

Fonseca, Jaisson Gustavo da.
F676a Aplicação do Paediatric Index of Mortality 2 em Unidade de Terapia Intensiva Pediátrica no Brasil [manuscrito]. / Jaisson Gustavo da Fonseca. - - Belo Horizonte: 2012.
98f.: il.
Orientador: Alexandre Rodrigues Ferreira.
Área de concentração: Saúde da Criança e do Adolescente.
Dissertação (mestrado): Universidade Federal de Minas Gerais, Faculdade de Medicina.

1. Terapia Intensiva. 2. Avaliação de Resultados (Cuidados de Saúde). 3. Índice de Gravidade de Doença. 4. Mortalidade. 5. Unidades de Terapia Intensiva Pediátrica. 6. Dissertações Acadêmicas. I. Ferreira, Alexandre Rodrigues. II. Universidade Federal de Minas Gerais, Faculdade de Medicina. III. Título.

NLM: WS 366

UNIVERSIDADE FEDERAL DE MINAS GERAIS
FACULDADE DE MEDICINA

JAISSON GUSTAVO DA FONSECA

APLICAÇÃO DO PEADIATRIC INDEX OF MORTALITY 2 EM UNIDADE DE
TERAPIA INTENSIVA PEDIÁTRICA NO BRASIL

BELO HORIZONTE

2012

JAISSON GUSTAVO DA FONSECA

APLICAÇÃO DO PAEDIATRIC INDEX OF MORTALITY 2 EM UNIDADE DE
TERAPIA INTENSIVA PEDIÁTRICA NO BRASIL

Dissertação apresentada ao Programa de Pós-Graduação em Ciências da Saúde da Faculdade de Medicina da Universidade Federal de Minas Gerais como requisito parcial para a obtenção de Título de Mestre em Ciências da Saúde.

Área de concentração: Saúde da Criança e do Adolescente.

Orientador: Professor Alexandre Rodrigues Ferreira

BELO HORIZONTE

2012

JAISSON GUSTAVO DA FONSECA

APLICAÇÃO DO PAEDIATRIC INDEX OF MORTALITY 2 EM UNIDADE DE
TERAPIA INTENSIVA PEDIÁTRICA NO BRASIL

Dissertação apresentada ao Programa de Pós-Graduação em Ciências da Saúde da Faculdade de Medicina da Universidade Federal de Minas Gerais como requisito parcial para a obtenção de Título de Mestre em Ciências da Saúde.

Área de concentração: Saúde da Criança e do Adolescente

Data da aprovação:

Banca examinadora:

Henrique de Assis Fonseca Tonelli
Hospital da Clínicas - UFMG

Vandack Alencar Nobre Júnior
Faculdade de Medicina - UFMG

Alexandre Rodrigues Ferreira
Orientador

UNIVERSIDADE FEDERAL DE MINAS GERAIS

Reitor: Prof. Clélio Campolina Diniz

Vice-Reitora: Profa. Rocksane de Carvalho Norton

Pró-Reitor de Pós-Graduação: Prof. Ricardo Santiago Gomez

Pró-Reitor de Pesquisa: Prof. Renato de Lima Santos

FACULDADE DE MEDICINA

Diretor: Prof. Francisco José Penna

Vice-Diretor: Prof. Tarcizo Afonso Nunes

Coordenador do Centro de Pós-Graduação: Prof. Manoel Otávio da Costa Rocha

Subcoordenadora do Centro de Pós-Graduação: Profa. Teresa Cristina de Abreu Ferrari

Chefe do Departamento de Pediatria: Profa. Benigna Maria de Oliveira

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS DA SAÚDE

Área de Concentração em Saúde da Criança e do Adolescente

Coordenadora: Profa. Ana Cristina Simões e Silva

Colegiado:

Prof^a. Ana Cristina Simões e Silva

Prof. Cássio da Cunha Ibiapina

Prof. Eduardo Araújo de Oliveira

Prof. Francisco José Penna

Prof. Jorge Andrade Pinto

Prof^a. Ivani Novato Silva

Prof. Marcos José Burle de Aguiar

Prof^a. Maria Cândida Ferrarez Bouzada Viana

Maria de Lourdes Melo Baeta (Repr. Discente)

LISTA DE TABELAS

Revisão da Literatura

Tabela 1 - Níveis de discriminação baseado na área sob a curva ROC.

Tabela 2 - Variáveis utilizadas para o cálculo do PRISM III.

Tabela 3 - Comorbidades específicas e motivações para admissão utilizadas para o cálculo do PRISM III.

Tabela 4 - Variáveis utilizadas para o cálculo do PIM2.

Tabela 5 - Diagnósticos de alto e baixo risco utilizados para o cálculo do PIM2.

Tabela 6 - Comparação entre as mortes observadas e previstas no estudo de Thukral e col.

Tabela 7 - Comparação entre mortes observadas e previstas em 10 faixas de risco crescentes no estudo de Canonero e col.

Tabela 8 - Estudos comparando o desempenho PIM2 versus PRISM III.

Tabela 9 - Estudos avaliando o desempenho do PIM2 em cenários diferentes.

Tabela 10 - Sistemas orgânicos e variáveis utilizadas no PELOD.

Tabela 11 - As quatro gerações de modelos prognósticos em UTI na população adulta.

Tabela 12 - As três gerações de modelos prognósticos em UTI na população pediátrica.

Tabela 13 - Potenciais usos dos escores de acuidade.

Tabela 14 - Mortalidade e taxa de mortalidade padronizada utilizando o PRISM III e o PIM2 da população admitida em mais 90 UTIP dos Estados Unidos no ano de 2010.

Artigo original

Tabela 1 - Distribuição dos sistemas orgânicos relacionados às causas das admissões clínicas.

Tabela 2 - Distribuição dos tipos de condição crônica complexa na população.

Tabela 3 - Mortalidade, PIM2 médio, TMP e média e mediana do tempo de permanência nos subgrupos.

Tabela 4 - Calibração do escore entre as diferentes faixas de risco.

Tabela 5 - Discriminação e calibração nos subgrupos de pacientes.

Tabela 6 – Discriminação e calibração em diferentes faixas etárias.

LISTA DE GRÁFICOS

Revisão da Literatura

Gráfico 1 – Taxa de mortalidade padronizada de uma hipotética UTIP através do tempo.

Gráfico 2 – Figura em forma de sino da relação entre a gravidade da doença e o potencial para responder a uma intervenção.

Artigo original

Gráfico 1 – Curva ROC de toda a população do estudo.

LISTA DE ABREVIATURAS

ANVISA	Agência Nacional de Vigilância Sanitária
APACHE	Acute Physiology and Chronic Health Evaluation
APACHE II	Acute Physiology and Chronic Health Evaluation II
APACHE III	Acute Physiology and Chronic Health Evaluation III
BE	Basic Excess
BiPAP	Bi Level Positive Airway Pressure
COEP	Comitê de Ética em Pesquisa
CPAP	Continuos Positive Airway Pressure
exp	exponencial
FiO ₂	Fração inspirada de oxigênio
FPTC	Fora de Possibilidade Terapêutica Curativa
HC	Hospital das Clínicas
H1N1	Subtipo de Influenza A
IC	Intervalo de Confiança
INR	International Normalized Ratio
LOD	Logistic Organ Dysfunction
mm	milímetro
mmHg	milímetro de Mercúrio
mmol/L	milimol por litro
MPM	Mortality Probability Model
MPM III	Mortality Probability Model III
PaCO ₂	Pressão arterial de dióxido de carbono
PaO ₂	Pressão arterial de oxigênio
PAS	Pressão Arterial Sistólica
PCR	Parada cardiorrespiratória
PICUEs	Pediatric Intensive Care Unit Evaluations
PIM	Paediatric Index of Mortality
PIM2	Paediatric Index of Mortality 2
PELOD	Pediatric Logistic Organ Dysfunction
P-MODS	Pediatric Multiple Dysfunction Score
PRISM	Pediatric Risk of Mortality

PRISM III	Pediatric Risk of Mortality III
PROWESS	Protein C Worldwide Evaluation in Severe Sepsis
PSI	Physiologic Stability Index
ROC	Receiver Operator Characteristic
SAPS	Simplified Acute Physiology Score
SAPS II	Simplified Acute Physiology Score II
SAPS III	Simplified Acute Physiology Score III
SDMO	Síndrome de Disfunção de Múltiplos Órgãos
SIDA	Síndrome de Imunodeficiência Adquirida
SMR	Standardized Mortality Ratio
SOFA	Sequential Organ Failure Assessment
TGO	Transaminase Glutâmico Oxalacética
TGP	Transaminase Glutâmico Piruvica
TISS	Therapeutic Intervention Severity Score
TP	Tempo de Protrombina
TTPA	Tempo de Tromboplastina Parcial Ativada
UCI	Unidade de Cuidado Intermediário
UFMG	Universidade Federal de Minas Gerais
UTI	Unidade de Terapia Intensiva
UTIP	Unidade de Terapia Intensiva Pediátrica
VM	Ventilação Mecânica
VPS	Virtual Pediatric Intensive Care Systems

Sistemas de Escores de Acuidade utilizados na Terapia Intensiva Pediátrica.

Resumo

Introdução: os Sistemas de Escores de Acuidade utilizados na Terapia Intensiva Pediátrica são ferramentas que quantificam a gravidade da doença e fornecem uma estimativa do desfecho de interesse de pacientes internados em unidades de terapia intensiva (UTI).

Objetivo: descrever o processo de desenvolvimento e validação, os tipos utilizados na terapia intensiva pediátrica, os principais usos e as limitações dos escores de acuidade. Com ênfase no escore de predição de desfecho Paediatric Index of Mortality 2 (PIM2) pelo fato de ter sido utilizado no artigo original desta dissertação. **Metodologia:** revisão não sistemática da literatura sobre os escores de acuidade utilizados na terapia intensiva pediátrica. A estratégia de busca e seleção de artigos utilizou o banco de dados Pubmed e as palavras-chave child, pediatric, critical care, intensive care, mortality, prediction, risk, score, severity of illness index. O período de busca ficou limitado aos últimos 10 anos. **Resultados:** o processo de desenvolvimento e validação dos sistemas de escore de acuidade é padronizado; a terapia intensiva pediátrica possui escores de predição de desfecho e de disfunção de órgãos validados; existem inúmeras propostas de uso para os escores de acuidade, com destaque para a utilização como ferramenta de comparação de desempenhos entre UTI (benchmarking) e para comparação de grupos durante a randomização em pesquisa científica; todos os escores de acuidade apresentam limitações que podem influenciar a análise dos resultados encontrados. **Conclusões:** os sistemas de escores de acuidade tem ganhado espaço na terapia intensiva pediátrica em decorrência das informações produzidas pelas suas estimativas que podem auxiliar no cuidado do paciente criticamente enfermo. Porém é importante reconhecer as limitações inerentes a todos os escores, com o intuito de poder analisar e utilizar as valiosas informações fornecidas por esses modelos de forma correta.

Palavras-chave: child, pediatric, critical care, intensive care, mortality, prediction, risk, score, severity of illness index.

Severity Scoring Systems used in the Pediatric Intensive Care.

Abstract

Introduction: severity Scoring Systems used in the Pediatric Intensive Care are tools that quantify the severity of the disease and provide an estimate of the outcome of interest from patients in intensive care units (ICU). **Objective:** to describe the development and validation process, the types used in pediatric intensive care, the main uses and limitations of acuity scores. With an emphasis on outcome prediction score Paediatric Index of Mortality 2 (PIM2) as results of that has been used in the original article of this thesis. **Methodology:** no systematic review of literature of the accuracy scores used in the pediatric intensive care. The search strategy and selection of articles used the database Pubmed and the keywords child, pediatric, critical care, intensive care, mortality, prediction, risk, score, severity of illness index. The search was limited to the last 10 years. **Results:** the process of developing and validating the accuracy of scoring systems is standardized, the pediatric intensive care has scores to predict outcome and organ dysfunction validated, there are numerous proposals for the use of scores of accuracy, especially for use as a tool for comparison of performance between ICU (benchmarking) and comparison groups during randomization in scientific research, all acuity scores have limitations that may influence the analysis of results. **Conclusions:** the severity scoring systems has gained importance in the pediatric intensive care as a result of information produced by their estimates that can assist in the care of critically ill patients. But it is important to recognize the limitations inherent in all the scores, in order to be able to analyze and use the valuable information provided by these models correctly.

Keywords: child, pediatric, critical care, intensive care, mortality, prediction, risk, score, severity of illness index.

Avaliação do Paediatric Index of Mortality 2 em Unidade de Terapia Intensiva Pediátrica no Brasil.

Resumo

Introdução: o Paediatric Index of Mortality 2 (PIM2) é um dos principais escores de predição de desfecho utilizados na terapia intensiva pediátrica. Diferenças das misturas de casos quando comparada com a população do desenvolvimento do escore podem resultar num desempenho inadequado do modelo quando aplicado em um novo cenário. Sendo recomendada a avaliação do modelo em novos cenários para verificar a persistência do desempenho do escore. **Objetivo:** avaliar o desempenho do escore PIM2 em uma Unidade de Terapia Intensiva Pediátrica (UTIP) no Brasil. **Metodologia:** estudo de coorte com 677 pacientes admitidos na UTIP entre 01 de Fevereiro de 2009 e 31 de Janeiro de 2011, com a avaliação do desempenho do escore através da discriminação e calibração. **Resultados:** as características mais marcantes da população do estudo foram o elevado número de admissões após procedimentos cirúrgicos e a alta prevalência de pacientes com condição crônica complexa e neoplasia. O PIM2 apresentou discriminação adequada indicada pela área sob a curva Receiver Operator Characteristic (ROC) de 0,84 (IC95% 0,810-0867) e calibração inadequada indicada pelo teste de ajuste de Hosmer-Lemeshow que evidenciou um qui-quadrado de 43,4056 ($p < 0,0001$). **Conclusão:** o PIM2 apresentou desempenho inadequado na amostra do estudo em decorrência da pobre calibração. A diferença da mistura de casos pode ser um dos fatores responsáveis pela calibração ruim, porém a influência da qualidade do cuidado ofertado não pode ser descartada.

Palavras-chave: mortality, intensive care, outcome assessment, Paediatric Index of Mortality 2, severity of illness index.

Implementation of the Paediatric Index of Mortality 2 in the Pediatric Intensive Care Unit in Brazil.

Abstract

Introduction: Paediatric Index of Mortality 2 (PIM2) is a major predictor of outcome scores used in the pediatric intensive care. Differences mixtures of cases compared with the development of the population can result in a performance score inappropriate model when applied in a new environment. It is recommended to evaluate the model in new scenarios to verify the persistence of performance scores. **Objective:** to evaluate the performance of PIM2 score in a Pediatric Intensive Care Unit (PICU) in Brazil. **Methodology:** a cohort study of 677 patients admitted to the pediatric ICU between February 1, 2009 and January 31, 2011, evaluating the performance of the score through discrimination and calibration. **Results:** the main characteristics of the study population were the high number of admissions after surgical procedures and the high prevalence of patients with complex chronic conditions and cancer. PIM2 showed adequate discrimination indicated by the area under the Receiver Operating Characteristic (ROC) of 0.84 (95%CI 0.810 to 0.867) and inadequate calibration indicated by the Hosmer-Lemeshow test showed that a chi-square 43,4056 ($p < 0.0001$). **Conclusion:** PIM2 showed poor performance in the study sample due to poor calibration. The mixture of cases can be one of the factors responsible for poor calibration, but the influence of the quality of care offered cannot be ruled out.

Keywords: mortality, intensive care, outcome assessment, Paediatric Index of Mortality 2, severity of illness index.

SUMÁRIO

1) Introdução da dissertação.....	1
2) Objetivos da dissertação.....	2
3) Revisão da literatura - Sistemas de Escores de Acuidade utilizados na Terapia Intensiva Pediátrica.	
3.1) Introdução.....	3
3.2) Objetivos.....	4
3.3) Metodologia.....	4
3.4) O que são os sistemas de escore de acuidade?.....	5
3.5) Desenvolvimento dos escores de acuidade.....	5
3.6) Avaliação dos escores de acuidade.....	8
3.7) Validação dos escores de acuidade.....	10
3.8) Tipos de escores de acuidade.....	12
3.9) Escores de predição de desfecho.....	13
3.10) Escores de disfunção de órgãos.....	29
3.11) Atualizações.....	31
3.12) Utilizações dos escores de acuidade.....	34
3.13) Limitações dos escores de acuidade.....	49
3.14) Conclusão.....	51
3.15) Referências Bibliográficas.....	52
4) Metodologia do artigo original	
4.1) Delineamento.....	62
4.2) População.....	62
4.3) Cenário.....	62
4.4) Coleta de dados.....	63
4.5) Variáveis estudadas.....	63
4.6) Calculo do PIM2.....	69
4.7) Análise estatística.....	70
4.8) Considerações éticas.....	71

5) Artigo original - Aplicação do Paediatric Index of Mortality 2 em Unidade de Terapia Intensiva Pediátrica no Brasil.

5.1) Introdução.....	72
5.2) Objetivos.....	73
5.3) Metodologia.....	73
5.4) Resultados.....	75
5.5) Discussão.....	82
5.6) Conclusão.....	91
5.7) Referências bibliográficas.....	91

6) Anexos

6.1) Anexo A – Parecer do comitê de ética.....	97
6.2) Anexo B – Termo de consentimento livre e esclarecido.....	98
6.3) Anexo C – Cópia da Ata da Defesa e da Declaração de Aprovação.....	103

1) Introdução da dissertação

A qualidade inadequada e o custo crescente dos cuidados de saúde oferecidos têm sido focos de interesse para a comunidade médica. A unidade de terapia intensiva (UTI) é um dos cenários onde essas questões têm sido mais debatidas em decorrência do impacto significativo da qualidade do cuidado nos resultados do paciente (morbidade, resultado funcional e mortalidade) e dos altos custos gerados pelo cuidado ao paciente criticamente doente.

Há um aumento do uso de indicadores de qualidade como instrumento de melhoria dos serviços de saúde ofertados. A taxa de mortalidade padronizada (TMP) é um dos principais indicadores utilizados na terapia intensiva, sendo calculada dividindo a taxa de mortalidade observada pela taxa de mortalidade prevista por um escore de predição de desfecho. Na terapia intensiva pediátrica os principais escores de predição de desfecho utilizados são o Pediatric Risk of Mortality (PRISM), o Paediatric Index of Mortality (PIM) e suas respectivas atualizações.

O uso do Paediatric Index of Mortality 2 (PIM2), a última versão do PIM, tem sido incentivado por ser um escore de fácil utilização, eficiente e de domínio público. Desde a publicação do PIM2 em 2003 vários estudos foram publicados avaliando o desempenho do escore em populações e cenários diferentes do estudo de desenvolvimento, com a maioria dos estudos evidenciando desempenho adequado do escore.

No Brasil ainda não foram publicados estudos avaliando o PIM2, sendo necessária a avaliação do desempenho do escore neste cenário antes de seu emprego como ferramenta de melhoria de qualidade dos cuidados de saúde oferecidos pelos serviços de terapia intensiva pediátrica existentes no país.

A presente dissertação segue as orientações do Centro de Pós-Graduação em Ciências da Saúde – Saúde da Criança e do Adolescente da Faculdade de Medicina da Universidade Federal de Minas Gerais e será apresentada em quatro partes.

A primeira parte refere-se aos objetivos da dissertação, a segunda parte é a revisão da literatura e será apresentada no formato convencional conforme o regulamento de formato de teses e dissertações do Programa de Saúde da Criança e do Adolescente. A terceira parte refere-se à metodologia completa utilizada no artigo original. A quarta parte é o artigo original, que será submetido ao periódico Pediatric Critical Care para avaliação de publicação. As normas de publicação deste periódico encontram-se disponíveis no endereço eletrônico: <http://journals.lww.com/pccmjournal/Pages/informationforauthors.aspx>.

2) Objetivos da dissertação

2.1) Objetivos do artigo de revisão

Descrever o processo de desenvolvimento e validação, os tipos utilizados na terapia intensiva pediátrica, os principais usos e as limitações dos escores de acuidade. Com ênfase no escore de predição de desfecho Paediatric Index of Mortality 2 (PIM2) em decorrência de ter sido utilizado no artigo original desta dissertação.

2.2) Objetivos do artigo original

2.2.1) Objetivo geral

Avaliar o desempenho do escore de predição de desfecho PIM2 na Unidade de Terapia Intensiva Pediátrica (UTIP) do Hospital das Clínicas da Universidade Federal de Minas Gerais (HC-UFGM).

2.2.2) Objetivos específicos

- A- Caracterizar a população atendida na UTIP do HC-UFGM.
- B- Avaliar a discriminação e calibração do escore PIM2 na população estudada.
- C- Comparar a mortalidade encontrada na população estudada com a mortalidade prevista pelo escore PIM2.

3) Revisão da literatura - Sistemas de Escores de Acuidade utilizados na Terapia Intensiva Pediátrica.

3.1) Introdução

A qualidade nos cuidados de saúde pode ser definida como o grau em que os serviços de saúde para indivíduos e populações aumentam a probabilidade do resultado de saúde desejado e são consistentes com o conhecimento profissional atual (1).

Há um aumento do interesse e preocupação com a qualidade inadequada e o custo crescente dos cuidados de saúde, tornando necessária a implementação de estratégias para a melhoria dos serviços de saúde (2). Essas questões são particularmente importantes na terapia intensiva em decorrência do impacto significativo da qualidade do cuidado nos resultados para o paciente (morbidade, resultado funcional e mortalidade) e dos altos custos gerados pelo cuidado ao paciente criticamente doente (2).

A mensuração tornou-se aspecto essencial no processo de melhoria da qualidade, com organizações de cuidados de saúde focando esforços na seleção e no uso de indicadores de qualidade (1).

A função do cuidado intensivo pediátrico é melhorar o resultado de crianças que são criticamente doentes ou que requerem cuidados intensivos após procedimentos médicos (3). Este resultado pode ser avaliado como sobrevivência, resultado funcional, ou qualidade de vida (4). A mortalidade é uma medida confiável e simples de resultado em terapia intensiva, porém a análise do dado na forma bruta pode produzir interpretações equivocadas sobre a qualidade dos cuidados de saúde de determinado serviço (5).

Intuitivamente compreendemos que se alguns pacientes são mais doentes do que outros, eles potencialmente terão pior sobrevida e outros desfechos objetivos, além de maior utilização de recursos como tempo de permanência na unidade. Hipoteticamente se perguntássemos a uma pessoa leiga para que unidade de terapia intensiva (UTI) ela levaria o seu filho criticamente doente, uma com mortalidade de 1,2% ao ano ou uma com mortalidade de 9% ao ano, a resposta seria que a criança seria levada para o primeiro serviço.

Entretanto como afirmar que a unidade que tem menor mortalidade, porém atende pacientes com baixa complexidade, fornece melhor cuidado de saúde do que a unidade com maior mortalidade que atende pacientes com alta complexidade. A diferença de mortalidade

pode estar relacionada à mistura (tipos) de casos e a gravidade da doença dos pacientes ao invés de diferenças de cuidados ofertados.

Por essa razão indicadores de qualidade devem isolar os fatores sobre controle da unidade de observação (por exemplo, UTI) dos fatores relacionados aos pacientes (mistura de casos e gravidade da doença) (6). Para esta finalidade geralmente é utilizada uma metodologia chamada de ajuste de risco que quantifica os fatores relacionados ao paciente (7).

O ajuste de risco envolve a contabilização de características dos pacientes que influenciam na qualidade da medida, mas não estão sob o controle do provedor (7). Na medida em que as técnicas de ajuste de risco são capazes de ajustar as diferenças na população estudada, permite que as comparações de qualidade sejam válidas (7).

Na terapia intensiva o ajuste de risco dos pacientes é geralmente produzido com o uso dos sistemas de escores de acuidade (3). Esses escores contabilizam as características dos pacientes atendidos na UTI (variáveis fisiológicas, laboratoriais, clínicas e demográficas) para quantificar seu grau de gravidade, permitindo uma comparação padronizada (8).

3.2) Objetivos

Essa revisão da literatura tem o objetivo descrever o processo de desenvolvimento e validação, os tipos utilizados na terapia intensiva pediátrica, os principais usos e as limitações dos escores de acuidade. Com ênfase no escore de predição de desfecho Paediatric Index of Mortality 2 (PIM2) em decorrência de ter sido utilizado no artigo original desta dissertação.

3.3) Metodologia

Revisão não sistemática da literatura ou revisão narrativa sobre os escores de acuidade utilizados na terapia intensiva pediátrica. A estratégia de busca e seleção de artigos utilizou o banco de dados Pubmed e as palavras-chave child, pediatric, critical care, intensive care, mortality, prediction, risk, score, severity of illness index. O período de busca ficou limitado aos últimos 10 anos. Além disso, foi realizada uma verificação das referências citadas em artigos originais e de revisão sobre o tema.

3.4) O que são os sistemas de escore de acuidade?

Os sistemas de escore de acuidade utilizados na terapia intensiva são ferramentas que quantificam a gravidade da doença e fornecem uma estimativa do desfecho de interesse de pacientes internados em UTI (5). O principal desfecho utilizado/estimado é a mortalidade, porém há escores que utilizam outros desfechos como o tempo de permanência na UTI e tempo de ventilação mecânica (9).

São escores compostos constituídos de um grupo de variáveis, que caracterizam a condição clínica do paciente (10). As variáveis utilizadas na maioria dos escores podem ser agrupadas em cinco categorias: idade (não utilizada em escores para população pediátrica), comorbidades, anormalidades fisiológicas, diagnósticos agudos, e intervenções (5).

Essas variáveis foram associadas a aumento de risco do desfecho de interesse após admissão em UTI nos estudos para produção dos escores, sendo designada para cada variável uma pontuação referente ao seu peso preditivo. O número de pontos atribuídos a cada variável é proporcional a sua capacidade de predizer o desfecho de interesse (10). Por exemplo, se no estudo para produção de um determinado escore a variável que avalia a condição hemodinâmica mostrar uma maior associação com o desfecho do que a variável que avalia a condição respiratória, para esta variável será atribuído um maior número de pontos. Os pontos de todas as variáveis incorporadas em um determinado escore são então computados para obter o risco do desfecho (10).

3.5) Desenvolvimento dos sistemas de escore de acuidade

O objetivo dos sistemas de escore de acuidade é determinar a probabilidade de um específico desfecho (por exemplo, mortalidade) com diferentes combinações de variáveis de predição em uma população bem definida (pacientes admitidos em UTI) (11).

A produção destes modelos é composta de algumas etapas de desenvolvimento: seleção do desenho do estudo, seleção da amostra, definição do número de pacientes requeridos, e seleção das variáveis de predição e de desfecho (11).

A coorte prospectiva é o desenho do estudo que responde de forma mais adequada ao objetivo e é a mais utilizada no desenvolvimento dos sistemas de escore de acuidade, permitindo a medição ideal de variáveis de predição e desfecho (11). A amostra do estudo

deve incluir pacientes em risco de desenvolver o desfecho de interesse (11), sendo no caso dos sistemas de escore de acuidade constituída de pacientes admitidos em UTI.

A característica multivariável da pesquisa de prognóstico torna difícil estimar o tamanho da amostra requerido, não havendo um método direto para isto (11). Afessa e col relatam que para ser generalizável o desenvolvimento do escore requer um grande banco de dados coletados de representativas UTI (9), porém não há conhecimento do que significaria uma amostra grande (5). Alguns estudos têm sugerido que para cada variável de predição candidata exista pelo menos 10 eventos (desfecho), embora em estudos mais recentes é relatado que este número possa ser menor em certas circunstâncias (12,13).

A seleção de variáveis de predição (independentes) apropriadas e a definição de uma variável de desfecho (dependente) clara e relevante estão entre as etapas mais importantes do desenvolvimento dos modelos (2).

As variáveis de predição (independentes) são aquelas variáveis que compõem o escore (2). São características dos pacientes que influenciam na probabilidade de ocorrência do desfecho, como idade, comorbidades, diagnósticos agudos e anormalidades fisiológicas (9). Devendo ser rotineiramente disponíveis, confiáveis e independentes de intervenções realizadas na UTI (9). São selecionadas e ranqueadas (peso relativo) subjetivamente por consenso de especialistas ou objetivamente usando métodos estatísticos (9).

O termo ranqueamento refere-se ao conceito que cada variável preditora tem um peso relativo proporcional a sua capacidade de predizer certo desfecho (10). Como exemplo, em um determinado modelo X de predição de mortalidade, a variável hemodinâmica pode ter uma maior influência para predizer o desfecho (maior poder preditivo) do que a variável respiratória, com isso essa variável terá um maior número de pontos na equação para o cálculo do escore.

Nos modelos mais antigos a seleção e ranqueamento das variáveis predictoras eram realizadas por consenso de especialistas. O Physiologic Stability Index (PSI), que foi o primeiro escore de predição de desfecho utilizado em pediatria, é um exemplo onde as variáveis de predição foram escolhidas e ranqueadas pelos pesquisadores que desenvolveram o modelo (14). Foram selecionadas 34 variáveis de sete sistemas fisiológicos (cardiovascular, respiratório, neurológico, hematológico, renal, gastrointestinal e metabólico) e para cada variável foi atribuída uma pontuação de 1 (anomalia na variável de leve intensidade), 3 (anomalia na variável de média intensidade), e 5 (anomalia na variável de grave intensidade) de acordo com a opinião dos pesquisadores que desenvolveram o modelo.

Nos modelos mais recentes a seleção e o ranqueamento são realizados através de métodos estatísticos. A análise multivariável é o padrão estatístico atual para seleção de variáveis, bem como para a determinação dos pesos relativos das variáveis na equação do modelo (2). A regressão logística multivariada é mais frequentemente usada para desfechos dicotômicos (por exemplo, sobrevivência/morte), já a análise de regressão linear ou multivariável por mínimos quadrados é mais frequentemente usada para variáveis contínuas (por exemplo, tempo de permanência) (2).

O Paediatric Index of Mortality 2 (PIM2), que é escore descritor de desfecho mais recente publicado para a população pediátrica, utilizou no seu desenvolvimento duas etapas para seleção e ranqueamento das variáveis de predição (15). Na primeira etapa foi examinada a associação entre um conjunto de variáveis individuais e mortalidade utilizando o teste do qui-quadrado para variáveis dicotômicas e o teste de Copas seguido pelo teste U de Mann-Whitney para variáveis contínuas. Na segunda etapa as variáveis que demonstraram associação foram testadas através da regressão logística para verificar a persistência da associação e o peso relativo de cada variável selecionada.

O outro componente principal no desenvolvimento do modelo é a definição da variável de desfecho (2). É importante quando no desenvolvimento ou avaliação de um sistema de escore de acuidade que os modelos tenham uma variável de desfecho clinicamente relevante e claramente definida que cumpra determinadas necessidades clínicas, de pesquisa, ou para outros propósitos (por exemplo, qualidade dos cuidados oferecidos) (2).

A principal variável de desfecho utilizada é a mortalidade em curto prazo (na UTI ou hospitalar) (16), tanto para modelos utilizados na população adulta quanto pediátrica, pelo fato que o principal objetivo da terapia intensiva é assegurar sobrevivência com morbidade mínima em pacientes com doenças ameaçadoras da vida e por ser uma variável com definição precisa (5).

Porém outras variáveis de desfecho têm sido incorporadas aos modelos com o intuito de satisfazer a necessidade de conhecimento mais refinado da evolução dos pacientes nas UTI, como o tempo de permanência na UTI utilizado num modelo para população adulta (17). Na população pediátrica Ruttimann e col utilizaram o escore de predição de desfecho Pediatric Risk of Mortality III (PRISM III) para desenvolver um modelo de predição de tempo de permanência na UTI que demonstrou validação adequada, porém de forma independente do escore original (18).

Após a seleção das variáveis de predição com seus pesos relativos já definidos é criada uma equação que será utilizada para o cálculo do escore e probabilidade do desfecho. O PIM2 disponibiliza a equação necessária para seu cálculo, sendo um bom exemplo para observar a contribuição relativa de cada variável (15):

$$\text{PIM2} = [0,01395 \times (A-120)] + (3,0791 \times B) + (0,2888 \times C) + (0,104 \times D) + (1,3352 \times E) - (0,9282 \times F) - (1,0244 \times G) + (0,7507 \times H) + (1,6829 \times I) - (1,5777 \times J) - 4,8841 *$$

$$\text{Probabilidade de morte} = \exp(\text{PIM2}) / [1 + \exp(\text{PIM2})]$$

* A = pressão arterial sistólica; B = reação pupilar a luz; C= relação PaO₂/FiO₂; D= excesso de bases; E= ventilação mecânica; G= admissão no pós-operatório; H= circulação extracorpórea; I= paciente com diagnóstico de alto risco; J= paciente com diagnóstico de baixo risco.

Completado as etapas de desenvolvimento de um sistema de escore de acuidade, o seu desempenho deve ser avaliado e validado (5).

3.6) Avaliação dos sistemas de escore de acuidade

Os sistemas de escore de acuidade devem ser submetidos ao mesmo exame minucioso que drogas e tecnologias antes de serem utilizados em decisões que impactem no serviço de saúde ofertado e no cuidado individual de pacientes (9).

Os modelos devem diferenciar entre sobreviventes e não sobreviventes, serem precisos em todas as faixas de risco e confiáveis (fornecerem estimativas idênticas e reproduzíveis para um paciente independente do observador) (19). Também devem ser dinâmicos, refletindo a mudança de tratamentos e mistura de casos através do tempo (9).

O desempenho dos escores de acuidade é geralmente avaliado através da discriminação e calibração (2). Discriminação é a capacidade do modelo para distinguir entre grupos de desfechos (2), ou seja, é uma medida da eficácia do modelo em discriminar entre aqueles que experimentaram o desfecho designado (por exemplo, morte) daqueles que não experimentaram o desfecho (por exemplo, alta da UTI) (5). A discriminação é comumente avaliada pela área sobre a curva receiver operator characteristic (ROC) (20). Sendo a curva

ROC construída a partir da sensibilidade (predição de morte) e da especificidade (predição de alta) de cada paciente da amostra estudada (5).

Se um modelo designa o desfecho ao acaso (randomicamente) a área sob a curva ROC será de 0,5, enquanto um modelo que designa o desfecho de forma perfeita (100% de sensibilidade e taxa de erro falso positivo de 0%) a área sob a curva ROC será de 1 (tabela1) (9). Esta medida deve permanecer alta ($>0,8$) quando o desempenho do modelo é avaliado em um conjunto de dados independentes (2), porém existem autores sugerindo que uma área sob a curva ROC de 0,7 pode ser considerada adequada para avaliação de desempenho (21).

Tabela 1 - Níveis de discriminação baseado na área sob a curva ROC

Área sob a curva ROC	Nível de discriminação
1,00	Perfeito
0,90-0,99	Excelente
0,80-0,89	Muito bom
0,70-0,79	Bom
0,60-0,69	Moderado
$<0,60$	Ruim

Calibração refere-se à correlação entre o desfecho previsto e atual para uma completa faixa de risco (9), ou seja, é o grau de correspondência em diferentes níveis de probabilidade entre a probabilidade de um desfecho (por exemplo, morte) como previsto pelo score e a frequência observada do desfecho (10).

A análise estatística da calibração é usualmente realizada com o teste de ajuste de Hosmer-Lemeshow (22). A questão estatística é se as discrepâncias entre o desfecho observado e esperado são estatisticamente significantes. Um valor de p para calibração menor do que 0,05 sugere que há uma diferença estatisticamente significativa entre o desfecho previsto e o observado, que é o oposto do que é desejado (10).

Há diferenças nos valores de p relatados pelos autores para que a calibração seja considerada adequada. Tibby e col relatam que é tradicionalmente considerado que a calibração é adequada se o valor de p é maior que 0,10 (23). Já Keegan e col (16) e Afessa e col (9) relatam que a calibração é considerada boa se o valor do p do teste Hosmer-Lemeshow é maior que 0,05. Um valor de p maior é provavelmente ainda melhor (10).

Normalmente a calibração é realizada pela categorização dos pacientes em 10 grupos, de acordo com a probabilidade de ocorrência do desfecho prevista pelo método (9). Porém

existem estudos que utilizam uma menor quantidade de grupos, geralmente 5 (24). Esses grupos podem utilizar intervalos da probabilidade de desfecho iguais (por exemplo, 0-10%, 10-20%, etc), sendo chamado de teste de Hosmer-Lemeshow com estatística H; ou podem utilizar intervalos de probabilidades diferentes (por exemplo, 0-1%; 1-5%; etc), sendo chamado de teste de Hosmer-Lemeshow com estatística C (22).

Após a categorização é realizada uma comparação entre os desfechos (óbitos e sobreviventes) esperados e observados pelo teste do qui-quadrado (25). Podendo ser calculado como $\sum (O - E)^2/E$, no qual O é o observado e E é o esperado número de sobreviventes e mortos em cada intervalo de risco (26). A partir do valor do qui-quadrado é calculado o valor de significância (22).

O teste de Hosmer-Lemeshow pode ser afetado pelo tamanho da amostra (27), com o valor do p sendo menor numa amostra muito grande e maior em pequenas amostras, levando a uma inapropriada estimativa da calibração (28). Com isso Shann sugere colocar mais peso na inspeção da tabela do teste de Hosmer-Lemeshow porque uma diferença pequena e sem importância clínica numa amostra grande irá ter o mesmo valor de p que uma diferença grande e clinicamente importante em uma pequena amostra (29).

O valor de p também pode ser afetado pela disposição dos sujeitos na categorização dos grupos, como demonstrado por Bertolini e col que obtiveram aproximadamente um milhão de diferentes valores de p quando usaram todas as possíveis disposições dos sujeitos em uma amostra de quase 1400 pacientes (30).

Quando avaliada em um conjunto de dados independentes, a calibração inadequada pode ser resultado de um modelo inadequado, porém também pode ser um indicador de diferença de qualidade de cuidado entre o conjunto de dados independentes e o conjunto de dados de referência utilizado no desenvolvimento do modelo (2).

Com isso apesar da extensa literatura sobre a metodologia utilizada para avaliação e validação dos sistemas de escores, permanecem várias questões relacionadas ao desenvolvimento e análise dos modelos que podem afetar a confiabilidade das conclusões (31).

3.7) Validação dos sistemas de escore de acuidade

São geralmente utilizadas três estratégias para validar um escore de acuidade, com níveis de rigorosidade crescente (32):

- a) Validação interna – uma abordagem comum é dividir o conjunto de dados de forma randômica (frequentemente 2:1), desenvolvendo o modelo utilizando a primeira parte (chamada de conjunto de treinamento), e usando a segunda parte para avaliar a acurácia preditiva. Esta abordagem irá tender a dar resultados otimistas em decorrência da similaridade dos dois conjuntos de dados.
- b) Validação temporal – uma alternativa é avaliar o desempenho do modelo em pacientes subsequentes do mesmo centro. Não é diferente em princípio da divisão de um único conjunto de dados pelo tempo, com isso ocorrerão clara similaridades entre os dois conjuntos de pacientes. Entretanto é uma avaliação prospectiva do modelo, com dados independentes dos dados originais usados no processo de desenvolvimento.
- c) Validação externa – examina a generalização do modelo, para isso é necessário o uso de novos dados coletados de uma apropriada população de pacientes de um diferente centro. Questões fundamentais de desenho da validação externa, como seleção da amostra e tamanho tem recebido limitada atenção, não havendo consenso sobre o tema.

Em decorrência da diversidade de estratégias de validação dos modelos há uma discussão sobre qual seria mais adequada. Altman e col relatam que ao considerar a adequabilidade de um modelo prognóstico não é suficiente demonstrar que ele prediz com sucesso desfecho nos dados do desenvolvimento inicial, é necessário evidenciar que o modelo funciona bem para outros grupos de pacientes (33). Afessa e col relatam que se os dados da validação originam-se da mesma população dos dados do desenvolvimento do modelo, os resultados podem não ser reproduzíveis em outras populações (9). Le Gall e col relatam que a validação deve ser realizada numa população diferente da que foi utilizada para desenvolver o escore (21).

Marcin e col relatam que a validação do escore deve ser feita preferencialmente pelos desenvolvedores usando dados que não foram utilizados no desenvolvimento do modelo (2). Bouch e col também relatam que a validação deve ser realizada numa população diferente da que foi utilizada para desenvolver o escore, porém consideram que a divisão randômica da

amostra original utilizada para o desenvolvimento do modelo é uma forma de alcançar este objetivo (5).

Já Pollack relata que a avaliação do desempenho de um escore numa amostra independente (seja uma UTI, um grupo de UTI, ou um país) não é uma “validação” do método (34). Pelo contrário, é uma comparação da relação entre as variáveis de predição e o desfecho em duas amostras. O desempenho será bom quando a relação das variáveis de predição com o resultado nas duas amostras é semelhante, e será ruim quando a relação não for semelhante.

A maioria dos sistemas de escores publicados, tanto para a população adulta quanto pediátrica, utiliza a validação interna como metodologia para avaliar o desempenho. O PRISM III e o PIM2, que são os principais modelos prognósticos utilizados em pediatria também utilizaram a validação interna (15,35).

3.8) Tipos de sistemas de escores de acuidade da doença

Os sistemas de escores de acuidade da doença utilizados para avaliação de pacientes internados em UTI podem ser divididos em dois grupos: escores que avaliam a severidade da doença na admissão e utilizam-na para prever desfecho; e escores que avaliam a presença e a severidade da disfunção de órgãos (36). Essa revisão da literatura tem como foco os escores que avaliam a severidade da doença na admissão, porém para facilitar a discussão do tema foi considerado necessário realizar uma introdução sobre os escores que avaliam a severidade da disfunção de órgãos.

A nomenclatura utilizada na literatura para identificar estes dois grupos é muito variável. Os escores que avaliam a severidade da doença na admissão podem ser identificados como escores de predição de desfecho (36), modelos de severidade da doença (9), modelos prognósticos (16), escores de predição de mortalidade (2) ou escores prognósticos (10). Já os escores que avaliam a severidade da disfunção de órgãos podem ser identificados como escores de disfunção de órgãos (2,36), modelos de disfunção de órgãos (9) ou escores de desfecho (10).

Nesta dissertação serão utilizados os termos escores de predição de desfecho e escores de disfunção de órgãos para definir os escores que avaliam a severidade da doença na admissão e escores que avaliam a severidade da disfunção de órgãos, respectivamente.

3.9) Escores de predição de desfecho

Os escores de predição de desfecho avaliam a severidade da doença na admissão e utilizam-na para prever desfecho (36). Para avaliar a severidade da doença os escores consideram algumas comorbidades e distúrbios fisiológicos apresentados pelo paciente na admissão na UTI, geralmente durante as primeiras 24 horas de internação (10). Sendo a mortalidade o principal desfecho utilizado, porém há escores que também predizem outros desfechos como tempo de permanência na UTI (9).

Na população pediátrica os principais escores de predição de desfecho atualmente utilizados são o Pediatric Risk of Mortality (PRISM) (37), o Paediatric Index of Mortality (PIM) (38), e suas atualizações respectivamente (15,35).

3.9.1) Pediatric Risk of Mortality

O Pediatric Risk of Mortality (PRISM) foi o primeiro sistema de escore de predição de desfecho desenvolvido para a população pediátrica, estando atualmente na sua terceira geração.

A primeira versão do PRISM foi publicada em 1984 e tinha o nome de Physiologic Stability Index (PSI) (14). Foi desenvolvido a partir de dados 423 admissões consecutivas de uma única UTIP multidisciplinar. Era composto de 34 variáveis de sete sistemas fisiológicos (cardiovascular, respiratório, neurológico, hematológico, renal, gastrointestinal e metabólico). Apresentava um bom poder preditivo e discriminatório, porém não era um instrumento prático, sendo muito trabalhoso de ser executado em função do excessivo número de variáveis fisiológicas a serem utilizadas (38).

Em 1988 Pollack e col publicaram a segunda geração do modelo, o escore PRISM (37), que também é chamado PRISM II por alguns intensivistas (10). Foi desenvolvido a partir do PSI para reduzir o número de variáveis fisiológicas necessárias para avaliação de risco de mortalidade e para obter um objetivo peso (ponderação objetiva) das variáveis restantes (37). O escore foi produzido através da análise de dados de 1415 pacientes de quatro unidades de terapia intensiva pediátricas (UTIP), e testado numa amostra diferente de 1227 pacientes de seis UTIP. Era composto de 14 variáveis fisiológicas rotineiramente medidas com 23 intervalos de variação. Apresentava discriminação (área sob a curva ROC= 0,92) e calibração (Hosmer-Lemeshow $X^2 = 0,80$ $p = > 0,95$) adequadas.

A versão atual do modelo, denominada PRISM III, foi publicada em 1996 (35). Foi desenvolvida a partir de dados de 11.165 pacientes de 32 UTIP dos Estados Unidos da América. A coleta de dados ocorreu em dois períodos, entre os anos de 1989 e 1992 e entre os anos de 1992 e 1994.

O escore é composto de 17 variáveis, sendo divididas em quatro grupos: sinais vitais cardiovasculares/neurológicos, gases sanguíneos e ácido-base, testes químicos e testes hematológicos (tabela 2). Também são coletados dados referentes às específicas comorbidades e motivação da admissão (tabela 3).

Tabela 2 - Variáveis utilizadas para o cálculo do PRISM III

Sinais vitais	Gases sanguíneos e ácido-base	Testes químicos	Testes hematológicos
Pressão arterial sistólica	Acidose	Glicose	Glóbulos brancos
Temperatura	pH	Potássio	Plaquetas
Estado de consciência	pCO ₂	Creatinina	TP ou TTPa
Frequência cardíaca	CO ₂ total	Ureia	
Reflexo pupilar	pO ₂		

Tabela 3 - Comorbidades específicas e motivações para admissão utilizadas para o cálculo do PRISM III

Doença cardiovascular não cirúrgica
Anomalia cromossomal
Câncer
Admissão prévia em UTIP
Parada cardiorrespiratória pré-admissão na UTIP
Pós-operatório
Diabetes aguda (por exemplo, cetoacidose diabética)
Admissão de pacientes já internados em outros setores

A coleta das variáveis dos pacientes pode ser realizada durante as primeiras 12 horas após a admissão (PRISM III-12) ou durante as primeiras 24 horas após a admissão (PRISM III-24). Os valores mais anormais de cada variável são coletados para o cálculo do escore. A equação necessária para estimar a mortalidade prevista não é de domínio público por ser um modelo patenteado, com isso os usuários têm de pagar para obter essa equação.

Os autores relatam que o uso do PRISM III-12 é interessante para avaliações de qualidade uma vez que, ao diminuir o tempo de coleta de dados, há uma melhor separação entre o período de observação do período de tratamento (viés de tratamento precoce), enquanto o PRISM III-24 é mais preciso para avaliações individuais de risco de mortalidade (35).

O escore foi submetido a uma validação interna para avaliar o desempenho, porém não está especificado como foi feita a divisão dos pacientes entre a amostra de desenvolvimento e validação do escore. O modelo apresentou capacidade de discriminação e calibração adequada, tanto para o PRISM III-12 (área sob a curva ROC= 0,941; Hosmer-Lemeshow $X^2 = 4,992$ $p = 0,4168$) quanto para o PRISM-24 (área sob a curva ROC= 0,944; Hosmer-Lemeshow $X^2 = 3,993$ $p = 0,5504$).

O PRISM III é um escore com ampla aceitação em decorrência de sua validação adequada em uma amostra populacional grande envolvendo uma série de diferentes UTIP (10), sendo considerado o modelo padrão contra os quais outros escores são comparados (39). É muito utilizado para descrever a gravidade dos casos nos diferentes braços de ensaios clínicos randomizados com crianças criticamente doentes (10).

Entretanto alguns problemas são destacados na literatura. O viés do tratamento precoce é o principal problema (10). O PRISM III inclui dados das primeiras 12 ou 24 horas na UTIP, e não somente da admissão; com isso um melhor tratamento dado no início, antes que os primeiros dados (variáveis) sejam observados, pode melhorar os mais anormais valores coletados. Por exemplo, ventilação mecânica adequada em pacientes com insuficiência respiratória deve melhorar a PaO₂, PCO₂ e pH se estas variáveis são medidas poucas horas após a admissão na UTI. Isto significa que o valor médio do PRISM III será menor nesta UTI do que seria para pacientes similares em uma UTI menos eficiente (com atraso na instituição da ventilação ou ventilação inadequada). Isso pode resultar em uma percepção que os pacientes são menos graves, isto é com menor probabilidade de morte, do que realmente são (38).

Outro problema levantado é relacionado ao fato que os dados dos PRISM III são coletados durante as primeiras 12 ou 24 horas na UTIP, ou até a morte do paciente (10). Como uma significativa proporção de crianças gravemente doentes morre durante o primeiro dia de sua estadia na UTI (>40% em um estudo na Austrália), a coleta por um maior período de tempo deve inflar a capacidade do PRISM III para prever a morte (38). Em outras palavras,

existe o perigo de que o escore esteja realmente fazendo o diagnóstico de morte em vez de prever o evento (38).

Também é considerado como fraqueza o fato do PRISM III ser patenteado, sendo necessária a compra de seu software por um valor elevado, além de uma taxa anual de licença de uso (10). Com isso o seu uso pode ficar inviabilizado em países em desenvolvimento, nos quais recursos para a saúde são geralmente escassos. Outro possível problema é relacionado ao grande número de variáveis necessárias para o cálculo do escore, dificultando a realização rotineira do escore, o que pode criar mais uma tarefa num ambiente de grande carga de trabalho (39).

Os desenvolvedores do PRISM III já realizaram recalibrações (vide item - Atualizações) do modelo através da coleta de dados das UTIP que possuem o software do escore (2). A última recalibração do modelo ocorreu recentemente com dados de aproximadamente 15.000 pacientes (2), já estando disponível para os proprietários do PRISM III (40). Não há informações disponíveis sobre o desenvolvimento e publicação de uma nova atualização do modelo.

3.9.2) Paediatric Index of Mortality

O Paediatric Index of Mortality (PIM) foi publicado em 1997, tendo o objetivo de ser uma alternativa ao PRISM (38). Foi desenvolvido através de dados de 5695 crianças admitidas em oito UTIPs da Austrália e Reino Unido. Era composto de oito variáveis coletadas (primeira medida de cada variável) no período compreendido entre o primeiro contato do médico da UTIP com o paciente até a primeira hora após a admissão na unidade. Apresentava discriminação (área sob a curva ROC= 0,90) e calibração (Hosmer-Lemeshow $X^2 = 8,73$ $p= 0,37$) adequadas.

Em 2003 foi publicada a atualização do modelo, chamada de Paediatric Index of Mortality 2 (PIM2) (15). Foi desenvolvido a partir de dados de 20787 admissões de 14 UTIP da Austrália, Nova Zelândia e Reino Unido. A coleta de dados ocorreu entre 1997 e 1999.

O escore é composto de 10 variáveis (tabelas 4 e 5), sendo utilizado o primeiro valor de cada variável medida no período compreendido entre o primeiro contato do médico da UTIP com o paciente até 1 hora após a admissão na unidade. A equação necessária para estimar a mortalidade prevista é de domínio público, existindo endereços eletrônicos na Internet com um software que realiza o cálculo instantaneamente (10).

Tabela 4 – Variáveis utilizadas para o cálculo do PIM2

Pressão arterial sistólica
Reação pupilar à luz
Relação PaO ₂ /FiO ₂
Basic Excess (excesso de base)
Ventilação mecânica
Admissão eletiva
Recuperação de cirurgia ou procedimentos
Admissão após circulação extracorpórea
Diagnóstico de alto risco
Diagnóstico de baixo risco

Tabela 5 – Diagnósticos de alto e baixo risco utilizados para o cálculo do PIM2

Diagnóstico de alto risco	Diagnóstico de baixo risco
Parada cardiorrespiratória pré-admissão	Asma como causa da admissão
Imunodeficiência combinada grave	Bronquiolite como causa da admissão
Leucemia ou Linfoma após primeira indução	Crupe como causa da admissão
Hemorragia cerebral espontânea	Apneia obstrutiva do sono como causa da admissão
Cardiomiopatia ou Miocardite	Cetoacidose diabética como causa da admissão

Foi realizada uma validação interna para avaliar o desempenho, com os dados das admissões de sete UTIP sendo utilizados para o desenvolvimento do modelo (9450 admissões) e os dados das restantes sete UTIP para o teste do score (11337 admissões). O modelo apresentou capacidade de discriminação e calibração adequada, tanto na amostra utilizada para testar o modelo (área sob a curva ROC= 0,90; Hosmer-Lemeshow $X^2 = 8,14$ $p= 0,42$) quanto na amostra total do estudo (área sob a curva ROC= 0,90; Hosmer-Lemeshow $X^2 = 11,56$ $p= 0,17$).

Alguns pontos fortes do score são destacados: desenvolvimento e validação adequados, ser um score de domínio público, e não ter o viés de tratamento precoce por incluir dados referentes apenas à primeira hora de admissão (10). Entre as fraquezas destacadas estão: o score foi pouco testado fora dos países onde foi desenvolvido (10), e uma das variáveis de predição é uma intervenção médica (ventilação mecânica) que pode ser

sujeita a um viés secundário a diferentes limiares de intervenção (a indicação de ventilação mecânica pode variar entre serviços ou mesmo entre médicos) (34).

3.9.3) Estudos sobre aplicação do PIM2

A partir da publicação em 2003 do PIM2, vários estudos foram realizados avaliando o desempenho do escore em populações e cenários diferentes do estudo original.

Slater e col publicaram em 2004 um estudo avaliando o PIM2 em 10 UTIP na Austrália e Nova Zelândia (26). A população do estudo foi de 11.296 pacientes admitidos entre janeiro de 2000 e dezembro de 2001. A taxa de mortalidade geral foi 4,3%. O escore apresentou discriminação adequada (área sob a curva ROC= 0,90), porém uma calibração inadequada (Hosmer-Lemeshow $X^2 = 17,53$ $p = 0,025$). A taxa de mortalidade padronizada (TMP) foi de 0,97 (IC95% 0,86-1,05). Os autores destacaram que as UTIP participantes deste estudo também participaram do estudo de desenvolvimento do PIM2, com isso um bom desempenho do escore poderia ser influenciado pela similaridade das amostras populacionais entre os estudos.

Mestrovic e col publicaram em 2005 um estudo avaliando o PIM2 em uma UTIP na Croácia (41). A população do estudo foi de 591 pacientes admitidos entre junho de 2002 e julho de 2004. A idade média da população foi 62 meses, o tempo médio de permanência nas unidades foi 2 dias, e a taxa de mortalidade geral foi 5,9%. A maioria da pacientes (73,3%) apresentava uma condição clínica como causa da admissão, sendo o traumatismo craniano a principal causa de óbito (20%). O escore apresentou discriminação adequada (área sob a curva ROC= 0,97), porém uma calibração inadequada (Hosmer-Lemeshow $X^2 = 19,59$ $p = 0,012$). A TMP foi de 1, não sendo calculado o intervalo de confiança.

Brady e colaboradores publicaram em 2006 um estudo avaliando o PIM2 em 18 UTIP do Reino Unido (25). A população do estudo foi de 10197 pacientes admitidos entre março de 2001 e fevereiro de 2002. A idade média da população foi 16,8 meses, o tempo médio de permanência nas unidades foi 1,9 dias, e a taxa de mortalidade geral foi 6,2%. O escore apresentou discriminação adequada (área sob a curva ROC= 0,84), porém uma calibração inadequada (Hosmer-Lemeshow $X^2 = 39,8$ $p = <0,001$). Não foi relatado a TMP, e nem fornecidos os dados necessários para o cálculo.

Os autores destacaram que o escore tendia a subestimar risco nas admissões de baixo e médio risco (25). Os autores também destacaram uma mortalidade adicional de 3%,

considerando o período até a alta hospitalar, contabilizando uma mortalidade hospitalar de 9,2% (25).

Thukral e col publicaram em 2006 um estudo avaliando o PIM2 em uma UTIP na Índia (42). A população do estudo foi de 215 pacientes admitidos entre julho de 2002 e julho de 2003. A idade média da população foi de 18 meses, o tempo médio de permanência na unidade foi de 4 dias, e a taxa de mortalidade geral foi de 32%. O escore apresentou discriminação (área sob a curva ROC= 0,81) e calibração (Hosmer-Lemeshow $X^2 = 7,64$ $p= 0,47$) adequadas. Porém os autores destacam que o escore previu menos mortes do que as observadas em todos os níveis de probabilidade (tabela 6), com isso a TMP observada foi elevada (mortes previstas = 48,59; mortes observadas = 76; TMP=1,57 IC95% 1,24-1,59).

Tabela 6 – Comparação entre as mortes observadas e previstas no estudo de Thukral e colaboradores.

Probabilidade de morte	< 1 %	1-5%	5-10%	10-15%	15-20%	20-25%	25-30%	30-50%	>50%
Mortes observadas	0	4	15	4	2	4	1	17	29
Mortes previstas	0,022	1,058	5,157	2,563	1,475	2,422	1,142	8,681	26,147

Os autores também destacaram a quantidade elevada de pacientes com desnutrição (48,8%) e com doenças crônicas (49,3%), sendo observada uma piora da capacidade de previsão do escore quando eram comparadas populações com ou sem desnutrição grave (TMP = 1,95 e 1,45 respectivamente) e populações com ou sem doenças crônicas (TMP = 1,77 e 1,37 respectivamente) (42). Durante a discussão sobre as possíveis causas das diferenças entre a mortalidade prevista e a observada, os autores relatam que esta diferença poderia ser secundária a diferença entre as características das amostras, diferenças de alocação de recursos, e diferenças entre práticas das unidades (unidades onde foi desenvolvido o escore e a unidade do estudo acima).

Eulmesekian e col publicaram em 2007 um estudo avaliando o PIM2 em uma UTIP na Argentina (24). A população do estudo foi de 1574 pacientes admitidos entre janeiro de 2004 e dezembro de 2005. A idade média da população foi de 45 meses, o tempo médio de permanência nas unidades foi de 2 dias, e a taxa de mortalidade geral foi de 2,6%. A maioria dos pacientes (62,2%) apresentava uma condição clínica como causa da admissão. O escore apresentou discriminação adequada (área sob a curva ROC= 0,90), porém calibração inadequada (Hosmer-Lemeshow $X^2 = 12,3$ $p= 0,348$). Para o conjunto total da população o escore previu mais mortes do que as observadas (mortes previstas = 48,1; mortes observadas = 41; TMP=0,85 IC95% 0,6-1,1). Os autores relataram que entre as possíveis razões para

explicar a calibração inadequada estão a baixa taxa de mortalidade e a diferença da mistura de casos em relação a população utilizada no desenvolvimento do escore (porém os autores não realizaram uma caracterização minuciosa da sua amostra).

Prieto-Espunes e col publicaram em 2007 um estudo avaliando o PIM2 em duas UTIP na Espanha (43). A população do estudo foi de 241 pacientes admitidos entre novembro de 2003 e junho de 2004. A idade média da população foi de 62,6 meses, o tempo médio de permanência nas unidades foi de 7,3 dias, e a taxa de mortalidade geral foi de 4,1%. A maioria dos pacientes (61%) apresentava uma condição clínica como responsável pela admissão, sendo as doenças do sistema respiratório o diagnóstico mais comum na admissão entre as causas clínicas. O escore apresentou discriminação (área sob a curva ROC= 0,87) e calibração (Hosmer-Lemeshow $X^2 = 4,8730$ $p= 0,8461$) adequadas. Entretanto o escore previu menos mortes do que as observadas (mortes previstas = 11; mortes observadas = 10; TMP=0,91 IC95% 0,35-1,48).

Qureshi e col publicaram em 2007 um estudo avaliando o PIM2 em uma UTIP no Paquistão (44). A população do estudo foi de 101 pacientes admitidos entre abril e junho de 2006. A idade média da população foi de 45,6 meses, o tempo médio de permanência nas unidades foi de 12,1 dias, e a taxa de mortalidade geral foi de 28,7 %. A maioria dos pacientes era constituída de pacientes clínicos (86%) e apresentava desnutrição (55,4%). O escore apresentou discriminação (área sob a curva ROC= 0,88) e calibração (Hosmer-Lemeshow $X^2 = 9,65$ $p= 0,29$) adequadas. A TMP foi de 1,4 (mortes previstas = 20,69; mortes observadas = 29; IC95% 0,77-2,0). Diferente do estudo de Thukral e colaboradores (42) a desnutrição não foi associada a uma piora da capacidade de previsão do escore. Os autores também comentaram que entre as possíveis causas para a elevada taxa de mortalidade observada estão o grande número de pacientes não cirúrgicos (diferente dos países desenvolvidos onde a proporção de pacientes cirúrgicos é maior) e um maior grau de instabilidade clínica dos pacientes (a mortalidade prevista foi três vezes maior do que em estudos prévios). Os autores não comentaram sobre o papel da estrutura, equipamentos, pessoal e práticas clínicas utilizadas na unidade como fatores responsáveis pela mortalidade e TMP elevada.

Wolfler e col publicaram em 2007 um estudo avaliando o PIM2 em 18 UTIP na Itália (45). A população do estudo foi de 3266 pacientes admitidos entre março de 2004 e fevereiro de 2005. A idade média da população foi de 10 meses, o tempo médio de permanência nas unidades foi de 2 dias, e a taxa de mortalidade geral foi de 5,2%. A maioria dos pacientes apresentava uma condição cirúrgica como causa da admissão. O escore apresentou

discriminação (área sob a curva ROC= 0,89) e calibração (Hosmer-Lemeshow $X^2 = 9,86$ $p= 0,26$) adequadas. A TMP foi de 0,89 (mortes previstas = 192; mortes observadas = 171; IC95% 0,71-1,09). Os autores destacaram que entre os pacientes com condições clínicas (46,8% do total da amostra), 28,9% tinham uma doença crônica e apresentaram uma mortalidade de 15,6%. Entre os pacientes classificados como apresentando imunodeficiência (pacientes que receberam quimioterapia ou corticoides, com infecção por síndrome da imunodeficiência adquirida, com imunodeficiência congênita, ou que foram transplantados) apresentaram uma mortalidade de 17%.

Hwang e col publicaram em 2008 um estudo avaliando o PIM2 em uma UTIP na Coreia do Sul (46). A população do estudo foi de 190 pacientes admitidos entre janeiro de 2003 e dezembro de 2007. A idade média da população foi de 63 meses, o tempo médio de permanência nas unidades foi de 2,5 dias, e a taxa de mortalidade geral foi de 14,7 %. A maioria da população (77,8%) apresentava um condição clínica como causa da admissão. O escore apresentou discriminação (área sob a curva ROC= 0,85) e calibração (Hosmer-Lemeshow $X^2 = 14,98$ $p= 0,308$) adequadas. A TMP foi de 1,0 (mortes previstas = 27,98; mortes observadas = 28; não sendo calculado o intervalo de confiança).

Canonero e col publicaram em 2010 um estudo avaliando o PIM2 em uma UTIP na Argentina (47). A população do estudo foi de 416 pacientes admitidos entre janeiro e dezembro de 2008. A idade média da população foi de 36 meses, o tempo médio de permanência nas unidades foi de 2 dias, e a taxa de mortalidade geral foi de 6,6%. O escore apresentou discriminação (área sob a curva ROC= 0,88) e calibração (Hosmer-Lemeshow $X^2 = 14,194$ $p= 0,077$) adequadas. Porém o escore previu mais mortes do que as observadas no pacientes com menor risco e menos mortes do que as observadas no pacientes com maior risco (tabela 7). Com isso apesar do modelo prever adequadamente o número de mortes no conjunto total da população (mortes previstas = 28,9; mortes observadas = 29; TMP=1,0 IC95% 0,55-1,55), nos pacientes de maior risco o escore apresenta uma piora do desempenho (mortes previstas = 20,6; mortes observadas = 28; TMP=1,38). Os autores relatam que o achado de uma calibração adequada avaliada pelo teste de Hosmer-Lemeshow em contraponto a observação de uma previsão inadequada quando é analisada a tabela com as faixas de risco pode ser decorrente do pequeno número da amostra e de mortes, o que pode influenciar no valor de p do teste de Hosmer-Lemeshow. Por isso os autores destacaram a utilidade da análise da tabela de faixas de risco como ferramenta auxiliar para avaliação da calibração.

Tabela 7 – Comparação entre mortes observadas e previstas em 10 faixas de risco crescentes no estudo de Canonero e col.

Faixas de risco	1	2	3	4	5	6	7	8	9	10
Mortes observadas	0	1	0	0	0	1	3	4	5	15
Mortes previstas	2,4	1,2	1,4	1,9	1,6	1,7	1,7	2	2,8	11,9

Hariharam e col publicaram em 2011 um estudo avaliando o PIM2 em uma UTIP em Trinidad e Tobago (48). A população do estudo foi de 163 pacientes admitidos entre janeiro e dezembro de 2008. A idade média da população foi de 60 meses, o tempo médio de permanência nas unidades foi de 5,4 dias, e a taxa de mortalidade geral foi de 5,5 %. A maioria dos pacientes (mais de 84%) apresentava uma condição clínica como causa da admissão, sendo as doenças respiratórias a principal categoria de diagnóstico entre os pacientes. O escore apresentou discriminação (área sob a curva ROC= 0,82) e calibração (Hosmer-Lemeshow $X^2 = 5,64$ $p = 0,58$) adequadas. A TMP foi de 0,89 (mortes previstas = 10,1; mortes observadas = 9; não sendo calculado o intervalo de confiança).

Ng e col publicaram em 2011 um estudo avaliando o PIM2 em três UTIP em Hong Kong (49). A população do estudo foi de 540 pacientes admitidos entre janeiro e dezembro de 2006. A idade média da população foi de 24 meses e a taxa de mortalidade geral foi de 1,3 %. A maioria dos pacientes (75,9%) apresentava uma condição clínica como causa da admissão, tendo as doenças do sistema respiratório a principal causa das internações (31,7%). O escore apresentou discriminação adequada (área sob a curva ROC= 0,90), mas a calibração não foi realizada devido ao insuficiente número de mortes. Entretanto foi observado que modelo previu mais mortes do que as observadas (mortes previstas = 14,2; mortes observadas = 7; TMP = 0,49 IC95% 0,13-0,86).

Imamura e col publicaram em 2011 um estudo avaliando o PIM-2 em uma UTIP no Japão (50). A população do estudo foi de 2536 pacientes admitidos entre janeiro de 2008 e dezembro de 2010. A idade média da população foi de 44,3 meses, o tempo médio de permanência nas unidades foi de 6,7 dias, e a taxa de mortalidade geral foi de 2,6 %. A maioria dos pacientes era constituída por pacientes cirúrgicos (67,7%), com a maioria dos procedimentos (83,7%) não sendo cirurgias cardíacas. O escore apresentou discriminação (área sob a curva ROC= 0,92) e calibração (Hosmer-Lemeshow $X^2 = 4,8$ $p = 0,44$) adequadas. Entretanto o escore previu mais mortes do que as observadas (mortes previstas = 86,9; mortes observadas = 67; TMP=0,77 IC95% 0,59-0,96). Os autores destacaram que a mortalidade dos pacientes clínicos foi significativamente maior do que dos pacientes cirúrgicos (5,7% x 1,2%

respectivamente), sendo relatado que esta diferença poderia ser secundária ao fato que um grande número das admissões de pacientes cirúrgicos era apenas para monitoramento de procedimentos de baixa ou média complexidade (como tonsilectomia, adenoidectomia, reparo de fissura labial e de palato). Os autores também destacaram que entre as possíveis explicações para a baixa mortalidade e TMP encontrada estão a desatualização do escore (os dados utilizados para produção do modelo foi coletados há mais de 10 anos atrás, com grandes avanços na qualidade do cuidado realizados desde então), o fato que maioria da população estudada era constituída de pacientes com baixo risco de morte (87% dos pacientes tinha probabilidade de morte menor do que 5%) e ao grande número de pacientes cirúrgicos.

Também foi realizado um estudo avaliando o PIM2 em uma população específica de pacientes pediátricos submetidos a cirurgias cardíacas (51). A população do estudo foi de 9208 pacientes, com os dados coletados durante o período entre janeiro de 2005 e setembro de 2007. O escore apresentou discriminação adequada (área sob a curva ROC= 0,80), porém calibração (Hosmer-Lemeshow $X^2 = 47,3$ $p = <0,0001$) inadequada. A idade média da população foi de 39,8 meses, o tempo médio de permanência nas unidades foi de 7,4 dias, a taxa de mortalidade geral foi de 3,4% e a TMP foi de 0,8 (IC95% 0,77 – 0,83). Embora a discriminação tenha sido adequada segundo os critérios geralmente utilizados para avaliar o desempenho, os autores consideraram o desempenho do escore inadequado (2). Entre as possíveis causas levantadas para explicar este achado estão a diferença na mistura de casos em decorrência da especificidade da população do estudo e diferenças na estrutura, pessoal e práticas clínicas.

A análise dos estudos avaliando o desempenho do PIM2 em populações e cenários diferentes do artigo original evidenciou uma discriminação adequada em todos os trabalhos (área sob a curva ROC maior do que 0,8), porém em quatro estudos a calibração foi inadequada ($p < 0,05$ no teste de Hosmer-Lemeshow) (tabela 9). Com isso a maioria dos autores considerou que o PIM-2 poderia ser utilizado como escore preditor de desfecho nas populações e cenários pesquisados.

Entre os estudos avaliando o PIM2, alguns também avaliaram simultaneamente outros escores. Foram encontrados três estudos com objetivo de comparar o desempenho do PIM2 e PRISM III, os escores de predição de desfecho mais recentes (tabela 8).

Tabela 8 – Estudos comparando o desempenho PIM2 versus PRISM III

Autor	PIM2	PRISM III
Slater e col		
<i>Discriminação (1)</i>	0,90 IC95% 0,88-0,91	0,93 IC95% 0,92-0,94
<i>Calibração (2)</i>	$X^2 = 17,53$ $p = 0,025$	$X^2 = 89,1$ $p = < 0,0001$
Brady e col		
<i>Discriminação (1)</i>	0,84 IC95% 0,82-0,86	0,88 IC95% 0,87-0,89 (3)
<i>Calibração (2)</i>	$X^2 = 39,8$ $p = < 0,01$	Não foi realizada
Hwang e col		
<i>Discriminação (1)</i>	0,85 IC95% 0,77-0,93	0,79 IC95% 0,68-0,89
<i>Calibração (2)</i>	$X^2 = 14,986$ $p = 0,308$	$X^2 = 12,899$ $p = 0,456$

- (1) Discriminação avaliada através da área sob a curva ROC
- (2) Calibração avaliada através do teste de Hosmer-Lemeshow
- (3) Versão customizada pelos autores do estudo

Slater e col, além do PIM2, avaliaram o desempenho do PIM, PRISM e PRISM III (26). O PRISM e o PRISM III apresentaram discriminação adequada (área sob a curva ROC= 0,90 e 0,93 respectivamente) e calibração inadequada (Hosmer-Lemeshow $X^2 = 446,56$ $p = < 0,0001$; e $89,1$ $p = < 0,0001$ respectivamente). O PIM e o PIM2 apresentaram discriminação adequada (área sob a curva ROC= 0,89 e 0,90 respectivamente) e calibração inadequada (Hosmer-Lemeshow $X^2 = 36,41$ $p = < 0,0001$; e $17,53$ $p = 0,025$ respectivamente). Os autores concluíram que todos os escores discriminaram adequadamente entre os pacientes que morreram dos que sobreviveram, três escores (PIM, PRISM e PRISM III) previram mais mortes do que as observadas, e que o ajuste entre a mortalidade observada e esperada foi melhor com o PIM2 no cenário do estudo.

Entretanto é importante destacar dois pontos que podem interferir na comparação do desempenho dos escores no estudo acima (26). O primeiro ponto é que tanto a amostra quanto o período de coleta de dados utilizados para avaliar o PIM, PRISM e PRISM III (15.670 pacientes; entre janeiro de 1997 e dezembro de 1999) foram diferentes dos utilizados para avaliar o PIM2 (11.296 pacientes; entre janeiro de 2000 e dezembro de 2001), só sendo similares as UTIP participantes do estudo. O segundo ponto é o fato que as UTIP participantes do estudo também participaram do estudo de desenvolvimento do PIM2, com

isso pode ocorrer uma tendência a um melhor desempenho deste modelo em decorrência das similaridades entre as populações.

Brady e col em estudo publicado em 2006, além do PIM2, também propuseram avaliar o PIM, PRISM e PRISM III (25). Entretanto os autores do PRISM III não disponibilizaram a equação para o cálculo do score, impossibilitando a avaliação do desempenho do score com uso da equação original. Na tentativa de avaliar o PRISM III foi desenvolvida uma customização do modelo, ou seja, foi realizada uma reestimativa do peso relativo das variáveis de predição na amostra específica do estudo utilizando regressão logística com a criação de uma nova equação.

O PRISM no estudo acima apresentou discriminação adequada (área sob a curva ROC= 0,82) e calibração inadequada (Hosmer-Lemeshow $X^2 = 541,9$ $p = < 0,001$). O PIM e o PIM2 apresentaram discriminação adequada (área sob a curva ROC= 0,81 e 0,84 respectivamente) e calibração inadequada (Hosmer-Lemeshow $X^2 = 86,7$ $p = < 0,001$; e 39,8 $p = < 0,01$ respectivamente). Já o PRISM III customizado apresentou discriminação adequada (área sob a curva ROC= 0,88), porém em decorrência de ser uma customização a calibração foi avaliada através do método de Cox (com adequada calibração) ao invés do teste de Hosmer-Lemeshow. Portanto o objetivo de comparar o desempenho do PIM2 e PRISM III com suas equações originais não foi alcançado.

Hwang e col em estudo publicado em 2008, além da avaliação do PIM2, também avaliaram o desempenho do PRISM III (46). O PIM2 apresentou uma melhor discriminação do que o PRISM III (área sob a curva ROC= 0,85 e 0,79 respectivamente), porém o PRISM III apresentou uma melhor calibração do que o PIM2 (Hosmer-Lemeshow $X^2 = 12.899$ $p = 0,456$; e 14.986 $p = 0,308$ respectivamente). Entretanto é importante destacar o pequeno tamanho da amostra (190 pacientes) em relação à maioria dos estudos publicados, o que pode interferir com a avaliação do desempenho.

A análise dos três estudos comparando PIM2 e PRISM III evidencia falhas nas metodologias dos trabalhos que podem interferir na avaliação e comparação do desempenho dos modelos, o que impossibilita afirmar que um modelo é melhor do que o outro. Lacroix e col relatam que mais estudos são necessários antes que se possa concluir que o PRISM III é melhor que o PIM-2 ou vice-versa (10).

Entre os estudos encontrados avaliando o PIM2 que também avaliavam outros escores simultaneamente, alguns pesquisadores compararam o desempenho dos escores mais recentes com as suas versões anteriores.

Slater e col observaram que o PIM2 (área sob a curva ROC= 0,90; Hosmer-Lemeshow $X^2 = 17,53$ $p = 0,025$) e o PRISM III (área sob a curva ROC= 0,93; Hosmer-Lemeshow $X^2 = 89,1$ $p = < 0,0001$) tem um desempenho superior em relação ao PIM (área sob a curva ROC= 0,89; Hosmer-Lemeshow $X^2 = 36,41$ $p = < 0,0001$) e ao PRISM (área sob a curva ROC= 0,90; Hosmer-Lemeshow $X^2 = 446,56$ $p = < 0,0001$) (26).

Brady e col também observaram desempenho superior do PIM2 em relação ao PIM quando foi avaliada a discriminação (área sob a curva ROC= 0,84 e 0,81 respectivamente) e calibração (Hosmer-Lemeshow $X^2 = 39,8$ $p = < 0,01$; e 86,7 $p = < 0,001$ respectivamente) (25).

Thukral e col avaliaram o desempenho do PIM e PRISM, além do PIM2 (42). Os dois escores apresentaram um desempenho inferior ao PIM2 quando foi avaliada a discriminação (área sob a curva ROC= 0,82 e 0,80 respectivamente) e calibração (Hosmer-Lemeshow $X^2 = 21,77$ $p = 0,005$; e 10,82 $p = 0,21$ respectivamente).

Qureshi e col também avaliaram o PRISM, além do PIM2 (44). Sendo observado que o PIM2 apresentava melhor discriminação (área sob a curva ROC= 0,88 e 0,78 respectivamente), porém o PRISM apresentava melhor calibração (Hosmer-Lemeshow $X^2 = 7,49$ $p = 0,49$; e 9,65 $p = 0,29$ respectivamente).

Canonero e col avaliaram o PRISM, além do PIM2 (47). Como no estudo de Qureshi (44), foi observado que o PIM2 tinha melhor discriminação (área sob a curva ROC= 0,88 IC95% 0,82-0,95; e 0,85 IC95% 0,78-0,92 respectivamente), porém o PRISM apresentava melhor calibração (Hosmer-Lemeshow $X^2 = 5,93$ $p = 0,54$; e 14,19 $p = 0,07$ respectivamente).

Ng e col observaram que o PIM2 apresentava discriminação superior ao PIM (área sob a curva ROC= 0,90 IC95% 0,71-1,00; e 0,88 IC95% 0,70-1,00 respectivamente), porém não realizaram avaliação da calibração em decorrência da pequena quantidade de mortes (49).

A análise dos estudos comparando o desempenho entre as versões evidenciou na maioria das ocasiões o desempenho superior dos escores mais recentes, fortalecendo a tese da necessidade de atualizações frequentes dos escores em decorrência da observação que mudanças na demografia dos pacientes, na prevalência das doenças e nas práticas utilizadas na UTI através do tempo pode alterar o desempenho dos modelos (vide item - Atualizações) (36).

A partir dos achados relatados nos estudos encontrados, algumas observações podem ser produzidas: 1) a maioria dos autores considerou que o PIM2 poderia ser utilizado como escore de predição de desfecho nas populações e cenários pesquisados; 2) não há dados para afirmar que o PRISM III é melhor que o PIM2 ou vice-versa; e 3) os escores mais recentes

apresentam desempenho superior em relação às versões anteriores, tornando necessária a utilização das versões mais recentes dos modelos (PIM2 e PRISM III).

Além disso, os autores destacaram duas características do PIM2 que facilitariam a difusão de sua utilização em relação ao PRISM III: domínio público da equação necessária para o cálculo da probabilidade de morte e facilidade da coleta das variáveis utilizadas no escore (poucas variáveis e curto período de observação do paciente).

Através de todas estas observações pode ser levantada a mesma conclusão que Piva e col: o PIM2 pode ser considerado uma ferramenta útil, eficiente e sem custo para ser utilizada em pacientes admitidos em UTIP, principalmente de países em desenvolvimento (52).

Tabela 9 – Estudos avaliando o desempenho do PIM2 em cenários diferentes.

Autor	Ano	País	Pacientes	Mortalidade	Discriminação (1)	Calibração (2)
Slater	2004	Austrália/Nova Zelândia	10.834	4,30%	0,90 (0,88-0,91)	X ² = 17,53 p= 0,025
Mestrovic	2005	Croácia	591	5,90%	0,97 (0,96-0,99)	X ² = 19,59 p= 0,012
Brady	2006	Inglaterra	9.475	6,20%	0,84 (0,82-0,86)	X ² = 39,8 p= <0,001
Thukral	2006	Índia	215	32%	0,81 (0,75-0,87)	X ² = 7,64 p= 0,47
Prieto Espuñe	2006	Espanha	241	4,10%	0,87 (0,78-0,95)	X ² = 4,87 p= 0,8461
Wolfler	2007	Itália	3.266	5,20%	0,89 (0,86-0,91)	X ² = 9,86 p= 0,26
Eulmesekian	2007	Argentina	1.574	2,70%	0,9 (0,89-0,92)	X ² = 12,2 p= 0,0348
Qureshi	2007	Paquistão	138	28,70%	0,88 (0,81-0,94)	X ² = 9,65 p= 0,29
Hwang	2008	Coréia do Sul	190	14,70%	0,85 (0,77-0,93)	X ² = 14,98 p= 0,308
Canonero	2010	Argentina	416	6,60%	0,88 (0,82-0,95)	X ² = 14,19 p= 0,077
Hariharan	2011	Trinidad e Tobago	163	5,50%	0,82 (0,72-0,92)	X ² = 5,64 p= 0,58
Ng	2001	Hong Kong (China)	540	1,30%	0,90 (0,73-1,00)	não realizada
Imamura	2010	Japão	2536	2,60%	0,92 (0,89-0,96)	X ² = 4,8 p= 0,44

(1) Discriminação avaliada através da área sob a curva ROC

(2) Calibração avaliada através do teste de Hosmer-Lemeshow

3.10) Escores de disfunção de órgãos

A disfunção de múltiplos de órgãos é a principal causa de morbidade e mortalidade na UTI (9). Avaliação inicial e sequencial da disfunção de órgãos fornecem informações sobre o prognóstico do paciente, bem como sobre a efetividade de tratamento (9).

Os escores de disfunção de órgãos avaliam a presença e a severidade da disfunção de órgãos (36), sendo desenvolvidos para descrever a gravidade da doença durante a permanência do paciente na UTI (10). Além desta função primária de avaliação de disfunção de órgãos, estes modelos possuem equações que estimam a probabilidade de morte dos pacientes (10).

Para realizar estas funções são coletados diariamente dados sobre distúrbios fisiológicos de sistemas orgânicos específicos desde admissão até o desfecho (morte) ou alta da UTI (10). A maioria dos escores utilizados na população adulta ou pediátrica indica valores para seis sistemas orgânicos: respiratório, cardiovascular, renal, hematológico, hepático, e sistema nervoso central (9).

Na população pediátrica os principais escores de disfunção de órgãos utilizados são o Pediatric Logistic Organ Dysfunction e o Pediatric Multiple Organ Dysfunction Score.

3.10.1) Pediatric Logistic Organ Dysfunction

O Pediatric Logistic Organ Dysfunction (PELOD) foi o primeiro escore de disfunção de órgãos publicado para a população pediátrica (53), sendo o escore mais utilizado (2).

Foi desenvolvido a partir de dados de 594 admissões consecutivas de três UTIP da França e Canadá, no período de janeiro a maio de 1997. Com posterior validação externa do escore com dados de 1806 pacientes de sete UTIP da França, Canadá e Suíça (54).

O modelo apresentou desempenho adequado (avaliação de probabilidade de morte) na amostra de desenvolvimento (área sob a curva ROC= 0,98; Hosmer-Lemeshow X^2 com $p= 0,44$), porém na amostra de validação foi observado uma

calibração inadequada (área sob a curva ROC= 0,91; Hosmer-Lemeshow X^2 para $p < 0,0001$) (55).

O escore é composto de 12 variáveis de seis sistemas orgânicos (tabela 10), sendo utilizado o valor mais alterado de cada variável para o cálculo do modelo.

Tabela 10 – Sistemas orgânicos e variáveis utilizadas no PELOD

<i>Sistema respiratório</i>
Relação PaO ₂ /FiO ₂
PaCO ₂
Ventilação mecânica
<i>Sistema cardiovascular</i>
Frequência cardíaca
Pressão sistólica
<i>Sistema neurológico</i>
Escala de coma de Glasgow
Reação pupilar
<i>Sistema hepático</i>
TGO ou TGP
Tempo de protrombina ou RNI
<i>Sistema renal</i>
Creatinina sérica
<i>Sistema hematológico</i>
Contagem de leucócitos
Contagem de plaquetas

Entre os pontos fracos destacados estão o pequeno tamanho da amostra para o desenvolvimento do modelo, a complexidade para o cálculo do sistema de escore, e a característica binária de algumas variáveis do modelo (isto é, apenas duas condições da variável são permitidas, normal ou disfunção) (56). Ao definir que a variável só pode ter estas duas condições, o escore não pode identificar que existe disfunção gradativa dos órgãos que está associado à aumento gradual na mortalidade (57). Além disso, o modelo apresenta calibração inadequada, sendo considerado um ponto desfavorável para o modelo (55). Já como pontos positivos são mencionados a metodologia adequada para o desenvolvimento do modelo e a utilização da validação externa (58).

3.10.2) Pediatric Multiple Organ Dysfunction Score

O Pediatric Multiple Organ Dysfunction Score (P-MODS) é o escore de disfunção de órgãos mais recentemente publicado para a população pediátrica, sendo uma alternativa ao PELOD (56).

Foi desenvolvido a partir de dados de 6456 pacientes pediátricos admitidos de uma única UTIP nos Estados Unidos entre Maio de 1996 e Dezembro de 1999. O escore é composto de cinco variáveis (ácido láctico, relação PaO₂/FiO₂, bilirrubina, fibrinogênio e ureia) representando cinco sistemas orgânicos (cardiovascular, respiratório, hepático, hematológico e renal) respectivamente.

Foi realizada validação interna do modelo, com metade da amostra (3228 pacientes) utilizada para o desenvolvimento do modelo e a outra metade (3228 pacientes) utilizada para validação do modelo. Tendo apresentado desempenho adequado tanto na amostra de desenvolvimento (área sob a curva ROC= 0,81; Hosmer-Lemeshow $X^2 = 2,54$ com $p= 0,637$) quanto na amostra de validação (área sob a curva ROC= 0,78; Hosmer-Lemeshow $X^2 = 4,73$ com $p= 0,317$).

Entre os pontos fracos estão a ausência de validação externa, o desempenho pior do que o PELOD (área sob a curva ROC= 0,78 X 0,91) e não possuir variável que avalie o sistema neurológico (58). Já como pontos positivos são mencionados a metodologia adequada para o desenvolvimento, o tamanho da amostra e a facilidade de cálculo (58).

3.11) Atualizações

Ao longo do tempo ocorrem mudanças na demografia dos pacientes, na prevalência das doenças e nas práticas utilizadas na UTI que alteram o desempenho dos escores de predição de desfecho (36), geralmente produzindo uma deterioração de sua acuidade (16).

As mudanças na demografia dos pacientes e na prevalência das doenças alteram a mistura de casos, o que pode influenciar no desempenho dos escores como demonstrado por Murphy-Filkins e col (59). Neste estudo os autores observaram que se certa porcentagem de um determinado tipo de paciente é ultrapassada, o modelo pode não ser acurado (59).

O aumento da proporção de pacientes pediátricos com comorbidades nas UTIP pode ser citado como exemplo das mudanças da mistura de casos ao longo do tempo. Odetola e col observaram um aumento na proporção de pacientes com comorbidades admitidos em UTIP nos Estados Unidos entre 1997 e 2006 (35% e 41% respectivamente).

A mudança na prática clínica é o outro fator que pode alterar o desempenho dos modelos prognósticos. Slater e col relatam que novos tratamentos e novas abordagens de manejo podem mudar a relação entre fisiologia e desfecho (15). Também relatam que mudanças nos critérios de indicação para admissão na UTI e nas atitudes para a indicação de início ou interrupção de suporte de vida podem alterar a relação entre doença e desfecho (15). Além disso, Lacroix e col relatam que a tendência secular de queda da mortalidade nas UTIP modifica a discriminação e a calibração dos escores em relação à mortalidade (10).

Para contrabalançar a deterioração da acuidade os modelos são frequentemente submetidos à customização pela criação de novas equações ou mudança dos pesos das variáveis constituintes (60,61). Em alguns casos, a adição de novas variáveis de predição pode ser necessária durante a customização (62,63).

O PRISM III através de um software chamado de Pediatric Intensive Care Unit Evaluations (PICUEs), que permite a coleta de dados e o cálculo dos desfechos de mais de 100 UTIP nos Estados Unidos, realiza recalibrações periódicas (customização) do escore com o objetivo manter a acuidade do modelo (2).

Entretanto a customização dos modelos parece ter um efeito limitado (9), como verificado em um estudo avaliando o desempenho dos escores de predição de desfecho (terceira geração) e suas customizações utilizadas na população adulta em UTI, no qual foi observado que os modelos mostravam uma discriminação adequada, porém uma calibração ruim (64). A melhora inicial do desempenho dos modelos antigos com a customização não foi mantida, já que estes modelos não refletiam a mistura de casos, o padrão de práticas e tratamentos atuais, sendo necessário o desenvolvimento da quarta geração dos escores (9).

Há dados limitados para recomendar com qual frequência o desempenho dos modelos devem ser avaliados, sendo recomendado por Keegan e col que os modelos devem ser periodicamente avaliados e atualizações podem ser necessárias a cada quatro anos (16). Pollack, um dos autores do PRISM, sugere que os algoritmos de predição desenvolvidos há mais de cinco anos são susceptíveis de serem ultrapassados (34).

A necessidade de atualizações fica evidenciada pelo número de gerações dos modelos. O modelo prognóstico mais utilizado na população adulta já está na sua quarta geração (65), com seus principais concorrentes já tendo três atualizações (66,67) (tabela 11). O tempo médio entre a penúltima (terceira geração) e última atualização (quarta geração) nestes modelos foi de 13 anos, com as versões atuais sendo publicadas entre os anos de 2005 e 2007.

Tabela 11 - As quatro gerações de modelos prognósticos em UTI na população adulta

Primeira geração	Segunda geração	Terceira geração	Quarta geração
APACHE I	APACHE II	APACHE III	APACHE IV
	SAPS I	SAPS II	SAPS III
	MPM I	MPM II	MPM III

Abreviações: APACHE, Acute Physiology and Chronic Health Evaluation; SAPS, Simplified Acute Physiology Score; MPM, Mortality Probability Model.

O número de atualizações dos principais modelos utilizados na população pediátrica é menor do que na população adulta (tabela 12). Outra diferença é o tempo médio entre a penúltima (segunda geração) e última atualização (terceira geração) que foi de 7 anos. Entretanto as versões atuais são mais antigas, sendo publicadas entre os anos de 1996 e 2003 (15,35).

Tabela 12 - As três gerações de modelos prognósticos em UTI na população pediátrica.

Primeira geração	Segunda geração	Terceira geração
PSI	PRISM	PRISM III
	PIM	PIM 2

Abreviações: PSI, Physiologic Stability Index; PRISM, Pediatric Risk of Mortality; PIM, Paediatric Index of Mortality.

Com isso a adequabilidade dos principais escores de predição de desfecho utilizados em pediatria pode ser questionada pela possibilidade de deterioração de seu desempenho com o tempo. Tendo como exemplo o PRISM III, modelo atualmente mais utilizado, que foi publicado em 1996 (35). A deterioração com o tempo pode ainda ser maior se levarmos em consideração que os dados utilizados para produção dos modelos

são ainda mais antigos, no caso de PRISM III são dados coletados de pacientes no período entre 1989 e 1994 (35).

3.12) Utilizações dos escores de acuidade

Além das informações sobre predição de desfecho (escores de predição de desfecho) e avaliação da função dos órgãos (escores de disfunção de órgãos), os sistemas de escores têm alguns outros potenciais usos (tabela 13).

Tabela 13 - Potenciais usos dos escores de acuidade (16)

Nível nacional

Benchmarking

Nível institucional

Uso interno para gestão de melhoria de qualidade, ao comparar o desempenho do hospital com a média nacional

Auto monitoramento por razões competitivas ou contratuais

Monitoramento por agências reguladoras ou seguros médicos

Ajustamento de desfecho em ensaios clínicos

Ajuda aos pacientes na seleção de diferentes hospitais

Nível médico (individual)

Melhoria de qualidade para médicos individuais

Uso institucional das informações de desfecho de médicos individuais

Ajuda aos pacientes na seleção de médicos individuais

Uso de desempenho inadequado persistente para sanções ou retirada de licença

Nível paciente

Ajuda o paciente no processo de tomada de decisão

Alocação de recursos

3.12.1) Benchmarking e indicador de qualidade

Embora a questão da qualidade no cuidado de saúde oferecido não seja nova, o foco atual está voltado para mensuração como ferramenta de melhoria dos serviços ofertados (6). A qualidade nos cuidados de saúde pode ser definida como o grau em que

os serviços de cuidados de saúde para indivíduos e populações aumentam a probabilidade do resultado de saúde desejado e são consistentes com o conhecimento profissional atual (1). Na terapia intensiva pediátrica, o objetivo principal é assegurar sobrevivência com mínima morbidade em crianças com doenças ameaçadoras da vida (8).

A melhoria da qualidade nos cuidados de saúde concentra-se em mover a curva de desempenho para frente (aumento de desempenho) pela adoção das melhores práticas (68), sendo a mensuração um dos aspectos fundamentais deste processo (69).

A mensuração da qualidade do serviço de saúde pode ser realizada através dos indicadores de qualidade, que podem ser divididos em dois grupos: indicadores de desfecho e indicadores de processo (70). Indicadores de processo investigam se o cuidado de saúde correto foi oferecido, enquanto os indicadores de desfecho avaliam como ficou o paciente após receber o cuidado de saúde (7).

O primeiro conjunto de indicadores de qualidade para Terapia Intensiva Pediátrica que foram submetidos à Joint Commission on Accreditation of Healthcare Organizations publicado em 2005 constitui um bom exemplo sobre o tema (6):

- Taxa de mortalidade padronizada (TMP)
- Tempo de permanência na UTI ajustada para a severidade
- Taxa de readmissões não planejadas
- Processo de avaliação de dor na admissão e periódica
- Adoção de práticas de segurança com o uso de medicações
- Adoção de práticas de prevenção de infecção de cateter venoso central

A taxa de mortalidade padronizada e o tempo de permanência na UTI ajustada para a severidade necessitam dos sistemas de ajustamento de risco para a realização das suas medidas, o que é alcançado através dos escores de predição de desfecho. Sem esse ajustamento uma comparação válida não pode ser realizada entre serviços que atendem pacientes com gravidade da doença e mistura de casos diferentes (8).

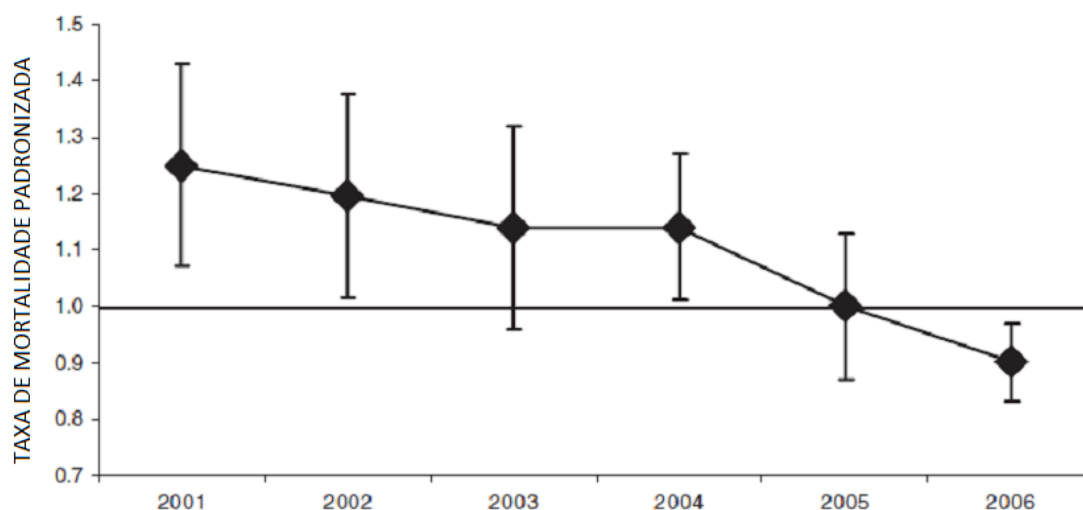
Como o principal objetivo da terapia intensiva é assegurar sobrevivência com mínima morbidade, a taxa de mortalidade padronizada constitui no principal indicador de qualidade utilizado (71). É calculada dividindo a taxa de mortalidade observada pela taxa de mortalidade prevista, conforme calculada pelos escores de predição de desfecho

(2). Entretanto a precisão desta inferência dependerá da habilidade dos escores para prever mortalidade (8).

A TMP é considerada um indicador de desfecho, medindo a efetividade do serviço em alcançar o principal objetivo da terapia intensiva, a sobrevivência (8). Este indicador deve ser apresentado com seu intervalo de confiança (IC) de 95% (72). Se o IC 95% da TMP incluir 1, o desempenho é considerado médio. Se o IC 95% da TMP não incluir 1, TMP menor do que 1 e maior do que 1 são considerados como tendo um desempenho bom ou ruim, respectivamente. Os intervalos de confiança do TMP são calculados em função do tamanho da amostra, sendo importante ao considerar uma taxa específica bem como a amplitude do intervalo de confiança (2).

Na figura abaixo é dado um exemplo hipotético da taxa de mortalidade padronizada em UTIP através de um período de tempo. A TMP desta unidade hipotética é significativamente pior do que o esperado em 2001, 2002 e 2004; não tem diferença significativa do esperado em 2003 e 2005; e é significativamente melhor do que o esperado em 2006 (2).

Gráfico 1 – Taxa de mortalidade padronizada de uma hipotética UTIP através do tempo.



O outro indicador de qualidade comumente utilizado na Terapia Intensiva é o tempo de permanência na UTI ajustada para a severidade, também chamado de taxa de duração de permanência padronizada (2). É calculada dividindo a duração de permanência observada pela duração de permanência esperada, conforme calculada pelos escores de predição de desfecho (2).

A taxa de duração de permanência padronizada é considerada um indicador de processo, medindo a eficiência do serviço (8). Esse indicador pode ser usado como um substituto da utilização de recursos, e assim podem ajudar a determinar que unidades utilizam uma menor quantidade de recursos para tratar comparáveis pacientes (73).

Embora ainda não implementado, a Joint Commission on Accreditation of Healthcare Organizations tem um plano para requerer que os hospitais nos Estados Unidos publiquem relatórios com sua mortalidade e duração de permanência ajustadas como ferramenta para melhoria de qualidade das UTI (16). No Brasil foi publicada em 2006 pela Agência Nacional de Vigilância Sanitária uma norma técnica (Regulamento Técnico para funcionamento de Unidades de Terapia Intensiva e Unidades de Cuidados Intermediários) que determina (74):

- As equipes das UTI e Unidades de Cuidados Intermediários (UCI) devem avaliar todo paciente por meio do Sistema de Classificação de Severidade da Doença ou outro Índice Prognóstico que venha substituí-lo.
- As equipes das UTI e UCI devem correlacionar a mortalidade geral de sua unidade à mortalidade geral esperada de acordo com Sistema de Classificação de Severidade da Doença ou outro Índice Prognóstico que venha substituí-lo, mantendo registro destes dados.

Porém a fiscalização para cumprimento da norma técnica geralmente não ocorre, conseqüentemente a maioria dos serviços de terapia intensiva no território brasileiro não utilizam uma ferramenta que pode ser útil no processo de melhoria dos cuidados de saúde oferecidos.

A partir da produção dos indicadores é possível realizar comparações entre desempenhos de unidades (2). Ao ressaltar diferenças de desempenhos entre UTI, indicando que unidades tem bom ou mau desempenho, abre-se uma janela de oportunidade para melhoria de qualidade (8).

Esta oportunidade surge ao permitir que se entenda porque certa unidade fornece um melhor desempenho. Ao identificar os motivos que produziram um melhor desempenho, essas informações podem ser compartilhadas com as outras unidades. Este processo é definido como Benchmarking e tem sido cada vez mais utilizado no processo de melhoria da qualidade nos cuidados de saúde (2).

Benchmarking pode ser definido como um processo em que níveis de desempenho de organizações são observados e então comparados com padrões internos ou externos, com a intenção de definir as “melhores práticas” (2). Pode ser externo ou interno, estabelecendo um padrão ou um ponto de referência contra o qual os outros são medidos (2).

O Benchmarking externo permite comparações diretas entre hospitais ou UTI, entretanto uma comparação indireta também pode ser realizada ao comparar o desempenho observado e o padrão histórico (2). Já o benchmarking interno permite aos médicos, UTI, hospitais ou sistemas de saúde realizar uma auto comparação, tipicamente ao longo do tempo (2).

O funcionamento adequado de um determinado Benchmarking depende de algumas características dos indicadores de qualidade utilizados para fazer as comparações (71):

- Clinicamente relevante.
- Válido (uma mudança na mensuração deve resultar numa mudança no desfecho de interesse).
- Robusto (uma pequena mudança na mensuração não deve resultar numa grande mudança no Benchmarking).
- Comparável (a mensuração é realizada da mesma maneira, e tem o mesmo significado entre as UTI comparadas).
- Padronizado (é possível comparar as mensurações entre diferentes hospitais).
- Anônimo (a mensuração não deve identificar pacientes ou profissionais envolvidos).
- Baseados em dados acessíveis, precisos e recolhidos regularmente.
- Fácil realização (devem produzir mínimo trabalho adicional).

A taxa de mortalidade padronizada e o tempo de permanência na UTI ajustada para a severidade possuem estas características e são os indicadores de qualidade mais comumente utilizados no processo de benchmarking (9). Já existem algumas organizações que oferecem a oportunidade de benchmarking para populações pediátricas e neonatais (3), porém não há informações sobre este tipo de organização no Brasil.

A Virtual Pediatric Intensive Care Systems (VPS) é uma dessas organizações, sendo constituída de uma rede de mais de 100 UTIP dos Estados Unidos que submetem

dados de cada admissão para um sistema de banco de dados eletrônico (8). Esse banco de dados é fonte para o entendimento do cuidado intensivo pediátrico, definindo práticas e padrões, gerando benchmarking e relatórios comparativos, e organizando colaboração em pesquisas (tabela 14). Atualmente, os bancos de dados da VPS contêm mais de 400.000 admissões, todos com dados ajustado por escores de predição de desfecho (40).

Tabela 14 – Mortalidade e taxa de mortalidade padronizada utilizando o PRISM III e o PIM 2 da população admitida em mais 90 UTIP dos Estados Unidos no ano de 2010. As populações se sobrepõem, mas não são idênticas (8).

	PRISM III	PIM 2
Total de casos	52.721	57.765
Total de mortes	1.719	1.678
Taxa de mortalidade	2,80%	2,90%
Mortalidade prevista	2,68%	3,21%
TMP	1,05	0,90

A aplicação apropriada do benchmarking em nível nacional ou regional pode fornecer informações confiáveis para sistemas públicos de saúde, seguradoras, fornecedores de cuidados de saúde e pacientes (9). Entretanto, é necessário suporte e pressão pelos governos, empresas prestadoras de serviços de saúde e hospitais associado a colaboração/adoção pelos profissionais responsáveis pela gestão das unidades (9).

Porém existem algumas limitações quando a qualidade dos cuidados de saúde oferecidos é avaliada através dos indicadores de qualidade que utilizam os escores de predição de desfecho para seu cálculo. O ajustamento de risco produzido pelos escores pode ser afetado por alguns fatores que não podem ser medidos normalmente, portanto influenciando em seus resultados (36).

O viés de tempo de execução (lead time) é um dos principais fatores que pode interferir nos resultados, sendo definido como a imprecisão na predição de risco que pode ocorrer quando o tratamento e a mensuração são iniciados em diferentes períodos (75). Este viés é um problema relevante nos escores de predição de desfecho utilizados na terapia intensiva, porque a maioria dos pacientes recebe alguma forma de tratamento (por exemplo, tratamento para estabilização recebido nos departamentos de emergência

ou enfermaria) antes da avaliação da gravidade de doença dentro da UTI (75). Com isso a TMP de certa UTI pode refletir o tratamento oferecido antes da admissão, em vez do desempenho da unidade onde o paciente foi admitido. Se o tratamento oferecido antes da admissão for adequado, o escore calculado pode sugerir uma menor gravidade e mortalidade do que é na verdade apresentado pelo paciente (75).

Pollack relata que as mensurações de tempo de permanência não quantificam as preferências dos médicos e pacientes, situação social do paciente, e a disponibilidade de leitos, fatores que podem interferir na decisão da alta para outra unidade (2).

Kahn e col demonstraram que o aumento da porcentagem de pacientes admitidos provenientes de outros hospitais pode significativamente enviesar a TMP calculada (76). Murphy-Filkins e col observaram que a proporção de diferentes tipos de pacientes (mistura de pacientes) pode influenciar o desempenho dos escores (59). Se certa porcentagem de um determinado tipo de paciente é ultrapassada, como pode ocorrer em UTI altamente especializada, o modelo pode não ser acurado.

Em decorrência desta possível imprecisão do ajustamento de risco produzido pelos escores, Wetzel e col sugerem que as comparações serão mais úteis e fidedignas quando as unidades avaliadas forem similares em relação a alguns fatores (por exemplo, mistura de casos, modelo de financiamento, tamanho da unidade, idade da população atendida) (8).

Pollack sugere que o mais importante uso dos métodos de ajustamento de risco é para benchmarking interno (34). A avaliação sequencial dentro de uma única unidade minimizaria as questões relacionadas a diferenças de amostras, integridade dos dados, e viés de tempo de execução, permitindo o acompanhamento do progresso da unidade (34). Como o objetivo é a melhoria contínua baseada em avaliações confiáveis de desempenho, os escores de acuidade podem ser considerados instrumentos úteis para realizar este objetivo (77).

3.12.2) Apoio à tomada de decisão clínica

A discussão sobre a utilização dos escores para auxílio de tomada de decisão clínica tem ganhado espaço na literatura. Porém a maioria dos estudos e debates está restrita a população adulta, sendo encontradas poucas referências sobre o tema voltado para a população pediátrica.

Três potenciais aplicações desses modelos para tomadas de decisão envolvendo pacientes individualmente tem sido propostas (78): ferramenta de triagem de pacientes para admissão em UTI; ferramenta de ajuda na decisão de alta da UTI; e ferramenta de auxílio aos médicos na discussão sobre limitação terapêutica em pacientes em UTI. Em decorrência da escassez de leitos de UTI em muitos hospitais, a prevenção de admissões desnecessárias e a transferência de pacientes que não necessitam de cuidados intensivos são iniciativas importantes (16).

Afessa e col sugerem que os escores de predição de desfecho para a população adulta Mortality Probability Model III (MPM III) e Simplified Acute Physiology Score III (SAPS III) têm o potencial para serem usados como ferramentas de apoio à decisão de admissão na UTI, já que a maioria de suas variáveis de predição é acessível na admissão (16). A racionalidade desta abordagem é que a estratificação de risco baseado nos escores de acuidade pode identificar pacientes que tem baixo risco de morte e em quem intervenções de “sustentação de vida” realizadas nas UTI são improváveis de serem requeridas (79). A identificação de tais pacientes pode permitir que seu tratamento seja realizado num ambiente fora da UTI, diminuindo a pressão por leitos nas UTI e de custos (internações em UTI produzem maiores custos financeiros do que em outros setores) (79). Porém não foram encontrados estudos avaliando os modelos como ferramenta de triagem para admissão, conforme relatado/proposto acima.

Outra abordagem para o uso dos modelos prognóstico como ferramenta de triagem tem sido proposta nas situações de epidemias e catástrofes com vítimas em massa, no qual a escassez de leitos pode ser ainda mais evidente.

No processo de organização de manejo da pandemia de influenza A H1N1, a Força Tarefa para Cuidados Intensivos em Massa do Departamento de Saúde dos Estados Unidos sugeriu o uso do escore de disfunção de órgãos para a população adulta Sequential Organ Failure Assessment (SOFA) no processo de triagem (80). O escore SOFA entraria como critério de exclusão do algoritmo de triagem, pacientes com alto risco de morte (80% de probabilidade de morrer pelo escore) não seriam admitidos para cuidados em UTI. Abordagem similar foi utilizada na recomendação do serviço de saúde público britânico sobre pandemia de influenza, no qual o escore SOFA é utilizado como parte de uma ferramenta de triagem para complementar o julgamento clínico (81).

Porém, Guest e col encontraram que a aplicação da ferramenta de triagem proposta pelo serviço de saúde britânico na população de um determinado hospital resultaria na recusa de admissão de 116 (46%) dos 255 pacientes admitidos no período

do estudo, dos quais 45 (39%) sobreviveram (82). A conclusão foi que a ferramenta de triagem proposta falhou em identificar adequadamente pacientes que poderiam se beneficiar de cuidados intensivos.

A Força Tarefa para epidemia de influenza e desastre em massa da Sociedade Europeia de Terapia Intensiva recomendou recentemente o uso de escore SOFA como ferramenta de priorização para admissão, porém enfatizando que esta abordagem tem suas limitações e que ainda não foi validada como instrumento de triagem (83).

A segunda proposta para uso dos modelos seria como ferramenta de auxílio na decisão de alta da UTI. Idealmente, a alta da UTI para uma unidade de menor nível de cuidado é apropriada quando o status fisiológico do paciente tenha estabilizado e a necessidade de monitoramento e cuidado intensivo não é mais necessário, ou quando o status fisiológico do paciente tenha deteriorado e intervenções ativas não são mais planejadas (84). Portanto, a adequação clínica da alta da UTI pode ser baseada na reversão da condição patológica aguda ou ausência de expectativas de reversão associado aos benefícios esperados da transferência (84). Seguindo esse raciocínio, alguns autores propõem a utilização dos escores na verificação se os pacientes apresentam as condições necessárias para alta da unidade (85).

Zimmerman e col combinaram o escore descritor de desfecho APACHE III e elementos do Therapeutic Intervention Severity Score (TISS), que é um escore para avaliar o nível de intervenções realizadas nos pacientes, para gerar uma estimativa de probabilidade de o paciente necessitar de tratamento ativo no próximo dia (85). Os autores sugeriram que esta estimativa poderia ser útil na tomada de decisão sobre alta, baseado na suposição que aqueles pacientes improváveis de requerer tratamento ativo de suporte a vida (baixo risco de mortalidade e baixo nível de intervenções) também seriam improváveis de serem readmitidos se recebessem alta da UTI. Entretanto, pelo fato que as altas da UTI no estudo não foram baseadas no risco previsto pelo escore, a utilidade desta abordagem permanece não provada (78).

A Força Tarefa para Cuidados Intensivos em Massa do Departamento de Saúde dos Estados Unidos, no documento sobre estratégias de manejo da pandemia de influenza A H1N1, também sugere o uso do escore SOFA como critério de exclusão para continuação de tratamento em UTI (80). O raciocínio empregado é que em pacientes com grande chance de morrer ($SOFA > 15$ em qualquer momento durante a internação) ou pacientes com necessidade de cuidados intensivos por tempo prolongado ($SOFA > 5$ por cinco dias ou mais com tendência de elevação ou estabilização) o

tratamento na UTI seria descontinuado, com o objetivo de realocar os escassos recursos disponíveis para outros pacientes com maiores chances de sobrevivência e com necessidade de menor tempo de internação (possibilitando que um maior número de vidas seja salvo).

Outra possível utilização dos escores de predição de desfecho relacionada a decisão de alta da UTI seria o uso dos modelos como fator de risco para readmissão na UTI após a alta da unidade. Frost e col publicaram uma meta-análise para avaliar a relação entre a gravidade da doença dos pacientes adultos na UTI e risco de readmissão na unidade durante a mesma hospitalização (86). A gravidade da doença medida na admissão ou na alta da UTI (utilizando os escores APACHE II, APACHE III, SAPS e SAPS II) foi maior em pacientes readmitidos do que nos não readmitidos. O risco de readmissão aumentava em 43% a cada aumento do desvio padrão no escore, independente se medido na admissão ou na alta.

A terceira possibilidade levantada para o uso dos modelos é como ferramenta de auxílio aos médicos na discussão sobre limitação terapêutica em pacientes em UTI. Os escores poderiam identificar um subgrupo de pacientes que seriam improváveis de sobreviver e em quem a continuação de tratamento na unidade poderia ser considerada fútil (87). A futilidade do tratamento pode ser definida quando o objetivo pretendido não pode ser alcançado (88), na terapia intensiva isto significaria que o tratamento oferecido não evitaria a morte do paciente antes da alta hospitalar (89).

Alguns pesquisadores têm utilizado os sistemas de escore como uma ferramenta para dar objetividade ao debate que envolve a definição de futilidade médica, sendo utilizado como critério de inclusão, e também existem estudos que tem avaliado se determinados valores de probabilidades de morte calculados pelos escores de predição de desfecho poderiam ser usados para definir limitação terapêutica (87).

Halevy e col em estudo para avaliar a frequência de pacientes adultos admitidos em UTI no qual o tratamento poderia ser considerado fútil utilizaram a probabilidade de morte maior que 90% calculada pelo APACHE II como um dos critérios de inclusão (90). Goh e col em estudo para determinar a frequência de pacientes pediátricos admitidos em UTIP no qual o tratamento poderia ser considerado fútil também utilizaram um risco morte maior que 90% calculada pelo PRISM II como um dos critérios de inclusão (91).

Afessa e col realizaram um estudo com objetivo para determinar se um aumento do escore APACHE III no terceiro dia de internação poderia identificar potenciais

tratamentos ineficazes (92). Foram avaliados 302 pacientes adultos com probabilidade de morte maior ou igual a 80 % na admissão. Em 34 pacientes (11,3%) ocorreu aumento do escore no terceiro dia de internação, e somente 2 dos 34 pacientes (6%) sobreviveram para alta hospitalar, comparado com uma taxa de sobrevivência de 43% (115/268) nos pacientes que não apresentaram aumento do escore no terceiro dia. Sendo concluído pelos autores que o escore pode identificar potenciais tratamentos ineficazes.

Porém existem alguns problemas com uso dos escores com esses objetivos (87). Os sistemas de escore quantificam o risco de morte, portanto para o uso desta metodologia seria necessário designar a probabilidade de morte (ponto de corte) no qual o tratamento do paciente seria considerado inútil (87). Essa tarefa é considerada complicada pelo fato que a probabilidade de morte calculada pelos escores de predição de desfecho tem desempenho inadequado em um nível individual, impedindo seu uso na tomada de decisão clínica (9).

Em declaração de consenso do Comitê de Ética da Sociedade de Cuidados Intensivos publicada em 1997 foi afirmado que o uso de sistemas de escores como um guia exclusivo para a tomada de decisões sobre iniciar ou continuar a prestar cuidados intensivos é inapropriado (88).

Entretanto existe um medicamento chamado drotrecogina alfa ativada, utilizado no tratamento de pacientes com sepse grave, que é licenciada nos Estados Unidos para uso somente em pacientes adultos que tenham um alto risco de morte, que são aqueles com escore APACHE II maior que 25 (36). Esta recomendação é resultado da análise de subgrupo dos dados do ensaio clínico PROWESS (Protein C Worldwide Evaluation in Severe Sepsis), que mostrou que pacientes com um escore APACHE II maior que 25 foram mais susceptíveis de beneficiar da droga (93).

Apesar das inúmeras utilidades propostas para o uso dos modelos com relação a tomadas de decisão envolvendo pacientes individualmente, estudos validando estas propostas ainda não foram publicados. Com isso a maioria da literatura disponível é contra essa possível utilização.

Afessa e col relatam que embora a maioria dos modelos prognósticos tenha um bom desempenho ao nível da população (grupo), sua pobre calibração em um nível individual tem impedido o seu uso na prática clínica diária (9). Mesmo modelos de gravidade de doença que demonstram bom desempenho para descrever pacientes em conjunto, não apresentam mesmo desempenho em nível individual (9). Lemeshow e col relatam que as probabilidades de mortalidade hospitalar podem fornecer informações

significativas aos médicos quando se discute o prognóstico com o pacientes e/ou seus familiares, entretanto estas probabilidades não devem ser empregadas para tomada de decisões de tratamento em nível individual (94).

Keegan e col relatam que os modelos prognósticos terão de serem submetidos ao mesmo tipo de avaliação rigorosa utilizada com medicamentos antes de serem utilizados em decisões que tem impacto na prestação de serviços de saúde e assistência individual (16). Keegan e col citam um editorial (95) no qual é afirmado que enquanto aguardamos estudos demonstrando melhoria nas decisões médicas com uso dos modelos prognóstico, estes devem ser usados como “o homem bêbado usa o poste, para apoio em vez de iluminação” na tomada de uma decisão clínica.

3.12.3) Uso em pesquisa científica

Os sistemas de escore estão sendo cada vez mais incorporados aos desenhos de estudos clínicos. Os modelos têm sido usados para comparar populações de pacientes em estudos clínicos e até mesmo para identificação de pacientes elegíveis para inclusão (36).

O uso mais comum dos sistemas de escore (e com menos controvérsias sobre sua adequabilidade) é relacionado à função de comparação de grupos. Os escores de predição de desfecho são frequentemente usados para verificar se a randomização funcionou e se os grupos de um ensaio clínico randomizado são equiparáveis no tempo zero, enquanto os escores de disfunção de órgãos podem ser usados no mesmo cenário para descrever mudanças após o tempo zero nas condições clínicas dos participantes do ensaio clínico randomizado (10).

O principal desfecho utilizado em estudos realizados na terapia intensiva adulta e pediátrica é a mortalidade, em decorrência do fato que a morte é um dado concreto e que é muito difícil enviesar este desfecho (10). Porém tem aumentado a percepção que a mortalidade por si só é inadequada como medida de desfecho para estudos intervencionais em pacientes adultos, o que tem levado estudos a incluir um escore de disfunção de órgãos como parte da avaliação do paciente, para que o efeito sobre a morbidade também possa ser avaliada (36).

Na população pediátrica a utilização da mortalidade como única medida de desfecho tem criado um problema de viabilidade na realização de ensaios clínicos randomizados (58). A taxa de mortalidade é baixa se comparada a população adulta: 20

a 40% em UTI para pacientes adultos, enquanto é comumente menor que 10 % em UTIP (10). Esta baixa taxa de mortalidade acarreta na necessidade de aumento significativo no tamanho da amostra requerida para realizar estudos pediátricos completos (58). O número de pacientes requeridos pode ser tão grande que pode ser impossível alcançá-los (10).

A incidência da taxa de SDMO nas UTIP é entre 18 e 25%, sendo significativamente maior do que a taxa de mortalidade (10). Portanto, faz sentido o questionamento se os escores de descritores podem ser bons substitutos do desfecho de morte para estudos clínicos realizados em UTIP (10).

Em decorrência destes achados a disfunção de órgãos e subsequentemente os escores de disfunção de órgãos estão sendo considerados como substitutos do desfecho de mortalidade em estudos clínicos realizados em UTIP, já que possuem as características necessárias para realizar esta função (10): a) há uma relação entre a medida padrão ouro e o desfecho substituto, pela observação de que SDMO é presente em quase todas as crianças criticamente doentes que morrem; b) a taxa de prevalência ou incidência do desfecho substituto deve ser significativamente maior do que a medida padrão ouro (a taxa de SDMO é quase três vezes maior que a mortalidade nas UTIP); e c) o valor científico do desfecho substituto deve ser bem avaliado, o que ocorreu com as validações realizadas com o PELOD e PMOD.

A outra utilização dos sistemas de escore durante o desenvolvimento de pesquisas médicas é na identificação de pacientes elegíveis para estudos clínicos, porém o uso desta metodologia levanta questões sobre sua adequabilidade (96).

Esta função já foi aplicada a drogas disponíveis e atualmente esta sendo usada em ensaios terapêuticos de novos agentes (96). A drotrecogina alfa ativada é um exemplo de como os sistemas de escore são utilizados neste contexto. Esta droga é licenciada nos Estados Unidos para uso somente em pacientes adultos com sepse grave que tenham um alto risco de morte, que são aqueles com escore APACHE II maior que 25. Esta recomendação é resultado da análise de subgrupo dos dados do ensaio clínico PROWESS (Protein C Worldwide Evaluation in Severe Sepsis), que mostrou que pacientes com um escore APACHE II maior que 25 foram mais susceptíveis de beneficiar da droga (93).

Uma abordagem do uso dos sistemas de escore para identificação de pacientes elegíveis para estudos clínicos considera uma relação linear entre a gravidade da doença e o potencial para responder a uma intervenção. A racionalidade desta abordagem é a

proposição que a intervenção deve ser limitada ao grupo de pacientes mais doentes, que presumivelmente seriam os mais prováveis de se beneficiarem (96).

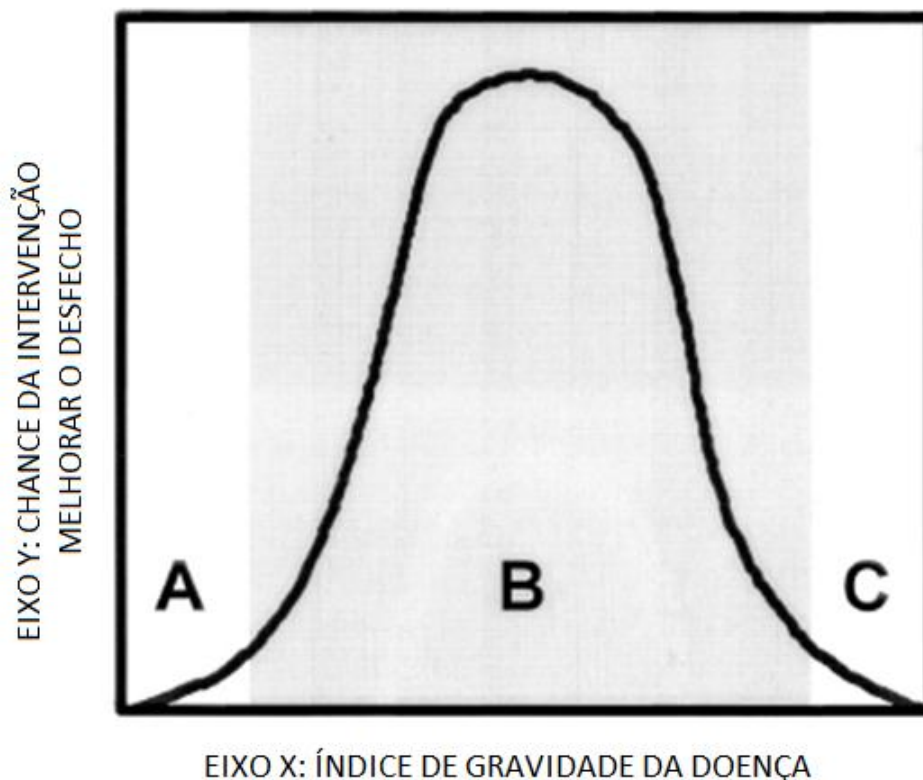
Eichacker e col relataram que a eficácia do tratamento com drogas anti-inflamatórias durante o tratamento de sepse foi dependente do risco de morte, com a observação de que o tratamento foi mais eficaz em pacientes com choque séptico com maiores chances de morrer e foi prejudicial naqueles com baixo risco (97). Sendo sugerido que ensaios clínicos subsequentes deveriam incorporar critérios de inclusão e exclusão tais como os escores de predição de desfecho para garantir que os pacientes prováveis de serem beneficiados sejam incluídos e aqueles prováveis de serem prejudicados sejam excluídos.

Porém esta relação linear entre gravidade e potencial resposta a terapia observada não é encontrada em todos os casos. Em um estudo sobre a relação entre o prognóstico de pacientes com pneumonia associada à ventilação mecânica, uso de antibióticos adequados e gravidade da doença do paciente, o tratamento empírico inadequado só foi associado com mau prognóstico em pacientes adultos com o escore de disfunção de órgãos Logistic Organ Dysfunction (LOD) menor ou igual a quatro (indicativo de baixo risco de mortalidade), portanto antibióticos podem ser mais efetivos em pacientes menos gravemente doentes (98). Em outro estudo sobre estratégia restritiva ou liberal de transfusão de concentrado de hemácias em pacientes criticamente doentes, foi observado que a mortalidade foi significativamente menor com a estratégia restritiva entre pacientes que eram menos gravemente doentes (APACHE II menor ou igual a 20) (99).

Outra abordagem do uso dos sistemas de escore para identificação de pacientes elegíveis para estudos clínicos considera a relação entre a gravidade da doença e o potencial para responder a uma intervenção como tendo uma forma de sino ao invés de linear (96). Nos extremos da gravidade da doença, a chance de detectar um benefício de sobrevivência de qualquer nova terapia é limitada com pacientes com alta probabilidade de sobreviver (gravidade baixa) ou de morrer (gravidade alta), independente da intervenção terapêutica (96).

Cohen e col em artigo sobre estratégias para desenvolvimento de estudos em pacientes com sepse relatam que a chance de uma determinada intervenção melhorar o desfecho de casos de sepse leve ou extremamente grave (zonas A e C de gráfico 2) é tão pequena, que é pensado por alguns especialistas que tais casos devem ser excluídos dos ensaios clínicos (100).

Gráfico 2 – Figura em forma de sino da relação entre a gravidade da doença e o potencial para responder a uma intervenção.



Porém também há autores que são contra o uso dos sistemas de escore na identificação de pacientes elegíveis para estudos clínicos, Vincent e colaboradores levantam alguns argumentos (96):

- a) Escores de predição de desfecho não foram projetados para esta finalidade, foram desenvolvidos para avaliar o risco de morte hospitalar e calibrados para maximizar a capacidade preditiva através de um espectro de doença ao invés de detectar diferenças de resposta clínica a uma intervenção específica para uma determinada doença.
- b) Escores de predição de desfecho genéricos não são medidas exclusivas de gravidade de disfunção fisiológica, são avaliadores de risco que combinam medidas de gravidade potencialmente modificáveis (escore de fisiologia aguda) e fatores de risco não modificáveis como idade e comorbidades. O valor do escore pode ser profundamente influenciado pelos fatores não modificáveis, mas

isto não necessariamente reflete no potencial para benefício de uma intervenção terapêutica. O escore APACHE utilizado para a população adulta indica pontos para a idade e comorbidades, com isso quando os outros componentes do escore tiveram valores semelhantes, os pacientes mais velhos terão uma maior pontuação do que os mais novos. Portanto, pacientes mais velhos terão intrinsecamente uma maior probabilidade de receber intervenções que são reservadas para uso em pacientes que apresentam um determinado ponto de corte.

- c) Os sistemas de escores são observador-dependentes, com alguns estudos mostrando que médicos diferentes calculam diferentes valores do escore do mesmo dado (101–103), e até o mesmo médico pode calcular diferentes valores quando revisa o mesmo dado em diferentes momentos (104).
- d) Os escores de predição de desfecho foram desenhados e validados para serem calculados com dados obtidos das primeiras 24 horas na UTI. Alguns estudos estão usando estes escores calculados com dados do período fora das 24 horas iniciais de sua chegada à UTI, neste período muitas das anormalidades fisiológicas e bioquímicas podem ter sido corrigidas ou controladas, alterando o cálculo do escore. Além disso, o uso destes escores quando o paciente já está na unidade nunca foi apropriadamente validado.

Por fim Vincent e col concluem que os escores de disfunção de órgãos podem ser melhores dos que os escores de predição de desfecho na função de identificação de pacientes elegíveis para estudos clínicos, porém afirmam que isto não é provado e necessita de mais investigação (96). O uso de escores de disfunção de órgãos pode evitar alguns problemas relacionados aos pontos dados à comorbidades (não fazem parte das variáveis coletadas) e período de coleta de dados restrito às 24 horas iniciais (o escore é calculado diariamente), mas algumas das limitações relatadas também podem ser aplicadas aos mesmos (96).

3.13) Limitações dos sistemas de escores de acuidade

Existem algumas limitações inerentes aos sistemas de escore de acuidade (9). Entre as limitações encontradas estão erros e vieses relacionados à mistura de casos, erros na coleta e registro dos dados, falhas no desenvolvimento e validação dos modelos (9).

Um sistema de escore de acuidade somente prediz com precisão a mortalidade para um grupo de pacientes se a mistura de casos for similar à utilizada no desenvolvimento do modelo (9). Vincent e col relatam que todos os sistemas de escores de acuidade somente predizem da melhor forma o comportamento de um grupo de pacientes que exatamente combinam com os pacientes da amostra de desenvolvimento dos modelos (36). Essa hipótese é fortalecida pela observação realizada por Murphy-Filkins e col de que quando certa porcentagem de um determinado tipo de paciente é ultrapassada, o modelo perde precisão (59).

Outra observação que fortalece esta hipótese é que embora os modelos tenham sido desenvolvidos a partir de grandes amostras populacionais, em quase todos os casos quando eles são aplicados a novas populações a calibração se deteriora, embora a discriminação dificilmente modifique (36). Slater e col compararam o desempenho do PRISM III e PIM2 numa amostra populacional diferente da amostra de desenvolvimento dos modelos, sendo observado que os dois escores apresentaram capacidade de discriminação adequada, porém calibração inadequada (26).

Fatores como viés de tempo de execução, localização pré-admissão, diagnósticos agudos, reserva fisiológica, e preferências dos pacientes para limitação terapêutica também podem influenciar a taxa de mortalidade (16), todos podendo ser considerados relacionados à mistura de casos.

A precisão de qualquer sistema de escore de acuidade também é altamente dependente da qualidade da coleta e registro dos dados (36). Podem ocorrer diferenças sistemáticas na forma que os escores são aplicados na prática clínica em relação ao momento do desenvolvimento e validação que interfiram na confiabilidade dos resultados, como (87): definições e regras da coleta de dados, critérios de inclusão/exclusão de pacientes, período de tempo para coleta de dados, método de coleta de dados, mensuração do desfecho, e a natureza da manipulação dos dados antes da análise. Por exemplo, a coleta de dados por período maior ou menor do que o estipulado pelo escore (escores que determinam que os dados sejam coletados durante a primeira hora após a admissão) podem sistematicamente sub ou superestimar a probabilidade do desfecho.

As etapas do desenvolvimento e validação dos sistemas de escore de acuidade podem conter algumas falhas de metodologia que influenciarão na sua confiabilidade. Todos os modelos para população adulta foram desenvolvidos em UTI selecionadas sem randomização, comprometendo a generalização dos achados (16), situação também

presente nos escores para a população pediátrica. Bouch e col relatam que os escores devem ser desenvolvidos e validados contra uma coorte real de pacientes, mas é difícil ter certeza se a coorte utilizada é representativa de toda a população de pacientes criticamente enfermos, sendo questionado se é possível existir uma população representativa de UTI (5).

A metodologia estatística utilizada para avaliar a calibração, geralmente o teste de Hosmer-Lemeshow, pode ser influenciada por vários fatores, incluindo o número de covariáveis e tamanho da amostra, acarretando em possíveis limitações de interpretação de desempenho (9,16).

A forma de validação dos escores pode ser outro ponto de questionamento, escores que não realizaram validação externa podem ter comprometimento de seus achados (33). Tamanho da amostra também tem grande influência na validade dos escores: amostras pequenas tendem a prejudicar a capacidade do modelo de distinguir e avaliar com segurança diferentes grupos de pacientes (5). Amostras grandes são claramente necessárias, mas o tamanho desta população não é conhecido (5).

Também podem ocorrer limitações na interpretação dos resultados, relacionada à possibilidade dos escores tornarem-se obsoletos (87). Dado o significativo período de tempo entre o desenvolvimento do modelo e sua utilização, é possível que muitos fatores (práticas de manejo e tecnologias) que influenciam a relação das variáveis e desfecho tenham mudado, alterando a precisão dos modelos (5). O PIM2 (15), que é escore de predição de desfecho mais recente para a população pediátrica, é um bom exemplo. Os dados utilizados para seu desenvolvimento são do período entre 1997 e 1999.

Além das possíveis limitações comentadas nesta seção, há de se destacar as limitações relacionadas aos potenciais utilizações dos sistemas de escore de acuidade já relatadas na seção referente ao tema.

3.14) Conclusão

Os escores de acuidade utilizados na terapia intensiva têm o objetivo de prever desfecho (escores de predição de desfecho) ou avaliar disfunção de órgãos (escores de disfunção de órgãos).

Porém desde o desenvolvimento dos primeiros escores outros potenciais usos têm sido propostos, com destaque para utilização como ferramenta de comparação de

desempenhos entre UTI (benchmarking) e para comparação de grupos durante a randomização em pesquisa científica. Entretanto é importante reconhecer as limitações inerentes a todos os escores, com o intuito de poder analisar e utilizar as valiosas informações fornecidas por esses modelos de forma correta.

Na última década os esforços para melhoria da qualidade do cuidado de saúde tem ganhado destaque, com grande interesse na utilização dos indicadores de qualidade como ferramenta deste processo. Com isso os escores de predição de desfecho ganham importância, já que são os instrumentos que permitem o ajuste de risco dos pacientes, permitindo comparações válidas entre práticas e serviços.

A terapia intensiva pediátrica já conta com escores adequadamente desenvolvidos e validados que podem auxiliar nesse processo contínuo de melhoria da qualidade de cuidados de saúde oferecidos, abrindo-se uma janela de oportunidades para os profissionais de saúde envolvidos com o cuidado de pacientes pediátricos criticamente doentes.

3.15) Bibliografia

1. Sachdeva RC, Jain S. Making the Case to Improve Quality and Reduce Costs in Pediatric Health Care. *Pediatric Clinics of North America*. 2009 Aug;56(4):731–43.
2. Marcin JP, Pollack MM. Review of the acuity scoring systems for the pediatric intensive care unit and their use in quality improvement. *Journal of intensive care medicine*. 2007 May-Jun;22(3):131–40.
3. Slonim AD, Pollack MM. Integrating the Institute of Medicine's six quality aims into pediatric critical care: relevance and applications. *Pediatric critical care medicine : a journal of the Society of Critical Care Medicine and the World Federation of Pediatric Intensive and Critical Care Societies*. 2005 May;6(3):264–9.
4. Beal AC, Co JPT, Dougherty D, Jorsling T, Kam J, Perrin J, et al. Quality measures for children's health care. *Pediatrics*. 2004 Jan;113(1 Pt 2):199–209.
5. Bouch DC, Thompson JP. Severity scoring systems in the critically ill. *Continuing Education in Anaesthesia, Critical Care & Pain*. 2008 Aug 22;8(5):181–5.
6. Scanlon MC, Mistry KP, Jeffries HE. Determining pediatric intensive care unit quality indicators for measuring pediatric intensive care unit safety. *Pediatric critical care medicine : a journal of the Society of Critical Care Medicine and the*

- World Federation of Pediatric Intensive and Critical Care Societies. 2007 Mar;8(2 Suppl):S3–10.
7. Kuhlthau K, Ferris TGG, Iezzoni LI. Risk adjustment for pediatric quality indicators. *Pediatrics*. 2004 Jan;113(1 Pt 2):210–6.
 8. Wetzel RC, Sachedeva R, Rice TB. Are all ICUs the same? *Paediatric anaesthesia*. 2011 Jul;21(7):787–93.
 9. Afessa B, Gajic O, Keegan MT. Severity of illness and organ failure assessment in adult intensive care units. *Critical care clinics*. 2007 Jul;23(3):639–58.
 10. Lacroix J, Cotting J. Severity of illness and organ dysfunction scoring in children. *Pediatric critical care medicine : a journal of the Society of Critical Care Medicine and the World Federation of Pediatric Intensive and Critical Care Societies*. 2005 May;6(3 Suppl):S126–34.
 11. Moons KGM, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? *BMJ*. 2009 Feb 23;338(feb23 1):b375–b375.
 12. Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in medicine*. 1996 Feb 28;15(4):361–87.
 13. Laupacis A, Sekar N, Stiell IG. Clinical prediction rules. A review and suggested modifications of methodological standards. *JAMA : the journal of the American Medical Association*. 1997 Feb 12;277(6):488–94.
 14. Yeh TS, Pollack MM, Ruttimann UE, Holbrook PR, Fields AI. Validation of a physiologic stability index for use in critically ill infants and children. *Pediatric research*. 1984 May;18(5):445–51.
 15. Slater A, Shann F, Pearson G. PIM2: a revised version of the Paediatric Index of Mortality. *Intensive care medicine*. 2003 Feb;29(2):278–85.
 16. Keegan MT, Gajic O, Afessa B. Severity of illness scoring systems in the intensive care unit. *Critical care medicine*. 2011 Jan;39(1):163–9.
 17. Zimmerman JE, Kramer AA, McNair DS, Malila FM, Shaffer VL. Intensive care unit length of stay: Benchmarking based on Acute Physiology and Chronic Health Evaluation (APACHE) IV. *Critical care medicine*. 2006 Oct;34(10):2517–29.
 18. Ruttimann UE, Patel KM, Pollack MM. Length of stay and efficiency in pediatric intensive care units. *The Journal of pediatrics*. 1998 Jul;133(1):79–85.
 19. Watts CM, Knaus WA. The case for using objective scoring systems to predict intensive care unit outcome. *Critical care clinics*. 1994 Jan;10(1):73–89; discussion 91–2.

20. Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology*. 1983 Sep;148(3):839–43.
21. Lemeshow S, Le Gall JR. Modeling the severity of illness of ICU patients. A systems update. *JAMA : the journal of the American Medical Association*. 1994 Oct 5;272(13):1049–55.
22. Lemeshow S, Hosmer DW. A review of goodness of fit statistics for use in the development of logistic regression models. *American journal of epidemiology*. 1982 Jan;115(1):92–106.
23. Tibby SM, Murdoch IA. Calibration of the paediatric index of mortality score for UK paediatric intensive care. *Archives of disease in childhood*. 2002 Jan;86(1):65; author reply 65–6.
24. Eulmesekian PG, Pérez A, Minces PG, Ferrero H. Validation of pediatric index of mortality 2 (PIM2) in a single pediatric intensive care unit of Argentina. *Pediatric critical care medicine : a journal of the Society of Critical Care Medicine and the World Federation of Pediatric Intensive and Critical Care Societies*. 2007 Jan;8(1):54–7.
25. Brady AR, Harrison D, Black S, Jones S, Rowan K, Pearson G, et al. Assessment and optimization of mortality prediction tools for admissions to pediatric intensive care in the United kingdom. *Pediatrics*. 2006 Apr;117(4):e733–42.
26. Slater A, Shann F. The suitability of the Pediatric Index of Mortality (PIM), PIM2, the Pediatric Risk of Mortality (PRISM), and PRISM III for monitoring the quality of pediatric intensive care in Australia and New Zealand. *Pediatric critical care medicine : a journal of the Society of Critical Care Medicine and the World Federation of Pediatric Intensive and Critical Care Societies*. 2004 Sep;5(5):447–54.
27. Ridley S. Severity of illness scoring systems and performance appraisal. *Anaesthesia*. 1998 Dec;53(12):1185–94.
28. Zhu BP, Lemeshow S, Hosmer DW, Klar J, Avrunin J, Teres D. Factors affecting the performance of the models in the Mortality Probability Model II system and strategies of customization: a simulation study. *Critical care medicine*. 1996 Jan;24(1):57–63.
29. Shann F. Are we doing a good job: PRISM, PIM and all that. *Intensive care medicine*. 2002 Feb;28(2):105–7.
30. Bertolini G, D'Amico R, Nardi D, Tinazzi A, Apolone G. One model, several results: the paradox of the Hosmer-Lemeshow goodness-of-fit test for the logistic regression model. *Journal of epidemiology and biostatistics*. 2000 Jan;5(4):251–3.

31. Mourouga P. Does it fit? Is it good? Assessment of scoring systems. *Current Opinion in Critical Care*. 2000;6:176–80.
32. Altman DG, Royston P. What do we mean by validating a prognostic model? *Statistics in medicine*. 2000 Feb 29;19(4):453–73.
33. Altman DG, Vergouwe Y, Royston P, Moons KGM. Prognosis and prognostic research: validating a prognostic model. *BMJ*. 2009 May 28;338(may28 1):b605–b605.
34. Pollack MM. It is what it was. *Pediatric critical care medicine : a journal of the Society of Critical Care Medicine and the World Federation of Pediatric Intensive and Critical Care Societies*. 2011 Mar;12(2):228–9.
35. Pollack MM, Patel KM, Ruttimann UE. PRISM III: an updated Pediatric Risk of Mortality score. *Critical care medicine*. 1996 May;24(5):743–52.
36. Vincent J-L, Moreno R. Clinical review: scoring systems in the critically ill. *Critical care (London, England)*. 2010 Jan;14(2):207.
37. Pollack MM, Ruttimann UE, Getson PR. Pediatric risk of mortality (PRISM) score. *Critical care medicine*. 1988 Nov;16(11):1110–6.
38. Shann F, Pearson G, Slater A, Wilkinson K. Paediatric index of mortality (PIM): a mortality prediction model for children in intensive care. *Intensive care medicine*. 1997 Feb;23(2):201–7.
39. Poonam Bhadoria AGB. Severity Scoring Systems in Paediatric Intensive Care Units. *Indian Journal of Anaesthesia*. 2008;52(Suppl (5)):663–75.
40. Gorman M, Sachdeva R, Rice T, Wetzel R. Support for PICUs and the PRISM scoring algorithms. *Pediatric critical care medicine : a journal of the Society of Critical Care Medicine and the World Federation of Pediatric Intensive and Critical Care Societies*. 2009 Jan;10(1):143–4.
41. Mestrovic J, Kardum G, Polic B, Omazic A, Stricevic L, Sustic A. Applicability of the Australian and New Zealand Paediatric Intensive Care Registry diagnostic codes and Paediatric Index of Mortality 2 scoring system in a Croatian paediatric intensive care unit. *European journal of pediatrics*. 2005 Dec;164(12):783–4.
42. Thukral A, Lodha R, Irshad M, Arora NK. Performance of Pediatric Risk of Mortality (PRISM), Pediatric Index of Mortality (PIM), and PIM2 in a pediatric intensive care unit in a developing country. *Pediatric critical care medicine : a journal of the Society of Critical Care Medicine and the World Federation of Pediatric Intensive and Critical Care Societies*. 2006 Jul;7(4):356–61.
43. Prieto Espuñes S, López-Herce Cid J, Rey Galán C, Medina Villanueva A, Concha Torre A, Martínez Cambolor P. [Prognostic indexes of mortality in pediatric intensive care units]. *Anales de pediatría (Barcelona, Spain : 2003)*. 2007 Apr;66(4):345–50.

44. Qureshi AU, Ali AS, Ahmad TM. Comparison of three prognostic scores (PRISM, PELOD and PIM 2) at pediatric intensive care unit under Pakistani circumstances. *Journal of Ayub Medical College, Abbottabad : JAMC*. 2007 Apr-Jun;19(2):49–53.
45. Wolfler A, Silvani P, Musicco M, Salvo I. Pediatric Index of Mortality 2 score in Italy: a multicenter, prospective, observational study. *Intensive care medicine*. 2007 Aug;33(8):1407–13.
46. Hwang HS, Lee NY, Han SB, Kwak GY, Lee SY, Chung SY, et al. Performance effectiveness of pediatric index of mortality 2 (PIM2) and pediatric risk of mortality III (PRISM III) in pediatric patients with intensive care in single institution: Retrospective study. *Korean Journal of Pediatrics*. 2008;51(11):1158.
47. Canonero I, Figueroa A, Cacciamano A, Olivier E, Cuestas E. [Validation of PRISM and PIM2 scores of mortality in a pediatric intensive care unit in Cordoba]. *Archivos argentinos de pediatría*. 2010 Oct;108(5):427–33.
48. Hariharan S, Krishnamurthy K, Grannum D. Validation of Pediatric Index of Mortality-2 scoring system in a pediatric intensive care unit, Barbados. *Journal of tropical pediatrics*. 2011 Feb;57(1):9–13.
49. Ng DK, Miu T-yat, Chiu W-keung, Hui N-tat, Chan C-hong. Validation of Pediatric Index of Mortality 2 in three pediatric intensive care units in Hong Kong. *Indian journal of pediatrics*. 2011 Dec;78(12):1491–4.
50. Imamura T, Nakagawa S, Goldman RD, Fujiwara T. Validation of pediatric index of mortality 2 (PIM2) in a single pediatric intensive care unit in Japan. *Intensive care medicine* . 2012 Apr;38(4):649-54
51. Czaja AS, Scanlon MC, Kuhn EM, Jeffries HE. Performance of the Pediatric Index of Mortality 2 for pediatric cardiac surgery patients. *Pediatric critical care medicine : a journal of the Society of Critical Care Medicine and the World Federation of Pediatric Intensive and Critical Care Societies*. 2011 Mar;12(2):184–9.
52. Garcia PCR, Piva JP. Pediatric Index of Mortality 2 (PIM2)--A prognostic tool for developing countries: Easy, efficient, and free! *Pediatric critical care medicine : a journal of the Society of Critical Care Medicine and the World Federation of Pediatric Intensive and Critical Care Societies*. 2007 Jan;8(1):77–8.
53. Leteurtre S, Martinot A, Duhamel A, Gauvin F, Grandbastien B, Nam TV, et al. Development of a pediatric multiple organ dysfunction score: use of two strategies. *Medical decision making : an international journal of the Society for Medical Decision Making*. 1999 Oct-Dec;19(4):399-410.
54. Leteurtre S, Martinot A, Duhamel A, Proulx F, Grandbastien B, Cotting J, et al. Validation of the paediatric logistic organ dysfunction (PELOD) score: prospective, observational, multicentre study. *Lancet*. 2003 Jul 19;362(9379):192–7.

55. Leteurtre S, Duhamel A, Grandbastien B, Lacroix J, Leclerc F. Paediatric logistic organ dysfunction (PELOD) score. *Lancet*. 2006 Mar 18;367(9514):897; author reply 900–2.
56. Graciano AL, Balko JA, Rahn DS, Ahmad N, Giroir BP. The Pediatric Multiple Organ Dysfunction Score (P-MODS): Development and validation of an objective scale to measure the severity of multiple organ dysfunction in critically ill children*. *Critical Care Medicine*. 2005 Jul;33(7):1484–91.
57. Marshall JC, Cook DJ, Christou NV, Bernard GR, Sprung CL, Sibbald WJ. Multiple organ dysfunction score: a reliable descriptor of a complex clinical outcome. *Critical care medicine*. 1995 Oct;23(10):1638–52.
58. Dominguez TE, Huh JW. Do we need another pediatric severity of illness score? *Critical care medicine*. 2005 Jul;33(7):1643–5.
59. Murphy-Filkins R, Teres D, Lemeshow S, Hosmer DW. Effect of changing patient mix on the performance of an intensive care unit severity-of-illness model: how to distinguish a general from a specialty intensive care unit. *Critical care medicine*. 1996 Dec;24(12):1968–73.
60. Beck DH, Smith GB, Pappachan JV. The effects of two methods for customising the original SAPS II model for intensive care patients from South England. *Anaesthesia*. 2002 Aug;57(8):785–93.
61. Moreno R, Apolone G. Impact of different customization strategies in the performance of a general severity score. *Critical care medicine*. 1997 Dec;25(12):2001–8.
62. Le Gall JR, Neumann A, Hemery F, Bleriot JP, Fulgencio JP, Garrigues B, et al. Mortality prediction using SAPS II: an update for French intensive care units. *Critical care (London, England)*. 2005 Jan;9(6):R645–52.
63. Rivera-Fernández R, Vázquez-Mata G, Bravo M, Aguayo-Hoyos E, Zimmerman J, Wagner D, et al. The Apache III prognostic system: customized mortality predictions for Spanish ICU patients. *Intensive care medicine*. 1998 Jun;24(6):574–81.
64. Ohno-Machado L, Resnic FS, Matheny ME. Prognosis in critical care. *Annual review of biomedical engineering*. 2006 Jan;8:567–99.
65. Zimmerman JE, Kramer AA, McNair DS, Malila FM. Acute Physiology and Chronic Health Evaluation (APACHE) IV: hospital mortality assessment for today's critically ill patients. *Critical care medicine*. 2006 May;34(5):1297–310.
66. Moreno RP, Metnitz PGH, Almeida E, Jordan B, Bauer P, Campos RA, et al. SAPS 3--From evaluation of the patient to evaluation of the intensive care unit. Part 2: Development of a prognostic model for hospital mortality at ICU admission. *Intensive care medicine*. 2005 Oct;31(10):1345–55.

67. Higgins TL, Teres D, Copes WS, Nathanson BH, Stark M, Kramer AA. Assessing contemporary intensive care unit outcome: an updated Mortality Probability Admission Model (MPM0-III). *Critical care medicine*. 2007 Mar;35(3):827–35.
68. Miles P. Health information systems and physician quality: role of the American board of pediatrics maintenance of certification in improving children's health care. *Pediatrics*. 2009 Jan;123 Suppl:S108–10.
69. Randolph G, Esporas M, Provost L, Massie S, Bundy DG. Model for improvement - Part Two: Measurement and feedback for quality improvement efforts. *Pediatric clinics of North America*. 2009 Aug;56(4):779–98.
70. Brook RH, McGlynn EA, Cleary PD. Quality of health care. Part 2: measuring quality of care. *The New England journal of medicine*. 1996 Sep 26;335(13):966–70.
71. Woodhouse D, Berg M, van der Putten J, Houtepen J. Will benchmarking ICUs improve outcome? *Current opinion in critical care*. 2009 Oct;15(5):450–5.
72. Hosmer DW, Lemeshow S. Confidence interval estimates of an index of quality performance based on logistic regression models. *Statistics in medicine*. 1995 Oct 15;14(19):2161–72.
73. Iwashyna TJ, Kramer AA, Kahn JM. Intensive care unit occupancy and patient outcomes. *Critical care medicine*. 2009 May;37(5):1545–57.
74. Agência Nacional de Vigilância Sanitária. Agência Nacional de Vigilância Sanitária - Regulamento Técnico para funcionamento de Unidades de Terapia Intensiva e Unidades de Cuidados Intermediários. 2006.
75. Tunnell RD, Millar BW, Smith GB. The effect of lead time bias on severity of illness scoring, mortality prediction and standardised mortality ratio in intensive care--a pilot study. *Anaesthesia*. 1998 Nov;53(11):1045–53.
76. Kahn JM, Kramer AA, Rubenfeld GD. Transferring critically ill patients out of hospital improves the standardized mortality ratio: a simulation study. *Chest*. 2007 Jan;131(1):68–75.
77. Afessa B, Keegan MT, Hubmayr RD, Naessens JM, Gajic O, Long KH, et al. Evaluating the performance of an institution using an intensive care unit benchmark. *Mayo Clinic proceedings*. Mayo Clinic. 2005 Feb;80(2):174–80.
78. Hyzy RC. ICU Scoring and Clinical Decision Making. *Chest*. 1995 Jun 1;107(6):1482–3.
79. Zimmerman JE, Wagner DP, Knaus WA, Williams JF, Kolakowski D, Draper EA. The use of risk predictions to identify candidates for intermediate care units. Implications for intensive care utilization and cost. *Chest*. 1995 Aug;108(2):490–9.

80. Devereaux AV, Dichter JR, Christian MD, Dubler NN, Sandrock CE, Hick JL, et al. Definitive care for the critically ill during a disaster: a framework for allocation of scarce resources in mass critical care: from a Task Force for Mass Critical Care summit meeting, January 26-27, 2007, Chicago, IL. *Chest*. 2008 May;133(5 Suppl):51S-66S.
81. Health D of. *Pandemic flu: a national framework for responding to an influenza pandemic*. 2007.
82. Guest T, Tantam G, Donlin N, Tantam K, McMillan H, Tillyard A. An observational cohort study of triage for critical care provision during pandemic influenza: “clipboard physicians” or “evidenced based medicine”? *Anaesthesia*. 2009 Nov;64(11):1199–206.
83. Sprung CL, Zimmerman JL, Christian MD, Joynt GM, Hick JL, Taylor B, et al. Recommendations for intensive care unit and hospital preparations for an influenza epidemic or mass disaster: summary report of the European Society of Intensive Care Medicine’s Task Force for intensive care unit triage during an influenza epidemic or mas. *Intensive care medicine*. 2010 Mar;36(3):428–43.
84. Capuzzo M, Moreno RP, Alvisi R. Admission and discharge of critically ill patients. *Current opinion in critical care*. 2010 Oct;16(5):499–504.
85. Zimmerman JE, Wagner DP, Draper EA, Knaus WA. Improving intensive care unit discharge decisions: supplementing physician judgment with predictions of next day risk for life support. *Critical care medicine*. 1994 Sep;22(9):1373–84.
86. Frost SA, Alexandrou E, Bogdanovski T, Salamonson Y, Davidson PM, Parr MJ, et al. Severity of illness and risk of readmission to intensive care: a meta-analysis. *Resuscitation*. 2009 May;80(5):505–10.
87. HERRIDGE M. Prognostication and intensive care unit outcome: the evolving role of scoring systems. *Clinics in Chest Medicine*. 2003 Dec;24(4):751–62.
88. Consensus statement of the Society of Critical Care Medicine’s Ethics Committee regarding futile and other possibly inadvisable treatments. *Critical care medicine*. 1997 May;25(5):887–91.
89. Rapoport J, Teres D, Lemeshow S. Can futility be defined numerically? *Critical care medicine*. 1998 Nov;26(11):1781–2.
90. Halevy A, Neal RC, Brody BA. The low frequency of futility in an adult intensive care unit setting. *Archives of internal medicine*. 1996 Jan 8;156(1):100–4.
91. Goh AY, Mok Q. Identifying futility in a paediatric critical care setting: a prospective observational study. *Archives of disease in childhood*. 2001 Mar;84(3):265–8.

92. Afessa B, Keegan MT, Mohammad Z, Finkielman JD, Peters SG. Identifying potentially ineffective care in the sickest critically ill patients on the third ICU day. *Chest*. 2004 Dec;126(6):1905–9.
93. Bernard GR, Vincent JL, Laterre PF, LaRosa SP, Dhainaut JF, Lopez-Rodriguez A, et al. Efficacy and safety of recombinant human activated protein C for severe sepsis. *The New England journal of medicine*. 2001 Mar 8;344(10):699–709.
94. Teres D, Lemeshow S. Why severity models should be used with caution. *Critical care clinics*. 1994 Jan;10(1):93–110; discussion 111–5.
95. TPN and APACHE. *Lancet*. 1986 Jun 28;1(8496):1478.
96. Vincent J-L, Opal SM, Marshall JC. Ten reasons why we should NOT use severity scores as entry criteria for clinical trials or in our treatment decisions. *Critical care medicine*. 2010 Jan;38(1):283–7.
97. Eichacker PQ, Parent C, Kalil A, Esposito C, Cui X, Banks SM, et al. Risk and the efficacy of antiinflammatory agents: retrospective and confirmatory studies of sepsis. *American journal of respiratory and critical care medicine*. 2002 Nov 1;166(9):1197–205.
98. Clec'h C, Timsit J-F, De Lassence A, Azoulay E, Alberti C, Garrouste-Orgeas M, et al. Efficacy of adequate early antibiotic therapy in ventilator-associated pneumonia: influence of disease severity. *Intensive care medicine*. 2004 Jul;30(7):1327–33.
99. Hébert PC, Wells G, Blajchman MA, Marshall J, Martin C, Pagliarello G, et al. A multicenter, randomized, controlled clinical trial of transfusion requirements in critical care. Transfusion Requirements in Critical Care Investigators, Canadian Critical Care Trials Group. *The New England journal of medicine*. 1999 Feb 11;340(6):409–17.
100. Cohen J, Guyatt G, Bernard GR, Calandra T, Cook D, Elbourne D, et al. New strategies for clinical trials in patients with sepsis and septic shock. *Critical care medicine*. 2001 Apr;29(4):880–6.
101. Chen LM, Martin CM, Morrison TL, Sibbald WJ. Interobserver variability in data collection of the APACHE II score in teaching and community hospitals. *Critical care medicine*. 1999 Sep;27(9):1999–2004.
102. Polderman KH, Jorna EM, Girbes AR. Inter-observer variability in APACHE II scoring: effect of strict guidelines and training. *Intensive care medicine*. 2001 Aug;27(8):1365–9.
103. Booth FV, Short M, Shorr AF, Arkins N, Bates B, Qualy RL, et al. Application of a population-based severity scoring system to individual patients results in frequent misclassification. *Critical care (London, England)*. 2005 Oct 5;9(5):R522–9.

104. Polderman KH, Christiaans HM, Wester JP, Spijkstra JJ, Girbes AR. Intra-observer variability in APACHE II scoring. *Intensive care medicine*. 2001 Sep;27(9):1550–2.

4) Metodologia do artigo original

4.1) Delineamento

Estudo de coorte dividido em duas partes, braços retrospectivo e prospectivo.

4.2) População

4.2.1) Critérios de inclusão

Foram incluídos todos os pacientes admitidos na UTIP do HC-UFMG entre 01 de Fevereiro de 2009 e 31 de Janeiro de 2011. Os pacientes admitidos entre 01 de Fevereiro de 2009 e 31 de Janeiro de 2010 constituíam o braço retrospectivo e os pacientes admitidos entre 01 de Fevereiro de 2010 e 31 de Janeiro de 2011 constituíam o braço prospectivo.

4.2.2) Critérios de exclusão

Foram excluídos do estudo todos os pacientes menores de 30 dias de vida ou maiores de 18 anos, pacientes com óbito constatado nas duas primeiras horas de internação, pacientes com suspeita de morte cerebral na admissão com posterior confirmação, e pacientes considerados fora de possibilidade terapêutica curativa (FPTC).

4.3) Cenário

A UTIP do HC-UFMG possui 10 leitos pediátricos que atende pacientes provenientes do pronto-atendimento, enfermaria, e bloco cirúrgico do próprio hospital, além de pacientes externos encaminhados via regulação da central de leitos. Esta em funcionamento de acordo com o Regulamento Técnico para funcionamento de Unidades de Terapia Intensiva e Unidades de Cuidados Intermediários estabelecido pela Agência Nacional de Vigilância Sanitária (ANVISA) (11).

4.4) Coleta de dados

Os dados foram coletados pelo próprio pesquisador através de informações contidas no prontuário médico e na folha de registro da enfermagem, de forma retrospectiva (braço 1) e prospectiva (braço 2).

4.5) Variáveis estudadas

As variáveis estudadas constituíam em três grupos: variáveis para caracterização dos pacientes, variáveis para o cálculo do PIM2 e variáveis de desfecho.

Todas as variáveis estudadas consistiam de informações do exame clínico, da história pregressa ou da história da moléstia atual. Com a exceção da pressão parcial de oxigênio no sangue arterial (PaO₂) e excesso de bases (basic excess = BE) necessárias para o cálculo do PIM2, que necessitam de coleta de sangue arterial para sua mensuração.

Não houve nenhuma instrução para equipe médica responsável pelo cuidado dos pacientes sobre a necessidade de coleta de PaO₂ e BE na população estudada. Ocorria mensuração destas variáveis se o médico responsável pelo paciente considerasse o conhecimento dos valores da PaO₂ e BE importantes para condução do caso. Nos pacientes em que não ocorria mensuração destas variáveis, a PaO₂ e o BE era registrados como zero, conforme orientação dos autores do escore.

A coleta de sangue era realizada por um técnico do laboratório do hospital treinado para essa atividade. As dosagens laboratoriais foram realizadas no próprio laboratório do HC-UFG.

4.5.1) Variáveis para caracterização dos pacientes

1- Idade

A idade do paciente no momento da admissão foi registrada em meses.

2 - Gênero

O paciente foi registrado como do gênero masculino ou feminino.

3- Tipo de admissão

A admissão foi registrada como sendo secundária a uma condição clínica (pacientes clínicos) ou cirúrgica (pacientes cirúrgicos).

4- Tipos de condição clínica dos pacientes clínicos

Entre os pacientes clínicos foi realizada uma classificação baseada em qual sistema orgânico estava associado a condição responsável pela admissão do paciente na UTIP.

No registro foi considerado a principal razão (a patologia que estava provocando maior desarranjo fisiológico no paciente) para admissão do paciente na UTIP, em decorrência do fato que alguns pacientes tinham mais de uma patologia com necessidade de cuidados intensivos.

O paciente foi classificado como tendo uma condição associada ao sistema:

- Hematológico ou oncológico
- Cardiovascular
- Respiratório
- Renal
- Hepático
- Neurológico
- Após parada cardiorrespiratória
- Outras causas

5- Tipos de condição cirúrgica dos pacientes cirúrgicos

O paciente foi classificado como tendo uma condição cirúrgica:

- Cardiovascular
- Neurológica
- Cirurgia geral pediátrica
- Ortopédica
- Transplantes
- Cateterismo cardíaco
- Outras causas.

6- Presença de condição crônica complexa

A presença de condição crônica complexa foi registrada quando o paciente apresentava qualquer condição médica em que era esperada uma duração da patologia de pelo menos 12 meses (exceto quando paciente evoluía para óbito), e que compromete algum

sistema orgânico ou órgão o suficiente para necessitar de cuidado de uma especialidade pediátrica e provavelmente necessite de hospitalização em um hospital de cuidados terciários (12). A condição crônica complexa do paciente era classificada de acordo com a classificação desenvolvida por Feudtner e col (12):

- Mal-formações neuromusculares
- Mal-formações cardiovasculares
- Respiratório
- Renal
- Gastrointestinal
- Hematológico ou imunológico
- Metabólico
- Outros defeitos genéticos ou congênitos
- Neoplasias

7- Caráter da admissão

A admissão era classificada como eletiva ou não eletiva. Sendo considerada eletiva uma condição na qual uma admissão na UTIP ou uma cirurgia pode ser postergada por mais de 6 horas sem efeitos adversos para o paciente, definição similar à utilizada no PIM2.

8- Presença de Neoplasia

O paciente era classificado como tendo ou não uma neoplasia. Entre os pacientes com neoplasia foi registrado o tipo de neoplasia:

- Leucemia/Linfoma
- Tumores do sistema nervoso central (SNC)
- Tumores torácicos
- Tumores abdominais
- Tumores ósseos
- Outras neoplasias

9- Utilização de ventilação mecânica

O paciente foi registrado como tendo utilizado ventilação mecânica (VM) quando em qualquer momento durante a permanência na UTIP, o paciente fosse ventilado com dispositivo de VM através de tubo traqueal ou traqueostomia. Não foi considerado

como tendo utilizado VM quando o paciente utilizou ventilação não invasiva por máscara ou pronga nasal.

4.5.2) Variáveis para cálculo do PIM2

Foi realizado o registro das variáveis para o cálculo do PIM2 conforme as orientações disponíveis no artigo original:

1- Pressão Arterial Sistólica (PAS)

A PAS foi registrada em mililitros de mercúrio (mmHg). Na ausência da informação da PAS do paciente foi registrado um valor de 120. Exceto nas seguintes situações: no paciente em parada cardíaca, foi registrado a PAS com um valor igual a zero; no paciente com instabilidade hemodinâmica e pressão arterial tão baixa que não permitiu mensuração, foi registrado um valor de 30.

2- Reação pupilar a luz

Nos pacientes com pupilas maiores do que 3 mm ou fixas bilateralmente em resposta a luz foi registrado um valor de 1. Nos outros tipos de reação a luz ou na ausência da informação, foi registrado um valor igual a zero. Não foi registrado como achado anormal se a alteração da reação a luz fosse decorrente da utilização de drogas, presença de toxinas ou lesão de globo ocular.

3- Relação da fração inspirada de oxigênio (FiO₂) e pressão parcial de oxigênio no sangue arterial

O valor da FiO₂ era registrado em porcentagem e a PaO₂ em mmHg, com posterior divisão do primeiro pela segundo componente. Se os valores eram desconhecidos, a relação era registrada como zero.

4- Excesso de base (basic excess = BE)

O valor absoluto do BE foi registrado em mmol/L. Na ausência da informação do BE do paciente foi registrado um valor de zero.

5- Ventilação Mecânica

Nos pacientes que utilizarem VM em qualquer momento durante a primeira hora na UTI foi registrado um valor de 1, nos pacientes que não utilizaram foi registrado um valor igual a zero. Ventilação Mecânica inclui CPAP nasal/mascará ou BiPAP ou ventilação com pressão negativa.

6- Admissão eletiva

Na admissão considerada eletiva foi registrado um valor de 1, nas classificadas como não eletivas (urgentes) foi registrado um valor igual a zero. Admissão eletiva incluiu admissão após cirurgia eletiva ou após procedimento eletivo (exemplo inserção de acesso central), ou monitoramento eletivo, ou revisão de ventilação domiciliar. Uma admissão em UTI ou uma cirurgia foi considerada eletiva se poderia ser postergada por mais de 6 horas sem efeitos adversos.

7- Recuperação de cirurgias ou de procedimentos

Nos pacientes em que o principal motivo para admissão foi a recuperação de cirurgia ou de procedimento era registrado um valor de 1, nas demais admissões foi registrado um valor igual a zero. Recuperação de cirurgias ou de procedimentos inclui procedimentos radiológicos e cateterismos cardíacos. Não inclui pacientes admitidos do bloco cirúrgico quando a recuperação da cirurgia não é a principal razão para admissão na UTI (exemplo paciente com lesão cerebral que é admitido proveniente do bloco cirúrgico após inserção de um monitor de pressão intracraniana, nesse caso a principal razão de admissão é a lesão cerebral).

8- Admissão após circulação extra-corpórea

Nos pacientes submetidos a circulação extra-corpórea foi registrado um valor de 1, nos pacientes que não foram submetidos ao procedimento será registrado um valor igual a zero.

9- Diagnóstico de alto risco

Nos pacientes apresentado algum dos diagnósticos relacionados abaixo foi registrado um valor de 1, nos demais pacientes foi registrado um valor igual a zero.

[A] Parada cardíaca precedendo admissão na UTI

[B] Imunodeficiência combinada grave

[C] Leucemia ou Linfoma após primeira indução com quimioterapia

- [D] Hemorragia cerebral espontânea
- [E] Cardiomiopatia ou miocardite
- [F] Síndrome do coração esquerdo hipoplásico
- [G] Doença hepática é a principal razão para admissão na UTI
- [F] Doença neuro-degenerativa

O diagnóstico de parada cardíaca precedendo a admissão incluiu parada cardíaca ocorrida dentro e fora do hospital. Requeria ausência de pulso documentada ou necessidade de compressão torácica externa, porém não incluía história pregressa de parada cardíaca.

No diagnóstico de hemorragia cerebral o sangramento deveria ser espontâneo (exemplo aneurisma ou malformação arteriovenosa). Não incluiu hemorragia cerebral traumática ou hemorragia intracraniana que não fosse intracerebral (exemplo hemorragia subdural).

O diagnóstico de Síndrome do coração esquerdo hipoplásico incluiu pacientes de qualquer idade, mas somente foram considerados os casos em que o procedimento de Norwood ou equivalente era ou foi necessário no período neonatal para a sobrevivência do paciente.

O diagnóstico de Insuficiência Hepática aguda ou crônica deveria ser considerado a principal razão para admissão na UTI. Incluiu pacientes admitidos para recuperação de Transplantes Hepáticos por Insuficiência Hepática aguda ou crônica.

O diagnóstico de Doença neurodegenerativa necessitava de uma história de perda progressiva dos marcos de desenvolvimento ou um diagnóstico firmado.

10- Diagnósticos de baixo risco

Nos pacientes apresentado algum dos diagnósticos relacionados abaixo foi registrado um valor de 1, nos demais pacientes foi registrado um valor igual a zero.

- [A] Asma é a principal razão para admissão na UTI
- [B] Bronquiolite é a principal razão para admissão na UTI
- [C] Crupe é a principal razão para admissão na UTI
- [D] Apnéia obstrutiva do sono é a principal razão para admissão na UTI
- [E] Cetoacidose diabética é a principal razão para admissão na UTI

O diagnóstico de Bronquiolite incluiu pacientes que apresentam com desconforto respiratório ou apnéia quando o diagnóstico clínico foi Bronquiolite.

O diagnóstico de Apnéia obstrutiva do sono incluiu pacientes admitidos após Adenoidectomia ou Tonsilectomia em quem a Apnéia obstrutiva do sono foi a principal razão para admissão na UTI.

4.5.3) Variáveis de desfecho

1- Evento de saída do UTIP

Era registrado se o paciente recebeu alta da UTIP ou evoluiu para óbito na unidade.

2- Tempo de internação

Era registrado o número de dias que o paciente permaneceu na UTIP antes da alta para outro setor de menor complexidade ou o número de dias que o paciente permaneceu na unidade até evoluir para óbito. Pacientes que evoluíram para óbito durante as primeiras 24 horas de internação foram computados com tendo permanecido 1 dia na unidade.

4.6) Cálculo do PIM2

A probabilidade de morte de cada paciente do estudo foi calculada através das instruções disponibilizadas pelos autores do score. Para estimar a probabilidade de morte eram necessárias duas etapas: o cálculo do valor do PIM2 e o cálculo da probabilidade de morte a partir do valor obtido do PIM2.

O valor do PIM2 foi obtido através da equação disponibilizada pelos autores que utiliza as variáveis descritas no item 3.5.2 (Variáveis para cálculo do PIM2):

$$\text{PIM2} = [0.01395 \times (A-120)] + (3.0791 \times B) + [0.2888 \times (100 \times C/D)] + (0.104 \times E) + (1.3352 \times G) - (0.9282 \times H) - (1.0244 \times I) + (0.7507 \times J) + (1.6829 \times L) - (1.5770 \times M) - 4.8841 *$$

* Sendo as letras utilizadas na equação correspondentes as variáveis:

A = PAS; B = Reação pupilar a luz; C = FiO₂; D = PaO₂; E = Base excess; G = Ventilação mecânica; H = Procedimento eletivo; I = Recuperação de cirurgias ou de

procedimentos; J = Admissão após bypass cardíaco; L = Diagnóstico de alto risco; M = Diagnóstico de baixo risco.

Com a substituição das letras pelo valor numérico correspondente a cada variável conforme instrução no item 2.5.2 (Variáveis para cálculo do PIM2).

A partir do valor do PIM2 era calculada a probabilidade de morte através de outra equação disponibilizada pelos autores do escore:

$$\text{Probabilidade de morte (em porcentagem)} = \exp(\text{PIM2}) / [1 + \exp(\text{PIM2})]$$

4.6.1) Exemplo de cálculo do PIM2

Paciente com síndrome do coração esquerdo hipoplásico é admitido na UTIP para recuperação pós-operatória de procedimento eletivo de Norwood. Na admissão ele estava intubado e foi colocado em ventilação mecânica. A primeira PAS registrada era 55 mmHg, PaO₂ era 110 mmHg, FiO₂ era 50% (0,5), base excess era -6. As pupilas eram reativas à luz. O diagnóstico de baixo risco não se aplica a este caso.

$$\begin{aligned} \text{PIM2} &= [0.01395 \times (55-120)] + (3.0791 \times 0) + [0.2888 \times (100 \times 0,5/110)] + (0.104 \times 6) \\ &+ (1.3352 \times 1) - (0.9282 \times 1) - (1.0244 \times 1) + (0.7507 \times 1) + (1.6829 \times 1) - (1.5770 \times 0) \\ &- 4.8841 = -1,4059 \end{aligned}$$

$$\text{Probabilidade de morte} = \exp(-1,4059) / [1 + \exp(-1,4059)] = 0.1969 \text{ ou } 19.7\%$$

4.7) Análise estatística

A análise descritiva da caracterização dos pacientes foi realizada através da média e desvio padrão para variáveis contínuas com distribuição normal, e mediana e intervalo interquartil para variáveis contínuas sem distribuição normal. As variáveis contínuas de distribuição normal foram avaliadas pelo teste t de student e expressas através da média e desvio-padrão (DP). As variáveis contínuas sem distribuição normal foram expressas através das medianas e intervalo interquartil 25-75% (IQ25-75%) e comparadas pelo teste não paramétrico de Mann-Whitney. A comparação da distribuição de variável dicotômica foi analisada através do teste de qui-quadrado, com

correção de Yates, ou do teste exato de Fisher, bicaudal, se necessário. A probabilidade de significância foi considerada significativa quando inferior a 0,05 ($p < 0,05$). A avaliação do desempenho do escore foi realizada através da discriminação e calibração (13). A discriminação foi avaliada através do cálculo da área sob a curva ROC (receiver operator characteristic), sendo considerada adequada quando a área foi maior do que 0,7. A calibração foi avaliada através do teste de ajuste de Hosmer-Lemeshow, sendo considerada adequada quando o valor de p do teste Hosmer-Lemeshow é maior do que 0,05. No teste de Hosmer-Lemeshow os pacientes foram divididos em 10 grupos com probabilidades de risco crescente e número similar de casos em cada grupo, de forma similar ao estudo de desenvolvimento do PIM2.

A comparação entre a mortalidade encontrada na população estudada com a mortalidade prevista pelo escore PIM2 foi realizada dividindo a taxa de mortalidade observada pela taxa de mortalidade prevista (conforme calculada pelo escore), chamada de taxa de mortalidade padronizada (TMP). A TMP foi apresentada com seu intervalo de confiança (IC) de 95%. Se o IC 95% da TMP incluiu 1, o desempenho foi considerado médio; se o IC 95% da TMP apresentou o limite superior menor do que 1 o desempenho foi considerado bom, e se o IC 95% da TMP apresentou o limite inferior maior do que 1 o desempenho foi considerado ruim. Todas as análises foram realizadas através do software SPSS versão para Windows 12.

4.8) Considerações éticas

O estudo foi aprovado pelo Comitê de Ética em Pesquisa da Universidade Federal de Minas Gerais (COEP-UFMG), número do parecer ETIC 0436.0.203.000-09.

5) Artigo original: Aplicação do Paediatric Index of Mortality 2 em Unidade de Terapia Intensiva Pediátrica no Brasil.

5.1) Introdução

A qualidade inadequada e o custo crescente dos cuidados de saúde oferecidos têm sido focos de interesse para a comunidade médica (1). A unidade de terapia intensiva (UTI) é um dos cenários onde essas questões têm sido mais debatidas em decorrência do impacto significativo da qualidade do cuidado nos resultados do paciente (morbidade, resultado funcional e mortalidade) e dos altos custos gerados pelo cuidado ao paciente criticamente doente (2).

Há um aumento do uso de indicadores de qualidade como instrumento de melhoria dos serviços de saúde ofertados (3). A taxa de mortalidade padronizada (TMP) é um dos principais indicadores utilizados na terapia intensiva, sendo calculada dividindo a taxa de mortalidade observada pela taxa de mortalidade prevista por um escore de predição de desfecho (4). Na terapia intensiva pediátrica os principais escores de predição de desfecho utilizados são o Pediatric Risk of Mortality (PRISM) (5), o Paediatric Index of Mortality (PIM) (6) e suas atualizações respectivamente (7,8).

O uso do Paediatric Index of Mortality 2 (PIM2), a última versão do PIM, tem sido incentivado por ser um escore de fácil utilização, eficiente e de domínio público (9). Desde a publicação do PIM2 em 2003 vários estudos foram publicados avaliando o desempenho do escore em populações e cenários diferentes do estudo de desenvolvimento, com a maioria dos estudos evidenciando desempenho adequado do escore (9).

Os autores do PIM2 relatam que diferenças das características e diagnósticos de uma população quando comparada com a população do desenvolvimento do escore podem resultar num desempenho inadequado do modelo quando avaliado em um novo cenário (8), sendo recomendada a avaliação dos escores em cenários e populações diferentes dos utilizados na produção do modelo com intuito de verificar a permanência do desempenho adequado (10).

No Brasil ainda não foram publicados estudos avaliando o PIM2, sendo necessária a avaliação do desempenho do escore neste cenário antes de seu emprego como ferramenta de melhoria de qualidade dos cuidados de saúde oferecidos pelos serviços de terapia intensiva pediátrica existentes no país (10).

5.2) Objetivos

Avaliar o desempenho (discriminação e calibração) do escore preditor de desfecho Paediatric Index of Mortality 2 na Unidade de Terapia Intensiva Pediátrica (UTIP) do Hospital das Clínicas da Universidade Federal de Minas Gerais (HC-UFGM). Comparar a mortalidade encontrada na população estudada com a mortalidade prevista pelo escore PIM2.

5.3) Metodologia

Estudo de coorte dividido em duas partes, braços retrospectivo e prospectivo no qual foram incluídos todos os pacientes admitidos na UTIP do HC-UFGM entre 01 de Fevereiro de 2009 e 31 de Janeiro de 2011. Os pacientes admitidos entre 01 de Fevereiro de 2009 e 31 de Janeiro de 2010 constituíam o braço retrospectivo e os pacientes admitidos entre 01 de Fevereiro de 2010 e 31 de Janeiro de 2011 constituíam o braço prospectivo.

Foram excluídos do estudo todos os pacientes menores de 30 dias de vida ou maiores de 18 anos, pacientes com óbito constatado nas duas primeiras horas de internação, pacientes com suspeita de morte cerebral na admissão com posterior confirmação, e pacientes considerados fora de possibilidade terapêutica curativa (FPTC).

A UTIP do HC-UFGM possui 10 leitos pediátricos e atende pacientes provenientes do pronto-atendimento, enfermaria, bloco cirúrgico e em menor frequência pacientes encaminhados via regulação da central de leitos. Está em funcionamento de acordo com o Regulamento Técnico para funcionamento de Unidades de Terapia Intensiva e Unidades de Cuidados Intermediários estabelecido pela Agência Nacional de Vigilância Sanitária (ANVISA) (11).

5.3.1) Variáveis estudadas

As variáveis estudadas constituíam em três grupos: variáveis para caracterização dos pacientes, variáveis para o cálculo do PIM2 e variáveis de desfecho. Na caracterização dos pacientes foram avaliadas as variáveis: idade; gênero; tipo de admissão (clínica ou cirúrgica); tipos de patologias clínicas; tipos de patologias

cirúrgicas; presença de condição crônica complexa; presença de neoplasia; condição eletiva ou não eletiva da admissão; e utilização de ventilação mecânica invasiva. No registro do tipo de patologia clínica foi considerada a principal razão para admissão do paciente na UTIP (a patologia que estava provocando maior desarranjo fisiológico no paciente), em decorrência do fato que alguns pacientes tinham mais de uma patologia com necessidade de cuidados intensivos.

O paciente foi classificado como tendo uma patologia clínica: cardiovascular; respiratória; renal; hepática; neurológica; pós-parada cardiorrespiratória; oncológica ou hematológica. O paciente foi classificado como tendo uma patologia cirúrgica: cardiovascular; neurológica; cirurgia geral; ortopédica; transplantes; cateterismo cardíaco.

A presença de condição crônica complexa foi registrada quando o paciente apresentava qualquer condição médica em que era esperada uma duração da patologia de pelo menos 12 meses (exceto quando paciente evoluía para óbito), e que compromete algum sistema orgânico ou órgão o suficiente para necessitar de cuidado de uma especialidade pediátrica e provavelmente necessitar de hospitalização em um hospital de cuidados terciários (12). A condição crônica complexa do paciente era classificada de acordo com a classificação desenvolvida por Feudtner e col (12): malformações neuromusculares; malformações cardiovasculares; respiratório; renal; gastrointestinal; hematológico ou imunológico; metabólico; outros defeitos genéticos ou congênitos; e neoplasias.

A admissão foi registrada como eletiva quando uma admissão na UTIP ou uma cirurgia poderia ser postergada por mais de 6 horas sem efeitos adversos para o paciente (8).

Foi realizado o registro das variáveis para o cálculo do PIM2 conforme as orientações disponíveis no artigo original (8). A probabilidade de morte de cada paciente do estudo foi calculada através das instruções disponibilizadas pelos autores do score (8). Já as variáveis de desfecho avaliadas foram: evento de saída do UTIP (alta ou óbito) e tempo de internação.

5.3.2) Análise estatística

A análise descritiva da caracterização dos pacientes foi realizada através da média e desvio padrão para variáveis contínuas com distribuição normal, e mediana e

intervalo interquartil para variáveis contínuas sem distribuição normal. As variáveis contínuas de distribuição normal foram avaliadas pelo teste t de student e expressas através da média e desvio-padrão (DP). As variáveis contínuas sem distribuição normal foram expressas através das medianas e intervalo interquartil 25-75% (IQ25-75%) e comparadas pelo teste não paramétrico de Mann-Whitney. A comparação da distribuição de variável dicotômica foi analisada através do teste de qui-quadrado, com correção de Yates, ou do teste exato de Fisher, bicaudal, se necessário. A probabilidade de significância foi considerada significativa quando inferior a 0,05 ($p < 0,05$). A avaliação do desempenho do escore foi realizada através da discriminação e calibração (13). A discriminação foi avaliada através do cálculo da área sob a curva ROC (receiver operator characteristic) (14), sendo considerada adequada quando a área foi maior do que 0,7 (15). A calibração foi avaliada através do teste de ajuste de Hosmer-Lemeshow (14), sendo considerada adequada quando o valor de p do teste Hosmer-Lemeshow é maior do que 0,05 (15). No teste de Hosmer-Lemeshow os pacientes foram divididos em 10 grupos com probabilidades de risco crescente e número similar de casos em cada grupo, de forma similar ao estudo de desenvolvimento do PIM2 (8).

A comparação entre a mortalidade encontrada na população estudada com a mortalidade prevista pelo escore PIM2 foi realizada dividindo a taxa de mortalidade observada pela taxa de mortalidade prevista (conforme calculada pelo escore), chamada de taxa de mortalidade padronizada (TMP) (4). A TMP foi apresentada com seu intervalo de confiança (IC) de 95%. Se o IC 95% da TMP incluiu 1, o desempenho foi considerado médio; se o IC 95% da TMP apresentou o limite superior menor do que 1 o desempenho foi considerado bom, e se o IC 95% da TMP apresentou o limite inferior maior do que 1 o desempenho foi considerado ruim (16). Todas as análises foram realizadas através do software SPSS versão para Windows 12.

O estudo foi aprovado pelo Comitê de Ética em Pesquisa da Universidade Federal de Minas Gerais (COEP-UFMG), número do parecer ETIC 0436.0.203.000-09.

5.4) Resultados

Ocorreram 756 admissões na UTIP do HC-UFMG entre 01 de Fevereiro de 2009 e 31 de Janeiro de 2011. Setenta e nove admissões não preencheram os critérios de inclusão do estudo: 63 pacientes tinham idade menor que 30 dias de vida, 10 pacientes foram considerados fora de possibilidade terapêutica curativa, 4 pacientes evoluíram

para óbito durante as 2 primeiras horas após a admissão na unidade, e 2 pacientes apresentavam suspeita diagnóstica de morte encefálica na admissão confirmada posteriormente. No total 677 admissões foram incluídas no estudo, não ocorrendo perda de seguimento ou dados de nenhuma admissão.

A mediana da idade dos pacientes foi de 4,9 anos (intervalo interquartil 25-75% 1,1-10,2 anos), com 51,5% (349) do gênero masculino. Os pacientes foram admitidos na UTIP após a realização de procedimentos cirúrgicos em 52,7% (357) dos casos. No total 568 pacientes (83,9%) apresentavam uma condição crônica complexa, 171 (25,2%) apresentavam uma neoplasia e 370 (54,6%) necessitaram de ventilação mecânica.

A população do estudo apresentava a seguinte distribuição etária: pacientes com idade menor do que 12 meses correspondiam a 22,9% do total da amostra, idade entre 13 e 60 meses eram 28,5%, entre 61 e 120 meses eram 23,2%, e os pacientes com idade entre 121 e 215 meses eram 25,4%.

Entre os pacientes admitidos após a realização de procedimentos cirúrgicos, 95% (340) dos pacientes foram submetidos a procedimentos eletivos. As clínicas cirúrgicas responsáveis foram a Cirurgia Geral Pediátrica (105; 29,41%), Neurocirurgia (91; 25,49%), Cirurgia Cardiovascular (84; 23,54%), Hemodinâmica (cateterismo cardíaco) (39; 10,92%); Transplante (17; 4,76%); Ortopédica (9; 2,52%) e outras (12; 3,36%).

Os principais procedimentos cirúrgicos realizados pela Cirurgia Geral Pediátrica foram ressecção de tumores abdominais (20; 19%), laparotomia por abdome agudo (17; 16,1%) e ressecção de tumores torácicos (13; 12,3%). A Neurocirurgia apresentou a ressecção de tumores do SNC (42; 46,1%), a colocação de dispositivo para derivação ventrículo-peritoneal (14; 15,3%) e a correção de Craniossinostose (13; 14,2%) como os principais procedimentos realizados. A maioria (78; 92,8%) dos procedimentos da Cirurgia Cardiovascular foram correções de cardiopatias congênitas. Na categoria de transplantes foram incluídos transplantes de rim (10 casos) e fígado (7 casos).

Os pacientes clínicos corresponderam a 47,3% (320/677) do total da amostra, com 97,5% (312) dos pacientes apresentando uma indicação de admissão não eletiva. Os três sistemas orgânicos responsáveis pela maioria dos pacientes clínicos admitidos na unidade foram os sistemas respiratório (134; 41,87%), circulatório (76; 23,75%) e neurológico (32; 10%) (tabela 1).

Tabela 1 – Distribuição dos sistemas orgânicos relacionados às causas das admissões clínicas.

Motivo da admissão dos pacientes clínicos	Valor
Respiratório	41,87% (134/320)
Circulatório	23,75% (76/320)
Neurológico	10% (32/320)
Outros	7,81% (25/320)
Hepático	5,93% (19/320)
Pós-PCR	5,62% (18/320)
Onco-hematológica	2,81% (9/320)
Renal	2,18% (7/320)

Abreviatura: PCR = Parada Cardiorrespiratória

As principais causas de admissão dentro da categoria do sistema respiratório foram pneumonia (61; 45,5%), obstrução de via aérea superior por múltiplas causas (17; 12,6%) e asma (10; 7,4%). As causas de admissão mais encontradas na categoria do sistema circulatório foram choque séptico (44; 57,8%) e insuficiência cardíaca descompensada (10; 13,1%). No sistema neurológico os diagnóstico mais observados foram *status epilepticus* (9; 28,1%), meningite/encefalite (6; 18,7%), acidente vascular cerebral (5; 15,6%) e hipertensão intracraniana por hidrocefalia descompensada ou por crescimento de tumor do SNC (5; 15,6%).

A maioria (568; 83,9%) dos pacientes apresentava uma condição crônica complexa. As três categorias de condições crônicas responsáveis pela maioria dos casos foram as Neoplasias (171; 30,10%), Malformações Cardiovasculares (154; 27,11%) e Malformações Neuromusculares (64; 11,26%) (tabela 2).

Na categoria das Neoplasias, os tumores do Sistema Nervoso Central (69; 40,35%) e as Leucemias/Linfomas (54; 31,57%) foram os principais diagnósticos encontrados. As Cardiopatias Congênitas foram responsáveis pela maioria dos diagnósticos (142; 92,20%) na categoria das Malformações Cardiovasculares. Na categoria das Malformações Neuromusculares a hidrocefalia (20; 31,2%), a paralisia cerebral (9; 14%), a mielomeningocele (7; 10,9%) e a malformação de vasos cerebrais (7; 10,9%) foram os diagnósticos mais comuns.

Tabela 2 – Distribuição dos tipos de condição crônica complexa na população.

Tipo de condição crônica complexa	Porcentagem/valor absoluto
Neoplasias	30,10% (171/568)
Malformações cardiovasculares	27,11% (154/568)
Malformações neuromusculares	11,26% (64/568)
Gastrointestinal	9,68% (55/568)
Renal	8,62% (49/568)
Respiratório	4,75% (27/568)
Outros defeitos genéticos ou congênitos	3,52% (20/568)
Hematológico ou imunológico	3,34% (19/568)
Metabólico	1,58% (9/568)

A taxa de mortalidade geral foi de 9,7% (66/677), com as taxas de mortalidade dos diferentes subgrupos contidas na tabela 5. A mortalidade dos pacientes clínicos apresentou diferença com significância estatística em relação aos pacientes cirúrgicos (17,8% versus 2,5%; $p < 0,0001$). Não ocorreu diferença com significância estatística entre a mortalidade dos pacientes com condição crônica complexa e sem condição crônica (10,3% versus 6,4%; $p = 0,27$). Não ocorreu diferença com significância estatística entre a mortalidade dos pacientes com neoplasia e sem neoplasia (12,8% versus 8,6%; $p = 0,15$).

A mediana da idade dos pacientes que morreram foi de 71 meses ou 5,9 anos, sendo 56,1% (37) do gênero feminino. Entre os pacientes que faleceram, 86,4% (57) apresentava um diagnóstico de patologia clínica como causa da admissão, e 89,3% (59) tinha uma condição crônica complexa. Todos os pacientes que morreram utilizaram ventilação mecânica. O mediana do tempo de permanência na UTIP dos pacientes que faleceram foi de 4 dias. As causas de admissão mais comuns entre os pacientes que morreram foram choque séptico (17; 25,7%), pneumonia (11; 16,6%) e encefalopatia hepática (6; 9%).

Entre os pacientes que faleceram que apresentavam uma condição crônica complexa, as categorias mais prevalentes foram Neoplasias (22; 37,2%), doenças gastrointestinais (23,7%) e doenças imuno-hematológicas (8; 13,5). Na categoria das doenças gastrointestinais, o diagnóstico de insuficiência hepática crônica foi o responsável por 85,7% dos casos.

A taxa de mortalidade também apresentou diferença entre as faixas etárias estudadas: pacientes com idade menor do que 12 meses apresentaram mortalidade de 5,8% (9/155), pacientes com idade entre 13 e 60 meses apresentaram mortalidade de 10,8% (21/193), pacientes entre 61 e 120 meses apresentaram mortalidade de 8,2% (13/157), e os pacientes com idade entre 121 e 215 meses apresentaram mortalidade de 13,3% (23/172).

A probabilidade de morte média da população do estudo estimada pelo PIM2 da foi de 5,9%, sendo de 24,4% entre os pacientes que morreram e de 4,1% entre os pacientes que sobreviveram. A probabilidade de morte média nos diferentes subgrupos está descrita na tabela 3. A probabilidade de morte estimada pela PIM2 apresentou diferença com significância estatística entre os pacientes clínicos e cirúrgicos (10,2% versus 1,9%; $p < 0,0001$). Não ocorreu diferença com significância estatística da probabilidade de morte estimada pelo PIM2 quando comparados os pacientes com e sem condição crônica complexa (5,9% versus 5,6%; $p = 0,5$) e quando comparados os grupos com e sem neoplasia (6,9% versus 5,5%; $p = 0,16$).

A TMP de toda a população do estudo foi de 1,65 (mortes observadas = 66; mortes estimadas pelo PIM2 = 39,9; IC95% 1,26-2,04). Nos pacientes cirúrgicos a TMP foi de 1,26 (mortes observadas = 9; mortes estimadas pelo PIM2 = 7,1; IC95% 0,44-2,08), já nos pacientes clínicos foi de 1,73 (mortes observadas = 57; mortes estimadas pelo PIM2 = 32,8; IC95% 1,28-2,18). A TMP dos pacientes com condição crônica complexa foi de 1,75 (mortes observadas = 59; mortes estimadas pelo PIM2 = 33,7; IC95% 1,31-2,19), enquanto dos pacientes sem condição crônica complexa foi de 1,14 (mortes observadas = 7; mortes estimadas pelo PIM2 = 6,1; IC95% 0,29-1,99). Nos pacientes com neoplasia a TMP foi de 1,84 (mortes observadas = 22; mortes estimadas pelo PIM2 = 11,9; IC95% 1,07-2,61), já nos pacientes sem neoplasia foi de 1,57 (mortes observadas = 44; mortes estimadas pelo PIM2 = 28; IC95% 1,11-2,03).

O tempo médio de permanência na unidade foi de 7,9 dias e a mediana foi de 4 dias. O tempo médio e a mediana de permanência nos subgrupos estão descritos na tabela 3. Os pacientes cirúrgicos apresentaram uma menor mediana de tempo de permanência na unidade do que os pacientes clínicos (2 versus 5,5 dias; $p < 0,0001$). Não ocorreu diferença com significância estatística da mediana do tempo de permanência entre os pacientes com e sem condição complexa ($p = 0,84$) e entre os pacientes com e sem neoplasia ($p = 0,10$).

Tabela 3 – Mortalidade, PIM2 médio, TMP e média e mediana do tempo de permanência nos subgrupos.

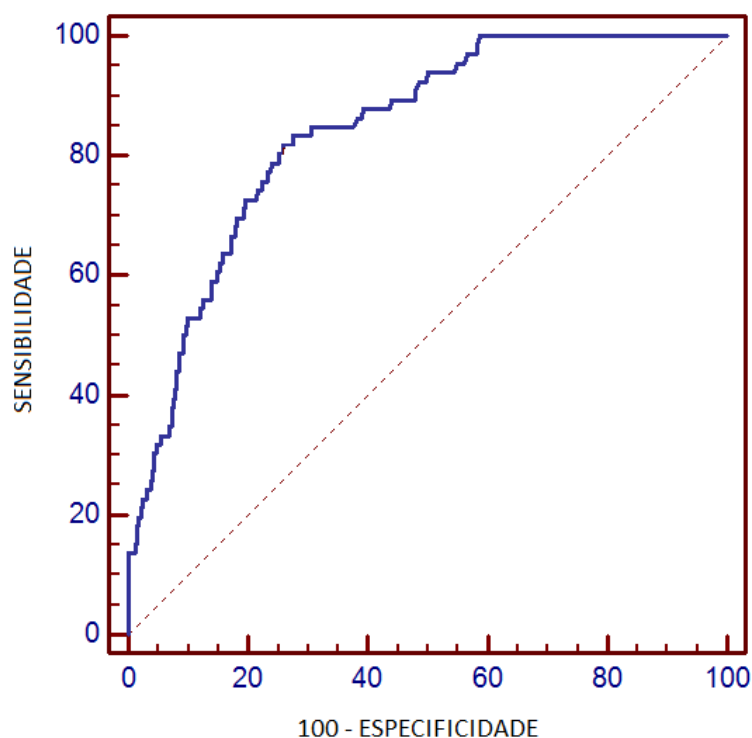
Categorias	Mortalidade	PIM2 médio	TMP	Tempo de permanência*
Geral	9,7%	5,9%	1,65	7,9 dias/ 4 dias
Cirúrgicos	2,5%	1,9%	1,26	4,7 dias/ 2 dias
Clínicos	17,8%	10,2%	1,73	11,4 dias/ 5,5 dias
Condição crônica complexa	10,3%	5,9%	1,75	8,1 dias/ 3dias
Sem condição crônica complexa	6,4%	5,6%	1,14	6,5 dias/ 3 dias
Neoplasia	12,8%	6,9%	1,84	8,2 dias/ 3 dias
Sem Neoplasia	8,6%	5,5%	1,57	7,8 dias/ 3 dias

Abreviatura: TMP = Taxa de Mortalidade Padronizada

* Média e mediana

A discriminação do escore avaliada através da curva ROC evidenciou uma área sob a curva ROC de 0,84 (IC95% 0,810-0867) (gráfico 1), sendo considerada adequada.

Gráfico 1 – Curva ROC de toda a população do estudo



A calibração do escore avaliada através do teste de ajuste de Hosmer-Lemeshow evidenciou um qui-quadrado de 43,4056 com 8 graus de liberdade ($p < 0,0001$) (tabela 4), sendo considerada inadequada.

Tabela 4 – Calibração do escore entre as diferentes faixas de risco.

Faixas de risco	Óbitos		Sobreviventes		Total
	Observados	Esperados	Observados	Esperados	
0,1-0,13%	0	3,839	64	60,161	64
0,14-0,22%	0	4,091	68	63,909	68
0,23-0,57%	0	4,137	68	63,863	68
0,58-0,99%	3	4,422	68	66,578	71
1,0-1,33%	4	4,374	65	64,626	69
1,34-1,98%	3	4,487	66	64,513	69
1,99-3,68%	4	4,67	64	63,33	68
3,73-6,07%	12	5,205	56	62,795	68
6,11-15,26%	18	6,948	50	61,052	68
15,54-99,25%	22	23,829	42	40,171	64

Teste de ajuste de Hosmer-Lemeshow $X^2 = 43,4056$, 8 graus de liberdade, $p < 0,0001$

Também foi realizada a discriminação e a calibração do escore nos seguintes subgrupos de pacientes: pacientes cirúrgicos, pacientes clínicos, pacientes com comorbidades, pacientes sem comorbidades, pacientes com neoplasias e pacientes sem neoplasias (tabela 5). Além disso, foi realizada a discriminação e calibração do escore em diferentes faixas etárias: ≤ 12 meses, entre 13 e 60 meses, entre 61 e 120 meses, e entre 121 e 213 meses (tabela 6).

Tabela 5 – Discriminação e calibração nos subgrupos de pacientes.

Categorias	Área sob a curva ROC	X^2 do teste de HL	p do teste de HL
Geral	0,840 (0,810-0,867)	43,405	<0,0001
Cirúrgicos	0,897 (0,861-0,926)	6,674	0,5721
Clínicos	0,743 (0,691-0,79)	12,584	0,127
Condição crônica complexa	0,826 (0,792-0,856)	36,38	<0,0001
Sem condição crônica complexa	0,944 (0,883-0,979)	7,089	0,527
Neoplasia	0,872 (0,812-0,918)	11,025	0,2002
Sem Neoplasia	0,823 (0,787-0,855)	27,786	0,0005

Abreviaturas: ROC = Receiver Operator Characteristic; HL = Hosmer-Lemeshow

Todos os subgrupos apresentaram discriminação adequada, com o subgrupo de pacientes sem condição crônica complexa apresentando o melhor poder discriminatório e o subgrupo dos pacientes clínicos apresentando o pior poder discriminatório. A calibração foi adequada para todos os subgrupos, exceto os subgrupos de pacientes sem neoplasias e de pacientes com condição crônica complexa. O subgrupo dos pacientes cirúrgicos apresentou a melhor calibração entre os subgrupos, enquanto o subgrupo dos pacientes com condição crônica complexa apresentou a pior calibração.

Tabela 6 – Discriminação e calibração em diferentes faixas etárias.

Faixas etárias	Área sob a curva ROC	X^2 do teste de HL	<i>p</i> do teste de HL
≤ 12 meses	0,859 (0,794-0,910)	13,359	0,1001
13 - 60 meses	0,763 (0,691-0,821)	9,135	0,331
61 – 120 meses	0,864 (0,800-0,913)	8,940	0,3473
121 - 213 meses	0,873 (0,817-0,921)	27,139	0,0007

Abreviaturas: ROC = Receiver Operator Characteristic; HL = Hosmer-Lemeshow

Todas as faixas etárias apresentaram discriminação adequada, com a faixa etária entre 121 e 213 meses apresentando a melhor discriminação e a faixa etária entre 13 e 60 meses apresentando a pior discriminação. Todas as faixas etárias apresentaram calibração adequada, exceto a faixa etária entre 121 e 213 meses. A faixa etária entre 61 e 120 meses apresentou a melhor calibração.

5.5) Discussão

Os indicadores de qualidade de saúde têm ganhado espaço no processo de melhoria dos serviços oferecidos para a população (17), sendo uma ferramenta bastante utilizada pelos gestores e profissionais de saúde (3). A taxa de mortalidade padronizada é o principal indicador de qualidade usado na terapia intensiva, sendo calculada dividindo a mortalidade observada pela mortalidade prevista por um escore de predição de desfecho (4).

O PIM2 e o Pediatric Risk of Mortality III (PRISM III) são os escores de predição de desfecho mais utilizados na terapia intensiva pediátrica, com os estudos

produzidos até o momento não permitindo afirmar que um dos escores tem desempenho superior (13). O PIM2 é um escore de domínio público, enquanto o PRISM III é patenteado, sendo necessário à compra de um software, além de uma taxa anual de licença de uso (18). Com isso a utilização do PIM2 tem ganhado espaço nas UTIP em todo mundo, principalmente em países em desenvolvimento onde a escassez de recursos financeiros é mais evidente (9).

Entretanto antes de iniciar a utilização de um escore de predição de desfecho num novo cenário é recomendada a avaliação do desempenho do modelo nesta nova população (10), sendo esse o objetivo principal do presente estudo. Os autores do PIM2 relatam que é possível que o escore não funcione bem num cenário onde as características ou diagnósticos dos pacientes sejam substancialmente diferentes dos pacientes do desenvolvimento do modelo (8).

A população estudada apresentou como principais características uma predominância de admissões de pacientes cirúrgicos, e uma alta proporção de pacientes com diagnóstico de condição crônica complexa e neoplasia. A alta prevalência de pacientes com condição crônica complexa (83,9%) pode ser considerada a característica mais marcante desta população, condição não encontrada no estudo de desenvolvimento do PIM2 (8). Nos estudos pesquisados avaliando o PIM2 apenas Thukral e col (19) e Wolfler e col (20) relatam a presença de doenças crônicas em suas populações, 49,3% e 28,9% respectivamente, prevalência bem abaixo da observada em nossa casuística.

Essa característica encontrada em nossa população pode estar relacionada aos avanços no cuidado médico nos últimos anos e principalmente na terapia intensiva pediátrica e neonatal, que tem resultado em melhoria nas taxas de sobrevivência de pacientes com doenças anteriormente intratáveis, levando a um aumento de pacientes portadores de condições crônicas complexas, que apresentam maior risco de internações em UTIP do que a população geral (21,22). Odetola e col relatam aumento da proporção de pacientes com comorbidades admitidos em UTIP nos Estados Unidos entre 1996 e 2006 (35% e 41% respectivamente), com a observação que este tipo de paciente tem uma maior mortalidade e consumo de recursos do que os pacientes sem comorbidades (23).

No estudo de desenvolvimento do PIM2 os pacientes cirúrgicos constituíam 44,7% da amostra, diferente de nossa casuística na qual as admissões de pacientes cirúrgicos foi o mais prevalente (52,7%). A predominância de admissões envolvendo procedimentos cirúrgicos em estudos avaliando o PIM2 foi observada apenas nos

trabalhos de Wolfler e col (50%) (20) e de Imamura e col (67,7%) (24). Nos demais trabalhos avaliando o PIM2 a porcentagem de pacientes cirúrgicos variou entre 14 e 47,8% (19,25–34).

Em relação às admissões de pacientes com causas clínicas a predominância de doenças que acometiam o sistema respiratório (41,8%) foi semelhante ao estudo de desenvolvimento do PIM2 no qual também a principal causa foram doenças do sistema respiratório (39%) (8). No entanto, no estudo de desenvolvimento do PIM2 ocorreram admissões decorrentes de condições relacionadas a traumatismos (16,7%) (8), fato não observado em nosso estudo, uma vez que nossa instituição não é referência para crianças e adolescentes vítimas de traumatismo.

A prevalência aumentada de causas relacionadas às doenças do sistema cardiovascular cujo predomínio foi o choque séptico, difere do estudo de desenvolvimento do PIM2 onde causas neurológicas foram a segunda em importância (8). Essa diferença está associada ao grande número de pacientes com câncer submetidos à quimioterapia e que evoluíram com neutropenia febril na enfermaria do HC-UFMG, condição reconhecida como um fator de risco importante para o choque séptico, morte e aumento do tempo de internação (35).

A outra característica marcante da população foi a observação que 25,2% dos pacientes tinham o diagnóstico de câncer, fato não observado no estudo de desenvolvimento do PIM2 (8). Nos estudos avaliando o PIM2 a prevalência de crianças com diagnóstico com câncer tem sido bem abaixo da nossa casuística, Thukral e col (19) relataram que 5,6% da casuística constituída de pacientes com câncer, e Canonero e col (32) observaram 1,7% dos pacientes tinham o diagnóstico de câncer.

A maioria dos pacientes do estudo (54,6%) utilizou ventilação mecânica em algum período durante a internação na unidade. Não há informações sobre a proporção de pacientes que utilizou VM na amostra de desenvolvimento do PIM2 (8). A VM instituída de forma aguda é uma indicação absoluta para admissão na UTIP, sendo considerada um indicador da utilização adequada dos leitos em terapia intensiva (36). Nos estudos pesquisados avaliando o PIM2 a proporção de pacientes que utilizaram VM variou entre 21 e 86,5% (19,20,24,25,27–34), com dois autores relatando uma proporção de utilização de VM maior do que 50%. Thukral e col (19) relataram que 86,5% e Hariharam e col (33) relataram que 84% dos pacientes do estudo utilizaram VM.

Na análise dos desfechos foi evidenciado que a mortalidade geral (9,7%) era compatível com as taxas observadas atualmente na terapia intensiva pediátrica que tem variado entre 5 a 10% (37). A mortalidade dos pacientes clínicos apresentou diferença com significância estatística quando comparada com a mortalidade dos pacientes cirúrgicos (17,8% versus 2,5%; $p < 0,0001$), o que está associado à maior gravidade da doença dos pacientes clínicos indicada pela probabilidade de morte estimada pela PIM2 (10,2% versus 1,9%; $p < 0,0001$). A diferença de probabilidade de morte pode estar associada ao fato que em 95% dos procedimentos cirúrgicos foram eletivos, enquanto somente 2,5% dos pacientes clínicos foram admitidos de forma eletiva na unidade. Imamura e col em estudo avaliando o PIM2 observaram diferença de mortalidade significativa entre os pacientes clínicos e cirúrgicos (5% versus 1,2%), sendo atribuída à menor gravidade dos pacientes cirúrgicos que eram submetidos a procedimentos eletivos de média complexidade na maioria dos casos (24).

Não ocorreu diferença com significância estatística da mortalidade dos pacientes com condição crônica complexa quando comparada com a dos pacientes sem condição crônica complexa (10,3% versus 6,4%; $p = 0,27$), entretanto houve tendência para uma maior mortalidade no grupo com condição crônica complexa. A probabilidade de morte estimada pelo PIM2 também não apresentou diferença com significância estatística quando comparados os pacientes com e sem condição crônica complexa (5,9% versus 5,6%; $p = 0,5$). Canonero e col relatam que os escores de predição de desfecho utilizados na pediatria não consideram a presença de doenças crônicas como fator de risco, embora existam evidências que esses pacientes apresentam maior risco de morte (32).

Wolfler e col em estudo avaliando o PIM2 observaram uma mortalidade elevada nos pacientes com doença crônica quando comparada com a mortalidade geral (15,6% e 5,2% respectivamente) (20). Odetola e col observaram uma mortalidade significativamente maior nos pacientes com comorbidades em relação aos pacientes sem comorbidades admitidos nas UTIP dos Estados Unidos em 1997 (12,5% versus 8,6%) e 2006 (10,8% versus 7,8%) (23). Os autores sugeriram que esta diferença podia estar associada à menor gravidade da doença do grupo sem comorbidade, porém o estudo não utilizou nenhum método para mensurar a gravidade (23).

A mortalidade dos pacientes com neoplasia não apresentou diferença com significância estatística quando comparado com os pacientes sem neoplasia (12,8% versus 8,6%; $p = 0,15$), entretanto houve tendência para uma maior mortalidade no grupo com neoplasia. Também não houve diferença com significância estatística de gravidade

da doença estimada pelo PIM2 quando foram comparados os grupos com e sem neoplasia (6,9% versus 5,5%; $p=0,16$). Taxas de mortalidade elevadas em pacientes com neoplasia quando comparadas com pacientes sem neoplasia admitidos em UTIP são relatadas na literatura, com mortalidade entre os pacientes com câncer variando entre 23% e 55% (38–40). Meyer e col relatam que os escores de predição de desfecho são inadequados para prever mortalidade de pacientes pediátricos com câncer admitidos em UTIP (41).

O estudo evidenciou que os pacientes com condição crônica complexa ou neoplasia apresentaram tendência para uma maior mortalidade que os pacientes sem estas condições, entretanto sem significância estatística. Na literatura também são observadas maiores taxas de mortalidade nos pacientes com comorbidade ou neoplasia quando comparados com pacientes sem estas condições. A ausência de diferença estatística no estudo pode ser secundária ao tamanho da amostra.

Ao mesmo tempo não foi observada diferença com significância estatística ou tendência de uma maior probabilidade de morte estimada pelo PIM2 nos grupos com condição crônica complexa ou neoplasia quando comparados com os pacientes sem estas condições. Com isso levantamos a hipótese que o score pode não ser adequado para avaliar pacientes com condição crônica complexa ou neoplasia.

Na análise dos desfechos, o tempo médio de permanência da UTIP do HC-UFMG foi de 7,9 dias, informação não encontrada no estudo de desenvolvimento do PIM2 (8). O tempo de permanência na unidade sem ajuste de risco para a gravidade da doença apresenta as mesmas falhas que a taxa de mortalidade sem ajuste de risco quando são realizadas comparações entre populações (2). Ruttimann e col utilizaram o escore de predição de desfecho PRISM III para desenvolver um modelo de predição de tempo de permanência na UTIP que demonstrou validação adequada, porém de forma independente do escore original (42). Não foram encontrados estudos de forma similar com a utilização do PIM2. Entre os estudos pesquisados que avaliavam o PIM2, o tempo médio de permanência na UTIP variou entre 4 e 12,1 dias (19,20,24–34), porém qualquer comparação sem um ajuste de risco torna as conclusões menos fidedignas.

Foi observado uma menor mediana de tempo de permanência na unidade nos pacientes cirúrgicos (2 versus 5,5 dias; $p<0,0001$), o que pode ser associado à menor gravidade da doença estimada pelo PIM2 e à maior proporção da condição eletiva das admissões. No entanto não houve diferença com significância estatística da mediana do

tempo de permanência entre os pacientes com e sem condição complexa ($p=0,84$) e entre os pacientes com e sem neoplasia ($p=0,10$) (tabela 3).

Odetola e col observaram que os pacientes com comorbidades têm um maior tempo médio de permanência na UTIP que os pacientes sem comorbidades (23). Embora existam evidências que pacientes pediátricos com câncer tenham internações hospitalares prolongadas quando comparados com o restante da população (43), há poucos estudos sobre internações em UTIP. Pound e col observaram que crianças com câncer admitidos em uma UTIP para tratamento de choque séptico apresentaram menor tempo de internação na unidade do que pacientes sem câncer (44).

Na análise do desempenho do PIM2 foram utilizadas a discriminação e calibração do score, que são as duas medidas objetivas utilizadas para avaliar o desempenho dos scores de predição de desfecho (15). A discriminação do PIM2 no estudo foi adequada (área sob a curva ROC de 0,84 IC95% 0,810-0,867), próxima a descrita na validação interna do PIM2, onde foi evidenciado uma área sob a curva ROC de 0,90 (IC95% 0,89-0,92) (8). Em todos os estudos que avaliaram o PIM2 a discriminação foi considerada adequada, com a área sob a curva ROC variando entre 0,81 e 0,97.

Na avaliação da calibração, o teste de Hosmer-Lemeshow tem sido o utilizado em estudos que avaliam o PIM2 em populações diferentes da amostra de desenvolvimento do score (19,20,24-34), porém alguns autores têm utilizado diferentes categorizações da amostra em grupos com probabilidade de morte diferentes como 0-1%, 1-5%, 5-15%, 15-30%, >30% (29). Estas diferenças de categorizações dos pacientes podem interferir nas comparações dos desempenhos entre os estudos, podendo a análise ser afetada pela disposição dos sujeitos na categorização dos grupos, como demonstrado por Bertolini e col, que obtiveram aproximadamente um milhão de diferentes valores de p quando usaram todas as possíveis disposições dos sujeitos em uma amostra de quase 1400 pacientes (45). Por isso, optamos por realizar a categorização dos pacientes de forma similar ao estudo de desenvolvimento do PIM2, que consistia na categorização dos pacientes em 10 grupos com probabilidades de risco crescente e número similar de casos em cada grupo.

Na avaliação da nossa população quando avaliada a calibração do PIM2, esta foi considerada inadequada pelo teste de Hosmer-Lemeshow ($p < 0,0001$), o que difere do estudo original de desenvolvimento do score onde na validação interna foi adequada com um valor de $p = 0,42$ (8). Em quatro estudos encontrados avaliando o PIM2 foi

observada calibração inadequada do modelo, Slater e col em uma população australiana encontraram um valor de $p < 0,025$ e associam que a melhora do cuidado oferecido através do tempo, diferenças entre os casos, e limiares para começar e descontinuar o suporte intensivo foram os fatores que poderiam ter influenciado a calibração inadequada (25). Embora o PIM2 tenha apresentado calibração inadequada, os autores consideraram que o escore é apropriado para ser usado como instrumento de monitoramento de qualidade.

Mestrovic e col na Croácia (26) e Brady e col na Inglaterra (27) também encontraram calibração inadequada quando avaliaram o PIM2. No estudo inglês, os autores comentam que para um escore ter uma boa calibração quando testado em uma nova população, todos os fatores que influenciam o desfecho devem ser incluídos no escore ou ter a mesma distribuição na amostra de validação e na amostra de desenvolvimento do modelo. Eulmeseikian e col ($p= 0,0348$), quando avaliaram o PIM2 na Argentina, relataram que as principais razões para explicar a calibração inadequada foram a baixa taxa de mortalidade e a diferença na mistura de casos, no entanto consideraram que o PIM2 apresentou um bom desempenho na amostra do estudo (29).

Em decorrência de possíveis fatores “externos” que podem influenciar no valor do p do teste de Hosmer-Lemeshow, como o tamanho da amostra, Shann sugere analisar a tabela de contingência do teste como parte da avaliação da calibração (46). No estudo fica evidenciado que o escore superestimou a mortalidade nas faixas de menor risco (riscos entre 0,1-0,13%; 0,14-0,22%; e 0,23-0,57%) e subestimou a mortalidade nas faixas de médio risco (riscos entre 3,73-6,07% e 6,11-15,26%) (tabela 6), fortalecendo a hipótese de calibração inadequada do PIM2 em nossa casuística.

A interpretação dos resultados do teste de ajuste de Hosmer-Lemeshow quando escores de predição de desfecho são aplicados em novas populações deve ser realizada com cuidado (46). Marcin e col relatam que quando a calibração é avaliada em amostra independente da amostra de desenvolvimento do escore, uma calibração inadequada pode ser resultado de um escore inadequado, porém também pode ser um indicador de diferenças na qualidade de cuidado ofertado entre as amostras do estudo e de desenvolvimento do escore (2).

Em situações na qual a avaliação do escore numa amostra independente evidencia um desempenho inadequado algumas questões devem ser respondidas (47): A diferença de desempenho está relacionada com o pessoal responsável pelo cuidado dos pacientes? As opções terapêuticas e diagnósticas têm acessibilidade e uso similares? O

sistema de saúde do qual a UTIP faz parte é eficiente? A coleta dos dados foi adequada? O escore não é apropriado para uma diferente mistura de casos?

O desenho do estudo não permite responder as três primeiras perguntas que estão relacionadas à qualidade do cuidado de saúde ofertado na UTIP do HC-UFMG. Porém, pode ser destacado que a unidade faz parte de um sistema público de saúde de um país em desenvolvimento, que enfrenta dificuldades relacionadas ao financiamento, estrutura, pessoal e organização operacional que podem afetar negativamente na qualidade do cuidado de saúde ofertada.

A metodologia empregada na coleta de dados do estudo seguiu as orientações propostas pelos autores do PIM2 (8). Deve ser destacado que a população avaliada no presente estudo apresenta características diferentes em relação ao estudo de desenvolvimento do escore, e que duas das características mais marcantes (condição crônica complexa e neoplasia) não foram avaliadas no estudo de desenvolvimento do escore, o que pode ter interferido no desempenho do PIM2 em nosso estudo. Os autores do PIM2 relatam que existe a possibilidade do escore não funcionar adequadamente num cenário onde as características ou diagnósticos dos pacientes são substancialmente diferentes da amostra de desenvolvimento do modelo (8).

Slater e col relatam que quando a calibração é inadequada em uma nova população, a conclusão apropriada é que os resultados nesta nova população são diferentes dos resultados observados na população de desenvolvimento do escore, sendo inapropriado concluir que o escore é inválido apenas com o uso deste critério (25). Murphy-Filkins e col observaram que se certa porcentagem de determinado tipo de paciente é ultrapassada, o escore pode não ser acurado (48).

Slater e col relatam que além de calibrar bem através de diferentes faixas de risco, o escore deve calibrar adequadamente em diferentes grupos diagnósticos (25). Se um escore tem uma previsão inadequada para um determinado tipo de paciente, o desempenho global do modelo pode ser influenciado pela proporção de pacientes admitidos com este tipo de condição (25). No estudo de desenvolvimento do PIM2 a avaliação da calibração entre diferentes grupos diagnósticos não foi feita através do teste de Hosmer-Lemeshow, e sim através da taxa de mortalidade padronizada (8). Porém para testar a hipótese levantada por Ridley que o teste de Hosmer-Lemeshow pode ser afetado pelo tamanho da amostra (49), optamos por também realizar este teste nos subgrupos de pacientes.

A discriminação foi adequada em todos os subgrupos e em todas as faixas etárias. A calibração foi adequada na maioria dos subgrupos, apenas os subgrupos com condição crônica complexa e sem neoplasias apresentaram calibração inadequada. Com a observação que os subgrupos com calibração adequada tinham uma menor quantidade de pacientes (entre 109 e 357 pacientes) do que os subgrupos com calibração inadequada (entre 506 e 568 pacientes). Esse achado está de acordo com a hipótese levantada por Zhu e col que relatam que o p do teste de Hosmer-Lemeshow será pequeno em amostras muito grandes, e será o oposto em amostras muito pequenas (50).

Entretanto essa observação não é encontrada na análise da calibração entre as diferentes faixas etárias, onde foi adequada na maioria das faixas etárias, exceto no subgrupo com idade entre 121 e 215 meses, que possuía uma menor quantidade de pacientes em relação aos subgrupos com calibração adequada. Com isso, embora o tamanho da população possa ter afetado o valor do p do teste de Hosmer-Lemeshow no estudo, esse fator não pode ser considerado o único responsável.

A taxa de mortalidade padronizada de 1,65 (IC95% 1,26-2,04) indica que a qualidade do cuidado de saúde ofertado pela UTIP do HC-UFMG no período entre fevereiro de 2009 e janeiro de 2011 foi pior do que a qualidade do cuidado de saúde oferecido pelas UTIP que participaram do estudo de desenvolvimento do PIM2 no período entre 1997 e 1999. A TMP é o principal indicador de qualidade do cuidado de saúde na terapia intensiva pediátrica, fazendo parte do primeiro conjunto de indicadores de qualidade que foram submetidos à Joint Commission on Accreditation of Healthcare Organizations e publicado em 2005 (3). A precisão deste indicador dependerá da habilidade do escore em prever mortalidade na população estudada (4).

O desempenho da UTIP do HC-UFMG foi considerado médio nos pacientes sem condição crônica complexa (SMR= 1,14 IC95% 0,29-1,99) e nos pacientes cirúrgicos (SMR= 1,26 IC95% 0,44-2,08), indicando que a qualidade do cuidado de saúde oferecido para estes dois subgrupos na UTIP do HC-UFMG no período do estudo foi similar a qualidade do cuidado de saúde ofertado pelas UTIP participantes do estudo de desenvolvimento do PIM2 no período entre 1997 e 1999.

Nos restante dos subgrupos (pacientes clínicos, pacientes com condição crônica complexa, pacientes com neoplasia e pacientes sem neoplasia) o desempenho da UTIP do HC-UFMG foi considerado ruim, indicando uma diferença da qualidade do cuidado de saúde ofertado favorável as UTIP participantes do estudo de desenvolvimento do PIM2. Além de diferenças da qualidade do cuidado de saúde oferecido, a hipótese já

levantada que o escore pode não ser capaz de prever morte adequadamente em determinados tipos de pacientes é pertinente. Thukral e col observaram uma piora da capacidade de previsão do PIM2 quando eram comparadas populações com e sem doença crônica (SMR = 1,77 e 1,37 respectivamente) (19).

O estudo não é capaz de responder algumas questões levantadas na discussão que poderiam explicar a calibração inadequada do PIM2 e o desempenho ruim da UTIP do HC-UFG quando avaliada através da TMP. Com isso uma comparação com outras unidades através da TMP (benchmarking externo) fica inviabilizada. Entretanto o PIM2 ainda pode ser utilizado no processo de melhoria da qualidade de cuidado de saúde oferecido na UTIP do HC-UFG. Pollack relata que o uso mais importante dos escores é para benchmarking interno, ou seja, comparação do desempenho da unidade através do tempo (47). A avaliação seqüencial dentro de uma única unidade minimizaria questões relacionadas à diferença de misturas de casos, estrutura, equipamentos e recursos humanos (47).

5.6) Conclusão

As características mais marcantes da população do estudo foram o elevado número de admissões após procedimentos cirúrgicos e a alta prevalência de pacientes com condição crônica complexa e neoplasia.

O desempenho do PIM2 não foi adequado na população do estudo em decorrência da calibração inadequada do escore. A taxa de mortalidade padronizada indicou que a qualidade do cuidado ofertado na UTIP do HC-UFG no período do estudo foi pior do que a ofertada pelas UTIP participantes do desenvolvimento do PIM2. A diferença da mistura de casos pode ser um dos fatores responsáveis pela calibração e TMP ruim, porém a influência da qualidade do cuidado ofertado não pode ser descartada.

5.7) Bibliografia

1. Sachdeva RC, Jain S. Making the Case to Improve Quality and Reduce Costs in Pediatric Health Care. *Pediatric Clinics of North America*. 2009 Aug;56(4):731–43.

2. Marcin JP, Pollack MM. Review of the acuity scoring systems for the pediatric intensive care unit and their use in quality improvement. *Journal of intensive care medicine*. 2007 May-Jun;22(3):131–40.
3. Scanlon MC, Mistry KP, Jeffries HE. Determining pediatric intensive care unit quality indicators for measuring pediatric intensive care unit safety. *Pediatric critical care medicine : a journal of the Society of Critical Care Medicine and the World Federation of Pediatric Intensive and Critical Care Societies*. 2007 Mar;8(2 Suppl):S3–10.
4. Woodhouse D, Berg M, van der Putten J, Houtepen J. Will benchmarking ICUs improve outcome? *Current opinion in critical care*. 2009 Oct;15(5):450–5.
5. Pollack MM, Ruttimann UE, Getson PR. Pediatric risk of mortality (PRISM) score. *Critical care medicine*. 1988 Nov 16(11):1110–6.
6. Shann F, Pearson G, Slater A, Wilkinson K. Paediatric index of mortality (PIM): a mortality prediction model for children in intensive care. *Intensive care medicine*. 1997 Feb;23(2):201–7.
7. Pollack MM, Patel KM, Ruttimann UE. PRISM III: an updated Pediatric Risk of Mortality score. *Critical care medicine*. 1996 May;24(5):743–52.
8. Slater A, Shann F, Pearson G. PIM2: a revised version of the Paediatric Index of Mortality. *Intensive care medicine*. 2003 Feb;29(2):278–85.
9. Garcia PCR, Piva JP. Pediatric Index of Mortality 2 (PIM2)--A prognostic tool for developing countries: Easy, efficient, and free! *Pediatric critical care medicine : a journal of the Society of Critical Care Medicine and the World Federation of Pediatric Intensive and Critical Care Societies*. 2007 Jan;8(1):77–8.
10. Altman DG, Vergouwe Y, Royston P, Moons KGM. Prognosis and prognostic research: validating a prognostic model. *BMJ*. 2009 May 28;338(may28 1):b605–b605.
11. Agência Nacional de Vigilância Sanitária. Agência Nacional de Vigilância Sanitária - Regulamento Técnico para funcionamento de Unidades de Terapia Intensiva e Unidades de Cuidados Intermediários. 2006.
12. Feudtner C, Hays RM, Haynes G, Geyer JR, Neff JM, Koepsell TD. Deaths attributed to pediatric complex chronic conditions: national trends and implications for supportive care services. *Pediatrics*. 2001 Jun;107(6):E99.
13. Lacroix J, Cotting J. Severity of illness and organ dysfunction scoring in children. *Pediatric critical care medicine : a journal of the Society of Critical Care Medicine and the World Federation of Pediatric Intensive and Critical Care Societies*. 2005 May;6(3 Suppl):S126–34.
14. Keegan MT, Gajic O, Afessa B. Severity of illness scoring systems in the intensive care unit. *Critical care medicine*. 2011 Jan;39(1):163–9.

15. Afessa B, Gajic O, Keegan MT. Severity of illness and organ failure assessment in adult intensive care units. *Critical care clinics*. 2007 Jul;23(3):639–58.
16. Hosmer DW, Lemeshow S. Confidence interval estimates of an index of quality performance based on logistic regression models. *Statistics in medicine*. 1995 Oct 15;14(19):2161–72.
17. Randolph G, Esporas M, Provost L, Massie S, Bundy DG. Model for improvement - Part Two: Measurement and feedback for quality improvement efforts. *Pediatric clinics of North America*. 2009 Aug;56(4):779–98.
18. Poonam Bhadoria AGB. Severity Scoring Systems in Paediatric Intensive Care Units. *Indian Journal of Anaesthesia*. 2008;52(Suppl (5)):663–75.
19. Thukral A, Lodha R, Irshad M, Arora NK. Performance of Pediatric Risk of Mortality (PRISM), Pediatric Index of Mortality (PIM), and PIM2 in a pediatric intensive care unit in a developing country. *Pediatric critical care medicine : a journal of the Society of Critical Care Medicine and the World Federation of Pediatric Intensive and Critical Care Societies*. 2006 Jul;7(4):356–61.
20. Wolfler A, Silvani P, Musicco M, Salvo I. Pediatric Index of Mortality 2 score in Italy: a multicenter, prospective, observational study. *Intensive care medicine*. 2007 Aug;33(8):1407–13.
21. Burns KH, Casey PH, Lyle RE, Bird TM, Fussell JJ, Robbins JM. Increasing prevalence of medically complex children in US hospitals. *Pediatrics*. 2010 Oct;126(4):638–46.
22. Simon TD, Berry J, Feudtner C, Stone BL, Sheng X, Bratton SL, et al. Children with complex chronic conditions in inpatient hospital settings in the United States. *Pediatrics*. 2010 Oct;126(4):647–55.
23. Odetola FO, Gebremariam A, Davis MM. Comorbid illnesses among critically ill hospitalized children: Impact on hospital resource use and mortality, 1997-2006. *Pediatric critical care medicine : a journal of the Society of Critical Care Medicine and the World Federation of Pediatric Intensive and Critical Care Societies*. 2010 Jul;11(4):457–63.
24. Imamura T, Nakagawa S, Goldman RD, Fujiwara T. Validation of pediatric index of mortality 2 (PIM2) in a single pediatric intensive care unit in Japan. *Intensive care medicine*. 2012 Apr;38(4):649-54.
25. Slater A, Shann F. The suitability of the Pediatric Index of Mortality (PIM), PIM2, the Pediatric Risk of Mortality (PRISM), and PRISM III for monitoring the quality of pediatric intensive care in Australia and New Zealand. *Pediatric critical care medicine : a journal of the Society of Critical Care Medicine and the World Federation of Pediatric Intensive and Critical Care Societies*. 2004 Sep;5(5):447–54.

26. Mestrovic J, Kardum G, Polic B, Omazic A, Stricevic L, Sustic A. Applicability of the Australian and New Zealand Paediatric Intensive Care Registry diagnostic codes and Paediatric Index of Mortality 2 scoring system in a Croatian paediatric intensive care unit. *European journal of pediatrics*. 2005 Dec;164(12):783–4.
27. Brady AR, Harrison D, Black S, Jones S, Rowan K, Pearson G, et al. Assessment and optimization of mortality prediction tools for admissions to pediatric intensive care in the United kingdom. *Pediatrics*. 2006 Apr;117(4):e733–42.
28. Prieto Espuñes S, López-Herce Cid J, Rey Galán C, Medina Villanueva A, Concha Torre A, Martínez Camblor P. [Prognostic indexes of mortality in pediatric intensive care units]. *Anales de pediatría (Barcelona, Spain)*: 2003. 2007 Apr;66(4):345–50.
29. Eulmesekian PG, Pérez A, Mincez PG, Ferrero H. Validation of pediatric index of mortality 2 (PIM2) in a single pediatric intensive care unit of Argentina. *Pediatric critical care medicine: a journal of the Society of Critical Care Medicine and the World Federation of Pediatric Intensive and Critical Care Societies*. 2007 Jan;8(1):54–7.
30. Qureshi AU, Ali AS, Ahmad TM. Comparison of three prognostic scores (PRISM, PELOD and PIM 2) at pediatric intensive care unit under Pakistani circumstances. *Journal of Ayub Medical College, Abbottabad: JAMC*; 2007 Apr-Jun;19(2):49–53.
31. Hwang HS, Lee NY, Han SB, Kwak GY, Lee SY, Chung SY, et al. Performance effectiveness of pediatric index of mortality 2 (PIM2) and pediatric risk of mortality III (PRISM III) in pediatric patients with intensive care in single institution: Retrospective study. *Korean Journal of Pediatrics*. 2008;51(11):1158.
32. Canonero I, Figueroa A, Cacciamano A, Olivier E, Cuestas E. [Validation of PRISM and PIM2 scores of mortality in a pediatric intensive care unit in Cordoba]. *Archivos argentinos de pediatría*. 2010 Oct;108(5):427–33.
33. Hariharan S, Krishnamurthy K, Grannum D. Validation of Pediatric Index of Mortality-2 scoring system in a pediatric intensive care unit, Barbados. *Journal of tropical pediatrics*. 2011 Feb;57(1):9–13.
34. Ng DK, Miu T-yat, Chiu W-keung, Hui N-tat, Chan C-hong. Validation of Pediatric Index of Mortality 2 in three pediatric intensive care units in Hong Kong. *Indian journal of pediatrics*. 2011 Dec;78(12):1491–4.
35. Pound CM, Johnston DL, Armstrong R, Gaboury I, Menon K. The morbidity and mortality of pediatric oncology patients presenting to the intensive care unit with septic shock. *Pediatric blood & cancer*. 2008 Nov;51(5):584–8.
36. Mickell JJ, Furgiuele TL. Acute mechanical ventilation: an index of the absolute need for pediatric intensive care unit beds. *Critical care medicine*. 1988 May;16(5):504–9.

37. Namachivayam P, Shann F, Shekerdemian L, Taylor A, van Sloten I, Delzoppo C, et al. Three decades of pediatric intensive care: Who was admitted, what happened in intensive care, and what happened afterward. *Pediatric critical care medicine : a journal of the Society of Critical Care Medicine and the World Federation of Pediatric Intensive and Critical Care Societies*. 2010 Sep;11(5):549–55.
38. Dursun O, Hazar V, Karasu GT, Uygun V, Tosun O, Yesilipek A. Prognostic factors in pediatric cancer patients admitted to the pediatric intensive care unit. *Journal of pediatric hematology/oncology*. 2009 Jul;31(7):481–4.
39. Akhtar N, Fadoo Z, Panju S, Haque A. Outcome and prognostic factors seen in pediatric oncology patients admitted in PICU of a developing country. *Indian journal of pediatrics*. 2011 Aug;78(8):969–72.
40. Owens C, Mannion D, O’Marcaigh A, Waldron M, Butler K, O’Meara A. Indications for admission, treatment and improved outcome of paediatric haematology/oncology patients admitted to a tertiary paediatric ICU. *Irish journal of medical science*. 2011 Mar;180(1):85–9.
41. Meyer S, Gottschling S, Biran T, Georg T, Ehlayil K, Graf N, et al. Assessing the risk of mortality in paediatric cancer patients admitted to the paediatric intensive care unit: a novel risk score? *European journal of pediatrics*. 2005 Sep;164(9):563–7.
42. Ruttimann UE, Patel KM, Pollack MM. Length of stay and efficiency in pediatric intensive care units. *The Journal of pediatrics*. 1998 Jul;133(1):79–85.
43. Hendrickson KC, Rimar J. Patterns of hospital resource utilization of children with leukemia and CNS tumors: a comparison of children who survive and those who die within 3 years of diagnosis. *Nursing economic\$*. 2009 Jan-Feb;27(1):35–44.
44. Pound CM, Johnston DL, Armstrong R, Gaboury I, Menon K. The morbidity and mortality of pediatric oncology patients presenting to the intensive care unit with septic shock. *Pediatric blood & cancer*. 2008 Nov;51(5):584–8.
45. Bertolini G, D’Amico R, Nardi D, Tinazzi A, Apolone G. One model, several results: the paradox of the Hosmer-Lemeshow goodness-of-fit test for the logistic regression model. *Journal of epidemiology and biostatistics* . 2000 Jan;5(4):251–3.
46. Shann F. Are we doing a good job: PRISM, PIM and all that. *Intensive care medicine*. 2002 Feb;28(2):105–7.
47. Pollack MM. It is what it was. *Pediatric critical care medicine : a journal of the Society of Critical Care Medicine and the World Federation of Pediatric Intensive and Critical Care Societies*. 2011 Mar;12(2):228–9.

48. Murphy-Filkins R, Teres D, Lemeshow S, Hosmer DW. Effect of changing patient mix on the performance of an intensive care unit severity-of-illness model: how to distinguish a general from a specialty intensive care unit. *Critical care medicine*. 1996 Dec;24(12):1968–73.
49. Ridley S. Severity of illness scoring systems and performance appraisal. *Anaesthesia*. 1998 Dec;53(12):1185–94.
50. Zhu BP, Lemeshow S, Hosmer DW, Klar J, Avrunin J, Teres D. Factors affecting the performance of the models in the Mortality Probability Model II system and strategies of customization: a simulation study. *Critical care medicine*. 1996 Jan;24(1):57–63.



**UNIVERSIDADE FEDERAL DE MINAS GERAIS
COMITÊ DE ÉTICA EM PESQUISA - COEP**

Parecer nº. ETIC 0436.0.203.000-09

**Interessado(a): Prof. Alexandre Rodrigues Ferreira
Departamento de Pediatria
Faculdade de Medicina - UFMG**

DECISÃO

O Comitê de Ética em Pesquisa da UFMG – COEP aprovou, no dia 21 de outubro de 2009, o projeto de pesquisa intitulado "**Aplicação do Pediatric Index of Mortality (PIM2) em unidade de terapia intensiva pediátrica no Brasil**" bem como o Termo de Consentimento Livre e Esclarecido.

O relatório final ou parcial deverá ser encaminhado ao COEP um ano após o início do projeto.

**Profa. Maria Teresa Marques Amaral
Coordenadora do COEP-UFMG**

**TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO
(PESQUISA ENVOLVENDO CRIANÇAS ATÉ 6 ANOS)**

Estudo: Aplicação do Pediatric Index of Mortality 2 (PIM2) em Unidade de Terapia Intensiva Pediátrica no Brasil.

Senhores pais ou responsáveis,

Seu filho foi admitido no Centro de Terapia Intensiva Pediátrica do Hospital das Clínicas da Universidade Federal de Minas Gerais para cuidados médicos. Estamos realizando um estudo para avaliar um método de prever o prognóstico (escore de gravidade – PIM 2) de pacientes admitidos em Unidade de Terapia Intensiva Pediátrica. Para isso, estamos coletando alguns dados do exame clínico e resultados de exames de laboratório, para aplicar esse método em todos os pacientes admitidos neste serviço, após a concordância dos responsáveis legais do paciente. A aplicação deste método não implica em nenhum exame ou procedimento adicional aos rotineiramente empregados na assistência do paciente.

Esse método tem a intenção de ser uma ferramenta que ajude a equipe médica a prever a evolução do paciente após a sua admissão na unidade, auxiliando na decisão de tratamentos que serão fornecidos ao paciente e também fornecendo informações sobre a qualidade de tratamento ofertada em nosso serviço.

Todos os dados a respeito da participação da criança serão mantidos em segredo, sendo informados apenas aos pacientes, pais e responsáveis e à equipe médica que presta assistência ao paciente. É importante deixar claro que a participação do paciente não é obrigatória. Caso haja recusa na participação desse estudo, o paciente continuará recebendo os mesmos cuidados que normalmente lhe são fornecidos, sem nenhum prejuízo.

Favor preencher as seguintes informações, caso concorde com a participação do seu filho:

Confirmo que estou ciente das explicações fornecidas e que é de minha livre vontade que concordo com a participação de _____ no estudo “Aplicação do Pediatric Index of Mortality 2 (PIM2) em unidade de terapia intensiva pediátrica no Brasil”. Também confirmo que não estou recebendo nenhum tipo de incentivo financeiro por esta participação e que me foi fornecida uma segunda via deste termo.

Belo Horizonte, ____/____/____

()pai, ()mãe ou ()responsável

Em caso de dúvida, estaremos à disposição para maiores esclarecimentos.

Jaisson Gustavo da Fonseca

Endereço e telefone do pesquisador:

Rua Gonçalves Dias, 1899, apto.704, Bairro de Lourdes, Belo Horizonte Tel.: 3353-4888 ou 9108-5295

Comitê de Ética em Pesquisa da UFMG (COEP) – tel (31) 3409-4592 Av. Antônio Carlos, nº 6627
Unidade Administrativa II – 2º andar – Campus Pampulha. Belo Horizonte, MG – CEP 31270-901.

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO
(PESQUISA ENVOLVENDO CRIANÇAS ENTRE 7 A 12 ANOS)

Estudo: Aplicação do Pediatric Index of Mortality 2 (PIM2) em unidade de terapia intensiva pediátrica no Brasil.

Senhores pais ou responsáveis,

Seu filho foi admitido no Centro de Terapia Intensiva Pediátrica do Hospital das Clínicas da Universidade Federal de Minas Gerais para cuidados médicos. Estamos realizando um estudo para avaliar um método de prever o prognóstico (escore de gravidade – PIM 2) de pacientes admitidos em Unidade de Terapia Intensiva Pediátrica. Para isso, estamos coletando alguns dados do exame clínico e resultados de exames de laboratório, para aplicar esse método em todos os pacientes admitidos neste serviço, após a concordância dos responsáveis legais do paciente. A aplicação deste método não implica em nenhum exame ou procedimento adicional aos rotineiramente empregados na assistência do paciente.

Esse método tem a intenção de ser uma ferramenta que ajude a equipe médica a prever a evolução do paciente após a sua admissão na unidade, auxiliando na decisão de tratamentos que serão fornecidos ao paciente e também fornecendo informações sobre a qualidade de tratamento ofertada em nosso serviço.

Todos os dados a respeito da participação da criança serão mantidos em segredo, sendo informados apenas aos pacientes, pais e responsáveis e à equipe médica que presta assistência ao paciente. É importante deixar claro que a participação do paciente não é obrigatória. Caso haja recusa na participação desse estudo, o paciente continuará recebendo os mesmos cuidados que normalmente lhe são fornecidos, sem nenhum prejuízo.

Favor preencher as seguintes informações, caso concorde com a participação do seu filho:

Confirmando que estou ciente das explicações fornecidas e que é de minha livre vontade que _____ concordo com a participação de _____ no estudo “Aplicação do Pediatric Index of Mortality 2 (PIM2) em unidade de terapia intensiva pediátrica no Brasil”. Também confirmo que não estou recebendo nenhum tipo de incentivo financeiro por esta participação e que me foi fornecida uma segunda via deste termo.

Belo Horizonte, ____/____/____

()pai, ()mãe ou ()responsável

Texto a ser lido para a criança que participará do estudo:

Você está internado neste hospital para tratar um problema de saúde, estamos fazendo uma pesquisa para estudar uma forma que nos ajude a perceber a chance de você melhorar do problema de saúde. Para isso coletamos dados dos exames de sangue e de quando

examinamos você, mas não será realizado nenhum exame a mais. Caso você aceite participar do estudo, após aprovação dos seus pais, você poderia assinar seu nome abaixo.

Belo Horizonte, ____/____/____

Criança

Em caso de dúvida, estaremos à disposição para maiores esclarecimentos.

NOMES DOS PESQUISADORES RETIRADOS.

Comitê de Ética em Pesquisa da UFMG (COEP) – tel (31) 3409-4592
Av. Antônio Carlos, nº 6627 Unidade Administrativa II – 2º andar – Campus Pampulha.
Belo Horizonte, MG – CEP 31270-901.

**TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO
(PESQUISA ENVOLVENDO ADOLESCENTES ENTRE 13 A 17 ANOS)**

PARA OS PAIS RESPONSÁVEIS

Estudo: Aplicação do Pediatric Index of Mortality 2 (PIM2) em unidade de terapia intensiva pediátrica no Brasil.

Senhores pais ou responsáveis,

Seu filho foi admitido no Centro de Terapia Intensiva Pediátrica do Hospital das Clínicas da Universidade Federal de Minas Gerais para cuidados médicos. Estamos realizando um estudo para avaliar um método de prever o prognóstico (score de gravidade – PIM 2) de pacientes admitidos em Unidade de Terapia Intensiva Pediátrica. Para isso, estamos coletando alguns dados do exame clínico e resultados de exames de laboratório, para aplicar esse método em todos os pacientes admitidos neste serviço, após a concordância dos responsáveis legais do paciente. A aplicação deste método não implica em nenhum exame ou procedimento adicional aos rotineiramente empregados na assistência do paciente.

Esse método tem a intenção de ser uma ferramenta que ajude a equipe médica a prever a evolução do paciente após a sua admissão na unidade, auxiliando na decisão de tratamentos que serão fornecidos ao paciente e também fornecendo informações sobre a qualidade de tratamento ofertada em nosso serviço.

Todos os dados a respeito da participação da criança serão mantidos em segredo, sendo informados apenas aos pacientes, pais e responsáveis e à equipe médica que presta assistência ao paciente. É importante deixar claro que a participação do paciente não é obrigatória. Caso haja recusa na participação desse estudo, o paciente continuará recebendo os mesmos cuidados que normalmente lhe são fornecidos, sem nenhum prejuízo.

Favor preencher as seguintes informações, caso concorde com a participação do seu filho:

Confirmando que estou ciente das explicações fornecidas e que é de minha livre vontade que concordo com a participação de _____ no estudo “Aplicação do Pediatric Index of Mortality 2 (PIM2) em unidade de terapia intensiva pediátrica no Brasil”. Também confirmo que não estou recebendo nenhum tipo de incentivo financeiro por esta participação e que me foi fornecida uma segunda via deste termo.

Belo Horizonte, ____/____/____

()pai, ()mãe ou ()responsável

Para os pacientes adolescentes:

Estudo: Aplicação do Pediatric Index of Mortality 2 (PIM2) em unidade de terapia intensiva pediátrica no Brasil.

Ao paciente,

Você foi admitido no Centro de Terapia Intensiva Pediátrica do Hospital das Clínicas da Universidade Federal de Minas Gerais para cuidados médicos. Estamos realizando um estudo para avaliar um método de prever o prognóstico (escore de gravidade – PIM 2) de pacientes admitidos em Unidade de Terapia Intensiva Pediátrica. Para isso, estamos coletando alguns dados do exame clínico e resultados de exames de laboratório, para aplicar esse método em todos os pacientes admitidos neste serviço, após a concordância dos pacientes. A aplicação deste método não implica em nenhum exame ou procedimento adicional aos rotineiramente empregados na assistência do paciente.

Esse método tem a intenção de ser uma ferramenta que ajude a equipe médica a prever a evolução do paciente durante sua internação, auxiliando na decisão de tratamentos que serão dados ao paciente e também fornecendo informações sobre a qualidade de tratamento oferecida.

Todos os dados a respeito da participação do paciente serão mantidos em segredo, sendo informados apenas aos pacientes, pais e responsáveis e à equipe médica que presta assistência ao paciente. É importante deixar claro que a participação do paciente não é obrigatória. Caso haja recusa na participação desse estudo, você continuará recebendo os mesmos cuidados que normalmente lhe são fornecidos, sem nenhum prejuízo.

Favor preencher as seguintes informações, caso concorde com a sua participação:

Confirmando que estou ciente das explicações fornecidas e que é de minha livre vontade que concordo com a participação no estudo “Aplicação do Pediatric Index of Mortality 2 (PIM2) em Unidade de Terapia Intensiva Pediátrica no Brasil”. Também confirmo que não estou recebendo nenhum tipo de incentivo financeiro por esta participação e que me foi fornecida uma segunda via deste termo.

Belo Horizonte, ____/____/____

Paciente

Em caso de dúvida, estaremos a disposição para maiores esclarecimentos.

NOMES DOS PESQUISADORES RETIRADOS

Comitê de Ética em Pesquisa da UFMG (COEP) – tel (31) 3409-4592
Av. Antônio Carlos, nº 6627 Unidade Administrativa II – 2º andar – Campus Pampulha.
Belo Horizonte, MG – CEP 31270-901.



FACULDADE DE MEDICINA
CENTRO DE PÓS-GRADUAÇÃO

Av. Prof. Alfredo Balena 190 / sala 533
Belo Horizonte - MG - CEP 30.130-100
Fone: (031) 3409.9641 FAX: (31) 3409.9640
cpq@medicina.ufmg.br



DECLARAÇÃO

A Comissão Examinadora abaixo assinada, composta pelos Professores Doutores: Alexandre Rodrigues Ferreira, Vandack Alencar Nobre Júnior e Henrique Assis Fonseca Tonelli aprovou a dissertação de mestrado intitulada: "APLICAÇÃO DO PAEDIATRIC INDEX OF MORTALITY 2 EM UNIDADE DE TERAPIA INTENSIVA PEDIÁTRICA NO BRASIL" apresentada pelo mestrando JAISSON GUSTAVO DA FONSECA para obtenção do título de Mestre pelo Programa de Pós-Graduação em Ciências da Saúde – Área de Concentração em Saúde da Criança e do Adolescente da Faculdade de Medicina da Universidade Federal de Minas Gerais, realizada em 28 de maio de 2012.

Prof. Alexandre Rodrigues Ferreira

Orientador

Prof. Vandack de Alencar Nobre Júnior

Prof. Henrique de Assis Fonseca Tonelli



FACULDADE DE MEDICINA
CENTRO DE PÓS-GRADUAÇÃO

Av. Prof. Alfredo Balena 190 / sala 221
Belo Horizonte - MG - CEP 30130-100
Fone: (31) 3408 5511 FAX: (31) 3405 5540
graduacao@ufmg.br



ATA DA DEFESA DE DISSERTAÇÃO DE MESTRADO de JAISSON GUSTAVO DA FONSECA nº de registro 2010662924. Às treze horas e trinta minutos, do dia vinte e oito de maio de dois mil e doze, reuniu-se na Faculdade de Medicina da UFMG, a Comissão Examinadora de dissertação indicada pelo Colegiado do Programa, para julgar, em exame final, o trabalho intitulado: "APLICAÇÃO DO PAEDIATRIC INDEX OF MORTALITY 2 EM UNIDADE DE TERAPIA INTENSIVA PEDIÁTRICA NO BRASIL", requisito final para a obtenção do Grau de Mestre em Ciências da Saúde: Saúde da Criança e do Adolescente, pelo Programa de Pós-Graduação em Ciências da Saúde: Saúde da Criança e do Adolescente. Abrindo a sessão, o Presidente da Comissão, Prof. Alexandre Rodrigues Ferreira, após dar a conhecer aos presentes o teor das Normas Regulamentares do trabalho final, passou a palavra ao candidato para apresentação de seu trabalho. Seguiu-se a arguição pelos examinadores, com a respectiva defesa do candidato. Logo após, a Comissão se reuniu sem a presença do candidato e do público para julgamento e expedição do resultado final. Foram atribuídas as seguintes indicações:

Prof. Alexandre Rodrigues Ferreira - Orientador	instituição: UFMG	Indicação: <u>Aprovado</u>
Prof. Vandack de Alencar Nobre Júnior	instituição: UFMG	Indicação: <u>Aprovado</u>
Prof. Henrique de Assis Fonseca Tonelli	instituição: UFMG	Indicação: <u>Aprovado</u>

Pelas indicações o candidato foi considerado Aprovado

O resultado final foi comunicado publicamente ao candidato pelo Presidente da Comissão. Nada mais havendo a tratar, o Presidente encerrou a sessão e lavrou a presente ATA, que será assinada por todos os membros participantes da Comissão Examinadora. Belo Horizonte, 28 de maio de 2012.

Prof. Alexandre Rodrigues Ferreira / Orientador Alexandre Rodrigues Ferreira

Prof. Vandack de Alencar Nobre Júnior Vandack de Alencar Nobre Júnior

Prof. Henrique de Assis Fonseca Tonelli Henrique de Assis Fonseca Tonelli

Prof. Ana Cristina Simões e Silva/Coordenadora Ana Cristina Simões e Silva

Prof. Ana Cristina Simões e Silva
Coordenadora do Centro de Pós-Graduação em
Ciências da Saúde: Saúde da Criança e do Adolescente
Faculdade de Medicina - UFMG

Obs.: Este documento não terá validade sem a assinatura e carimbo do Coordenador

ANCS
CONFERE COM ORIGINAL
Centro de Pós-Graduação
Faculdade de Medicina - UFMG