

**UNIVERSIDADE FEDERAL DE MINAS GERAIS
FACULDADE DE MEDICINA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS APLICADAS À SAÚDE DO
ADULTO**

SARA MONTEIRO DE MORAES

**AVALIAÇÃO DA VALIDADE DO *GLOBAL TRIGGER TOOL* DO *INSTITUTE FOR
HEALTHCARE IMPROVEMENT (IHI)* PARA IDENTIFICAÇÃO DE EVENTOS
ADVERSOS EM PACIENTES ADULTOS HOSPITALIZADOS**

Belo Horizonte
2022

SARA MONTEIRO DE MORAES

AVALIAÇÃO DA VALIDADE DO *GLOBAL TRIGGER TOOL* DO *INSTITUTE FOR HEALTHCARE IMPROVEMENT* PARA IDENTIFICAÇÃO DE EVENTOS ADVERSOS EM PACIENTES ADULTOS HOSPITALIZADOS

Tese apresentada ao Programa de Pós-Graduação em Ciências Aplicadas à Saúde do Adulto da Universidade Federal de Minas Gerais, como requisito para a obtenção do grau de Doutor.

Orientadora: Prof^a. Dr^a. Teresa Cristina de Abreu Ferrari

Co-orientadora: Dr^a. Aline Maria Rezende Beleigoli

Belo Horizonte
2022

M827a Moraes, Sara Monteiro de.
Avaliação da validade do Global Trigger Tool do Institute for Healthcare Improvement (IHI) para identificação de eventos adversos em pacientes adultos hospitalizados [recursos eletrônicos]. / Sara Monteiro de Moraes. - - Belo Horizonte: 2022.
148f.: il.
Formato: PDF.
Requisitos do Sistema: Adobe Digital Editions.

Orientador (a): Teresa Cristina de Abreu Ferrari.
Coorientador (a): Alline Maria Rezende Beleigoli.
Área de concentração: Gestão da Assistência em Saúde.
Tese (doutorado): Universidade Federal de Minas Gerais, Faculdade de Medicina.

1. Segurança do Paciente. 2. Dano ao Paciente. 3. Gestão de Riscos. 4. Gestão em Saúde. 5. Qualidade da Assistência à Saúde. 6. Dissertação Acadêmica. I. Ferrari, Teresa Cristina de Abreu. II. Beleigoli, Alline Maria Rezende. III. Universidade Federal de Minas Gerais, Faculdade de Medicina. IV. Título.

NLM: WX 185

Bibliotecário responsável: Fabian Rodrigo dos Santos CRB-6/2697

Universidade Federal de Minas Gerais

Reitora: Profa. Sandra Regina Goulart Almeida

Vice-Reitor: Prof. Alessandro Fernandes Moreira

Pró-reitora de Pós-graduação: Profa. Isabela Almeida Pordeus

Pró-reitor de Pesquisa: Prof. Fernando Marcos dos Reis

Faculdade de Medicina

Diretora: Profa. Alamanda Kfoury Pereira

Vice-Diretora: Profa. Cristina Gonçalves Alvim

Departamento de Clínica Médica

Chefe do departamento: Eliane Viana Mancuzo

Programa de Pós-Graduação em Ciências Aplicadas à Saúde do Adulto

Coordenadora: Teresa Cristina de Abreu Ferrari

Subcoordenadora: Profa. Luciana Costa Faria

Membros do Colegiado:

Profa. Cláudia Alves Couto

Profa. Gilda Aparecida Ferreira

Profa. Karina Braga Gomes Borges

Profa. Luciana Costa Faria

Profa. Luciana Diniz Silva

Profa. Melissa Orlandin Premaor

Profa. Teresa Cristina de Abreu Ferrari



UNIVERSIDADE FEDERAL DE MINAS GERAIS
FACULDADE DE MEDICINA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS APLICADAS À SAÚDE DO ADULTO

FOLHA DE APROVAÇÃO

AVALIAÇÃO DA VALIDADE DO GLOBAL TRIGGER TOOL DO INSTITUTE FOR HEALTHCARE IMPROVEMENT (IHI) PARA IDENTIFICAÇÃO DE EVENTOS ADVERSOS EM PACIENTES ADULTOS HOSPITALIZADOS

SARA MONTEIRO DE MORAES

Tese de Doutorado defendida e aprovada, no dia dois de setembro de dois mil e vinte e dois, pela Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em Ciências Aplicadas à Saúde do Adulto da Universidade Federal de Minas Gerais constituída pelos seguintes professores doutores:

Teresa Cristina de Abreu Ferrari - Orientadora
UFMG

Alline Maria Rezende Beleigoli - Coorientadora
Universidade Flinders

Sandhi Maria Barreto
UFMG

Luciana de Gouvêa Viana
UFMG

Silvana Márcia Bruschi Kelles
PUC Minas

Helena Barreto dos Santos
HCPA

Paulo Sousa
ENSP-NOVA

Belo Horizonte, 02 de setembro de 2022.



Documento assinado eletronicamente por **Alline Maria Rezende Beleigoli**, **Usuário Externo**, em 05/09/2022, às 11:50, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Luciana de Gouvea Viana**, **Professora do Magistério Superior**, em 05/09/2022, às 12:59, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Silvana Márcia Bruschi Kelles**, **Usuária Externa**, em 05/09/2022, às 16:38, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Teresa Cristina de Abreu Ferrari**, **Professora do Magistério Superior**, em 05/09/2022, às 17:54, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Helena Barreto dos Santos**, **Usuária Externa**, em 21/09/2022, às 19:41, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Paulo Jorge dos Santos Sousa**, **Usuário Externo**, em 22/09/2022, às 14:28, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Sandhi Maria Barreto**, **Membro de comissão**, em 27/09/2022, às 09:56, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **1716031** e o código CRC **5A0BF58A**.

Agradecimentos

Agradeço à Universidade Federal de Minas Gerais, professores, técnicos administrativos e colaboradores. Meu carinho e orgulho por fazer parte dessa instituição pública e de excelência.

À Profa. Teresa Ferrari e à Profa. Alline Beleigoli, minhas orientadoras e parceiras desde a concepção do projeto de pós-graduação.

À Nathália Mansur, Bruna Mattos e Paloma de Oliveira, colegas de pós-graduação com quem compartilhei a experiência acadêmica.

Aos alunos de iniciação científica João Marcos Azevedo, Thais Alves, Caio Lôbo e Yane Andrade, que contribuíram com zelo e compromisso para a coleta e organização dos dados.

Aos alunos de graduação vinculados ao projeto de extensão Capítulo Open School IHI, que não mediram esforços para fazer a interlocução com as áreas de cuidado e com os profissionais da assistência.

Aos colegas do Hospital das Clínicas da UFMG, que entenderam a importância de identificarmos os incidentes de cuidado no contexto de aprendizado e busca por melhorias e aceitaram nosso convite para compartilharem as vivências e desafios na oferta de um cuidado seguro e de qualidade.

À equipe da América Latina do IHI, em especial ao Dr. Paulo Borem, que acolheu a solicitação para o uso da ferramenta GTT no contexto de pesquisa e proporcionou oportunidades para que eu pudesse conhecer a Ciência da Melhoria e me apaixonar pela força transformadora que ferramentas de qualidade podem ter nos serviços de saúde e na saúde das populações.

Aos meus amigos e familiares, que conviveram com minhas angústias e ausências e me apoiaram carinhosamente nesses últimos anos.

RESUMO

A segurança do paciente é um princípio fundamental do cuidado e um componente crítico na oferta de assistência em saúde de qualidade. Seu aprimoramento depende da capacidade de compreender a frequência de ocorrência e os tipos de eventos adversos, com o intuito de priorizar ações de melhoria. Para isso, são necessárias estratégias de mensuração válidas, reprodutíveis e custo-efetivas. O *Global Trigger Tool (GTT)* do *Institute for Healthcare Improvement (IHI)* é uma ferramenta que propõe a estimativa da ocorrência de eventos adversos por um método simples, barato e de fácil execução. Ele tem sido utilizado nos últimos anos como estratégia de mensuração em segurança do paciente por diversas instituições e alguns programas nacionais. Embora a maior capacidade do *GTT* em identificar eventos adversos em comparação a outros métodos já tenha sido demonstrada, há poucos dados sobre sua acurácia. Além disso, estudos sugerem que os eventos adversos que geram danos mais leves ao paciente são mais difíceis de serem identificados pelo *GTT* e que a exclusão desses eventos poderia aumentar a validade do método.

Objetivos Avaliar a validade do *GTT* para identificação de eventos adversos em pacientes adultos internados considerando-se todos os eventos adversos e o subgrupo de eventos adversos com maior dano ao paciente. **Métodos** Trata-se de um estudo de teste diagnóstico, no qual o *GTT* representa o teste índice e a identificação de eventos adversos representa a condição de interesse. Devido à inexistência de um teste padrão ouro, um método padrão referência composto foi construído. O método padrão referência utilizou diferentes estratégias para identificação da ocorrência de eventos adversos, combinando a coleta de informações durante as internações e a análise retrospectiva dos dados clínicos e administrativos. Ambos os testes foram aplicados em uma amostra aleatória de 211 internações de pacientes adultos ocorridas em outubro/novembro de 2016 em um hospital público e de grande porte de Belo Horizonte, Brasil. A validade do *GTT* foi avaliada por meio do coeficiente *Kappa (K)* ponderado e de medidas clássicas de acurácia – sensibilidade, especificidade e acurácia global. A classificação dos eventos adversos quanto à gravidade do dano foi baseada no índice de categorização de erros de medicação do *National Coordinating Council for Medication Error Reporting and Prevention (NCC MERP)* adaptado pelo *IHI*, sendo os eventos de menor gravidade classificados como “E” e os de maior gravidade

como “F-I”. **Resultados** Foram identificados 176 eventos adversos em 67 internações pelo método padrão referência e 129 eventos adversos em 76 internações pelo *GTT*, resultando em taxas de 126 e 93 eventos adversos/1.000 pacientes-dia, respectivamente. Houve concordância substancial entre os métodos para identificação de internações com evento adverso, com $K=0,61$ (Intervalo de Confiança [IC] 95% 0,49 a 0,71), e concordância discreta para identificação de cada evento adverso, com $K=0,09$ (IC 95% 0,02 a 0,19). Considerando o subgrupo de eventos adversos de maior dano (F-I), houve concordância substancial entre os métodos, com $K=0,80$ (IC 95% 0,70 a 0,88) para identificação de internações com evento adverso e $K=0,74$ (IC 95% 0,64 a 0,83) para identificação de cada evento adverso. Os dados de sensibilidade, especificidade e acurácia global do *GTT* para a identificação de eventos adversos individuais foram, respectivamente, 0,41 (IC 95% 0,34;0,49), 0,68 (IC 95% 0, 60;0,74) e 0,54 (IC 95% 0,49; 0,60) para todos os eventos adversos, independentemente da categorização de dano, e 0,85 (IC 95% 0,72;0,93), 0,88 (IC 95% 0,82;0,92) e 0,87 (IC 95% 0,82;0,91) para o subgrupo de eventos adversos categorizados como dano F-I. Entre os eventos adversos perdidos pelo *GTT* estão, principalmente, eventos relacionados aos cuidados de enfermagem, como os relacionados a acesso venoso periférico e cateteres gástricos/entéricos. **Conclusão** O *GTT* mostrou-se um método válido para identificação de eventos adversos em pacientes adultos hospitalizados. Sua acurácia aumenta quando os eventos adversos que geram danos menores para os pacientes não são contados. Entre os principais eventos adversos perdidos pelo *GTT* estão aqueles relacionados aos cuidados de enfermagem. Portanto, o *GTT* deve ser utilizado em conjunto com outras estratégias de mensuração para alcançar resultados representativos do perfil de qualidade da assistência prestada e, assim, nortear as melhores estratégias de melhoria.

Palavras-chave: Segurança do Paciente, Eventos Adversos, Global Trigger Tool.

ABSTRACT

Patient safety is a fundamental principle of care and a critical component of quality healthcare. Its improvement depends on the ability to understand the frequency of occurrence and types of adverse events, in order to prioritize improvement actions. For this purpose, valid, reproducible and cost-effective measurement strategies are needed. The Global Trigger Tool (GTT) of the Institute for Healthcare Improvement (IHI) is a tool that proposes the estimation of the occurrence of adverse events through a simple, inexpensive and easy-to-execute method. It has been used in recent years as a measurement strategy in patient safety by several institutions and some national programs. Although the greater ability of the GTT to identify adverse events compared to other methods has already been demonstrated, there are few data on its accuracy. Furthermore, studies suggest that the minor harm AEs are more difficult to be identified by the GTT and that the exclusion of these events could increase the validity of the method. **Objectives** To assess the validity of the GTT for identifying adverse events in adult hospitalized patients for all adverse events and for the subgroup of adverse events with greater harm to the patient. **Methods** This is a diagnostic test study, in which the GTT is the index test and the identification of adverse events represents the condition of interest. Due to the inexistence of a gold standard test, a composite reference standard method was constructed combining the collection of information during hospitalizations and retrospective administrative and clinical data analysis for AEs identification. Both tests were applied to a random sample of 211 hospitalizations of adult patients, occurred in Oct/Nov 2016 at a large public hospital in Belo Horizonte, Brazil. The validity of the GTT was evaluated using weighted Kappa coefficient and classical measures of accuracy – sensitivity, specificity and global accuracy. The classification of adverse events in terms of severity of harm was based on the National Coordinating Council for Medication Error Reporting and Prevention (NCC MERP) Index for Categorizing Errors adapted by the IHI, with less severe events classified as “E” and those with greater severity gravity as “F-I”. **Results** A total of 176 adverse events were identified in 67 admissions using the reference standard method and 129 adverse events in 76 admissions using the GTT, resulting in rates of 126 and 93 adverse events /1,000 patient-days, respectively. There was substantial agreement between the methods for identification of admissions with adverse event ($K=0.61$; 95% confidence interval [CI] 0.49 to 0.71),

and slight agreement for identification of each adverse event ($K=0.09$; 95%CI 0.02 to 0.19). Considering the subgroup of adverse events of greater harm (F-I), there was substantial agreement between the methods, with $K=0.80$ (95% CI 0.70 to 0.88) for identification of admissions with adverse event and $K=0.74$ (95% CI 0.64 to 0.83) for identification of each adverse event. The sensitivity, specificity and global accuracy data of the GTT for the identification of individual adverse events were, respectively, 0.41 (95% CI 0.34;0.49), 0.68 (95% CI 0.60;0.74) and 0.54 (95% CI 0.49;0.60) for all adverse events, regardless of the harm categorization, and 0.85 (95% CI 0.72;0.93), 0.88 (95% CI 0.82;0.92) and 0.87 (95% CI 0.82;0.91) for the subgroup of adverse events categorized as harm F-I. Among the adverse events lost by the GTT are mainly adverse events related to nursing care, such as those related to peripheral venous access and gastric/enteric catheters. **Conclusion** The GTT proved to be a valid method for identifying adverse events in hospitalized adult patients. Its accuracy increases when minor harm AEs are not counted. Among the main adverse events missed by the GTT are those related to nursing care. Therefore, the GTT should be used in conjunction with other measurement strategies to achieve results that are representative of the quality of care provided and, thus, guide the best improvement strategies.

Key-words: Patient Safety, Adverse Events, Global Trigger Tool.

LISTA DE FIGURAS

Figura 1 - Tabela 2 x 2 para avaliação da acurácia de um teste diagnóstico.	48
Figura 2 - Fluxograma de orientação para utilização de métodos para avaliação de testes diagnósticos em situações “sem padrão ouro”.	56
Figura 3 – Método padrão referência para identificação de eventos adversos	84
Figura 4 – Método <i>GTT</i> para identificação de eventos adversos	87

LISTA DE TABELAS

Tabela 1 - Prevalência dos eventos adversos, proporção de concordância e coeficiente de Kappa ajustado por prevalência do <i>GTT</i> e método padrão referência construído para todos os eventos adversos (E-I) e para o subgrupo de eventos mais graves (F-I).....	120
---	-----

LISTA DE QUADROS

Quadro 1 – Alternativas metodológicas para construção de um padrão referência em estudos de avaliação de validade de testes diagnósticos “sem padrão ouro”	52
Quadro 2 – As quatro fases de estudos de testes diagnósticos.....	71
Quadro 3 - Escala e instruções fornecidas aos médicos revisores para julgar a existência de um evento adverso	74
Quadro 4 - Classificação dos eventos adversos quando à gravidade do dano	75
Quadro 5 - Valores de Kappa e respectivos níveis de concordância	88

SUMÁRIO

1	Considerações iniciais.....	16
1.1	Introdução	16
1.2	Fundamentação teórica.....	19
1.2.1	O impacto dos cuidados inseguros em saúde	19
1.2.2	Aprender com os erros: a base do movimento pela segurança do paciente	21
1.2.3	A importância de medir para melhorar e a escolha de métodos adequados de medição em melhoria.....	24
1.2.3.1	Análise de dados rotineiramente já coletados	26
1.2.3.2	Revisão retrospectiva de prontuários	28
1.2.3.3	Trigger Tools	33
1.2.4	<i>Global Trigger Tool</i> como estratégia de medição em melhoria.....	34
1.2.4.1	Adaptações do GTT	36
1.2.5	A avaliação do <i>Global Trigger Tool</i> como um teste diagnóstico	41
1.2.5.1	Confiabilidade ou precisão	41
1.2.5.2	Validade ou acurácia.....	45
1.2.6	Avaliação da acurácia de testes diagnósticos em situações sem padrão ouro	47
1.3	Referências bibliográficas	57
2	Objetivos	69
2.1	Objetivo Geral.....	69
2.2	Objetivos específicos	69
3	Método	70
3.1	Desenho do estudo.....	70
3.2	População do estudo	71
3.3	Cálculo amostral	72
3.4	Definições	73

3.5 Variáveis	75
3.6 Aplicação dos testes índice e referência.....	77
3.6.1 Método padrão referência	78
3.6.2 Global Trigger Tool.....	84
3.7 Estatística	87
3.8 Considerações éticas.....	89
3.9 Conflitos de interesse	90
4 Resultados	91
5 Considerações finais	117
6 Anexos	131
ANEXO A: Planilha original do GTT-IHI.....	131
ANEXO B: Quesitos para avaliação do nível de confiança de que foi o cuidado prestado que causou a lesão ou dano ao paciente	132
7 Apêndices.....	133
APÊNDICE A: Formulário de coleta de dados gerais dos pacientes	133
APÊNDICE B: Protocolo para formação de padrão referência — Roteiro para busca ativa de eventos adversos no Bloco Cirúrgico	135
APÊNDICE C: Protocolo para formação de padrão referência — Roteiro para busca ativa de eventos adversos no Bloco Obstétrico e Maternidade.....	136
APÊNDICE D: Protocolo para formação de padrão referência — Roteiro para busca ativa de eventos adversos no Setor de Endoscopia	137
APÊNDICE E: Protocolo para formação de padrão referência — Roteiro para busca ativa de eventos adversos nas Enfermarias.....	138
APÊNDICE F: Protocolo para formação de padrão referência — Roteiro para busca ativa de eventos adversos nas Terapias Intensivas	139
APÊNDICE G: Protocolo para formação de padrão referência — relação de resultados de exames laboratoriais rastreados	142
APÊNDICE H: Protocolo para formação de padrão referência — Formulário para busca ativa de incidentes.....	142

APÊNDICE I: Protocolo para formação de padrão referência — Formulário para caracterização dos Eventos Adversos	144
APÊNDICE J: “Folha de Trabalho” final da pesquisa	147
APÊNDICE K: “Sumário da Revisão” final da pesquisa.....	148

1 CONSIDERAÇÕES INICIAIS

1.1 INTRODUÇÃO

Mais de duas décadas se passaram desde que o Institute of Medicine divulgou os emblemáticos relatórios sobre segurança e qualidade nos serviços de saúde – “*To Err Is Human: building a safer health system*” e “*Crossing the Quality Chasm: A New Health Care for the 21st Century*” (Institute of Medicine, 2000; Institute of Medicine, 2001). Na época, estimou-se que cerca de 100 mil pessoas morriam por ano em hospitais dos Estados Unidos devido a danos causadas pelos cuidados em saúde, com gastos anuais associados à ocorrência de eventos adversos evitáveis – incluindo perda de renda, invalidez e despesas médicas – entre 17 e 29 milhões de dólares (Institute of Medicine, 2000).

Esses estudos trouxeram à tona as falhas dos sistemas de saúde e impulsionaram uma ousada agenda mundial para melhorar a qualidade do cuidado (Institute of Medicine, 2001). O marco inicial do movimento pela segurança do paciente foi o lançamento da Aliança Mundial para a Segurança do Paciente (*World Alliance for Patient Safety*) pela Organização Mundial de Saúde (OMS) em outubro de 2004 (WHO, 2004). Os objetivos desse programa – que passou a se chamar *Patient Safety Program* – eram, entre outros, organizar os conceitos e as definições sobre segurança do paciente; propor medidas para reduzir os riscos e mitigar os eventos adversos; coordenar, disseminar e acelerar as melhorias na segurança do paciente; e promover o gerenciamento de riscos na assistência à saúde para prevenir danos ao paciente em todo o mundo (WHO, 2005; WHO, 2017).

Destacam-se ainda grandes campanhas regionais e nacionais, com apoio de governos e instituições sem fins lucrativos, que se desenvolveram ao longo dos últimos 20 anos. Entre elas estão as campanhas “100.000 Vidas” e “5 Milhões de Vidas”, que foram promovidas pelo *Institute of Healthcare Improvement (IHI)* nos Estados Unidos entre 2004 e 2008 e envolveram milhares de hospitais (IHI, 2006), e o *Prevent Harm, Work Safely*, que propôs a redução do número de eventos adversos evitáveis em todos os hospitais holandeses entre 2008 e 2012 através da implementação de um Sistema de Gestão da Segurança (Zegers et al., 2009). No Brasil, o Ministério da Saúde instituiu, em 2013, o Programa Nacional de

Segurança do Paciente e a obrigatoriedade de implantação do Núcleo de Segurança do Paciente (Brasil, 2013; Brasil, 2014).

Todas essas iniciativas propunham o monitoramento e a divulgação transparente de dados sobre segurança dos cuidados oferecidos pelas instituições de saúde com a finalidade de alcançar uma redução mensurável da mortalidade e da ocorrência de eventos adversos como uma medida confiável de melhoria na segurança do paciente (WHO, 2002; IHI, 2006; WHO 2010). No entanto, os métodos mais amplamente utilizados para estimativa da ocorrência de eventos adversos até então, como a notificação voluntária e a revisão retrospectiva clássica de prontuários, mostravam-se pouco sensíveis ou caros e demorados (Sari et al., 2007; Vincent, Burnett e Carthey, 2013). Isso limitava a capacidade de demonstrar a evolução de indicadores ligados à segurança dos pacientes.

Como alternativa, o *IHI* propôs, em 2003, o *Global Trigger Tool (GTT)*, uma ferramenta prática para medir eventos adversos em todas as áreas do cuidado de paciente adultos hospitalizados, que seria capaz de, objetivamente, detectar e monitorar os danos relacionados aos cuidados em saúde ao longo do tempo (McCannon et al., 2007; Classen et al., 2008; Griffin e Resar, 2009; Good et al., 2011). O *IHI* descreve o *GTT* como uma metodologia válida e confiável para detecção de danos em hospitais individuais, que inova por apresentar-se como um método estruturado e baseado em revisão de pequena amostra de prontuários, em um processo contínuo (Griffin e Resar, 2009). Essas características possibilitariam o acompanhamento da evolução das taxas de eventos adversos dentro de uma organização, assim como a avaliação do impacto das intervenções para melhoria na qualidade do cuidado (Classen et al., 2008; Griffin e Resar, 2009; Sharek et al., 2011). Ele é reconhecido como a única ferramenta de medição da segurança do paciente que pretende atender a esse propósito (Mattsson et al., 2013).

O *GTT* deriva das *Triggers Tools*, um grupo de ferramentas para identificação de eventos adversos por meio da revisão de prontuários que utiliza um método em duas etapas. A primeira é caracterizada por uma avaliação direcionada em busca de marcadores pré-definidos de alto risco para a ocorrência de eventos adversos, os chamados *triggers* (Jick H, 1974; Classen et al., 2008; Griffin e Resar, 2009).

Apenas os prontuários nos quais foram identificados esses *triggers* passam por revisões completas (segunda etapa), reduzindo despesas com recursos humanos e requisitos de tempo. Pelas vantagens metodológicas de menor custo para aplicação, maior agilidade nas revisões e potencial capacidade de detectar mudanças nos indicadores de segurança do paciente ao longo do tempo, o *GTT* foi amplamente adotado por hospitais, sistemas de saúde e entidades governamentais em vários países (Naessens et al., 2009; Landrigan et al., 2010; Classen et al., 2011; Good et al., 2011; Kennerly et al., 2011).

Apesar do uso difundido do *GTT*, existem poucos dados publicados sobre as propriedades de medição e utilidade dessa ferramenta e esses são restritos a avaliar a concordância entre diferentes avaliadores para encontrar *triggers* e detectar danos (Hilbert et al., 2016; Hanskamp-Sebregts et al., 2016). Desde a concepção da ferramenta *GTT* havia uma clara preocupação quanto a confiabilidade entre os revisores e o impacto da subjetividade da avaliação dos incidentes nos resultados (Griffin e Resar, 2009). Os estudos encontraram concordância moderada entre os revisores para a identificação de interações com eventos adversos, o que foi considerado aceitável, e o método, portanto, confiável (Classen et al., 2008; Naessens et al., 2010; Sharek et al., 2011).

Treinamento, constância entre membros da equipe e incentivo para que o grupo de revisores estabelecessem juntos diretrizes claras sobre definições sobre *triggers*, eventos e classificação de dano foram apontados como possíveis soluções para que resultados consistentes e comparáveis ao longo do tempo fossem obtidos, pelo menos no nível interno de uma organização (Classen et al., 2008). Entretanto, como não existe um método padrão ouro para detectar eventos adversos, a verdadeira acurácia do *GTT* permanece desconhecida (Klein et al., 2020). Alguns autores destacam que a limitação do tempo para a revisão inicial em busca dos *triggers* e a recomendação de que o prontuário não seja lido “da primeira à última página” podem favorecer a não identificação de alguns eventos, em especial aqueles com grau de dano mais leve (Griffin e Resar, 2009; Mattsson et al., 2013), mas não foram encontrados estudos nos quais foram abordadas as propriedades do método entre subgrupos de eventos adversos de diferentes classificações de dano.

Nesse contexto, acredita-se que a avaliação da acurácia do *GTT* como um método para identificação de eventos adversos em pacientes adultos hospitalizados possa subsidiar a tomada de decisão pelos gestores dos sistemas de saúde na escolha da ferramenta mais adequada, ou combinação delas, para realização da gestão de risco assistencial e desenvolvimento de estratégias que apoiem ações de melhorias na qualidade do cuidado. Adicionalmente, uma análise da acurácia de subgrupos, dividindo os eventos adversos identificados quanto ao grau de dano, pode ajudar a esclarecer se o *GTT* apresenta maior sensibilidade para eventos de maior gravidade para o paciente, ou seja, quando os eventos adversos de menor gravidade são excluídos.

1.2 FUNDAMENTAÇÃO TEÓRICA

1.2.1 O IMPACTO DOS CUIDADOS INSEGUROS EM SAÚDE

Os danos ao paciente fazem parte dos cuidados em saúde desde que eles começaram a ser praticados, mas sua extensão começou a ser medida e quantificada de forma sistemática apenas nas últimas décadas (Slawomirski, Auraaen e Klazinga, 2017). O rápido desenvolvimento científico na área da saúde a partir da segunda metade do século XX tem estreita relação com esse fato, uma vez que tornou o cuidado mais complexo, invasivo e dependente de uma fina interação entre fatores humanos e tecnológicos, expondo, de forma crescente, os pacientes ao risco de sofrerem eventos adversos (Chantler C, 1999; Leape e Berwick, 2005). Soma-se a isso a pressão por produção, longas jornadas de trabalho dos profissionais que prestam assistência e a predominância de um cuidado fragmentado e não coordenado (James, 2013).

Pode-se dizer que o movimento moderno de segurança do paciente começou nos últimos anos do século 20 e ganhou impulso ao longo das primeiras duas décadas do novo século. O relatório americano *“To Err Is Human: building a safer health system”* ressaltou a escala do problema, os paralelos com outras indústrias de alto risco e as fragilidades dos sistemas de saúde (Institute of Medicine, 2000). Na mesma época, uma série de estudos em diferentes países avaliaram a extensão dos danos causados pelos cuidados na assistência hospitalar (Brennan et al., 1991; Wilson et al., 1999; Vincent, Neale e Woloshynowych, 2001).

Vale ressaltar que os primeiros estudos em segurança do paciente se concentraram em eventos mais raros e trágicos, muitas vezes associados a atos de negligência (Brennan et al., 1991). Com o passar dos anos e o avanço nas discussões sobre critérios aceitáveis de segurança e qualidade do cuidado em saúde, tornou-se claro que a frequência de danos era muito maior do que se pensava anteriormente. Danos provocados pelo cuidado que eram normalizados até então – como a infecção relacionada à assistência em saúde, eventos adversos medicamentosos, complicações de procedimentos cirúrgicos e danos relacionados a quedas e lesões por pressão – passaram a ser entendidos como eventos adversos (Vincent, Burnett e Carthey, 2013).

É comumente relatado que cerca de um em cada 10 pacientes hospitalizados sofrem danos, sendo que, em média, metade deles poderiam ser evitados (WHO, 2017; Schwendimann et al., 2018). Em um estudo sobre frequência e evitabilidade de eventos adversos em oito países de baixa e média renda, a taxa de eventos adversos entre todas as internações hospitalares foi em torno de 8%, dos quais 83% poderiam ter sido evitados e 30% levaram à morte (Wilson et al., 2012).

Estima-se que eventos adversos sejam a 14^a causa de morbidade e mortalidade no mundo, o que é comparável a doenças como tuberculose e malária (Jha et al., 2013; Slawomirski, Aaraaen e Klazinga, 2017). As falhas nos cuidados em saúde são apontadas como a terceira causa de óbito nos Estados Unidos, atrás apenas das doenças cardiovasculares e das neoplasias (Makary e Daniel, 2016), e a responsável por 4% de todas as mortes nos países em desenvolvimento (OECD, 2018).

Estudos sobre custos em saúde associados a cuidados inseguros mostram que 15% dos gastos e das atividades hospitalares podem ser atribuídos a tratamentos de condições resultantes de incidentes de cuidado, sendo a maior carga associada a eventos adversos comuns, como infecções relacionadas à assistência à saúde, tromboembolismo venoso, lesões por pressão, erros de medicação e diagnóstico incorreto ou tardio (Slawomirski, Aaraaen e Klazinga, 2017). Os custos agregados, que somam os gastos diretos em saúde aos prejuízos referentes a litígio e perda

da capacidade e da produtividade no trabalho de pacientes e cuidadores, chegam a trilhões de dólares americanos a cada ano (WHO, 2017; OECD 2020).

Por outro lado, o custo para prevenir falhas nos sistemas de saúde e danos aos pacientes mostram-se insignificantes em comparação aos gastos resultantes de cuidados inseguros. Por exemplo, melhorias direcionadas para protocolos de segurança levaram a uma economia estimada de 28 bilhões de dólares americanos entre 2010 e 2015 nos hospitais do Medicare nos Estados Unidos (AHRQ, 2016).

No Brasil, estudos sobre o dano gerado ao paciente como consequência de um cuidado inseguro referem-se principalmente ao ambiente hospitalar, com taxas que variam, a depender da metodologia aplicada, de 7,3% a 40,9% das internações com pelo menos um evento adverso (Mendes et al., 2009; Moraes et al., 2021). Um estudo no qual se avaliou apenas internações clínicas de pacientes adultos em um hospital da rede pública em Minas Gerais demonstrou aumento médio no custo total de internação de mais de 98% nas admissões que apresentaram algum tipo de evento adverso comparado com aquelas nas quais não foi identificado nenhum dano ao paciente relacionado ao cuidado (Keulen, 2017).

1.2.2 APRENDER COM OS ERROS: A BASE DO MOVIMENTO PELA SEGURANÇA DO PACIENTE

Um dos maiores desafios que os sistemas de saúde enfrentam hoje é oferecer cuidados seguros e de qualidade em ambientes complexos, pressurizados e em rápida movimentação (Leape e Berwick, 2005; James, 2013). Desde o início do movimento atual pela segurança do paciente, foi reconhecido que uma etapa fundamental para reduzir os riscos associados aos cuidados de saúde era aprender com as situações que deram errado (Donaldson, 2002). Isso envolve não apenas reconhecer que as falhas nos sistemas de saúde existem, mas criar estratégias para detecção, prevenção e mitigação da ocorrência de eventos adversos.

Os sistemas de notificação são o método mais amplamente utilizado para identificação de incidentes nos cuidados em saúde (WHO, 2005; Doupi, 2009). Eles foram adaptados de outras indústrias de alto risco, nas quais ocupam, há muitos

anos, o papel central dos programas de segurança (WHO, 2020). Por meio deles, as organizações de saúde recebem e apuram relatos sobre situações inseguras ou danos relacionados aos cuidados oferecidos, suas possíveis causas e fatores contribuintes. As informações são então analisadas e geram conhecimento para formular ações para melhorar os processos e evitar que incidentes semelhantes aconteçam novamente.

A maioria dos sistemas de notificação de incidentes é aberta a todos os profissionais de saúde e os relatos podem ser voluntários ou obrigatórios, e, até mesmo, anônimos. Como a lógica dos sistemas de notificação é fornecer informações para aprendizagem, eles devem incluir, sempre que relevante, situações ou circunstâncias que potencialmente poderiam resultar em dano, mesmo que efetivamente ele não tenha ocorrido (OECD, 2018). Além de representarem uma via de comunicação sobre alertas na qualidade do cuidado e fornecerem dados sobre riscos na área de saúde, os sistemas de notificação criam oportunidades de resposta a pacientes e familiares sobre as ações que as instituições tem promovido para tornar o cuidado mais seguro (WHO, 2020).

A fim de melhorar a segurança para futuros pacientes por meio do compartilhamento de lições aprendidas pela análise de eventos já relatados, muitos países implementaram sistemas de notificação nos quais os eventos adversos, suas causas e qualquer informação contextual relevante são registrados em um repositório central. No Brasil, por exemplo, há um sistema eletrônico de notificações, o NOTIVISA, que é alimentado localmente pelos Núcleos de Segurança do Paciente e está sob responsabilidade da Agência Nacional de Vigilância Sanitária (ANVISA). O NOTIVISA tem por objetivo o registro e processamento de dados sobre incidentes de cuidado e queixas técnicas em todo o território nacional, fornecendo informações para a tomada de decisões em nível municipal, estadual, distrital e federal (Brasil, 2015).

O Brasil, assim como outros países, como a Austrália e o Reino Unido, mantém uma lista de incidentes de notificação obrigatória que inclui eventos adversos graves e potencialmente evitáveis (Brasil, 2015; ACSQHC, 2020; NHS, 2018). Essa prática amplia a perspectiva de aprendizado das notificações para um patamar de

governança e responsabilização, e, o mais importante, declara padrões mínimos aceitáveis de segurança em saúde (Vincent, Burnett e Carthey, 2013).

Apesar de vantagens como viabilidade de execução, mesmo em cenários com restrição de recursos, e menores custos, a notificação de incidentes não pode ser encarada como um mecanismo autônomo para reduzir o risco e melhorar a segurança. Ela precisa ser parte de uma estratégia que envolve um processo investigativo profundo para expor as fontes de risco para pacientes, a compreensão sobre como o dano ocorre e a implementação de mudanças com base nesses aprendizados para melhorar a segurança nos cuidados (Blais et al., 2008; WHO, 2020).

Além disso, para que os sistemas de notificação alcancem seus objetivos é fundamental que as instituições adotem uma cultura justa. Sem ela, predomina um ambiente de medo e apreensão entre os profissionais de saúde, o que reduz a disposição de relatar os incidentes e prejudica o processo de aprendizagem e de melhoria (WHO, 2020).

O erro humano é e sempre será uma realidade. Em uma estrutura de cultura justa, o foco está em abordar questões sistêmicas que contribuem para falhas e danos. Embora os profissionais sejam responsabilizados por desrespeitar ativamente protocolos e procedimentos, o relato de incidentes de cuidado é incentivado e os profissionais são apoiados quando os sistemas falham. Em uma cultura verdadeiramente justa, todos profissionais – assistenciais ou não – têm poder e não têm medo de expressar preocupações sobre ameaças à segurança do paciente (American College of Healthcare Executives [ACHE] and IHI/ National Patient Safety Foundation [NPSF] Lucian Leape Institute, 2017).

Promover essa cultura requer foco na segurança não apenas como uma iniciativa de melhoria fundamental, mas como um valor central que está totalmente incorporado a nossas organizações de saúde e que tem as lideranças como grandes impulsionadores. Os líderes devem possuir uma compreensão completa dos princípios e comportamentos de uma cultura justa e estar comprometidos em ensiná-los e modelá-los. Esses comportamentos incluem transparência, trabalho

em equipe, comunicação ativa, civilidade e feedback direto e oportuno e devem ser compreendidos e aplicados igualmente a todos os profissionais, independentemente do cargo ou função (ACHE e IHI/NPSF Lucian Leape Institute, 2017).

Outras questões que afetam a notificação de incidentes na área de saúde abrangem o escopo do que deve ser relatado e o confuso papel que ela exerce como estratégia de medição (Macrae, 2016). Critérios amplos para notificação, descritos frequentemente como "quaisquer incidentes não intencionais ou inesperados que levaram ou poderiam ter levado a danos", e a busca por números elevados de relatos, interpretados como reflexo de uma cultura de segurança forte, podem gerar volumes não manejáveis de informações (Hutchinson et al., 2009; Macrae, 2016). Em tais circunstâncias, pode não haver tempo e recursos suficientes para realizar a análise necessária, e, assim como acontece com outros bancos de dados na área de saúde, as atividades relacionadas a coleta, armazenamento e análise de dados são priorizadas frente ao seu uso para melhorias (Macrae, 2016; WHO, 2020).

Como estratégia de medição, os números ou taxas de incidentes relatados oferecem uma maneira particularmente pobre de medir o desempenho de segurança, já que esses sistemas detectam apenas uma pequena fração dos eventos adversos, em torno de 7 a 15%, e são influenciados por uma série de fatores cognitivos, sociais e organizacionais (Vincent C, 2007; Sari et al., 2007; Blais et al., 2008; Shojania KG, 2008; Mitchell et al., 2016). Apesar disso, gráficos de taxas de eventos identificados por sistemas de notificação continuam comumente sendo usados como medidas de segurança em serviços de saúde (Macrae, 2016).

1.2.3 A IMPORTÂNCIA DE MEDIR PARA MELHORAR E A ESCOLHA DE MÉTODOS ADEQUADOS DE MEDIÇÃO EM MELHORIA

Embora haja consenso sobre a necessidade de prevenir os danos relacionados aos cuidados em saúde e melhorar a segurança do paciente, ainda há pouco acordo sobre qual a medida mais adequada em segurança do paciente (Classen et al.,

2008; Borzecki e Rosen, 2020). Como exposto, os sistemas de notificação de incidentes, apesar de desempenharem papel fundamental na aprendizagem para a melhoria, não são adequados para quantificar as taxas de incidentes e danos ao longo do tempo (Sari et al., 2007; Vincent, Burnett e Carthey, 2013).

Existem diversas fontes e instrumentos reconhecidos para estudar a ocorrência de incidentes de cuidado e eventos adversos além da notificação, como os relatórios de autópsias, análise de reclamações por negligência médica, observação direta, vigilância clínica prospectiva, dados administrativos e de codificação e revisão de prontuários médicos (WHO, 2002; Thomas et al., 2003; Michel et al., 2004; OCDE, 2018). Todos eles apresentam vantagens e limitações e a escolha de um método ou outro vai depender tanto de fatores contextuais, como o objetivo da medição, a disponibilidade de recursos e os custos envolvidos, quanto das propriedades intrínsecas de cada método, que incluem o potencial de abrangência, o detalhamento das informações obtidas e a precisão e acurácia dos resultados (WHO, 2002; Andermann et al., 2013).

Por exemplo, quando o objetivo de medir a segurança do paciente é detectar o universo de incidentes de segurança que ocorrem, então medidas mais amplas, abrangendo grandes populações, como aquelas baseadas em dados administrativos, podem ser preferíveis. Por outro lado, o uso dos resultados para elaboração de relatórios públicos e pagamento por desempenho requer a utilização de métodos que favoreçam a precisão em vez da abrangência, como aqueles com denominadores restritos ou baseados na revisão de prontuários médicos (Borzecki e Rosen, 2020).

As discussões sobre qual seria o método mais apropriado para medir os requisitos de segurança do paciente são frequentemente prejudicadas por uma tentativa simplista de identificar a “melhor” metodologia ou o “melhor” instrumento. Por exemplo, frequentemente o valor da revisão retrospectiva de prontuários é questionado por ela não fornecer dados sobre fatores humanos e sobre outras questões importantes não usualmente registradas. Contudo, esses estudos não têm como objetivo primário fornecer tais informações e sim avaliar a natureza e a escala dos danos (WHO, 2002).

Um dos principais objetivos da medição em segurança do paciente é subsidiar ações de melhoria. Nesse caso, uma estratégia de medição adequada deve fornecer informações que amparem o estabelecimento de prioridades, a elaboração de ideias de mudança e a avaliação do real impacto dos esforços de melhoria (Classen et al., 2011). Atividades de medição importantes incluem o desenvolvimento de definições claras, com critérios de inclusão e de exclusão; protocolos de coleta de dados; estratégias de amostragem; e procedimentos para coleta, registro e análise dos dados. (WHO, 2015).

Na prática, as estratégias de medição relacionadas a ações de melhoria devem favorecer a coleta de dados periódicos de forma sustentada ao longo do tempo e ter uma articulação direta entre as medições que estão sendo coletadas e o que se pretende alcançar. Considerando que coleta, análise e elaboração de relatórios podem consumir tempo e recursos valiosos da equipe, apenas dados relevantes devem ser coletados e, sempre que possível, deve-se utilizar uma amostra para a extrapolação de dados (WHO, 2015). Para uma análise longitudinal, os dados obtidos devem ser plotados ao longo do tempo, utilizando-se, por exemplo, gráficos de tendências (Langley et al. 2011).

Instrumentos pouco acurados para quantificar os danos relacionados aos cuidados em saúde podem levar a interpretações equivocadas sobre problemas de segurança do paciente e, subsequentemente, à implementação de intervenções inadequadas de melhoria (Hanskamp-Sebregts et al., 2016). Além de melhorar a segurança do paciente, dados confiáveis e válidos devem ser a base de uma relação transparente entre prestadores de cuidados, fontes pagadoras e pacientes (Denis, 2014; Werner e Asch, 2005).

1.2.3.1 Análise de dados rotineiramente já coletados

Uma fonte reconhecida de informações sobre a segurança do cuidado nos serviços de saúde, especialmente hospitais, são os dados coletados rotineiramente. Eles contêm um rico conjunto de elementos, como informações demográficas; códigos de diagnóstico, procedimentos e tratamentos instituídos; documentação

assistencial; planejamento e monitoramento dos cuidados prestados; e resultados de exames complementares. Por meio de uma análise direcionada desses documentos, pode-se identificar, sistematicamente, uma série de lapsos de segurança e de eventos adversos ao longo do tempo e em diferentes ambientes de cuidado (OECD, 2018).

Os bancos de dados administrativos e de codificações relativos aos cuidados em saúde são fontes importantes para medição do dano. Contudo, a capacidade de definir indicadores de qualidade e segurança confiáveis e comparáveis depende de uma robusta infraestrutura de dados subjacente, respeito ao uso de codificações padronizadas (que inclua diagnósticos secundários, por exemplo) e sinalização do início da condição (para identificar se tal morbidade se iniciou durante uma hospitalização) (OECD, 2018).

De forma pioneira, em 2003 a *Agency for Healthcare Research and Quality* (AHRQ) liderou nos Estados Unidos a formulação de um conjunto básico de indicadores de segurança do paciente, conhecidos como PSI – sigla para o termo em inglês “*Patient Safety Indicators*”. Este utiliza dados administrativos, como códigos selecionados de diagnósticos secundários da Classificação Internacional de Doenças (CID) ou de procedimentos, para detectar potenciais eventos adversos evitáveis intra-hospitalares. Por exemplo, um paciente cirúrgico com um diagnóstico secundário de septicemia estreptocócica (CID-9 038.0) seria considerado como tendo sepse pós-operatória (PSI 13) (Naessens et al., 2009). Outros exemplos de PSI são: pneumotórax iatrogênico, insuficiência respiratória pós-operatória, embolia pulmonar pós-operatória ou trombose venosa profunda, deiscência da ferida pós-operatória e punção ou laceração acidental (AHRQ, 2003).

Por utilizarem dados administrativos da alta hospitalar, que apresentam reconhecida variabilidade na codificação e inconsistências em relação a diagnósticos e procedimentos, eles foram vistos inicialmente como “indicadores”, em vez de medidas definitivas de qualidade e segurança (Vincent, Burnett e Carthey, 2013). Um estudo conduzido por Rosen e colaboradores (2012) mostrou baixo valor preditivo positivo para alguns PSIs individuais, variando de 28% (IC 95% 15-43%) para fratura de quadril pós-operatória a 87% (IC 95% 79-92%) para

deiscência de ferida pós-operatória. As razões apontadas pelos pesquisadores para os falsos positivos incluíram condições que estavam presentes na admissão, erros e falta de especificidade de codificação.

Pela praticidade, grupos em todo o mundo adaptaram os PSIs do AHRQ para uso em seus próprios sistemas de saúde (Raleigh et al., 2008; von Plessen, Kodal e Anhøj, 2012; Vincent, Burnett e Carthey, 2013). Tentando aumentar a validade dos dados fornecidos pelos PSIs, o Medicare, sistema de saúde americano, exigiu que, a partir de outubro de 2007, os hospitais indicassem se os diagnósticos estavam “presentes na admissão” (POA, sigla para o termo em inglês “*present on admission*”). Um ano depois, esse mesmo sistema de saúde designou uma lista específica do que chamou de “condições adquiridas no hospital” (HACs, sigla para o termo em inglês “*hospital acquired conditions*”) – situações ou condições indesejáveis que afetam um paciente e que foram decorrentes dos cuidados em saúde inseguros – e parou de fornecer reembolso por serviços prestados para tratá-los (Anand, Kranker e Chen, 2019).

Apesar das críticas quanto à acurácia dos dados, outros programas foram desenvolvidos com base nos PSIs e HACs para incentivar a melhoria da qualidade e da segurança do paciente por meio de redução de reembolso aos serviços de saúde pelos custos relacionados a ocorrência de eventos adversos (OECD, 2018; Padula et al., 2020). Embora a monitorização desses indicadores possa apoiar iniciativas de melhoria de qualidade e ser útil no monitoramento de tendências (Vincent, Burnett e Carthey, 2013), muitos defendem que melhorias adicionais de codificação são necessárias antes que dados administrativos como os PSIs possam ser utilizados para estimar o real dano causado por sistemas de saúde ou para subsidiar programas de pagamento por desempenho (Rosen et al., 2012). Além disso, a utilização desses sistemas apenas se aplicaria em países que possuem um forte sistema de codificação.

1.2.3.2 Revisão retrospectiva de prontuários

A revisão retrospectiva do prontuário do paciente é uma outra estratégia para identificação de dano baseada na análise de dados usualmente coletados. Ela é

mais informativa que a avaliação restrita de dados administrativos e de codificação, porém mais laboriosa (OECD, 2018). Originalmente foi utilizada apenas para fins epidemiológicos, uma vez que fornece uma boa indicação da natureza, incidência e impacto econômico dos eventos adversos. Entretanto, técnicas de revisão mais recentes favorecem a obtenção de informações valiosas sobre causa e fatores associados (WHO, 2002). É o método para medir a prevalência de eventos adversos mais exaustivamente estudado e foi a base para relatórios relevantes sobre segurança do paciente (Institute of Medicine, 2000; Hanskamp-Sebregts et al., 2016).

Esse método demanda avaliação cuidadosa dos registros médicos por revisores treinados, o que é feito, predominantemente, de forma manual. Isso torna a revisão de prontuários uma ferramenta para identificação de eventos adversos mais demorada e cara que as demais (Vincent, Burnett e Carthey, 2013). Foi inicialmente utilizada apenas em contexto de pesquisa e de grandes inquéritos e, pelo tempo requerido para as análises, apresentava resultados que podiam não refletir mais o momento atual da organização (Unbeck et al., 2013). Com a experiência e o refinamento, o método foi sendo adaptado para produzir, de forma relativamente mais barata, análises da ocorrência do dano mais próximas do tempo real e algumas organizações passaram a realizar revisões periódicas e a utilizar os resultados como parte da estratégia de qualidade e melhoria (Vincent, Burnett e Carthey, 2013).

Revisões sistemáticas realizadas na última década (Hanskamp-Sebregts et al., 2016; Schwendimann et al., 2018; Klein et al., 2020) indicam que os métodos de revisão retrospectiva de prontuários mais frequentemente utilizados na atualidade são o Método Harvard (Brennan et al., 1991; Leape et al., 1991) e o *Global Trigger Tool* do *Institute of Healthcare Improvement* (Griffin e Resar, 2009). Apesar de semelhanças no processo de revisão – que acontece em duas etapas, sendo a primeira baseada na busca direcionada por dicas sobre a possível ocorrência de um evento adverso – eles possuem diferenças metodológicas que têm impacto nos custos de implementação e na aplicabilidade.

O Método de Harvard, descrito originalmente no estudo que ficou conhecido como *Harvard Medical Practice Study* (Hiatt et al., 1989; Brennan et al., 1991), define evento adverso como uma “lesão causada pelo tratamento médico (em vez da doença de base) e que prolongou a hospitalização, produziu uma deficiência no momento da alta ou ambos” e engloba tanto atos de comissão quanto de omissão de cuidados. Na primeira etapa do processo de revisão, enfermeiros treinados realizam a leitura detalhada dos prontuários e selecionam aqueles nos quais pelo menos um dos 18 rastreadores pré-estabelecidos foram encontrados. Se um prontuário for examinado como positivo, dois médicos o revisam, de forma independente, em busca de evidências de eventos adversos. Entre os rastreadores estão, por exemplo, retorno para sala cirúrgica durante o período de internação, parada cardíaca ou respiratória, transferência para uma unidade de cuidado especial ou para um outro hospital de cuidados agudos e indicação de litígio descrita em prontuário (Hiatt et al., 1989).

O método sofreu algumas modificações e foi utilizado em estudos subsequentes na Austrália (Wilson et al., 1999), Reino Unido (Vincent, Neale e Woloshynowych, 2001), Nova Zelândia (Davis et al., 2001), Canadá (Baker et al., 2004), Espanha (Aranaz-Andrés et al., 2008), Suécia (Soop et al., 2009), Brasil (Mendes et al., 2009), Portugal (Sousa et al., 2018), entre outros. A maioria deles previa a utilização de escalas para apoiar as decisões dos revisores médicos quanto a relação de causalidade, gravidade e evitabilidade dos incidentes e todos mantiveram o princípio básico de identificar eventos adversos por meio da busca de critérios numa primeira etapa da revisão dos prontuários. Por questões de custos e subsidiados por estudos anteriores que não demonstraram impacto na acurácia dos resultados, alguns protocolos de aplicação do método passaram a utilizar apenas um profissional médico na segunda etapa da revisão (De Vries et al., 2008; Zegers et al., 2010; Wilson et al., 2012).

Cabe ressaltar que, apesar do estudo original focar principalmente em questões médico-legais e de negligência, as adaptações mais recentes ampliaram as definições considerando-se as novas perspectivas sobre segurança do paciente (Unbeck et al., 2013). Essas diferenças conceituais sobre erros e danos em saúde se refletem nas taxas encontradas pelos pesquisadores: no primeiro estudo foram

identificados eventos adversos em 3,7% das internações avaliadas (Brennan et al., 1991), enquanto que nos estudos mais recentes as taxas foram, em geral, acima de 10% (Wilson et al., 1999; Vincent, Neale e Woloshynowych, 2001; Davis et al., 2001; Soop et al., 2009; Sousa et al., 2018).

Em uma iniciativa coordenada pela Organização Mundial da Saúde, esse método foi adaptado e utilizado em um estudo que incluiu 26 hospitais de oito países: Egito, Jordânia, Quênia, Marrocos, Tunísia, Sudão, África do Sul e Iêmen (Wilson et al., 2012). Os objetivos do estudo eram avaliar a frequência e a natureza dos eventos adversos em países de baixa e média renda e examinar a validade do uso da revisão retrospectiva de prontuários em sistemas de saúde com poucos recursos, nos quais a documentação dos registros clínicos pode ser menos abrangente. No geral, os pesquisadores identificaram eventos adversos em 8,2% dos 15.548 prontuários avaliados, com variação entre os países de 2,5% a 18,4%. Entre os eventos identificados, 83% foram considerados evitáveis, 30% foram associados a morte do paciente e cerca de 34% foram atribuídos a erros terapêuticos em situações clínicas relativamente não complexas. A taxa global de ocorrência de eventos adversos encontrada nesse estudo se aproxima do resultado obtido pelo estudo brasileiro conduzido por Mendes e colaboradores (2009), um dos primeiros fora do bloco dos países desenvolvidos, que identificou eventos adversos em 7,6% das internações avaliadas, sendo 69% dos eventos adversos considerados evitáveis.

Quanto à viabilidade e à validade da utilização da revisão retrospectiva de prontuários como instrumento para identificação de eventos adversos e medição em segurança do paciente em países em desenvolvimento e em transição, os autores fizeram algumas observações. Dos prontuários selecionados, 86% estavam disponíveis para revisão, mas a falta de notas de enfermagem, de relatórios de patologia e de registros de procedimentos, entre outras incompletudes, pode ter tido algum efeito em ambas as fases da revisão. Eles destacaram a menor taxa de triagem positiva na revisão primária, que foi de cerca de 22%, em comparação com os 26 a 44% obtidos em outros estudos (Brennan et al., 1991; Wilson et al., 1999; Vincent, Neale e Woloshynowych, 2001; Baker et al., 2004) e fizeram algumas suposições sobre as possíveis causas. Por exemplo, o

menor número de vezes no qual o critério readmissão foi encontrado – com média de 7,5% (variação de 1,3-19,9%) versus cerca de 23% em outros estudos (Wilson et al. 1999; Baker et al., 2004) – poderia significar que muitos hospitais iniciaram um novo prontuário a cada vez que um paciente internou ou que as readmissões de fato não ocorreram e esse critério pode não ser tão útil em países em desenvolvimento.

Considerando-se que o desempenho dos avaliadores também pode levar à variação nas taxas de eventos adversos identificados, um protocolo rigoroso de treinamento e supervisão dos revisores foi instituído. Dados sobre a confiabilidade da revisão primária foram disponibilizados em seis dos oito países e o índice de Kappa foi superior a 0,75 em cinco deles. Sobre a validade, os pesquisadores reconheceram que as taxas de eventos adversos poderiam estar subestimadas e associaram esse fato à perda de eventos menos graves ou à uma possível exclusão equivocada de eventos menos evitáveis por confusão por parte dos revisores quanto a definição operacional adotada.

Apesar das dificuldades relatadas pelos autores, a revisão retrospectiva de prontuários se mostrou satisfatória para obter informações importantes sobre a ocorrência de eventos adversos e a qualidade dos cuidados oferecidos em países em transição e em desenvolvimento frente a outros métodos disponíveis (Wilson et al., 2012). Ao mesmo tempo, o estudo reforça a existência de algumas limitações inerentes ao método, já amplamente reconhecidas, que podem ter impacto na validade dos dados obtidos.

Entre as limitações críticas está que a capacidade de determinar o que ocorreu depende exclusivamente da documentação e dos registros em prontuário médico. Como resultado, nem todos os incidentes podem ser capturados e podem faltar detalhes para avaliação da evitabilidade e dos fatores causais, informações fundamentais para os esforços de melhoria (Wong et al., 2015). Além disso, há questionamentos sobre a confiabilidade entre avaliadores, a qual, como em outros métodos, sofre forte impacto da subjetividade e da ausência de definições e classificações padronizadas em segurança do paciente (Thomas e Petersen, 2003). No entanto, os principais fatores que limitam o uso amplo e rotineiro dessa

metodologia pelas instituições de saúde são os custos e recursos de tempo envolvidos, o que fez com que esse tipo de revisão ficasse restrito, quase que exclusivamente, a contextos de pesquisa (Resar, Rozich e Classen, 2003; Classen et al., 2011).

1.2.3.3 Trigger Tools

Em 1999, o *IHI*, em uma iniciativa conhecida como Projeto “*Idealized Design of the Medication System*”, convocou um painel de especialistas para discutirem estratégias para tornar o processo de medicação mais seguro. O grupo concluiu que os métodos vigentes na época para detecção e medição de eventos adversos medicamentosos não eram nem eficazes, nem práticos (Classen et al., 2008). Fundamentados nos trabalhos de Classen, Kilbridge e colaboradores (Classen et al., 1991; Classen et al., 1992; Kilbridge e Classen, 2002), o painel adaptou e desenvolveu uma metodologia baseada na busca de gatilhos (*triggers*, em inglês), por meio da revisão de prontuários médicos, para detectar eventos adversos medicamentosos em pacientes hospitalizados (Classen et al., 2008).

A revisão de mais de 2.800 prontuários em 86 hospitais que utilizaram, preliminarmente, essa nova abordagem para detecção de eventos adversos medicamentosos mostrou uma taxa de 2,67 eventos adversos medicamentosos por 1.000 doses de medicamentos aplicadas. Vinte e quatro por cento das admissões experimentaram pelo menos um evento adverso relacionado à medicação, uma incidência muito maior do que havia sido encontrada em estudos anteriores (Rozich, Haraden e Resar, 2003).

“*Triggers*” são definidos, nesse contexto, como condições sentinela que apresentam forte relação com a ocorrência de um evento adverso (Naessens et al., 2010) e são equivalentes aos “rastreadores” do Método Harvard de revisão de prontuários. Embora o conceito de *trigger* tenha sido introduzido ainda na década de 70, quando um grupo de trabalho liderado por Jick desenvolveu um processo de coleta de informações sobre toxicidade de drogas a partir da identificação de informações-chave nos registros médicos (Jick, 1974), essa variação do método de revisão de prontuários para detecção de eventos adversos se consolidou nos anos

2000 e ficou conhecida, genericamente, como “*Trigger Tools*” (Classen et al., 2008).

Utilizando uma abordagem semelhante àquela utilizada para identificação de eventos adversos medicamentosos, o *IHI* desenvolveu uma variedade de *Trigger Tools* para a detecção de dano ao paciente em diferentes ambientes de cuidado, incluindo as unidades de terapia intensiva adulto, de cuidados perinatais, cuidados cirúrgicos e, em trabalhos colaborativos, unidade de terapia intensiva neonatal e cuidado de pacientes pediátricos (Resar, Simmonds e Haraden, 2006; *IHI*, 2005; Griffin e Classen, 2008; Sharek et al., 2006; Takata et al., 2008; Unbeck et al., 2014). Em 2003, a partir das experiências acumuladas, o *IHI* propôs o *Global Trigger Tool (GTT)*, uma ferramenta para detecção e monitoramento de eventos adversos em quase todos os espectros de atendimento a pacientes adultos hospitalizados, com exceção do cuidado em saúde mental e de reabilitação (Griffin e Resar, 2009).

A metodologia do *GTT* para detecção de eventos adversos é uma adaptação da metodologia clássica de revisão de prontuários do estudo de Harvard (Leape et al., 1991). Ela mantém a característica de revisão em duas etapas, sendo a primeira a busca por informações-chave, que evidenciam a possibilidade de ocorrência de um evento adverso, e a segunda a validação dos achados por profissional médico (Resar, Rozich e Classen, 2003). Entretanto, alterações foram realizadas para tornar o método mais prático, eficiente e confiável, viabilizando seu uso fora do ambiente de pesquisa. Destacam-se a limitação de 20 minutos por prontuário para revisão inicial, o uso de amostragem e a retirada da avaliação quanto a evitabilidade, negligência e tipo de erro (Classen et al., 2008).

1.2.4 GLOBAL TRIGGER TOOL COMO ESTRATÉGIA DE MEDIÇÃO EM MELHORIA

A metodologia do *GTT* se baseia na revisão retrospectiva de prontuários por um grupo treinado de revisores em duas etapas. Na primeira etapa, dois revisores primários, geralmente formada por profissionais enfermeiros, com experiência clínica e conhecimento sobre o formato dos prontuários utilizados pela instituição,

avaliam individualmente cada prontuário buscando por *triggers*. A cada *trigger* identificado, eles avaliam a relação com a possível ocorrência de um evento adverso e fazem a classificação quanto a gravidade do dano. Todo esse processo não deve durar mais que 20 minutos por prontuário avaliado. Ao final da avaliação individual, a dupla de revisores primários estabelece um consenso dos achados. Na segunda etapa da revisão, esse consenso é apresentado a um revisor médico, que valida os achados (Griffin e Resar, 2009).

A ferramenta em si consta de uma planilha com 54 *triggers* divididos em 6 módulos: cuidado, medicação, cirúrgico, cuidados intensivos, cuidado perinatal e departamento de emergência (Anexo A). Os *triggers* representam dicas de que um evento adverso pode ter acontecido. Alguns deles são eventos adversos por definição, como infecções relacionadas à assistência em saúde ou lacerações de terceiro ou quarto grau ocorridas durante a assistência ao parto. Todos os prontuários devem ser examinados quanto a presença dos *triggers* dos módulos cuidado e medicação e os demais módulos devem ser utilizados apenas quando aplicáveis (Griffin e Resar, 2009).

A definição utilizada pelo *GTT* para evento adverso é: “lesão física não intencional resultante de ou que teve os cuidados de saúde como fator contribuinte, que requer monitoramento, tratamento ou hospitalização adicionais, ou que resulta em morte”. Apesar de reconhecer que eventos atribuídos à omissão de tratamentos baseados em evidências devam ser foco dos esforços de melhoria, o *GTT* inclui apenas os eventos adversos relacionados à prestação ativa de cuidados. Os eventos adversos que já estavam presentes na admissão hospitalar devem ser contados e indicados ao relatar os resultados (Griffin e Resar, 2009).

O principal diferencial do *GTT* apontado pelo *IHI* e por alguns estudiosos é que, além de ser um método prático e custo-efetivo para detecção de eventos adversos, ele possui os requisitos críticos para ser utilizado como ferramenta de medição para subsidiar internamente atividades de melhoria em segurança do paciente (Classen et al., 2008; Griffin e Resar, 2009; Sharek et al., 2011). Essa propriedade está relacionada, principalmente, a sua estratégia de amostragem e de apresentação de resultados.

O *IHI* propõe que o *GTT* seja aplicado em uma amostra aleatória pequena, mas contínua ao longo do tempo – 10 prontuários a cada 2 semanas – e que os resultados obtidos sejam apresentados como taxas – porcentagem de admissões com pelo menos um evento adverso, eventos adversos por 100 admissões e eventos adversos por 1.000 pacientes-dia – e representados em gráficos de tendência (Griffin e Resar, 2009). Esse tipo de representação gráfica, utilizada com frequência por profissionais da área de melhoria da qualidade, permite compreender, de forma objetiva e com mínima complexidade matemática, se as mudanças efetuadas em um processo ou sistema provocam melhorias ao longo do tempo (Perla, Provost e Murray, 2011).

Por outro lado, assim como outros métodos que se baseiam na revisão retrospectiva de prontuários, o *GTT* também sofre impacto da qualidade dos registros médicos. Além disso, apesar da limitação do tempo para a revisão inicial em busca dos *triggers* e da recomendação de que o prontuário não seja lido “da primeira à última página” serem fatores que favoreçam a aplicabilidade da ferramenta, há uma reconhecida perda de sensibilidade para eventos com danos mais leves (Griffin e Resar, 2009; Mattsson et al., 2013).

1.2.4.1 Adaptações do GTT

O *GTT* foi utilizado inicialmente em alguns hospitais e sistemas de saúde dos Estados Unidos (Adler et al., 2008; Good et al. 2011; Kennerly et al., 2011). Posteriormente foi traduzido e adaptado para uso em diversos países, como Suécia (Soop et al., 2009), Dinamarca (von Plessen, Kodal e Anhøj, 2012), Noruega (Deilkås et al., 2015), Palestina (Najjar et al., 2013); Turquia (Kurutkan et al., 2015); Itália (Mortaro et al., 2017; Parrinello et al., 2019); Alemanha (Brösterhaus et al., 2020) e Portugal (Pierdevara et al., 2020).

Tornou-se uma metodologia amplamente utilizada para a revisão sistemática e geral de eventos adversos em registros de pacientes, com maior capacidade de detecção de dano em comparação a outros métodos, como a notificação de incidentes e o PSI (Naessens et al., 2009; Classen et al., 2011; Rutberg et al.,

2014; Mull et al., 2015). O *GTT* apresenta taxas de eventos superiores também às aquelas relatadas nos estudos que utilizaram o Método Harvard de revisão de prontuários, com valores que variam entre 15% e pouco mais de 40% das internações com pelo menos um evento adverso (Naessens et al., 2009; Landrigan et al., 2010; Sharek et al., 2011; Kennerly et al., 2013; Mattsson et al., 2013; Garret et al., 2013; Najjar et al., 2013; Kurutkan et al., 2015; Mull et al., 2015; Mortaro et al., 2017; Moraes et al., 2021).

Tem sido considerado um método prático para estimar a ocorrência de eventos adversos em pacientes hospitalizados e avaliar a eficácia das intervenções para melhorar a segurança do paciente ao longo do tempo, com vantagens frente a outros métodos quanto a desempenho, custo, disponibilidade e sustentabilidade (Sharek et al., 2011). Essas características fizeram com que países como a Suécia e a Noruega adotassem o *GTT* como estratégia de medição oficial em seus programas nacionais de segurança do paciente (Soop et al., 2009; Nilsson et al., 2018; Deilkås et al., 2015).

O *GTT* foi proposto como um método simples, o qual poderia ser utilizado, com mínimas adaptações, por qualquer instituição de cuidado hospitalar de pacientes adultos. O *IHI* previa que entre as alterações necessárias estariam a modificação de alguns *triggers*, em especial daqueles referentes a medicamentos, para que fossem incluídos itens padronizados para uso local. O elemento norteador para a substituição era entender a intenção por trás de cada *trigger* (Griffin e Resar, 2009).

No entanto, hospitais e sistemas de saúde viram necessidade de adaptações mais complexas para tornar o método mais factível, informativo e reproduzível. Tentando reduzir custos e tempo de revisão, um grupo de pesquisadores ligado ao Baylor Health Care System, um sistema integrado de prestação de serviços de saúde no norte do Texas, Estados Unidos, propôs o uso de apenas um revisor primário, o que lhes permitiria a revisão de duas a três vezes mais registros utilizando-se os mesmos recursos (Good et al., 2011; Kennerly et al., 2013; Garrett et al., 2013). Para não perder em sensibilidade e especificidade, utilizaram enfermeiros revisores profissionalmente treinados, o que garantiu concordância moderada a substancial ($Kappa=0,62$, IC 95% 0,54;0,70) para identificação de internações com eventos

adversos e credibilidade dos resultados frente às lideranças dos hospitais e equipes clínicas (Kennerly et al., 2013; Sharek et al., 2011). Outros estudos aplicaram diferentes grupos de revisores primários, como estudantes da pós-graduação em saúde pública (Mortaro et al., 2017) ou alunos de períodos mais avançados do curso de graduação em medicina (Moraes et al., 2021).

Na tentativa de ampliar as informações fornecidas pela ferramenta no processo de aprendizagem para melhoria da segurança do paciente, o mesmo grupo ligado ao Baylor Health Care System propôs a inclusão da avaliação quanto a evitabilidade do dano, potenciais fatores contribuintes para a ocorrência de um evento adverso e uma breve descrição narrativa dos eventos no formato “Situação-Contexto-Avaliação” (Good et al., 2011; Kennerly et al., 2013). Além disso, foram incluídos os eventos adversos relacionados a omissão de cuidados e os critérios de inclusão foram modificados para internações de pacientes adultos com três dias ou mais de tempo de permanência (Good et al., 2011; Kennerly et al., 2013).

A ampliação do tempo de permanência mínimo de 24 horas – critério de inclusão da ferramenta original – para três dias de internação baseava-se na percepção de que as taxas de eventos adversos eram maiores nas hospitalizações mais prolongadas. Entre as admissões com mais de três dias de duração, a taxa de eventos adversos foi de 23,2 por 100 admissões, enquanto que entre as admissões com menos de três dias, foi de 5,5 por 100 admissões. A exceção foram os eventos adversos relacionados aos cuidados obstétricos e perinatais, para os quais o grupo indicou uma ferramenta específica, com foco no período inicial após a admissão (Kennerly et al., 2013).

Sobre a inclusão da avaliação da evitabilidade do dano, o grupo não tinha a intenção de excluir os eventos considerados “não evitáveis” ou utilizar essa informação para fins disciplinares. O foco era auxiliar as instituições na identificação de padrões de eventos adversos nos quais as intervenções para melhoria poderiam ser mais eficazes, embora essa avaliação adicionasse maior grau de subjetividade à ferramenta por depender do julgamento do revisor médico baseado nos dados registrados em prontuário (Kennerly et al., 2013).

O grupo ligado ao Baylor Health Care System propôs ainda um ajuste metodológico no processo de amostragem. Para adequar os seus objetivos com a capacidade da organização em aplicar a ferramenta, eles utilizaram uma estratégia de amostragem que variava com o tamanho do hospital. Eles selecionaram mensalmente 10 prontuários para hospitais com menos de 6.000 internações anuais, 35 prontuários para hospitais com mais de 21.000 internações anuais e uma amostra de 2% para hospitais com número de internações anuais entre esses dois valores (Good et al., 2011; Kennerly et al., 2013).

Entretanto, era exatamente o pequeno tamanho da amostra sugerida pelo *IHI*, de 10 prontuários quinzenais, que preocupava alguns estudiosos. Isso se justifica porque pequenas amostras podem aumentar a variabilidade aleatória das taxas de eventos adversos identificadas a cada período, o que reduziria a precisão não só na detecção dessas taxas, mas também na avaliação da evolução da ocorrência de eventos adversos ao longo do tempo numa instituição (Mattsson et al., 2013; Landrigan et al., 2010; Shojania e Thomas, 2013).

Um estudo realizado em um hospital da Noruega investigou o impacto do aumento do número de prontuários revisados quinzenalmente, de 10 para 70, entre as internações realizadas no período de um ano (Mevik et al., 2016). Na amostra maior, foram identificados 1,45 (IC 95% 1,07-1,97; $p=0,02$) vezes mais eventos adversos por 1.000 pacientes-dia do que na amostra menor – 39,3 (IC 95% 35,8-43,1) versus 27,2 (IC 95% 20,3-36,4). Quando as taxas de eventos adversos por 1.000 pacientes-dia foram representadas em gráficos de tendência, houve maior variabilidade na menor amostra. Apesar da diferença na taxa de eventos encontrada pelos pesquisadores entre as duas amostras, as distribuições de tipo e nível de dano não diferiram significativamente e as infecções relacionadas à assistência em saúde foram o tipo de evento adverso mais comum em ambas.

O processo de amostragem tem impacto também na avaliação de eventos adversos mais raros, em especial aqueles mais graves. Por exemplo, alguns estudos que analisaram amostras menores não identificaram nenhum evento adverso que provocou ou contribuiu para a morte do paciente (Mull et al., 2015; Mevik et al., 2016; Moraes et al., 2021). Nesse sentido, Haukland e colaboradores

(2019) propuseram o uso do *GTT* como um método para a revisão dos óbitos e obtenção de informações sobre a ocorrência de eventos adversos graves que contribuíram para a morte do paciente.

Algumas instituições desenvolveram sistemas informatizados para coleta de dados (Good et al., 2011). Embora o *GTT* possa ser aplicado com sucesso em um contexto baseado em papel, a ferramenta eletrônica de coleta de dados oferece vantagens substanciais, como a estruturação da revisão e coleta de dados para obtenção de informações mais completas e a transmissão direta dos resultados para bases de dados (Good et al., 2011).

Com o avanço dos prontuários eletrônicos, alguns pesquisadores investiram também na automatização da identificação dos *triggers*, o que poderia fornecer uma abordagem mais padronizada e econômica que a revisão manual. Mevik e colaboradores (2019) descreveram um método modificado do *GTT* que substitui a primeira etapa de revisão por uma busca automatizada pelos *triggers* em prontuários eletrônicos baseada em algoritmos e inteligência artificial. Ao comparar o método *GTT* modificado com o original em uma amostra de 1.233 prontuários de internações em um hospital de médio porte na Noruega, os pesquisadores obtiveram uma taxa de 35 eventos adversos por 1.000 pacientes-dia para ambos, com redução de tempo total de revisão de 411 para 23 horas.

Além de economizar recursos, o uso de sistemas de identificação automática de *triggers* pode aumentar a confiabilidade do método e tornar os resultados mais comparáveis ao reduzir a variabilidade entre *triggers* identificados, uma vez que utilizam processos padronizados de busca por algoritmos e não sofrem restrição de tempo (Naessens et al., 2010; Unbeck e Schildmeijer, 2013). Além disso, podem fornecer uma plataforma para identificar pacientes em risco de eventos adversos em tempo real e serem utilizados para melhorar o resultado clínico, otimizar o tratamento e reduzir o ônus do dano ao paciente (Sammer et al., 2017; Mevik et al., 2019).

1.2.5 A AVALIAÇÃO DO *GLOBAL TRIGGER TOOL* COMO UM TESTE DIAGNÓSTICO

O *GTT*, assim como outros métodos utilizados para identificação de eventos adversos, funciona como um teste diagnóstico. Para um teste diagnóstico ser considerado útil, ele deve ser avaliado quanto a confiabilidade e validade. A confiabilidade mede a precisão de um teste, ou seja, a possibilidade de se obter um mesmo resultado quando o teste é repetido, seja pelo mesmo ou por outro avaliador. Já a validade mede se o resultado do teste é verdadeiro, ou seja, sua acurácia. Além dessas duas propriedades, a escolha de um método diagnóstico também envolve questões de custo e de implementação e a avaliação do efeito que os seus resultados podem provocar em medidas de cuidado em saúde (Hulley *et al.*, 2015).

Embora o *GTT* permita a identificação de mais eventos adversos do que outros métodos e seja empregado por muitos hospitais e sistemas de saúde ao redor do mundo, alguns pesquisadores apontam fragilidades metodológicas e de implementação e questionam o real valor dessa ferramenta como estratégia de medição em segurança do paciente (Landrigan *et al.*, 2010; von Plessen, Kodal e Anhøj, 2012; Mattsson *et al.*, 2013; James, 2013; Shojania e Thomas, 2013). Há poucos dados publicados sobre as propriedades de medição e utilidade dessa ferramenta e estes são restritos a avaliar a concordância entre diferentes avaliadores na identificação de *triggers* e detecção de danos (Mattsson *et al.*, 2013).

1.2.5.1 Confiabilidade ou precisão

Uma medida de alta precisão é aquela reprodutível, isto é, cujos os resultados são semelhantes a cada aferição (Hulley *et al.*, 2015). Os principais questionamentos quanto aos critérios de precisão do *GTT* se referem à reprodutibilidade entre os revisores primários para identificação dos *triggers*, ponto fundamental para identificação de um evento adverso, e ao impacto da subjetividade do avaliador para estabelecer, em algumas circunstâncias, a relação de causalidade entre o dano e o cuidado *versus* a ocorrência de um desfecho desfavorável atribuído à evolução natural da doença. Estudos que envolveram diferentes equipes de revisão

mostraram uma variação substancial nas taxas de danos relatadas (Sharek et al., 2011; von Plessen, Kodal e Anhøj, 2012; Moraes et al., 2021), com concordância apenas moderada entre os avaliadores, levando a diferentes conclusões sobre o processo de segurança do paciente (Schildmeijer et al., 2012; Mattsson et al., 2013).

O efeito da variabilidade entre os revisores sobre as medidas de resultado padrão do *GTT* foi avaliado por Mattsson e colaboradores (2013). A ferramenta foi aplicada em uma amostra idêntica de 10 prontuários a cada quinzena por duas equipes de revisores independentes durante o ano de 2010 no departamento de oncologia de um hospital universitário na Dinamarca. Apenas 31% dos eventos adversos foram identificados por ambas as equipes e foram encontradas diferenças na categorização sobre o dano de eventos idênticos. A concordância moderada (Kappa=0,45, IC 95% 0,30-0,52) para identificação de pelo menos um evento adverso entre as equipes deu origem a diferentes conclusões sobre a evolução do processo de segurança do paciente ao longo do tempo. A análise de Bland-Altman (Bland e Altman, 1986) para a taxa de eventos adversos por 1.000 pacientes-dia sugeriu pouco erro sistemático, mas grande erro aleatório.

A importância desse último dado reside no fato de que um dos pontos fortes apontados para o uso do *GTT* como estratégia de medição em segurança do paciente é justamente a possibilidade de avaliá-la como um processo, por meio da representação dos resultados em gráficos de tendência e da análise da evolução deles ao longo do tempo. Se a análise da evolução dos resultados gerados pelo *GTT* ao longo do tempo for considerada o dado de maior valor fornecido pelo *GTT*, pode-se argumentar que a identificação de eventos idênticos pode não ser muito importante, desde que a conclusão entre as equipes permaneça a mesma (Mattsson et al., 2013).

Resultados como esses confrontam o grau de precisão do *GTT* e levantam dúvidas sobre a capacidade do método permitir a identificação de uma mudança real no nível de dano ao paciente em uma organização. Para manter os dados robustos e comparáveis ao longo do tempo, seria necessário alcançar um alto grau de reprodutibilidade entre os revisores, garantindo que as variações nas taxas de

eventos adversos fossem devidas, exclusivamente, a mudanças externas ao sistema de medição, como, por exemplo, resultado de ações de melhoria no cuidado em saúde ofertado.

O próprio manual original do *IHI* sobre o *GTT* ressalta que a ferramenta pode apresentar baixa confiabilidade intra e interexaminador, especialmente quando um grande número de revisores é empregado (Griffin e Resar, 2009). Treinamento, constância entre membros da equipe e incentivo para que o grupo de revisores estabeleçam juntos diretrizes claras sobre definições de *triggers*, eventos e classificação da gravidade do dano, foram apontados como possíveis soluções para aumentar a confiabilidade, pelo menos internamente em uma organização (Classen et al., 2008; Griffin e Resar, 2009; von Plessen, Kodal e Anhøj, 2012; Mattsson et al., 2013; Mortaro et al., 2017).

Um fato que intriga pesquisadores da área é que algumas instituições que adotaram o *GTT* como estratégia de medição não foram capazes de detectar reduções nas taxas de eventos adversos ao longo do tempo, apesar de terem implementado, concomitantemente, várias iniciativas de melhoria da qualidade (Landrigan et al., 2010; Rutberg et al., 2014; Mortaro et al., 2017). Esses resultados podem significar que as ações instituídas não provocaram melhorias concretas na segurança do paciente e, por isso, os danos ao paciente continuavam comuns, entretanto, não é possível descartar que o *GTT* falhou em capturar uma mudança real (Landrigan et al., 2010).

Em contraste, Garrett e colaboradores (2013) perceberam uma redução progressiva nas taxas de eventos adversos em um grande sistema de saúde americano ao longo de um período de três anos de acompanhamento. Nesse caso, destaca-se que a categorização dos eventos por natureza permitiu a priorização de iniciativas voltadas para melhoria, como manejo glicêmico e lesões por pressão, direcionando os esforços para os problemas de qualidade de maior impacto nas taxas de dano e de maior potencial de alcançar avanços positivos.

Estudos em segurança do paciente que abordam a implementação de melhorias são menos comuns. Frequentemente poucas informações sistemáticas sobre como

as lideranças da organização utilizam os dados sobre os eventos adversos para monitorar ou melhorar a segurança do paciente estão disponíveis (Deilkås et al., 2015). Mattsson e colaboradores (2013) sugeriram que a manutenção das taxas de eventos adversos obtidas pelo *GTT* num mesmo patamar ao longo do tempo no hospital em estudo poderia estar relacionada a uma lenta taxa de implementação de mudanças em nível local. Esses autores propuseram mais foco nos esforços para a tradução de intervenções de segurança baseadas em evidências para a prática clínica.

Frente a essas contradições, enquanto alguns pesquisadores desestimulam o uso do *GTT* como estratégia de medição devido às inconsistências quanto à sua precisão (Mattsson et al., 2013), outros afirmam que as fragilidades estão relacionadas apenas a questões de implementação do processo de revisão e não são inerentes ao método (Von Plessen, Kodal e Anhøj, 2012). Sendo assim, essas fragilidades poderiam ser minimizadas ou corrigidas com treinamento sistemático e constância entre revisores, além do uso de definições operacionais claras e da padronização dos procedimentos de revisão (Classen et al., 2011; Naessens et al., 2010; Schildmeijer et al., 2012; von Plessen, Kodal e Anhøj, 2012; Mortoro et al., 2017).

A baixa confiabilidade entre os revisores e, portanto, a baixa precisão do método, tem sido apontada como o principal fator que limita o uso do *GTT* como uma estratégia de medição em segurança do paciente para comparar diferentes instituições quanto a indicadores de qualidade (Schildmeijer et al., 2012; Mattsson et al., 2013). Embora o *GTT* não tenha sido projetado para comparar o desempenho entre hospitais ou serviços de saúde quanto à qualidade e segurança dos cuidados ofertados (Griffin e Resar, 2009), é comum encontrar estudos que o façam.

Deilkås e colaboradores (2017) apresentaram a primeira comparação exploratória entre taxas e tipos de eventos adversos obtidos pelo *GTT* entre diferentes países num estudo que incluiu hospitais da Suécia e da Noruega. Esses países têm condições e contextos estruturais semelhantes para a saúde e têm buscado cooperação e aprendizagem mútua. Utilizando dados de amostras aleatórias

representativas de internações ocorridas no ano de 2013 em ambos países, nenhuma diferença significativa na taxa geral de internações com pelo menos um evento adverso foi encontrada (13,0% [IC 95% 11,7-14,3%] na Noruega e 14,4% [IC 95% 12,6-16,3%] na Suécia), mas houve diferença nas taxas de eventos quanto a natureza, com predomínio de complicações cirúrgicas em hospitais noruegueses e maiores taxas de lesões por pressão, quedas e "outros" eventos adversos nos hospitais suecos.

As comparações entre países com base na análise de prontuários médicos exigem que as amostras analisadas sejam representativas para o tipo de atendimento hospitalar em cada país e que artifícios estatísticos sejam utilizados para equilibrar diferenças nas características populacionais, caso existam. Ainda assim, pode haver grande variação entre as equipes, em especial se houver diferenças no processo de revisão de uma país para outro, ou mesmo de uma instituição para outra dentro de um mesmo país (Deilkås et al., 2017).

1.2.5.2 Validade ou acurácia

Para ser introduzido na prática, um teste diagnóstico deve ser avaliado quanto à validade analítica, validade clínica e utilidade clínica (Bossuyt et al., 2012; Burke, 2014). A validade analítica é determinada durante a construção do teste e esclarece se o teste funciona em condições experimentais. Por outro lado, a validade clínica, também chamada de acurácia diagnóstica, responderá se o teste funciona, em condições reais, na população de pacientes de interesse. Já a utilidade clínica responderá se a aplicação do teste na prática pode levar a melhorias nos resultados de saúde.

Os estudos de acurácia diagnóstica avaliam a capacidade do teste de discriminar, com exatidão, os pacientes com ou sem a condição de interesse (Malett et al., 2012; Hulley et al., 2015). A acurácia de um teste diagnóstico geralmente é definida em termos de suas características operacionais, incluindo sensibilidade, especificidade, valores de predição positivos e negativos e razões de verossimilhança (Hulley et al., 2015). Ao entender a ocorrência de um evento adverso como um caso positivo e os métodos para identificação de eventos

adversos como testes diagnósticos, então esses testes também poderiam ser descritos com essas características.

Em uma revisão sistemática publicada em 2016, não foram encontrados estudos que tenham avaliado a acurácia do *GTT* (Hanskamp-Sebregts et al., 2016). A inexistência de um padrão ouro para identificação de eventos adversos é apontada como o grande fator limitador para a existência de pesquisas dessa natureza (Classen et al., 2008; Mattsson et al., 2013).

A maioria dos estudos nos quais o *GTT* foi avaliado como ferramenta diagnóstica o comparou com outros métodos imperfeitos para identificação de eventos adversos, algumas vezes elencando um deles como referência, sem considerar as fragilidades inerentes a cada método (Michel et al., 2004). Alguns consideraram a revisão retrospectiva de prontuários como o padrão referência (Michel et al., 2004; Classen et al., 2011). Outros mediram a “efetividade” de um método por meio da proporção de casos identificados em relação a uma lista de referência que incluía os eventos adversos identificados por qualquer um dos métodos comparados (Michel et al., 2004; Naessens et al., 2009; Mull et al., 2015).

Outra fragilidade metodológica de alguns estudos desenvolvidos com o objetivo de avaliar o desempenho de métodos para identificação de eventos adversos é que eles tentam estabelecer o melhor método para identificação de eventos adversos utilizando taxas como variável principal, em especial o percentual de admissões com pelo menos um evento adverso (Michel et al., 2004; Moraes et al., 2021). Dessa forma, a identificação individual de cada evento adverso, naquelas situações nas quais foram identificados mais de um evento adverso em uma mesma internação, não é valorizada.

De modo geral, os resultados dos estudos apontam que diferentes eventos adversos são detectados por métodos diferentes e que nenhum deles parece fornecer uma detecção completa dos eventos, o que seria obtido por uma combinação de diferentes métodos (Naessens et al., 2009; Classen et al.; 2011). Esses estudos sugerem também que todos os métodos, incluindo o *GTT*, carecem de especificidade (Naessens et al., 2009; Classen et al.; 2011).

1.2.6 AVALIAÇÃO DA ACURÁCIA DE TESTES DIAGNÓSTICOS EM SITUAÇÕES SEM PADRÃO OURO

As medidas de acurácia são obtidas ao se comparar os resultados do teste diagnóstico em avaliação, chamado de teste índice, com os resultados de um teste padrão ouro. Assume-se que o padrão ouro é a técnica que oferece a aferição do resultado verdadeiro de certa característica, ou seja, determina, com exatidão, a presença ou ausência de uma doença ou condição de interesse (Hulley et al., 2015). Contudo, pesquisadores que avaliam a acurácia de testes diagnósticos frequentemente encontram situações nas quais o padrão ouro não está disponível em todos os pacientes, o padrão ouro é imperfeito ou não existe um padrão ouro aceito. Essas situações são chamadas de “situações sem padrão ouro” (Rutjes et al., 2007).

Na ausência de um padrão ouro verdadeiro, o recurso mais frequentemente empregado é utilizar o melhor método disponível, não necessariamente sem erros, para determinar a presença ou ausência da condição de interesse, o chamado padrão referência (Rutjes et al., 2007). Dessa forma, um teste ideal, porém não factível, é substituído pelo melhor teste possível. Entretanto, a imperfeição do padrão referência implica em risco de viés, que é conhecido como viés do padrão referência (Whiting et al., 2004; Zhou, Obuchowski e McClish, 2002; Trikalinos e Balion, 2012).

No paradigma clássico da avaliação de testes diagnósticos com resultados dicotômicos, os resultados de um estudo de acurácia podem ser resumidos em uma tabela 2x2 e várias medidas podem ser calculadas a partir dessa tabela (Figura 1). Todas essas medidas têm em comum a necessidade de uma classificação de pacientes com e sem a condição de interesse. Em outras palavras, os resultados do teste índice são verificados comparando-os com o resultado do padrão ouro ou referência.

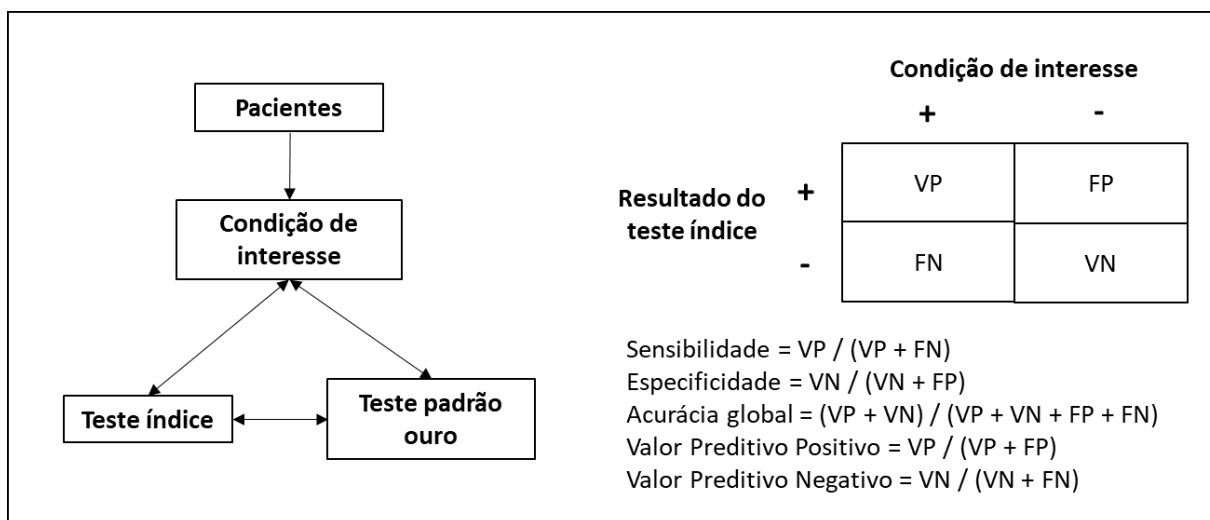


Figura 1 - Tabela 2 x 2 para avaliação da acurácia de um teste diagnóstico.

Figura adaptada de Rutjes et al., 2007. VP: verdadeiro positivo; FP: falso positivo; FN: falso negativo; VN: verdadeiro negativo.

Idealmente, o protocolo de avaliação de um teste diagnóstico deveria seguir os seguintes critérios: (a) o padrão ouro/referência fornece uma classificação livre de erros; (b) todos os resultados do teste índice são verificados pelo mesmo padrão ouro/referência; (c) os resultados do teste índice e padrão referência são interpretados sem o conhecimento do resultado do outro teste; e (d) o teste índice e o padrão ouro/referência são realizados ao mesmo tempo ou em um intervalo curto o suficiente para eliminar as alterações no status da condição alvo (Knottnerus e Muris, 2003; Rutjes et al., 2007).

Entretanto, pesquisadores podem encontrar situações nas quais o procedimento ideal de verificação não pode ser alcançado. Por exemplo, o padrão ouro/referência pode falhar em indicar a presença ou ausência da condição de interesse. Isso pode acontecer mesmo que o padrão ouro/referência não tenha erro analítico, como nas circunstâncias em que a condição alvo não produz as mudanças bioquímicas de interesse ou quando alguns tumores são perdidos durante a propedêutica por imagem em oncologia porque estão abaixo do nível de detecção. Falhas na execução metodológica do padrão ouro/referência e erros de interpretação por parte dos avaliadores são fontes adicionais de classificação incorreta, o que pode ser reduzido pelo estabelecimento de padronização rigorosa do método e adesão

estrita ao protocolo ou por um melhor treinamento dos avaliadores (Rutjes et al., 2007).

Além disso, para várias condições de interesse, não existe um padrão ouro/referência baseado em alterações histológicas ou bioquímicas e a condição é definida por uma combinação de critérios desenvolvidos por pesquisadores individuais ou durante uma reunião de consenso. Essas classificações podem variar com o tempo ou entre os países e não estão isentas de erros (Rutjes et al., 2007). Nesses casos, para reduzir variabilidade inter e intraobservador, o padrão referência deve ser executado de forma adequada e interpretado usando critérios padronizados. Isso é especialmente importante quando o diagnóstico padrão depende de interpretações subjetivas. Sessões de treinamento e feedback permanente sobre o desempenho são importantes para alcançar maior grau de confiabilidade (Knottnerus e Muris, 2003).

Dados os muitos fatores potenciais que podem levar a erros na classificação da condição de interesse, é improvável que exista um padrão ouro perfeito na prática. Essas observações apoiam a tendência, indicada por alguns autores, pela preferência pelo uso da terminologia “padrão referência”, um termo mais neutro, indicando o melhor método disponível, em substituição ao termo “padrão ouro”, que sugere um método sem erro (Rutjes et al., 2007). Além disso, é recomendável que os pesquisadores sempre discutam a qualidade do padrão referência e as consequências potenciais de uma classificação incorreta (Knottnerus e Muris, 2003; Rutjes et al., 2007).

Na prática, qualquer discordância entre o padrão referência e o teste índice será rotulado como um resultado “falso” para o teste índice. Nas circunstâncias em que há erros de classificação pelo padrão referência, haverá mudanças na tabela de classificação 2x2 e o efeito da classificação incorreta pelo padrão referência implicará em viés para cima ou para baixo nas estimativas da acurácia do diagnóstico (Hulley et al., 2015). A direção do viés irá depender se os erros do teste índice e do padrão referência imperfeito estão correlacionados. Se os erros forem positivamente correlacionados, por exemplo, ambos dizem que o teste é negativo na presença da condição de interesse, isso aumentará erroneamente a

concordância nas tabelas 2x2 e as estimativas de acurácia serão superestimadas. Por outro lado, se os erros forem negativamente correlacionados, como na circunstância em que o teste índice acertou e o padrão referência errou o status da condição de interesse, isso reduzirá erroneamente a concordância nas tabelas 2x2 e as estimativas de acurácia serão subestimadas (Pepe, 2004; Rutjes et al., 2007; Trikalinos e Balion, 2012).

Vários métodos têm sido propostos para lidar com situações em que o padrão referência é imperfeito, está parcialmente indisponível ou nas quais não existe um padrão referência aceito. Rutjes e colaboradores (2007), Trikalinos e Balion (2012) e Umemneku Chikere e colaboradores (2019) realizaram, de forma independente, estudos de revisão sistemática e descreveram algumas opções para avaliação do desempenho de um teste quando o padrão de referência é "imperfeito":

(a) Validar os resultados do teste índice

O paradigma de acurácia do teste diagnóstico é abandonado e o teste índice é avaliado quanto a capacidade de prever os resultados relevantes para o paciente. Isso quer dizer que, em vez de avaliar o desempenho diagnóstico ou de triagem do teste, o pesquisador quantifica o impacto, geralmente de longo prazo, das estratégias de gerenciamento do paciente (Rutjes et al., 2007; Trikalinos e Balion, 2012). Os estudos podem, por exemplo, relacionar os resultados do teste índice com eventos clínicos futuros ou avaliar se o teste pode prever quem se beneficiará mais de uma intervenção do que de outra. Como o paradigma de acurácia não é empregado, as medidas clássicas de acurácia são substituídas por outras medidas, como taxas de eventos, riscos relativos e outras medidas estatísticas de correlação (Rutjes et al., 2007).

(b) Imputar ou ajustar os dados ausentes no padrão referência

Existe um padrão referência aceitável, mas o seu resultado não é obtido em todos os pacientes. Os métodos desse grupo imputam ou ajustam as informações ausentes no subconjunto de pacientes sem desfecho no padrão referência (Rutjes et al., 2007; Umemneku Chikere et al., 2019).

- (c) Avaliar a concordância do teste índice e do padrão referência em vez de avaliar a acurácia do teste índice

Neste caso, o teste índice e o referência são tratados como dois métodos de medição alternativos e o estudo explora o quão bem um teste concorda com o outro, de modo que, por exemplo, um possa ser utilizado no lugar do outro. Avaliar a concordância pode ser a única opção significativa se nenhum dos testes comparados for uma escolha óbvia para um padrão referência. Para resultados de testes categóricos, pode-se resumir a extensão da concordância entre dois testes usando a estatística Kappa de Cohen (uma medida de concordância categórica que leva em conta a probabilidade de que algum acordo ocorrerá por acaso). No caso de resultados de testes contínuos, pode-se comparar diretamente as medições de um teste versus medições do outro teste por meio de artifícios estatísticos de regressão (Trikalinos e Balion, 2012).

- (d) Construir um padrão referência

Esses métodos combinam várias informações (resultados de testes) para construir um resultado padrão referência. Os grupos de pacientes recebem testes diferentes ou o mesmo conjunto de testes, após os quais, os resultados são combinados, por exemplo, por regra predefinida determinística, procedimento de consenso entre especialistas ou um modelo estatístico baseado em dados reais (Quadro 1) (Rutjes et al., 2007; Umemneku Chikere et al., 2019).

Quadro 1 – Alternativas metodológicas para construção de um padrão referência em estudos de avaliação de validade de testes diagnósticos “sem padrão ouro”

Método	Verificação diferencial	Análise discrepante	Padrão referência composto	Painel ou diagnóstico de consenso	Análise de classe latente
Descrição	Usa um segundo padrão referência, em geral, menos invasivo e oneroso, em parte da amostra.	Usa uma combinação de padrões referência de maneira sequencial. Um segundo teste referência é aplicado apenas aos resultados discrepantes da análise inicial.	Os resultados de vários testes imperfeitos são combinados para criar um padrão referência composto. Cada paciente recebe todos os testes que integram o padrão referência composto.	Grupo de especialistas produzem um diagnóstico de consenso com base em informações sistematizadas.	Usa um modelo estatístico para combinar diferentes informações (resultados de testes) para construir um padrão referência.
Modelo	Amostra de pacientes é dividida em subgrupos, que são verificados por padrões referência diferentes. A escolha do subgrupo que será testado com o padrão referência preferencial apresenta um padrão não aleatório e é baseada em critérios como um resultado positivo no teste índice, resultados positivos em outros testes ou a presença de fatores de risco para a condição de interesse.	Inicialmente, todos os pacientes são testados com o teste índice e um padrão referência imperfeito. Os resultados discordantes são retestados com um padrão referência adicional, chamado teste de resolver, com melhores propriedades discriminatórias do que o primeiro. Os resultados do teste de resolver são então usados para atualizar a tabela 2x2 final, com base na qual as estimativas de precisão para o teste de índice são calculadas.	Dois ou mais testes são aplicados em todos os pacientes e uma regra pré-especificada (determinística) é usada para classificar os pacientes como tendo a condição-alvo. Frequentemente os pesquisadores definem que a condição-alvo deve estar presente se um dos testes de referência for positivo.	Os especialistas podem discutir as informações sobre cada paciente em uma reunião e produzir um diagnóstico de consenso usando a votação por maioria em caso de desacordo; podem determinar o diagnóstico final de forma independente e discutir em reunião de consenso apenas casos discordantes; diagnósticos finais de especialistas individuais podem ser combinados usando um modelo estatístico.	Baseia-se em cálculos estatísticos de probabilidade, levando-se em consideração dados de prevalência, sensibilidade e especificidade dos testes utilizados para construir um padrão referência.
Pontos positivos	Permite obter uma verificação completa nas situações nas quais há dados não verificados ou quando não é possível aplicar o mesmo padrão referência em toda a amostra.	Método direto e fácil de aplicar, sem necessidade de conhecimentos estatísticos. Fornece uma alternativa eficiente para reduzir o número de pacientes que precisam ser testados com o melhor padrão de referência disponível quando esse padrão é invasivo ou caro.	O método é direto e fácil de entender. Permite combinar várias fontes de informação para avaliar se a condição alvo está presente. A regra pré-especificada para definir quando a condição-alvo está presente aumenta a transparência e evita problemas de incorporação e vies de work-up.	Abordagem flexível em determinar condições que são mal definidas. Opção de classificar as condições em um binário (condição-alvo presente ou ausente) e também de forma multinível (como em severo, moderado, leve, não doença, indeterminada, etc.).	São bem documentados, estatisticamente sólidos e têm sido aplicados em muitas áreas de pesquisa. A análise de classe latente é uma abordagem flexível que pode incorporar diferentes tipos de resultados de teste (dicotômicos, ordinais e contínuos).

Pontes negativos	Risco de viés de verificação diferencial, ou viés de padrão referência diferente, viés de investigação ou viés de seleção. Em geral, superestima medidas de acurácia. Para reduzir o risco de viés deve-se incorporar a verificação diferencial no projeto, especificando qual grupo de pacientes receberá um ou outro padrão referência.	O padrão de verificação é baseado no resultado do teste índice. Embora a análise discrepante forneça o status da condição de interesse para aqueles que são retestados, ela não fornece essa informação para aqueles que não foram retestados, que geralmente são a maioria.	Testes padrão referência diferentes podem definir a condição de interesse de forma diferente, introduzindo algum grau de viés. A inclusão de mais de dois testes referência no padrão referência composto pode, então, obscurecer a definição final da doença.	Pode haver baixa concordância inter e intraexaminadores. Os níveis de concordância para os especialistas devem ser medidos e relatados no estudo como parte da validação do método de consenso. A subjetividade do processo de classificação da doença. Pode ocorrer viés de incorporação se o resultado do teste do índice for divulgado aos especialistas.	Em uma análise de classe latente, a condição de interesse não é definida em um sentido clínico e os médicos podem se sentir desconfortáveis com o que os resultados representam. Na análise de classe latente os resultados do teste índice são frequentemente usados para construir o padrão de referência, o que viola o princípio da independência do teste índice.
Campo de aplicação	Vários testes diagnósticos aceitáveis estão disponíveis e podem servir como padrão referência.	Diverso. Porém, é recomendável evitar devido ao potencial risco de viés relacionado à incorporação implícita dos resultados do teste índice na definição do verdadeiro estado da condição de interesse.	Pode ser considerado em todas as situações onde um único padrão de ouro não existe, mas vários testes de referência imperfeitos estão disponíveis.	Alternativa atraente se um padrão referência aceito não existe e múltiplas fontes de informação precisam ser interpretadas de forma criteriosa para se chegar a um diagnóstico.	Aplicável para variáveis latentes, que são aquelas nas quais o resultado de interesse não pode ser medido diretamente.

- (e) Ajustar matematicamente as estimativas de sensibilidade e especificidade do teste índice para corrigir imperfeições do padrão referência

Há um padrão de referência preferido, mas esse padrão é conhecido por ser imperfeito. As soluções ajustam as estimativas de acurácia ou realizam análises de sensibilidade para examinar o impacto deste padrão referência imperfeito. O ajuste é baseado em dados externos, provenientes de pesquisas anteriores, sobre o grau de imperfeição. Os métodos de correção podem ser úteis se houver informações confiáveis sobre o grau de imperfeição do padrão referência e sobre a correlação dos erros entre o teste índice e o padrão referência (Rutjes et al., 2007; Trikalinos e Balion, 2012).

- (f) Calcular “estimativas ingênuas” da acurácia do teste índice e discutir em que direção eles são tendenciosos

Quando o padrão referência é imperfeito, ou seja, não espelha a verdade adequadamente, as estimativas típicas de acurácia calculadas a partir da tabela 2x2 podem ser imprecisas e, por esse motivo, foram denominadas de “estimativas ingênuas” de acurácia do teste por Trikalinos e Balion (2012). Outros autores utilizam os termos “co-positividade” e “co-negatividade”, ou ainda, “sensibilidade relativa” e “especificidade relativa” para se referirem a mesma situação (Buck et al., 1964; Gart e Buck, 1966; Line, Peters e Keenan, 1997; Ogusuk e Salem, 2004). Apesar de utilizarem as mesmas fórmulas para o cálculo, a preferência pelo uso desses termos dá destaque ao risco de viés nos resultados das medidas de acurácia relacionados ao uso de um padrão referência imperfeito.

Os pesquisadores devem decidir qual opção é mais adequada para avaliar o desempenho de um teste índice nas situações “sem padrão ouro” levando-se em consideração alguns fatores. O primeiro é que, possivelmente, vários testes padrão referência (imperfeitos) estejam disponíveis. Se uma escolha ideal não for óbvia, o pesquisador deve considerar o uso de mais de um padrão referência e descrever, de forma clara, quais são as implicações dessa escolha. Em segundo lugar, os pesquisadores devem selecionar qual(is) opção(ões) para avaliar o desempenho do teste é(são) apropriada(s). Elas não precisam ser mutuamente excludentes e, em alguns casos, podem ser complementares. Por fim, algumas opções dependem de

análise estatística especializada e não há nenhum dado empírico sobre os méritos e armadilhas dos métodos que se ajustam matematicamente para um padrão referência “imperfeito” (Trikalinos e Balion, 2012).

Reitsma e colaboradores (2009) sugeriram um fluxograma de orientação para escolha de métodos para avaliação de testes diagnósticos em situações “sem padrão ouro” (Figura 2). Esse fluxograma não contempla todas as opções descritas acima, mas fornece informações valiosas aos pesquisadores na tomada de decisão sobre qual alternativa metodológica pode ser mais adequada em cada situação.

No âmbito da segurança do paciente, não há dados claros disponíveis sobre medidas de acurácia dos testes para identificação de eventos adversos e nenhum deles é apontado como padrão referência preferido. Nesse contexto, uma solução metodológica para avaliação da acurácia do *GTT*, segundo o estudo de Reitsma e colaboradores (2009), seria a construção de um padrão referência. A partir daí, seria possível calcular as estimativas ingênuas da acurácia do teste ou avaliar a concordância do teste índice e do padrão referência.

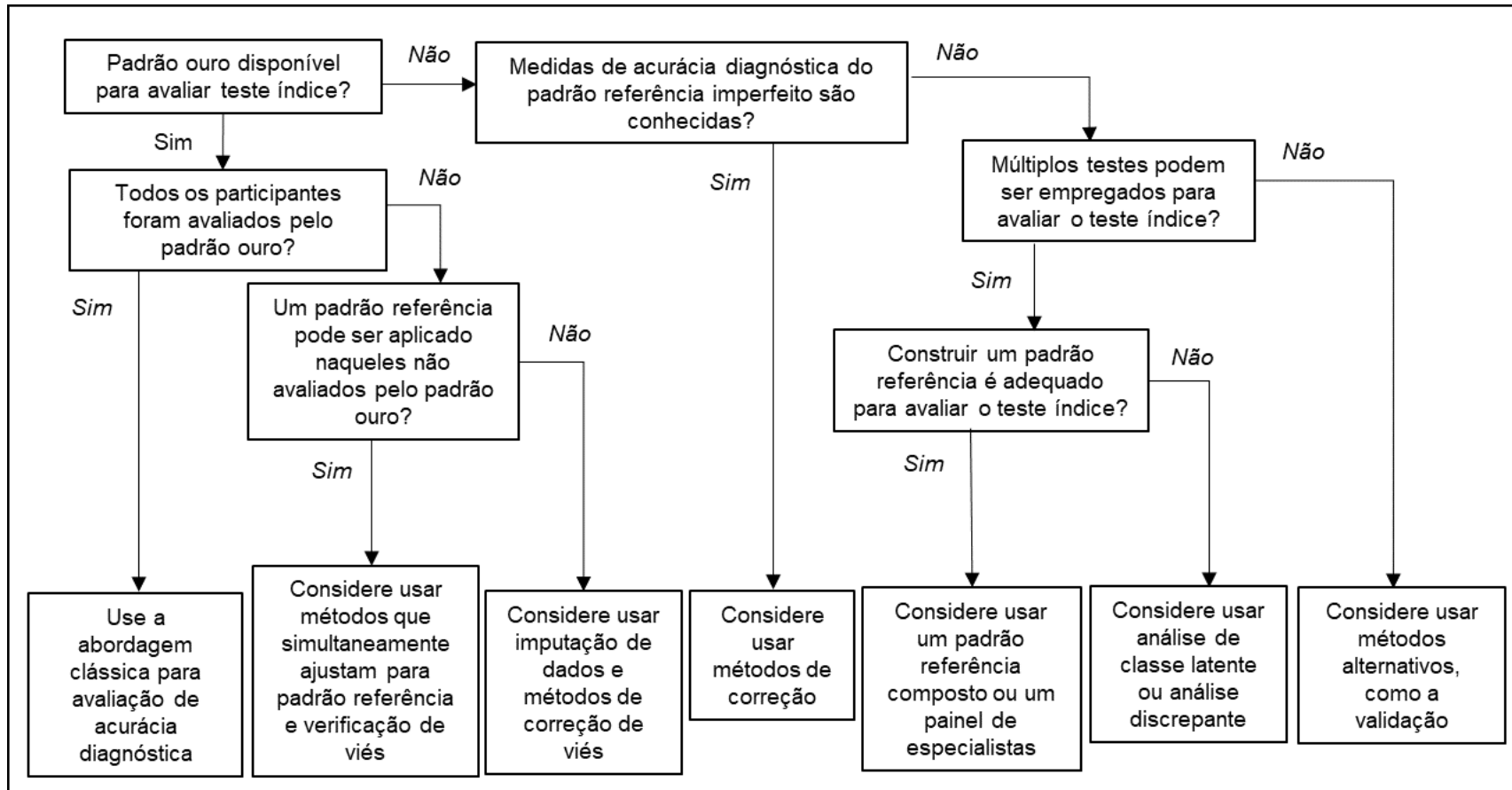


Figura 2 - Fluxograma de orientação para utilização de métodos para avaliação de testes diagnósticos em situações “sem padrão ouro”.

Figura adaptada de Reitsma e colaboradores (2009).

1.3 REFERÊNCIAS BIBLIOGRÁFICAS

Agency for Healthcare Research and Quality (AHRQ). AHRQ Quality Indicators: Guide to Patient Safety Indicators. Rockville, MD; March 2003.

Agency for Healthcare Research and Quality (AHRQ). National Scorecard on Rates of Hospital-Acquired Conditions 2010 to 2015: Interim Data From National Efforts to Make Health Care Safer. Agency for Healthcare Research and Quality. 2016. Available at: www.ahrq.gov/sites/default/files/wysiwyg/professionals/quality-patientsafety/pfp/2015-natl-scorecard-hac-rates.pdf. Accessed Mar 03, 2021.

Adler L, Denham CR, McKeever M, et al. Global trigger tool: Implementation basics. *J Patient Saf* 2008;4:245–9.

American College of Healthcare Executives (ACHE) and Institute for Healthcare Improvement (IHI)/ National Patient Safety Foundation (NPSF) Lucian Leape Institute. *Leading a Culture of Safety: A Blueprint for Success*. Boston, MA: American College of Healthcare Executives and Institute for Healthcare Improvement; 2017.

Anand P, Kranker K, Chen AY. Estimating the hospital costs of inpatient harms. *Health Serv Res*. 2019 Feb;54(1):86-96.

Andermann A, Wu AW, Lashoher A, et al. Patient Safety Research Training and Education Expert Working Group of WHO Patient Safety. Case studies of patient safety research classics to build research capacity in low- and middle-income countries. *Jt Comm J Qual Patient Saf*. 2013 Dec;39(12):553-60.

Aranaz-Andrés JM, Aibar-Remón C, Vitaller-Murillo J, et al. ENEAS work group. Incidence of adverse events related to health care in Spain: results of the Spanish National Study of Adverse Events. *J Epidemiol Community Health*. 2008 Dec;62(12):1022-9.

Australian Commission on Safety and Quality in Health Care (ACSQHC). Australian Sentinel Event List (version 2). Sydney, 2020.

Baker GR, Norton PG, Flintoft V, Blais R, et al. The Canadian Adverse Events Study: the incidence of adverse events among hospital patients in Canada. *CMAJ*. 2004 May 25;170(11):1678-86.

Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;1:307–10.

Blais R, Bruno D, Bartlett G, Tamblyn R. Can we use incident reports to detect hospital adverse events? *Journal of Patient Safety*. 2008; 4(1): 9–12.

Borzecki AM, Rosen AK. Is there a 'best measure' of patient safety? *BMJ Qual Saf.* 2020 Mar;29(3):185-188.

Bossuyt PM, Reitsma JB, Linnet K, Moons KG. Beyond diagnostic accuracy: the clinical utility of diagnostic tests. *Clinical chemistry.* 2012; 58(12):1636–43.

Brasil. Agência Nacional de Vigilância Sanitária. Resolução RDC nº 36, de 25 de julho de 2013. Institui ações para a segurança do paciente em serviços de saúde e dá outras providências. *Diário Oficial da República Federativa do Brasil, Brasília (DF),* 2013 jul 26; Seção 1:32.

Brasil. Ministério da Saúde. Documento de referência para o Programa Nacional de Segurança do Paciente / Ministério da Saúde; Fundação Oswaldo Cruz; Agência Nacional de Vigilância Sanitária. – Brasília: Ministério da Saúde, 2014. 40 p.

Brasil. Agência Nacional de Vigilância Sanitária. Nota Técnica GVIMS/GGTES/ANVISA nº 01/2015. Orientações gerais para a notificação de eventos adversos relacionados à assistência à saúde [Internet]. Brasília; 2015 jan 12 [citado 2017 mar 12]. 27 p.

Brennan TA, Leape LL, Laird NM, Herbert L, Localio AR, Lawthers AG et al. Incidence of adverse events and negligence in hospitalised patients: results of the Harvard Practice Study. *N Engl J Med.* 1991;324(6):370–6.

Brösterhaus M, Hammer A, Kalina S, Grau S, Roeth AA, Ashmawy H, Groß T, Binnebösel M, Knoefel WT, Manser T. Applying the Global Trigger Tool in German Hospitals: A Pilot in Surgery and Neurosurgery. *J Patient Saf.* 2020 Dec;16(4):e340-e351.

Buck AA, Sadun EH, Anderson RI, Shaffa E. Comparative studies of some immunologic screening tests for schistosomiasis in Ethiopia. *Am J Epidemiol.* 1964 July;80(1):75-84.

Burke W. Genetic tests: clinical validity and clinical utility. *Current protocols in human genetics.* 2014; 81(1):9.15. 1–9. 8.

Classen DC, Lloyd RC, Provost L, Griffin FA, Resar R. Development and Evaluation of the Institute for Healthcare Improvement Global Trigger Tool. *J Patient Saf.* 2008; 4(3):169-177.

Classen DC, Resar R, Griffin F, Federico F, Frankel T, Kimmel N, Whittington JC, Frankel A, Seger A, James BC. 'Global trigger tool' shows that adverse events in hospitals may be ten times greater than previously measured. *Health Aff (Millwood).* 2011 Apr;30(4):581-9. Erratum in: *Health Aff (Millwood).* 2011 Jun;30(6):1217.

Classen DC, Pestotnik SL, Evans RS, et al. Description of a computerized adverse drug event monitor using a hospital information system. *Hosp Pharm*. 1992;27:774Y783.

Classen DC, Pestotnik SL, Evans RS, et al. Computerized surveillance of adverse drug events in hospital patients. *JAMA*. 1991;266:2847Y2851.

Chantler C. The role and education of doctors in the delivery of health care. *Lancet*. 1999 Apr 3;353(9159):1178-81.

Davis P, Lay-Yee R, Briant R, et al. Adverse events in New Zealand public hospitals: principal findings from a national survey. Occasional paper no 3. New Zealand: Ministry of Health, 2001.

Deilkås E, Bukholm G, Lindstrøm J, et al. Monitoring adverse events in Norwegian hospitals from 2010 to 2013. *BMJ Open* 2015;5: e008576.

Deilkås ET, Risberg MB, Haugen M, Lindstrøm JC, Nylén U, Rutberg H, Michael S. Exploring similarities and differences in hospital adverse event rates between Norway and Sweden using Global Trigger Tool. *BMJ Open*. 2017 Mar 20;7(3):e012492.

Denis J-L. Accountability in healthcare organizations and systems. *Healthc Policy* 2014;10:8–11.

De Vries EN, Ramrattan MA, Smorenburg SM, Gouma DJ, Boermeester MA. The incidence and nature of in-hospital adverse events: a systematic review. *Qual Saf Health Care*. 2008;17(3):216–23.

Donaldson L. An organisation with a memory. *Clin Med (Lond)*. 2002 Sep-Oct;2(5):452-7.

Doupi P. National reporting systems for patient safety incidents: a review of the situation in Europe. Helsinki: National Institute for Health and Welfare, 2009.

Garrett PR Jr, Sammer C, Nelson A, Paisley KA, Jones C, Shapiro E, Tonkel J, Housman M. Developing and implementing a standardized process for global trigger tool application across a large health system. *Jt Comm J Qual Patient Saf*. 2013 Jul;39(7):292-7.

Gart JJ, Buck AA. Comparison of a screening test and a reference test in epidemiologic studies. II. A probabilistic model for the comparison of diagnostic tests. *Am J Epidemiol*. 1966 May;83(3):593-602.

Good VS, Saldaña M, Gilder R, Nicewander D, Kennerly DA. Large-scale deployment of the Global Trigger Tool across a large hospital system:

refinements for the characterisation of adverse events to support patient safety learning opportunities. *BMJ Qual Saf.* 2011 Jan;20(1):25-30.

Griffin FA, Classen DC. Detection of adverse events in surgical patients using the Trigger Tool approach. *Qual Saf Health Care.* 2008 Aug;17(4):253-8.

Griffin FA, Resar RK. *IHI Global Trigger Tool for Measuring Adverse Events.* IHI Innovation Series. 2nd ed. Cambridge, MA: Institute for Healthcare Improvement; 2009.

Hanskamp-Sebregts M, Zegers M, Vincent C, et al. Measurement of patient safety: a systematic review of the reliability and validity of adverse event detection with record review. *BMJ Open* 2016;6:e011078.

Haukland EC, Mevik K, von Plessen C, Nieder C, Vonen B. Contribution of adverse events to death of hospitalised patients. *BMJ Open Qual.* 2019 Feb 13;8(1):e000377.

Hibbert PD, Molloy CJ, Hooper TD, Wiles LK, Runciman WB, Lachman P, Muething SE, Braithwaite J. The application of the Global Trigger Tool: a systematic review. *Int J Qual Health Care.* 2016 Dec 1;28(6):640-649.

Hiatt HH, Barnes BA, Brennan TA, Laird NM, Lawthers AG, Leape LL, Localio AR, Newhouse JP, Peterson LM, Thorpe KE, et al. A study of medical injury and medical malpractice. *N Engl J Med.* 1989 Aug 17;321(7):480-4.

Hulley SB et al. *Delineando a pesquisa Clínica: uma abordagem epidemiológica.* 4ª edição. Artmed, 2015. 386p.

Hutchinson A, Young TA, Cooper KL, et al. Trends in healthcare incident reporting and relationship to safety and quality data in acute hospitals: results from the National Reporting and Learning System. *Qual Saf Health Care* 2009;18:5–10.

Institute for Healthcare Improvement (IHI). *IHI Perinatal Trigger Tool.* 2005. Available at: www.ihl.org/resources/Pages/Tools/PerinatalTriggerTool.aspx. Accessed in Mar 05, 2021.

Institute for healthcare Improvement (IHI). *5 Million Lives Campaign.* Cambridge, 2006. Available at: www.ihl.org/about/Documents/5MillionLivesCampaignCaseStatement.pdf. Accessed Feb 22, 2021.

Institute of Medicine (US) Committee on Quality of Health Care in America. *Crossing the Quality Chasm: A New Health System for the 21st Century.* Washington (DC): National Academies Press (US); 2001.

Institute of Medicine (US) Committee on Quality of Health Care in America; Kohn LT, Corrigan JM, Donaldson MS, editors. *To Err is Human: Building a Safer Health System*. Washington (DC): National Academies Press (US); 2000.

James JT. A new, evidence-based estimate of patient harms associated with hospital care. *J Patient Saf*. 2013;9(3):122-128.

Jha AK, Larizgoitia I, Audera-Lopez C, Prasopa-Plaizier N, Waters H, Bates DW. The global burden of unsafe medical care: analytic modelling of observational studies. *BMJ Qual Saf*. 2013 Oct;22(10):809-15.

Jick H. Drugs--remarkably nontoxic. *N Engl J Med*. 1974 Oct 17;291(16):824-8.

Kennerly D, et al. Journey to no preventable risk: The Baylor Health Care System patient safety experience. *Am J Med Qual*. 2011;26(1): 43–52.

Kennerly DA, Saldaña M, Kudyakov R, da Graca B, Nicewander D, Compton J. Description and evaluation of adaptations to the global trigger tool to enhance value to adverse event reduction efforts. *J Patient Saf*. 2013 Jun;9(2):87-95.

Keulen, MSLV. *Eventos adversos: incidência e impacto econômico em um hospital de grande porte*. Dissertação (Mestrado em Saúde Coletiva) – Faculdade de Medicina, Universidade Federal de Juiz de Fora, Juiz de Fora, 2017.

Kilbridge P, Classen D. Surveillance for adverse drug events: history, methods and current issues. *VHA Res Series*. 2002;3:1Y48.

Klein DO, Rennenberg RJMW, Koopmans RP, Prins MH. A Systematic Review of Methods for Medical Record Analysis to Detect Adverse Events in Hospitalized Patients. *J Patient Saf*. 2020 Mar 11.

Knottnerus JA, Muris JW. Assessment of the accuracy of diagnostic tests: the cross-sectional study. *J Clin Epidemiol*. 2003 Nov;56(11):1118-28.

Kurutkan MN, Usta E, Orhan F, et al. Application of the IHI Global Trigger Tool in measuring the adverse event rate in a Turkish healthcare setting. *Int J Risk Saf Med*. 2015;27:11–21.

Landrigan CP, Parry GJ, Bones CB, Hackbarth AD, Goldmann DA, Sharek PJ. Temporal trends in rates of patient harm resulting from medical care. *N Engl J Med*. 2010 Nov 25;363(22):2124-34. Erratum in: *N Engl J Med*. 2010 Dec 23;363(26):2573.

Langley GJ, Moen RD, Nolan KM, Nolan TW, Norman CL, Loyde PP. Tradução Petenate A. *Modelo de Melhoria: uma abordagem prática para melhorar o*

desempenho organizacional. Mercado de Letras. Campinas, 2011. 584 páginas.

Leape LL, Brennan TA, Laird N, et al. The nature of adverse events in hospitalized patients. Results of the Harvard medical practice study II. *N Engl J Med*. 1991;324:377Y384.

Leape LL, Berwick DM. Five years after To Err Is Human: what have we learned? *JAMA*. 2005 May 18;293(19):2384-90.

Line BR, Peters TL, Keenan J. Diagnostic test comparisons in patients with deep venous thrombosis. *J Nucl Med*. 1997 Jan;38(1):89-92.

Macrae C. The problem with incident reporting. *BMJ Qual Saf*. 2016 Feb;25(2):71-5.

Makary MA, Daniel M. Medical error-the third leading cause of death in the US. *BMJ*. 2016 May 3;353:i2139.

Malett S, Haligan S, Matthew Thompson GP, Collins GS, Altman DG. Interpreting diagnostic accuracy studies for patient care. *BMJ (Online)*. 2012; 345(7871).

Mattsson TO, Knudsen JL, Lauritsen J, Brixen K, Herrstedt J. Assessment of the global trigger tool to measure, monitor and evaluate patient safety in cancer patients: reliability concerns are raised. *BMJ Qual Saf*. 2013 Jul;22(7):571-9.

McCannon CJ, Hackbarth AD, Griffin FA. Miles to go: an introduction to the 5 Million Lives Campaign. *Jt Comm J Qual Patient Saf*. 2007 Aug;33(8):477-84.

Mendes W, Martins M, Rozenfeld S et al. The assessment of adverse events in hospitals in Brazil. *Int J Qual Health Care* 2009 Aug;21(4):279-84.

Mevik K, Griffin FA, Hansen TE, Deilkås ET, Vonen B. Does increasing the size of bi-weekly samples of records influence results when using the Global Trigger Tool? An observational study of retrospective record reviews of two different sample sizes. *BMJ Open*. 2016 Apr 25;6(4):e010700.

Mevik K, Hansen TE, Deilkås EC, Ringdal AM, Vonen B. Is a modified Global Trigger Tool method using automatic trigger identification valid when measuring adverse events? *Int J Qual Health Care*. 2019 Aug 1;31(7):535-540.

Michel P, Quenon JL, de Sarasqueta AM, Scemama O. Comparison of more methods for estimating rates of adverse events and rates of preventable adverse events in acute care hospitals. *BMJ*. 2004 Jan 24;328(7433):199.

Mitchell I, Schuster A, Smith K, et al. Patient safety reporting: a qualitative study of thoughts and perceptions of experts 15 years after 'To Err is Human'. *BMJ Qual Saf* 2016;25:92–9.

Moraes SM, Ferrari TCA, Figueiredo NMP, Almeida TNC, Sampaio CCL, Andrade YCP, Benedito RO, Beleigoli AM. Assessment of the reliability of the IHI Global Trigger Tool: new perspectives from a Brazilian study. *Int J Qual Health Care*. 2021 Mar 18;33(1):mzab039.

Mortaro A, Moretti F, Pascu D, Tessari L, Tardivo S, Pancheri S, Marta G, Romano G, Mazzi M, Montresor P, Naessens JM. Adverse Events Detection Through Global Trigger Tool Methodology: Results From a 5-Year Study in an Italian Hospital and Opportunities to Improve Interrater Reliability. *J Patient Saf*. 2017 Jun 9.

Mull HJ, Brennan CW, Folkes T, et al. Identifying previously undetected harm: piloting the Institute for Healthcare Improvement's Global Trigger Tool in the Veterans Health Administration. *Qual Manag Health Care*. 2015;24:140–146.

Naessens JM, Campbell CR, Huddleston JM, et al. A comparison of hospital adverse events identified by three widely used detection methods. *Int J Qual Health Care* 2009;21:301–7.

Naessens JM, O'Byrne TJ, Johnson MG, Vansuch MB, McGlone CM, Huddleston JM. Measuring hospital adverse events: assessing inter-rater reliability and trigger performance of the Global Trigger Tool. *Int J Qual Health Care*. 2010 Aug;22(4):266-74.

Najjar S, Hamdan M, Euwema MC, et al. The Global Trigger Tool shows that one out of seven patients suffers harm in Palestinian hospitals: challenges for launching a strategic safety plan. *Int J Qual Health Care*. 2013;25:640–647.

National Health Service (NHS). Never Events policy and framework – revised January 2018. London, 2018.

Nilsson L, Borgstedt-Risberg M, Soop M, Nylén U, Ålenius C, Rutberg H. Incidence of adverse events in Sweden during 2013-2016: a cohort study describing the implementation of a national trigger tool. *BMJ Open*. 2018 Mar 30;8(3):e020833.

OECD. Measuring patient safety. Opening the Black Box. 2018

OECD. The economics of patient safety. From analysis to action. 2020.

Ogusuk MM, Salem JI. Analysis of different primers used in the PCR method: diagnosis of tuberculosis in the state of Amazonas. *Brazil J Bras Pnemol*. 2004;30:343–9.

Padula WV, Black JM, Davidson PM, Kang SY, Pronovost PJ. Adverse Effects of the Medicare PSI-90 Hospital Penalty System on Revenue-Neutral Hospital-Acquired Conditions. *J Patient Saf.* 2020 Jun;16(2):e97-e102. doi: 10.1097/PTS.0000000000000517. PMID: 30110019.

Parrinello V, Grasso E, Saglimbeni G, Patanè G, Scalia A, Murolo G, Lachman P. Assessing the development and implementation of the Global Trigger Tool method across a large health system in Sicily. *F1000Res.* 2019 Mar 7;8:263.

Pepe MS. Incomplete data and imperfect reference tests. In Pepe MS, editor. *The statistical evaluation of medical tests for classification and prediction.* Oxford: Oxford University Press; 2004. pp. 168–213.

Perla RJ, Provost LP, Murray SK. The run chart: a simple analytical tool for learning from variation in healthcare processes. *BMJ Qual Saf.* 2011 Jan;20(1):46-51.

Pierdevara L, Porcel-Gálvez AM, Ferreira da Silva AM, Barrientos Trigo S, Eiras M. Translation, Cross-Cultural Adaptation, and Measurement Properties of the Portuguese Version of the Global Trigger Tool for Adverse Events. *Ther Clin Risk Manag.* 2020 Dec 3;16:1175-1183.

Raleigh VS, Cooper J, Bremner SA, Scobie S. Patient safety indicators for England from hospital administrative data: case-control analysis and comparison with US data. *BMJ* 2008;337:a1702. doi:10.1136/bmj.a1702.

Reitsma JBR A. W. S.; Khan K. S.; Coomarasamy A.; Bossuyt P. M. A review of solutions for diagnostic accuracy studies with an imperfect or missing reference standard. *Journal of Clinical Epidemiology.* 2009; 62(8):797–806.

Resar RK, Rozich JD, Classen D. Methodology and rationale for the measurement of harm with trigger tools. *Qual Saf Health Care* 2003;12(Suppl 2):ii39–45.

Resar RK, Simmonds T, Haraden CR. A trigger tool to identify adverse events in the intensive care unit. *Jt Comm J Qual Patient Saf.* 2006;32(10):585Y590.

Rosen AK, Itani KM, Cevasco M, Kaafarani HM, Hanchate A, Shin M, Shwartz M, Loveland S, Chen Q, Borzecki A. Validating the patient safety indicators in the Veterans Health Administration: do they accurately identify true safety events? *Med Care.* 2012 Jan;50(1):74-85.

Rozich JD, Haraden CR, Resar RK. Adverse drug event trigger tool: a practical methodology for measuring medication related harm. *Qual Saf Health Care.* 2003;12(3):194Y200.

Rutberg H, Borgstedt Risberg M, Sjudahl R, et al. Characterisations of adverse events detected in a university hospital: a 4-year study using the Global Trigger Tool method. *BMJ open*. 2014;4:e004879.

Rutjes AWS, Reitsma JB, Coomarasamy A, Khan KS, Bossuyt PMM. Evaluation of diagnostic tests when there is no gold standard. A review of methods. *Health Technol Assess* 2007;11(50).

Sammer C, Miller S, Jones C, Nelson A, Garrett P, Classen D, Stockwell D. Developing and Evaluating an Automated All-Cause Harm Trigger System. *Jt Comm J Qual Patient Saf*. 2017 Apr;43(4):155-165.

Sari AB, Sheldon TA, Cracknell A, Turnbull A. Sensitivity of routine system for reporting patient safety incidents in an NHS hospital: retrospective patient case note review. *BMJ*. 2007 Jan 13;334(7584):79.

Schwendimann R, Blatter C, Dhaini S, Simon M, Ausserhofer D. The occurrence, types, consequences and preventability of in-hospital adverse events - a scoping review. *BMC Health Serv Res*. 2018 Jul 4;18(1):521.

Schildmeijer K, Nilsson L, Arestedt K, Perk J. Assessment of adverse events in medical care: lack of consistency between experienced teams using the global trigger tool. *BMJ Qual Saf*. 2012 Apr;21(4):307-14.

Schildmeijer K, Nilsson L, Perk J, Arestedt K, Nilsson G. Strengths and weaknesses of working with the Global Trigger Tool method for retrospective record review: focus group interviews with team members. *BMJ Open*. 2013 Sep 24;3(9):e003131.

Sharek PJ, Horbar JD, Mason W, et al. Adverse events in the neonatal intensive care unit: development, testing and findings of a NICU-focused trigger tool to identify harm in North American NICUs. *Pediatrics*. 2006;118(4):1332Y1340.

Sharek PJ, Parry G, Goldmann D, Bones K, Hackbarth A, Resar R, Griffin FA, Rhoda D, Murphy C, Landrigan CP. Performance characteristics of a methodology to quantify adverse events over time in hospitalized patients. *Health Serv Res*. 2011 Apr;46(2):654-78.

Shojania KG. The frustrating case of incident-reporting systems. *Qual Saf Health Care* 2008;17:400–2.

Shojania KG, Thomas EJ. Trends in adverse events over time: why are we not improving? *BMJ Qual Saf*. 2013 Apr;22(4):273-7.

Slawomirski L, Auraaen A, Klazinga N. The economics of patient safety. Organisation for Economic Co-operation and Development (OECD), 2017.

Soop M, Fryksmark U, Köster M, et al. The incidence of adverse events in Swedish hospitals: a retrospective medical record review study. *Int J Qual Health Care* 2009;21:285–91.

Sousa P, Uva AS, Serranheira F, Uva MS, Nunes C. Patient and hospital characteristics that influence incidence of adverse events in acute public hospitals in Portugal: a retrospective cohort study. *Int J Qual Health Care*. 2018 Mar 1;30(2):132-137.

Takata GS, Mason W, Taketomo C, Logsdon T, Sharek PJ. Development, testing, and findings of a pediatric-focused trigger tool to identify medication-related harm in US children's hospitals. *Pediatrics*. 2008 Apr;121(4):e927-35.

Thomas EJ, Petersen LA. Measuring errors and adverse events in health care. *J Gen Intern Med*. 2003 Jan;18(1):61-7.

Trikalinos TA, Balion CM. Chapter 9: options for summarizing medical test performance in the absence of a "gold standard". *J Gen Intern Med*. 2012 Jun;27 Suppl 1(Suppl 1):S67-75.

Umemneku Chikere CM, Wilson K, Graziadio S, Vale L, Alen AJ. Diagnostic test evaluation methodology: A systematic review of methods employed to evaluate diagnostic tests in the absence of gold standard - An update. *PLoS One*. 2019 Oct 11;14(10):e0223832.

Unbeck M, Schildmeijer K. Is detection of adverse events affected by record review methodology? An evaluation of the 'Harvard Medical Practice Study' method and the 'Global Trigger'. *Patient Saf* 2013;7:10.

Unbeck M, Lindemalm S, Nydert P, Ygge BM, Nylén U, Berglund C, Härenstam KP. Validation of triggers and development of a pediatric trigger tool to identify adverse events. *BMC Health Serv Res*. 2014 Dec 21;14:655.

Vincent C. Incident reporting and patient safety. *BMJ*. 2007;334(7584):51.

Vincent C, Burnett S, Carthey J. The measurement and monitoring of safety. The Health Foundation. London, 2013.

Vincent C, Neale G, Woloshynowych M. Adverse events in British hospitals: preliminary retrospective record review. *BMJ*. 2001;322:517.

von Plessen C, Kodal AM, Anhøj J. Experiences with global trigger tool reviews in five Danish hospitals: an implementation study. *BMJ Open*. 2012 Oct 12;2(5):e001324.

Werner RM, Asch DA. The unintended consequences of publicly reporting quality information. *JAMA* 2005;293:1239–44.

Whiting P, Rutjes AW, Reitsma JB, Glas AS, Bossuyt PM, Kleijnen J. Sources of variation and bias in studies of diagnostic accuracy: a systematic review. *Ann Intern Med* 2004;140:189–202.

Wilson RM, Harrison BT, Gibberd RW, Hamilton JD. An analysis of the causes of adverse events from the Quality in Australia Health Care Study. *Medical Journal of Australia*. 1999;170:411–515.

Wilson RM, Michel P, Olsen S, Gibberd RW, Vincent C, El-Assady R, Rasslan O, Qsous S, Macharia WM, Sahel A, Whittaker S, Abdo-Ali M, Letaief M, Ahmed NA, Abdellatif A, Larizgoitia I; WHO Patient Safety EMRO/AFRO Working Group. Patient safety in developing countries: retrospective estimation of scale and nature of harm to patients in hospital. *BMJ*. 2012 Mar 13;344:e832.

Wong BM, Dyal S, Etchells EE, Knowles S, Gerard L, Diamantouros A, Mehta R, Liu B, Baker GR, Shojania KG. Application of a trigger tool in near real time to inform quality improvement activities: a prospective study in a general medicine ward. *BMJ Qual Saf*. 2015 Apr;24(4):272-81.

World Health Organization (WHO). *Assessing and tackling patient harm: a methodological guide for data-poor hospitals*. Geneva, 2010.

World Health Organization (WHO). *Patient safety incident reporting and learning systems: technical report and guidance*. Geneva, 2020.

World Health Organization (WHO). *Patient Safety: Making health care safer*. Geneva, 2017.

World Health Organization (WHO). *Patient Safety: Rapid Assessment Methods for Assessing Hazards*. Report of the WHO working group meeting. Geneva, 2002.

World Health Organization (WHO). World Health Organization. Regional Office for the Eastern Mediterranean. *Patient safety tool kit*. 2015.

World Health Organization (WHO). World Alliance for Patient Safety: forward programme. Geneva, 2004.

World Health Organization (WHO). World Alliance for Patient Safety. *WHO draft guidelines for adverse event reporting and learning systems: from information to action*. Geneva 2005.

Zegers M, de Bruijne MC, Wagner C, Groenewegen P, van der Wal G, de Vet H. The inter-rater agreement of retrospective assessments of adverse events does not improve with two reviewers per patient record. *J Clin Epidemiol* 2010;63:94-102.

Zegers M, de Bruijne MC, Wagner C, et al. Adverse events and potentially preventable deaths in Dutch hospitals: results of a retrospective patient record review study. *Qual Saf Health Care* 2009;18:297–302.

Zhou X-H, Obuchowski NA, McClish DK. Methods for correcting imperfect standard bias. *Statistical methods in diagnostic medicine*. New York: Wiley; 2002. pp. 359–95.

2 OBJETIVOS

2.1 OBJETIVO GERAL

Avaliar a validade da ferramenta *Global Trigger Tool (GTT)* proposta pelo *Institute for Healthcare Improvement (IHI)* para identificação de eventos adversos em pacientes adultos internados em ambiente hospitalar.

2.2 OBJETIVOS ESPECÍFICOS

- Construir um método padrão referência para identificação de eventos adversos;
- avaliar a validade do *GTT* para identificação de eventos adversos em pacientes adultos hospitalizados em relação ao método padrão referência;
- avaliar a validade do *GTT* para identificação de eventos adversos com dano maior aos pacientes adultos hospitalizados em relação ao método padrão referência.

3 MÉTODO

3.1 DESENHO DO ESTUDO

Trata-se de estudo transversal para avaliação da validade de uma ferramenta para identificação de eventos adversos em pacientes adultos internados. Comporta-se, portanto, como um estudo de teste diagnóstico, no qual o *GTT* é o teste índice e a condição de interesse é a ocorrência de eventos adversos. Devido à inexistência de um padrão ouro reconhecido,¹ um padrão referência foi construído a partir da combinação de diferentes estratégias para identificação de eventos adversos.

Assim como na pesquisa clínica, os estudos de teste diagnóstico possuem diferentes fases, nas quais se propõe responder a diferentes perguntas, usando, para isso, diferentes delineamentos de estudo. Considerando-se as fases das pesquisas que avaliam testes diagnósticos descritas por Sackett e Haynes (Quadro 2),² este estudo pode ser entendido como um estudo de fase III, no qual se avalia, entre pacientes nos quais é clinicamente sensato suspeitar da condição de interesse, se o resultado do teste distingue aqueles com e sem a condição. Esse desenho de estudo parte de uma amostra de população clinicamente representativa para indivíduos que serão testados na prática clínica e não se sabe, com antecedência, quem possui ou não a condição de interesse.

¹ Hanskamp-Sebregts M, Zegers M, Vincent C, et al. Measurement of patient safety: a systematic review of the reliability and validity of adverse event detection with record review. *BMJ Open* 2016;6:e011078.

² Sackett DL, Haynes RB. The architecture of diagnostic research. In: Knottnerus JA, editor. *The evidence base of clinical diagnosis*. London: BMJ Books; 2002. p. 19–38.

Quadro 2 – As quatro fases de estudos de testes diagnósticos

Fase do estudo	Pergunta que o estudo busca responder	Descrição do estudo
Fase I	Os resultados dos testes em pacientes acometidos diferem daqueles em indivíduos normais?	Parte-se de um grupo de pacientes que sabidamente têm a condição de interesse e um grupo de pessoas que definitivamente não a tem.
Fase II	Os pacientes com determinados resultados de testes têm maior probabilidade de ter a condição de interesse?	Parte de grupos de pacientes com testes diagnósticos positivos e negativos e avalia a presença da condição de interesse.
Fase III	Os resultados dos testes distinguem os pacientes com e sem a condição de interesse entre aqueles nos quais é clinicamente sensato suspeitar do transtorno?	Parte-se de uma população na qual é adequado suspeitar da ocorrência da condição de interesse e aplica-se o teste índice e o teste padrão ouro/referência.
Fase IV	Os pacientes submetidos ao teste diagnóstico se saem melhor do que pacientes semelhantes não testados?	O valor final de um teste diagnóstico é medido nos resultados de saúde que se seguem das intervenções diagnósticas e terapêuticas que os resultados do teste precipitam.

Adaptado de Sackett e Haynes³

3.2 POPULAÇÃO DO ESTUDO

O estudo foi realizado no Hospital São Vicente de Paulo, prédio principal do complexo do Hospital das Clínicas da Universidade Federal de Minas Gerais (HC/UFGM). Trata-se de um hospital de grande porte, universitário, público e geral, localizado na região centro-sul de Belo Horizonte, onde são realizadas atividades de ensino, pesquisa e assistência. É referência no sistema municipal e estadual de saúde no atendimento aos pacientes portadores de doenças de média e alta complexidade, incluindo os oncohematológicos e transplantados. À época do estudo, o hospital possuía mais de 500 leitos, sendo 90 de terapia intensiva, com média de 1.500 internações e 1.600 cirurgias por mês, segundo dados administrativos de 2016.

Os critérios de inclusão na população do estudo foram pacientes com 18 anos ou mais que permaneceram internados no hospital por mais de 24 horas entre os dias quatro de outubro a dois de novembro de 2016, período definido como o corte transversal da pesquisa. Foram excluídas as internações cujos

³ Sackett DL, Haynes RB. The architecture of diagnostic research. In: Knottnerus JA, editor. The evidence base of clinical diagnosis. London: BMJ Books; 2002. p. 19–38.

prontuários estavam incompletos ou indisponíveis para avaliação dos revisores em qualquer uma das etapas da pesquisa. Considerou-se prontuário incompleto aquele que não continha elementos críticos do registro, como sumário de alta, prescrições e porções significativas das evoluções médicas e de enfermagem, relevantes para o entendimento do processo de doença e dos desfechos clínicos do paciente.

Os critérios de inclusão e exclusão seguiram as recomendações do *IHI* para a aplicação do *GTT*.⁴ Ressalta-se que não foram assinalados como critérios de exclusão formais as internações psiquiátricas e para fins de reabilitação, como propõe o documento original do *GTT*, uma vez que essas modalidades de internação não integram o escopo de atendimento da instituição onde a pesquisa foi conduzida. Caso um mesmo indivíduo tivesse duas internações no período correspondente, ele poderia ser selecionado aleatoriamente mais de uma vez.

3.3 CÁLCULO AMOSTRAL

Para o cálculo do tamanho da amostra, foram utilizados os seguintes parâmetros: nível de confiança de 95%, margem de erro de 5%, tamanho da população de 1.500 (média mensal histórica de internações da instituição) e proporção mínima de eventos adversos esperada na população de 20%, resultando em uma amostra de 212 internações. A estimativa da proporção mínima esperada na população foi baseada em dados de estudos anteriores que utilizaram o *GTT* e reportaram taxas de ocorrência de eventos adversos entre 7,2% e 27,0% das internações.⁵

Devido a características relacionadas ao fluxo de prontuários dentro da instituição e frequente indisponibilidade deles para consulta pelos pesquisadores, foram aleatorizadas 268 internações, o que corresponde a cerca de 125% do total da amostra pretendido. Cada prontuário foi solicitado ao

⁴ Griffin FA, Resar RK. *IHI Global Trigger Tool for Measuring Adverse Events*. IHI Innovation Series. 2nd ed. Cambridge, MA: Institute for Healthcare Improvement; 2009.

⁵ Hanskamp-Sebregts M, Zegers M, Vincent C, et al. Measurement of patient safety: a systematic review of the reliability and validity of adverse event detection with record review. *BMJ Open* 2016;6:e011078.

Serviço de Arquivo Médico e Estatística (SAME) — serviço responsável pela guarda e disponibilização dos prontuários para fins de pesquisa e de assistência — pelo menos em duas ocasiões diferentes. Após a segunda negativa, outro prontuário era solicitado, seguindo-se a ordem de internações previamente definida pela aleatorização.

Seleção aleatória simples da amostra foi realizada por meio do programa *R version 3.0.1 (2013-05-16) Copyright (C) 2013 The R Foundation for Statistical Computing*.

3.4 DEFINIÇÕES

Foi considerado evento adverso “qualquer dano físico, não intencional, resultado direto de um cuidado em saúde, ou que teve o cuidado em saúde como fator contribuinte, e que requer tratamento adicional, hospitalização ou prolongamento do tempo de internação, ou que ocasione a morte do paciente”.⁶ Ressalta-se que, diferentemente do proposto pelo *IHI* no manual do *GTT*, foram incluídos no escopo deste estudo os eventos adversos resultantes de atos de omissão que tinham clara relação com a ocorrência de danos para o paciente. Entre esses eventos inclui-se o atraso para a realização de propedêutica ou tratamentos bem indicados devido a falhas organizacionais, como filas para agendamento ou indisponibilidade de profissionais habilitados ou outros recursos para realizar determinado tipo de cuidado.

Detectar eventos adversos é especialmente desafiador devido à relação causal do dano com o cuidado ser muitas vezes invisível. Quando um paciente experimenta um desfecho ruim, geralmente presume-se que o resultado foi causado pelo processo da doença subjacente. Embora isso muitas vezes seja verdade, uma análise aprofundada frequentemente determinará que os cuidados em saúde foram, pelo menos, parcialmente responsáveis. No entanto, essa determinação depende da interpretação de um conjunto de informações disponíveis no momento da análise, frequentemente relatos ou registros retrospectivos e incompletos, por um ou mais especialistas.

⁶ Griffin FA, Resar RK. *IHI Global Trigger Tool for Measuring Adverse Events*. IHI Innovation Series. 2nd ed. Cambridge, MA: Institute for Healthcare Improvement; 2009.

A subjetividade do julgamento do(s) revisor(es) sobre a presença de dano e da atribuição dele ao cuidado é uma limitação de todos os métodos para identificação de eventos adversos. Isso pode interferir tanto na precisão quanto na acurácia de um teste. Estratégias para suavizar essa limitação incluem a padronização do processo de revisão, uso de definições operacionais claras (em especial sobre conceito de dano e classificações em segurança do paciente) e treinamento e constância dos revisores.⁷

Nesse sentido, para reduzir o grau de subjetividade do julgamento do revisor médico na determinação da ocorrência de um evento adverso foi utilizada uma escala de 4 pontos para determinar o nível de confiança de que o dano poderia ser atribuído ao cuidado em saúde, ao invés do processo de doença do paciente (Quadro 3). Foram considerados eventos adversos aqueles classificados como “2” ou “3”, ou seja, quando havia evidência moderada à certa de que o dano poderia ser atribuído ao cuidado. Os eventos classificados como “0” ou “1” foram descartados. Uma série de perguntas adaptada do *Canadian Adverse Events Study*⁸ e traduzida para o Português por Mendes e colaboradores⁹ foi proposta para guiar o raciocínio do médico revisor (Anexo B).

Quadro 3 - Escala e instruções fornecidas aos médicos revisores para julgar a existência de um evento adverso

Após análise de informações sobre o contexto clínico e os cuidados oferecidos ao paciente, qual o nível de confiança que você tem de que foi o **CUIDADO EM SAÚDE** que causou a lesão ou dano?

0 - Praticamente nenhuma evidência de que o cuidado foi a causa da lesão ou dano.

1 - Evidência mínima a modesta de que o cuidado foi a causa da lesão ou dano.

2 - Evidência moderada a forte de que o cuidado foi a causa da lesão ou dano.

Adaptado de Baker e colaboradores⁷ e Mendes e colaboradores⁸

⁷ Hulley SB et al. Delineando a pesquisa Clínica: uma abordagem epidemiológica. 4ª edição. Artmed, 2015. 386p.

⁸ Baker GR, Norton PG, Flintoft V, Blais R, et al. The Canadian Adverse Events Study: the incidence of adverse events among hospital patients in Canada. CMAJ. 2004 May 25;170(11):1678-86.

⁹ Mendes W, Martins M, Rozenfeld S et al. The assessment of adverse events in hospitals in Brazil. *Int J Qual Health Care* 2009 Aug;21(4):279-84.

Todos os eventos identificados, independentemente por qual método, foram classificados quanto à natureza e quanto à gravidade do dano. Para essa última, foi utilizada a adaptação realizada pelo *IHI* do índice de categorização de erros de medicação do *National Coordinating Council for Medication Error Reporting and Prevention (NCC MERP)*.¹⁰ Essa adaptação consistiu na exclusão de categorias que não contemplavam dano ao paciente e de qualquer referência a erros contidas nas descrições das categorias, ampliando a relação causal do dano para qualquer fator relacionado ao cuidado e que não poderia ser atribuído à evolução da doença. A classificação final quanto ao dano para o paciente está descrita no Quadro 4. Quando um mesmo evento identificado pelos dois métodos foi classificado em categorias diferentes de gravidade do dano, foi mantida a classificação de maior gravidade.

Quadro 4 - Classificação dos eventos adversos quando à gravidade do dano

E - dano temporário ao paciente, com necessidade de intervenção; F - dano temporário ao paciente, com necessidade de iniciar ou prolongar hospitalização; G - dano permanente ao paciente; H - intervenção necessária para sustentar a vida; I: morte do paciente.
--

Traduzido de Griffin e Resar, 2009¹¹

3.5 VARIÁVEIS

Dados descritivos para caracterização das internações contidas na amostra foram recolhidas dos prontuários por meio do preenchimento do “Formulário 1” (Apêndice A). Parte das informações contidas nesse formulário foram utilizadas por outros pesquisadores em projetos de pesquisa complementares, nos quais foram avaliados outros elementos da segurança do paciente.

¹⁰ NCC MERP: National Coordinating Council for Medication Error Reporting and Prevention. NCC MERP Index for Categorizing Medication Errors Algorithm; 1996 [revised February 20, 2001]. NCC MERP; 2001.

¹¹ Griffin FA, Resar RK. IHI Global Trigger Tool for Measuring Adverse Events. IHI Innovation Series. 2nd ed. Cambridge, MA: Institute for Healthcare Improvement; 2009.

No contexto deste estudo, as internações foram caracterizadas quanto a(o):

- idade do paciente no momento da admissão;
- gênero;
- classificação de gravidade e complexidade do paciente no momento da admissão dado pelo Índice de Comorbidade de Charlson original¹² e classificação adaptada segundo valores de Charlson (“Baixo” para resultados de Charlson de zero, “Médio” para resultados iguais a um ou dois, “Alto” para resultados iguais a três ou quatro e “Muito alto” para resultados iguais ou maiores que cinco);
- duração da internação em dias;
- caráter da internação — urgente ou eletiva;
- motivo da internação — cirúrgico, clínico, obstétrico.

As variáveis de interesse em um estudo de avaliação da validade de um teste são os resultados que podem ser obtidos por esse teste. Os testes diagnósticos comumente apontam a presença ou ausência de uma condição, frequentemente uma doença. Utilizando-se o mesmo racional, pode-se assumir que, ao aplicar um teste para identificação de eventos adversos em uma amostra de internações, os resultados possíveis são: haverá internações nas quais serão identificados eventos adversos e outras nas quais nenhum evento adverso será identificado. Estudos prévios que avaliaram o *GTT* como método diagnóstico utilizaram essa métrica.^{13,14}

Entretanto, a informação sobre a identificação ou não de um ou mais eventos adversos não fornece dados qualificados e quantitativos sobre o dano experimentado pelo paciente. Seria o mesmo que dizer que um paciente possui uma ou mais doenças sem especificar qual ou quais. Considerando-se que o

¹² Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis*. 1987;40(5):373-83. doi: 10.1016/0021-9681(87)90171-8. PMID: 3558716.

¹³ Naessens JM, Campbell CR, Huddleston JM, et al. A comparison of hospital adverse events identified by three widely used detection methods. *Int J Qual Health Care* 2009;21:301–7.

¹⁴ Classen DC, Resar R, Griffin F, Federico F, Frankel T, Kimmel N, Whittington JC, Frankel A, Seger A, James BC. 'Global trigger tool' shows that adverse events in hospitals may be ten times greater than previously measured. *Health Aff (Millwood)*. 2011 Apr;30(4):581-9. Erratum in: *Health Aff (Millwood)*. 2011 Jun;30(6):1217.

mesmo paciente pode apresentar nenhum, um, dois, três, ou inúmeros eventos adversos em uma única internação, a avaliação da validade de um método que se propõe identificar “eventos adversos em internações”, e não “internações com eventos adversos”, deveria, idealmente, considerar cada evento adverso detectado como um desfecho único e individual.

Para contemplar as duas abordagens descritas, aquela utilizada em outros estudos semelhantes da área e a que entendemos ser a mais adequada, a condição de interesse “eventos adversos” foi avaliada por meio de duas unidades de estudo distintas: 1) por internação de paciente, na qual a ocorrência de um ou mais eventos adversos numa internação do paciente contava sempre como um único caso positivo, e 2) por evento adverso, na qual cada evento adverso identificado numa internação contava como um caso positivo. Em ambas, as internações de pacientes nas quais nenhum evento adverso foi detectado foram contadas cada uma como um desfecho negativo.

3.6 APLICAÇÃO DOS TESTES ÍNDICE E REFERÊNCIA

O teste índice (*GTT*) e o padrão referência construído foram aplicados na mesma amostra selecionada de maneira independente, garantindo o cegamento entre os avaliadores. A aplicação dos testes seguiu as características metodológicas de cada um deles, sendo exclusivamente retrospectiva para o *GTT*, por meio da análise de prontuários, e formada por uma combinação de estratégias no método padrão referência construído.

Embora o *GTT* permita a identificação de eventos adversos durante todo o período de internação, o método padrão referência construído teve a análise restrita ao período do corte transversal. Por esse motivo, foram contabilizados apenas os eventos adversos ocorridos nesse período, independentemente do início ou término da internação, ou anteriores a ela, desde que diretamente relacionados à internação atual, como aqueles que causaram danos temporários ou permanentes, exigindo novos tratamentos ou intervenções.

3.6.1 Método padrão referência

Considerando-se a ausência de um padrão ouro bem definido para identificação de eventos adversos,¹⁵ optou-se pela elaboração de um protocolo específico para a construção de um padrão referência.^{16,17} Fundamentados nas conclusões de estudos prévios que apontaram que, em geral, diferentes métodos identificam diferentes eventos,^{18,19,20,21} estratégias distintas de busca foram combinadas para identificação de potenciais danos ao paciente gerados pelo cuidado em saúde nas mais diversas situações e cenários durante uma internação.

O protocolo de construção do padrão referência incluiu elementos de métodos consagrados, como entrevistas com profissionais, revisão de registros e dados clínicos individuais, análise de relatórios de dados agrupados e avaliação de notificações voluntárias de incidentes. Procurou-se, assim, superar as fragilidades individuais de cada método, principalmente no que diz respeito à subnotificação, à ausência de registro adequado em prontuários e às divergências de dados administrativos.^{22,23}

¹⁵ Hanskamp-Sebregts M, Zegers M, Vincent C, et al. Measurement of patient safety: a systematic review of the reliability and validity of adverse event detection with record review. *BMJ Open* 2016;6:e011078.

¹⁶ Rutjes AWS, Reitsma JB, Coomarasamy A, Khan KS, Bossuyt PMM. Evaluation of diagnostic tests when there is no gold standard. A review of methods. *Health Technol Assess* 2007;11(50).

¹⁷ Umemneku Chikere CM, Wilson K, Graziadio S, Vale L, Alen AJ. Diagnostic test evaluation methodology: A systematic review of methods employed to evaluate diagnostic tests in the absence of gold standard - An update. *PLoS One*. 2019 Oct 11;14(10):e0223832.

¹⁸ Naessens JM, Campbell CR, Huddleston JM, et al. A comparison of hospital adverse events identified by three widely used detection methods. *Int J Qual Health Care* 2009;21:301–7.

¹⁹ Classen DC, Resar R, Griffin F, Federico F, Frankel T, Kimmel N, Whittington JC, Frankel A, Seger A, James BC. 'Global trigger tool' shows that adverse events in hospitals may be ten times greater than previously measured. *Health Aff (Millwood)*. 2011 Apr;30(4):581-9. Erratum in: *Health Aff (Millwood)*. 2011 Jun;30(6):1217.

²⁰ Kennerly DA, Kudyakov R, da Graca B, et al. Characterization of adverse events detected in a large health care delivery system using an enhanced global trigger tool over a five-year interval. *Health Serv Res*. 2014;49: 1407–1425.

²¹ Rutberg H, Borgstedt Risberg M, Sjodahl R, et al. Characterisations of adverse events detected in a university hospital: a 4-year study using the Global Trigger Tool method. *BMJ open*. 2014;4:e004879.

²² World Health Organization (WHO). Patient Safety: Rapid Assessment Methods for Assessing Hazards. Report of the WHO working group meeting. Geneva, 2002.

²³ Michel P, Quenon JL, de Sarasqueta AM, Scemama O. Comparison of more methods for estimating rates of adverse events and rates of preventable adverse events in acute care hospitals. *BMJ*. 2004 Jan 24;328(7433):199.

O método padrão referência foi executado em duas etapas: 1) busca por indícios que sugeriam a ocorrência de eventos adversos nas internações, chamados aqui genericamente de rastreadores; 2) confirmação do possível evento adverso.

Etapa 1: busca por rastreadores

Os rastreadores foram selecionados com base em metodologias e protocolos de pesquisa de estudos publicados na área e na percepção dos pesquisadores de oportunidades de obtenção de dados de interesse dentro do contexto institucional e englobaram diferentes fontes e meios de coleta, descritos abaixo. Destaca-se que a busca em todos os tipos de fontes e a identificação dos rastreadores se restringiu aos dias de internação delimitados no corte transversal do estudo, ou seja, entre os dias quatro de outubro e dois de novembro de 2016, chamado aqui de “período da pesquisa”. A depender do tipo de fonte, essa busca por rastreadores foi realizada em tempo corrente, durante a internação, ou de maneira retrospectiva.

a) Rastreadores de cuidado quase em tempo real obtidos por entrevistas com profissionais assistenciais

Os profissionais que prestavam assistência direta aos pacientes, principalmente enfermeiros e técnicos de enfermagem, mas também médicos e outros membros da equipe multiprofissional, foram entrevistados ao final de cada plantão durante os 30 dias correspondentes ao período da pesquisa, incluindo finais de semana e plantões noturnos. Por meio de um roteiro semiestruturado, os entrevistadores perguntaram sobre situações que pudessem indicar a ocorrência de um evento adverso. Esses critérios de triagem foram selecionados de acordo com as particularidades dos setores visitados e estão descritos nos Apêndices B a F. Os setores de cuidado incluíram as enfermarias de internação, pronto socorro, unidades de terapia intensiva, bloco cirúrgico, setores de radiologia e de endoscopia. Os profissionais foram questionados sobre, por exemplo, a ocorrência de reações adversas a medicamentos, procedimentos invasivos e possíveis complicações, quedas ou lesões por pressão, problemas com acessos

vasculares ou outros dispositivos invasivos, necessidade de avaliação pelo plantão ou ocorrência de atrasos na prestação de cuidados, entre outros.

O grupo de entrevistadores foi composto por 18 alunos de graduação de três cursos da área da saúde — enfermagem, fonoaudiologia e medicina, que se revezaram nas visitas aos setores de atendimento ao paciente adulto da instituição. Eles foram supervisionados por três profissionais médicos e receberam treinamento teórico-prático com duração de cerca de 10 horas. Essa etapa da coleta de dados contou com o apoio da alta direção do hospital, de lideranças clínicas e de membros da unidade de qualidade e segurança do paciente, que validaram e reforçaram, junto aos profissionais da ponta, a importância de se identificar falhas no processo de cuidado em saúde no intuito de promover melhorias para os pacientes.

b) Revisão diária das prescrições

Diariamente, entre os dias quatro de outubro e dois de novembro de 2016, um grupo de 12 alunos de graduação do curso de farmácia, treinados e supervisionados por uma farmacêutica sênior, avaliaram as prescrições e os registros dos enfermeiros assistenciais e técnicos de enfermagem de todos os pacientes em todos os setores de internação (enfermarias e terapias intensivas) e pronto socorro. Os pesquisadores foram orientados a buscar dados que sugerissem a ocorrência de efeitos adversos a medicamentos ou erros de medicação, como a suspensão abrupta de medicamentos, ausência de administração apesar da prescrição e medicamentos em falta.

c) Resultados de exames laboratoriais

Rastreadores laboratoriais incluíram resultados de exames alterados, selecionados pelo seu potencial de refletir alterações orgânicas decorrentes da ocorrência de um evento adverso, como aumento nos níveis de creatinina, alterações em sódio e elevações em provas de coagulação (Apêndice G). A busca por esses dados foi realizada de forma retrospectiva pelo sistema eletrônico de controle, registro e liberação de resultados de exames laboratoriais adotado pela instituição por período (entre os dias

quatro de outubro e dois de novembro de 2016) e por tipo de exame laboratorial e filtrados pelos valores de interesse.

d) Prescrição de antibióticos e dados sobre infecções relacionadas à assistência em saúde

Avaliação dos relatórios oficiais fornecidos pela Comissão de Controle de Infecções Hospitalares (CCIH) sobre infecções relacionadas à assistência em saúde que ocorreram durante o período da pesquisa, associada à revisão de todos os pedidos de autorização para início de antibióticos, condição fundamental para dispensação de qualquer medicamento antimicrobiano pela farmácia da instituição.

e) Intercorrências durante cirurgias, cuidados obstétricos ou procedimentos invasivos

Foram revisados os boletins operatórios, folhas de sala e *checklists* de cirurgia segura de procedimentos invasivos, cirúrgicos e obstétricos, realizados no período da pesquisa. Buscou-se por evidências de mudança de procedimento proposto, duração do procedimento excedendo o tempo previsto, descrição de intercorrências durante o procedimento, necessidade de transfusão de sangue, uso de medicamentos selecionados ou qualquer outra evidência de ocorrência de incidente de cuidado. Além disso, dados administrativos foram avaliados para a identificação de cancelamentos ou reabordagem precoce.

f) Dados da Agência Transfusional

Informações sobre notificações recebidas e apuradas pela Agência Transfusional local que confirmaram a ocorrência de eventos adversos relacionados ao uso de hemocomponentes e hemoderivados no período da pesquisa.

g) Notificação voluntária

Revisão dos dados de notificação voluntária registrados por meio do sistema de notificação eletrônico oficial da instituição para as internações ocorridas no período e fornecidos pelo Núcleo de Segurança do Paciente.

h) Reinternação precoce

Revisão do motivo de reinternação hospitalar precoce, ou seja, que ocorreu num prazo inferior a 30 dias da última alta. As informações foram obtidas a partir de dados administrativos.

i) Avaliação dos óbitos

Avaliação retrospectiva das possíveis causas dos óbitos ocorridos no período por meio da avaliação dos registros em prontuário.

j) Achados durante análise de prontuários em outras etapas da pesquisa

Qualquer evidência clara de ocorrência de evento adverso notada pelos pesquisadores durante a manipulação do prontuário para coleta de dados individuais ou durante a análise de outros rastreadores identificados na formação do padrão referência. É importante ressaltar que não inclui os eventos encontrados durante a aplicação do *GTT*.

Destaca-se que o conceito de “rastreadores” adotado no protocolo para formação do padrão referência é semelhante ao que define os “*triggers*” do *GTT*, ou seja, são dicas ou evidências que sinalizam para a possibilidade de que tenha ocorrido algum evento adverso durante o cuidado do paciente. Essa diferenciação nominal foi adotada apenas com a finalidade de destacar a independência metodológica do padrão referência construído dentro do protocolo da pesquisa em relação à ferramenta em análise. Apesar de semelhança na definição, eles não correspondem exatamente aos mesmos elementos de triagem, sendo muito mais amplos no protocolo para formação do padrão referência.

Os dados sobre os rastreadores foram coletados por meio do Formulário 2A (Apêndice H), de maneira a garantir o sigilo tanto de dados de identificação do paciente quanto do profissional que forneceu a informação. Cada internação poderia ter uma infinidade de rastreadores positivos, sendo que vários deles poderiam corresponder a apenas um evento adverso.

Por questões de viabilidade metodológica, a busca pelos rastreadores descritos na fase 1 do método padrão referência ocorreu de maneira indistinta para todos os indivíduos da população. A fase 2, relacionada a confirmação de um possível evento, se restringiu às internações selecionadas aleatoriamente na amostra.

Etapa 2: confirmação e classificação do evento adverso

Na segunda etapa, que ocorreu entre junho e outubro de 2017, de forma concomitante, porém independente, à aplicação do *GTT*, dois profissionais médicos, com formação em clínica médica e mais de cinco anos de experiência, se revezaram para avaliar os prontuários das internações que tiveram pelo menos um rastreador positivo. Eles buscavam por informações adicionais, além daqueles fornecidas pelos rastreadores, para validar a ocorrência de dano ao paciente relacionado ao cuidado.

Após a confirmação de um evento adverso, classificavam-no quanto à escala de confiança de que o dano foi resultado do cuidado em saúde e quanto à gravidade do dano. Em caso de dúvida, os médicos poderiam consultar um ao outro ou, em situações que demandavam conhecimento específico, acionar médicos especialistas colaboradores. As informações referentes aos eventos adversos identificados foram registradas no Formulário 2B (Apêndice I).

As etapas metodológicas do padrão referência construído estão resumidas na Figura 3:

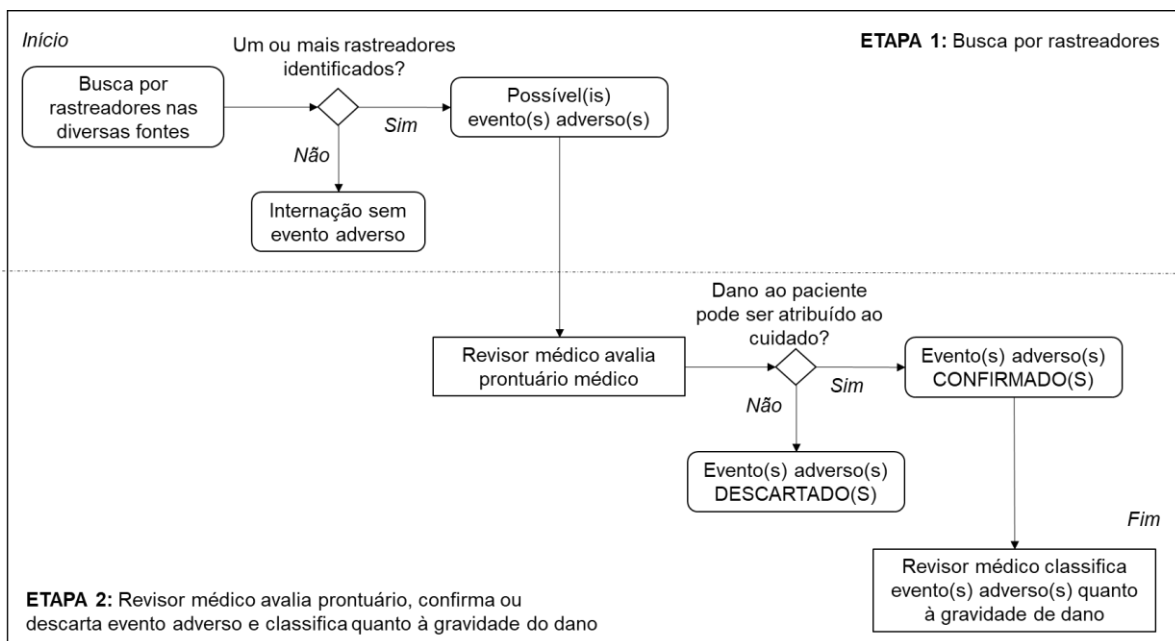


Figura 3 – Método padrão referência para identificação de eventos adversos

É importante ressaltar que precauções foram tomadas para reduzir a chance de vieses. Uma revisão de prontuários formal com uso de *triggers* semelhantes ao utilizado no *GTT* não foi incluída no método padrão referência e os revisores de um método não tiveram acesso aos resultados obtidos pelo outro método, evitando-se, respectivamente, vies de incorporação e vies de análise.

3.6.2 Global Trigger Tool

Na época do estudo, não havia uma versão oficial do *GTT* em português. A tradução e adaptação da ferramenta foi realizada utilizando-se como documentos base o manual original do *GTT* e uma versão da ferramenta em Português, traduzida e cedida para uso na pesquisa pelo time da América Latina do *IHI*.

O *IHI* recomenda que as traduções e adaptações sejam simples e voltadas para a viabilização do processo da busca pelos *triggers* durante a revisão de prontuários em cada instituição. As principais adaptações necessárias se relacionam, em geral, a elementos de cunho prático, como por exemplo a substituição de medicamentos não padronizados por medicamentos padronizados, levando-se em consideração a intenção por trás de cada *trigger*.

Nesse sentido, as definições dos *triggers* foram mantidas o mais próximo possível daquelas descritas na segunda edição da versão original do *GTT* e apenas pequenos ajustes foram necessários. A versão final dos formulários do *GTT* utilizados no estudo podem ser consultadas nos Apêndices J e K — “Folha de Trabalho” e “Sumário da Revisão”, respectivamente. Essa fase do estudo foi compartilhada com outros projetos de pesquisa relacionados e detalhes sobre a adaptação e aplicação do *GTT* no contexto deste protocolo já foram descritos em outra publicação.²⁴

A equipe de revisores foi constituída por uma dupla de revisores primários formada por dois acadêmicos de medicina, um do quarto o outro do quinto ano do curso, e por dois profissionais médicos com formação em clínica médica e mais de 10 anos de experiência em cuidado hospitalar, que exerceram o papel de revisor médico, se revezando, de acordo com a disponibilidade, para validar os dados coletados pelos revisores primários. Todos os revisores tinham familiaridade com o formato do prontuário médico e com o modelo assistencial oferecido pela instituição. A qualificação dos revisores incluiu a leitura individual do documento oficial do *GTT*, seguida de um treinamento teórico-prático de 12 horas com foco no significado de cada *trigger*, na compreensão do conceito de evento adverso e na classificação de dano. Dez prontuários médicos foram avaliados por todos os revisores, incluindo cinco exemplos comentados do *IHI* e cinco registros da instituição. Todos os resultados foram discutidos em conjunto a fim de reduzir a variação de abordagem. Os revisores receberam um manual descrevendo cada *trigger* e apontando em quais partes do prontuário institucional eles provavelmente poderiam ser encontrados.

A aplicação do *GTT* ocorreu entre junho e outubro de 2017, de forma concomitante à fase 2 da aplicação do método padrão referência, mantendo-se, contudo, a independência e o cegamento na aplicação dos dois testes. Cerca de 10 internações foram revisadas por semana.

²⁴ Moraes SM, Ferrari TCA, Figueiredo NMP, Almeida TNC, Sampaio CCL, Andrade YCP, Benedito RO, Beleigoli AM. Assessment of the reliability of the IHI Global Trigger Tool: new perspectives from a Brazilian study. *Int J Qual Health Care*. 2021 Mar 18;33(1):mzab039.

Numa primeira etapa, cada revisor primário teve um tempo máximo de 20 minutos por prontuário para, individualmente, identificar os *triggers*, analisar a ocorrência de possíveis eventos adversos, classificar os danos e anotar os achados na Folha de Trabalho. Posteriormente, a dupla de revisores primários se reuniu para discutir os achados e registrar o consenso no Sumário da Revisão. Os revisores primários foram instruídos a apresentar os achados sobre cada possível evento adverso de forma descritiva e detalhada, destacando o contexto, data de ocorrência, dano causado e intervenções necessárias.

Destaca-se que o desempenho da dupla de acadêmicos de medicina foi comparado ao desempenho de uma dupla de enfermeiros experientes, composição mais frequentemente recomendada para a formação do grupo de revisores primários.²⁵ Testes de confiabilidade utilizando-se o kappa foram aplicados, com resultados que demonstraram não haver diferenças significativas no desempenho das duas duplas.²⁶

Na segunda etapa, o revisor médico, com base nas conclusões dos revisores primários descritas no Sumário da Revisão, validou os eventos adversos identificados e a classificação quanto ao dano. Para reduzir o grau de subjetividade no julgamento sobre a relação do dano com o cuidado, eles foram instruídos a utilizar a escala de confiança descrita previamente no quadro 3 e os eventos classificados como 0 ou 1 foram descartados. Os revisores médicos puderam consultar os prontuários médicos para esclarecimentos e contaram com o apoio de um grupo de médicos especialistas, os quais foram consultados em situações complexas, quando não era evidente a atribuição do dano ao cuidado em saúde.

O fluxo de aplicação do *GTT* está demonstrado na Figura 4.

²⁵ Griffin FA, Resar RK. IHI Global Trigger Tool for Measuring Adverse Events. IHI Innovation Series. 2nd ed. Cambridge, MA: Institute for Healthcare Improvement; 2009.

²⁶ Moraes SM, Ferrari TCA, Figueiredo NMP, Almeida TNC, Sampaio CCL, Andrade YCP, Benedito RO, Beleigoli AM. Assessment of the reliability of the IHI Global Trigger Tool: new perspectives from a Brazilian study. *Int J Qual Health Care*. 2021 Mar 18;33(1):mzab039.

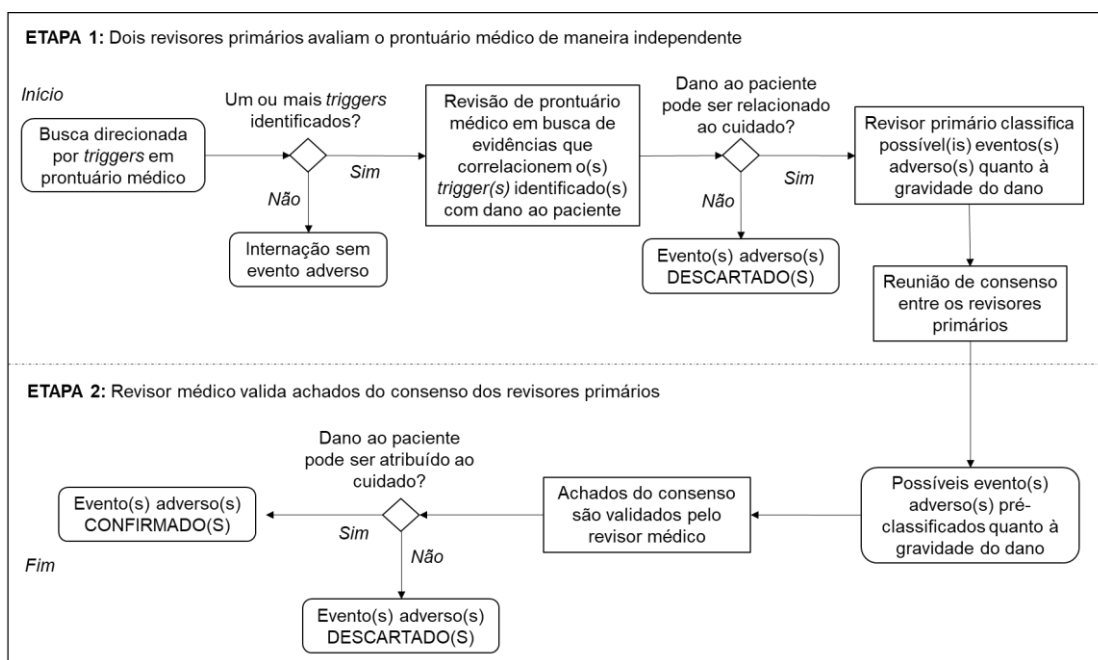


Figura 4 – Método GTT para identificação de eventos adversos

3.7 ESTATÍSTICA

Frequências absolutas e relativas foram utilizadas para descrever as variáveis qualitativas e medidas de tendência central e dispersão para as variáveis quantitativas. A frequência de ocorrência de eventos adversos foi apresentada como o número de eventos adversos por 1.000 pacientes-dia (número total de eventos adversos identificados / número total de dias de internação avaliados somando-se todas as internações x 1.000). Teste não paramétrico para duas amostras relacionadas, o teste de McNemar, foi utilizado para comparar o desempenho dos dois testes – o teste índice e o padrão referência construído – para identificar eventos de diferentes classificações de gravidade do dano.

Considerando-se que não há um padrão ouro/referência preferido para identificação eventos adversos e que não se conhece as medidas de acurácia do teste padrão referência que foi construído dentro deste protocolo de pesquisa, optou-se por avaliar a validade do GTT utilizando-se duas estratégias: 1) medindo-se o nível de concordância do GTT com o padrão

referência e 2) calculando-se “estimativas ingênuas” de acurácia diagnóstica do *GTT* em relação ao padrão referência.²⁷

A análise do nível de concordância foi realizada através do coeficiente de concordância Kappa de Cohen ajustado por prevalência (*PABAK*), que busca ajustar o coeficiente Kappa para o impacto que desbalanço entre a prevalência das condições “presença de evento adverso” e “ausência de evento adverso” possam ter.²⁸ O valor do coeficiente de Kappa ajustado por prevalência varia de “menos um” (quando a proporção de concordância é igual a zero) a “um” (quando a proporção de concordância é igual a um) e é “igual a zero” quando a proporção de concordância é de 0,50. Quanto mais próximo de “um” for seu valor, maior é o indicativo de que existe concordância entre os métodos e, quanto mais próximo de “zero”, maior é o indicativo de que a concordância é puramente aleatória. A interpretação usual para o valor do coeficiente de Kappa ajustado é a mesma utilizada para a interpretação do coeficiente de Kappa e está descrito no Quadro 5.

Quadro 5 - Valores de Kappa e respectivos níveis de concordância

Valor de Kappa	Nível de Concordância
Entre -1 e 0	insignificante (<i>poor</i>)
Entre 0 e 0,2	fraco (<i>slight</i>)
Entre 0,21 e 0,4	razoável (<i>fair</i>)
Entre 0,41 e 0,6	moderado (<i>moderate</i>)
Entre 0,61 e 0,8	forte (<i>substantial</i>)
Entre 0,81 e 1	quase perfeito (<i>almost perfect</i>)

Adaptado de Landis e Kock (1977)²⁹.

O resultado do teste em estudo, ou seja, a variável de interesse, foi considerado dicotômico e o cálculo das “estimativas ingênuas” de acurácia foi

²⁷ Trikalinos TA, Balion CM. Chapter 9: options for summarizing medical test performance in the absence of a "gold standard". *J Gen Intern Med.* 2012 Jun;27 Suppl 1(Suppl 1):S67-75.

²⁸ Byrt T, Bishop J, Carlin JB. Bias, prevalence and kappa. *J Clin Epidemiol.* 1993 May;46(5):423-9.

²⁹ Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*, v. 33, n. 1, p. 159–74, Mar 1977.

realizado utilizando-se tabelas 2x2 contendo os resultados obtidos pelo *GTT* e pelo padrão referência construído, a partir das quais foram calculadas a sensibilidade, especificidade e acurácia global.

A avaliação de validade foi realizada utilizando-se duas unidades de estudos: 1) por internação e 2) por evento adverso. Nessa segunda situação, foram considerados como “resultados positivos” a soma de todos os eventos adversos encontrados por cada método e como “resultados negativos” a soma dos eventos identificados que o método em questão não foi capaz de encontrar, acrescido do número de internações que não tiveram nenhum evento adverso identificado por nenhum método.

Alguns pesquisadores destacam que os eventos adversos da categoria E de gravidade do dano são os mais difíceis de detectar e categorizar ao utilizar o *GTT* e que a exclusão desses eventos poderia aumentar a confiabilidade do método em estudos de precisão.³⁰ Para verificar a existência de possíveis diferenças na capacidade do *GTT* identificar eventos adversos de diferentes perfis de gravidade, a validade da ferramenta foi avaliada tanto para identificação de eventos adversos de qualquer gravidade, quanto para o subgrupo de eventos de maior gravidade, classificados como F, G, H ou I.

O software utilizado para as análises foi o R (versão 3.5.0).

3.8 CONSIDERAÇÕES ÉTICAS

A pesquisa foi aprovada pelo Comitê de Ética em Pesquisa da Universidade Federal de Minas Gerais e por meio do sistema eletrônico Plataforma Brasil (CAAE: 52532715.7.0000.5149). Foi concedida dispensa da obtenção do Termo de Consentimento Livre e Esclarecido (TCLE) uma vez que as informações foram obtidas por meio de avaliação de dados administrativos e de prontuário e manejadas de forma anônima, sem contato direto com pacientes ou familiares e sem instituição de nenhuma intervenção. A participação dos

³⁰ Mattsson TO, Knudsen JL, Lauritsen J, Brixen K, Herrstedt J. Assessment of the global trigger tool to measure, monitor and evaluate patient safety in cancer patients: reliability concerns are raised. *BMJ Qual Saf.* 2013 Jul;22(7):571-9.

profissionais de saúde nas entrevistas para coleta de rastreadores quase em tempo real foi voluntária e o sigilo sobre a informação e a fonte foi garantido.

3.9 CONFLITOS DE INTERESSE

A doutoranda atuou como voluntária como Líder Mundial de Capítulos na *Open School IHI* entre setembro de 2017 e agosto de 2019 e fez parte de um grupo de voluntários do Brasil e de Portugal para a adaptação para o português do White Paper "*IHI Global Trigger Tool for Measuring Adverse Events*" em 2019.

4 RESULTADOS

Os resultados estão apresentados no formato de artigo intitulado “The accuracy of Global Trigger Tool is higher for identification of adverse events of greater harm”. Esta é a versão original do artigo, que foi publicado no International Journal for Quality in Health Care, em 28 de fevereiro de 2023, após breve revisão. A versão final do artigo pode ser acessada em:

https://academic.oup.com/intqhc/advance-article-abstract/doi/10.1093/intqhc/mzad005/7008723?utm_source=advanceaccess&utm_campaign=intqhc&utm_medium=email

Int J Qual Health Care. 2023 Feb 28;34(1):mzad005.

doi: 10.1093/intqhc/mzad005.

THE ACCURACY OF GLOBAL TRIGGER TOOL IS HIGHER FOR IDENTIFICATION OF ADVERSE EVENTS OF GREATER HARM

Sara Monteiro Moraes¹

Teresa Cristina Abreu Ferrari¹

Alline Beleigoli²

¹ Universidade Federal de Minas Gerais Faculdade de Medicina, Departamento de Clínica Médica, Belo Horizonte, Brazil

² Flinders University, Caring Futures Institute, Adelaide, Australia

Abstract

Background Global Trigger Tool (GTT) of the Institute for Healthcare Improvement (IHI) has been used as a measurement strategy in patient safety by several institutions and national programs. Although the greater ability of the GTT to identify adverse events (AEs) compared to other methods has already been demonstrated, there are few data on its accuracy and studies suggest lower sensitivity for minor AEs. This study aimed to assess the validity of the GTT for identifying AEs in adult inpatients for all AEs and for the subgroup of AEs with greater harm to the patient, classified as F-I on the IHI-GTT adapted version of the National Coordinating Council for Medication Error Reporting and Prevention (NCC MERP) Index for Categorizing Errors.

Methods In this diagnostic test study, GTT is the index test and identification of AEs (yes/no) represents the condition of interest. Due to the lack of a gold standard test, a composite reference standard method (RSM) was developed. RSM combined real-time (during hospitalizations) and retrospective search of medical records and administrative data for screening criteria and AEs. Both tests were applied to a random sample of 211 hospitalizations of adult inpatients during Oct/Nov 2016 in a large public hospital in Belo Horizonte, Brazil. The validity of the GTT was evaluated using sensitivity, specificity and global accuracy.

Results A total of 176 AEs were identified in 67 admissions using RSM and 129 AEs in 76 admissions using GTT, resulting in rates of 126 and 93 AEs/1,000 patient-days, respectively. Sensitivity, specificity and global accuracy of the GTT for identification of individual AEs were, respectively, 0.41 (95% confidence interval [CI] 0.34;0.49), 0.68 (95%CI 0, 60;0.74) and 0.54 (95%CI 0.49;0.60) for all AEs, regardless of the harm categorization, and 0.85 (95%CI 0.72;0.93), 0.88 (95%CI 0.82;0.92) and 0.87 (95%CI 0.82;0.91) for the subgroup of AEs categorized as harm F-I. Among the main AEs missed by the GTT are AEs related to nursing care, such as those related to peripheral venous access and gastric/enteric catheters.

Conclusion GTT proved to be a valid method for identifying AEs in adult inpatients. Its accuracy increases when minor harm AEs are not counted. Among the main adverse events missed by the GTT are those related to nursing care. Therefore, the GTT should be used in conjunction with other measurement strategies to achieve results that are representative of the quality profile of the care provided and, thus, guide the best improvement strategies.

Key words

Patient safety, Quality measurement, Adverse events, Global trigger tool

Introduction

One of the major challenges healthcare systems face today is providing safe and quality care in complex, pressurized and fast-moving environments.[1] A reliable measurement strategy is required to determine whether efforts to enhance safety are resulting in overall improvements.[2,3]

The Global Trigger Tool (GTT) of the Institute for Healthcare Improvement (IHI) is a simple, inexpensive, and easy-to-execute method for estimating the occurrence of AEs in adult hospitalized patients.[4,5] Although its greater ability to identify AEs compared to other methods has already been demonstrated,[6,7,8,9,10] there are few data on its accuracy.[3,11,12] Moreover, studies suggest that the minor harm AEs are more difficult to be identified by the GTT and that the exclusion of these events could increase the validity of the method.[2,5]

The aim of this study was to assess the validity of the GTT for identification of AEs in adult inpatients for all AEs and for the subgroup of AEs with greater patient harm.

Methods

Study design, population and sample

This is a diagnostic test study, with a cross-sectional design. The assessment of the validity of a test should be performed by comparing its results with the best test available for the same diagnostic purpose, considered the gold standard.[13] Due to the lack of a recognized gold standard test to identify AEs,[2-4,11,12] a reference standard method (RSM) was developed.[14-16]

The study was performed in a 500-bed, general, public, university hospital in Belo Horizonte, Brazil, offering tertiary and quaternary care. It has a hybrid medical record, with both electronic and paper-based documentation. Patients aged 18 years or older who were hospitalized for more than 24 hours between October 4 and November 2 of 2016 were eligible. Patients whose records were incomplete or unavailable were excluded.

Sample size calculation considered a confidence level of 95%, margin of error of 5%, population size of 1,500 (historical monthly average of hospitalizations) and minimum expected proportion of AEs in the population of 20%, which was based on previous studies showing AEs rates between 7.2% and 27.0% of

hospitalizations.[3] This resulted in a sample of 212 admissions. An intentional random oversampling of about 25% was added due to potential losses leading to a final sample of 268 admissions.

Definitions

AE was defined as “an unintended physical injury resulting from or contributed to by medical care which requires additional monitoring, treatment or hospitalization, or which results in death”.^[5] AEs resulting from acts of omission that were clearly related to the occurrence of harm to the patient were counted. These events include delays in the provision of health services, such as carrying out diagnosis or treatments, due to organizational issues.

All AEs were classified according to the nature and harm. The IHI-GTT adapted version of the National Coordinating Council for Medication Error Reporting and Prevention (NCC MERP) Index for Categorizing Errors was used to classify AE-related harm into five categories: (E) temporary harm and required intervention; (F) temporary harm and required initial or prolonged hospitalization; (G) permanent harm; (H) intervention required to sustain life; (I) death.^[5,17]

Although the GTT allows the identification of AEs during the entire period of hospitalization, the RSM analysis was restricted to the cross-sectional period. For this reason, were counted only AEs that occurred in this period, regardless of the beginning or end of hospitalization, or those prior to it, provided they were directly related to the current admission, such as those that caused temporary or permanent harm, requiring new treatments or interventions.

Reference standard test

Some strategies for evaluating the performance of a test when the gold standard is imperfect or non-existent were described.^[14-16] Previous studies showed that different AEs are identified by different methods and that a combination of methods achieves better sensitivity.^[6-10] Therefore, in this study the researchers created a composite RSM that included elements of different strategies for AE identification, combining the collection of information

during hospitalizations and retrospective analysis of clinical and administrative data.

The RSM was performed in two stages: 1) search for screening criteria that suggested the occurrence of AEs; 2) confirmation of the possible AE. The screening criteria sought in stage one covered different sources and collection methods and are described in box 1.

Box 1 – Different sources and collection methods of screening criteria included in the reference standard method

a) Semi-structured interviews with healthcare professionals

Professionals who provide direct care to patients, mainly nurses, were interviewed at the end of each shift between October 4 and November 2 of 2016 (cross-sectional period), including weekends and night shifts. Using a semi-structured script, the interviewers asked about situations that could indicate the occurrence of an AE. These screening criteria were selected according to the particularities of the sectors visited and are described in supplementary material 1. Professionals were asked about, for example, the occurrence of adverse drug reactions, invasive procedures and possible complications, falls or pressure injuries, problems with vascular access or other medical devices, the need for evaluation by the medical shift or the occurrence of delays in the provision of care, among others.

The group of interviewers consisted of 18 undergraduate students from three courses in the health area — nursing, speech therapy and medicine, who took turns visiting the institution's adult patient care sectors. They were supervised by three medical professionals and received 10-hour training that included theoretical topics and practical training.

This stage of data collection was supported by the hospital's administrative and clinical leadership, which motivated health professionals about the importance of identifying gaps in healthcare to promote improvements to patients.

b) Review of prescriptions

Daily, during the cross-sectional period, a group of 12 undergraduate pharmacy students, trained and supervised by a senior pharmacist, reviewed the prescriptions and notes of the nursing staff of patients who were in the emergency department or hospitalized in the wards and intensive therapies in search of information that would suggest the occurrence of adverse effects or medication errors.

c) Results of laboratory tests

Retrospective search in the electronic system of the clinical analysis laboratory for altered test results, selected for their potential to reflect organic changes resulting from the occurrence of an AE, such as increased creatinine levels, changes in sodium levels and elevations in coagulation tests (supplementary material 2).

d) Request to start antibiotics and data on healthcare-associated infections

Evaluation of official reports provided by the Hospital Infection Control Committee on healthcare-associated infections that occurred during the cross-sectional period, associated with the review of all requests for authorization to start antibiotics, a fundamental condition for dispensing any antimicrobial by the institution's pharmacy.

e) Review of operative notes of surgical and obstetric procedures

The operative notes of procedures that occurred during the cross-sectional period were reviewed looking for information that suggested the occurrence of AEs. Screening criteria included change in proposed

procedure, duration of procedure exceeding the expected time, description of complications during the procedure, blood transfusion requirement, administration of selected medications or any other evidence of an incident of care. In addition, administrative data were evaluated to identify cancellations or early surgical reoperation.

e) Transfusion Agency Reports

Information provided by the local Transfusion Agency on the occurrence of AEs related to the use of blood components and blood products during the cross-sectional period.

g) Voluntary reporting

Review of voluntary notifications registered in the institution's electronic notification system regarding hospitalizations that occurred in the cross-sectional period.

h) Early hospital readmission

Retrospective evaluation of the causes of early hospital readmission, that is, a hospital admission that occurred within 30 days of the last discharge. The information was obtained from administrative data.

h) Review of the cause of death

Retrospective evaluation of the causes of deaths that occurred during the cross-sectional period through the review of medical records.

j) Findings in medical records during the collection or analysis of other data

Any clear evidence of the occurrence of an AE noted by the researchers while handling the record to collect data on the patient or to verify other possible AEs identified in step one of the reference standard method. Does not include AEs found through GTT.

AE: Adverse event. GTT: Global Trigger Tool

Subsequently, in stage 2, physicians reviewed charts from admissions that were positive for at least one screening criterion, looking for additional information to validate the occurrence of harm to the patient related to healthcare. After confirming an AE, they rated it on the harm categorization. The team of physicians was composed of two Internal Medicine specialists with more than five years of experience. They were trained on the concepts and classifications used in the method and could consult each other or a medical specialist.

To reduce the subjectivity of the medical reviewer's judgment in determining the occurrence of an AE, a four-point scale (0-3) was used to determine the level of confidence that the harm could be attributed to healthcare, rather than the patient's disease process. Only those classified as 2-3 (moderate to certain evidence that the harm could be attributed to the healthcare) were considered AEs. Questions adapted from Baker and collaborators [18] and Mendes and collaborators [19] were used to support the medical reviewer's judgment.

GTT test

At the time of the study, there was no official version of the GTT in Portuguese. The tool translation and adaptation was based on the original GTT whitepaper and a Portuguese version of it translated by the IHI Latin American team. No substantial changes were made to general content of the original version. This process was described elsewhere.[20]

The review team consisted of a pair of primary reviewers composed of medical students of the fourth and fifth years, and by two senior physician specialists in Internal Medicine, who alternated, depending on availability, to validate the findings of the primary reviewers. All reviewers were familiar with the medical record and the model of care offered by the institution. The qualification of the reviewers included the individual reading of the GTT whitepaper, followed by a 12-hour theoretical-practical training, focused on AE concept understanding, meaning of each trigger and harm categorization. Ten medical records were evaluated by all reviewers, including five IHI's commented examples and five institution's medical records.

The GTT was applied following the IHI protocol. In a first step, each primary reviewer was allowed a maximum of 20 minutes per record to search for triggers, identify possible AEs, classify them according to the harm category and record the findings. Subsequently, the pair of primary reviewers discussed individual findings and recorded the consensus. In the second step, medical reviewer, based on the findings of the primary reviewers, confirmed the AEs and rectified harm categories when necessary. Medical reviewers could consult medical records for clarification and had the support of a group of specialist physicians, who were consulted in complex situations, when the attribution of harm to healthcare was not evident. Similar to the RSM, medical reviewers applied the four-point scale to determine the level of confidence that harm could be attributed to health care and only those rated 2-3 were considered AEs.

Analysis and statistics

The sample was characterized using absolute and relative frequencies for qualitative variables and measures of central tendency and dispersion for quantitative variables. McNemar's test was used to compare the performance of the GTT and the RSM to identify AEs of different harm categories. The frequency of occurrence of AEs was presented as the number of AEs per 1,000 patient-days.

The test results were considered dichotomous and the calculation of accuracy was performed using 2x2 tables. The validity of the GTT was assessed using the classic measures of accuracy – sensitivity, specificity and global accuracy.[14,15,21]

The results were demonstrated using two different study units: 1) per patient admission, in which the occurrence of one or more AEs in a hospitalization counted as a single positive case, and 2) per AE, in which each AE identified in an admission counted as a positive case. In both, patient admissions in which no AE was detected were counted as a negative case each. To verify the existence of possible differences in the ability of the GTT to identify AEs of different harm categories, the validity of the tool was evaluated for AEs in

general (E-I) and for the subgroup of events of greater harm, classified as F, G, H or I (F-I).

The software used for the analysis was R (version 3.5.0).

Results

Preliminarily, a total of 1,172 admissions were considered eligible. Of the 268 records selected, 49 could not be accessed due to remote storage of large records and their use for other purposes, such as medical care, audits or billing. Of the 219 available records, 20 were deemed as not meeting the eligibility criteria. An extra round of randomization selected 12 additional admissions leading to 211 admissions.

Among the 211 patients in the final sample, 134 (63.5%) were women, 63 (29.9%) were ≥ 60 years old, 172 (81.5%) have an urgent hospitalization and 47 (22.3%) were classified as Charlson Comorbidity Index ≥ 3 (considered high or very high). Regarding the reason for hospitalization, 90 (42.7%) were surgical, 76 (36.0%) clinical and 45 (21.3%) obstetric. The mean length of stay was 12.2 days (SD 18.6) and 1,393 patient-days were evaluated in the study.

There were identified 627 occurrences matching the screening criteria by the RSM, of which 274 (43.7%) were considered AEs. One AE could be related to one or more screening criteria. The most frequent sources of information were the semi-structured interviews with healthcare professionals (n=357), followed by requests to start antibiotics (n=85). The frequency of occurrence and positivity for AEs of the screening criteria identified using the RSM by type of source is described in Table 1.

Table 1 – Frequency of occurrence of the reference standard method screening criteria by type of source and percentage of them that were confirmed as adverse event

Type of source on the screening criteria	Total	Confirmed as adverse events (%)
Semi-structured interviews with healthcare professionals	357	165 (46.2)
Request to start antibiotics	85	29 (34.1)
Results of laboratory tests	49	11 (22.5)
Review of operative notes	34	9 (26.5)
Review of prescriptions	33	15 (45.5)
Findings in medical records during the collection or analysis of other data	21	21 (100.0)
Voluntary reporting	21	11 (52.4)
Early hospital readmission	11	3 (27.3)
Review of obstetric notes	7	2 (28.6)
Report on healthcare-associated infections	6	6 (100.0)
Transfusion Agency Reports	2	2 (100.0)
Review of the cause of death	1	0 (0.0)
Total	627	274 (43.7)

A total of 176 AEs were identified in 67 admissions using the RSM and 129 AEs in 76 admissions using the GTT, resulting in rates of 126 and 93 AEs/1,000 patient-days, respectively. Seventy-two AEs were identified by both methods. There is no significant difference between the methods in the identification of AEs in general ($p=0.10$). However, when analyzing subgroups, more AEs categorized as harm “E” were identified using the RSM ($p<0.005$), while the GTT was superior in identifying AEs of greater harm (F-I) ($p=0.02$) (Table 2). No AE resulted in patient death.

Table 2 - Adverse events by method of identification and category of harm

Adverse events by category of harm*	Reference standard	GTT	P-value**
E	124	64	<0.005
F	42	49	0.11
G	3	8	--
H	7	8	--
I	0	0	--
All category of harm (E-I)	176	129	0.10
Greater harm (F-I)	52	65	0.02

*Category of harm by the NCC MERP-IHI adapted version: (E) temporary harm and required intervention; (F) temporary harm and required initial or prolonged hospitalization; (G) permanent harm; (H) intervention required to sustain life; (I) death.

** McNemar Test.

Table 3 shows the frequency of AEs identified by nature using the two methods, considering all AEs and the subgroup of AEs with greater harm (F-I). In general, the most frequent AE were those related to peripheral venous access, medication, surgical/anesthetics, infections and delays in the provision of health services. None of the 16 AEs related to gastric/enteric catheters were identified through the GTT and only two of the 67 (3%) AEs related to peripheral venous access were evidenced by the index test. All AEs categorized in these two types were classified as harm "E" and mostly referred to accidental removal.

Table 3 – Frequency of occurrence of adverse events by nature and by method of identification in absolute number and in percentage considering all category of harm (E-I) and the subgroup of events of greater harm (F-I)

Adverse events by nature	All (E-I)		Greater harm (F-I)	
	Reference standard (%)	GTT-IHI (%)	Reference standard (%)	GTT-IHI (%)
Peripheral venous accesses	66 (37.5)	2 (1.6)	0 (0.0)	0 (0.0)
Medication	29 (16.5)	46 (35.7)	10 (19.2)	12 (18.5)
Surgical / anesthetics	19 (10.8)	27 (20.9)	13 (25.0)	21 (32.3)
Infections	19 (10.8)	22 (17.1)	13 (25.0)	13 (20.0)
Delays in the provision of health services	12 (6.8)	12 (9.3)	12 (23.1)	12 (18.5)
Gastric/enteric catheters	16 (9.1)	0 (0.0)	0 (0.0)	0 (0.0)
Transfusion of blood products	2 (1.1)	6 (4.7)	0 (0.0)	1 (1.5)
Phlebitis	3 (1.7)	2 (1.6)	0 (0.0)	0 (0.0)
Pressure injury	1 (0.6)	3 (2.3)	0 (0.0)	0 (0.0)
Radiotherapy	1 (0.6)	3 (2.3)	1 (1.9)	3 (4.6)
Airway	2 (1.1)	0 (0.0)	2 (3.9)	0 (0.0)
Central vascular accesses	2 (1.1)	0 (0.0)	0 (0.0)	0 (0.0)
Bladder catheter	0 (0.0)	2 (1.6)	0 (0.0)	2 (3.1)
Obstetric care	1 (0.6)	1 (0.8)	1 (1.9)	1 (1.5)
Fall	1 (0.6)	1 (0.8)	0 (0.0)	0 (0.0)
Transplants	1 (0.6)	1 (0.8)	0 (0.0)	0 (0.0)
Dialysis therapy	0 (0.0)	1 (0.8)	0 (0.0)	0 (0.0)
Skin injury from mechanical restraint	1 (0.6)	0 (0.0)	0 (0.0)	0 (0.0)
Total	176	129	52	65

The results of the accuracy of the GTT compared to the RSM are shown in Table 4. Considering only the subgroup of AEs of greater harm (F-I), the sensitivity, specificity and global accuracy data were, respectively, 0.90 (95% confidence interval [CI] 0.77;0.97), 0.90 (95%CI 0.84;0.94) and 0.90 (95%CI 0.85;0.94) for the study unit "per patient admission" and 0.85 (95%CI 0.72;0.93), 0.88 (95%CI 0.82;0.92) and 0.87 (95%CI 0.82;0.91) for the study unit "per EA". When all AEs were evaluated (E-I), the validity estimates were significantly lower for the study unit "per AE", with a sensitivity of 0.41 (95%CI

0.34;0.49), specificity of 0.68 (95%CI 0.60;0.74) and global accuracy of 0.54 (95%CI 0.49;0.60).

Table 1 - Sensitivity, specificity and global accuracy of the GTT in relation to the reference standard test for the identification of adverse events considering all category of harm (E-I) and the subgroup of events of greater harm (F-I)

	"Per patiente admission" study unit		"Per adverse event" study unity	
	All (E-I) n=211	Greater harm (F-I) n=211	All (E-I) n=352	Greater harm (F-I) n=226
Sensitivity *	0.76 (0.64;0.86)	0.90 (0.77;0.97)	0.41 (0.34;0.49)	0.85 (0.72;0.93)
Specificity *	0.83 (0.75;0.88)	0.90 (0.84;0.94)	0.68 (0.60;0.74)	0.88 (0.82;0.92)
Global accuracy *	0.81 (0.75;0.86)	0.90 (0.85;0.94)	0.54 (0.49;0.60)	0.87 (0.82;0.91)

*95% confidence interval.

Discussion

Statement of principal findings

The GTT showed satisfactory sensitivity, specificity and global accuracy for detection of AEs in inpatients when compared to a reference method that combined real-time screening and retrospective search for screening criteria and AEs in medical records and administrative data. The GTT accuracy was higher when AEs with minor harm were not counted.

Interpretation within the context of the wider literature

Although widely used by individual institutions and some national patient safety programs, there is no clear data on its accuracy.[3,11,12] The scarcity of studies that address the accuracy of the GTT and other methods for identifying AEs is justified by the difficulty in establishing a gold standard test.[2-4,11,12] Classen and collaborators compared the results obtained through the GTT with those obtained through a detailed retrospective review of medical records and administrative data and found a sensitivity of 94.9% and specificity of 100% for identifying admissions with AEs.[7] However, the reference standard test used has similar limitations to the GTT, which may have resulted in overestimated accuracy rates.

An important concern about the accuracy of the GTT and other methods that are based on retrospective review of medical records refers to their dependence on the quality of notes and the possible lack of recording of AEs, especially those that caused minor harm to the patient.[22] To assess the supposed impact of this on the GTT accuracy, it was essential that the RSM overcome this limitation and, for this reason, the interview with professionals was included as a search strategy.[23] The screening criteria identified through this source led to the confirmation of 127 of the 176 AEs of the RSM (72.2%), and they were the only source for the identification of 103 AEs (87 categorized as harm “E”).

Considering the study unit "per AE", there is a significant difference in the GTT accuracy measures when evaluating all AEs in relation to the subgroup of AEs with greater harm. Of the 124 AEs categorized as “E” identified by the RSM, 96 were not identified through the GTT. Of these, notes were found in the medical records of 85; 71 of which in nursing notes, part of the record not prioritized in the GTT review process.[5,24,25] This loss of sensitivity for minor AEs has already been reported by other researchers and can be justified by issues inherent to the GTT methodology, which restricts the time for reviewing medical records and recommends that it is not meant to be read “from the first page to the last page”, [2,5] more than due to the low quality of the records.[23]

In this context, it is important to note that this study was conducted in a reference university hospital, which has been striving to achieve quality standards in care. These characteristics may have influenced the quality of the medical records and, consequently, the GTT accuracy measures. In addition, during the cross-sectional period, professionals were approached on the topic of patient safety and this may have improved the quality of notes regarding AEs.

Analyzing only the events identified through the GTT, as in other studies,[6-9,11] the most common natures of AEs were those related to medication, surgery/anesthetics and infections. The failure to identify a large number of AEs related to nursing care through the GTT method, including those related to peripheral venous access and gastric/enteric catheters, corroborates the

impression reported by experienced reviewers that the GTT is primarily focused on harm related to care performed by physicians.[24]

Strengths and limitations

The combination of strategies for identifying AEs in the RSM was a strength of the research and aimed to overcome the individual weaknesses of each method, especially regarding underreporting, lack of adequate registration in medical records and divergences in administrative data.[23] Despite the efforts made by the researchers to create a method “almost perfect”, the reference standard test used in this study is suboptimal. There were identified situations in which the GTT was correct and the RSM was wrong regarding the occurrence of AEs. This bias led to an inadequate reduction in agreement in 2x2 tables and results in underestimated accuracy measures.[13-15]

Precautions were taken to reduce the chance of bias. The index test and the RSM were applied independently by groups of exclusive reviewers, ensuring blinding of the results, and both used the same definitions and classifications. Although the RSM includes data from medical record as a source of information, a systematic search for triggers as in the GTT was not used.

Although primary reviewers with little experience on GTT were employed, we do not consider this as a limitation. Studies have shown that reviewers' training and experience increase the reliability between them,[26,27] which can also impact the validity of the tool. However, the inter-examiner reliability of this pair of primary reviewers was previously described with substantial reliability results for the identification of hospitalizations with AEs in relation to a pair of experienced nurses,[20] which is comparable to other studies that evaluated experienced reviewers.[2,25,26]

Implications for policy, practice and research

The results of this study reinforce the validity of the GTT for identification of AEs of greater harm, but, on the other hand, they emphasize its inability for minor harm AEs identification. These events, although individually causing less impact

to the patient, become relevant due to the frequency they occur, consuming significant staff time and costly resources.[28] In this study, AEs related to peripheral venous accesses and gastric/enteric catheters together correspond to 35.6% of all AEs identified and, therefore, should be seen as indicators of poor quality care and target of improvement actions, which includes appropriate measurement strategies.[28-30]

Although it was not the objective of the study, the analysis of the AEs identified by the different sources in the RSM allows inferences about possible complementary methods to identify certain groups of AEs that are usually missed by the GTT, such as the semi-structured interview with the professionals who provide direct patient care. This method was used as a strategy to identify AEs related to feeding tube complications [29] and can be an alternative to direct observation which is not suitable for global assessment of adverse events, with possible gains in safety culture through greater involvement of direct care professionals.[23] However, further studies are needed on the potential of this method to identify different types of AEs, sampling strategies, possible biases and costs involved.

Conclusions

The GTT proved to be a valid method for identifying AEs in hospitalized adult patients. Its accuracy increases when minor harm AEs are not counted. Among the main adverse events missed by the GTT are those related to nursing care. Therefore, it should be used in combination with other measurement strategies to achieve results that are representative of the quality profile of the care provided and, thus, guide the best improvement strategies.

Ethics

The research was approved by the Local Research Ethics Committee (CAAE:52532715.7.0000.5149).

Funding

The authors received no specific funding for this work.

Conflict of interests

The lead author acted as a volunteer as a Global Chapter Leader at the IHI Open School between September 2017 and August 2019 and as one of the authors of the Portuguese translation of the White Paper "IHI Global Trigger Tool for Measuring Adverse Events" in 2019.

Acknowledgement

We are grateful to all the volunteers, who dedicated their time in all phases of data collection in this study, and to the IHI Latin America team and the Hospital leaders for their support.

Data sharing statement

The data underlying this article will be shared on reasonable request to the corresponding author.

References

1. James JT. A new, evidence-based estimate of patient harms associated with hospital care. *J Patient Saf.* 2013;9(3):122-128.
2. Mattsson TO, Knudsen JL, Lauritsen J, et al. Assessment of the global trigger tool to measure, monitor and evaluate patient safety in cancer patients: reliability concerns are raised. *BMJ Qual Saf.* 2013 Jul;22(7):571-9.
3. Hanskamp-Sebregts M, Zegers M, Vincent C, et al. Measurement of patient safety: a systematic review of the reliability and validity of adverse event detection with record review. *BMJ Open* 2016;6:e011078.
4. Classen DC, Lloyd RC, Provost L, et al. Development and Evaluation of the Institute for Healthcare Improvement Global Trigger Tool. *J Patient Saf.* 2008; 4(3):169-177.
5. Griffin FA, Resar RK. IHI Global Trigger Tool for Measuring Adverse Events. IHI Innovation Series. 2nd ed. Cambridge, MA: Institute for Healthcare Improvement; 2009.
6. Naessens JM, Campbell CR, Huddleston JM, et al. A comparison of hospital adverse events identified by three widely used detection methods. *Int J Qual Health Care* 2009;21:301–7.
7. Classen DC, Resar R, Griffin F, et al. 'Global trigger tool' shows that adverse events in hospitals may be ten times greater than previously measured.

- Health Aff (Millwood). 2011 Apr;30(4):581-9. Erratum in: Health Aff (Millwood). 2011 Jun;30(6):1217.
8. Kennerly DA, Kudyakov R, da Graca B, et al. Characterization of adverse events detected in a large health care delivery system using an enhanced global trigger tool over a five-year interval. *Health Serv Res*. 2014;49: 1407–1425.
 9. Rutberg H, Borgstedt Risberg M, Sjodahl R, et al. Characterisations of adverse events detected in a university hospital: a 4-year study using the Global Trigger Tool method. *BMJ open*. 2014;4:e004879.
 10. Mull HJ, Brennan CW, Folkes T, et al. Identifying previously undetected harm: piloting the Institute for Healthcare Improvement's Global Trigger Tool in the Veterans Health Administration. *Qual Manag Health Care*. 2015;24:140–146.
 11. Hibbert PD, Molloy CJ, Hooper TD, et al. The application of the Global Trigger Tool: a systematic review. *Int J Qual Health Care*. 2016 Dec 1;28(6):640-649.
 12. Klein DO, Rennenberg RJMW, Koopmans RP, et al. A Systematic Review of Methods for Medical Record Analysis to Detect Adverse Events in Hospitalized Patients. *J Patient Saf*. 2020 Mar 11.
 13. Hulley SB et al. *Designing clinical research*, 4th Edition. 4th Edition. Lippincott Williams & Wilkins, 2013. 367p.
 14. Rutjes AWS, Reitsma JB, Coomarasamy A, et al. Evaluation of diagnostic tests when there is no gold standard. A review of methods. *Health Technol Assess* 2007;11(50).
 15. Trikalinos TA, Balion CM. Chapter 9: options for summarizing medical test performance in the absence of a "gold standard". *J Gen Intern Med*. 2012 Jun;27 Suppl 1(Suppl 1):S67-75.
 16. Umemneku Chikere CM, Wilson K, Graziadio S, et al. Diagnostic test evaluation methodology: A systematic review of methods employed to evaluate diagnostic tests in the absence of gold standard - An update. *PLoS One*. 2019 Oct 11;14(10):e0223832.
 17. National Coordinating Council for Medication Error Reporting and Prevention. NCC MERP Index for Categorizing Medication Errors Algorithm; 1996 [revised February 20, 2001]. NCC MERP; 2001.
 18. Baker GR, Norton PG, Flintoft V, et al. The Canadian Adverse Events Study: the incidence of adverse events among hospital patients in Canada. *CMAJ*. 2004 May 25;170(11):1678-86.
 19. Mendes W, Martins M, Rozenfeld S et al. The assessment of adverse events in hospitals in Brazil. *Int J Qual Health Care* 2009 Aug;21(4):279-84.

20. Moraes SM, Ferrari TCA, Figueiredo NMP, et al. Assessment of the reliability of the IHI Global Trigger Tool: new perspectives from a Brazilian study. *Int J Qual Health Care*. 2021 Mar 18;33(1):mzab039.
21. Knottnerus JA, Muris JW. Assessment of the accuracy of diagnostic tests: the cross-sectional study. *J Clin Epidemiol*. 2003 Nov;56(11):1118-28.
22. Wilson RM, Michel P, Olsen S, et al. WHO Patient Safety EMRO/AFRO Working Group. Patient safety in developing countries: retrospective estimation of scale and nature of harm to patients in hospital. *BMJ*. 2012 Mar 13;344:e832.
23. Michel P. Strengths and weaknesses of available methods for assessing the nature and scale of harm caused by the health system: Literature review. Geneva: WHO; 2004.
24. Schildmeijer K, Nilsson L, Perk J, et al. Strengths and weaknesses of working with the Global Trigger Tool method for retrospective record review: focus group interviews with team members. *BMJ Open*. 2013 Sep 24;3(9):e003131.
25. Sharek PJ, Parry G, Goldmann D, et al. Performance characteristics of a methodology to quantify adverse events over time in hospitalized patients. *Health Serv Res*. 2011 Apr;46(2):654-78.
26. Schildmeijer K, Nilsson L, Arestedt K, et al. Assessment of adverse events in medical care: lack of consistency between experienced teams using the global trigger tool. *BMJ Qual Saf*. 2012 Apr;21(4):307-14.
27. Naessens JM, O'Byrne TJ, Johnson MG, et al. Measuring hospital adverse events: assessing inter-rater reliability and trigger performance of the Global Trigger Tool. *Int J Qual Health Care*. 2010 Aug;22(4):266-74.
28. Galazzi A, Adamini I, Consonni D, Roselli P, Rancati D, Ghilardi G, Greco G, Salinaro G, Laquintana D. Accidental removal of devices in intensive care unit: An eight-year observational study. *Intensive Crit Care Nurs*. 2019 Oct;54:34-38.
29. Gimenes FRE, Baracioli FFLR, Medeiros AP, et al. Factors associated with mechanical device-related complications in tube fed patients: A multicenter prospective cohort study. *PLoS One*. 2020 Nov 19;15(11):e0241849.
30. Marsh N, Webster J, Mihala G, et al. Devices and dressings to secure peripheral venous catheters to prevent complications. *Cochrane Database Syst Rev*. 2015 Jun 12;(6):CD011070.

Supplementary material 1: Semi-structured scripts used by interviewers during visits to different care units in search of screening criteria for identifying adverse events during the first stage of the Standard Reference Method.

SURGERY CENTER

During surgery

- Bleeding more than expected.
- Burn (by cautery).
- Injury, repair or removal of organs.
- Surgery on the wrong side, on the wrong patient, wrong procedure.
- Object forgotten in the patient.
- Unplanned intraoperative radiography.
- Longer-than-expected surgeries.
- Change of proposed procedure.

Surgical context

- Reinterventions in recent surgeries.

Clinical/anesthetic complications

- Clinical complication during surgery (heart attack, stroke, other).
- Cardiorespiratory arrest.
- Intra or postoperative death.
- Intubation or reintubation or some type of ventilatory support in the anesthesia recovery room.
- Unplanned transfer to intensive care.
- Use of epinephrine, norepinephrine, naloxone or flumazenil.

Medicines/blood components

- Reactions.
- Medication errors.

Organizational

- Loss of biological samples.
- Cancellation of surgeries.
- Incomplete filling of administrative and clinical forms.
- Equipment failure.
- Resource for care not available (equipment, medication, intensive care bed).

Others

- Any other incident related to surgical center care.

MATERNITY**Obstetric care**

- Conversion to cesarean section.
- Third or fourth-degree tear during childbirth
- Instrumented delivery in normal delivery.
- Kristeller maneuver.
- Blood transfusion.
- Use of oxytocin and misoprostol outside the context of labor induction
- General anesthesia.
- Intercurrences with sedation/anesthesia during a procedure.
- Maternal death.
- Unplanned organ removal, injury or repair (includes hysterectomy).
- Any other obstetric complication.

General care

- Falls.
- Start of antibiotics (except prophylactic ones).
- Venous access: bruising at puncture sites, loss/obstruction, phlebitis and cellulite.
- Hypoglycemia (less than 60mg/dL) or hyperglycemia (greater than 300mg/dL).
- Clinical complications, evaluation by medical shift or transfer to intensive care.
- Unplanned return to the delivery or operating room.
- Invasive procedure (central catheter, dialysis catheter, chest drain).
- Any patient on dialysis (only if she was not previously on dialysis).
- Drowsiness or agitation, mechanical restraint required.
- Prolonged hospitalization (longer than 5 days).
- Patient with some degree of dependence.
- Intercurrences related to the use of devices such as probe, drain or tracheostomy.

Medicines/blood components

- Reactions.
- Medication errors

Organizational

- Loss of biological samples.
- Incomplete filling of administrative and clinical forms.
- Equipment failure.
- Resource for care not available (equipment, medication, intensive care bed).

Others

- Any other occurrence or situation that may indicate the occurrence of an incident related to health care.

ENDOSCOPIC PROCEDURES

- Intercurrences related to the procedure: bleeds, perforations, preparation-related (fasting hypoglycemia, dehydration from laxatives).
- Intercurrences related to sedation/anesthesia.
- Medication Errors and Adverse Drug Events.
 - Falls.
- Wrong patient, wrong procedure.
- Loss of biological sample.
- Cancellation of procedures.
- Equipment failure.
- Resource for care not available (equipment, medication, intensive care bed).
- Any other complication or incident of care.

EMERGENCY ROOM AND CLINICAL AND SURGICAL WARDS

General care

- Falls.
- New or worsening pressure injury.
- Hypoglycemia (less than 60 mg/dL) or hyperglycemia (>400mg/dL).
- Diarrhea or constipation.
- Initiation of dialysis during hospitalization.
- **Venous access**
 - Hematomas at puncture sites.
 - Phlebitis and cellulite.
 - Loss or obstruction of peripheral venous access requiring a new puncture.
 - Multiple attempts to obtain peripheral venous access and/or a long time without venous access, with delay in the administration of intravenous medicament.
 - Accidental loss or obstruction of central catheters.
- **Probes and drains**
 - Accidental removal/loss/displacement.
 - Obstruction.
 - Diet aspiration.
- **Tracheostomy**
 - Obstruction or semi-obstruction of the cannula.
 - Accidental removal/loss/displacement.
- **Neurological**
 - New neurological alterations.
 - Agitation or drowsiness.
 - Mechanical containment
- **Infection**
 - Diagnosis of a new infection or initiation of antibiotics.
- **Invasive procedures:**
 - Puncture of dialysis cateter or central venous access catheter.
 - Chest drain.
 - Unplanned surgical intervention.

Medicines/blood components

- Reactions.
- Medication erros.

Clinical complications

- Clinical complications, evaluation by medical shift or transfer to intensive care.
- Cardiorespiratory arrest or death.
- Intubation or initiation of Non-Invasive Ventilation.
- Use of noradrenaline, dobutamine, nitroglycerin or sodium nitroprusside.

Organizational

- Loss of biological samples.
- Equipment failure.
- Resource for care not available (equipment, medication, intensive care bed).

Others

- Any other occurrence or situation that may indicate the occurrence of an incident related to health care.

INTENSIVE THERAPY

Medicines/blood components

- Reactions.
- Medication errors.

Airway/mechanical ventilation

- Orotracheal intubation: indication, complications during the procedure, reintubation, pneumothorax.
- Airway-related complications: endotracheal tube obstruction, accidental extubation, pneumothorax, unplanned mechanical ventilation disconnection, bronchoaspiration, atelectasis with repercussions on saturation, infections.
- Tracheostomy care: accidental decannulation, obstruction, strictures, fistulas and airway perforations, complications related to the tracheostomy procedure.

Vascular accesses

- Complications of puncture of vascular accesses: hemorrhage, pneumothorax or hemothorax.
- Accidental loss, obstruction.
- Catheter-associated thrombosis.
- Infection, phlebitis.

Probes and drains

- Infection or trauma / obstruction / accidental removal.

General care/prophylaxis

- Thromboembolism.
- New or worsening pressure injury of pre-existing.
- Fall.
- Bleeding.
- Hypoglycemia.
- Cardiopulmonary arrest/death
- New neurological alterations, agitation or drowsiness, mechanical containment.

Surgeries

- Complications of surgery (bleeds, infections, clinical complications - heart attack, stroke, etc., any other complication during or after a surgical procedure that may be related to the procedure).
- Unplanned surgical re-approach.

Readmissions

- Readmission to intensive care.

Equipment

- Equipment failure
- Equipment unavailability

Others

- Any other occurrence or situation that may indicate the occurrence of an incident related to health care.

Supplementary material 2: List of laboratory tests used as screening criteria in the first stage of the Reference Standard Method.

Laboratory test	Normal reference values	Results considered as a positive screening criterion
Alanine Transaminase (ALT)	7 - 35 U/L (♀); 7 - 50 U/L (♂)	> 345 U/L
Aspartate Transferase (AST)	15 - 46 U/L	> 230 U/L
Amylase	30 - 110 U/L	> 330 U/L
Lipase	23 - 300 U/L	> 900 U/L
Creatinine	0.52 - 1.04 mg/dL (♀) e 0.66 – 1.25 mg/dL (♂)	Increase of more than 1mg/dL in relation to the previous exam OR result equal to or greater than twice the previous exam.
Urea	15 - 36 mg/dL (♀) e 19 - 43 mg/dL (♂)	> 170 mg/dL
Sodium	137 - 145 mmol/L	< 120 mmol/L OR > 160 mmol/L OR change of 10 mmol/L or more up or down in less than 24 hours.
Potassium	3.5 - 5.1 mmol/L	> 6.5 mmol/L
Glucose	Fasting normal range in adults: 55 – 99 mg/dL	< 60 e > 400 mg/dL
Hemoglobin	12 - 16 g/dL (♀); 13 – 17.5 g/dL (♂)	< 5 g/dL
Lactate	0.7 - 2.5 mmol/L (venous); < 1.2 mmol/L (arterial)	> 4 mmol/L
Arterial pH	7.35 - 7.45	< 7.15
Bicarbonate	22 - 30 mmol/L	< 12 mmol/L
Arterial pCo2	32 - 48 mmHg	> 60 mmHg
Arterial pO2	83 - 108 mmHg	> 120 mmHg
Activated partial thromboplastin time (aPTT)	26 - 30 seconds	> 120 seconds
International Normalized Ratio (INR)	In general, 2 – 3 or 2.5 - 3.5 in patients using warfarin	> 5
Toxin A and B (Clostridium difficile infection)	Negative	Positive
Digoxin	Therapeutic range 0.8 to 2.0 ng/mL	> 2.0 ng/mL
Carbamazepine	Therapeutic range 4 - 12 µg/mL	< 4 e > 12 µg/mL
Phenytoin	Therapeutic range 10 - 20 µg/mL	< 10 e > 20 µg/mL
Phenobarbital	Therapeutic range 15 - 40 µg/mL	< 15 e > 40 µg/mL
Valproic acid	Therapeutic range 50 - 120 µg/mL	< 50 e > 120 µg/mL
Serum levels of vancomycin	Therapeutic range 10.0 – 20.0 µg/mL	<10 e > 20 µg/mL

(♀): women; (♂): men; >: greater than; <: less than.

5 CONSIDERAÇÕES FINAIS

O objetivo deste estudo foi avaliar a validade do *GTT* como método de identificação de eventos adversos. Embora amplamente utilizado em instituições individuais e em alguns programas nacionais de segurança do paciente, não há dados claros sobre sua acurácia.^{31,32,33} Os resultados mostram que o *GTT*, quando comparado ao método padrão referência construído, apresenta sensibilidade, especificidade e acurácia global satisfatórias, especialmente para identificação de eventos adversos com maior gravidade de dano ao paciente.

A avaliação da validade de instrumentos diagnósticos sem padrão ouro bem estabelecido apresenta desafios metodológicos que podem tornar esse tipo de estudo pouco factível ou muito trabalhoso. No campo da segurança do paciente, as limitações para definição de um padrão ouro ou referência para a identificação de eventos adversos envolvem não apenas aspectos relacionados ao método, mas também questões conceituais.

A falta de uma terminologia comum em segurança do paciente é um obstáculo crítico para compartilhar e agregar dados. Apesar dos esforços para unificar a taxonomia em segurança do paciente,³⁴ não há consenso na literatura e os estudos tendem a apresentar divergências conceituais e classificações não padronizadas.³⁵ Isso torna a tarefa de comparar o desempenho de diferentes métodos ou do mesmo método aplicado a diferentes contextos ainda mais

³¹ Hanskamp-Sebregts M, Zegers M, Vincent C, et al. Measurement of patient safety: a systematic review of the reliability and validity of adverse event detection with record review. *BMJ Open* 2016;6:e011078.

³² Hibbert PD, Molloy CJ, Hooper TD, et al. The application of the Global Trigger Tool: a systematic review. *Int J Qual Health Care*. 2016 Dec 1;28(6):640-649.

³³ Klein DO, Rennenberg RJMW, Koopmans RP, et al. A Systematic Review of Methods for Medical Record Analysis to Detect Adverse Events in Hospitalized Patients. *J Patient Saf*. 2020 Mar 11.

³⁴ Conceptual framework for the International Classification for Patient Safety, Version 1.1: final technical report. Geneva: World Health Organization; 2009. Disponível em https://apps.who.int/iris/bitstream/handle/10665/70882/WHO_IER_PSP_2010.2_eng.pdf?sequence=1&isAllowed=y, acesso em 11 de maio de 2022.

³⁵ Forster AJ, Bernard B, Drösler SE, Gurevich Y, Harrison J, Januel JM, Romano PS, Southern DA, Sundararajan V, Quan H, Vanderloo SE, Pincus HA, Ghali WA. A World Health Organization field trial assessing a proposed ICD-11 framework for classifying patient safety events. *Int J Qual Health Care*. 2017 Aug 1;29(4):548-556.

desafiadora. Diante da ausência de um sistema de classificação de eventos adversos por natureza simples e de fácil interpretação, optamos por utilizar uma classificação própria, que permitisse a compreensão dos eventos adversos mais frequentemente identificados ou perdidos pelo *GTT*.

Além disso, a decisão final sobre o dano ser atribuído ao cuidado em saúde e não a evolução da doença fica, geralmente, a cargo de um revisor médico e está sujeita a vieses de interpretação e de registro. Assim como realizado em outros estudos, tentamos reduzir a subjetividade desse julgamento acrescentando a escala de confiança sobre a existência de um evento adverso e as perguntas para guiar o raciocínio do médico revisor.

Quanto ao processo de avaliação da validade propriamente dito, a construção de um método padrão referência é uma alternativa no delineamento de estudos de teste diagnóstico quando um padrão ouro não existe ou não está disponível.^{36,37,38} Entretanto, é muito provável que os pesquisadores tenham que lidar com testes imperfeitos e necessitem lançar mão de estratégias alternativas para analisar os dados obtidos, para além das medidas clássicas de acurácia.

Uma busca detalhada na literatura, que incluiu revisões sistemáticas sobre o *GTT*, resultou em apenas um estudo que apresentou dados de validade do método. Classen e colaboradores encontraram sensibilidade de 94,9% para detectar pacientes com pelo menos um evento adverso e especificidade de 100% para detectar pacientes sem eventos ao comparar os resultados do *GTT* com os achados de uma revisão retrospectiva detalhada de dados administrativos e de prontuário realizada por um grupo de especialistas.³⁹

³⁶ Reitsma JBR A. W. S.; Khan K. S.; Coomarasamy A.; Bossuyt P. M. A review of solutions for diagnostic accuracy studies with an imperfect or missing reference standard. *Journal of Clinical Epidemiology*. 2009; 62(8):797–806.

³⁷ Rutjes AWS, Reitsma JB, Coomarasamy A, Khan KS, Bossuyt PMM. Evaluation of diagnostic tests when there is no gold standard. A review of methods. *Health Technol Assess* 2007;11(50).

³⁸ Umehneku Chikere CM, Wilson K, Graziadio S, Vale L, Alen AJ. Diagnostic test evaluation methodology: A systematic review of methods employed to evaluate diagnostic tests in the absence of gold standard - An update. *PLoS One*. 2019 Oct 11;14(10):e0223832.

³⁹ Classen DC, Resar R, Griffin F, et al. 'Global trigger tool' shows that adverse events in hospitals may be ten times greater than previously measured. *Health Aff (Millwood)*. 2011 Apr;30(4):581-9. Erratum in: *Health Aff (Millwood)*. 2011 Jun;30(6):1217.

Nesse estudo, os pesquisadores buscaram agregar fontes de dados e construíram um padrão referência. Entretanto, a busca por informações sobre a possível ocorrência de eventos adversos se restringiu a revisão retrospectiva de documentos e dados, o que pode ter gerado resultados de sensibilidade e especificidade superestimados. Além disso, no estudo são apresentados os dados de validade apenas para o desfecho internações com ou sem eventos adversos, não sendo possível analisar as propriedades do método na identificação de cada evento adverso.

Um dos pontos fortes do presente estudo foi a elaboração de um método padrão referência abrangente, que combinou diferentes fontes reconhecidas para obtenção de informações sobre a ocorrência de dano relacionado ao cuidado em saúde, incluindo dados administrativos e clínicos agrupados, além de entrevistas com profissionais da assistência direta.⁴⁰ Essa estratégia visava superar as fragilidades individuais de cada método, principalmente quanto à subnotificação, falta de registro adequado em prontuário e divergências em dados administrativos.

Contudo, apesar dos esforços para criar um método “quase perfeito”, o teste padrão referência utilizado é subótimo. Foram identificadas situações nas quais o *GTT* estava correto e o método padrão referência estava errado quanto à ocorrência de eventos adversos. Isso gera redução inadequada na concordância em tabelas 2x2 e resulta em medidas de acurácia subestimadas.^{41,42,43}

Frente às falhas apresentadas pelo padrão referência construído e embasados na literatura e na opinião de especialistas em Estatística,⁴⁴ propôs-se a

⁴⁰ Michel P. Strengths and weaknesses of available methods for assessing the nature and scale of harm caused by the health system: Literature review. Geneva: WHO; 2004.

⁴¹ Hulley SB et al. Designing clinical research, 4th Edition. 4th Edition. Lippincott Williams & Wilkins, 2013. 367p.

⁴² Rutjes AWS, Reitsma JB, Coomarasamy A, et al. Evaluation of diagnostic tests when there is no gold standard. A review of methods. Health Technol Assess 2007;11(50).

⁴³ Trikalinos TA, Balion CM. Chapter 9: options for summarizing medical test performance in the absence of a "gold standard". J Gen Intern Med. 2012 Jun;27 Suppl 1(Suppl 1):S67-75.

⁴⁴ Trikalinos TA, Balion CM. Chapter 9: options for summarizing medical test performance in the absence of a "gold standard". J Gen Intern Med. 2012 Jun;27 Suppl 1(Suppl 1):S67-75.

avaliação da validade por meio de concordância e coeficiente de *Kappa* e pelo cálculo de “estimativas ingênuas” das medidas clássicas de acurácia. A opção por apresentar no artigo os dados de validade por meio de sensibilidade, especificidade e acurácia global e discutir o possível viés do método padrão referência, que poderia implicar em resultados subestimados, foi pela provável maior familiaridade com essas medidas por parte do público interessado – gestores e profissionais dos setores de segurança e qualidade dos serviços de saúde.

Concordância e coeficiente de *Kappa*, por se tratarem de medidas comumente empregadas na avaliação da reprodutibilidade, outra dimensão avaliada em estudos de teste diagnóstico,⁴⁵ poderiam gerar estranheza e acrescentar pouca informação no âmbito da lacuna científica que o presente estudo pretendia preencher. Apesar de não apresentados no artigo, os dados de prevalência, proporção de concordância e *Kappa* ajustado por prevalência foram calculados e estão apresentados na Tabela 1. Para a identificação dos eventos adversos com danos maiores ao paciente, categorizados como dano “F” até “I”, houve concordância forte entre os métodos. Por outro lado, quando todos os eventos adversos foram incluídos na análise, independentemente da classificação de dano, a concordância foi forte para identificação de internações com eventos adversos (unidade de estudo “por internação”), porém foi fraca para a identificação de cada evento adverso (unidade de estudo “por evento adverso”).

Tabela 2 - Prevalência dos eventos adversos, proporção de concordância e coeficiente de Kappa ajustado por prevalência do GTT e método padrão referência construído para todos os eventos adversos (E-I) e para o subgrupo de eventos mais graves (F-I)

Eventos adversos	Unidade de estudo “por internação”		Unidade de estudo “por evento adverso”	
	Todos (E-I) n=211	Maior dano (F-I) n=211	Todos (E-I) n=352	Maior dano (F-I) n=226
Prevalência	0,436	0,275	0,662	0,323
Proporção de concordância	0,806	0,900	0,543	0,872

⁴⁵ Hulley SB et al. Delineando a pesquisa Clínica: uma abordagem epidemiológica. 4ª edição. Artmed, 2015. 386p.

Coeficiente <i>Kappa</i> ajustado por prevalência*	0,61 (0,49;0,71)	0,80 (0,70;0,88)	0,09 (-0,02;0,19)	0,74 (0,64;0,83)
---	------------------	------------------	-------------------	------------------

* Intervalo de confiança de 95%.

Estes dados reforçam os resultados obtidos por meio das medidas de acurácia clássica, indicando que o *GTT* se apresenta como um método válido para identificação de internações com eventos adversos, com maior sensibilidade para identificação de eventos com danos mais graves ao paciente. Isso significa que, ao se avaliar cada evento adverso, o *GTT* tende a perder os incidentes de cuidado de menor dano para o paciente.

A menor capacidade de identificação de eventos categorizados como dano “E” por meio do *GTT* já foi relatada por outros pesquisadores.^{46,47} Apesar da má qualidade dos registros ser frequentemente citada como a maior limitação dos métodos baseados na revisão retrospectiva de dados,⁴⁸ os achados do presente estudo mostram que características inerentes ao método podem justificar, em grande parte, essas perdas. Dos 124 eventos adversos categorizados como dano “E” identificados pelo método padrão referência, 96 não foram identificados pelo *GTT*. Destes, foram encontradas anotações em prontuário de 85 eventos, sendo que em 71 deles os registros constavam em anotações de enfermagem.

O método *GTT* não prevê a identificação de todos os eventos adversos.⁴⁹ A limitação no tempo de revisão, assim como a estratégia de amostragem, favorece aspectos de viabilidade e aplicabilidade em detrimento de acurácia e reprodutibilidade, visando a redução dos custos e dos recursos envolvidos no uso da ferramenta.

⁴⁶ Griffin FA, Resar RK. IHI Global Trigger Tool for Measuring Adverse Events. IHI Innovation Series. 2nd ed. Cambridge, MA: Institute for Healthcare Improvement; 2009.

⁴⁷ Mattsson TO, Knudsen JL, Lauritsen J, et al. Assessment of the global trigger tool to measure, monitor and evaluate patient safety in cancer patients: reliability concerns are raised. *BMJ Qual Saf.* 2013 Jul;22(7):571-9.

⁴⁸ Michel P. Strengths and weaknesses of available methods for assessing the nature and scale of harm caused by the health system: Literature review. Geneva: WHO; 2004.

⁴⁹ Griffin FA, Resar RK. IHI Global Trigger Tool for Measuring Adverse Events. IHI Innovation Series. 2nd ed. Cambridge, MA: Institute for Healthcare Improvement; 2009.

Frente a limitação do tempo de 20 minutos imposta aos revisores primários para avaliação do prontuário em busca dos *triggers*, o *IHI* recomenda que os registros não sejam lidos “da primeira à última página”, priorizando a revisão dos códigos de diagnósticos e procedimentos, sumário de alta, resultados de exames, prescrições, notas de cirurgia e anestesia e, por último, notas da enfermagem e da equipe multidisciplinar.⁵⁰ Portanto, para internações com maior duração, dificilmente os revisores conseguirão revisar todas as notas de enfermagem, ou mesmo registros médicos e do restante da equipe multiprofissional, e as porções do prontuário que trazem dados agregados, como sumário de alta e codificações, descrevem, usualmente, apenas os eventos e circunstâncias de maior relevância para o paciente.

Apesar de não ter sido o objetivo do estudo, ao se avaliar as fontes de dados incluídas no método padrão referência e o perfil de eventos adversos identificados por cada uma delas, nota-se que muitos dos eventos adversos perdidos pelo *GTT* foram identificados por meio de entrevistas com os profissionais da assistência. Os rastreadores identificados por meio dessa fonte levaram a confirmação de 127 dos 176 eventos adversos do método padrão referência (72,2%), sendo que eles foram a única fonte para identificação de 103 (87 eventos adversos categorizados como dano “E”, 10 como “F”, 2 como “G” e 4 como “H”).

A entrevista com profissionais da assistência é um método bem estabelecido, especialmente útil para a identificação de incidentes de cuidado ocorridos fora das instituições de saúde e de eventos adversos com baixa rastreabilidade nos registros, como quedas ou eventos relacionados à anestesia.⁵¹ Foi utilizado como estratégia para identificar eventos adversos em diversos estudos e pode ser uma alternativa à observação direta, que não é factível para avaliação global de eventos adversos, com possíveis ganhos na cultura de segurança por

⁵⁰ Griffin FA, Resar RK. *IHI Global Trigger Tool for Measuring Adverse Events*. IHI Innovation Series. 2nd ed. Cambridge, MA: Institute for Healthcare Improvement; 2009.

⁵¹ Michel P. Strengths and weaknesses of available methods for assessing the nature and scale of harm caused by the health system: Literature review. Geneva: WHO; 2004.

meio de maior envolvimento dos profissionais da assistência direta.^{51,52} No entanto, são necessários mais estudos sobre o potencial desse método para identificar diferentes tipos de eventos adversos, estratégias de amostragem, possíveis vieses e custos envolvidos.

Ao se avaliar a natureza dos eventos perdidos pelo *GTT*, nota-se que muitos deles estão relacionados a cuidados de enfermagem, como perda ou remoção inadvertida de acessos venosos periféricos e cateteres gástricos e entéricos. A falha do *GTT* em detectar eventos relacionados ao cuidado de enfermagem corrobora os achados de Schildmeijer e colaboradores sobre a impressão relatada por revisores experientes de que o *GTT* é voltado principalmente para os danos ligados às ações realizadas pelos médicos.⁵³

Os eventos adversos relacionados ao cuidado de enfermagem, embora tendam a gerar menor dano ao paciente como evento individual, tornam-se relevantes devido à frequência em que ocorrem, consumindo tempo significativo da equipe e aumentando os custos da internação.⁵⁴ A inserção de um cateter venoso periférico é o procedimento invasivo mais frequentemente realizado em hospitais. Os acessos venosos geralmente falham antes que o tratamento intravenoso seja concluído, com taxa de falha ou reinicialização não programada entre 33% e 69%.⁵⁵ As causas de falha mais comumente identificadas são deslocamento parcial ou remoção acidental, flebite, oclusão, infiltração, vazamento e, raramente, infecção. As taxas de remoção acidental de cateteres gástricos/entéricos variam entre 10,2 e 73,9 remoções por 1.000

⁵² Gimenes FRE, Baracioli FFLR, Medeiros AP, et al. Factors associated with mechanical device-related complications in tube fed patients: A multicenter prospective cohort study. *PLoS One*. 2020 Nov 19;15(11):e0241849.

⁵³ Schildmeijer K, Nilsson L, Perk J, Arestedt K, Nilsson G. Strengths and weaknesses of working with the Global Trigger Tool method for retrospective record review: focus group interviews with team members. *BMJ Open*. 2013 Sep 24;3(9):e003131.

⁵⁴ Galazzi A, Adamini I, Consonni D, Roselli P, Rancati D, Ghilardi G, Greco G, Salinaro G, Laquintana D. Accidental removal of devices in intensive care unit: An eight-year observational study. *Intensive Crit Care Nurs*. 2019 Oct;54:34-38.

⁵⁵ Marsh N, Webster J, Mihala G, Rickard CM. Devices and dressings to secure peripheral venous catheters to prevent complications. *Cochrane Database of Systematic Reviews* 2015, Issue 6. Art. No.: CD011070.

dias de dispositivo em pacientes em terapia intensiva.^{56,57} Um estudo prospectivo multicêntrico realizado no Brasil com dados de 447 pacientes adultos hospitalizados que fizeram uso de cateteres gástricos/entéricos demonstrou 191 eventos adversos relacionados ao uso do dispositivo em 116 pacientes, sendo que 134 correspondiam a remoção inadvertida e 25 a obstrução do tubo, resultando, na maioria, em dano leve.⁵⁸ A instituição de medidas simples nos processos de cuidado, em geral, leva a redução significativa das taxas de incidentes.^{59,60}

Apesar de estudos direcionados para detecção desses tipos de eventos adversos indicarem altas taxas de ocorrência, os estudos mais abrangentes raramente os destacam individualmente. Kennerly e colaboradores publicaram dados sobre os resultados encontrados em cinco anos de uso do *GTT* em um sistema de saúde norte americano. Os eventos adversos associados ao uso de acesso venoso periférico corresponderam a 4,2% de todos os eventos identificados, todos categorizados como dano “E”.⁶¹ Em outro estudo, eventos adversos associados ao uso de cateteres foram categorizados como “*device complications*” e representaram 13% do total de eventos adversos identificados por uma combinação de métodos, incluindo o *GTT*, atrás apenas dos eventos adversos de natureza cirúrgica/anestésica.⁶²

⁵⁶ Carrión MI, Ayuso D, Marcos M, Paz Robles M, de la Cal MA, Alía I, Esteban A. Accidental removal of endotracheal and nasogastric tubes and intravascular catheters. *Crit Care Med*. 2000 Jan;28(1):63-6. doi: 10.1097/00003246-200001000-00010. PMID: 10667500.

⁵⁷ Galazzi A, Adamini I, Consonni D, Roselli P, Rancati D, Ghilardi G, Greco G, Salinaro G, Laquintana D. Accidental removal of devices in intensive care unit: An eight-year observational study. *Intensive Crit Care Nurs*. 2019 Oct;54:34-38. doi: 10.1016/j.iccn.2019.06.002. Epub 2019 Jun 22.

⁵⁸ Gimenes FRE, Baracioli FFLR, Medeiros AP, Prado PRD, Koepp J, Pereira MCA, Trivisani CB, Rabeh SAN, Souza FB, Miasso AI. Factors associated with mechanical device-related complications in tube fed patients: A multicenter prospective cohort study. *PLoS One*. 2020 Nov 19;15(11):e0241849.

⁵⁹ Motta APG, Rigobello MCG, Silveira RCCP, Gimenes FRE. Nasogastric/nasoenteric tube-related adverse events: an integrative review. *Rev Lat Am Enfermagem*. 2021 Jan 8;29:e3400. doi: 10.1590/1518-8345.3355.3400. PMID: 33439952; PMCID: PMC7798396.

⁶⁰ Carrión MI, Ayuso D, Marcos M, Paz Robles M, de la Cal MA, Alía I, Esteban A. Accidental removal of endotracheal and nasogastric tubes and intravascular catheters. *Crit Care Med*. 2000 Jan;28(1):63-6.

⁶¹ Kennerly DA, Kudyakov R, da Graca B, et al. Characterization of adverse events detected in a large health care delivery system using an enhanced global trigger tool over a five-year interval. *Health Serv Res*. 2014;49: 1407–1425.

⁶² Mull HJ, Brennan CW, Folkes T, et al. Identifying previously undetected harm: piloting the Institute for Healthcare Improvement's Global Trigger Tool in the Veterans Health Administration. *Qual Manag Health Care*. 2015;24:140–146.

No presente estudo, os eventos adversos relacionados a acessos venosos periféricos e cateteres gástricos/entéricos corresponderam juntos a 35,6% de todos os eventos adversos identificados pelo *GTT*, todos eles categorizados como dano “E” e a maioria descrita como remoção inadvertida ou perda. Considerando-se a alta prevalência, eventos dessas naturezas devem ser vistos como indicadores de assistência de má qualidade e ser alvo de ações de melhoria, o que inclui estratégias de mensuração adequadas.

Algumas medidas foram tomadas para reduzir o risco de viés durante o delineamento do estudo e sua execução. O *GTT* e o método padrão referência construído foram aplicados de forma independente por grupos de revisores exclusivos, garantindo o cegamento dos resultados, e em ambos foram utilizadas as mesmas definições e classificações. Embora o método padrão referência incluía dados do prontuário como fonte de informação, não foi utilizada uma busca sistemática de rastreadores como no *GTT*, evitando-se viés de incorporação.

Além disso, o grupo que utilizou o *GTT* seguiu os parâmetros propostos pelo *IHI*, descritos no manual original da ferramenta. As exceções estão relacionadas ao perfil da dupla de revisores primários e a inclusão de eventos adversos relacionados a atos de omissão. Ambas as situações foram acompanhadas pelos pesquisadores principais e provavelmente não impactaram nos resultados obtidos neste estudo.

O *IHI* descreve que o grupo de revisores primários deve ser formado por profissionais com formação clínica e conhecimento sobre o conteúdo e organização do prontuário da instituição, bem como sobre os processos de cuidado. São citados como possíveis componentes desse grupo profissionais enfermeiros, farmacêuticos e fisioterapeutas, com destaque para enfermeiros experientes como melhores revisores.⁶³ Além disso, estudos têm demonstrado

⁶³ Griffin FA, Resar RK. *IHI Global Trigger Tool for Measuring Adverse Events*. IHI Innovation Series. 2nd ed. Cambridge, MA: Institute for Healthcare Improvement; 2009.

que o treinamento e a experiência dos revisores aumentam a confiabilidade entre eles,^{64,65} o que também pode impactar na validade da ferramenta.

Na presente pesquisa, a dupla de revisores primários foi escolhida com base em fatores relacionados a disponibilidade e custos. Apesar de terem sido empregados acadêmicos de medicina com pouca experiência na aplicação do *GTT*, não consideramos isso como uma limitação. Como não foram encontrados dados na literatura sobre o desempenho desse perfil de revisores, a confiabilidade interexaminador da dupla foi avaliada, com concordância forte para a identificação de interações com eventos adversos em relação a uma dupla de enfermeiros experientes e razoável em relação aos achados dos revisores médicos,⁶⁶ o que é comparável a outros estudos nos quais foram avaliados revisores experientes.^{67,68,69}

A inclusão de eventos adversos relacionados a omissão foi uma estratégia utilizada também em outros estudos e por instituições que implementaram o *GTT* como ferramenta de medição em segurança do paciente^{70,71}. A opção por incluir esses eventos está relacionada ao seu potencial de gerar aprendizado

⁶⁴ Schildmeijer K, Nilsson L, Arestedt K, et al. Assessment of adverse events in medical care: lack of consistency between experienced teams using the global trigger tool. *BMJ Qual Saf.* 2012 Apr;21(4):307-14.

⁶⁵ Naessens JM, O'Byrne TJ, Johnson MG, et al. Measuring hospital adverse events: assessing inter-rater reliability and trigger performance of the Global Trigger Tool. *Int J Qual Health Care.* 2010 Aug;22(4):266-74.

⁶⁶ Moraes SM, Ferrari TCA, Figueiredo NMP, et al. Assessment of the reliability of the IHI Global Trigger Tool: new perspectives from a Brazilian study. *Int J Qual Health Care.* 2021 Mar 18;33(1):mzab039.

⁶⁷ Sharek PJ, Parry G, Goldmann D, et al. Performance characteristics of a methodology to quantify adverse events over time in hospitalized patients. *Health Serv Res.* 2011 Apr;46(2):654-78.

⁶⁸ Schildmeijer K, Nilsson L, Arestedt K, et al. Assessment of adverse events in medical care: lack of consistency between experienced teams using the global trigger tool. *BMJ Qual Saf.* 2012 Apr;21(4):307-14.

⁶⁹ Mattsson TO, Knudsen JL, Lauritsen J, et al. Assessment of the global trigger tool to measure, monitor and evaluate patient safety in cancer patients: reliability concerns are raised. *BMJ Qual Saf.* 2013 Jul;22(7):571-9.

⁷⁰ Hibbert PD, Molloy CJ, Hooper TD, Wiles LK, Runciman WB, Lachman P, Muething SE, Braithwaite J. The application of the Global Trigger Tool: a systematic review. *Int J Qual Health Care.* 2016 Dec 1;28(6):640-649.

⁷¹ Kennerly DA, Saldaña M, Kudryakov R, da Graca B, Nicewander D, Compton J. Description and evaluation of adaptations to the global trigger tool to enhance value to adverse event reduction efforts. *J Patient Saf.* 2013 Jun;9(2):87-95.

institucional, impulsionando melhoria nos cuidados.⁷² Embora o conceito de omissão aplicado a ocorrência de eventos adversos seja entendido como algo menos objetivo que os atos de comissão e, portanto, sujeito a maior grau de variabilidade entre diferentes avaliadores, a conceituação utilizada neste estudo seguiu diretrizes claras e foi aplicada igualmente em ambos os testes. Foram incluídos eventos adversos resultantes de atos de omissão quando era possível estabelecer relação direta entre a não oferta de um cuidado em saúde com a ocorrência de dano para o paciente. Como exemplos pode-se citar situações nas quais houve aumento no tempo da internação devido a indisponibilidade de um exame ou procedimento ou o estado de saúde do paciente agravou devido a questões de desabastecimento e falta de medicamentos no sistema público de saúde.

Este estudo não foi desenhado para avaliar prevalência e taxas de eventos adversos. Devido a fatores relacionados à metodologia aplicada ao teste padrão referência, que teve a análise restrita ao período de corte transversal, nem todos os dias das internações incluídas na amostra foram avaliados. Sendo assim, consideramos que seria adequado apresentar os resultados apenas como “número de eventos adversos/1.000 pacientes dias”. A não disponibilização de alguns prontuários para os pesquisadores pode ter impactado nesses dados, uma vez que a maioria deles se referia a internações prolongadas, nas quais frequentemente são encontrados um maior número de eventos adversos. Entretanto, entendemos que esse fato tem pouco ou nenhum impacto na avaliação da validade, uma vez que o mesmo universo de internações foi avaliado por ambos os métodos.

É importante destacar que esse estudo foi desenvolvido em um hospital de referência para pacientes de cuidado terciário e quaternário no estado de Minas Gerais, incluindo transplantes e oncohematologia. Trata-se de um hospital universitário, que tem se esforçado nos últimos anos para alcançar metas de qualidade internacionais, tanto no cuidado, quanto no registro em prontuário. Essas características, tanto relacionadas ao perfil de gravidade dos

⁷² de Vries EN, Ramrattan MA, Smorenburg SM et al. The incidence and nature of in-hospital adverse events: a systematic review. *Qual Saf Health Care* 2008;17:216–23.

pacientes, quanto ao modelo de cuidado da instituição, podem ter interferido nos resultados obtidos, inclusive com impacto na avaliação da validade da ferramenta.

Apesar de demandar recursos de tempo e profissionais treinados e experientes para revisão manual dos prontuários, o que implica em custos, o *GTT* é o método, entre os mais sensíveis, que requer menores esforços e que apresenta maior potencial de sustentabilidade a longo prazo.⁷³ Adler e colaboradores estimaram que o custo para a implementação do *GTT* em um sistema de saúde nos Estados Unidos foi de 710 dólares para o treinamento inicial da equipe de revisores e de 611 dólares ao mês por hospital para manutenção das atividades de revisão.⁷⁴ Nossa experiência em empregar acadêmicos de medicina de anos finais do curso como revisores primários mostrou-se uma alternativa de menor custo, com bons resultados de confiabilidade quando comparados a enfermeiros experientes.⁷⁵

Entretanto, a detecção adequada da incidência de eventos adversos corresponde apenas a uma pequena parcela das atividades para melhoria da qualidade dos cuidados. É fundamental que nos programas de segurança do paciente sejam criadas estratégias para transformar os dados em aprendizado e melhorar os cuidados e os resultados para os pacientes.⁷⁶ Qualquer avaliação de custo-efetividade deve considerar não só os custos de implementação da ferramenta, mas também os benefícios das atividades em melhoria da qualidade associadas.⁷⁷

⁷³ Naessens JM, O'Byrne TJ, Johnson MG, et al. Measuring hospital adverse events: assessing inter-rater reliability and trigger performance of the Global Trigger Tool. *Int J Qual Health Care*. 2010 Aug;22(4):266-74.

⁷⁴ Adler L, Denham CR, McKeever M, et al. Global trigger tool: Implementation basics. *J Patient Saf* 2008;4:245–9.

⁷⁵ Moraes SM, Ferrari TCA, Figueiredo NMP, et al. Assessment of the reliability of the IHI Global Trigger Tool: new perspectives from a Brazilian study. *Int J Qual Health Care*. 2021 Mar 18;33(1):mzab039.

⁷⁶ Klein DO, Renneberg RJMW, Koopmans RP, Prins MH. A Systematic Review of Methods for Medical Record Analysis to Detect Adverse Events in Hospitalized Patients. *J Patient Saf*. 2020 Mar 11.

⁷⁷ Mull HJ, Brennan CW, Folkes T, et al. Identifying previously undetected harm: piloting the Institute for Healthcare Improvement's Global Trigger Tool in the Veterans Health Administration. *Qual Manag Health Care*. 2015;24:140–146.

Embora o *GTT* tenha se mostrado válido para identificação de eventos adversos, em especial os de maior gravidade, outras propriedades que não foram abordadas neste estudo precisam ser consideradas na avaliação da aplicabilidade do método. Alguns estudiosos já demonstraram preocupação quanto ao impacto que a pequena amostra quinzenal e que a moderada confiabilidade intra e interexaminadores obtida na maioria dos estudos nos quais se avaliou a reprodutibilidade poderia ter nas conclusões das análises da evolução das taxas de eventos adversos, obtidas ao longo do tempo.^{78,79,80,81,82,83} Essas fragilidades metodológicas e de implementação colocam à prova o real valor dessa ferramenta como estratégia de medição em segurança do paciente e precisam ser mais bem abordadas em estudos futuros.

Para concluir, o *GTT* mostrou ser um método válido para identificação de eventos adversos em pacientes adultos internados. Contudo, ele não foi capaz de identificar todos os eventos adversos, deixando de detectar principalmente aqueles classificados como dano “E” e os relacionados a alguns cuidados de enfermagem. Por isso, não recomendamos que o *GTT* seja utilizado como método único de identificação de eventos adversos. Outras ferramentas de medição devem ser associadas a ele para alcançar resultados representativos do perfil de segurança e qualidade nos cuidados em saúde oferecidos e, assim, orientar as melhores estratégias para melhoria.

⁷⁸ Mevik K, Griffin FA, Hansen TE, Deilkås ET, Vonen B. Does increasing the size of bi-weekly samples of records influence results when using the Global Trigger Tool? An observational study of retrospective record reviews of two different sample sizes. *BMJ Open*. 2016 Apr 25;6(4):e010700.

⁷⁹ Mattsson TO, Knudsen JL, Lauritsen J, Brixen K, Herrstedt J. Assessment of the global trigger tool to measure, monitor and evaluate patient safety in cancer patients: reliability concerns are raised. *BMJ Qual Saf*. 2013 Jul;22(7):571-9.

⁸⁰ Landrigan CP, Parry GJ, Bones CB, Hackbarth AD, Goldmann DA, Sharek PJ. Temporal trends in rates of patient harm resulting from medical care. *N Engl J Med*. 2010 Nov 25;363(22):2124-34. Erratum in: *N Engl J Med*. 2010 Dec 23;363(26):2573.

⁸¹ Shojania KG, Thomas EJ. Trends in adverse events over time: why are we not improving? *BMJ Qual Saf*. 2013 Apr;22(4):273-7.

⁸² von Plessen C, Kodal AM, Anhøj J. Experiences with global trigger tool reviews in five Danish hospitals: an implementation study. *BMJ Open*. 2012 Oct 12;2(5):e001324.

⁸³ Schildmeijer K, Nilsson L, Arestedt K, Perk J. Assessment of adverse events in medical care: lack of consistency between experienced teams using the global trigger tool. *BMJ Qual Saf*. 2012 Apr;21(4):307-14.

As limitações inerentes aos diferentes métodos de identificação de eventos adversos desqualificam as métricas abrangentes de segurança do paciente, como as taxas gerais de eventos adversos geradas individualmente por cada um deles. Medidas personalizadas para danos específicos ao paciente, que valorizem a compreensão da ocorrência de eventos adversos por natureza, tem sido propostas para detectar o progresso ao longo do tempo e diferenças na segurança do cuidado oferecido entre as instituições.⁸⁴ Nesse sentido, desde que sejam seguidas definições e classificações padronizadas em segurança do paciente, o *GTT* pode representar uma das fontes de dados para a medida de eventos específicos de interesse.

⁸⁴ Shojania KG, Marang-van de Mheen PJ. Identifying adverse events: reflections on an imperfect gold standard after 20 years of patient safety research. *BMJ Qual Saf.* 2020 Apr;29(4):265-270.

6 ANEXOS

ANEXO A: PLANILHA ORIGINAL DO GTT-IHI

Cares Module Triggers		+ Event Description and Harm Category (E-I)		Medication Module Triggers		+ Event Description and Harm Category (E-I)	
C1	Transfusion or use of blood products			M1	<i>Clostridium difficile</i> positive stool		
C2	Code/arrest/rapid response team			M2	Partial thromboplastin time greater than 100 seconds		
C3	Acute dialysis			M3	International Normalized Ratio (INR) greater than 6		
C4	Positive blood culture			M4	Glucose less than 50 mg/dl		
C5	X-ray or Doppler studies for emboli or DVT			M5	Rising BUN or serum creatinine greater than 2 times baseline		
C6	Decrease of greater than 25% in hemoglobin or hematocrit			M6	Vitamin K administration		
C7	Patient fall			M7	Benadryl (Diphenhydramine) use		
C8	Pressure ulcers			M8	Romazicon (Flumazenil) use		
C9	Readmission within 30 days			M9	Naloxone (Narcan) use		
C10	Restraint use			M10	Anti-emetic use		
C11	Healthcare-associated infection			M11	Over-sedation/hypotension		
C12	In-hospital stroke			M12	Abrupt medication stop		
C13	Transfer to higher level of care			M13	Other		
C14	Any procedure complication						
C15	Other						
Surgical Module Triggers							
S1	Return to surgery			I1	Pneumonia onset		
S2	Change in procedure			I2	Readmission to intensive care		
S3	Admission to intensive care post-op			I3	In-unit procedure		
S4	Intubation/reintubation/BitPap in Post Anesthesia Care Unit (PACU)			I4	Intubation/reintubation		
S5	X-ray intra-op or in PACU						
S6	Intra-op or post-op death						
S7	Mechanical ventilation greater than 24 hours post-op						
S8	Intra-op epinephrine, norepinephrine, naloxone, or romazicon						
S9	Post-op troponin level greater than 1.5 ug/ml						
S10	Injury, repair, or removal of organ						
S11	Any operative complication						
Perinatal Module Triggers							
P1	Tetburaline use						
P2	3rd- or 4th-degree lacerations						
P3	Platelet count less than 50,000						
P4	Estimated blood loss > 500 ml (vaginal) or > 1,000 ml (C-section)						
P5	Specialty consult						
P6	Oxytocic agents						
P7	Instrumented delivery						
P8	General anesthesia						
Emergency Department Module Triggers							
E1	Readmission to ED within 48 hours						
E2	Time in ED greater than 6 hours						

Patient Identifier _____ Total Events _____ Total LOS _____ Write descriptions of the events in greater detail on reverse of Worksheet.

ANEXO B: QUESITOS PARA AVALIAÇÃO DO NÍVEL DE CONFIANÇA DE QUE FOI O CUIDADO PRESTADO QUE CAUSOU A LESÃO OU DANO AO PACIENTE

Causalidade
<p>Há relação temporal que indique que lesão ou dano foi consequência do cuidado? <input type="checkbox"/> Não <input type="checkbox"/> Sim</p> <p>Há alguma anotação no prontuário indicando que o cuidado prestado tenha causado a lesão ou dano?</p> <p><input type="checkbox"/> Não <input type="checkbox"/> Sim</p> <p>Há alguma anotação no prontuário sugerindo a possibilidade de lesão ou dano decorrente da doença de base do paciente? <input type="checkbox"/> Não <input type="checkbox"/> Sim</p> <p>O momento de ocorrência do evento sugere que a lesão ou dano está relacionada com o tratamento?</p> <p><input type="checkbox"/> Provável <input type="checkbox"/> Possível <input type="checkbox"/> Improvável</p> <p>O momento de ocorrência do evento sugere que a lesão ou dano está relacionada à ausência de tratamento?</p> <p><input type="checkbox"/> Provável <input type="checkbox"/> Possível <input type="checkbox"/> Improvável</p> <p>Existe alguma outra explicação plausível indicando a causa da lesão ou dano? <input type="checkbox"/> Não <input type="checkbox"/> Sim <input type="checkbox"/> Possível</p> <p>Antes da ocorrência da lesão ou dano, houve oportunidade para alguma intervenção que poderia tê-la evitada?</p> <p><input type="checkbox"/> Não <input type="checkbox"/> Sim <input type="checkbox"/> Possível</p> <p>A ausência ou a demora no tratamento é causa reconhecida desse tipo de lesão ou dano?</p> <p><input type="checkbox"/> Amplamente reconhecida <input type="checkbox"/> Reconhecida por outros especialistas <input type="checkbox"/> Não <input type="checkbox"/> Não se aplica</p> <p>A ausência ou a demora no diagnóstico é causa reconhecida desse tipo de lesão ou dano?</p> <p><input type="checkbox"/> Amplamente reconhecida <input type="checkbox"/> Reconhecida por outros especialistas <input type="checkbox"/> Não <input type="checkbox"/> Não se aplica</p> <p>O tratamento dado ao paciente é reconhecido como causa desse tipo de lesão ou dano?</p> <p><input type="checkbox"/> Amplamente reconhecida <input type="checkbox"/> Reconhecida por outros especialistas <input type="checkbox"/> Não <input type="checkbox"/> Não se aplica</p> <p>A lesão ou dano em questão é reconhecida como complicação esperada da doença de base do paciente?</p> <p><input type="checkbox"/> Amplamente reconhecida <input type="checkbox"/> Reconhecida por outros especialistas <input type="checkbox"/> Não <input type="checkbox"/> Não se aplica</p> <p>A lesão ou dano foi reconhecida durante a internação índice? <input type="checkbox"/> Não <input type="checkbox"/> Sim</p> <p>Se ‘Sim’, Adotaram-se as medidas apropriadas durante a internação índice?</p> <p><input type="checkbox"/> Não <input type="checkbox"/> Sim <input type="checkbox"/> Não foi necessária qualquer medida</p> <p>Se ‘Sim’, A lesão ou dano respondeu às medidas tomadas?</p> <p><input type="checkbox"/> Não <input type="checkbox"/> Sim <input type="checkbox"/> Possivelmente <input type="checkbox"/> Não se aplica</p> <p>>>> Após as devidas considerações dos detalhes clínicos do cuidado prestado ao paciente, independentemente das possibilidades de prevenção, e considerando as suas respostas às questões acima, responda o item “Relação de causa”.</p>

Adaptação e tradução do *Canadian Adverse Events Study I* (Baker et al., 2004) por Mendes e colaboradores (Mendes et al., 2009).

7 APÊNDICES

APÊNDICE A: FORMULÁRIO DE COLETA DE DADOS GERAIS DOS PACIENTES

	Número do paciente na pesquisa				
Formulário 1 - Informações gerais dos pacientes					
Iniciais do nome: _____	Prontuário: _____	Data de nascimento: ___/___/____			
Data da internação: ___/___/2016	Data da alta: ___/___/____	Tempo de internação: _____			
Dados da internação					
Internação prévia no HC nos últimos 3 meses: ___/0- Não ___/1- Sim Alta em: ___/___/2016					
Internação no PA: ___/0-Não ___/1-Sim Tempo: _____ dias					
Internação em CTI: ___/0-Não ___/1-Sim Tempo: _____ dias					
Dados clínicos do paciente					
Diagnóstico principal de entrada: _____					
Motivo da internação: _____					
Diagnóstico principal de saída: _____					
Situação da alta: ___/1-Alta ___/2-Óbito ___/3-Transferência para outra instituição					
Grau de suporte clínico:					
___/1-Completo ___/2-não explicitado ___/3-Alguma limitação de suporte ou não invasivo					
Grau de dependência (avaliação inicial do enfermeiro da terapia intensiva):					
___/1-Totalmente dependente ___/2-Parcialmente dependente ___/3-Independente ___/4- Informação ausente					
Dados indiretos de grau de dependência	Admissão: 1º registro	Admissão CTI (até 24h)	48-72h da adm. no CTI	1º dado após alta de CTI	Último registro antes alta/óbito
Data					
Escala de Morsen					
Queda últimos meses					
Diagnóstico secundário					
Terapia endovenosa					
Ajuda deambular					
Estado mental					
Marcha/transferência					
Escala de Braden					
Percepção sensorial					
Nutrição					
Mobilidade					
Atividade					
Umidade					
Fricção e cisalhamento					
Tabagismo: ___/0- Não ___/1- Sim ___/2- Não Informado					
Etilismo: ___/0- Não ___/1- Sim ___/2- Não Informado					
Histórico medicamentoso					
Paciente apresenta alguma alergia a medicamento: ___/0-Não ___/1-Sim ___/2-Não informado					
Descrever: _____					
Uso de medicamentos domiciliares: ___/0-Não ___/1-Sim ___/2- Não informado					
Anotar fármaco/posologia:					

Comorbidades à admissão	
<input type="checkbox"/>	Insuficiência Cardíaca congestiva
<input type="checkbox"/>	IAM prévio
<input type="checkbox"/>	Angina
<input type="checkbox"/>	Insuficiência cardíaca isquêmica
<input type="checkbox"/>	Arritmia
<input type="checkbox"/>	Doença valvar
<input type="checkbox"/>	Doença pulmonar circulatória
<input type="checkbox"/>	Doença vascular periférica
<input type="checkbox"/>	Hipertensão
<input type="checkbox"/>	Hemiplegia/paralisia
<input type="checkbox"/>	AVC ou AIT prévios
<input type="checkbox"/>	Epilepsia
<input type="checkbox"/>	Parkinson
<input type="checkbox"/>	Afasia
<input type="checkbox"/>	Outras desordens neurológicas
<input type="checkbox"/>	Usa oxigênio domiciliar
<input type="checkbox"/>	Doença pulmonar crônica
<input type="checkbox"/>	Diabetes SEM lesão de órgão alvo
<input type="checkbox"/>	Diabetes COM lesão de órgão alvo
<input type="checkbox"/>	Diabetes com doença vascular periférica***
<input type="checkbox"/>	Hipotireoidismo
<input type="checkbox"/>	Doença renal
<input type="checkbox"/>	Doença hepática sem varizes/hipertensão portal
<input type="checkbox"/>	Hepatites crônicas, cirrose biliar
<input type="checkbox"/>	Doença hepática com varizes/hipertensão portal
<input type="checkbox"/>	Doença péptica ulcerosa
<input type="checkbox"/>	Colite ulcerativa ou ulceração isolada
<input type="checkbox"/>	Sangramento de TGI prévio (ulceroso)
<input type="checkbox"/>	AIDS/HIV
<input type="checkbox"/>	Linfoma
<input type="checkbox"/>	Cancer metastático (tumor sólido metastático)
<input type="checkbox"/>	Tumor sólido sem metástase
<input type="checkbox"/>	<i>Lupus eritematoso sistêmico</i>
<input type="checkbox"/>	<i>Artrite reumatóide</i>
<input type="checkbox"/>	<i>Esclerose sistêmica</i>
<input type="checkbox"/>	<i>Polimialgia reumática</i>
<input type="checkbox"/>	Doença reumatológica (<i>que não classificada acima</i>)
<input type="checkbox"/>	Coagulopatia
<input type="checkbox"/>	Obesidade (sem especificação)
<input type="checkbox"/>	Perda peso
<input type="checkbox"/>	Hiponatremia ou hipernatremia
<input type="checkbox"/>	Hipopotassemia ou hiperpotassemia
<input type="checkbox"/>	Acidose ou alcalose
<input type="checkbox"/>	Desidratação
<input type="checkbox"/>	Hipovolemia
<input type="checkbox"/>	Outros distúrbios hidroeletrólíticos
<input type="checkbox"/>	Anemia ferropriva secundária a perda sangue (crônica)
<input type="checkbox"/>	Anemia por deficiência (dieta, não especificada)
<input type="checkbox"/>	Abuso do álcool
<input type="checkbox"/>	Abuso drogas
<input type="checkbox"/>	Psicoses
<input type="checkbox"/>	Demência
<input type="checkbox"/>	Depressão
<input type="checkbox"/>	Celulite ou úlcera pele
<input type="checkbox"/>	Usa varfarina

*** marcar também doença vascular periférica

Outras comorbidades: _____

Pesquisador: _____

APÊNDICE B: PROTOCOLO PARA FORMAÇÃO DE PADRÃO REFERÊNCIA — ROTEIRO PARA BUSCA ATIVA DE EVENTOS ADVERSOS NO BLOCO CIRÚRGICO

Roteiro – Busca de Incidentes SETOR BLOCO CIRÚRGICO
Rastreadores cirúrgicos
Ato cirúrgico
<ul style="list-style-type: none"> - Sangramentos além do previsto. - Queimadura (por cautério). - Lesão, reparo ou remoção de órgãos. - Cirurgia no lado errado, no paciente errado, procedimento errado. - Objeto esquecido no paciente. - Rx intra operatório não programado. - Cirurgias mais longas que o previsto. - Mudança de procedimento proposto.
Contexto cirúrgico
<ul style="list-style-type: none"> - Reintervenção em cirurgias recentes.
Complicações clínicas/anestésicas
<ul style="list-style-type: none"> - Complicação clínica durante cirurgia (IAM, AVC, outro). - PCR (parada cardiorrespiratória). - Morte intra ou pós-operatória. - Intubação ou reintubação ou algum tipo de suporte ventilatório na sala de recuperação anestésica. - Transferência para CTI, caso tenha sido não programada e previamente indicada. - Uso de epinefrina, noraepinefrina, naloxone ou flumazenil.
Medicamentos/hemocomponentes
<ul style="list-style-type: none"> - Reações à infusão de algum medicamento. - Reação durante ou logo após a infusão de hemocomponente. - Erros de medicação: falhas de prescrição/administração de medicamentos.
Organizacional
<ul style="list-style-type: none"> - Perda de material biológico coletado para exame. - Cirurgias canceladas. - Não preenchimento ou preenchimento incompleto da folha de sala. - Não preenchimento ou preenchimento incompleto do checklist da cirurgia segura. - Falha de algum equipamento do setor com prejuízo para a assistência. - Necessidade de algum recurso para o cuidado que não estava disponível (seja equipamento, medicamento ou mesmo vaga em CTI).
Outros
<ul style="list-style-type: none"> - Qualquer outro tipo de complicação dentro do bloco cirúrgico

APÊNDICE C: PROTOCOLO PARA FORMAÇÃO DE PADRÃO REFERÊNCIA — ROTEIRO PARA BUSCA ATIVA DE EVENTOS ADVERSOS NO BLOCO OBSTÉTRICO E MATERNIDADE

Roteiro – Busca de Incidentes SETOR BLOCO OBSTÉTRICO E MATERNIDADE
<p>Rastreadores Cuidado Obstétrico</p> <ul style="list-style-type: none"> - Conversão de parto normal para cesariana. - Lacerações 3º ou 4º grau em parto normal. - Parto instrumentalizado em parto normal (uso de fórceps). - Manobra de Kristeller. - Qualquer necessidade de hemotransfusão. - Uso de ocitotocina e misoprostol fora do contexto de indução do parto (<i>são drogas uterotônicas e ajudam a reduzir o sangramento da mãe nos casos e atonia uterina</i>). - Necessidade de anestesia geral. - Intercorrências com sedação/anestesia durante algum procedimento. - Morte materna. - Remoção, lesão ou reparo não planejado de órgão (inclui histerectomia). - Qualquer outra intercorrência obstétrica. <p>- O Bloco Obstétrico algumas vezes é usado para fazer outros procedimentos ginecológicos. Pergunte se houve algum no plantão que está se encerrando e se houve alguma intercorrência.</p>
<p>Rastreadores de cuidados gerais</p> <ul style="list-style-type: none"> - Quedas. - Necessidade de início de antibiótico (exceto os profiláticos). - Acesso venoso: <ul style="list-style-type: none"> Hematomas em locais de punção Perda/obstrução Flebites e celulites Algum paciente em uso de cateter venoso central - Hipoglicemias (menor que 60mg/dL) ou hiperglicemias (maior que 300mg/dL). - Intercorrência clínica, com necessidade de acionar o plantão médico de qualquer especialidade ou necessidade de transferência para CTI. - Retorno não planejado à sala de parto ou de cirurgia para alguma intervenção. - Necessidade de reabordagem cirúrgica. - Realização de procedimento invasivo (cateter central, cateter de diálise, dreno de tórax). - Alguma paciente em diálise (apenas se já não era previamente dialítica). - Sonolência mais importante ou agitação. Necessidade de contenção no leito. - Alguma paciente com internação prolongada (<i>maior que 5 dias</i>). - Paciente com algum grau de dependência. - Em uso de algum dispositivo como sonda, dreno ou traqueostomia. Algum intercorrência com eles.
<p>Medicamentos/Hemocomponentes (ESPECIFICAR QUAL MEDICAMENTO/HEMOCOMPONENTE ENVOLVIDO)</p> <ul style="list-style-type: none"> - Reação à infusão de algum medicamento. - Falhas de prescrição/administração de medicamentos. - Realização de transfusão de algum hemocomponente e se houve alguma reação durante a infusão.
<p>Organizacional</p> <ul style="list-style-type: none"> - Perda de material biológico coletado para exame - Falha de algum equipamento do setor com prejuízo para a assistência de algum paciente. - Necessidade de algum recurso para o cuidado que não estava disponível (seja equipamento, medicamento ou mesmo vaga em CTI). - Não preenchimento ou preenchimento incompleto da folha de sala.
<p>Outros</p> <ul style="list-style-type: none"> - Qualquer outra intercorrência ou evento que possa indicar um possível incidente relacionado ao cuidado e a assistência médico-hospitalar.

APÊNDICE D: PROTOCOLO PARA FORMAÇÃO DE PADRÃO REFERÊNCIA — ROTEIRO PARA BUSCA ATIVA DE EVENTOS ADVERSOS NO SETOR DE ENDOSCOPIA

Roteiro – Busca de Incidentes SETOR ENDOSCOPIA

- Intercorrências relacionadas à realização de exames:
 - . Sangramentos
 - . Perfurações
 - . Relacionada ao preparo (hipoglicemia por jejum, desidratação por laxativos)
- Intercorrências por sedação/anestesia
 - . Cardiovascular
 - . SNC
- Reações adversas a medicamentos ou falhas de prescrição/administração/dispensação de medicamentos.
- Quedas.
- Realização de exames em pacientes trocados.
- Perda de material biológico coletado para exame.
- Exames de pacientes internados que estavam agendados, mas que não foram realizados (não cumpriu jejum mínimo exigido, por exemplo, ou outro motivo).
- Falha de algum equipamento do setor com prejuízo para a assistência.

- Recurso indisponível, seja equipamento, medicamento ou mesmo vaga em CTI.

- **Qualquer outra intercorrência ou evento adverso.**

APÊNDICE E: PROTOCOLO PARA FORMAÇÃO DE PADRÃO REFERÊNCIA — ROTEIRO PARA BUSCA ATIVA DE EVENTOS ADVERSOS NAS ENFERMARIAS

Roteiro – Busca de Incidentes SETOR ENFERMARIAS CLÍNICAS E CIRÚRGICAS
- Setores: PA (UDC e corredor), 2° Leste, 2° Sul, 3° leste, 7° Leste, 7° Norte, 8° Leste, 8° Sul, 9° Leste, 10° Sul
Cuidado
<ul style="list-style-type: none"> - Quedas. - Lesão por pressão nova ou em piora. - Hipoglicemias (menor que 60 mg/dL) ou hiperglicemias (>400mg/dL). - Diarreia ou constipação. - Início de diálise na internação. - Acesso venoso <ul style="list-style-type: none"> - Hematomas em locais de punção. - Flebites e celulites (infecções da pele e subcutâneo). - Perda ou obstrução de acesso venoso periférico com necessidade de nova punção. - Múltiplas tentativas de acesso venoso periférico e/ou longo tempo sem acesso com atraso de administração de drogas endovenosas. - Perda acidental ou obstrução de cateteres centrais. - Sondas e drenos (ESPECIFICAR) <ul style="list-style-type: none"> - Retirada/perda/deslocamento acidental. - Obstrução de sondas ou drenos com necessidade de reposicionamento. - Administração de dieta por sonda enteral mal posicionada - Aspiração de dieta - Traqueostomia <ul style="list-style-type: none"> - Obstrução ou semi-obstrução de cânula. - Perda de cânula, decanulação inadvertida. - Neurológico <ul style="list-style-type: none"> - Alterações neurológicas novas, que surgiram ao longo da internação. - Quadro de agitação ou sonolência / rebaixamento de sensório - Uso de contenção no leito - Infecção <ul style="list-style-type: none"> - Diagnóstico de nova infecção ou início de antibiótico de urgência. - Procedimentos invasivos (realizados no setor ou em bloco cirúrgico): <ul style="list-style-type: none"> - Punção de cateter de diálise ou de acesso venoso central. - Dreno de tórax. - Ou passagem/retorno ao bloco cirúrgico não programado.
Medicamentos/Hemocomponentes (ESPECIFICAR QUAL MEDICAMENTO/HEMOCOMPONENTE ENVOLVIDO)
<ul style="list-style-type: none"> - Reações durante ou logo após a infusão de medicamentos (Urticária, vermelhão no corpo, coceira, falta de ar, etc). Necessidade de interromper uma medicação no meio da infusão. - Erros de medicação - prescrição/dispensação/administração de medicamentos. - Intercorrências durante a transfusão de sangue e seus componentes.
Intercorrências graves
<ul style="list-style-type: none"> - Necessidade de acionar o plantão médico (qualquer especialidade) por piora do quadro clínico do paciente. - PCR (parada cardiorrespiratória) ou morte. - Intubação ou reconexão em ventilação mecânica não programada. - Início de VNI (Ventilação Não Invasiva). - Pacientes usando noradrenalina, dobutamina, nitroglicerina ou nitroprussiato de sódio. - Transferência para CTI.
Organizacional
<ul style="list-style-type: none"> - Perda/extravio de material biológico coletado para exame. - Falha de algum equipamento do setor com prejuízo para a assistência de algum paciente. - Necessidade de algum recurso para o cuidado que não estava disponível (seja equipamento, medicamento ou mesmo vaga em CTI).
Outros
<ul style="list-style-type: none"> - Qualquer outra intercorrência ou evento que possa indicar um possível incidente relacionado ao cuidado e assistência médico-hospitalar.

APÊNDICE F: PROTOCOLO PARA FORMAÇÃO DE PADRÃO REFERÊNCIA — ROTEIRO PARA BUSCA ATIVA DE EVENTOS ADVERSOS NAS TERAPIAS INTENSIVAS

Roteiro -Busca de incidentes SETOR DE TERAPIA INTENSIVA por categoria
<p>1. Medicação</p> <p><u>1.1. incidente relacionado a prescrição</u></p> <ul style="list-style-type: none"> - médico prescreveu medicação para paciente “x” em dados de paciente “y”? - dose medicação acima da desejada? (ex. Midazolam 15mg fez 50mg, bolus medicamentos em BIC) - farmácia dispensou medicamento diferente do prescrito pelo médico? - farmácia não mandou medicamento? - faltou medicamento no hospital? (Ex.: médico prescreveu ou queria prescreveu porem estoque zerado) - relacionada a prescrição verbal (pedido de urgência por médico) - falta de prescrição de algum item? - outro problema? <p><u>1.2. reação adversa a administração</u></p> <ul style="list-style-type: none"> - medicamento administrado para paciente “x” e foi prescrito para “y”? - anafilaxia, alergia, outra reação - via administração deveria ser IM e foi EV? Ou vice versa? Ou VO? SNE? - dose de medicação trouxe incidentes adversos? (ex. era para ser ampola de 15mg e fez de 50mg – sonolência excessiva; ou dose acabou sendo alta e trouxe alguma consequência ao paciente) - sonolência ou rebaixamento com necessidade de intervenção - hipotensão com necessidade de volume ou aminas - taquiarritmia atribuída a medicação - diarreia e constipação devido medicação (ex. opióide) <p>2. Via aérea/VM</p> <p><u>2.1. intubação nova</u> <input type="checkbox"/>SIM <input type="checkbox"/>NÃO (se não, pule 2.1 e vá para item 2.2)</p> <ul style="list-style-type: none"> - procedimento com intercorrência? (via aérea difícil -> várias tentativas de TOT, uso de via aérea avançada – mascara laríngea, intubação esôfago?) - hipotensão pos medicações pre TOT? - PCR peri TOT? - motivo TOT eletivo? Urgência por rebaixamento do sensório/ piora respiratória? - reintubação (ja estava com TOT outro momento nessa internação?) - pneumotórax pos TOT? <p><u>2.2. problemas relacionadas a TOT/VM</u></p> <ul style="list-style-type: none"> - obstrução do TOT? - extubação acidental? - pneumotórax associado a TOT/VM - desconexão na VM não programada - broncoaspiração - atelectasia com repercussão na saturação (virou o paciente e percebeu piora clinica e respiratória) <p><u>2.3. traqueostomia</u></p> <ul style="list-style-type: none"> - decanulação acidental - rolha? - estenoses, fistulas e perfurações de via aérea - intercorrências relacionadas a realização da TQT (sangramento, etc)

Roteiro -Busca de incidentes
SETOR DE TERAPIA INTENSIVA por categoria

3. acessos vasculares

3.1 Punção nova SIM NÃO (Central, PIA, CDL, swan ganz ou outro. Se não, pule para item 3.2)

- hemorragia com repercussão hemodinâmica/hematimétrica ou necessidade de transfusão durante a punção
- pneumotórax ou hemotórax decorrente de punção de acesso venoso central
- > se sim, perguntar se foi punção guiada por US.

3.2. Cateter prévio (incluindo periférico)

- Perda acidental, obstrução - com necessidade de nova punção
- Hematoma com repercussão (necessidade de transfusão ou instabilidade clínica/laboratorial ou sintomatologia intensa – precisou medicação para dor relacionada)
- Trombose associada a cateter
- infecção de cateter central em tratamento
- flebite
- não conseguiu periférico de jeito nenhum

4 Sondas e drenos

- SVD - infecção ou trauma relacionado / obstrução / retirada acidental
- SNE – deslocamento com necessidade de reposicionamento/ obstrução / retirada acidental

5. Cuidados/profilaxias

- pneumonia associada a VM (VM > 48 horas)
- TVP ou TEP em paciente internado prévio ou na enfermaria ou CTI
- não realização de algum cuidado (?)
- úlceras de pressão novas ou piora de pré-existente
- queda

6. Cirurgias (passagem pelo bloco cirúrgico não programada)

- Complicações de cirurgias (sangramentos, infecções, complicações clínicas – AVC, IAM, etc, qualquer outra complicação durante ou após procedimento cirúrgico que possa ter relação com o procedimento).

7. Hemorragias

- HDA ou HBD (que não seja motivo de admissão hospitalar - busca de LAMGD)
- hemorragia de outro sítio
- relacionada a diálise

8. Readmissões

- readmissão ao CTI
- readmissão hospitalar precoce (definir tempo)
- readmissão em terapia intensiva precoce

9. PCR/morte

- motivo
- esperado ou não

10. Aparelhos

- falha de equipamento existente (BIC, monitor, ECG, transdutor, PNI, aspirador, respirador, laringoscópio)
- precisava de equipamento mas não tinha (capnógrafo, bougie, monitor PIC)

Roteiro -Busca de incidentes
SETOR DE TERAPIA INTENSIVA por categoria

11. Hemocomponentes

(Transfusões de sangue? SIM NÃO – sem não, pule esse item)

- TRALI ou TACO (IRpA apos ou durante transfusão)
- reações cutâneas (hiperemia, prurido intenso)
- anafilaxia
- administração de hemocomponente em paciente “x” que deveria ser no “y”

12. Hipoglicemias

- glicemia <70 com ou sem sintomas, mas que tenha sido comunicada ou tenha intervenção (SGH, alimento)

13. Outro

- paciente retirou monitorização
- paciente agitado com necessidade de contenção no leito
- Sonolência ou rebaixamento do nível de consciência após procedimento ou não explicado por doença atual.
- **Qualquer outra intercorrência**

APÊNDICE G: PROTOCOLO PARA FORMAÇÃO DE PADRÃO REFERÊNCIA — RELAÇÃO DE RESULTADOS DE EXAMES LABORATORIAIS RASTREADOS

A
R

Resultado de exames laboratoriais rastreados

Exame laboratorial	Valores rastreados na pesquisa	Valores médios de referência
Ácido valpróico	< 50 e > 120 µg/mL	Terapêutico: 50,0 – 120,00 µg/mL
ALT (TGP)	> 345 U/L	< 35 U/L para ♀ e < 50 U/L para ♂
Amilase	> 330 U/L	30 – 110 U/L
AST(TGO)	> 230 U/L	15 – 46 U/L
Bicarbonato	< 12 mmol/L	22 - 30 mmol/L
Carbamazepina	< 4 e > 12 µg/mL	4,0 – 12,0 µg/mL
Creatinina	aumento 2x em relação ao resultado do exame anterior; ou aumento em 1,0 mg/dL de um exame em relação ao resultado do exame anterior	0,52 – 1,04 mg/dL para ♀ e < 0,66 – 1,25 mg/dL para ♂
Digoxina	> 2 ng/mL	0,8 a 2,0 ng/mL
Fenitoína	< 10 e > 20 µg/mL	Terapêutico 10,0 – 20,0 µg/mL
Fenobarbital	< 15 e > 40 µg/mL	Terapêutico 15,0 a 40,0 µg/mL
Glicose	< 60 e > 400 mg/dL	Adulto (jejum): 55 – 99 mg/dL
Hemoglobina	< 5,0 g/dL	12,0 – 16,0 g/dL para ♀ e 13,0 – 17,5 g/dL para ♂
Lactato	> 4 mmol/L	0,7 – 2,5 mmol/L para sangue venoso
Lipase	> 900 U/L	23 – 300 U/L
pCO ₂ (arterial)	> 60 mmHg	32 – 48 mmHg
pH (arterial)	< 7,15	7,35 – 7,45
pO ₂ (arterial)	> 120 mmHg	83 – 108 mmHg
Potássio	> 6,5 mmol/L	3,5 a 5,1 mmol/L
PTTa	> 120 segundos	26 – 30 segundos
RNI	> 5,0	2,0 a 3,0 ou 2,5 a 3,5 para pacientes anticoagulados
Sódio	< 120 ou > 160 mmol/L; ou variação de 10 mmol/L ou mais para mais ou para menos em menos de 24 horas.	137 – 145 mmol/L
Toxina A e B (pesquisa para <i>Clostridium</i>)	Positiva	Negativo
Ureia	> 170 mg/dL	15 – 36 mg/dL para ♀ e 19 – 43 mg/dL para ♂
Vancomicina	<10 e > 20 µg/mL	10,0 a 20,0 µg/mL

Número do Formulário	Número do incidente	Número do paciente na pesquisa
-------------------------	------------------------	-----------------------------------

Formulário 2A - Busca ativa de Incidentes

Identificação
 Iniciais do nome: _____ Prontuário: _____ Data de nascimento: ___/___/____
 Data da internação: ___/___/2016

Dados do evento suspeito
 Local de internação no momento de ocorrência do evento – Leito: _____ Andar/Setor: _____
 Local de ocorrência do evento: ___/1- mesmo da internação ___/2- Outro: _____
 Dia de ocorrência do evento: ___/___/2016. Turno: ___/1- Manhã ___/2- Tarde ___/3- Noite
 ___/4- Não se aplica /impossível identificar

Dado para suspeita (Rastreador de busca ativa):

Código: _____ (Tabela 1 – Rastreadores de busca ativa)

Informação obtida por meio de:

<p>1-Relato de profissional: ___/1a- Técnico de enfermagem ___/1b- Enfermeiro ___/1c- Plantão Médico ___/1d- Médico assistente ___/1e- Médico interconsultor ___/1f- Outros profissionais da assistência (multiprofissional): _____ ___/1g- Profissional não ligado a assistência (laboratório, Imagem, Banco de sangue, etc)</p> <p>___/4- Dado de resultado de exames ___/5- Notificações ___/6- Outro _____</p>	<p>2- Anotação de profissional em prontuário, prescrições ou formulários de avaliação próprios da internação. ___/2a- Técnico de enfermagem ___/2b- Enfermeiro ___/2c- Médico ___/2d- Acadêmico ___/2e- Outro profissional</p> <p>3- Relato de usuários: ___/3a- Relato do próprio paciente ___/3b- Relato de familiar do paciente envolvido ___/3c- Relato de outro paciente ou familiar de outro paciente</p>
--	---

Descrição do possível incidente:

Nome do 1º investigador: _____ **Data:** ___/___/2016
Nome do 2º investigador: _____ **Data:** ___/___/2016

INCIDENTE: ___/0- DESCARTADO (Interromper investigação)
 ___/1- CONFIRMADO (Prosseguir investigação em Formulário 2B)

APÊNDICE I: PROTOCOLO PARA FORMAÇÃO DE PADRÃO REFERÊNCIA — FORMULÁRIO PARA CARACTERIZAÇÃO DOS EVENTOS ADVERSOS

Número do paciente na pesquisa		Número do incidente		CTI ___/0- Não ___/1- Sim	
Formulário 2B - Caracterização do Incidente					
Nome do 2º investigador: _____			Data: ___/___/2017		
Tipo de Incidente*					
Possível Incidente: _____					
Dano					
Descrever o dano: _____					
Classificação segundo índice de MERP NCC ^(1*) : _____					
Classificação OMS ^(2*) : _____					
Rastreadores:					
	Descrição do rastreador	Código⁽³⁾	Número⁽⁴⁾	Coleta⁽⁵⁾	R-I⁽⁶⁾
1					
2					
3					
4					
5					
6					
<small>⁽³⁾Tabela 1 – Tipo de Rastreador. / ⁽⁴⁾Identificação letra(s)-número do rastreador / ⁽⁵⁾Tipo de coleta: 1-Tempo real (P); Eletrônico 2-ATB; 3-BC (Bloco Cirúrgico); 4-BS (Banco de Sangue); 5- CCIH; 6-EE (Exames); 7-EO (Obstetria); 8-EOB (Óbito); 9-ER (readmissões); 10-EV (Vigihosp); 11-Nenhum dos anteriores: revisão do prontuário. / ⁽⁶⁾Correlação Rastreador - Incidente</small>					
Identificação em prontuário					
Tipo de formulário: _____			Data: ___/___/___		
Profissional: _____			Especialidade: _____		
Momento da ocorrência do evento adverso					
Dia: ___/___/___ Turno: ___/1- Manhã (7:01 às 13h) ___/2- Tarde (13:01 às 19h) ___/3- 1º turno da noite (19:01 às 0h) ___/4- 2º turno da noite (0:01 às 7h) ___/5-NA ou indeterminado					
___/1-Antes da internação índice.					
___/2-Durante a internação índice: ocorrência e/ou abordagem do evento durante o período da pesquisa.					
___/3-Durante a internação índice: ocorrência e/ou abordagem do evento fora do período da pesquisa.					
___/4-Depois da internação índice.					
___/9-Não se aplica ou não foi possível determinar					
Local internação: _____ Mesmo da internação? ___/1- Sim ___/9-NA ou indeterminado					
___/0- Não. Qual? _____					
Instituição de cuidado envolvido na ocorrência do evento adverso					
___/1-Mesma da internação – cuidado hospitalar.					
___/2-Mesma da internação – cuidado ambulatorial geral.					
___/3-Mesma da internação – cuidado ambulatorial especializado.					
___/4-Outra instituição - cuidado hospitalar.					
___/5-Outra instituição – cuidado ambulatorial geral.					
___/6-Outra instituição – cuidado ambulatorial especializado.					
___/7-Outra instituição – casa de repouso ou de cuidados prolongados.					
___/8-Outros: _____					
Possibilidade de relação do Evento Adverso com administração de medicamentos					
___/0- Não ___/1- Sim					

(7*) Relação de Causa

Com que nível de confiança você diria que CUIDADO PRESTADO causou a lesão ou dano? _____
 Ao responder a esta pergunta, sempre que for relevante, responda antes todas as questões que forem aplicáveis à situação contidas no Anexo (7).

Necessário consultar especialista? ___/0-Não ___/1-Sim. Opinião do especialista: _____

(8*) Classificação quanto EVITABILIDADE

Classifique, atribuindo pontuação numa escala de 6 pontos, seu grau de confiança nas evidências de possibilidade de evitabilidade. _____

Ao responder a esta pergunta, sempre que for relevante, responda antes todas as questões que forem aplicáveis à situação contidas no Anexo (8).

Necessário consultar especialista? ___/0-Não ___/1-Sim. Opinião do especialista: _____

Consequências organizacionais – Aumento dos recursos necessários para o doente (Possível mais de 1)

- ___/1- Aumento no tempo de permanência.
 ___/2- Admissão em setor de maior complexidade.
 ___/3- Tratamento/exames adicionais.
 ___/4- Perturbação do trabalho/atraso para os outros doentes.
 ___/5- Necessidade de pessoal adicional.
 ___/6- Necessidade de equipamento adicional.
 ___/7- Não foi possível caracterizar.

Fatores contribuintes (Possível mais de 1)

1-Improvável 2-Possivelmente relevante 3-Relevante 4-Muito relevante 5-Indeterminado

1- Falha de comunicação: ___/1a- Momento de transição de cuidado ___/1b- Outro

Descrever: _____

___/2- Equipamentos com defeito ou indisponíveis. Qual: _____

Descrever: _____

___/3- Problemas com agendamento ou prestação de serviços (inclui consultas e interconsultas, procedimentos, cirurgias, etc). Qual: _____

Descrever: _____

___/4- Funcionamento inadequado dos serviços de apoio terapêutico ou diagnóstico. Qual: _____

Descrever: _____

___/5- Equipe inadequada no momento da ocorrência do evento adverso. Descrever: _____

___/6-Falta de recursos essenciais. Qual: _____

Descrever: _____

___/7- Outros. Especificar: _____

Conclusão

___/1- Evento Adverso ___/2-Incidente sem dano ___/3- Quase falha ___/4- Circunstância notificável

___/5- Associado à doença ___/9- Indeterminado

Número do paciente na pesquisa		Número do incidente	
Consulta a especialista			
Especialidade: _____			
Pergunta: _____			

Parecer do especialista			
Nome do especialista: _____ Data: __/__/____			
Parecer: _____			

