiestruturados visto que todos os itens analisados estavam disposto em uma escala ordinal, discreta e finita (*Likert* de 5 níveis), além do desafio de definir a importância relativa entre os atributos e sua associação entre a avaliação de cada atributo e a avaliação geral relatada, visto que além de medir a satisfação dos clientes-consumidores com cada atributo, um desafio adicional pode ser definir a importância relativa entre os atributos (Fontnote et al., 2007), que pode ser obtida de duas formas: *Declarada* ou *Derivada*. A *Importância Declarada* é aquela na qual se solicita ao próprio entrevistado que avalie a importância dos atributos, para tal, geralmente se recorre a determinadas escalas, como a de *diferencial semântico*, de *ranking* e de *soma constante*; enquanto a *importância Derivada* é obtida a partir da associação por métodos estatísticos entre a avaliação de cada atributo e a avaliação geral.

Dessa forma, cientes de tais limitações, nesta investigação mediu-se a *Importância Declarada* por uma escala de *Likert* de 5 níveis (*diferencial semântico*) além dos cálculos da *Importância Derivada* através de técnicas de estatística descritiva e inferencial para avaliar os itens da escala de qualidade percebida concebida (tópicos 6.2 e 6.4).

No processo analítico, uma importante limitação foi o nível de respostas obtidas (233 - cerca de 5,6 % do universo de decisores) na consolidação final do painel de indicadores e dimensões de performance, apesar de três ciclos de envio de *e-mails*, e, também, o alcance de uma amostra limitada 2.547 usuários (25 a 30 usuários por unidade de saúde) na análise final da escala de qualidade e satisfação dos usuários do SUS.

Por fim, destaca-se a aplicação de questionários na pesquisa de campo com itens analisados validados apenas pelas médias ponderadas obtidas na análise estatística descritiva. Na transição das fases de consolidação dos instrumentos através de painéis de especialistas e grupos *Delphi* de profissionais e usuários mantiveram-se algumas variáveis que não foram validadas posteriormente nas análises inferenciais (indicadores e dimensões para o painel de indicadores e dimensões de performance e questões, subdimensões e macrodimensões para a escala de qualidade e satisfação de usuários), visto que esta análise foi realizada, *a posteriori*, nas respostas obtidas na pesquisa de campo, que poderia potencializar a análise dos possíveis

impactos de tais exclusões, possivelmente, uma das maiores limitações desta investigação que se pode propor para a continuidade deste estudo.

7.4. Investigação futura

Para a continuidade desta investigação pode-se sugerir a análise da generalização dos resultados obtidos pela aplicação dos Artefatos gerados e validados em outros extratos amostrais, preferencialmente por amostragem aleatória de gestores, profissionais e usuários. Propõem-se a ampliação da amostra em todos os extratos amostrais, a fim de permitir outros processos analíticos e comparações entre diferentes classes de gestores, profissionais e usuários, bem como de diferentes tipos de unidades de saúde.

Ao nível instrumental, propõe-se estruturar uma plataforma de aplicação digital (*online*) para otimizar o processo de aplicação e captação de respostas de especialistas, profissionais e usuários.

Outro avanço proposto é o aprofundamento analítico dos itens considerados importantes/relevantes (médias ponderadas elevadas) da análise descritiva, mas excluídos pela análise inferencial e, possivelmente, aplicar novos instrumentos apenas com os itens analisados e validados pelas análises inferenciais, que foram sintetizados pelos modelos finais estruturados na figura 7 (*proxy* E+F) e a figura 8 (*proxy* G) para os indicadores e dimensões do *proxy* de painel de indicadores e dimensões de performance e nas figuras 10 e 11 (*proxy* M+N) para as questões, subdimensões e macrodimensões para a escala de qualidade e satisfação de usuários do SUS.

Sugere-se, ainda, a análise aprofundada da correlação positiva (ou negativa) entre as perspectivas de Expectativa e Percepção que deve ser melhor apreciada, visto que não se pode asseverar posição sobre os achados obtidos, em especial, quanto a fragilidade dos resultados obtidos na perspectiva Expectativa, haja visto que não há consenso na literatura quanto a sua mensuração.

Além disso, em estudos futuros de novos instrumentos de qualidade e satisfação percebida, propõe-se explorar outras variáveis moderadoras tais como nível de renda, habitantes de regiões metropolitanas ou do campo, dentre outras que a literatura aponte como potencial fator moderador da qualidade/satisfação percebida pelos usuários de saúde.

7.5. Contributos desta investigação

No primeiro escopo do estudo, implementou-se um processo colaborativo pelo envolvimento dos decisores chaves (*decision making*), na estruturação de um *proxy* de painel de indicadores e dimensões de performance a fim de assegurar a análise e validação dos itens do painel de indicadores além de legitimidade para a usabilidade do instrumento final em organizações e serviços de saúde.

Na parte empírica, de início, nos estudos 1 a 3, os resultados foram agregados em subcategorias (domínios de avaliação e indicadores de Performance) e categorias (macrodomínios de avaliação e dimensões de performance) de análise da performance organizacional através de um processo empírico colaborativo de painéis de especialistas e grupos pilotos de gestores e de campo de decisores que livremente expressaram seus conhecimentos prévios e atribuíram importância/relevância para a tomada de decisão de cada um dos itens apresentados a partir de conceitos globais e genéricos apresentados, tendo sido estruturados e aplicadas análises descritivas (tópico 6.1.1.) em modelos inovadores de mensuração de performance com 24 domínios e seis macrodomínios de avaliação (*proxy C*), outro com 20 indicadores e sete dimensões de performance (*proxy E+F*) além de outro com 56 indicadores e sete dimensões de performance (*proxy G*).

Em seguida, aplicou-se métodos de análise inferencial (tópico 6.1.2.), a fim de verificar a validade das demais hipóteses propostas e consolidar os modelos desenhados nesta investigação para os itens de avaliação de performance (domínios e macrodomínios de avaliação e indicadores e dimensões de performance) (estudos 1 a 3), conforme métodos de análise inferencial detalhados no início do tópico 6, que dividiram-se em quatro análises de resultados: estudo das cargas fatoriais dos domínios *versus* macrodomínios de avaliação estudados (*proxy C*) e indicadores *versus* dimensões de performance (*proxy E+F* e *proxy G*), validação por critérios de qualidade e validade dos macrodomínios de avaliação (*proxy C*) e de dimensões de performance (*proxy E+F* e *proxy G*), análise da qualidade (ajustamento) do modelo final resultante, e, análise das correlações entre as variáveis e suas influências na Performance Global Percebida nas três *proxies* estudadas (*proxy C*, *proxy E+F* e *proxy G*).

No estudo 1, em síntese, nas fases 1 a 3, todos os 20 domínios e os quatro macrodomínios de avaliação, e, na fase 4, todos os 24 domínios e os cinco macrodomínios de avaliação, foram relevantes/importantes (média superior ou próxima de 4). Por sua vez, nas fases 1 a 3, não houve significância estatística para considerar como fatores mediadores os níveis de atenção, de complexidade e de carga função, na ampla maioria dos quatros

macrodomínios de avaliação. Na fase 4, não houve significância (estatística) para considerar como mediadores os níveis de atenção, de complexidade e de escolaridade de usuários), na ampla maioria dos cinco macrodomínios de avaliação, exceto para Quantitativos e Efeitos, para os níveis de atenção.

Por sua vez, os resultados obtidos mostraram que a Performance Global Percebida média foi de 4,16 e a mediana de 4,18 e que 75% das respostas atribuídas resultaram em valoração superior a 3,88 (cerca de 4, muito importante/relevante). Pode-se observar uma variabilidade na valoração média da variável latente Performance Global Percebida. Porém, não houve diferença significativa (valor- $p > 0,050$) para considerar as variáveis nível de atenção à saúde, de complexidade da assistência ou de cargo/função enquanto fator mediador da Performance Global Percebida, na *proxy C* ($n = 112$) (Tabela 63, apêndice 9, tópico 9.1.2.1).

Assim, da análise descritiva da *proxy C*, sintetizou-se que os domínios e macrodomínios da Performance Global Percebida obtiveram médias elevadas, sendo considerados relevantes/importantes pelos respondentes e que as três possíveis variáveis mediadoras analisadas não foram consideradas como fatores mediadores de nenhum dos macrodomínios e da Performance Global Percebida.

Na análise inferencial, nos resultados da fase 4, no modelo final, após análise e exclusão de sete domínios de avaliação, todos os demais 17 domínios de avaliação que permaneceram na análise, e, todos os cinco macrodomínios de avaliação apresentaram carga fatorial superior a 0,50 (ou suas permanências não impediram a validação convergente da AFC), na *proxy C*. A confiabilidade e a validade dos cinco macrodomínios de avaliação (Quantitativos, Qualitativos, Efeitos, Satisfação e Sistêmico/Estratégico) e da variável latente Performance Global Percebida foram asseguradas, uma vez que todos apresentaram validação convergente ($AVE > 0,40$), confiabilidade adequada ($A.C. > 0,60$ ou $C.C. > 0,60$), unidimensionalidade e validação discriminante ($VCM < AVE$). Foi observado que os parâmetros e índices de ajustamento e qualidade dos modelos de equações estruturais indicaram bom ajuste, já que a razão entre a estatística qui-quadrado e os graus de liberdade foi menor que 3, as estatísticas TLI e o CFI foram maiores que 0,80 e o RMSEA do foi menor que o limite máximo de 0,10. Para concluir a análise inferencial, a figura 63 sintetizou os resultados apresentados na Tabela 62 (apêndice 9, tópico 9.1.2.1), nos quais foi possível observar que os domínios de avaliação apresentaram maiores correlações com seus respectivos macrodomínios, na *proxy C* ($n = 112$) (Tabela 65, apêndice 9, tópico 9.1.2.2).

Diante destes resultados, a figura 6 (*proxy C*) sintetizou o ajuste da Análise Fatorial Confirmatória (AFC) e a modelagem para a variável latente Performance Global Percebida. Dessa forma, a ampla maioria das hipóteses iniciais (quadro 15, tópico 6.3.1.) foram validadas e confirmaram a validação dos itens de mensuração de performance testados (24 domínios e cinco macrodomínios de avaliação), na *proxy C*, que consolidou uma aproximação de um painel de indicadores e dimensões de performance pelos conhecimentos prévios dos decisores chaves, mas com reduzido alinhamento com a literatura.

Nos estudos 2 e 3, na visão de profissionais, gestores e decisores, em síntese, nas fases 1 a 3, todos os 20 indicadores e as sete dimensões de performance, e, na fase 4, todos os 56 indicadores e as sete dimensões de performance, foram muito relevantes/importantes (média superior ou próxima de 4). Por sua vez, nas fases 1 a 3, não houve tendência de mediação de valorações médias obtidas na ampla maioria das dimensões de performance quanto aos níveis de atenção à saúde, de complexidade da assistência e de cargo/função dos especialistas e profissionais, com algumas exceções: quanto aos níveis de atenção à saúde (exceto nas dimensões Produtividade Assistencial e Efetividade) e de cargo/função (exceto na dimensão Efetividade), na *proxy E+F*. Na fase 4, não houve tendência de mediação de valorações médias obtidas na ampla maioria das dimensões de performance quanto aos níveis de atenção à saúde, de complexidade da assistência e de cargo/função dos especialistas e profissionais, exceto na dimensão Produtividade Assistencial, na *proxy G*.

Por sua vez, os resultados obtidos mostraram que a Performance Global Percebida média foi de 4,04 e a mediana de 4,04 e que 75% das respostas atribuídas resultaram em valoração superior a 3,96 (cerca de 4, muito importante/relevante). Pode-se observar uma variabilidade na valoração média da variável latente Performance Global Percebida. Porém, não houve diferença significativa (valor- $p > 0,050$) para considerar as variáveis nível de atenção à saúde, de complexidade da assistência enquanto fator mediador da Performance Global Percebida. No entanto, houve diferença significativa (valor- $p = 0,019$) quanto a influência do nível de cargo/função na valoração da Performance Global Percebida. Ao fazerem-se as comparações múltiplas, houve diferença significativa (valor- $p = 0,015$) entre os de cargo/função Sistêmicos e Organizacionais, sendo que os indivíduos de cargo/função Sistêmicos apresentaram Performance Global Percebida significativamente maior, na *proxy E+F* ($n = 112$) (Tabela 65, apêndice 9, tópico 9.1.2.2).

Da mesma forma, os resultados obtidos mostraram que a Performance Global Percebida média foi de 4,22 e a mediana de 4,22 e que 75% das respostas atribuídas resultaram em valoração superior a 4,45 (acima de 4, muito importante/relevante). Pode-se observar uma

variabilidade na valoração média da variável latente Performance Global Percebida. Porém, não houve diferença significativa (valor- $p > 0,050$) para considerar as variáveis nível de atenção à saúde, de complexidade da assistência e de cargo/função ocupados enquanto fator mediador da Performance Global Percebida. na *proxy* G (n = 233)(Tabela 67, apêndice 9, tópico 9.1.2.3).

Assim, da análise descritiva das *proxies* E+F e G, sintetizou-se que os indicadores e dimensões da Performance Global Percebida obtiveram médias elevadas, sendo considerados muito relevantes/importantes pelos respondentes e que as três possíveis variáveis mediadoras analisadas puderam ser consideradas como fatores mediadores de nenhuma das macrodimensões e da Performance Global Percebida, com raras exceções. No entanto, houve diferença significativa (valor- $p = 0,019$) quanto a influência do nível de cargo/função, sendo que os indivíduos de cargo/função Sistêmicos apresentaram Performance Global Percebida significativamente maior, na *proxy* E+F.

Na análise inferencial, nos resultados da fase 3 (*proxy* E+F) e da fase 4 (*proxy* G), no modelo final, após a análise e exclusão de seis indicadores de performance e de uma dimensão de performance (Produtividade Assistencial), todos os demais 14 indicadores e as seis dimensões de performance que permaneceram na análise (Eficiência, Efetividade, Acesso, Impacto, Qualidade Técnica/Científica e Sistêmico/Estratégico) apresentaram carga fatorial superior a 0,50 (ou suas permanências não impediram a validação convergente da AFC), na *proxy* E+F. Por sua vez, na *proxy* G, após a análise e exclusão de sete indicadores de performance e de uma dimensão de performance (Eficiência), todos os demais 49 indicadores e as seis dimensões de performance que permaneceram na análise (Produtividade Assistencial, Efetividade, Acesso, Impacto, Qualidade Técnica/Científica e Sistêmico/Estratégico) apresentaram carga fatorial superior a 0,50 (ou suas permanências não impediram a validação convergente da AFC). Em ambas as análises (*proxy* E+F e *proxy* G) a confiabilidade e a validade das sete dimensões de performance e da variável latente Performance Global Percebida foram asseguradas, uma vez que todas apresentaram validação convergente ($AVE > 0,40$), confiabilidade adequada ($A.C. > 0,60$ ou $C.C. > 0,60$), unidimensionalidade e validação discriminante ($VCM < AVE$). Foi observado que os parâmetros e índices de ajustamento e qualidade dos modelos de equações estruturais indicaram bom ajuste, já que a razão entre a estatística qui-quadrado e os graus de liberdade foi menor que 3, as estatísticas TLI e o CFI foram maiores que 0,80 e o RMSEA do foi menor que o limite máximo de 0,10. Para concluir a análise inferencial, a figura 64 sintetizou os resultados apresentados na Tabela 64 (apêndice 9, tópico 9.1.2.2), nos quais foi possível observar que os indicadores de performance apresentaram

maiores correlações com suas respectivas dimensões, na *proxy* E+F (n = 112). Da mesma análise inferencial, a figura 65 sintetizou os resultados apresentados na Tabela 66 (apêndice 9, tópico 9.1.2.3), nos quais foi possível observar que os indicadores de performance apresentaram maiores correlações com suas respectivas macrodimensões, na *proxy* G (n = 233).

Diante destes resultados, a figura 7 (*proxy* E+F) e a figura 8 (*proxy* G) sintetizaram os ajustes da Análise Fatorial Confirmatória (AFC) e a modelagem para a variável latente Performance Global Percebida. Dessa forma, a ampla maioria das hipóteses iniciais (quadro 15, tópico 6.3.1.) foram validadas e confirmaram a validação dos itens de mensuração de performance testados (14 indicadores e seis dimensões de performance, na *proxy* E+F, e, 49 indicadores e seis dimensões de performance, na *proxy* G), que consolidou uma nova aproximação de um painel de indicadores e dimensões de performance pelos conhecimentos dos grupos *delphi* de gestores, com um primeiro realinhamento com a literatura (*proxy* E+F) e por fim, validou a versão final do *proxy* de painel de indicadores e dimensões de performance desta investigação, devidamente realinhado com os quadros conceituais (*framework*) e expansão de novos indicadores, sem alterações das dimensões de performance, que resultou no *proxy* de painel de indicadores e dimensões de performance definitivo (*proxy* G).

Dessa forma, concluiu-se o mapeamento das categorias de macrodomínios e domínios de avaliação, seguidos pela etapa de análise e consolidação de painéis de indicadores e dimensões de performance, além dos reflexos sobre uma *variável latente* Performance Global Percebida, medidos por suas magnitudes (cargas fatoriais), significâncias estatísticas e suas correlações em modelos estruturais validados (figuras 6, 7 e 8). Destacam-se os aspectos inéditos na literatura de avaliação de performance, da colaboração com decisores chaves na valoração da relevância/importância e estruturação de um *proxy* de painel de indicadores e dimensões de performance do SUS, além da validação multicêntrica de modelos finais através de análise de cargas fatoriais, validade e confiabilidade dos itens, e, ainda, dos testes de possíveis fatores mediadores na valoração.

Em outro escopo desta investigação, a partir das 22 questões e cinco dimensões do instrumental original SERVQUAL proposto por Parasuraman *et al.* (1991a) e demais literaturas consultadas, executou-se a ponderação de 40 constructos, 22 subdimensões de constructos e cinco macrodimensões de qualidade, com validação de grupos pilotos (195 especialistas e profissionais de saúde). Estes 40 constructos foram base para construção de um novo questionário estruturado com 20 questões fechadas, 13 subdimensões e seis macrodimensões de qualidade, validados por grupos pilotos (62 profissionais e gestores de saúde) e, Em seguida, consolidada por uma pesquisa de campo com 506 usuários do SUS, em atendimento em 15

unidades de saúde da cidade de Itaitinga-Ceará. Por fim, foi validada uma escala inovadora de medição de qualidade/satisfação percebida no SUS (escala *Quality Saude*) sendo aplicada em 2.547 usuários em atendimento distribuídos em 74 unidades do SUS de quatro cidades do Ceará.

Nos estudos 4 e 5, os resultados foram respectivamente agregados em subcategorias (constructos e questões validadas) e categorias (subdimensões de constructos e subdimensões de questões validadas) e macrocategorias (macrodimensões de qualidade) de mensuração de qualidade/satisfação percebida através de um processo empírico colaborativo de painéis de especialistas e grupos pilotos de gestores e de campo de usuários que livremente expressaram seus conhecimentos prévios e atribuíram importância/relevância para o julgamento de cada um dos itens apresentados a partir de conceitos globais e genéricos apresentados, tendo sido estruturados e aplicadas análises descritivas (tópico 6.2.1.) em modelos inovadores de mensuração de qualidade/satisfação percebida com 40 domínios, 22 subdimensões de domínios e cinco macrodimensões de qualidade (*proxy J*) e outro com 20 questões validadas, 13 subdimensões de questões validadas e seis macrodimensões de qualidade (*proxy M+N*). Algumas das hipóteses propostas quanto aos resultados descritivos foram verificadas.

Em seguida, aplicou-se métodos de análise inferencial (tópico 6.2.2.), a fim de verificar a validade das demais hipóteses propostas e consolidar os modelos desenhados nesta investigação dos itens de mensuração de qualidade percebida (constructos, questões validadas, subdimensões e macrodimensões de qualidade) (estudos 4 e 5), conforme métodos de análise inferencial detalhados no início do tópico 6, que dividiram-se em quatro análises de resultados: estudo das cargas fatoriais dos constructos *versus* subdimensões de constructos e macrodimensões de qualidade (*proxy J*) e questões validadas *versus* subdimensões e macrodimensões de qualidade (*proxies M e N*), validação por critérios de qualidade e validade das macrodimensões de qualidade (*proxy J*) e das macrodimensões de questões validadas de qualidade (*proxies M e N*), análise da qualidade (ajustamento) do modelo final resultante, e, análise das correlações entre as variáveis e suas influências na Qualidade Global Percebida nas três *proxies* estudadas (*proxy J e proxies M e N*).

No estudo 4, em síntese, nas fases 1 a 3, na visão de especialistas e profissionais, a ampla maioria das subcategorias (40 constructos), das categorias (22 subdimensões de constructos) e das macrocategorias (cinco macrodimensões de qualidade - Tangibilidade, Confiabilidade, Presteza, Segurança e Empatia), foram muito relevantes/importantes (média superior ou próxima de 4). Na fase 4, na visão de usuários, em síntese, a ampla maioria desses mesmos constructos, subdimensões de constructos e macrodimensões de qualidade, foram

relevantes/importantes (média superior ou próxima de 3). Por sua vez, nas fases 1 a 3, nas respostas de especialistas e profissionais, houve significância estatística para considerar como fatores mediadores os níveis de atenção, de complexidade e de cargo função, na ampla maioria das 22 subdimensões de constructos (entre 12 a 20, conforme o foco da análise) e de todas as cinco macrodimensões de qualidade. Na fase 4, nas respostas de usuários, na ampla maioria das 22 subdimensões de constructos, houve uma tendência de influência das variáveis mediadoras nível de atenção à saúde, nível de complexidade da assistência e nível de escolaridade de usuários (entre 14 a 18, conforme o foco da análise) e houve significância estatística para considerar o nível de atenção como fator moderador em todas as cinco macrodimensões de qualidade, mas não houve tendência de moderação dos níveis de complexidade assistencial e de escolaridade do usuário na maioria das cinco macrodimensões. Nas quatro fases do estudo não se aplicou nenhuma análise de possíveis reflexos na variável latente Qualidade Global Percebida.

Assim, da análise descritiva da *proxy* J, sintetizou-se que os 40 constructos, as 22 subdimensões e as cinco macrodimensões de constructos obtiveram médias elevadas, sendo considerados muito relevantes/importantes pelos respondentes e que as três possíveis variáveis mediadoras analisadas foram consideradas como fatores mediadores da ampla maioria das subdimensões (12 a 20, conforme o foco da análise), mas não houve tendência de moderação dos níveis de complexidade assistencial e de escolaridade do usuário na maioria das cinco macrodimensões, porém as valorações das cinco macrodimensões foram moderadas pelo nível de atenção de atendimento da unidade.

Explorou-se, ainda, uma análise descritiva da variável Qualidade Global Percebida, inclusive quanto as possíveis influências das variáveis nível de atenção à saúde, nível de complexidade da assistência e nível de escolaridade de usuários. Assim, a Tabela 104 (apêndice 9, tópico 9.2.2.1), mostra que a Qualidade Global Percebida média foi de 3,17 e a mediana foi de 3,01 e que 75% das respostas atribuídas resultaram em valoração superior a 3,00 (cerca de 3, importante/relevante). Pode-se observar uma variabilidade na valoração média da variável latente Qualidade Global Percebida. Houve diferença significativa (valor- $p \leq 0,050$) para considerar as variáveis nível de atenção à saúde e de escolaridade do usuário enquanto fator mediador da Qualidade Global Percebida, no entanto, não houve diferença significativa (valor- $p = 0,104$) quanto ao nível de complexidade na *proxy* J ($n = 506$). Destaca-se nos dois fatores mediadores que o nível Terciário de atenção obteve a menor valoração média (3,13), enquanto o nível Secundário de atenção a maior valoração média (3,23), por sua vez, o nível Médio de

escolaridade obteve a menor valoração média (3,11) e o nível Basilar a maior valoração média (3,25), na *proxy J*.

Na análise inferencial, nos resultados da fase 4, no modelo final, após a análise e exclusão de dois constructos, de sete subdimensões de constructos e de uma macrodimensão de qualidade (Tangibilidade), os demais 38 constructos, as demais 15 subdimensões de constructos e as demais quatro macrodimensões de qualidade que permaneceram na análise (Confiabilidade, Presteza, Segurança e Empatia) apresentaram carga fatorial superior a 0,50 (ou suas permanências não impediram a validação convergente da AFC), na *proxy J*. A confiabilidade e a validade das 22 subdimensões de constructos, das cinco macrodimensões de qualidade e da variável latente Qualidade Global Percebida foram asseguradas, uma vez que todas apresentaram validação convergente ($AVE > 0,40$), confiabilidade adequada ($A.C. > 0,60$ ou $C.C. > 0,60$), unidimensionalidade e validação discriminante ($VCM < AVE$). Foi observado que os parâmetros e índices de ajustamento e qualidade dos modelos de equações estruturais indicaram bom ajuste, já que a razão entre a estatística qui-quadrado e os graus de liberdade foi menor que 3, as estatísticas TLI e o CFI foram maiores que 0,80 e o RMSEA do foi menor que o limite máximo de 0,10. Para concluir a análise inferencial, as figuras 66 e 67 sintetizaram os resultados apresentados nas Tabelas 102 e 103, respectivamente (apêndice 9, tópico 9.2.2.1), nos quais foi possível observar que os constructos apresentaram maiores correlações com suas subdimensões de constructos e as subdimensões apresentaram maiores correlações com suas respectivas macrodimensões de qualidade percebida, na *proxy J* ($n = 506$).

Diante destes resultados, a figura 9 (*proxy J*) sintetizou o ajuste da Análise Fatorial Confirmatória (AFC) e a modelagem para a variável latente Qualidade Global Percebida. Dessa forma, a ampla maioria das hipóteses iniciais (quadro 16, tópico 6.3.2.) foram validadas e confirmaram a validação dos itens de mensuração de qualidade testados (38 constructos, 15 subdimensões de constructos e quatro macrodimensões de qualidade, na *proxy J*), que consolidou uma aproximação de um grupo dos itens de mensuração de qualidade, conforme os conhecimentos prévios dos especialistas e profissionais participantes (*proxy H+I*) e posterior análise e consolidação por 506 usuários (*proxy J*). Pode-se destacar que cerca de um terço das subdimensões de constructos não foram validadas, achados alinhados com a revisão proposta pelo painel de especialistas e profissionais para estruturar as questões do instrumento do estudo 5.

No estudo 5, em síntese, nas fases 1 a 3, na visão de especialistas e profissionais, a ampla maioria das subcategorias (20 questões validadas), das categorias (13 subdimensões de questões

validadas) e das macrocategorias (seis macrodimensões de qualidade - Tangibilidade, Confiabilidade, Presteza, Segurança, Empatia e Acessibilidade), foram relevantes/importantes (média superior ou próxima de 3). Na fase 4, na visão de usuários, em síntese, todas essas mesmas questões validadas, subdimensões de questões validadas e macrodimensões de qualidade foram muito relevantes/importantes (média superior a 3). Por sua vez, nas fases 1 a 3, nas respostas de especialistas e profissionais, houve significância estatística para considerar como fatores mediadores os níveis de atenção, de complexidade e de cargo função, na ampla maioria (entre 11 a 12, conforme o foco de análise) das 13 subdimensões de questões validadas, e, na ampla maioria (entre 5 e 6, conforme o foco de análise) das seis macrodimensões para perspectiva de Percepção e em raras exceções para a observação de Expectativa - tanto para subdimensões quanto para macrodimensões. Na fase 4, nas respostas de usuários, na ampla maioria das 13 subdimensões de questões validadas, houve uma tendência de influência das variáveis mediadoras nível de atenção à saúde, de complexidade da assistência e de escolaridade de usuários na ampla maioria (10 a 13, conforme o foco de análise) das 13 subdimensões de questões validadas e na ampla maioria (entre 4 a 6, conforme o foco de análise) das seis macrodimensões, para as perspectivas de Expectativa e de Percepção, sendo exceções que quanto ao nível de escolaridade para a observação de Expectativa houve significância para apenas três subdimensões e uma macrodimensão.

Pode-se destacar que a fase 4 do estudo 5 teve amostra ampliada de usuários, sendo alvo das verificações das hipóteses dos itens da escala de qualidade/satisfação percebida, e que, praticamente, todas as categorias de análises (13 subdimensões de questões validadas) e macrocategorias de análises (seis macrodimensões de qualidade) obtiveram significância estatística para os três fatores mediadores (nível de atenção à saúde, de complexidade da assistência e de cargo/função dos entrevistados), achados que consolidaram a revisão proposta pelos especialistas e profissionais na transição do estudo 4 para o 5, bem como para os três fatores mediadores (nível de atenção à saúde, de complexidade da assistência e de escolaridade do usuário), perspectivas da Percepção e para a Expectativa, com exceção do nível de escolaridade para a perspectiva de Expectativa de usuários na qual houve significância em apenas três subdimensões e uma macrodimensão pelo que se sugere aprofundar as análises quanto a Expectativa de usuários.

Foi possível observar, ainda, que houve correlação e positiva na ampla maioria (na *proxy* L - 10 e na *proxy* M+N - 12) das 12 subdimensões de questões validadas analisadas (Acessibilidade aos Serviços e Satisfação Global não foram mensuradas quanto a Expectativa que anulou a macrodimensão Acessibilidade, sem análises de correlações) e em todas as cinco

macrodimensões de qualidade (Tangibilidade, Confiabilidade, Presteza, Segurança e Empatia). Dessa forma, houve uma tendência que quanto maior foram os valores da Expectativa maior tenderam a ser os valores da Percepção obtidas nas subdimensões de questões validadas e macrodimensões de qualidade, em contraposição à associação negativa citada pela literatura de aplicação da escala SERVQUAL (E. Babakus & W. G. Mangold, 1992; Parasuraman et al., 1988, 1991a; Parasuraman et al., 1994), mas alinhada com outros estudos empíricos que suportam relação positiva entre *expectativas* e *satisfação* (Oliver & Linda, 1981; Swan & Trawick, 1981; Bearden e Teel, 1983), pelo que se sugere aprofundar as análises, possivelmente, quanto a Expectativa de usuários.

Explorou-se, ainda, uma análise descritiva da variável Qualidade Global Percebida, inclusive quanto as possíveis influências das variáveis nível de atenção à saúde, nível de complexidade da assistência e nível de escolaridade de usuários. Assim, a Tabela 109 (apêndice 9, tópico 9.2.2.2), mostra que a Qualidade Global Percebida média foi de 3,35 e a mediana de 3,00 e que 75% das respostas atribuídas resultaram em valoração superior a 3 (cerca de 4, muito importante/relevante). Pode-se observar uma variabilidade na valoração média da variável latente Qualidade Global Percebida. Houve diferença significativa (valor- $p \leq 0,050$) para considerar as variáveis nível de atenção à saúde, de complexidade da assistência e de escolaridade de usuários enquanto fator mediador da Qualidade Global Percebida, para a Percepção, na *proxy* M+N ($n = 2.547$). Destaca-se nos fatores mediadores que nível Primário de atenção apresentou a menor valoração média (2,99), enquanto o nível Secundário obteve a maior valoração média (3,63), enquanto o nível de Média complexidade da assistência apresentou a maior valoração média (3,47), e por sua vez, o nível Médio de escolaridade obteve a menor valoração média (3,31) e o nível Superior a maior valoração média (3,51).

Na análise inferencial, nos resultados da fase 4, no modelo final, após a análise e exclusão de uma questão validada para a perspectiva da Expectativa, as demais 17 questões (a última questão não foi mensurada para Expectativa por dificuldades na coleta de dados) e todas as 19 questões para a Percepção, além de todas as 13 subdimensões de questões validadas para a Percepção e 12 subdimensões para a Expectativa (que teve exclusão de uma subdimensão), e, ainda, todas as seis macrodimensões de qualidade percebida analisadas (Tangibilidade, Confiabilidade, Presteza, Segurança, Empatia e Acessibilidade – esta não foi mensurada para a Expectativa) apresentaram carga fatorial superior a 0,50 (ou suas permanências não impediram a validação convergente da AFC), na *proxy* M+N. A confiabilidade e a validade das 13 subdimensões de questões validadas, das seis macrodimensões de qualidade e da variável

latente Qualidade Global Percebida foram asseguradas, uma vez que todas apresentaram validação convergente ($AVE > 0,40$), confiabilidade adequada ($A.C. > 0,60$ ou $C.C. > 0,60$), unidimensionalidade e validação discriminante ($VCM < AVE$). Foi observado que os parâmetros e índices de ajustamento e qualidade dos modelos de equações estruturais indicaram bom ajuste (nas perspectivas de Expectativa e de Percepção), já que a razão entre a estatística qui-quadrado e os graus de liberdade foi menor que 3, as estatísticas TLI e o CFI foram maiores que 0,80 e o RMSEA do foi menor que o limite máximo de 0,10. Para concluir a análise inferencial, as figuras 68 e 69 sintetizaram os resultados das Tabelas 105 e 106 (Expectativa) e as figuras 70 e 71 os resultados das Tabelas 107 e 108 (Percepção) (apêndice 9, tópico 9.2.2.2), nos quais foi possível observar que as questões validadas apresentaram maiores correlações com suas respectivas subdimensões de questões validadas que por sua vez apresentaram maiores correlações com suas respectivas macrodimensões de qualidade percebida, na *proxy* M+N ($n = 2.547$).

Diante destes resultados, as figuras 10 e 11 (*proxy* M+N) sintetizaram os ajustes das Análises Fatoriais Confirmatórias (AFC) e as modelagens para as variáveis latentes Qualidade Global Percebida (Expectativa e Percepção). Dessa forma, a ampla maioria das hipóteses iniciais (quadro 16, tópico 6.3.2.) foram validadas e confirmaram a validação dos itens de mensuração de qualidade testados (20 questões validadas, 13 subdimensões de questões validadas e seis macrodimensões de qualidade, na *proxy* M+N).

Dessa forma, concluiu-se a tradução transcultural, a análise da adequação e a validação de itens da escala internacional SERVQUAL de qualidade e satisfação em busca de mapear os constructos de qualidade (ideias-força) relevantes, na visão de agentes chaves para avaliação de qualidade/satisfação em serviços de saúde, posteriormente, consolidados por adaptações em etapas e fases incrementais de validação de questões validadas, além de subdimensões e macrodimensões de qualidade e satisfação para usuários do SUS, e, ainda, dos reflexos sobre uma *variável latente* Qualidade Global Percebida, medidos por suas magnitudes (cargas fatoriais), significâncias estatísticas e suas correlações em modelos estruturais validados (figuras 9, 10 e 11). Destacam-se os aspectos inéditos na literatura de tradução e validação de questões, subdimensões e macrodimensões de qualidade e satisfação, pela colaboração com decisores chaves na valoração da relevância/importância e estruturação de uma escala inovadora de aferição de qualidade e satisfação no SUS (Quality Saúde), além da validação multicêntrica de modelos finais através de análise de cargas fatoriais, validade e confiabilidade dos itens, e, ainda, dos testes de possíveis fatores mediadores na valoração, tendo validado as três variáveis mediadoras estudadas.

Ressalte-se que a comprovação da tendência das variáveis nível de atenção à saúde, nível de complexidade da assistência, cargo/função – para especialistas e profissionais além do nível de escolaridade – apenas para usuários, funcionarem enquanto fatores mediadores da ampla maioria das 13 subdimensões de questões propostas e das cinco macrodimensões de qualidade originais da escala SERVQUAL, além da macrodimensão Acessibilidade adicionada, trata-se de um achado inovador desta investigação. Outro achado inovador foi que quase todas as subcategorias de análise (20 questões validadas), as categoriais (13 subdimensões de questões validadas) e as macrocategorias (seis macrodimensões de qualidade) analisadas foram validadas, após a revisão proposta pelos especialistas e profissionais na transição do estudo 4 para o 5. Deve-se registrar que a validação destas mesmas cinco macrodimensões da escala SERVQUAL para a saúde, ocorreu nos estudos de Babakus e Mangold (1992).

Pode-se concluir pela adequação do instrumento final, uma escala de medição de qualidade/satisfação percebida (questionário), com suas respectivas subescalas – 13 subdimensões e seis macrodimensões, enquanto um instrumento de gestão da qualidade/satisfação de unidades do SUS (*Quality Saude*), na visão de usuários, Artefato inovador no contexto de saúde pública e privada no Brasil.

Diante dos achados e contributos ressaltados desta investigação, em colaboração com especialistas, gestores e usuários, consolidou-se os painéis de indicadores e dimensões de performance, além de uma nova escala de aferição da qualidade/satisfação percebida no SUS (*Quality Saude*), derivada da SERVQUAL, através de provas de conceito, de valor e de utilidade próprias do método de *Design Research*.

Por fim, enquanto contribuição deste estudo, operacionalizou a validação multicêntrica, tendo construído modelos finais através de análise de cargas fatoriais, validade e confiabilidade dos itens, de *proxies* de painéis de indicadores com 453 decisores chaves e uma adaptação transcultural da escala SERVQUAL que resultou em uma nova escala de avaliação da qualidade/satisfação percebida, com subescalas de intervenção ou gestão da qualidade, tendo sido validada por 195 especialistas e gestores e, aplicada a 2.547 usuários de 74 unidades do SUS.

REFERÊNCIAS BIBLIOGRÁFICAS

- Acurcio, F. A., Cherchiglia, M. L., & Santos, M. A. (1991). Avaliação de qualidade de serviços de saúde. *Saúde em debate*, 33, 4.
- Aday, L. A. (1993). *Evaluating the Medical Care System: Effectiveness, Efficiency, and Equity* (H. A. Press Ed.): University of Michigan.
- Aharony, L., & Strasser, S. (1993). Patient satisfaction: what we know about and what we still need to explore. *Medical Care Review*, 50(1), 31.
- AHCPR. (1999). Quality measurement initiatives in the US Federal Government. In: Presentation at Commonwealth Fund.
- AIHW. (2008). Australia's Health. In.
- AIHW, A. I. O. H. A. W.-. (2000). *The seventh biennial health report of the Australian Institute of Health and Welfare*. Canberra: AIHW
- Akerman, M., & Nadanovsky, P. (1992). Evaluation of Health Services - What to Evaluate? *Caderno de Saúde pública*, 8(4), 5.
- Albuquerque, J. D. (2012). *Qualidade hospitalar: Uma análise da sua realidade em Campina Grande - Paraíba - Brasil*. (Doctoral). Universidade de Extremadura, Espanha: Badajoz.
- Alkin, M. C., & Christie, C. A. (2004). Evaluation Roots: Tracing Theorist's Views and Influences. *Sage publications*.
- Almeida, C., Braveman, P., Gold, M. R., Szwarcwald, C. L., Ribeiro, J. M., Mi, A., . . . Viacava, F. (2001). Methodological concerns and recommendations on policy consequences of the World Health Report 2000. *Lancet*, 357(9269), 5286.
- Almeida, C. M. (1999). Reforma do Estado e reforma de sistemas de saúde: experiências internacionais de tendências de mudança. *Ciência & Saúde Coletiva*, 4(2), 24.
- Alves, A. S. D. (2012). *Responsabilidade e Governança na Moderna Gestão Hospitalar*. (Doctoral). Universidade de Aveiro, Portugal: Aveiro.
- Amado, C. A. E. F., & Dyson, R. G. (2008). On comparing the performance of primary care providers. *European Journal of Operational Research*, 185, 18.
- Amado, C. A. E. F., & Santos, S. P. (2009). Challenges for performance assessment and improvement in primary health care: The case of the Portuguese health centres. *Health Policy*, 91, 14.
- APM/CREMESP. (1998). *Controle de Qualidade do Atendimento Médico-Hospitalar do Estado de São Paulo - Manual de Orientação aos Hospitais Participantes* (E. Atheneu Ed.). São Paulo.
- Arah, O. A., Custers, T., & Klazinga, N. S. (2003). *Updating the Key Dimensions of Hospital Performance: The Move Towards a Theoretical Framework* Paper presented at the 3rd Workshop on Hospital Performance Measurement Barcelona.
- Arah, O. A., Klazinga, N. S., Delnoij, D. M. J., Asbroek, A. H. A. T., & Custers, T. (2003). Conceptual frameworks for health systems performance: a quest for effectiveness, quality, and improvement *International Journal for Quality in Health Care*, 15(5), 22. doi:10.1093/intqhc/mzg049
- Arah, O. A., Westert, G. P., Hurst, J., Niek, S., & Klazinga, N. S. (2006). A Conceptual framework for the OECD Health Care Quality Indicators Project *International Journal for Quality in Health Care*, 9. doi:10.1093/intqhc/mzl024
- Arbuckle, J. L. (2008). *Amos 17 users' guide* (SPSS Ed.). Chicago, IL.
- Arcelay, A., Sanchez, E., & Hernandez, K. (1999). Self-assessment of all the health centres of a public health service through the European Model of total quality management *Int J Health Care Qual Assur Inc Leadersh Health Serv* 12, 5.
- Associação, & Medicina, P. d. (1998). *Controle da Qualidade do Atendimento Hospitalar no Estado de São Paulo* (Atheneu Ed.). São

Paulo: Atheneu.

- Atkinson, L. (1988). The measurement-statistics controversy: Factor analysis and subinterval data *Bulletin of the Psychonomic Society*, 26(4), 4.
- Austen, A. (2012). Stakeholders management in public hospitals in the context of resources *Management*, 16(2), 14. doi:10.2478/v10286-012-0067-8
- Ayres, J. R. d. C. M. (2002). *Epidemiologia, promoção da saúde e o paradoxo do risco* Paper presented at the V Congresso Brasileiro de Epidemiologia, Curitiba, PR, Brasil.
- Babakus, E., & Mangold, G. (1992). Adapting SERVQUAL Scale Hospital Service: Na empirical Investigation. *Health Services Research*, 22.
- Babakus, E., & Mangold, W. G. (1992). Adapting the SERVQUAL scale to Hospital Services: An empirical investigation. *Health Services Research*, 26(6), 767-786.
- Banker, R. D., Charnes, A., & Cooper, W. (1984). Models for estimating technical and scale efficiencies in data envelopment analysis. *Management Science*, 30(9), 15.
- Barbosa, L., & Neto, A. P. (2017). Ludwik Fleck (1896-1961) e translação do conhecimento: considerações sobre a genealogia de um conceito. *Saúde Debate*, 41(Especial), 13.
- Barreto, E. (2016). Rio tem a maior taxa de mortalidade no SUS em três décadas. *O Globo - Caderno Brasil*. Retrieved from <http://oglobo.globo.com/brasil/rio-tem-maior-taxa-de-mortalidade-no-sus-em-tres-decadas-18770526#ixzz4OrM31ntY> [29 de fevereiro de 2016]
- Barzelay, M. (1997). Central audit institutions and performance auditing: A comparative analysis of organizational strategies in the OECD. *Governance*, 10, 26.
- Batista, W. J., & Paiva, A. D. (2014). *A utilização do instrumento padrão de pesquisa de satisfação (IPPS) no diagnóstico de uma organização militar de saúde do Exército Brasileiro*. Paper presented at the XI Simpósio de Excelência em Gestão e Tecnologia-SEGTe Resende - RJ.
- Bearden, W. O., & Teel, J. E. (1983). Selected determinants of consumer satisfaction and complaint reports. *JMR, Journal of Marketing Research*, 20(1), 1.
- Becker, H. (1993). *Métodos de pesquisa em ciências sociais* (Hucitec Ed.). São Paulo: Hucitec.
- Bentler, P. M. (1990). Comparative Fit Indexes in Structural Models. *Psychological Bulletin*, 107(2), 238-246.
- Bentler, P. M., & Bonett, P. G. (1980). Significance Tests and Goodness-of-Fit in the Analysis of Covariance Structures. *Psychological Bulletin*, 88(3), 588-606.
- Bentler, P. M., & Chou, C.-P. (1987). Practical Issues in Structural Equation Modeling. *Sociological Methods & Research*, 16(1), 40. doi:10.1177/0049124187016001004
- Berry, L. L., Parasuraman, A., Zeithaml, V., & Adsit, D. (1994). Improving service quality in America: Lessons learned; Executive commentary. *The Academy of Management Executive*, 8(2), 21.
- Bitner, M. J. (1992). Evaluating service encounters: the effects of physical surroundings and employees. *Journal of Marketing*, 56(2).
- Bitner, M. J., Booms, B. H., & Mohr, L. A. (1994). Critical service encounters: The employee's viewpoint. *Journal of Marketing*, 58(4), 1.
- Bittencourt, A. S., Camacho, L. A. B., & Leal, M. C. O. (2006). O sistema de informação hospitalar e sua aplicação na saúde coletiva. *Cadernos de Saúde Pública*, 22(1), 19-30.
- Boccolini, C. S., Boccolini, P. d. M. M., Damacena, G. N., Ferreira, A. P. d. S., & Szwarcwald, C. L. (2016). Fatores associados à discriminação percebida nos serviços de saúde do Brasil: resultados da Pesquisa Nacional de Saúde, 2013 *Ciência & Saúde Coletiva*, 21(2), 8.

- Boggan, R., & Biklein, S. (1994). Características da investigação qualitativa. In P. Editora (Ed.), *Investigação qualitativa em educação: uma introdução à teoria e aos métodos* (pp. 47-51). Porto: Porto Editora.
- Boland, T., & Fowler, A. (2000). A systems perspective of performance management in public sector organisations. *International Journal of Public Sector Management*, 13(5), 30.
- Bollen, K. A., & Long, J. S. (1993). *Testing Structural Equation Models*. In S. F. Editions (Ed.), (Vol. 154, pp. 320).
- Bonilla, M., & Rubio, V. (2000). Sistemas de ajuste por riesgo. *Revista Medica Uruguay*, 16(2), 10.
- Borges, J. B. C., Carvalho, S. M. R., & Silva, M. A. M. (2006). Qualidade do serviço prestado aos pacientes de cirurgia cardíaca do Sistema Único de Saúde – SUS. *Revista Brasileira de Cirurgia Cardiovascular*, 25(2).
- Bosi, M. L., & Mercado, F. J. (2006). *Pesquisa qualitativa de serviços de saúde* (Vozes Ed.).
- Brandt, E., Schmidt, W., Dziewas, R., & Groene, O. (2005). Implementing the Health Promoting Hospitals Strategy through a combined application of the EFQM Excellence Model and the Balanced Scorecard In W. R. O. f. Europe (Ed.), *Health Promotion in Hospitals: Evidence and Quality Management* (pp. 80-89). Copenhagen: WHO Regional Office for Europe.
- Brasil. (1986a). *Relatório Final*. Retrieved from Brasília: http://conselho.saude.gov.br/biblioteca/relatorios/relatorio_8.pdf
- Brasil. (1986b). *VIII Conferência Nacional de Saúde - Relatório Final*. Retrieved from Brasília: http://conselho.saude.gov.br/biblioteca/relatorios/relatorio_8.pdf
- Brasil. (1988). *Constituição da República Federativa do Brasil de 1988*. Distrito Federal: Brasília: Congresso Nacional
- Brasil. (1990a). *Lei 8.080, de 19 de setembro de 1990*. Distrito Federal: Brasília: Congresso Nacional
- Lei nº 8.080, de 19 de setembro de 1990, Lei nº 8.080/1990 C.F.R. (1990b).
- Lei nº 8.142, de 12 de novembro de 1990, Lei nº 8.142/1990 C.F.R. (1990c).
- Brasil. (1994). *Uso e disseminação de informações em saúde: subsídios para elaboração de uma política de informações em saúde para o SUS*. Brasília: Ministério da Saúde
- Brasil. (1996a). *Portaria nº 2.390/GM de 11 de Dezembro 1996*.
- Portaria nº 2.390, de 11 de dezembro 1996, (1996b).
- Brasil. (1997a). *Portaria nº 820/GM, de 25 de junho de 1997*.
- Portaria nº 820, de 25 de junho de 1997, (1997b).
- Brasil. (1998). *Lei 9.656, de 03 de junho de 1998*. Distrito Federal: Brasília: Congresso Nacional
- Brasil. (2000a). *Portaria nº 443/GM, de 26 de Abril de 2000*.
- Portaria nº 443, de 26 de abril de 2000, (2000b).
- Brasil. (2002a). *Indicadores básicos para a saúde no Brasil: conceitos e aplicações*. Brasília: Ministério da Saúde
- Brasil. (2002b). *Portaria nº 312 de 02 de Maio de 2002*. Brasília: Ministério da Saúde
- Portaria nº 312, de 02 de maio de 2002, (2002c).
- Brasil. (2003a). *A Saúde na Opinião dos Brasileiros*. Retrieved from Brasília: http://bvsmis.saude.gov.br/bvs/publicacoes/opiniao_brasileiros1.pdf
- Brasil. (2003b). *Pacto de Indicadores da Atenção Básica: instrumento de negociação qualificador do processo de gestão do SUS*. Recife: Ministério da Saúde Retrieved from <http://www.scielo.br/pdf/rbsmi/v3n2/a13v03n2.pdf>
- Brasil. (2004). *Programa Nacional de Avaliação de Serviços de Saúde-PNASS*. Brasília: Ministério da Saúde

- Brasil. (2005). *Decreto nº 5.378 de 23 de fevereiro de 2005*. Retrieved from http://www.planalto.gov.br/ccivil_03/_ato2004-2006/2005/Decreto/D5378.htm
- Portaria nº 3.085, de 1 de dezembro de 2006, § 1 (2006a).
- Portaria nº 3.332, de 28 de dezembro de 2006, § 1 (2006b).
- Brasil. (2006c). *Sistema de Planejamento do SUS: uma construção coletiva*. Brasília: Ministério da Saúde
- Portaria nº 376, de 16 de fevereiro de 2007, § 1 (2007).
- Brasil. (2008a). *Indicadores básicos para a saúde no Brasil: conceitos e aplicações*. Retrieved from Brasília:
- Portaria nº 1.885, de 9 de setembro de 2008, § 1 (2008b).
- Brasil. (2009a). *A experiência brasileira em sistemas de informação em saúde*. Brasília: Ministério da Saúde
- Decreto nº 6.932, de 11 de dezembro de 2009, 6932 C.F.R. (2009b).
- Brasil. (2010a). *Portaria nº 4.279, de 30 de dezembro de 2010*.
- Brasil. (2010b). *Programa de Qualificação da Saúde Suplementar: Componente Qualificação das Operadoras*. Retrieved from Rio de Janeiro: http://www.ans.gov.br/images/stories/A_ANS/Transparencia_Institucional/Indicadores_de_qualidade/relatorio_gestao_ultimo_ano_base.pdf
- Brasil. (2010c). *Programa Nacional de Gestão Pública e Desburocratização – GESPÚBLICA. Manual de Avaliação de Satisfação do Usuário do Serviço Público*. Retrieved from Brasília: http://www.escoladegoverno.pr.gov.br/arquivos/File/Ministerio_do_Planejamento/ipp_s.pdf
- Decreto nº 7.508, de 28 de junho de 2011, (2011a).
- Brasil. (2011b). *IDSUS: Índice de Desempenho do Sistema de Saúde*. Brasília: Ministério da Saúde
- Brasil. (2011c). *Legislação Estruturante do SUS*. Brasília: Ministério da Saúde
- Portaria nº 1.171, de 19 de maio de 2011, (2011d).
- Brasil. (2011e). *Programa Nacional de Melhoria do Acesso e da Qualidade da Atenção Básica. Portaria 1.654, de 19 de julho de 2011*. Brasília, DF: Ministério da Saúde
- Brasil. (2011f). *Sistema Único de Saúde - SUS*. Brasília: Ministério da Saúde
- Brasil. (2012a). *Programa Nacional de Melhoria do Acesso e da Qualidade da Atenção Básica (PMAQ): manual instrutivo*. Brasília: Ministério da Saúde
- Brasil. (2012b). *Índice de Desempenho do Sistema Único de Saúde*. Ministério da Saúde Retrieved from <http://idsus.saude.gov.br>
- Brasil. (2013a). *A produção de conhecimento em Economia da Saúde: uma perspectiva bibliográfica (2004-2012)*. Retrieved from Brasília: www.saude.gov.br/economiasaude
- Brasil. (2013b). *Manual instrutivo do PMAQ-AB para as equipes de Atenção Básica*. Ministério da Saúde
- Portaria nº 2.135, de 25 de setembro de 2013, (2013c).
- Brasil. (2014). *Governos Estaduais no Federalismo Brasileiro*. Brasília: Instituto de Pesquisa Econômica Aplicada
- Brasil. (2015a). *A Gestão do SUS. Sistema de Saúde*. Brasília: Ministério da Saúde
- Brasil. (2015b). *A Gestão do SUS. Sistema de Saúde*. Brasília: Ministério da Saúde
- Brasil. (2015c). *Alternativas de gerência de unidades públicas de saúde*. Brasília: Ministério da Saúde
- Brasil. (2015d). *Alternativas de gerência de unidades públicas de saúde*. Brasília: Ministério da Saúde
- Portaria nº 28, de 08 de janeiro de 2015, (2015e).

- Brasil. (2015f). *Sistemas de Informação da Atenção à Saúde: Contextos Históricos, Avanços e Perspectivas no SUS*. Brasília: Ministério da Saúde
- Brasil. (2016a). *Mapa Assistencial da Saúde suplementar 2016*. Rio de Janeiro
- Brasil. (2016b). *Nota N° 791/2016/GMOA/GGRAS/DIPRO/ANS*. Retrieved from Rio de Janeiro:
- Brasil. (2016c). *Relatório do Rol de Indicadores IDSS do QUALISUS*. Retrieved from http://www.ans.gov.br/images/stories/prestadores/qualiss/Fichas_Indicadores_QUALISS.pdf (Lista instituída na RN 405-16)
- Brasil. (2016d). *Resolução normativa RN n° 416, de 22 de dezembro de 2016*. Retrieved from <http://www.ans.gov.br/component/legislacao/?view=legislacao&task=TextoLei&format=raw&id=MzM00A>
- Resolução normativa RN n° 405, de 9 de Maio de 2016, (2016e).
- Brasil. (2017a). *Autoavaliação para melhoria do acesso e da qualidade da atenção básica: AMAQ*. Brasília: Ministério da Saúde, Secretaria de Atenção à Saúde, Departamento de Atenção Básica Retrieved from http://189.28.128.100/dab/docs/portaldab/publicacoes/amaq_2017.pdf
- Decreto n° 9.094, 17 de julho de 2017, (2017b).
- Brasil. (2017c). *Lei n° 13.460 de 26 de Junho de 2017*. Distrito Federal: Brasília: Congresso Nacional Retrieved from http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2017/Lei/L13460.htm
- Brasil. (2017d). *Manual instrutivo para as equipes de atenção básica e NASF - Programa Nacional de Melhoria do Acesso e da Qualidade da Atenção Básica (PMAQ) - Terceiro ciclo - (2015-2017)*. Brasília: Ministério da Saúde, Secretaria de Atenção à Saúde, Departamento de Atenção Básica Retrieved from http://189.28.128.100/dab/docs/portaldab/documentos/Manual_Instrutivo_3_Ciclo_PMAQ.pdf
- Resolução n° 23, de 17 de agosto de 2017, (2017e).
- Resolução n° 37, de 22 de março de 2018, (2018).
- Braveman, P., Starfield, B., & Geiger, J. H. (2001). World Health Report 2000: how it removes equity from the agenda for public health monitoring and policy. *British Medical Journal*, 323, 4.
- Brook, R. H., & Lohr, K. (1985a). Efficacy, effectiveness, variations and quality: boundary-crossing research. *Med Care*, 23(5), 23.
- Brook, R. H., & Lohr, K. (1985b). Efficacy, Effectiveness, Variations and Quality: Boundary-Crossing Research. *Med Care*, 23(5), 23.
- Brook, R. H., & McGlynn, E. A. (1991). Maintaining quality of care. In H. U. Press (Ed.), *Health Services Research Key to Health Policy*. Cambridge: Haward University Press.
- Brunet, L., Brassard, A., & Corriveau, L. (1991). *Administration Scolaire et efficacité dans le organizations* (É. A. DÁRC Ed.). Montréal: Éditions Agence DÁRC.
- Bryson, J. (2004). What to do when stakeholders matter: Stakeholder identification and analysis techniques. *Public Management Review*, 6(1), 33.
- Bunge, M. (1974). Os conceitos do modelo: modelo na ciência teórica. In *Teoria e realidade* (Perspectiva ed., pp. 11-40). São Paulo: Perspectiva.
- Buono, A. (2005a). *Corporate social performance: a stakeholder approach*. *Proquest Psychology Journals*, 58(3), 5.
- Buono, A. (2005b). *Corporate social performance: a stakeholder approach* Corporate social performance: a stakeholder approach. *Proquest Psychology Journals*, 58(3), 5.
- Caldeira, P., & Faia-Correia, M. (2002). People, organizations and technology: Socio-technical systems and stakeholders theories. In *Proceedings of the Sixth International*

- Conference on the Social and Ethical Impacts of Information and Communication Technologies* (ETHICOMP ed., pp. 757-768): ETHICOMP.
- Cameron, K. (1980). Critical questions in assessing organizational effectiveness. *Organization Dynamics*, 6, 15.
- Cameron, K. (1981). Domains of organizational effectiveness in colleges and universities. *Management Journal*, 24, 23.
- Cameron, K. S., & Whetten, D. A. (1983). *Organizational Effectiveness: A Comparison of Multiple Models* (A. Press Ed.). Orlando: Academic Press.
- Campbell, S. M., Roland, M. O., & Buetow, S. A. (2000). Defining quality of care. *Social Science & Medicine*, 51, 15.
- Campos, F. E. (1988). Resolutividade – Uma Aproximação à Avaliação Qualitativa dos Serviços de Saúde *Imprensa Universitária*.
- Campos, R. T. O., & Furtado, J. P. (2011). *Desafios da Avaliação de Programas e Serviços em saúde: Novas Tendências e Questões Emergentes* (Editora Unicamp ed.). Campinas, São Paulo: Editora Unicamp.
- Carvalho, C. (2007). *Organizações, actores envolvidos e partes interessadas: Determinantes da saliência dos stakeholders e sua relação com o desempenho organizacional*. (Doctoral). Universidade de Coimbra, Coimbra, Portugal.
- Carvalho, C., & Gomes, A. (2000). Eficácia organizacional: Determinantes e dimensões. *Psicologica*, 25, 24.
- Carvalho, C., & Gomes, A. (2002). Eficácia organizacional: Construção de um instrumento de medida e questões em torno da sua avaliação. *Revista Psicologia e Educação*, 1(1/2), 23.
- Cassiani, S. H. B., & Rodrigues, L. P. (1996). A Técnica de Delphi e a Técnica de Grupo Nominal como estratégias de coleta de dados das pesquisas em enfermagem. *Acta Paulista de Enfermagem*, 9(3), 8.
- Cavallo, F., Traversa, G., Piergenteli, P., & Reuga, G. (1988). *Qualità ed Efficacia dei Servizi Sanitari* (F. Angeli/Cresa Ed.). Milano.
- CFM. (2014). Pesquisa revela que 92% da população está insatisfeita com a saúde no Brasil. Retrieved from http://portal.cfm.org.br/index.php?option=com_content&view=article&id=24984:data-folha-revela-93-da-populacao-esta-insatisfeita-com-a-saude-no-brasil&catid=3
- CFM. (2015). Para 93% da população, a saúde Brasil é considerada péssima, ruim ou regular. Retrieved from http://portal.cfm.org.br/index.php?option=com_content&view=article&id=25807:2015-10-13-14-55-36&catid=3
- CFM. (2018). *Opinião dos brasileiros sobre o atendimento público na área de saúde (2018)*. Retrieved from http://portal.cfm.org.br/images/PDF/datafolha_sus_cfm2018.pdf
- Chahal, H., & Kumari, N. (2011). Consumer Perceived Value and Consumer Loyalty in the Healthcare Sector. *Journal of Relationship Marketing*, 10(2), 15. doi:10.1080/15332667.2011.577729
- Chahal, H., & Kumari, N. (2012). Consumer perceived value -The development of a multiple item scale in hospitals in the Indian context. *International Journal of Pharmaceutical and Healthcare Marketing*, 6(2), 24.
- Champagne, F., Contandriopoulos, A. P., Broussele, A., Hartz, Z. M. A., & Denis, J. (2009). L.L'Evaluation dans le domaine de la santé: concepts et méthodes. In L. P. d. I. U. d. Montréal (Ed.), *L'Evaluation: concepts et méthodes* (pp. 35-36). Montréal: Les Presses de l'Université de Montréal.

- Chang, L. L., Stephen, W., & Northcott, D. N. (2002). The NHS Performance Assessment Framework: A “Balanced Scorecard” Approach? *Journal of Management in Medicine*, 16(5), 14.
- Chavez, M. R. (1981). *Los sistemas de evaluación en el sector salud. Limites para su desarrollo* (E. V Ed. Vol. 23).
- Chen, H. T. (1990). *Theory-Driven evaluations* (S. Publications Ed.). Newbury Park: Sage Publications.
- Chianca, T., & Youker, B. (2004). *La evaluación em América Latina y El Caribe. Visión general de los desarrollos recientes*.
- Churchill, G., & Surprenant, C. (1982). An investigation into the determinants of costumer satisfaction. *Journal of Marketing Research*, 19(4), 14.
- CIHI. (1999). *Roadmap Initiative . . . Launching the Process*. Ottawa.
- CIHI. (2000). *Canadian Health Information Roadmap Initiative Indicators Framework*. Ottawa.
- CIHI. (2001a). A portrait of Canada’s health care system In *Health Care in Canada*. Ottawa: CIHI.
- CIHI. (2001b). Health Indicators. Volume 2001 No. 3: December 2001. In (pp. 82-221). Ottawa: Statistics Canada and Canadian Institute for Health Information
- CIHI, C. I. f. H. I.-. (2001). *Health Care in Canadian*. Retrieved from Otawwa, Ontario: www.cihi.ca
- CIHR, C. I. o. H. R. (2004). *Knowledge Translation Strategy 2004-2009: Innovation in Action*. Retrieved from Ottawa: <http://www.cihr-irsc.gc.ca/e/26574.html>
- Clavier, C., Sénéchal, Y., & Potvin, L. (2011). A theory - based model of translation practices in public health participatory research. *Sociology of Health and Illness*, 34(5), 16.
- CNI, C. N. d. I.-. (2012). *Retratados da sociedade brasileira: saúde publica (2011)*. Retrieved from Brasília: <http://admin.cni.org.br/portal/data/files/00/FF80808134CE96AF0134D22242451D75/Retratos%20da%20Sociedade%20Brasileira%20Sa%C3%BAde%20P%C3%BAblica%20Janeiro%202012.pdf>
- CNI, C. N. d. I.-. (2018). *Retratados da sociedade brasileira: saúde publica (2018)*. Retrieved from Brasília: https://bucket-gw-cni-static-cms-si.s3.amazonaws.com/media/filer_public/18/7f/187f1473-2603-4b06-a8a1-070486293e98/retratosdasociedadebrasileira_44_saude.pdf
- Coelli, T., Rao, D. S. P., & Battese, G. E. (1999). *An Introduction to efficiency and productivity analysis* (K. A. Publishers Ed.): Kluwer Academic Publishers.
- Commission, A. (2000). *Aiming to Improve: The Principles of Performance Measurement*. London: Audit Commission
- Commission, J. (1998). Accreditation Program for Healthcare Services. In. Oakbrook Terrace: Joint commission.
- Committee, N. H. P. (2001). National Health Performance Framework Report. In. Brisbane: Queensland Health.
- Conlon, D., Dyne, L., & Milner, M. (2004). The effects of physical and social context on evaluations of captive, intensive service relation ships. *Academy of Management Journal*, 47(3), 23.
- Contandriopoulos, A. P. (2006). Avaliando a institucionalização da avaliação. *Ciência Saúde Coletiva*, 10(3), 7.
- Contandriopoulos, A. P., Champagne, F., Denis, J. L., & Pineault, R. (1997). A avaliação na área de saúde: Conceitos e métodos. In E. Fiocruz (Ed.), *Avaliação em Saúde: Dos*

- Modelos Conceituais à Prática na Análise da Implantação de programas* (pp. 29-47). Rio de Janeiro: Editora Fiocruz.
- Cooper, W. W., Seiford, L. M., & Tone, K. (2002). Data envelopment analysis – a comprehensive text with models, applications, references and DERA – solver software. In (Kluwer Academic Publishers ed.). Netherlands: Kluwer Academic Publishers.
- Costa, C., & Lopes, S. (2011). *Relatório de Avaliação do Desempenho dos Hospitais Públicos em Portugal Continental - 2010*. Lisboa: Escola Nacional de Saúde Pública
- Couto, R. C. (2012). *Avaliação do nível de conformidade com requisitos técnicos e legais de segurança assistencial e seus fatores associados em hospitais da rede do Estado de Minas Gerais - 2012*. (Doctoral). Universidade Federal de Minas Gerais, Minas Gerais.
- Couto, R. C. (2014). *Avaliação do nível de conformidade com requisitos técnicos e legais de segurança assistencial e seus fatores associados em hospitais da rede do Estado de Minas Gerais – 2012*. (Doctoral). Universidade Federal de Minas Gerais, Belo Horizonte.
- Couto, R. C., Serufo, J. C., Couto, R. C., Pedrosa, T. M. G., Andrade, L. O. F. d., & Serufo Filho, J. C. (2012). Requisitos técnicos e legais relacionados à gestão de risco assistencial na rede hospitalar do estado de minas gerais, brasil. In F. d. Medicina (Ed.), *Avaliação do nível de conformidade com requisitos técnicos e legais de segurança assistencial e seus fatores associados em hospitais da rede do estado de minas gerais – 2012*. Belo Horizonte: Universidade Federal de Minas Gerais.
- Cozzens, S. (1995). Performance Assessment and the National Science Foundation: Proposals for NSF's Response to the Government Performance and Results Act. *NSF Discussion Paper*.
- CQH. (2009). *Compromisso com a Qualidade Hospitalar - 3º Caderno de Indicadores CQH - 2009*. Retrieved from São Paulo: www.cqh.org.br/?q=node/322
- Cronin, J. R. J., & Taylor, A. S. (1992). Measuring service quality: a reexamination and a extension. *American Marketing Association*, 56.
- Cruz, W. B. S., & Melleiro, M. M. (2007). Análise da satisfação dos usuários de um hospital privado. *Revista Escola de Enfermagem da USP*.
- Da Rocha, C. M. V., Martinello, A. G., & Fagundes, A. G. B. (2012). Capacitação para o planejamento em saúde: uma contribuição à implementação do planejassus. *Revista Baiana de Saúde Pública*, 36(2), 560-576.
- Da Silva, F. J. C. P. (2014). *Avaliação da qualidade do serviço de saúde: a visão do cliente*. (Doctoral). Universidade de São Paulo, Ribeirão Preto.
- Da Silva, S. R. A., Gonçalves, M., Siqueira, P. C., & Da Silveira, C. A. C. (2008). As decisões de investimento na Fundação Hospitalar de Minas Gerais e seus reflexos nos indicadores de qualidade. *Revista de Administração Hospitalar e Inovação em Saúde*.
- Daibert, P. B. (2015). *Impacto econômico e assistencial das complicações relacionadas à internação hospitalar*. (Master). Universidade Federal de Minas Gerais, Minas Gerais.
- Dall'Agnol, C. M., & Trench, M. H. (1999). Grupos focais como estratégia metodológica em pesquisa na enfermagem. *Rev. Gaúcha Enf*, 20(1), 5-25.
- De Geyndt, W. (1970). Five approaches for assessing the quality of care *Hospital Administration*, 15, 22.
- De Moraes, R. M. (2014). *Um modelo para avaliação de sistemas de informação do SUS de abrangência nacional*. (Doctoral). Universidade de São Paulo, Ribeirão Preto.
- Denzin, N. K., & Lincoln, Y. S. (2006). *O planejamento da pesquisa qualitativa: teoria e abordagens* (Bookman Ed.). Porto Alegre: Bookman.

- Deslandes, S. F., & Gomes, R. (2004). A pesquisa qualitativa nos serviços de saúde: notas teóricas. In E. Vozes (Ed.), *Pesquisa qualitativa de serviços de saúde* (pp. 99-120): Editora Vozes.
- Development, O. f. E. C. a. (2001). OECD Health Data 2001. In. Paris: Organisation for Economic Cooperation and Development
- Dillon, W. R. (1997). *Good Science*. Paper presented at the Marketing Research.
- Doll, R. (1974). Surveillance and monitoring. *International Journal of Epidemiology* 3, 10.
- Donabedian, A. (1980a). Basic approaches to assessment: structure, process and outcome. In H. A. Press (Ed.), *Explorations in Quality Assessment and Monitoring* (Vol. 1, pp. 77-125). Michigan: Health Administration Press.
- Donabedian, A. (1980b). *Explorations in Quality Assessment and Monitoring. The Definition of Quality and Approaches to its Assessment* (H. A. Press Ed. Vol. 1): Ann Arbor.
- Donabedian, A. (1980c). The definition of quality: A Conceitual exploration In H. A. Press (Ed.), *Explorations in Quality Assessment and Monitoring* (Vol. 1, pp. 30). Michigan.
- Donabedian, A. (1980d). *The Definition of Quality and Approaches to its Assessment* (A. Press Ed.): Health Ann Arbor.
- Donabedian, A. (1988). The quality of care: how can it be assessed? *The Journal of the American Medical Association* 260(12), 6.
- Donabedian, A. (1990). The seven pillars of quality. *Archives of Pathology Laboratory Medicine*, 114, 4.
- Donabedian, A. (2003). *An introduction to quality assurance in health care*. New York: Oxford University Press
- Dos Reis, E. J. F. B., Dos Santos, F. P., De Campos, F. E., Acúrcio, F. A., Leite, M. T. T., Leite, M. L. C., . . . Dos Santos, M. A. (1990). Avaliação da Qualidade dos Serviços de Saúde: Notas Bibliográficas. *Cadernos de Saúde Pública*, 6(1), 12.
- Driscoll, C., & Starik, M. (2004). The primordial stakeholder: Advancing the Conceitual consideration of stakeholder status for the natural environment. *Journal of Business Ethics*, 49, 19.
- Dudziak, E. A. (2001). *A information Literacy e o papel educacional das bibliotecas*. (Master). Universidade de São Paulo, São Paulo.
- Déniz-Déniz, M., & Zárraga-Oberty, C. (2004). The Assessment of the stakeholders' environment in the new age of knowledge: An empirical study of the influence of the organisational structure. *Business Ethics: A European Review*, 13(4), 17.
- EFQM, E. F. f. Q. M.-. (1999). Healthcare working group, Strategic directions. In. Brussels: EFQM Representative Office.
- Efron, B., & Tibshirani, R. J. (1993). *An Introduction to the Bootstrap* (C. Hall Ed.).
- Escorel, S., Giovanella, L., Mendonça, M. H. M., & Senna, M. C. M. (2007). O Programa de Saúde da Família e a construção de um novo modelo para a atenção básica no Brasil. *Rev. Panam. Salud Públ. /Pan Am. J. Public Health*, 21(2), 110.
- Escrivão, A. (2004). *Uso de indicadores de saúde na gestão de hospitais públicos da região metropolitana de São Paulo*. Retrieved from São Paulo: http://gvpesquisa.fgv.br/sites/gvpesquisa.fgv.br/files/publicacoes/P00275_1.pdf
- Esperidião, M., & Trad, L. A. B. (2005). Avaliação de satisfação de usuários. *Ciência e saúde coletiva [online]*, 10, 10.
- Esperidião, M. A. (2009). *O usuário e o julgamento dos serviços de saúde*. (Doctoral). Universidade Federal da Bahia, Bahia.
- Etzioni, A. (1975). *Comparative Analysis of Complex Organizations* (F. Press Ed.). New York.

- Fals Borda, O., & Mora-Osejo, L. E. (2004). La superación del Eurocentrismo. Enriquecimiento del saber sistémico y endógeno sobre nuestro contexto tropical. *Polis, Revista de la Universidad Bolivariana*, 2(7).
- Farrel, M. J. (1957). The measure of productive efficiency. *Journal of the Royal Statistical Society* 120, 28.
- Felix, J. D., Zandonade, E., Amorim, M. H. C., & Castro, D. S. (2012). Avaliação da completude das variáveis epidemiológicas do Sistema de Informação sobre Mortalidade em mulheres com óbitos por câncer de mama na Região Sudeste - Brasil (1998 a 2007) *Ciência & Saúde Coletiva*, 17(4), 945-953.
- Fernandes, F. M. B., Ribeiro, J. M., & Moreira, M. R. (2011). Reflexões sobre Avaliação de Políticas de Saúde no Brasil. *Cad. Saúde Pública*, 27(9), 11.
- Ferrara, F. A., Acebal, E., & Paganini, J. (1976). *Medicina de la Comunidad* (I. Editorial Ed. 2a ed.). Buenos Aires.
- Ferreira, J. J., Raposo, M. L., Rodrigues, R. G., Dinis, A., & Paço, A. (2012). A model of entrepreneurial intention: An application of the psychological and behavioral approaches. *Journal of Small Business and Enterprise Development*, 19, 17.
- Ferreira, P. L. (1993). Family of Instruments to Evaluate Hospital quality: a pilot test. In Coimbra, Portugal: Faculty of Economics.
- Ferreira, P. L., Luz, A., Valente, S., Raposo, V. M., Godinho, P., & Felício, E. D. (2001). Determinantes da satisfação dos Pacientes dos cuidados primários: o caso de Lisboa e Vale do Tejo [*Determinants of satisfaction of primary care users: the case of Lisbon and Tagus Valley*]. *Revista Portuguesa de Saúde Pública*, 2, 9.
- Ferreira, P. L., & Mendes, J. (1997). Hospital patient satisfaction: a Portuguese experience. European Healthcare Management Association Conference Changing management in healthcare: new missions, conditions & skills. *EHMA*, 69.
- Ferreira, P. L., & Raposo, V. M. (2006). A governança em saúde e a utilização de indicadores de satisfação [*Health governance and the use of satisfaction indicators*]. *Revista Portuguesa de Clínica Geral*, 22, 12.
- Ferreira, P. L., Raposo, V. M., & Pisco, L. (2017). A voz dos utilizadores dos cuidados de saúde primários da região de Lisboa e Vale do Tejo, Portugal [*The voice of Primary Care patients in the Lisbon and Tagus Valley region of Portugal*]. *Ciência & Saúde Coletiva*, 22(3), 12. doi:10.1590/1413-81232017223.33252016
- Fetter, R. B., Shin, Y., Freeman, J. L., Averill, R. F. E., & Thompson, J. D. (1980). Case Mix Definition by Diagnosis-Related Groups. *Medical Care*, 18, 53.
- FHEMIG, F. H. d. E. d. M. G.-. (2009). *Do Planejamento aos Resultados*. Retrieved from http://www.fhemig.mg.gov.br/index.php/docman/Publicacoes-3/Publicacoes_Institucionais/3140-cartilha-planejamento-web-1/file
- FIOCRUZ. (2003). *Relatório PROADESS, 2003*. Retrieved from Rio de Janeiro: <http://www.proadess.icict.fiocruz.br/relatoriofinal.pdf>
- FIOCRUZ. (2011). *Relatório PROADESS, 2011*. Retrieved from Rio de Janeiro: <http://www.proadess.icict.fiocruz.br/index.php?pag=produ>
- FIOCRUZ. (2012). *Relatório PROADESS, 2012*. Retrieved from Rio de Janeiro: <http://www.proadess.icict.fiocruz.br/RelatorioProadess2012.pdf>
- Fletcher, A., et al. (1993). Implications for trials in progress of publication of positive results. *Lancet* 342, 5.
- Flexner, A. (1910). *Medical Education in United States and Canada: Report to Carnegie Foundation for Advancement of Teaching*. New York: Merrymount Press

- Fontnote, G., Henke, L., Carson, K., & Carson, P. P. (2007). Techniques for determining importance: Balancing scientific method and subjectivity. *Journal of Targeting, Measurement and Analysis for Marketing*, 15, 11. doi:10.1057/palgrave.jt.5750043
- Fornell, C., & Larcker, D. F. (1981). Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Journal of Marketing Research*, 18(1), 39-50.
- Frasão, G. (2017). Em oito meses, número de cirurgias eletivas cresceu 39,1% no Brasil. *Portal Ministério da Saúde*. Retrieved from <http://portalms.saude.gov.br/noticias/agencia-saude/42101-em-oito-meses-numero-de-cirurgias-eletivas-cresceu-39-1-no-brasil>
- Freeborn, D. K., & Greenlick, M. R. (1973). Evaluation of the performance of ambulatory care system: research requirements and opportunities. *Medical Care*, 11, 8.
- Freeman, R. (1984). *Strategic management: A stakeholder approach* (Pitman Ed.). Boston: Pitman.
- Frenk, J. Dimensions of health system reform. 27, 16.
- Fund, K. s. (1998). *Organisational Audits Standards* (K. s. Fund Ed.). London: King's Fund.
- Gastal, F. L., La Forgia, G., Gonçalves, J. K., Melo, J. C., Siqueira, L. P., & Soares, T. M. (2005). Qualidade em saúde e acreditação no Brasil – experiências acumuladas, análise crítica e resultados. Consultant report for World Bank and Organização de Acreditação. In. Brasília, DF.
- Gattinara, B. C., Ibacache, J., Puente, C., Giaconi, J., & Caprara, A. (1995). Percepcion de la comunidad acerca de la calidad de los servicios de salud públicos en los distritos Norte e Ichilo, Bolívia. *Caderno de Saúde Publica*, 11, 14.
- Gennip, E. (1998). *The accreditation schema for acute hospitals* (TNO Ed.). Delft: TNO.
- Geraedts, H. P., Montenarie, R., & van Rijk, P. P. (2001). The benefits of total quality management *Comput Med Imaging Graph*, 25, 4.
- Giese, J. L., & Cote, J. A. (2000). Defining Consumer Satisfaction. *Academy of Marketing Science Review*, 2000, 24.
- Gil, A. C. (2010). *Como elaborar projetos de pesquisa* (Atlas Ed. 5a ed.). São Paulo.
- Giovinazzo, R. A., & Fischmann, A. A. (2001). *Delphi eletrônico – Uma experiência de utilização da metodologia de pesquisa e seu potencial de abrangência regional*. Paper presented at the XIV Congresso Latinoamericano de Estratégia, Buenos Aires.
- Glunk, V., & Wilderom, C. (1999). Organizational effectiveness = corporate performance? Why and how two research approaches need to be merged. In ISPA (Ed.), *Readings in organization science – Organizational change in a changing context* (pp. 155-181). Lisboa: ISPA.
- Gomes, A. S. (2009). *Mortalidade hospitalar: Modelo preditivos de risco usando os dados do sistema de internações hospitalares do SUS*. (Doctor). Universidade Federal do Rio Grande do Sul, Porto Alegre.
- Gonçalo, C. R., Borges, M. L., Vignochi, L., Haubert, M. P., & Junges, F. M. (2010). *Relevância de um sistema de indicadores no processo de avaliação do desempenho do setor da saúde*. Paper presented at the XXX Encontro Nacional de Engenharia de Produção, São Carlos, SP, Brasil.
- Gouveia, G. C., Souza, W. V. d., Luna, C. F., Souza-Júnior, de, P. R. B., & Szwarcwald, C. L. (2009). Satisfação dos usuários do sistema de saúde brasileiro: fatores associados e diferenças regionais. *Rev. Bras. Epidemiol*, 12(3), 15.
- Green, L. W., & Kreuter, M. W. (1999). *Health Promotion Planning: An Education and Ecological Approach* (M. P. Company Ed.). Mountain View: Mayfield Publishing Company.

- GREEN, S. B., AKEY, T. M., FLEMING, K. K., HERSHBERGER, S. L., & MARQUIS, J. G. (1997). The effect of the number of scale points on chi square fit indices in confirmatory factor analysis. *Structural Equation Modeling*, 4(2), 13.
- Greenfield, S., Nicolucci, A., & Mattke, S. Selecting Indicators for the Quality of Diabetes Care at the Health Systems Level in OECD Countries In (Vol. 15): OECD Health Technical Paper.
- Greenfield, S., Nicolucci, A., & Mattke, S. (2004). Selecting Indicators for the Quality of Diabetes Care at the Health Systems Level in OECD Countries In (Vol. 15): OECD Health Technical Paper.
- Groene, O. (2008a). An international review of projects on hospital performance assessment. *International Journal for Quality in Health Care*, 20(3), 10. doi:10.1093/intqhc/mzn008
- Groene, O. (2008b). The World Health Organization Performance Assessment Tool for Quality Improvement in Hospitals (PATH): Na Analysis of the Pilot Implementation in 37 Hospitals. *International Journal for Quality in Health Care*, 20(3), 7.
- Grönroos, C. (1978). A Service-Oriented Approach to Marketing of Services. *European Journal of Marketing*, 12(8), 1.
- Grönroos, C. (1982). *Strategic management and marketing in the service sector* (S. S. o. E. a. B. A. r. reports Ed.). Helsinki: Swedish School of Economics and Business Administration research reports.
- Grönroos, C. (1988). Service Quality: The Six Criteria Of Good Perceived Service. 9, 10.
- Grönroos, C. (1990). Marketing Redefined. *Management Decision*, 28(8), 1. Retrieved from <http://proquest.umi.com/pqdweb?did=7101082&Fmt=7&clientId=13656&RQT=309&VName=PQD>
- Grönroos, C. (2003). *Marketing: gerenciamento e serviços* (Elsevier Ed.). Rio de Janeiro: Elsevier.
- Guba, E. G., & Lincoln, Y. S. (2011). *Avaliação de quarta geração* (B. Honorato, Trans. E. Unicamp Ed.). São Paulo: Editora Unicamp.
- Guimarães, T. B., & Tavares-De-Almeida, B. (2006). Da estratégia aos Resultados concretos: a experiência do governo de Minas. *Cadernos BDMG*, 12, 76.
- Gummesson, E. (1977). Marknadsföring och ninköp av konsulttjänster (marketing and Purchasing of Professional Services). In (Akademilitteratur ed.). Stockholm: Stockholm Universitet.
- Guttman, L. (1954). Some necessary conditions for common factor analysis. *Psychometrika*, 19, 149-161.
- Hair, J. F., William, B., Babin, B., & Anderson, R. E. (2009). *Análise multivariada de dados* (Bookman Ed. 6a ed.). Porto Alegre.
- Hanson, R. (1992). Determining attribute importance. *Quirk's Marketing Research Review*. Retrieved from http://www.quirks.com/articles/article.asp?arg_articleid=430
- Hart, S., & Sharma, S. (2004). Engaging fringe stakeholders for competitive imagination. *Academy of Management Executive*, 18(1), 12.
- Hart, W. L., & Bogan, E. (1992). *The Baldrige* (McGraw-Hill Ed.). New York: McGraw-Hill.
- Hartz, S., & John, J. (2008). Contribution of economic evaluation to decision making in early phases of product development: A methodological and empirical review *International Journal of Technology Assessment in Health Care*, 24(4), 9. doi:10.1017/S0266462308080616
- Hartz , Z. M. A., & Vieira-da-Silva, L. M. (2005). Avaliação em saúde. *EDUFBA*, 25.
- Health, D. o. (2001). *NHS Performance Indicators: A Consultation*. London: Department of Health

- Health, D. o. (2002). *NHS Performance Indicators, National Figures: February 2002*. London: Department of Health
- Henseler, J., Ringle, C. M., & Sinkovics, R. R. (2009). The use of partial least squares path modeling in international marketing. In E. G. P. Limited (Ed.), *New Challenges to International Marketing (Advances in International Marketing (Vol. 20, pp. 277–319)*.
- Hercos, B. V. S., & Berezovsky, A. (2004). *Qualidade do serviço oftalmológico prestado aos pacientes ambulatoriais, do Sistema Único de Saúde – SUS*. Paper presented at the XV Congresso Norte-Nordeste de Oftalmologia, Belo Horizonte, Minas Gerais – Brasil.
- Hermann, R., & Mattke, S. Selecting Indicators for the Quality of Mental Health Care at the Health Systems Level in OECD Countries In (Vol. 17): OECD Health Technical Paper
- Hermann, R., & Mattke, S. (2004a). Selecting Indicators for the Quality of Mental Health Care at the Health Systems Level in OECD Countries In (Vol. 17): OECD Health Technical Paper
- Hermann, R., & Mattke, S. (2004b). Selecting Indicators for the Quality of Mental Health Care at the Health Systems Level in OECD Countries In (Vol. 18): OECD Health Technical Paper
- Hirsch, P., & Levin, D. (1999). Umbrella advocates versus validity police: A life-cycle model. *Organization Science, 10*(2), 15.
- Hoffmeyer, U. D., & Kluwer, M. T. R. (1994). Financing Health Care. In (Vol. 1): Academic Publishers.
- Holland, K., & Fennell, S. (2000). Clinical governance is ‘ACE’ – using the EFQM excellence model to support baseline assessment *Int J Health Care Qual Assur Inc Leadersh Health Serv, 13*, 8.
- Hollander, M., & Wolfe, D. A. (1999). *Nonparametric Statistical Methods* (J. W. Sons Ed.). New York.
- Hood, C. (1990). Beyond the public bureaucracy state? Public administration in the 1990s', inaugural lecture. In (Vol. 16): London School of Economics.
- Hood, C. (1991). A public management for all seasons? *Public Administration, 69*, 3-19.
- Hsiao, C. W. (1995). A framework for assessing health financing strategies and the role of health insurance, in An International assessment of health care financing. Lessons for developing countries In.
- Hsiao, W. H., & Sidat, B. (2008). Health systems: Concepts and deterministic models of performance. In: Background paper prepared for the Workshop on Research Agendas on Global Health Systems, Harvard University.
- Hurst, J. (2000). Challenges for health systems in Member Countries of the Organisation for Economic Co-operation and Development. *Bull World Health Organ, 78*, 10.
- Hurst, J. (2002). Performance measurement and improvement in OECD health systems: overview of issues and challenges. In E. d. OCDE (Ed.), *Measuring up – improving health system performance in OECD countries* (pp. 20). Paris: OECD.
- Hutubessy, R. C. W. (2003). *Cost-effectiveness analysis for sector-wide priority setting in health*. (Doctoral). Erasmus University Rotterdam, Holand.
- Ibrahim, J. E. (2001). Performance indicators from all perspectives. *Int J Qual Health Care 13*(6), 2.
- Identification of a Set of Key Quality Indicators in the Hospital Setting Using the European Excellence Model (2001). In. Madrid: Club Gestión de Calidad
- Iezzoni, L. I. (1997). The Risks of Risk Adjustment. *JAMA, 278*(19), 8.
- Iezzoni, L. I. (2003). Risk adjustment for measuring health care outcomes. *Health Administration Press*.

- IHPS. (2005). *Risk adjustment methods and their relevance to “Pay-or-Play”: supplement E to the report: Challenges and alternatives for employer Pay-or-Play Program design: an implementation and alternative scenario analysis of California’s “Health Insurance Act of 2003”*. Washington, DC
- Jackson, S. (1999). Exploring the possible reasons why the UK Government commended the EFQM (European Foundation for Quality Management) excellence model as the framework for delivering governance in the new NHS *Int J Health Care Qual Assur Inc Leadersh Health Serv* 12, 10.
- Jackson, S., & Bircher, R. (2002). Transforming a run down general practice into a leading edge primary care organisation with the help of the EFQM excellence model *Int J Health Care Qual Assur Inc Leadersh Health Serv* 15, 13.
- JCAHO, J. C. o. A. o. H. O.-. (1997). *National Library of Healthcare Indicators™ Health Plan and Network Edition* Oakbrook Terrace: Joint Commission on Accreditation of Healthcare Organizations.
- Jones, J., & Hunter, D. (1995). Qualitative research: consensus methods for medical and health services reseach. *British Medical Journal*, 331(7001), 5.
- Junior, F. A. T. (2009). *Instrumentos de contratualização e a pactuação por resultados. Inovações no sistema de gestão do Sistema Único de Saúde (SUS)*. Paper presented at the II Congresso Consad de Gestão Pública, Brasília.
- Junior, F. A. T. (2013). *Alinhamento Estratégico: O processo de contratualização por resultados em Minas Gerais por meio do sistema GEICOM*. Paper presented at the VI Congresso Consad de Gestão Pública, Brasília.
- Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20, 141-151.
- Kaplan, R. S., & Norton, D. P. (1992). The Balanced Scorecard – Measures that Drive Performance. *Harvard Business Review*, 70(1), 8.
- Katz, D., & Kahn, R. (1977). *Psicologia social de las organizaciones* (Trilas Ed.). México: Trilas.
- Kazandjian, V. (2003). Accountability Through Measurement: A Global Health Care Imperative *ASQ Quality Press*, 2.
- Kazandjian, V., & Lied, T. (1999). Healthcare performance measurement: systems design and evaluation *ASQ Health Care Series*, 38.
- Kazandjian, V. A. (2003). Are performance indicators generic? The international experience of the Quality Indicator Project *Journal of Evaluation in Clinical Practice*, 9(2), 12.
- Klazinga, N. (2000). Re-engineering trust: the adoption and adaptation of four models for external quality assurance of health care services in western European health care systems *Int J Qual Health Care*, 12, 7.
- Knowles, J. C., Leighton, C., & Stinson, W. Measuring results of health sector reform for system performance: a handbook of indicators In. Bethesda Maryland: Partnerships for Health Reform, Special Initiatives Report No. 1
- Kopalle, P. K., & Lehmann, D. R. (2001). Strategic management of expectations: The role of disconfirmation sensitivity and perfectionism. *Journal of Marketing Research*, 38(3), 9.
- Kujala, J. (2001). Analysing moral issues in stakeholder relations. *Business Ethics: A European Review*, 10(3), 15.
- Kutzin, J. Towards Universal Health Care Coverage: A Goal-Oriented Framework for Policy Analysis. In.
- La Forgia, G. M., & Couttolenc, B. F. (2009). *Desempenho Hospitalar no Brasil: Em busca da Excelência* (E. Singular Ed.). São Paulo: Editora Singular.

- Lakatos, E. M., & Marconi, M. A. (2007). *Metodologia científica* (Atlas Ed. 7a ed.). São Paulo.
- Lambie, L., & Mattke, S. Selecting Indicators for the Quality of Cardiac Care at the Health Systems Level in OECD Countries In (Vol. 14): OECD Health Technical Paper.
- Lambie, L., & Mattke, S. (2004). Selecting Indicators for the Quality of Cardiac Care at the Health Systems Level in OECD Countries In (Vol. 14): OECD Health Technical Paper.
- Lane, D. S., & Kelman, H. R. (1975). Assessment of maternal health care quality: Conceptual and methodologic issues. *Medical Care*, 13, 17.
- Lawton, A., McKevitt, D., & Millar, M. (2000a). Coping with ambiguity: Reconciling external legitimacy and organizational implementation in performance measurement. *Public Money & Management*, 7.
- Lawton, A., McKevitt, D., & Millar, M. (2000b). Coping with ambiguity: Reconciling external legitimacy and organizational implementation in performance measurement. *Public Money & Management*, 7.
- Lemke, C. (2005). *Modelos de equações estruturais com ênfase em Análise Fatorial Confirmatória no Software Amos*. (Bacharel). Universidade federal do Rio Grande do Sul, Porto Alegre.
- Lichtig, L. K. (1986). *Hospital Information Systems for Case Mix Management* (J. W. a. Sons Ed.). New York: John Wiley & Sons.
- Lima, H. d. O., Lima, M. F., Campos, L. I., & Oliveira, B. R. d. (2010). Fortalecimento e Qualificação da Rede Hospitalar: Pro-Hosp. In *O Choque de Gestão na Saúde em Minas Gerais* (Vol. 324, pp. 27). Belo Horizonte Secretaria de Estado de Saúde de Minas Gerais.
- Lima, J. G. (2016). *Atributos da Atenção Primária nas regiões de saúde: uma análise dos dados do Programa Nacional de Melhoria do Acesso e da Qualidade da Atenção Básica*. FIOCRUZ, Rio de Janeiro
- Londono, J. L., & Frenk, J. (1997). Structured Pluralism: towards an innovative model for health system reform in Latin America *Health Policy*, 41, 36.
- Lopes, E. S., & Pardal, L. (2011). *Métodos e Técnicas de Investigação Social* (A. Editores Ed.).
- Lorenzo, S., Arcelay, A., & Bacigalupe, M. (2001). *Guide for Self-assessment of Healthcare Centres Using as a Reference the Self-assessment Model of the European Foundation for Quality Management (EFQM)* (MSD Ed.). Madrid: MSD.
- Lourença, P. R., Miguez, J., Gomes, A. D., & Carvalho, C. (2004). Eficácia grupal: Análise e discussão de um modelo multi-dimensional. *Psicologica*, 11.
- Macente, L. B., & Zandonade, E. (2010). Avaliação da completude do Sistema de Informação sobre Mortalidade por suicídio na região Sudeste, Brasil, no período de 1996 a 2007 *J Bras Psiquiatr*, 59(3), 173-181.
- Machado, J. P., Martins, A. C. M., & Martins, M. S. (2013). Avaliação da qualidade do cuidado hospitalar no Brasil: uma revisão sistemática. *Cad. Saúde Pública*, 29(6), 20.
- Malcolm Balridge National Quality Award, 1999 Criteria for Performance Excellence. (1999). In. Gaithersburg: NIST.
- Malik, A. M., & Schiesari, L. M. C. (1998). Qualidade na Gestão Local de Serviços e Ações de Saúde. In. São Paulo: Faculdade de Saúde Pública da Universidade de São Paulo
- Marin, H. F. (2010). Sistemas de Informação em Saúde: considerações gerais. *J. Health Inform*, 2(1), 17.
- Marques, A. J. S., Mendes, E. V., & Lima, H. O. (2010). Fortalecimento e qualificação da rede hospitalar – Pro-Hosp. In S. d. E. d. S. d. M. Gerais (Ed.), *O Choque de Gestão*

- em Minas Gerais: resultados na saúde*. Minas gerais: Secretaria de Estado de Saúde de Minas Gerais.
- Marshall, M., Leatherman, S., & Mattke, S. Selecting Indicators for the Quality of Health Promotion, Prevention and Primary Care at the Health Systems Level in OECD Countries In (Vol. 16): OECD Health Technical Paper
- Marshall, M., Leatherman, S., & Mattke, S. (2004). Selecting Indicators for the Quality of Health Promotion, Prevention and Primary Care at the Health Systems Level in OECD Countries In (Vol. 16): OECD Health Technical Paper
- Martinez, L. F., & Ferreira, A. I. (2008). *Análise de Dados com SPSS* (E. Editora Ed. 2a ed.). Lisboa.
- Marôco, J. (2014). *Análise de Equações Estruturais: fundamentos teóricos, software e aplicações* (ReportNumber Ed. 2a ed.). Pêro Pinheiro.
- Mattke, S. (2004). Monitoring and improving the technical quality of medical care: a new challenge for policymakers in OECD countries In P. Studies (Ed.), *Towards High Performing Health Systems*. Paris: OECD.
- Mattke, S., Epstein, A. M., & Leatherman, S. (2006). The OECD Health Care Quality Indicators Project: history and background *International Journal for Quality in Health Care*, 4. doi:10.1093/intqhc/mzl019
- Mattke, S., Kelley, E., & Scherer, P. (2004). *Health Care Quality Indicators Project Initial Indicators Report*. Retrieved from Paris: <http://www.oecd.org/dataoecd/1/34/36262514.pdf>
- Matus, C. (1997). *Política, Planejamento e Governo* (IPEA Ed. 3a ed.). Brasília: Tomo I.
- McGlynn, E. A. (1997). Six challenges in measuring the quality of health care. *Health Affairs*, 16, 15.
- McKee, M., & Healy, J. (2002). Pressures for Change. In O. University (Ed.), *Hospitals in a Changing Europe* (pp. 49). Buckingham: Open University.
- McLaughlin, J. A., & Jordan, G. B. (1999). Logic models: A tool for telling your program's performance history. *Evaluation and Program Planning*, 22(1), 8.
- McLoughlin, V., Leatherman, S., Fletcher, M., & Owen, J. W. (2001). Improving performance using indicators: Recent experiences in the United States, the United Kingdom, and Australia *Int Journal Qual Health Care* 13, 8.
- Medicine, I. o. (1990). Medicare: A Strategy for Quality Assurance. In (Vol. 2). Washington: National Academy Press.
- Medicine, I. o. (2001). *Crossing the quality chasm: a new health system for the 21st century*. Retrieved from Washington:
- Medina, M. G., Silva, G. A. P., Aquino, R., & Hartz, Z. M. A. (2005). Uso de modelos teóricos na avaliação em saúde-aspectos conceituais e operacionais. In EDUFBA (Ed.), *Avaliação em saúde* (pp. 41-63). Salvador: EDUFBA.
- Mellahi, K., & Wood, J. (2003). The role and potential of stakeholders in "holoow participation": Conventional stakeholder theory and institutionalism alternatives. *Business and Society Review*, 108(2), 20.
- Mendes, E. V. (2011). *As redes de atenção à saúde* (O. P.-A. d. Saúde Ed.). Brasília: Organização Pan-Americana da Saúde.
- Millar, J., & Mattke, S. Selecting Indicators for Patient Safety at the Health Systems Level in OECD Countries. In (Vol. 18): OECD Health Technical Paper.
- Millar, J., & Mattke, S. (2004). Selecting Indicators for Patient Safety at the Health Systems Level in OECD Countries. In (Vol. 18): OECD Health Technical Paper.
- Minas, G. (2010a). *Decreto nº 45.468, de 13 de setembro de 2010*. Belo Horizonte: Assembléia Legislativa de Minas Gerais

- Minas, G. (2010b). *O Choque de Gestão em Minas Gerais: resultados na saúde*. Belo Horizonte: Secretaria de Estado de Saúde de Minas Gerais
- Minas, G. (2012). *Resolução SES/MG n° 3.407, de 21 de Agosto de 2012*. Belo Horizonte: Governo do Estado de Minas Gerais
- Minayo, M. C. S. (2006). *O desafio do conhecimento: pesquisa qualitativa em saúde*. São Paulo: Hucitec.
- Minayo, M. C. S., Assis, S. G. d., & Sousa, E. R. (2010). *Avaliação por triangulação de métodos - Abordagem de programas sociais* (F. O. Cruz Ed.). Rio de Janeiro: Fundação Oswaldo Cruz.
- Mira, J. J., Lorenzo, S., Rodríguez-Marín, J., Aranaz, J., & Sitges, E. (1998). Application of the European improvement model to healthcare: benefits and limitations *Rev Calidad Asistencial*, 13, 6.
- Mitchell, R., Agle, B., & Wood, D. (1997). Toward a theory of stakeholder identification and salience: Defining the principle of who and what really counts. *Academy of Management Review*, 22(4), 34.
- Mittal, V., Ross, W. T. J., & Baldasare, P. M. (1998). The asymmetric impact of negative and positive attribute-level performance on overall satisfaction and repurchase intentions. *Journal of Marketing*, 62(1), 1.
- Moeller, J. (2001). The EFQM Excellence Model. German experiences with the EFQM approach in health care. *International Journal for Quality in Health Care*, 13(1), 5.
- Moeller, J., Breinlinger, J., O'Reilly, & Elser, J. (2000). Quality management in German health care—the EFQM Excellence Model *Int J Health Care Qual Assur Inc Leadersh Health Serv* 13, 5.
- Moeller, J., & Sonntag, H. G. (1998). Systematic analysis and controlling of health care organisations lead to numerical health care improvements. *Health Manpower Manage*, 24, 5.
- Monteiro, W. M. S. (2008). *Avaliação da qualidade dos serviços prestados em unidades básicas de saúde de fortaleza*. (Master). Universidade Estadual do Ceará, Fortaleza.
- Moracho, O., Colina, A., Amondarain, M. A., Aguirre, L., Ruiz-Álvarez, E., & Salgado, M. V. (2001). Practical experience of the external evaluation process with the EFQM Excellence Model in the Hospital of Zumarraga *Rev Calidad Asistencial*, 16, 8.
- Morin, E., Savoie, A., & Beaudin, G. (1994). *L'efficacité de l'organisation: Théories, représentations et mesures* (G. M. Éditeur Ed.). Montréal: Gaetan Morin Éditeur.
- Morra Imas, L. G., & Rist, R. C. (2009). *The Road to Results Designing and Conducting Effective Development Evaluations*. Retrieved from Washington: <https://openknowledge.worldbank.org/handle/10986/2699>
- Murray, C. J. L., & Frenk, J. (2000a). A framework for assessing the performance of health systems *Bulletin of the World Health Organization*, 78(6), 15.
- Murray, C. J. L., & Frenk, J. (2000b). *A WHO Framework for Health System Performance Assessment*. Retrieved from Geneva: <http://www.who.int/iris/handle/10665/66267>
- Nabitz, U. W., & Klazinga, N. S. (1999). EFQM approach and the Dutch Quality Award *International Journal of Health Care Quality Assurance*, 12(2), 6.
- Nabitz, U. W., & Walburg, J. A. (2000). Addicted to quality—winning the Dutch Quality Award based on the EFQM Model. *Int J Health Care Qual Assur Inc Leadersh Health Serv* 13, 7.
- Navarro, J., & Quijano, S. (2006). La efectividad organizativa. In Icaria (Ed.), *Dirección de recursos humanos y consultoría en las organizaciones* (pp. 297-326). Barcelona: Icaria.
- Navarro, V. (2000). Assessment of the World Health Report 2000. *Lancet*, 356, 4.

- NHCP, N. H. P. C. (2002). *National Report on Health Sector Performance Indicators 2001*. Brisbane
- NHS. (1999). *Quality and performance in the NHS: High level performance indicators*. London
- NHS. (2002). *NHS Performance Indicators – Acute NHS Hospital Trusts*. United Kingdom: National Health Service
- Noronha, J. C. (2001). Sistema de Informações Hospitalares como ajuste de risco em índices de desempenho. *Rev. Saúde Pública*, 35(2), 8.
- Noronha, M. F. (2001). *Classificação de hospitalizações em Ribeirão Preto: os Diagnosis Related Groups*. (Doctoral). USP, São Paulo.
- Noronha, M. F., Veras, C. T., Leite, I. C., Martins, M. S., Neto, F. B., & Silver, L. (1991). O desenvolvimento dos "Diagnosis Related Groups"- DRGs. Metodologia de classificação de pacientes hospitalares *Revista de Saúde Pública de São Paulo*, 25(3), 11.
- Novaes, H. M. D. (2000). Avaliação de Programas, Serviços e Tecnologias em Saúde. *Rev. Saúde Pública*, 34(5), 13. Retrieved from www.fsp.usp.br/rsp
- Nunnally, J. C., & Bernstein, I. H. (1994). The Assessment of Reliability. *Psychometric Theory*, 3, 248-292.
- OECD. An Assessment of Health System Performance across OECD Countries In: OECD DEELSA/ELSA/WPI(99)3.
- OECD. Health outcome measurement in OECD countries: toward outcome-oriented policy making In. Paris: OECD DEELSA/ELSA/WP1(98)6/ANN.
- OECD. (2001). *Performance measurement and performance management in OECD health systems* (47). Retrieved from www1.oecd.org
- OECD. (2013). *Health at a Glance 2013*. Retrieved from http://dx.doi.org/10.1787/health_glance-2013-en
- OECD. (2015a). *Health at a Glance 2015: OECD Indicators*. Retrieved from Paris: http://dx.doi.org/10.1787/health_glance-2015-en
- OECD. (2015b). *OECD Health Statistics 2015*. Retrieved from <http://dx.doi.org/10.1787/health-data-en>
- OECD. (2017). *Health at a Glance 2017: OECD Indicators*. Retrieved from Paris: http://dx.doi.org/10.1787/health_glance-2017-en
- OECD (Ed.) (2002). *Glossary of Key Terms in Evaluation and Results-Based Management*. Paris.
- Oliver, R. L. (1980). A Cognitive model of the antecedents and consequences of satisfaction decisions. *Journal of Marketing Research*, 17(4), 10.
- Oliver, R. L. (1981a). Measurement and evaluation of the satisfaction process in retail settings. *Journal of Retailing*, 57(3), 24.
- Oliver, R. L. (1981b). What in customer satisfaction? *Wharton Magazine*, 5(3), 6.
- Oliver, R. L. (1997). *Satisfaction, a behavioral perspective on the consumer* (Irwin/MacGraw-Hill Ed.): Irwin/MacGraw-Hill.
- Oliver, R. L., & Desarbo, W. S. (1988). Response Determinants in Satisfaction Judgments. *Journal of Consumer Research*, 14, 13.
- Oliver, R. L., & Linda, G. (1981). Effects of Satisfaction and Its Antecedents on Consumer Preference and Intention. *Advances in Consumer Research*, 8, 6.
- OMS. (1970). *Comite de Expertos de la OMS en Estadística Sanitaria. Indicadores estadísticos para la planificación y la evaluación de programas de salud pública* Retrieved from Ginebra:
- Onocko-Campos, R., & Furtado, J. P. (2005). A transposição das políticas de saúde mental no Brasil para a prática nos novos serviços. *Rev. Latinoam. Psicopatol. Fundam*, 1(1), 14.

- OR, Z. (2002). Improving the performance of health systems: from measures to action (a review of experiences in four countries). *Labour Market and Social Policy – Occasional Papers*, 57. doi:10.1787/18151981
- Organization, P. A. H. (1999). *Guidelines for the preparation of the health services system profile in the countries of the region* (P. A. H. Organization Ed.). Washington DC: Pan American Health Organization.
- OTA. (1978). *Assessing the Efficacy and Safety of Medical Technologies* (OTA Ed.): Washington.
- OTA (Ed.) (1988). *Hospital Mortality Rates*. Washington D.C: Congress of the United States: Office of Technology Assessment.
- Ovretveit, J. (1998). *Evaluation Health Intervention: an introduction to evaluation of health treatments, services, policies, and organizational* (O. U. Press Ed.). Buckingham: Open University Press.
- Paim, J. S., & Teixeira, C. F. (2006). Política, planejamento e gestão em saúde: balanço do estado da arte. *Rev. Saude Publica*, 40, 6.
- Papanicolas, I., & Smith, P. C. (2010). EuroREACH Framework for Health System Performance. In.
- Parasuraman, A., Berry, L. L., & Zeithaml, V. A. (1985a). A Conceptual Model of Service Quality and its Implications for Future Research. *Journal of Marketing*, 49, 10.
- Parasuraman, A., Berry, L. L., & Zeithaml, V. A. (1985b). A Conceptual Model of Service Quality and its Implications for Future Research. *Journal of Marketing*, 49, 41-50.
- Parasuraman, A., Berry, L. L., & Zeithaml, V. A. (1988). SERVQUAL: A Multiple-Item Scale for Measuring Consumer Perceptions of Service Quality. *Journal of Retailing*, 64, 29.
- Parasuraman, A., Berry, L. L., & Zeithaml, V. A. (1991a). Refinement and Reassessment of the SERVQUAL Scale. *Journal of Retailing*, 67(4).
- Parasuraman, A., Berry, L. L., & Zeithaml, V. A. (1991b). Understanding customer expectation of service. *Sloan Management Review*, 31, 10.
- Parasuraman, A., Zeithaml, V., & Berry, L. (1994). Reassessment of Expectations as a Comparison Standard in Measuring Service Quality: Implications for Further Research. *Journal of Marketing*, 58, 14.
- Patton, M. Q. (1997). *Utilization-Focused Evaluation: The New Century Text* (S. Publications Ed.). London: Sage Publications.
- Pedro, S., Ferreira, P. L., & Mendes, J. (2010). Aplicação de métodos estatísticos na avaliação da satisfação dos Pacientes com o internamento hospitalar. *Tourism & Management Studies*, 6, 15.
- Perez Arias, E. B., & Feller, J. J. (1983). El Control de los Sistemas de Atención Médica Conceptualización y Mecánica Operativa. *Medicina y Sociedad*, 6(6), 8.
- Pinto, R. (2010a). *Sistemas de informações hospitalares de Brasil, Espanha e Portugal – semelhanças e diferenças*. (Master). FioCruz, Rio de Janeiro.
- Pinto, R. (2010b). *Sistemas de informações hospitalares de Brasil, Espanha e Portugal – semelhanças e diferenças*. (Master). FioCruz, Rio de Janeiro.
- Pitt, D. J. (1999). Improving performance through self-assessment. *Int J Health Care Qual Assur Inc Leadersh Health Serv*, 12, 9.
- Podinovski, V. V. (2004). Production trade-offs and weight restrictions in data envelopment analysis. *Journal of the Operational Research Society*, 55(12), 12.
- Porterfield, J. D. (1976). Evaluation of Patients: Codman Revisited. *Acad. Med*, 52(1), 9.
- Ramalho, N. (2005). *Construção de um modelo metamórfico da eficácia organizacional: Papel estruturante da cultura de gestão*. (Doctoral). Instituto Superior de Ciências do Trabalho e da Empresa, Lisboa.

- Ramos, L. M. A., & De Oliveira, C. B. (2010). O choque de Gestão nos hospitais públicos de Minas Gerais - FHEMIG. In *O choque de gestão na saúde em Minas Gerais* (pp. 263-275). Belo Horizonte.
- Raposo, V. (2007). *Governança hospitalar - uma proposta Conceitual e metodológica para o caso português*. Universidade de Coimbra, Coimbra.
- Reilly, T., Zema, C., & Crofton, C. (2002). Providing performance information for consumers: experience from the United States. In *Measuring Up. Improving Health Systems Performance in OECD Countries* (pp. 97-116). Ottawa: OECD.
- Rerimann, B. (1982). Organizational Competence as a predictor of long run survival and growth. *Academy of Management Journal*, 25(2), 12.
- Ressel, L. B. (2008). O uso do grupo focal em pesquisa qualitativa. *Texto Contexto Enfermagem*, 17(4), 779-786.
- RIPSA, R. I. d. I. p. a. S. n. B.-. (2007). *A iniciativa Ripsa: origem e evolução (1995-2007)*. Retrieved from <http://www.ripsa.org.br/lildbi/docsonline/get.php?id=225>
- Rocha, T. A. H., Da Silva, N. C., Barbosa, A. C. Q., Amaral, P. V., Thumé, E., Rocha, J. V., . . . Facchini, L. A. (2017). Cadastro Nacional de Estabelecimentos de Saúde: evidências sobre a confiabilidade dos dados. *Ciencia & Saude Coletiva*, 23(1), 229-240.
- Rodrigues, W. W., & Torres, V. R. (1982). *Contribuição ao Desenvolvimento do Processo de Aviação em Serviços de Saúde*. Brasília: Ministério da Saúde
- Rodríguez, B. A., & Ferrándiz-Santos, B. (2004). Integration of the EFQM Model and the Hoshin Kanri deployment in a primary care area. *Rev Calidad Asistencial*, 19, 8.
- Rogers, E., & Wright, P. (1998). Measuring organizational performance in strategic human resource management: problems, prospects, and performance information markets. *Human Resource Management Review*, 8(3), 21.
- Rootman, I., Goodstadt, M., Hyndman, B., McQueen, D. V., Potvin, L., Rossi, P. H., & Freeman, H. E. (2004). *Evaluation, a systematic approach* (S. Publications Ed.). Beverly Hills: Sage Publications.
- Rosalem, V. (2013). *Análise das percepções dos principais atores da cadeia produtiva da saúde sobre a qualidade dos serviços prestados por hospitais no Estado de Goiás - Brasil*. (Doctoral). Escola de Administração de Empresas de São Paulo, São Paulo.
- Roski, J., & Gregory, R. Performance measurement for ambulatory care: moving towards a new agenda. *Int J Qual Health Care*, 13(6), 7.
- Rotta, C. S. G. (2004). *Utilização de indicadores de desempenho hospitalar como instrumento gerencial*. (Doctoral). São Paulo.
- Rowan, M. S. (2000). Logic models in primary care reform. navigating the evaluation. *Canadian Journal of Program Evaluation*, 15(2), 12.
- Rundall, T. G. (1992). Health planning and evaluation. In A. Lange (Ed.), *Public Health & Preventive Medicine* (13a ed., pp. 16). Connecticut.
- Samico, I., Felisberto, E., Figueiró, A. C., & Frias, P. G. (2010). *Avaliação em saúde: bases conceituais e operacionais* (Medbook Ed.). Rio de Janeiro: Medbook.
- Sampieri, R. H., Collado, C. F., & Lúcio, P. B. (2006). *Metodologia de pesquisa* (M. Hill Ed.). São Paulo: McGraw Hill.
- Santos, M. P. (1995). Avaliação da qualidade dos serviços públicos de atenção à saúde da criança sob a ótica do usuário. *Revista Brasileira de Enfermagem*, 48, 11.
- Santos, S. B. (2007). Perspectivas da avaliação na Política Nacional de Humanização em Saúde: aspectos conceituais e metodológicos. *Ciência & Saúde Coletiva*, 12(4), 12.
- Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In S. Publications (Ed.), *Latent Variables Analysis: Applications for Developmental Research* (pp. 399-419). Thousand Oaks, CA.

- Savin-Baden, M., & Major, C. H. (2013). *Qualitative research: the essential guide to theory and practice* (Routledge Ed.). Abingdon: Routledge.
- Savoie, A., & Morin, E. (2001). Representações da eficácia organizacional: desenvolvimentos recentes. *Psychologica*.
- Schramm, J. M. A., Oliveira, A. F., Leite, I. C., Valente, J. G., Gadelha, A. M. J., Portela, M. C., & Campos, M. R. (2004). Transição epidemiológica e o estudo de carga de doença no Brasil. *Ciência & Saúde Coletiva*, 9(4), 12.
- Scriven, M. (1991). *Evaluation thesaurus* (S. Publications Ed.). Newbury Park: Sage Publications.
- Seruf-Filho, J. C. S. (2014). *Avaliação da produtividade de hospitais brasileiros pela metodologia do diagnosis related Groups 145.710 altas em 116 hospitais*. (Master). Universidade de Minas Gerais, Minas Gerais.
- Services, D. o. H. a. H. (2000). Healthy People 2010. Understanding and Improving Health. In. Washington: Government Printing Office
- Shaughnessy, P., & Hittle, D. (2002). Overview of risk adjustment and outcome measures for home health agency OBQI reports: highlights of current approaches and outline of planned enhancements. *Center For Health Services Research*.
- Sheldon, T. (1998). Promoting health care quality: what role performance indicators? *Quality in Health Care*, 6.
- Sicotte, C., Champagne, F., Contandriopoulos, A., Barnsley, J., Beland, F., Leggat, S., . . . Brémond, M. (1998). A Conceptual framework for the analysis of health care organizations' performance *Health services management research*, 11(24).
- Silva, L. M. V., & Formigli, V. L. A. (1994). Avaliação em Saúde: Limites e Perspectivas. *Cad. Saúde Pública*, 10(1), 12.
- Simonet, D. (2011). The New Public Management Theory and the Reform of European Health Care Systems: An International Comparative Perspective *International Journal of Public Administration*, 34(12), 13. doi:10.1080/01900692.2011.603401
- Simón, R., Guix, J., Nualart, L., Surroca, R. M., & Carbonell, J. M. (2001). Use of several models as diagnostic and quality improvement tool: EFQM and Joint Commission. *Rev Calidad Asistencial*, 16, 5.
- Smee, C. H. (2002). Improving Value for Money in the United Kingdom National Health Service. In *Measuring UP: Improving Health System Performance in OECD Countries* (pp. 57-86). Paris: OECD Publications Service.
- Souza, L. E. P. F., Vieira-Da-Silva, L. M., & Hartz, Z. M. A. (2005). *Avaliação em Saúde: dos modelos teóricos à prática na avaliação de programas e sistemas de saúde*. Paper presented at the Conferência de consenso sobre a imagem-objetivo da descentralização da atenção à saúde no Brasil, Rio de Janeiro.
- Steiger, J. H., Shapiro, A., & Browne, M. W. (1985). On the multivariate asymptotic distribution of sequential chi-square statistics. *Psychometrika*, 50, 253–264.
- Stevens, P., Stokes, L., & O'Mahony, M. (2006). Metrics, targets and performance. *National Institute Economic Review*, 197(1), 13.
- Stewart, A. (2003). An investigation of the suitability of the EFQM Excellence Model for a pharmacy department within an NHS Trust *Int J Health Care Qual Assur Inc Leadersh Health Serv*, 16, 12.
- Stott, N., Kinnersley, P., & Elwyn, G. J. (1997). Measuring general practice based primary care generic outcomes. *Family Pract*, 14, 6.
- Strauss, A., & Corbin, J. (1990). *Basic of qualitative research: grounded theory and techniques* (Sage Ed.). London: Sage.
- Suchman, M. (1995). Managing legitimacy: strategic and institucional approaches. *Academy of Management Review*, 20(3), 40.

- Surprenant, C. F., & Solomon, M. R. (1987). Predictability and Personalization in the Service Encounter. *Journal of Marketing*, 51(2), 1.
- Swan, J. E., & Trawick, F. I. (1981). Disconfirmation of expectations and satisfaction with a retail service. *Journal of Retailing*, 57(3), 19.
- Szymanski, D. M., & Henard, D. H. (2001). Customer satisfaction: A meta-analysis of the empirical evidence. *Academy of Marketing Science. Journal*, 29(1), 1.
- Tamaki, E. M., Tanaka, O. Y., Felisberto, E., Alves, C. K. d. A., Junior, M. D., Bezerra, L. C. d. A., . . . De Miranda, A. S. (2012). Metodologia de construção de um painel de indicadores para o monitoramento e a avaliação da gestão do SUS. *Ciência & Saúde Coletiva*, 17(4), 11.
- Tanaka, O. Y., & Melo, C. (2008). Avaliação de serviços e programas de saúde para a tomada de decisão. In Atheneu (Ed.), *Saúde pública: bases conceituais* (pp. 119-131). São Paulo: Atheneu.
- Tanaka, O. Y., & Tamaki, E. M. (2012a). O papel da avaliação para a tomada de decisão na gestão de serviços de saúde. *Ciência & Saúde Coletiva*, 17(4), 8.
- Tanaka, O. Y., & Tamaki, E. M. (2012b). O papel da avaliação para a tomada de decisão na gestão de serviços de saúde. *Ciência & Saúde Coletiva*, 17(4), 8.
- Tenenhaus, M., Esposito Vinzi, V., Chatelin, Y., & Lauro, C. (2005). PLS path modeling. *Computational Statistics and Data Analysis*, 48, 159–205.
- Travassos, C., Noronha, J. C., & Martins, M. (1999). Mortalidade Hospitalar como Indicador de Qualidade: Uma Revisão. *Ciência & Saúde Coletiva*, 4(2).
- Trieschmann, J., Dennis, A., Northcraft, G., & Niemi-Jr, A. (2000). Serving multiple constituencies in business schools: M. B. A. program versus research performance. *Academy of Management Journal*, 43(6), 12.
- Triviños, A. N. S. (1987). *Introdução à pesquisa em ciências sociais: a pesquisa qualitativa em educação* (Atlas Ed.). São Paulo: Atlas.
- Tse, D. K., & Wilton, P. C. (1988). Models of Consumer Satisfaction: Na Extension. *Journal of Marketing Research*, 25, 9.
- Uchimura, K. Y., & Bosi, M. L. M. (2002). Qualidade e subjetividade na avaliação de programas e serviços em saúde *Cad. Saúde Pública*, 18(6), 9.
- UNDP. (2009). *Handbook on Planning, Monitoring and Evaluation for Development Results*.
- Unit, O. H. P. (2001). Health at a Glance. In Paris: Organisation for Economic Co-operation and Development
- US/DHHS. (2000). Healthy People 2010: Leading health indicators. In.
- Vallejo, P., Saura, R. M., Sunol, R., Kazandjian, V., Ureña, V., & Mauri, J. (2006). A proposed adaptation of the EFQM fundamental concepts of excellence to health care based on the PATH framework. *International Journal for Quality in Health Care*, 18(5), 9.
- Veillard, J. (2012). *Performance Management in Health Systems and Services Studies on its Development and Use at International, National/Jurisdictional, and Hospital Levels*. (Doctoral). Universiteit van Amsterdam, Amsterdam.
- Veillard, J., Champagne, F., Klazinga, N., Kazandjian, V., Arah, O. A., & Guisste, A. L. (2005). A performance assessment framework for hospitals: the WHO regional office for Europe PATH Project. *International Journal for Quality in Health Care*, 17(6), 9.
- Veras, C. M. T., & Martins, M. S. (1994). A confiabilidade dos dados nos formulários de autorização de internação hospitalar (AIH). *Cad. Saúde Pública*, 10, 18.
- Viacava, F., Almeida, C., Caetano, R., Fausto, M., Macinko, J., Martins, M., . . . Szwarcwald, C. L. (2004). Uma metodologia de avaliação do desempenho do sistema de saúde brasileiro. *Ciência & Saúde Coletiva*, 9(3), 14.

- Viacava, F., Ugá, M. A. D., Porto, S., Laguardia, J., & Moreira, R. d. S. (2012). Avaliação de Desempenho de Sistemas de Saúde: um modelo de análise. *Ciência & Saúde Coletiva*, 17(4), 23.
- Vieira-Da-Silva, L. M. (2005). Conceitos, abordagens e estratégias para a avaliação em saúde. In *Avaliação em saúde* (pp. 15-39). Salvador: EDUFBA.
- Vinagre, M. H. (2008). *A satisfação de Pacientes de serviços públicos de saúde: preditores e mediadores*. (Doctoral). Instituto Superior de Ciências do Trabalho e da Empresa- ISCTE, Portugal: Lisboa.
- Vitoria-Gasteiz. (2000). *Guide for Self-assessment of Healthcare Organizations (simplified version)* (O.-S. V. d. Salud Ed.).
- Volpato, L. F. (2014). *Utilização de uma ferramenta para análise da qualidade nos serviços públicos de saúde*. (Doctoral). Universidade estadual de campinas, Piracicaba
- Vuori, H. (1991). A qualidade da saúde. *Divulgação em saúde para debate*, 1, 9.
- Vuori, H. V. (1982). Quality Assurance of Health Services: Concepts and Methodology. In *Public Health in Europe* (Vol. 16, pp. 127). Copenhagen: World Health Organization, Regional Office for Europe.
- Walburg, J., Klazinga, N., & Nabitz, U. (2000). The EFQM excellence model: European and Dutch experiences with the EFQM approach in health care. *International Journal for Quality in Health Care*, 12(3), 11.
- Wheaton, B. (1987). Assessment of fit in overidentified models with latent variables. *Sociological Methods & Research*, 16(1), 118-154.
doi:10.1177/0049124187016001005
- WHO. (2000a). *The World Health Report 2000. Health Systems: Improving Performance*. Geneva: World Health Organization
- WHO. (2000b). *The World Health Report: Health System: Improving Performance*. Geneva: World Health Organisation
- WHO. (2003). *Health Systems Performance Assessment: Debates, Methods and Empiricism*. Geneva: World Health Organisation
- WHO (Ed.) (1981). *Health programme evaluation*. Geneva: World Health Organization.
- WHO, W. H. O. (2006). *Bridging the "Know-Do" Gap: Meeting on Knowledge Translation in Global Health*. Retrieved from Geneva:
- Williams, S. (2001). Increasing employees' creativity by training their managers. *Industrial and Commercial Training*, 33(2), 6.
- Wolfson, M., & Alvarez, R. (2002). Towards integrated and coherent health information systems for performance monitoring: The Canadian experience. In *Measuring UP: Improving Health System Performance in OECD Countries* (pp. 133-135). Paris: OECD Publications Service.
- Worthen, B. R., Sanders, J. R., & Fitzpatrick, J. L. (2004). *Avaliação de programas: Conceção e práticas* (G. Edusp Ed.). São Paulo: Edusp, Gente.
- Wright, J. T. C., & Giovinazzo, R. A. (2000). Delphi – Uma ferramenta de apoio ao planejamento prospectivo. *Cadernos de Pesquisas em Administração*, 1(12).
- Yeomans, K. A., & Golder, P. A. (1982). The Guttman-Kaiser Criterion as a Predictor of the Number of Common Factors. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 31(3), 221-229.
- Yi, Y. (1990). A Critical Review of Consumer Satisfaction. In A. M. Association (Ed.), *Review of Marketing* (pp. 68-123). Chicago: American Marketing Association.
- Yin, R. K. (2001). *Estudo de caso: planejamento e métodos* (Artmed Ed.). Porto Alegre: Artmed.
- Yin, R. K. (2009). *Case study research: design and methods* (S. Publications Ed.). Thousand Oaks: Sage Publications.

Zeithaml, V. A., Berry, L. L., & Parasuraman, A. (1993). The nature and determinants of customer expectations of service. *Academy of Marketing Science. Journal*, 21(1), 1.

