

ROBERTA MURIEL LONGO ROEPKE

**Comparação da performance preditiva de escores prognósticos
em pacientes vítimas de trauma internados em
unidade de terapia intensiva:
um estudo de validação externa de SAPS 3, SOFA, ISS e NISS**

Tese apresentada à Faculdade de Medicina
da Universidade de São Paulo para
obtenção do título de Doutor em Ciências

Programa de Anestesiologia, Ciências
Cirúrgicas e Medicina Perioperatória

Área de concentração: Medicina
Perioperatória

Orientador: Prof. Dr. Luiz Marcelo Sá
Malbouisson

São Paulo

2022

ROBERTA MURIEL LONGO ROEPKE

**Comparação da performance preditiva de escores prognósticos
em pacientes vítimas de trauma internados em
unidade de terapia intensiva:
um estudo de validação externa de SAPS 3, SOFA, ISS e NISS**

Tese apresentada à Faculdade de Medicina
da Universidade de São Paulo para
obtenção do título de Doutor em Ciências

Programa de Anestesiologia, Ciências
Cirúrgicas e Medicina Perioperatória

Área de concentração: Medicina
Perioperatória

Orientador: Prof. Dr. Luiz Marcelo Sá
Malbouisson

São Paulo

2022

Dados Internacionais de Catalogação na Publicação (CIP)

Preparada pela Biblioteca da
Faculdade de Medicina da Universidade de São Paulo

©reprodução autorizada pelo autor

Roecke, Roberta Muriel Longo

Comparação da performance preditiva de escores prognósticos em pacientes vítimas de trauma internados em unidade de terapia intensiva : um estudo de validação externa de SAPS 3, SOFA, ISS e NISS / Roberta Muriel Longo Roecke. -- São Paulo, 2022.

Tese(doutorado)--Faculdade de Medicina da Universidade de São Paulo.
Programa de Anestesiologia, Ciências Cirúrgicas e Medicina Perioperatória.
Área de Concentração: Medicina Perioperatória.

Orientador: Luiz Marcelo Sá Malbouisson.

Descritores: 1.Prognóstico 2. Índices de gravidade do trauma 3.Traumatismo múltiplo 4.Traumatismos craniocerebrais 5. Lesões encefálicas traumáticas 6.Cuidados críticos 7.Unidades de terapia intensiva

USP/FM/DBD-487/22

Responsável: Erinalva da Conceição Batista, CRB-8 6755

Dedicatória

A meus pais, **Silvana e Roberto**, a quem devo a pessoa que sou hoje. Obrigada pelo carinho, dedicação e incentivo contínuo à educação.

A minha irmã, **Giorgia**, pelo companheirismo e exemplo. Obrigada por sempre ter apoiado meus projetos de vida.

A meu marido, **Bruno**, pelo incentivo, companheirismo e amor nesses anos juntos. Obrigada pela paciência e apoio incondicional para alcançar este sonho.

Agradecimentos

Agradeço a meus amigos Renata Mello Guazzelli e Renato Daltro de Oliveira, que não só contribuíram diretamente ao sucesso deste projeto, mas pela amizade nos últimos anos.

Aos colegas da Unidade de Terapia Intensiva de Trauma e Emergência Cirúrgica do Hospital das Clínicas da FMUSP, pela paciência comigo na elaboração desta tese. Em especial, agradeço ao Dr. Estevão Bassi, pelo apoio e contribuições, desde o início deste projeto.

A meu orientador, Prof. Dr. Luiz Marcelo Sá Malbouisson, pela oportunidade e contribuições no desenvolvimento desta tese.

A todos os professores que participaram de minha formação acadêmica, pelo aprendizado e por estimularem minha curiosidade e instigarem-me a buscar sempre novos conhecimentos.

A todos os participantes anônimos desta pesquisa, vítimas da maior causa de incapacidade e morte de jovens em nosso País. Que a pesquisa clínica possa contribuir no cuidado aos pacientes vítimas de trauma, uma condição ainda tão negligenciada.

Nunca houve, como hoje, um momento com mais informação e menos conhecimento.

(George Steiner)

Normalização

Esta tese está de acordo com as seguintes normas, em vigor no momento desta publicação:

Referências: adaptado de *International Committee of Medical Journals Editors* (Vancouver).

Universidade de São Paulo. Faculdade de Medicina. Divisão de Biblioteca e Documentação. Guia de apresentação de dissertações, teses e monografias. Elaborado por Anneliese Carneiro da Cunha, Maria Julia de A. L. Freddi, Maria F. Crestana, Marinalva de Souza Aragão, Suely Campos Cardoso, Valéria Vilhena. 3ª ed. São Paulo: Divisão de Biblioteca e Documentação; 2011.

Abreviaturas dos títulos dos periódicos de acordo com *List of Journals Indexed in Index Medicus*.

Sumário

Lista de Abreviaturas e Siglas	ix
Lista de Tabelas	xi
Lista de Figuras	xii
Resumo	xiii
Abstract	xv
1 INTRODUÇÃO	1
2 OBJETIVO	5
3 REVISÃO DA LITERATURA.....	7
3.1 Modelos preditivos em trauma	8
3.1.1 Modelos anatômicos em trauma	8
3.1.2 Modelos fisiológicos em trauma	10
3.1.3 Modelos combinados em trauma	11
3.2 Modelos prognósticos em doentes críticos.....	13
3.2.1 Modelos prognósticos de gravidade de doença	13
3.2.2 Modelos prognósticos que reflitam disfunção orgânica	14
3.3 Recomendações atuais para avaliação da performance preditiva de modelos prognósticos	15
4 MÉTODOS.....	19
4.1 Aspectos éticos	20
4.2 Delineamento do estudo	20
4.3 Caracterização da unidade de terapia intensiva e do hospital.....	20
4.4 População do estudo.....	21
4.4.1 Critérios de inclusão.....	21
4.4.2 Critérios de exclusão	21
4.5 Desfechos	22
4.6 Coleta e manejo de dados	22
4.6.1 Extração de variáveis do banco de dados	22
4.6.2 Coleta retrospectiva de dados.....	25
4.6.3 Derivação de variáveis a partir do banco de dados	27
4.6.4 Revisão e qualidade dos dados	27

4.7	Análise estatística	28
4.7.1	Manejo de dados faltantes	32
5	RESULTADOS.....	33
5.1	Discriminação dos escores prognósticos	41
5.2	Calibração dos escores prognósticos	43
5.3	Utilidade clínica dos escores prognósticos	49
5.4	Análise estratificada, de acordo com a presença de traumatismo cranioencefálico grave	50
5.5	Análise de sensibilidade com imputação múltipla	52
6	DISCUSSÃO	53
6.1	Implicações para a prática clínica e avaliação de desfechos da UTI.....	58
6.2	Implicações para pesquisa.....	60
6.3	Pontos fortes e limitações	60
7	CONCLUSÕES	63
8	ANEXOS	65
	Anexo A - Template do banco de dados	66
	Anexo B - Tabela AIS	68
	Anexo C - Tabela SOFA.....	71
	Anexo D – Charlson Comorbidity Index	72
	Anexo E – Preenchimento do Checklist TRIPOD	73
9	REFERÊNCIAS	77

Listas

Lista de Abreviaturas e Siglas

AIS	<i>Abbreviated Injury Scale</i>
AUROC	<i>Area Under the Receiver Operating Characteristic</i>
CITL	<i>Calibration-in-the-large</i>
FMUSP	Faculdade de Medicina da Universidade de São Paulo
FR	Frequência respiratória
GCS	<i>Glasgow Coma Scale</i>
ISS	<i>Injury Severity Score</i>
ILP	Instituição de longa permanência
IC	Intervalo de Confiança
ICU	<i>Intensive care unit</i>
IIQ	Intervalo interquartil
NISS	<i>New Injury Severity Score</i>
PAD	Pressão arterial diastólica
PAM	Pressão arterial média
PAS	Pressão arterial sistólica
RTS	<i>Revised Trauma Score</i>
RASS	<i>Richmond Agitation Sedation Scale</i>
SOFA	<i>Sequential Organ Failure Assessment</i>
SAPS 3	<i>Simplified Acute Physiology Score 3</i>
TBI	<i>Traumatic brain injury</i>

TC	Tomografia computadorizada
TMP	Taxa de mortalidade padronizada
TRIPOD	<i>Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis</i>
TCE	Traumatismo cranioencefálico
UTI	Unidade de terapia intensiva

Lista de Tabelas

Tabela 1 -	Características dos pacientes	36
Tabela 2 -	Mecanismos específicos de trauma.....	37
Tabela 3 -	Segmentos corporais acometidos.....	38
Tabela 4 -	Comparações individuais de AUROC entre os escores	42
Tabela 5 -	Equações de predição de mortalidade, escore de Brier e medidas de calibração dos escores.....	43
Tabela 6 -	Análise de subgrupo, de acordo com a presença de TCE grave	50
Tabela 7 -	Análise de sensibilidade, após imputação múltipla de dados faltantes nos escores prognósticos	52

Lista de Figuras

Figura 1 - Fluxograma do estudo	34
Figura 2 - Histograma do SAPS 3 estratificado por desfecho hospitalar.....	39
Figura 3 - Histograma do SOFA total máximo em 24 horas estratificado por desfecho hospitalar	39
Figura 4 - Histograma do ISS estratificado por desfecho hospitalar	40
Figura 5 - Histograma do NISS estratificado por desfecho hospitalar.....	40
Figura 6 - Comparação de AUROC dos escores prognósticos.....	41
Figura 7 - Gráfico de calibração do SAPS 3.....	44
Figura 8 - Gráfico de calibração do SAPS 3 Equação América Central/Sul	44
Figura 9 - Gráfico de calibração do SAPS 3 Equação recalibrada.....	45
Figura 10 - Gráfico de calibração do SOFA total máximo nas 24 horas de admissão.....	46
Figura 11 - Gráfico de calibração do ISS	47
Figura 12 - Gráfico de calibração do NISS.....	47
Figura 13 - Gráfico de calibração comparando todos os escores	48
Figura 14 - Curvas de utilidade clínica dos escores.....	49
Figura 15 - Gráficos de calibração em pacientes com TCE grave. A: SAPS 3 Equação Geral; B: SAPS 3 Equação América Central/Sul; C: SAPS 3 Equação recalibrada; D: SOFA; E: ISS; F: NISS.....	51

Resumo

Roepke RML. *Comparação da performance preditiva de escores prognósticos em pacientes vítimas de trauma internados em unidade de terapia intensiva: um estudo de validação externa de SAPS 3, SOFA, ISS e NISS [tese]*. São Paulo: Faculdade de Medicina, Universidade de São Paulo; 2022.

INTRODUÇÃO: A performance preditiva dos escores anatômicos em comparação com escores de doença crítica, em geral, em pacientes vítimas de trauma, admitidos em unidade de terapia intensiva (UTI) é desconhecida. O objetivo desta análise é realizar a validação externa, comparando a performance preditiva para mortalidade hospitalar de escores de doença crítica em geral (SAPS 3 e SOFA total nas primeiras 24 horas de admissão na UTI) e, de escores anatômicos (ISS e NISS) nessa população. **MÉTODOS:** Coorte retrospectiva de pacientes vítimas de trauma admitidos em uma UTI especializada de centro terciário de São Paulo entre maio de 2012 e janeiro de 2016. Dados relativos à admissão e primeira evolução para cálculo de SAPS 3 e SOFA foram coletados prospectivamente no banco de dados da unidade. Os escores ISS e NISS foram calculados retrospectivamente com base em resultados de tomografia e prontuário médico. A performance para predição da mortalidade hospitalar foi avaliada pela discriminação, calibração e utilidade clínica pela curva de decisão. **RESULTADOS:** A coorte incluiu 1.053 pacientes admitidos por trauma na UTI, sendo 84,2% do sexo masculino e com idade média de 40 (± 18) anos. O mecanismo principal de trauma foi contuso (90,7%). Traumatismo cranioencefálico esteve presente em 67,8% dos pacientes, sendo grave em 43,3%. No momento da admissão na UTI, 846 pacientes (80,3%) estavam em ventilação mecânica invasiva e 644 (64,3%) em uso de droga vasoativa. Óbito hospitalar ocorreu em 251 pacientes (23,8%). Em relação aos escores prognósticos, a mediana de SAPS 3 foi 41; de SOFA total máximo nas 24 horas de admissão, 7; de ISS, 29; e NISS, 41. AUROCs (IC 95%) foram: SAPS 3 = 0,786 (0,756-0,817), SOFA total máximo em 24 horas de admissão = 0,807 (0,778-0,837), ISS = 0,616 (0,577-0,656) e NISS = 0,689 (0,649-0,729). Em comparações individuais, o SAPS 3 e o SOFA não foram diferentes, mas ambos foram melhores que os escores anatômicos ($p < 0,001$). O SAPS 3 na equação geral e na equação América Central/Sul apresentaram calibração imperfeita. Após recalibração do intercepto e *slope* da equação, o SAPS 3 recalibrado apresentou melhor calibração ao longo de todo o limiar de risco do desfecho, semelhante ao observado com SOFA. O SOFA total nas primeiras 24 horas de admissão na UTI apresentou calibração próxima de ideal, ao longo de ampla faixa de predição de risco e o maior benefício líquido na curva de decisão. **CONCLUSÃO:** Os escores anatômicos ISS e NISS têm performances preditivas baixas em pacientes críticos com trauma, sendo superados pelos escores SAPS 3 e SOFA total máximo de 24 horas de admissão. A gravidade de doença crítica em trauma

é mais bem caracterizada por disfunção orgânica e variáveis clínicas do que por lesões anatômicas.

Descritores: Prognóstico; Índices de gravidade do trauma; Traumatismo múltiplo; Traumatismos craniocerebrais; Lesões encefálicas traumáticas; Cuidados críticos; Unidades de terapia intensiva.

Abstract

Roepke RML. *Predictive performance for hospital mortality of SAPS 3, SOFA, ISS and new ISS in critically ill trauma patients: a validation cohort study* [thesis]. São Paulo: “Faculdade de Medicina, Universidade de São Paulo”; 2022.

INTRODUCTION: It's not known whether anatomical scores perform better than general critical care scores for trauma patients admitted to the intensive care unit (ICU). We compare the predictive performance for hospital mortality of general critical care scores (SAPS 3 and SOFA) with anatomical injury-based scores (injury severity score [ISS] and new ISS [NISS]). **METHODS:** Retrospective cohort study of patients admitted to a specialized trauma ICU from a tertiary hospital in São Paulo, Brazil between May, 2012 and January, 2016. We retrieved data from the ICU database for critical care scores and calculated ISS and NISS from chart data and whole body computed tomography results. We compared the predictive performance for hospital mortality of each model through discrimination, calibration, and decision-curve analysis. **RESULTS:** The sample comprised 1053 victims of trauma admitted to the ICU, with 84.2% male patients and mean age of 40 (± 18) years. Main injury mechanism was blunt trauma (90.7%). Traumatic brain injury (TBI) was present in 67.8% of patients; 43.3% with severe TBI. At the time of ICU admission, 846 patients (80.3%) were on mechanical ventilation and 644 (64.3%) on vasoactive drugs. Hospital mortality was 23.8% (251). Median SAPS 3 was 41; median maximum SOFA within 24 hours of admission, 7; ISS, 29; and NISS, 41. AUROCs (95% CI) were: SAPS 3 = 0.786 (0.756-0.817), SOFA = 0.807 (0.778-0.837), ISS = 0.616 (0.577-0.656) and NISS = 0.689 (0.649-0.729). In pairwise comparisons, SAPS 3 and SOFA did not differ, while both outperformed the anatomical scores ($p < 0.001$). SAPS 3 was miscalibrated for the general equation and for the customized Central, South America equation. After recalibration, SAPS 3 presented a better calibration throughout the entire range of expected risks, like that observed with SOFA. Maximum SOFA within 24 hours of admission presented the best calibration and net benefit in decision-curve analysis. **CONCLUSIONS:** Trauma-specific anatomical scores have fair performance in critically ill trauma patients and are outperformed by SAPS 3 and SOFA. Illness severity is best characterized by organ dysfunction and physiological variables than anatomical injuries.

Descriptors: Prognosis; Trauma severity indices; Multiple trauma; Craniocerebral trauma; Traumatic brain injuries; Critical care; Intensive care units.

1 Introdução

1 INTRODUÇÃO

Os modelos prognósticos são desenvolvidos para ajudar a prever a probabilidade de um evento. Em medicina intensiva, muitos escores prognósticos foram desenvolvidos para avaliar a gravidade da doença e a probabilidade de mortalidade hospitalar. O uso de escores prognósticos é importante por diversos motivos, incluindo a avaliação de resultados da Unidade de Terapia Intensiva (UTI) ao longo do tempo, *benchmarking* e aplicação clínica para a prática diária, quando especificamente validados com este fim. A correta aplicação de escores prognósticos em uma determinada situação necessita de validação externa, que é uma etapa pouco descrita na literatura. Esta deve ser realizada em subpopulações específicas, bem como em regiões geográficas e intervalos de tempo diferentes das descrições originais.

Muitos escores prognósticos foram criados especificamente para predição de eventos em pacientes vítimas de trauma de maneira geral, os quais podem ser classificados em anatômicos, fisiológicos ou combinados. Eles incluem variáveis precoces da apresentação do trauma e da admissão no hospital, como dados relativos à extensão e gravidade das lesões anatômicas sofridas e variáveis clínicas ou laboratoriais que reflitam desarranjos fisiológicos consequentes ao trauma.

Entre os escores anatômicos em trauma mais utilizados em estudos clínicos e epidemiológicos para caracterização de vítimas de trauma, estão os derivados do *Abbreviated Injury Scale* (AIS), como o *Injury Severity Score*

(ISS) ⁽¹⁾ e sua posterior modificação, o *New Injury Severity Score* (NISS). Entretanto, esses escores foram desenvolvidos baseados em opinião de especialistas, sem processo de derivação e validação baseados em dados e não incluem variáveis fisiológicas e de disfunção orgânica. Além disso, a performance preditiva desses escores em estratos mais graves, como em pacientes críticos, não foi adequadamente avaliada.

Por outro lado, para pacientes críticos em geral, entre os mais difundidos modelos prognósticos estão o *Simplified Acute Physiology Score 3* (SAPS 3) ^(2, 3), que combina variáveis demográficas, antecedentes e de acuidade da doença, e o *Sequential Organ Failure Assessment* (SOFA) ⁽⁴⁾, que avalia a presença e a gravidade de disfunção orgânica e sua evolução na terapia intensiva.

Embora o desenvolvimento de novos modelos prognósticos seja, frequentemente, considerado por pesquisadores, a validação externa dos escores preditivos é etapa fundamental e pode ser até mais relevante do que a criação de novos modelos.

A subpopulação de pacientes vítimas de trauma críticos é de grande relevância, em especial, em países de média renda como o Brasil, onde causas relacionadas a trauma estão entre em as dez maiores causas de morte e representam a segunda causa de incapacidade na população ⁽⁵⁾. A performance preditiva de diferentes escores prognósticos nessa subpopulação foi avaliada em análise de subgrupo de estudos maiores, onde foi demonstrada alguma descalibração ^(2, 6), ou em estudos menores com tamanhos amostrais relativamente pequenos e sem representatividade

de países de média renda ^(7, 8). Além disso, embora os escores anatômicos sejam amplamente utilizados para descrever a gravidade de vítimas de trauma em geral, faltam validações de desempenho especificamente naqueles pacientes com necessidade de admissão em UTI. Assim, a hipótese principal é que escores prognósticos que reflitam gravidade clínica no momento da admissão na UTI (como SAPS 3), ou que reflitam disfunção orgânica (como SOFA) predizem melhor mortalidade hospitalar nesta população, em comparação com escores anatômicos específicos (como ISS e NISS).

O objetivo deste estudo é realizar uma validação externa, comparando a performance de escores utilizados em doentes críticos em geral (SAPS 3 e SOFA) com escores anatômicos específicos para vítimas de trauma (ISS e NISS) para predição de mortalidade hospitalar em uma coorte retrospectiva de pacientes vítimas de trauma admitidos em uma UTI especializada no Brasil.

2 Objetivo

2 OBJETIVO

O objetivo primário do presente estudo é comparar a performance de diferentes escores prognósticos: SAPS 3, SOFA total máximo nas 24 horas de admissão na UTI, ISS e NISS, para predição de mortalidade hospitalar em pacientes ≥ 15 anos vítimas de trauma internados em UTI.

O objetivo secundário é comparar a performance preditiva desses escores estratificados de acordo com um subgrupo de interesse clínico, aqueles com presença de traumatismo cranioencefálico (TCE) grave.

3 Revisão da Literatura

3 REVISÃO DA LITERATURA

3.1 Modelos preditivos em trauma

A descrição dos modelos preditivos em trauma historicamente esteve relacionada à organização dos sistemas de atendimento às vítimas de trauma nos Estados Unidos da América, a partir da década de 1970. A necessidade de definir pacientes com lesões de maior gravidade, relacionadas à maior morbidade e mortalidade foi necessária para definir o planejamento de atendimento, de uso de recursos e da própria avaliação e benchmarking dos desfechos dos pacientes.

A combinação de variáveis, como: idade, comorbidades, sexo, mecanismo de trauma, variáveis fisiológicas e extensão e gravidade das lesões anatômicas é usada em diferentes escores preditivos em trauma. Eles podem ser classificados em anatômicos, fisiológicos e combinados e contemplam variáveis relativas ao atendimento inicial do paciente, no momento da admissão no hospital.

3.1.1 Modelos anatômicos em trauma

Com o objetivo inicial de avaliar e comparar lesões em pacientes vítimas de acidente automobilístico, em 1971, foi publicado o *Abbreviated Injury Scale* ⁽⁹⁾. Um grupo de especialistas de diversas áreas listou lesões possíveis e as graduou em gravidade, de acordo com segmentos corporais, na tentativa de padronizar uma ferramenta de avaliação das lesões. A

descrição inicial contemplava uma pequena quantidade de lesões possíveis, mas após sua publicação original, posteriores atualizações periódicas foram publicadas. O AIS classifica a gravidade individual das lesões em: 1 – leve, 2 – moderada, 3 – grave, 4 – severa, 5 – crítica e 6 – máxima (considerada intratável, não-sobrevivível), em diferentes segmentos corporais: cabeça e pescoço, face, tórax, abdome, pelve e extremidades e superfície externa.

Embora o AIS fornecesse uma ferramenta de classificação individual de lesões anatômicas, não representava o paciente com múltiplas lesões e sua relação com mortalidade. Em 1974, Baker *et al* ⁽¹⁾ reportaram uma coorte de 2.128 vítimas de acidente automobilístico e suas respectivas lesões, de acordo com a classificação pelo AIS. Observando a relação não linear entre os códigos pelo AIS e a mortalidade observada na coorte, descreveram o *Injury Severity Score*, definido como a soma dos quadrados dos maiores AIS, nas três áreas mais afetadas do corpo.

Embora tenha se tornado difundido, algumas limitações do ISS diminuem sua capacidade preditiva. Mesmo em pacientes com graves lesões restritas a um mesmo segmento corporal, apenas a lesão mais grave é considerada no escore. Além disso, nos pacientes com múltiplos segmentos acometidos, o ISS considera o maior código de um segundo segmento acometido, ainda que exista lesão de maior código AIS no mesmo segmento do corpo. Estas definições permitem que não sejam consideradas necessariamente as lesões anatômicas mais graves sofridas pelo paciente.

Ainda considerando a relação quadrática dos códigos de AIS com a mortalidade observada, foi realizada uma modificação simples do ISS em

1997⁽¹⁰⁾, com a descrição do NISS definido, como a soma dos quadrados dos maiores AIS, independente do segmento corporal acometido. O NISS demonstrou melhor predição de morbidade e mortalidade em diferentes coortes e cenários, sobretudo no trauma contuso (que, caracteristicamente, acomete mais de um segmento corporal), mas também no penetrante⁽¹¹⁻¹⁵⁾.

Do ponto de vista metodológico, uma importante limitação dos escores anatômicos derivados do AIS é sua construção baseada em consenso de especialistas, diferente de modelos derivados de dados cujas predições de mortalidade são pautadas em grandes bancos de dados⁽¹⁶⁾.

Apesar das limitações descritas, o ISS persiste até hoje como um dos escores de trauma mais utilizados mundialmente, sendo com frequência usado, como variável importante de caracterização clínica de pacientes em estudos observacionais ou ensaios clínicos.

3.1.2 Modelos fisiológicos em trauma

A partir da década de 1970, a resposta fisiológica ao trauma passou a ser reconhecida, como importante fator de gravidade e associação com desfechos piores⁽¹⁷⁾. Foram desenvolvidos modelos preditivos que incluíam variáveis fisiológicas relativas à apresentação inicial do paciente, seja em contexto pré-hospitalar/triagem ou na admissão hospitalar, especialmente, levando em conta a importância dos estados de choque, hipoperfusão tecidual e exame neurológico, como determinantes dos desfechos dos pacientes. Além disso, a avaliação da extensão anatômica das lesões pode

ser complexa e não rapidamente disponível em condições de urgente avaliação inicial das vítimas de trauma.

Um dos primeiros descritos, o *Trauma Score* ⁽¹⁸⁾, incluía pressão arterial sistólica (PAS), frequência respiratória (FR), *Glasgow Coma Scale* (GCS), tempo de enchimento capilar e esforço respiratório. Sua posterior modificação, o *Revised Trauma Score* (RTS) ⁽¹⁹⁾, excluiu tempo de enchimento capilar e esforço respiratório por dificuldade de avaliação no atendimento inicial, mostrando melhor predição de mortalidade do que sua primeira versão, sobretudo entre pacientes com traumatismo cranioencefálico. O mais recente *Physiologic Trauma Score*, que incluiu variáveis da síndrome da resposta inflamatória sistêmica combinados com GCS e idade, encontrou acurácia comparável a modelos anatômicos ou combinados mais complexos ⁽²⁰⁾.

3.1.3 Modelos combinados em trauma

Os modelos preditivos combinados unem a gravidade e extensão anatômica das lesões com variáveis fisiológicas que medem a resposta ao trauma, além de características do paciente que modulam essa resposta e que estão associadas ao desfecho, como sexo, idade e comorbidades. Entre os mais populares está o *Trauma Injury Severity Score*, que foi descrito em 1987, combinando idade, RTS e ISS, com equações propostas para traumas contusos e penetrantes ⁽²¹⁾. O *Revised Injury Severity Classification* ⁽²²⁾ e sua posterior atualização ⁽²³⁾ também apresentam boas capacidades preditivas e discriminativas.

Muitas outras características já foram estudadas e incluídas em diversos modelos preditivos combinados, como pupilas e reatividade pupilar, componente motor do GCS, acidose (através do *base excess*, lactato), coagulação (através de *International Normalized Ratio*), hemoglobina, necessidade de transfusão sanguínea, temperatura, parada cardiorrespiratória no atendimento pré-hospitalar, entre outras.

A capacidade preditiva de modelos fisiológicos ou combinados é considerada superior a modelos anatômicos em trauma, conforme demonstrado em várias coortes e metanálises ^(14, 24–26). No entanto, quando considerados pacientes vítimas de trauma com necessidade de admissão em UTI, conforme a proposta desta tese, podem existir limitações ao uso dos escores fisiológicos ou combinados. Se o momento de aplicação desses escores, ao invés de relativo à apresentação clínica na admissão hospitalar, for simplesmente transposto para o momento admissão na UTI, muitas dessas variáveis fisiológicas já refletiriam intervenções terapêuticas do atendimento hospitalar antes da chegada à UTI. Por exemplo: transfusão de hemoderivados, necessidade de procedimento para controle de sangramento, correção de distúrbios hidroeletrólíticos e ácido-básicos já podem ter ocorrido; ou medidas de suporte invasivo como sedação, ventilação mecânica invasiva e uso de droga vasoativa já foram instituídas.

De fato, a descrição dos próprios modelos preditivos acima citados leva em consideração a apresentação inicial do paciente e, dessa forma, não é adequado o uso de variáveis fisiológicas do momento de admissão na UTI para seu cálculo. O mesmo princípio aplica-se ainda no atendimento

inicial, quando algumas dessas variáveis já podem estar manipuladas, como é o caso da FR no paciente sedado e em ventilação mecânica invasiva desde o atendimento pré-hospitalar. Além disso, algumas variáveis são frágeis por sofrerem viés de aferição, notadamente, a FR.

3.2 Modelos prognósticos em doentes críticos

Em UTI, modelos prognósticos têm importante aplicação para quantificar a gravidade clínica e calcular a probabilidade de óbito, com consequente avaliação de resultados e comparação entre diferentes serviços (*benchmarking*) ⁽²⁷⁾. Outra aplicação relevante é para o planejamento da alocação de recursos, considerando o custo financeiro elevado do cuidado em terapia intensiva e frequente baixa disponibilidade de leitos.

Os escores prognósticos usados em terapia intensiva podem ser destinados à população de pacientes críticos em geral ou a subgrupos específicos. Os mesmos também podem ser divididos em escores de gravidade de doença crítica ou escores que reflitam disfunção orgânica.

3.2.1 Modelos prognósticos de gravidade de doença

Os modelos prognósticos relacionados à gravidade de doença mais utilizados são o *Acute Physiology and Chronic Health Evaluation* ⁽²⁸⁾, o *Mortality Probability Model* ⁽²⁹⁾ e o SAPS ^(2, 3).

O escore SAPS 3 combina variáveis antecedentes (idade e comorbidades), variáveis relacionadas à circunstância de admissão na UTI, variáveis fisiológicas e o uso de suportes orgânicos na primeira hora de admissão na UTI para predição de mortalidade hospitalar. O modelo foi atualizado para sua versão SAPS 3 em 2005. Foi desenvolvido com dados de uma coorte de 16.784 pacientes, incluindo UTIs de países em cinco continentes, para os quais foram geradas equações específicas de predição de mortalidade além de uma equação geral ^(2, 3). Permanece como um dos modelos prognósticos mais populares, sobretudo na Europa e no Brasil, sendo utilizado para cálculo de taxa de mortalidade padronizada (TMP) e avaliação dos resultados de UTI.

3.2.2 Modelos prognósticos que reflitam disfunção orgânica

O desenvolvimento de disfunções orgânicas é importante marcador de morbidade e mortalidade em UTI. O escore mais utilizado para quantificar a extensão e gravidade da disfunção orgânica é o SOFA ⁽⁴⁾. Embora não tenha sido descrito originalmente para predição de mortalidade, validações posteriores demonstraram associação do desenvolvimento de disfunção orgânica caracterizada pelo SOFA com piores desfechos em doença crítica em geral, e não somente entre os pacientes com sepse, conforme a descrição original ^(30–33).

Uma potencial vantagem do SOFA em relação aos escores de gravidade clínica, em geral, é a possibilidade de avaliação dinâmica, por meio de suas derivações, como SOFA diário máximo, delta SOFA e SOFA

total máximo na internação, por exemplo, diferente de outros escores que reflitam gravidade clínica, especificamente, no momento da admissão na UTI.

3.3 Recomendações atuais para avaliação da performance preditiva de modelos prognósticos

Modelos preditivos para diagnóstico ou prognóstico de condições clínicas têm se tornado cada vez mais importantes na era da medicina de precisão ou personalizada ⁽³⁴⁾. Uma busca simples pelo termo modelo preditivo na principal base de publicações científicas mostra uma explosão de inserções sobre o tema.

Embora exista um corpo de evidências robusto com recomendações sobre o desenvolvimento e validação de modelos preditivos ^(35, 36), revisões sistemáticas recentes mostram que a maioria dos estudos publicados apresentam falhas metodológicas importantes em sua construção e apresentação ^(37–39).

Muitas vezes, o interesse crescente no tópico, impulsionado pela pressão do meio acadêmico por publicações fez crescer o número de modelos preditivos desenvolvidos, e a maioria deles, além de conterem falhas metodológicas, não chegam a ser aplicados clinicamente. Nesse sentido, antes de poder ser aplicado de maneira adequada, é necessário que a performance preditiva dos modelos seja adequadamente testada em

uma amostra diferente da usada para seu desenvolvimento, etapa conhecida como validação externa.

A performance de um modelo é tipicamente avaliada pela de discriminação e calibração. A discriminação pode ser descrita como a capacidade do modelo preditivo distinguir o paciente que teve o evento do paciente que não teve o evento de interesse. Para eventos binários, é demonstrada pela estatística C, pela área sob a curva *Receiver Operator Characteristic* (AUROC). Esta medida é, frequentemente, reportada por estudos de validação externa, mas do ponto de vista de aplicabilidade clínica, tem interpretação limitada à beira-leito.

A calibração pode ser descrita, como a concordância entre os riscos estimados do desfecho e os observados. Tradicionalmente, é reportada pela proporção de riscos para grupos de pacientes com riscos similares, em geral, divididos em decis de risco, usando *Hosmer–Lemeshow goodness-of-fit test*. No entanto, esta abordagem não vem sendo mais recomendada, porque não indica a direção da perda de calibração, apenas indicado um valor de p das diferenças entre os grupos, que podem ser considerados arbitrários, e o valor de p é muito influenciado pelo tamanho da perda de calibração e tamanho amostral. Em vez disso, a calibração pode ser avaliada graficamente, onde uma predição ideal é avaliava visualmente, além dos valores do intercepto, também conhecidos como alfa ou ainda *calibration-in-the-large* (CITL), que comparam a média dos riscos preditos com a média dos riscos observados; e do valor beta, ou ângulo da curva de calibração ou *slope*, que compara todos os riscos observados com riscos

esperados. A calibração é, curiosamente, pouco reportada nos estudos de validação externa ^(37, 38). A publicação das equações de predição utilizadas também é de suma importância, considerando a reprodutibilidade da validação.

Talvez ainda mais importante do que discriminação e calibração seria uma medida de utilidade clínica, ou seja, o quanto o modelo preditivo adiciona à tomada de decisão, como modificação de condutas e melhora de desfechos. Novos métodos baseados na curva de decisão clínica foram introduzidos e têm sido aplicados sobretudo nos campos de cardiologia e oncologia ⁽⁴⁰⁾.

Outras características metodológicas importantes para realização da validação externa são o tamanho amostral e a quantidade e manejo de dados faltantes, frequentemente subótima na literatura. O tamanho amostral em modelos prognósticos vem sendo classicamente descritos, baseados no número de eventos mínimos para desenvolver ou validar um escore, além do número de eventos por variável incluída no modelo ^(41–44). Embora esses critérios tenham sido descritos para evitar sobreajuste do modelo, eles não são suficientes para a validação. Recentemente, Riley et al. apresentaram uma forma de calcular um tamanho de amostra mínimo que levava em conta não apenas avaliar a discriminação, mas também as medidas de calibração e a precisão dessas estimativas ⁽⁴⁵⁾. Para calcular a amostra desta forma, além de estimar a proporção do desfecho na amostra (i.e., número de eventos), seria necessário levar em conta a distribuição, média e variância do logito do modelo e designar a precisão desejada para o intervalo de

confiança da AUROC e do *slope* de calibração. Nesta tese, conforme descrito adiante, o cálculo amostral leva em conta estas recomendações mais recentes.

4 Métodos

4 MÉTODOS

4.1 Aspectos éticos

O estudo foi aprovado pela Comissão de Ética em Pesquisa do Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo, sob Parecer número 1.905.974, de 2 de fevereiro de 2017. Considerando a natureza observacional retrospectiva do estudo, obteve-se dispensa de Termo de Consentimento Livre e Esclarecido.

4.2 Delineamento do estudo

Estudo de coorte retrospectiva unicêntrica de pacientes vítimas de trauma admitidos na UTI especializada, no período entre maio de 2012 e janeiro de 2016.

4.3 Caracterização da unidade de terapia intensiva e do hospital

O estudo incluiu pacientes admitidos na UTI de Trauma e Emergência Cirúrgica do Departamento de III Clínica Cirúrgica do Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo (FMUSP), unidade de 17 leitos especializada em atendimento de vítimas de trauma. Os pacientes eram admitidos na UTI diretamente da unidade de emergência ou do centro cirúrgico.

O Hospital das Clínicas da FMUSP é referência no atendimento de traumatismo grave na região metropolitana de São Paulo, atendendo uma população estimada em 8.000.000 habitantes, recebendo pacientes por transporte terrestre e aéreo. O atendimento aos pacientes traumatizados dispõe de cirurgia de trauma, radiologia intervencionista, neurocirurgia, ortopedia, cirurgia vascular, plástica e bucomaxilofacial, além de UTI com equipe multidisciplinar dedicada a trauma. O atendimento inicial dos pacientes traumatizados é organizado, de acordo com as diretrizes do *Advanced Trauma Life Support*[®].

4.4 População do estudo

A população do estudo foi definida, de acordo com os critérios abaixo:

4.4.1 Critérios de inclusão

- Pacientes vítimas de trauma com necessidade de admissão em UTI.
- Idade maior ou igual a 15 anos.

4.4.2 Critérios de exclusão

- Pacientes vítimas de mecanismo de trauma por queimadura exclusiva;
- Pacientes vítimas de mecanismo de trauma com lesão hipóxico-isquêmica exclusiva (exemplo: afogamento, enforcamento);

- Paciente vítimas de trauma, cuja causa de admissão na UTI não foi relacionada ao trauma, como causa clínica (exemplo: sepse, tromboembolismo pulmonar).
- Readmissões à UTI.

4.5 Desfechos

O desfecho primário do estudo usado para comparar a performance preditiva dos escores foi a mortalidade hospitalar. Desfechos secundários para caracterização da amostra incluíram mortalidade na UTI, destino na alta da UTI, tempo de estadia na UTI (em dias), destino na alta hospitalar e tempo de estadia no hospital (em dias).

A performance preditiva dos escores foi avaliada pela discriminação, calibração e utilidade clínica, detalhadas a seguir.

4.6 Coleta e manejo de dados

4.6.1 Extração de variáveis do banco de dados

Na época relativa ao período de inclusão do estudo, a UTI de Trauma e Emergência Cirúrgica dispunha de plataforma de coleta de dados personalizada confeccionada pelo programa *Microsoft Access*®, contendo campos estruturados e não estruturados. Criada em fase anterior ao atual sistema de prontuário eletrônico hospitalar, a ferramenta de coleta de dados tinha como objetivo principal o registro de prontuário médico e de passagem

de plantão, com formulários de admissão e evolução diária, que eram, posteriormente, impressos e anexados ao prontuário físico. Como objetivos secundários, era utilizada como instrumento de gerenciamento da unidade e com granularidade que permitisse eventual finalidade de pesquisa. Um *template* dos campos contidos no banco de dados encontra-se no Anexo A.

As seguintes variáveis em relação à admissão dos pacientes foram extraídas do banco de dados:

- dados demográficos (sexo, data de nascimento, idade em anos);
- comorbidades: Diabetes mellitus, insuficiência cardíaca congestiva, doença pulmonar obstrutiva crônica, insuficiência coronariana, insuficiência renal crônica dialítica e não dialítica, neoplasia, infecção por HIV, outros;
- hábitos e vícios: tabagismo, etilismo;
- datas de admissão hospitalar e na UTI;
- causa de admissão na UTI: clínica, cirúrgica trauma, cirúrgica não trauma, readmissão;
- passagem prévia por outro serviço médico ou UTI;
- mecanismo de trauma geral: contuso e penetrante, além de descrição de mecanismos específicos;
- segmentos corporais acometidos no trauma: cranioencefálico, coluna, face, tórax, abdome, pelve, extremidades;
- variáveis relativas ao atendimento pré-hospitalar e ao atendimento inicial no Pronto-socorro: parada cardiorrespiratória, GCS, PAS, FR;

- intervenções terapêuticas recebidas antes da admissão na UTI: necessidade de cirurgia, transfusão de hemocomponentes, uso de droga vasoativa;
- exame físico na admissão à UTI:
 - uso de dispositivos invasivos: intubação orotraqueal;
 - neurológico: sedativos, escala de sedação de *Richmond Agitation Sedation Scale* (RASS), GCS, pupilas;
 - cardiovascular: drogas vasoativas, PAS, pressão arterial diastólica (PAD), pressão arterial média (PAM), Frequência Cardíaca (FC);
 - respiratório: suporte ventilatório, SatO₂;
- exames laboratoriais relativos à admissão na UTI: gasometria venosa e arterial, creatinina, uréia, leucócitos, plaquetas, bilirrubina total; e
- SAPS 3 calculado no momento da admissão e todas as variáveis do escore em campos estruturados.

Foram extraídas do banco de dados as seguintes variáveis da primeira evolução médica dos pacientes, relativas ao primeiro dia de internação na UTI:

- PAM máxima e mínima nas 24h, diurese;
- uso de dispositivos invasivos: intubação orotraqueal;
- neurológico: sedativos, escala de sedação de RASS, GCS, pupilas;
- cardiovascular: drogas vasoativas, PAS, PAD, PAM, FC;
- respiratório: suporte ventilatório, SatO₂;

- renal: necessidade de terapia substitutiva renal; e
- exames laboratoriais: gasometria venosa e arterial, creatinina, ureia, leucócitos, plaquetas e bilirrubina total.

4.6.2 Coleta retrospectiva de dados

Algumas variáveis não estavam disponíveis no banco de dados em sua forma original e foram coletadas retrospectivamente pelos investigadores com base em revisão de prontuário médico.

Os seguintes desfechos hospitalares foram coletados:

- datas de alta da UTI e hospitalar;
- destino na alta da UTI: óbito, enfermaria, transferência para outra UTI, transferência para outro hospital, transferência para instituição de longa permanência (ILP); e
- destino na alta hospitalar: óbito, alta para domicílio, transferência para outro hospital, transferência para ILP.

Com base nos desfechos, foram calculados a mortalidade na UTI e hospitalar e o tempo de estadia na UTI e hospitalar (em dias).

Os escores ISS e NISS foram calculados com base nas lesões anatômicas reportadas em campos não estruturados do banco de dados e na análise dos laudos de tomografias computadorizadas (TC) realizadas durante o atendimento inicial, antes da admissão em UTI. Neste centro de trauma, uma TC de corpo inteiro é indicada em pacientes hemodinamicamente estáveis, vítimas de trauma por mecanismos de trauma de alta energia (ex: quedas de > três metros de altura, acidente de

automóvel ou motocicleta em alta velocidade, ejeção de ocupante do veículo, deformidade importante do veículo, atropelamento, óbito na cena do acidente, entre outros). O exame não era indicado em pacientes hemodinamicamente instáveis com necessidade de intervenção cirúrgica para controle de hemorragia ou intervenção neurocirúrgica de urgência, também a critério dos médicos envolvidos no atendimento primário. Nesse caso, a tomografia de corpo inteiro era frequentemente, realizada após a cirurgia e antes da admissão na UTI, também de acordo com a estabilidade clínica do paciente. Alternativamente, TC de segmentos corporais específicos podem ser realizadas deste modo direcionado (ex: apenas TC crânio), também conforme julgamento clínico.

As lesões anatômicas foram classificadas por médicos independentes cegados aos desfechos dos pacientes, baseada em coleta padronizada após período inicial de treinamento em conjunto. As lesões sofridas foram classificadas de acordo com o AIS em: 1 – leve, 2 – moderada, 3 – grave, 4 – severa, 5 – crítica e 6 – máxima (considerada intratável ou não-sobrevivível), em seis segmentos corporais: cabeça e pescoço, face, tórax, abdome, pelve e extremidades e superfície externa. Uma tabela com a classificação das principais lesões traumáticas, de acordo com o AIS versão 2008 encontra-se no Anexo B. Com base nesses valores, foram calculados:

- ISS: definido como a soma dos quadrados dos maiores AIS, nos três segmentos mais afetados do corpo.
- NISS: definido como a soma dos quadrados dos três maiores AIS, independente qual segmento afetado do corpo.

4.6.3 Derivação de variáveis a partir do banco de dados

A partir de variáveis disponíveis no banco de dados, obteve-se informação para cálculo do escore SOFA total máximo nas primeiras 24 horas de internação na UTI ^(4, 31). Foram atribuídos valores de 0 a 4, de acordo com a gravidade de disfunção orgânica em cada sistema (cardiovascular, respiratório, hepático, renal, neurológico e coagulação), de acordo com a classificação disponível no Anexo C. Foram feitas tabulações cruzadas entre campos estruturados dos formulários de admissão e evolução médica diária para checar a consistência dos dados. Em caso de divergência ou dados faltantes nesses campos, foram checados os campos abertos do prontuário médico, sendo sempre considerados para classificação os piores valores ao longo das primeiras 24 horas de internação na UTI. Caso a informação não estivesse disponível em nenhum dos campos consultados, foram considerados como dado faltante.

Foram feitas tabulações cruzadas entre campos estruturados relativos às comorbidades, categorias do SAPS 3 relativas à presença de comorbidades e revisão dos campos abertos do prontuário médico para derivação do *Charlson Comorbidity Index* ⁽⁴⁶⁾, de acordo com classificação disponível no Anexo D. Em caso de divergência, considerou-se a informação contida nos campos abertos do prontuário.

4.6.4 Revisão e qualidade dos dados

Considerando o desenho retrospectivo do estudo, foi realizada checagem e correção de variáveis categóricas assinaladas com base em

tabulações cruzadas com demais informações disponíveis no banco de dados, inclusive com revisão de campos abertos, de forma a garantir a consistência dos dados. Dessa forma, além das checagens já citadas, foram checadas e adicionadas ou corrigidas classificações em relação a mecanismos de trauma, segmentos corporais acometidos e variáveis relativas ao atendimento pré-hospitalar e ao atendimento inicial no Pronto-socorro.

De forma específica e considerando a análise principal do estudo, todas as variáveis do SAPS 3 foram checadas e corrigidas se necessário, sendo o SAPS 3 recalculado retrospectivamente.

Valores extremos foram revisados e corrigidos quando incorretos. Quando encontrados dados discrepantes, foi realizada revisão das informações contidas em campos abertos do referido banco.

4.7 Análise estatística

O presente estudo foi elaborado e reportado, de acordo com as recomendações do *Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD)* ⁽³⁶⁾. O *checklist* utilizado encontra-se no Anexo E, onde estão detalhadas as páginas e tabelas do texto que contêm as recomendações.

Pelo desenho retrospectivo com base em banco de dados, não foi realizado cálculo formal de tamanho de amostra *a priori*. Considerando a disponibilidade do banco de dados com 1.053 pacientes incluídos no período

do estudo, sendo 251 eventos de interesse, e a proposta de validação externa dos escores prognósticos, a amostra referida contempla o número de 100-200 eventos para atingir poder estatístico adequado, conforme recomendado por alguns autores ⁽⁴¹⁻⁴⁴⁾.

Mais recentemente, foi proposta nova forma de cálculo de amostra para validações externas em modelos preditivos com desfechos binários, com estimativas específicas para discriminação, calibração e utilidade clínica. Este cálculo tem como pressupostos a proporção do evento na população da validação, os erros-padrão para cada estimativa, a calibração esperada e a variância do logito do modelo, além de potenciais limiares de riscos para tomada de decisão clínica ⁽⁴⁵⁾.

De acordo com o cálculo de amostra sugerido por *Riley et al.*, se considerarmos uma distribuição do SAPS 3 com média de 45, variância de 65, *skewness* de 0,5 e *curtose* de 3,0 (que se aproxima à distribuição do SAPS no histograma do artigo original), uma proporção de eventos de 23,7%, uma precisão do intervalo de confiança 95% para a AUROC de 0,1 (+/- 0,05) e para o *slope* de calibração de 0,4 (+/- 0,2), uma amostra de 302 pacientes com 76 eventos de interesse seria adequada para avaliar a discriminação e de 1.083 pacientes com 257 eventos de interesse seria adequada para avaliar a calibração, especialmente, quando algum grau de descalibração fosse esperado. Esta amostra aproxima-se muito deste valor e tem poder adequado inclusive para avaliar a calibração.

Variáveis contínuas foram analisadas como média e desvio padrão ou mediana e intervalo interquartil, conforme distribuição normal. Normalidade

foi avaliada por meio de histogramas e aplicação do teste de Shapiro-Wilk. Variáveis categóricas foram analisadas como frequência e proporção. Variáveis categóricas foram comparadas pelo teste de *Chi* quadrado e quantitativas pelo teste t ou Mann-Whitney, conforme apropriado.

De acordo com recomendações recentes ^(36, 43, 47–49), a performance preditiva dos modelos foi avaliada pela discriminação, calibração e utilidade clínica.

Discriminação refere-se à capacidade do modelo diferenciar o paciente que teve o desfecho (óbito hospitalar) daquele que não teve o desfecho (sobrevivente na alta hospitalar) e foi avaliada pela estatística C da AUROC. A comparação entre as AUROC foi feita pelo método de DeLong ⁽⁵⁰⁾. Foram realizadas também comparações individuais entre os escores, com cálculo de alfa de Cronbach com intervalo de confiança (IC) 95% pelo *bootstrap* com 100 repetições para avaliar a concordância entre os modelos. O escore de Brier foi calculado para avaliar a acurácia dos modelos, com valores entre 0 e 1, sendo os valores mais baixos quanto maior a acurácia.

A calibração refere-se à concordância entre desfechos observados e desfechos preditos. Foi avaliada pelo gráfico de calibração com método de alisamento linear local (Lowess). O CITL, compara a média dos riscos preditos com a média dos riscos observados, tem-se como valor ideal 0; no ângulo da curva de calibração ou *slope*, que compara todos os riscos observados com riscos esperados, tem-se como valor ideal 1. Nos gráficos, também observam-se grupos por decis de predição de risco do evento. Uma calibração imperfeita é avaliada visualmente pelo desvio dos valores ideais.

Também foi calculada a razão esperado/observado (E/O), com valor ideal de 1.

A utilidade clínica foi avaliada pela curva de decisão, considerando benefício líquido ao longo de diferentes limiares de probabilidade do desfecho, comparados com a estratégia tratar todos os pacientes que tiveram desfecho e tratar nenhum dos pacientes que não tiveram o desfecho ⁽⁴⁰⁾.

Para a análise da calibração e da curva de decisão, foi necessário determinar as probabilidades de mortalidade hospitalar. Para a predição de mortalidade, a partir dos escores, foram calculados o logito e a probabilidade de óbito esperada, conforme as equações originais Geral e América Central/Sul do SAPS 3 ^(2, 3). Também foi realizada recalibração do intercepto e do *slope* do SAPS 3. Para os escores SOFA, ISS e NISS, que não têm equações de predição de mortalidade descritas originalmente, foi aplicada a regressão logística para predição do desfecho na coorte para derivação da equação. Portanto, são apresentadas a validação aparente de SOFA, ISS, NISS e SAPS 3 recalibrados e a validação externa do SAPS 3 Geral e América Central/Sul.

Como análise secundária, entre os escores de melhor discriminação e calibração na análise principal, foi realizada a avaliação da performance preditiva dos modelos, estratificada de acordo com um subgrupo de interesse clínico, definido pela presença de TCE grave. Esta subpopulação foi selecionada por sua importância clínica na prática e pela possibilidade de

os escores apresentarem performance preditiva diferente de uma população geral de pacientes com trauma na UTI.

A análise estatística foi feita usando o *software* STATA®, versão 16⁽⁵¹⁾, incluindo pacotes de contribuição de usuários pmcalplot, brier e dca⁽⁴⁰⁾. Foi considerado significativo um valor de $p < 0,05$ e não foram realizados ajustes para múltiplas comparações.

4.7.1 Manejo de dados faltantes

Como estratégia inicial para o manejo de dados faltantes, foi realizada a revisão de informações em campos abertos do banco de dados, conforme explicado anteriormente, de modo a completar os dados com informações existentes no banco de dados, antes de considerá-los, realmente, faltantes.

Como a proporção de dados faltantes foi pequena ($< 5\%$), a análise primária foi realizada sobre os casos completos⁽⁵²⁾. Como análise de sensibilidade foi realizada imputação múltipla dos dados faltantes, assumindo que os dados foram faltantes ao acaso. A imputação múltipla foi realizada com o método de equações encadeadas, incluindo o desfecho, as variáveis exploradas faltantes (SAPS 3 e SOFA) e as variáveis auxiliares que poderiam ajudar a prever as variáveis exploradas faltantes (idade, GCS, ISS e NISS, Charlson). Foram realizadas dez imputações, utilizando o método de pareamento preditivo médio para o cálculo das imputações. Após, os dados foram analisados, conforme as regras de Rubin para bancos de dados imputados para corrigir os erros-padrão⁽⁵³⁾.

5 Resultados

5 RESULTADOS

No período entre maio de 2012 e janeiro de 2016, foram admitidos 1.984 pacientes na UTI. Destes, 1.053 pacientes foram admitidos por trauma e incluídos na coorte do estudo, conforme os critérios de elegibilidade (Figura 1). Houve 50 pacientes com dados faltantes nos escores SAPS 3 ou SOFA total nas primeiras 24 horas de admissão, sendo 1.003 pacientes incluídos na análise de casos completos. Não houve dados faltantes em relação aos desfechos analisados.

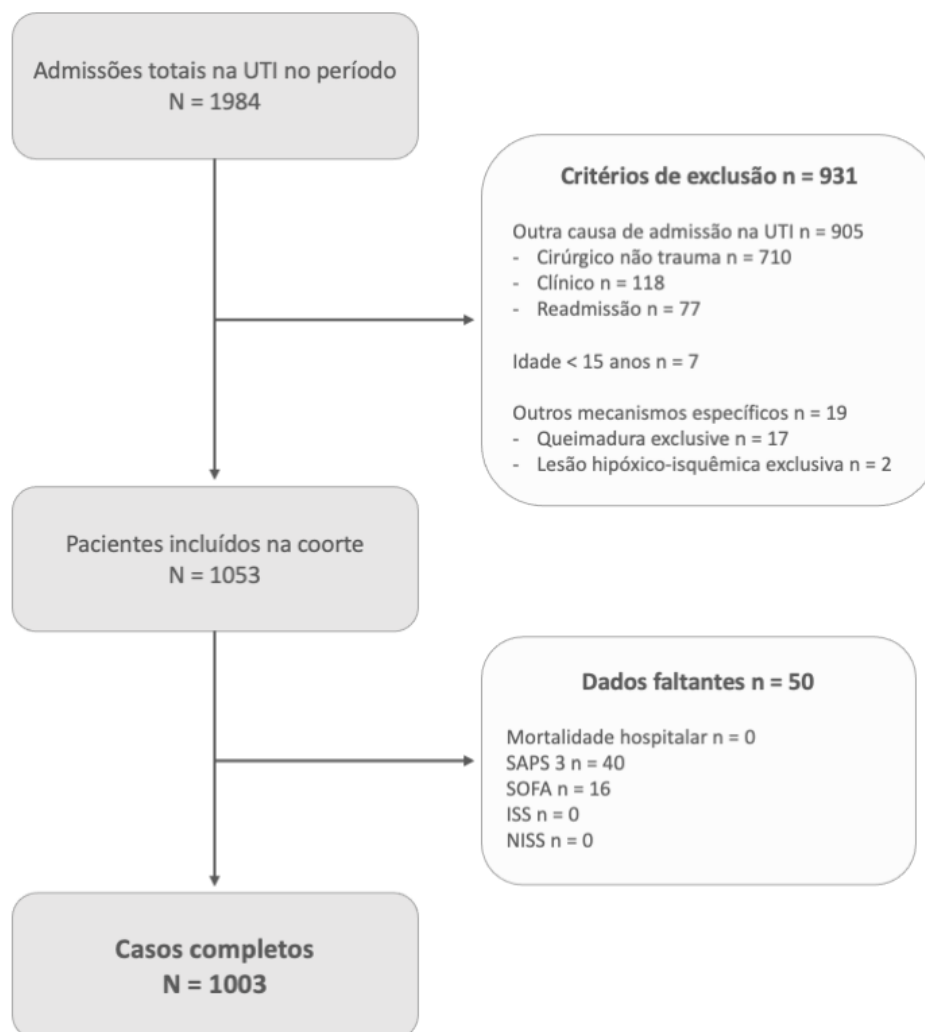


Figura 1 - Fluxograma do estudo

As características principais dos pacientes incluídos na coorte encontram-se nos dados da Tabela 1. A maioria dos pacientes era do sexo masculino (84,2%), com idade média de 40 (± 18) anos. O mecanismo principal de trauma foi contuso (90,7%), e acidentes de trânsito foram os mecanismos específicos mais frequentes (Tabela 2). Traumatismo cranioencefálico esteve presente em 67,8% dos pacientes, sendo grave em 43,3%. Os segmentos corporais acometidos no trauma encontram-se nos dados da Tabela 3. Tomografia de corpo inteiro foi realizada em 62,7% dos pacientes. No momento da admissão na UTI, 846 pacientes (80,3%) estavam em ventilação mecânica invasiva e 644 (64,3%) em uso de droga vasoativa. Óbito hospitalar ocorreu em 251 pacientes (23,8%). A mediana do tempo de estadia na UTI foi 8 dias; e hospitalar, 17 dias. Em relação aos escores prognósticos, a mediana do SAPS 3 foi 41; do SOFA total máximo nas 24 horas de admissão, 7; do ISS, 29; e do NISS, 41.

Tabela 1 - Características dos pacientes

	Coorte completa N = 1053	Sobreviventes N = 802 (76,2%)	Não Sobreviventes N = 251 (23,8%)	Valor de p
Idade, média (DP)	40 (18)	38 (16)	46 (20)	< 0,001
Sexo masculino, n (%)	887 (84,2%)	693 (86,4%)	194 (77,3%)	< 0,001
Charlson, mediana [IIQ]	0 [0, 0]	0 [0, 0]	0 [0, 0]	0,026
Comorbidades				0,009
Nenhuma, n (%)	907 (86,1%)	703 (87,7%)	204 (81,3%)	
Uma, n (%)	88 (8,4%)	64 (8%)	24 (9,6%)	
Duas ou mais, n (%)	58 (5,5%)	35 (4,4%)	23 (9,2%)	
Mecanismo de trauma				0,020
Contuso, n (%)	955 (90,7%)	718 (89,5%)	237 (94,4%)	
Penetrante, n (%)	98 (9,3%)	84 (10,5%)	14 (5,6%)	
Trauma cranioencefálico, n (%)	714 (67,8%)	512 (63,8%)	202 (80,5%)	< 0,001
TCE grave, n (%)	452 (43,3%)	293 (36,8%)	159 (63,9%)	< 0,001
TC corpo inteiro, n (%)	660 (62,7%)	499 (62,2%)	161 (64,1%)	0,58
Ventilação mecânica, n (%)	846 (80,3%)	604 (74,8%)	246 (97,6%)	< 0,001
Droga vasoativa, n (%)	677 (64,3%)	446 (55,6%)	232 (92%)	< 0,001
SAPS 3, mediana [IIQ]	41 [29, 53]	38 [26, 48]	53 [44, 64]	< 0,001
SOFA, mediana [IIQ]	7 [4, 10]	6 [4, 9]	11 [8, 13]	< 0,001
ISS, mediana [IIQ]	29 [25, 36]	25 [22, 34]	34 [25, 41]	< 0,001
NISS, mediana [IIQ]	41 [29, 50]	41 [27, 50]	50 [41, 57]	< 0,001
Tempo de estadia na UTI, dias mediana [IIQ]	8 [4, 16]	8 [4, 17]	6 [4, 12]	< 0,001
Tempo de estadia hospitalar, dias mediana [IIQ]	17 [9, 34]	22 [12, 40]	7 [5, 13]	< 0,001

Variáveis contínuas apresentadas como média e (DP) ou mediana e [IIQ], conforme apropriado. Variáveis categóricas apresentadas como frequência n e (%).

TCE: traumatismo cranioencefálico; TC: tomografia computadorizada; SAPS 3: Simplified Acute Physiology Score 3; SOFA: Sequential Organ Failure Assessment; ISS: Injury Severity Score; NISS: New Injury Severity Score.

Houve dados faltantes nas seguintes variáveis: idade (n=12), TCE grave (n=8), SAPS 3 (n=40), SOFA (n=16).

Tabela 2 - Mecanismos específicos de trauma

	Coorte completa N = 1053	Sobreviventes N = 802 (76,2%)	Não Sobreviventes N = 251 (23,8%)
Mecanismos contusos	955 (90,7%)	718 (89,5%)	237 (94,4%)
Trânsito – automóvel	114 (10,8)	94 (11,7%)	20 (8%)
Trânsito – motocicleta	208 (19,8%)	172 (21,4%)	36 (14,3%)
Trânsito – bicicleta	14 (1,3%)	12 (1,5%)	2 (0,8%)
Trânsito – atropelamento	248 (23,6%)	182 (22,7%)	66 (26,3%)
Trânsito – outros	2 (0,2%)	1 (0,1%)	1 (0,4%)
Queda de alta energia (≥ 3 m)	167 (15,9%)	111 (13,8%)	56 (22,3%)
Queda de baixa energia (< 3 m)	110 (10,4%)	80 (10%)	30 (12%)
Agressão	40 (3,8%)	31 (3,9%)	9 (3,6%)
Soterramento	13 (1,2%)	9 (1,1%)	4 (1,6%)
Outros mecanismos contusos	18 (1,7%)	14 (1,7%)	4 (1,6%)
Mecanismo contuso desconhecido	21 (2%)	12 (1,5%)	9 (3,6%)
Mecanismos penetrantes	98 (9,3%)	84 (10,5%)	14 (5,6%)
Ferimento por arma de fogo	67 (6,4%)	55 (6,9%)	12 (4,8%)
Ferimento por arma branca	29 (2,8%)	27 (3,4%)	2 (0,8%)
Outros mecanismos penetrantes	2 (0,2%)	2 (0,2%)	0 (0%)

m: metros

Tabela 3 - Segmentos corporais acometidos

	Coorte completa N = 1053	Sobreviventes N = 802 (76,2%)	Não Sobreviventes N = 251 (23,8%)	Valor de p
Trauma cranioencefálico	714 (67,8%)	512 (63,8%)	202 (80,5%)	< 0,001
Trauma de coluna	175 (16,6%)	141 (17,6%)	34 (13,5%)	0,13
Trauma de face	269 (25,5%)	203 (25,3%)	66 (26,3%)	0,76
Trauma de tórax	319 (30,3%)	242 (30,2%)	77 (30,7%)	0,88
Trauma de abdome	233 (22,1%)	173 (21,6%)	60 (23,9%)	0,44
Trauma de pelve	132 (12,5%)	92 (11,5%)	40 (15,9%)	0,062
Trauma de extremidades	382 (36,3%)	308 (38,4%)	74 (29,5%)	0,010

Nas Figuras 2 a 5 são apresentadas as distribuições dos escores prognósticos avaliados, estratificados pelo desfecho hospitalar. Os escores apresentaram distribuição unimodal. O SAPS 3 e o SOFA total máximo em 24 horas de admissão apresentaram melhor diferenciação entre as distribuições, de acordo com o desfecho hospitalar.

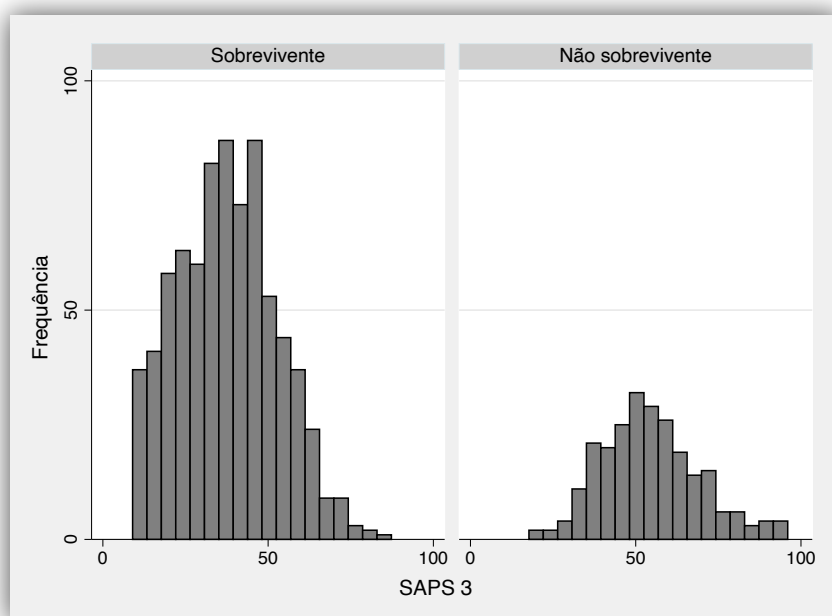


Figura 2 - Histograma do SAPS 3, estratificado por desfecho hospitalar

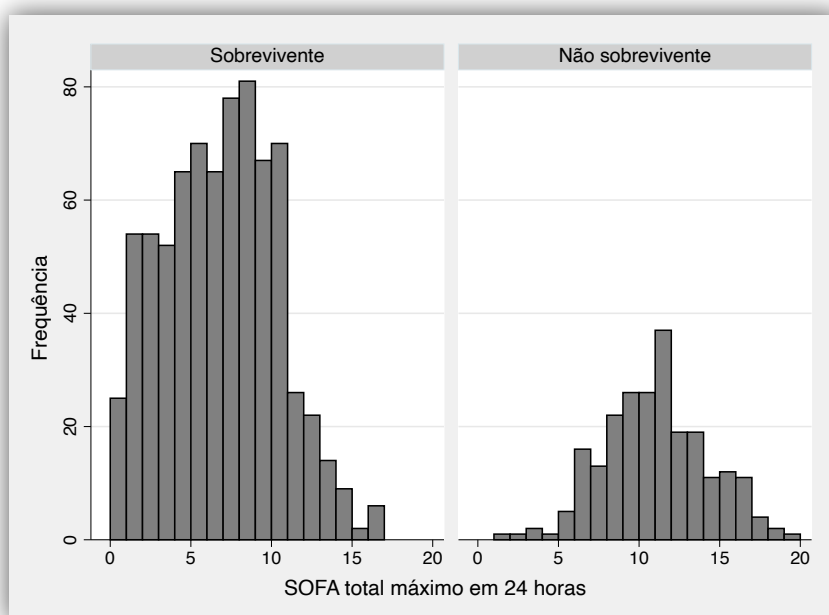


Figura 3 - Histograma do SOFA total máximo em 24 horas, estratificado por desfecho hospitalar

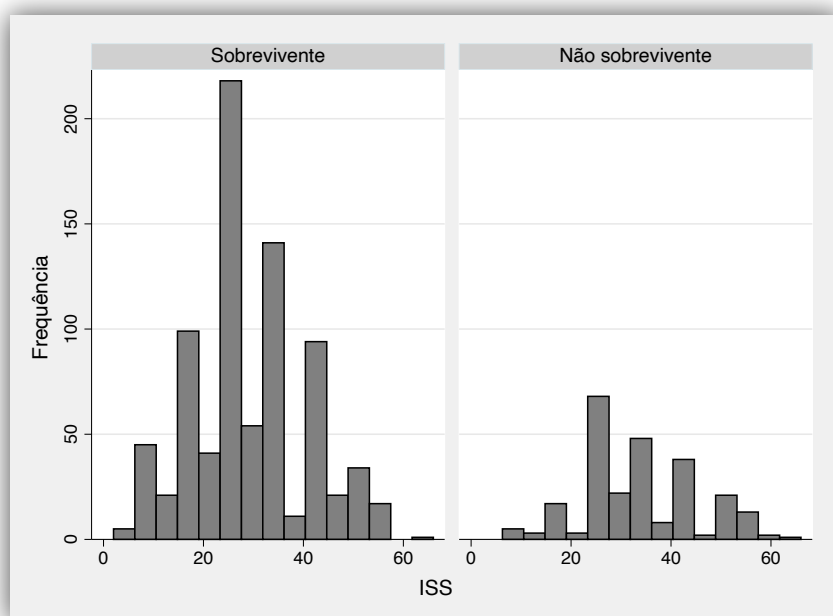


Figura 4 - Histograma do ISS, estratificado por desfecho hospitalar

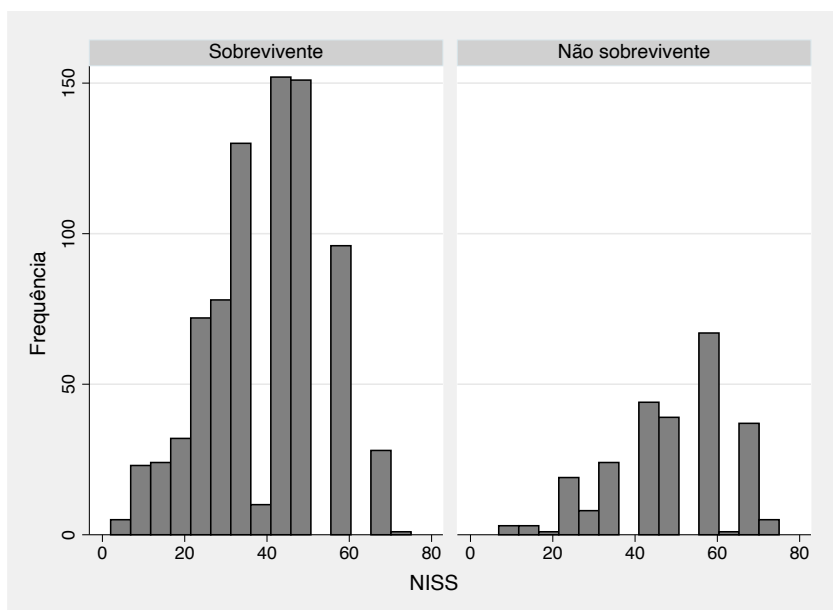


Figura 5 - Histograma do NISS, estratificado por desfecho hospitalar

5.1 Discriminação dos escores prognósticos

A discriminação dos modelos é apresentada na Figura 6. Foram obtidas as seguintes AUROC: SAPS 3 = 0,786 (CI 95% 0,756-0,817), SOFA total máximo em 24 horas de admissão = 0,807 (CI 95% 0,778-0,837), ISS = 0,616 (CI 95% 0,577-0,656) e NISS = 0,689 (CI 95% 0,649-0,729) com $p < 0,001$ em relação a um modelo não-discriminativo.

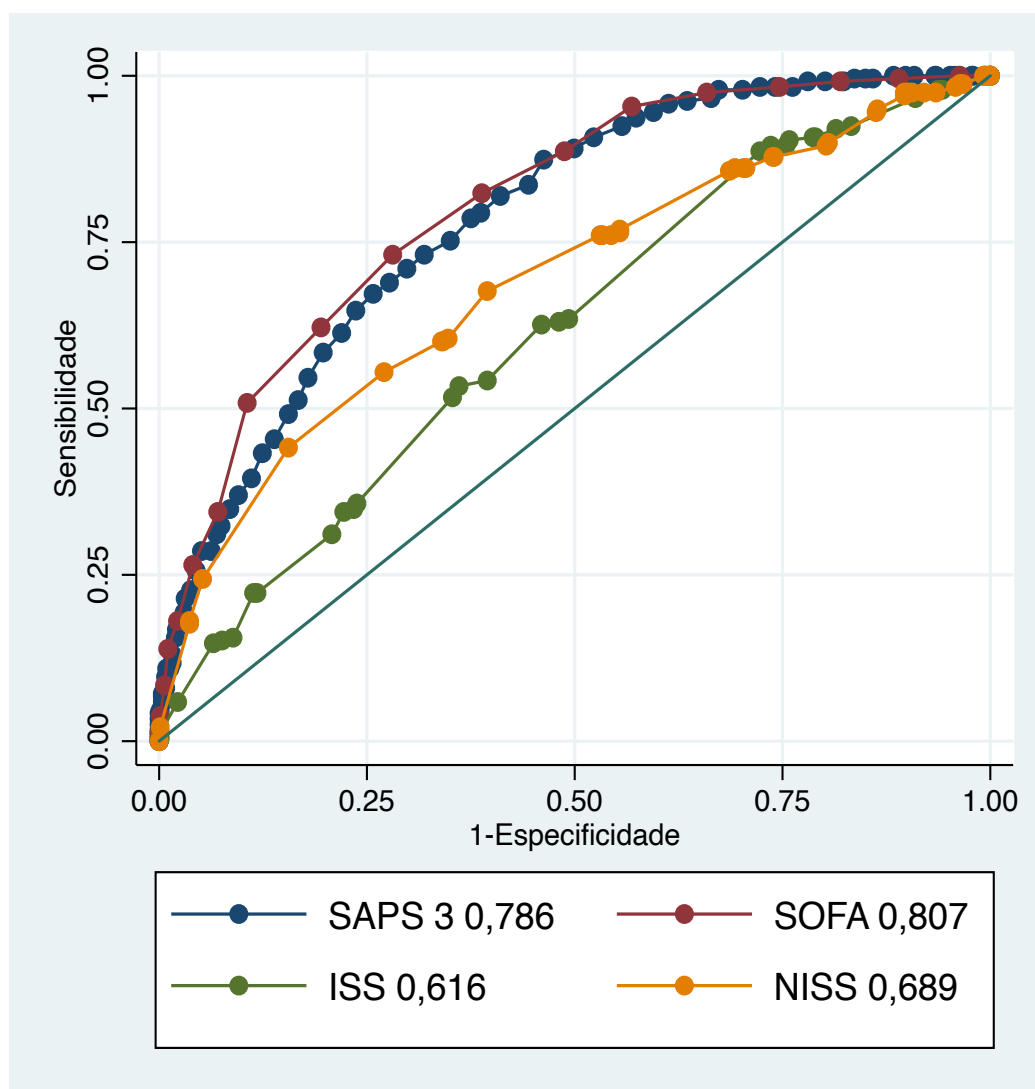


Figura 6 - Comparação de AUROC dos escores prognósticos

Os dados da Tabela 4 apresentam a comparação individual das AUROC, bem como a consistência entre os escores, medida pelo alfa de Cronbach. Nas comparações individuais, a discriminação do SAPS 3 e do SOFA não foi diferente, enquanto ambos os escores apresentaram melhor discriminação do que os escores anatômicos. Por outro lado, o NISS apresentou melhor discriminação do que o ISS. Os escores anatômicos apresentam boa consistência interna entre si, o que não foi observado entre os demais escores.

Tabela 4 - Comparações individuais de AUROC entre os escores

	SAPS 3	SOFA	ISS	NISS
SAPS 3	0,786 (0,756-0,817)	0,452 (0,429-0,476)	0,413 (0,352-0,473)	0,458 (0,397-0,523)
SOFA	0,198	0,807 (0,778-0,837)	0,313 (0,258-0,368)	0,259 (0,213-0,301)
ISS	<0,001	<0,001	0,616 (0,577-0,656)	0,802 (0,778-0,825)
NISS	<0,001	<0,001	<0,001	0,689 (0,649-0,729)

Nas diagonais em negrito, são apresentados os resultados da AUROC de cada escore, com IC 95% entre parênteses. Nas células amarelas inferiores, os valores de p das comparações individuais das AUROC entre os escores. Nas células azuis superiores, os valores do alfa de Cronbach de consistência entre os escores com os IC criados com 100 replicações por *bootstrap*.

SAPS 3: Simplified Acute Physiology Score 3; SOFA: Sequential Organ Failure Assessment; ISS: Injury Severity Score; NISS: New Injury Severity Score.

5.2 Calibração dos escores prognósticos

As equações de predição são apresentadas nos dados da Tabela 5 e os gráficos de calibração são apresentados nas Figuras 7 a 10.

Tabela 5 - Equações de predição de mortalidade, escore de Brier e medidas de calibração dos escores

Escore	Equação	Brier	E / O	CITL	Slope
SAPS 3, EG	Logit = $-32,6659 + \ln(\text{SAPS } 3 + 20,5958) * 7,3068$	0,158	0,63	0,82	0,68
SAPS 3, ACS	Logit = $-64,5990 + \ln(\text{SAPS } 3 + 71,0599) * 13,2322$	0,152	0,83	0,35	0,66
SAPS 3, RC	Logit = $-21,9761 + \ln(\text{SAPS } 3 + 20,5958) * 4,9831$	0,147	1,00	0,00	1,00
SOFA	Logit = $-4,0995 + (\text{SOFA}) * 0,3480$	0,139	1,01	-0,02	0,99
ISS	Logit = $-2,2814 + (\text{ISS}) * 0,0359$	0,177	1,00	0,01	0,95
NISS	Logit = $-3,3590 + (\text{NISS}) * 0,0503$	0,165	1,00	0,01	1,00

Os resultados de calibração de SOFA, ISS e NISS são medidas de validação aparente, considerando que a ausência de equação previamente descrita impede validação externa completa. A interpretação pode ser prejudicada e os gráficos de calibração (Figuras 7 a 12) podem auxiliar nessa avaliação.

E/O: razão esperado por observado; CITL: Calibration in the large; Brier: escore de Brier.

SAPS 3: Simplified Acute Physiology Score 3; SOFA: Sequential Organ Failure Assessment; ISS: Injury Severity Score; NISS: New Injury Severity Score; EG: Equação geral; ACS: América Central/Sul; RC: Equação recalibrada.

O SAPS 3 Equação geral e Equação América Central/Sul apresentaram calibração imperfeita. Para o SAPS 3 Equação geral, o risco predito de óbito hospitalar foi sistematicamente menor do que a mortalidade observada nesta amostra, especialmente, na faixa mais baixa de risco do desfecho (Figura 7).

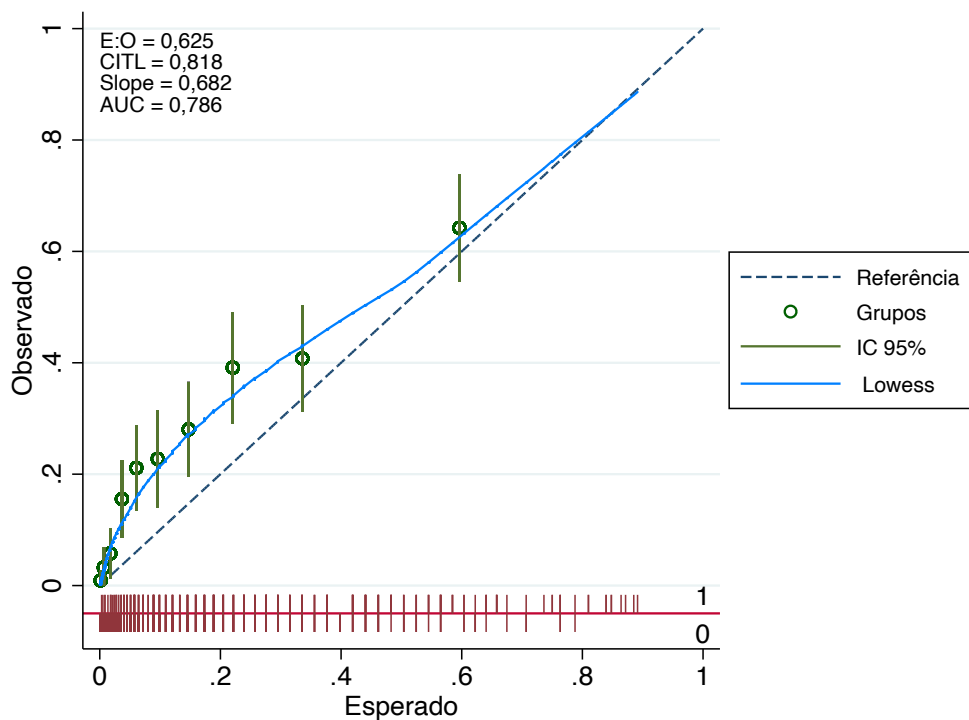


Figura 7 - Gráfico de calibração do SAPS 3

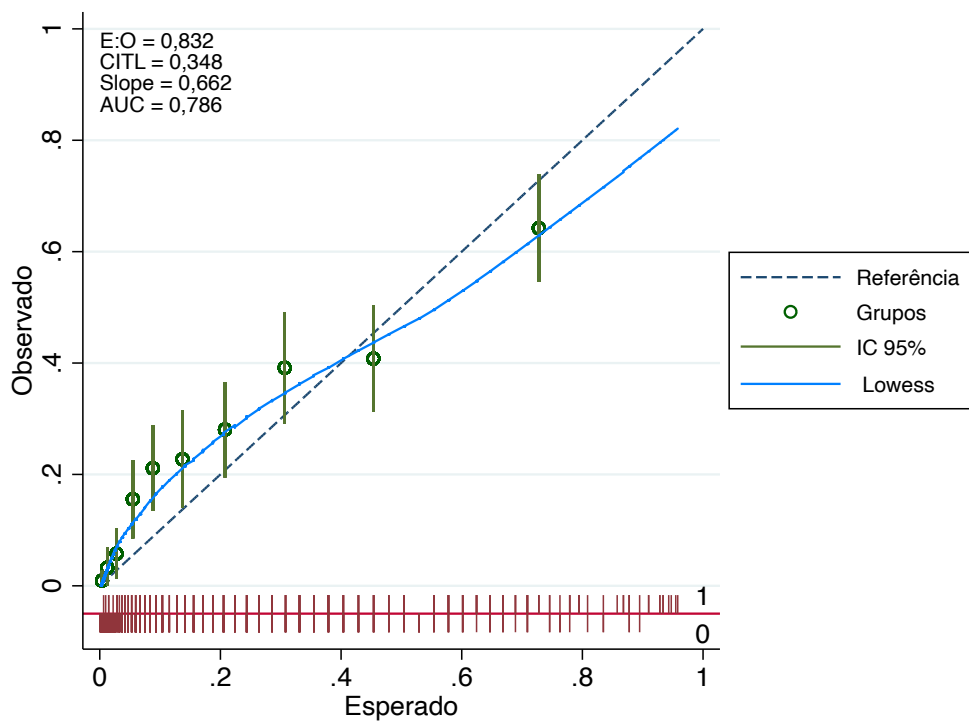


Figura 8 - Gráfico de calibração do SAPS 3 Equação América Central/Sul

Após recalibração do intercepto e *slope* da equação, o SAPS 3 Equação recalibrada apresentou melhor calibração ao longo de todo o limiar de risco do desfecho, semelhante ao observado com SOFA (Figuras 9 e 10). Ambos SAPS 3 e SOFA superestimaram a mortalidade no estrato de mais alto risco do desfecho.

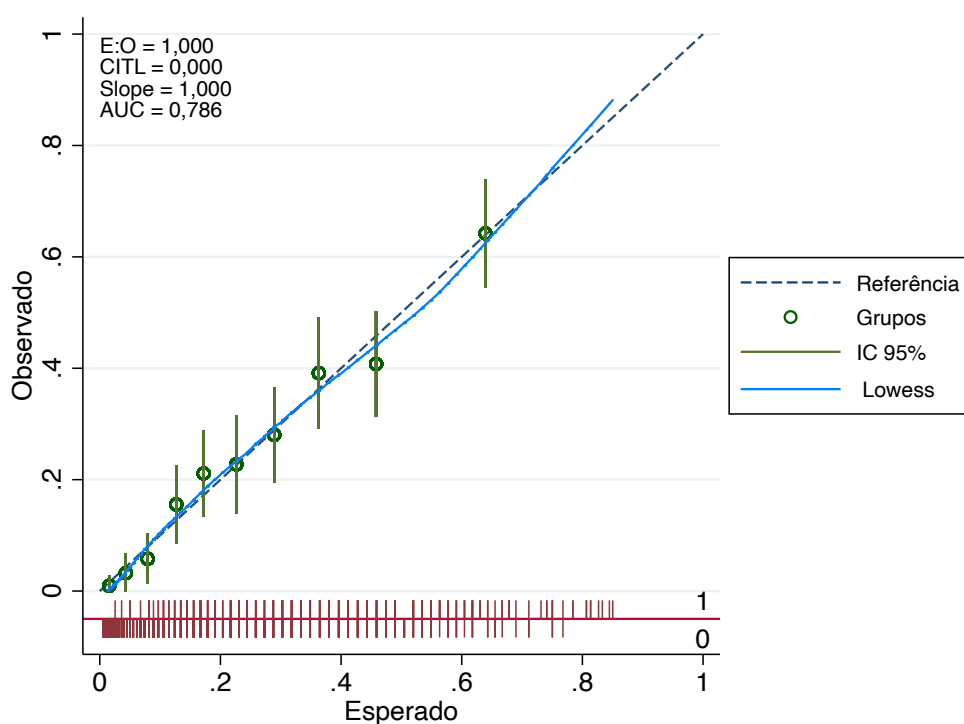


Figura 9- Gráfico de calibração do SAPS 3 Equação recalibrada

O SOFA total nas primeiras 24 horas de admissão na UTI apresentou calibração próxima de ideal, ao longo de ampla faixa de predição de risco (Figura 10). Nas probabilidades altas de evento, apresentou discreta perda de calibração, predizendo riscos mais elevados do que os observados.

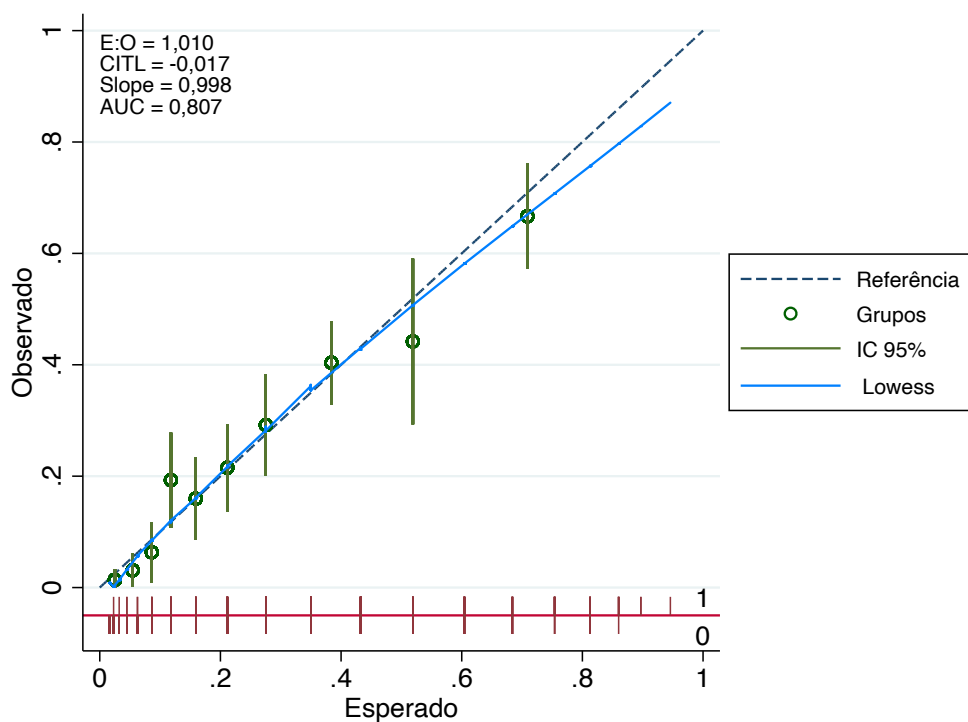


Figura 10 - Gráfico de calibração do SOFA total máximo nas 24 horas de admissão

ISS e NISS apresentaram calibração adequada, porém com faixas estreitas de predição de risco (Figuras 11 e 12).

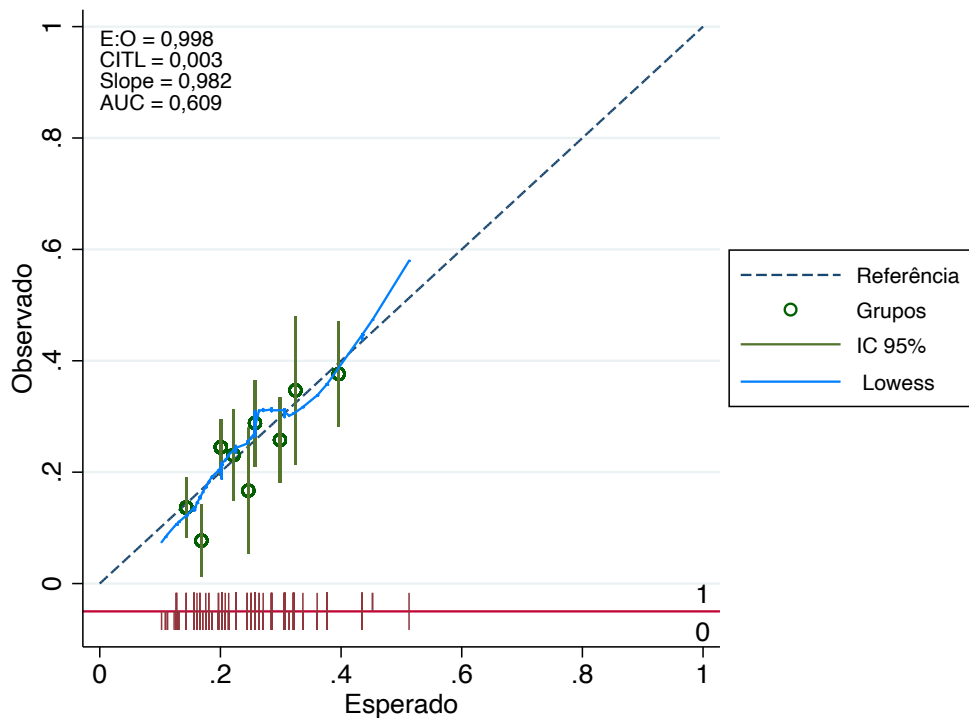


Figura 11 - Gráfico de calibração do ISS

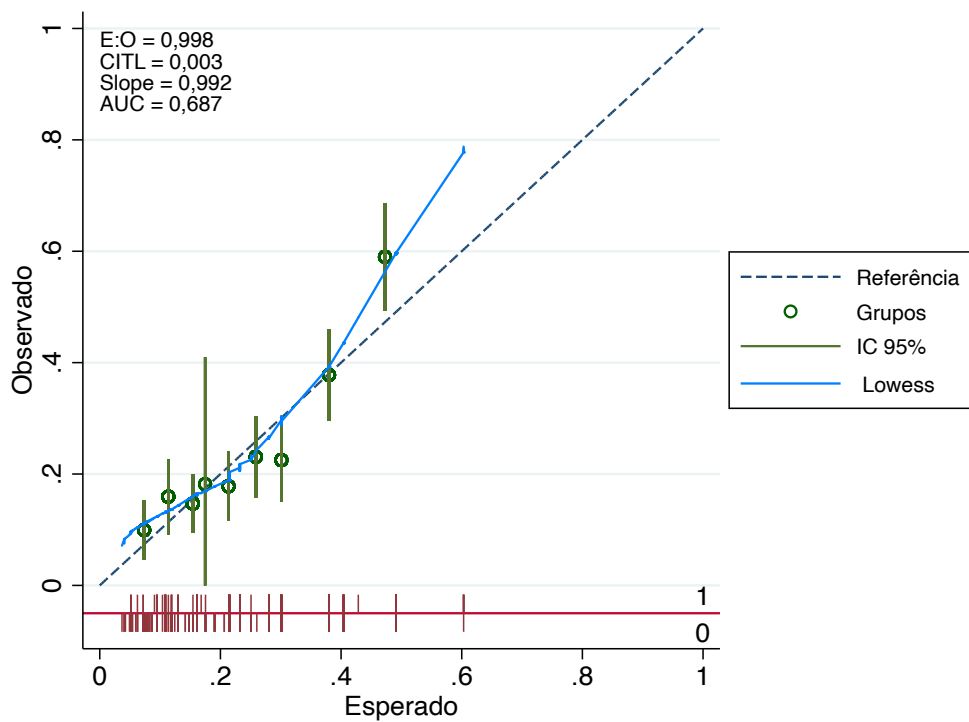


Figura 12 - Gráfico de calibração do NISS

Na Figura 13, estão apresentados os gráficos de calibração de todos os escores analisados, para fins de comparação.

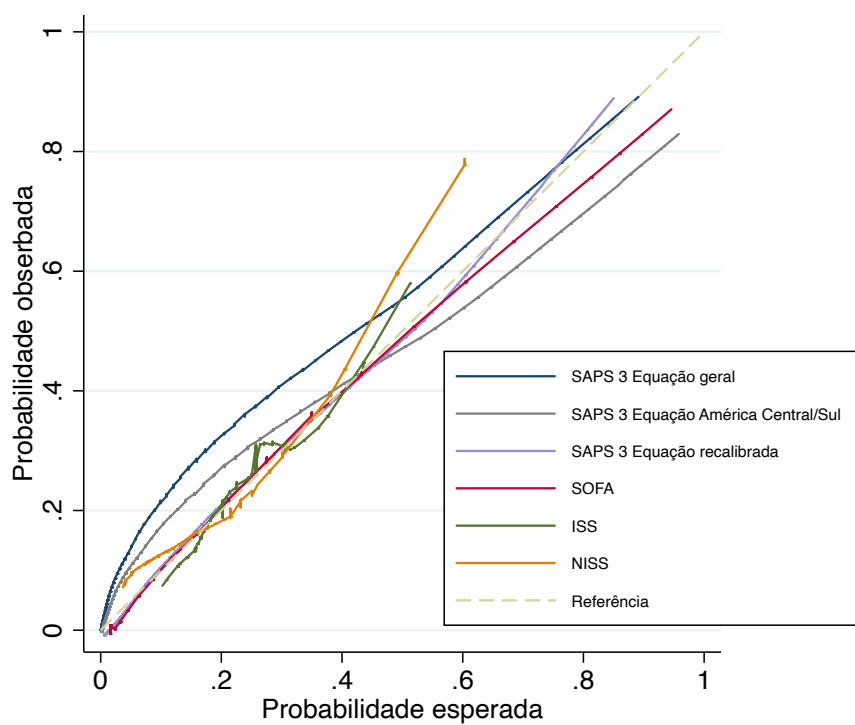


Figura 13 - Gráfico de calibração comparando todos os escores

5.3 Utilidade clínica dos escores prognósticos

As curvas de utilidade clínica dos escores prognósticos são apresentadas na Figura 13. O SAPS 3 e SOFA total máximo nas 24 horas de admissão na UTI apresentaram os maiores benefícios líquidos, quando comparados com os demais escores, independente do limiar de probabilidades. Ambos demonstraram maior benefício líquido na faixa intermediária de probabilidade dos desfechos, sobretudo o SOFA, de 10 a 50% de mortalidade hospitalar. ISS apresentou o menor benefício líquido na curva de utilidade clínica.

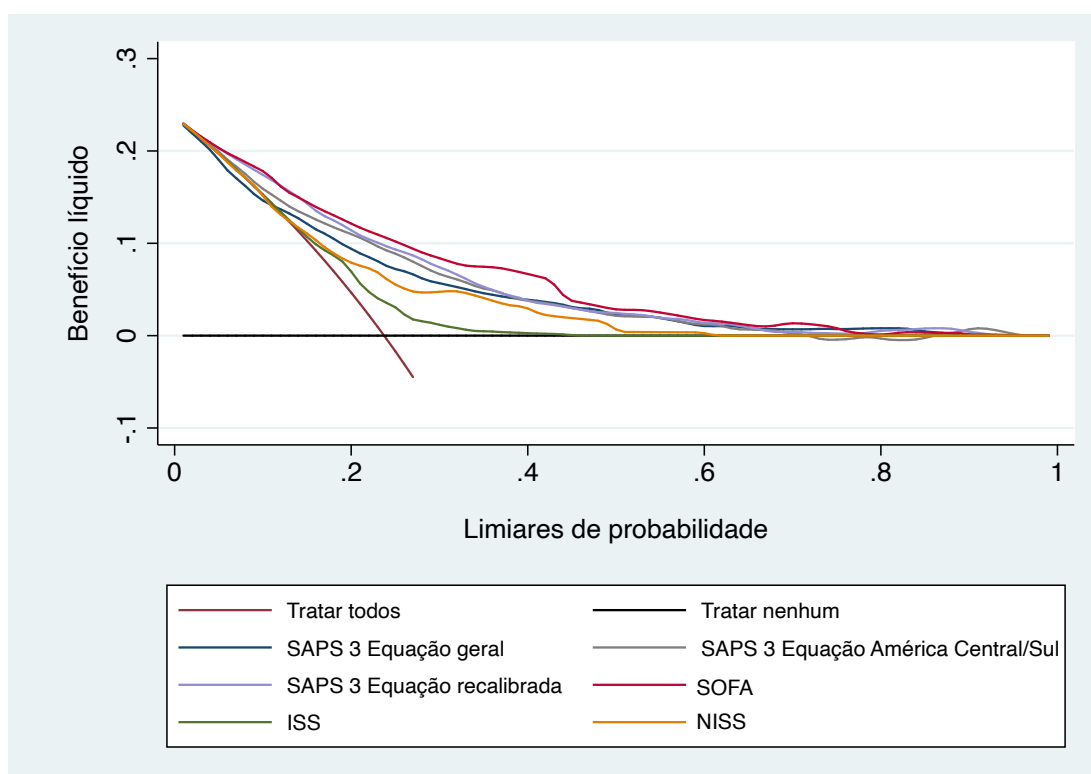


Figura 14 - Curvas de utilidade clínica dos escores

5.4 Análise estratificada de acordo com a presença de traumatismo cranioencefálico grave

A análise de subgrupo de acordo com a presença de TCE grave é apresentada nos dados da Tabela 6. De maneira geral, houve diminuição na acurácia em todos os modelos, com redução das AUROCs e aumento do escore de Brier.

O SOFA permaneceu o modelo mais discriminativo, sem grandes modificações na calibração.

Tabela 6 - Análise de subgrupo, de acordo com a presença de TCE grave

Escore	AUROC	Brier	E / O ¹	CITL ¹	Slope ¹
SAPS 3, EG	0,718	0,215	0,63	0,85	0,64
SAPS 3, ACS	0,718	0,204	0,83	0,37	0,60
SAPS 3, RC	0,718	0,196	0,94	0,12	0,94
SOFA	0,769	0,183	1,00	0,00	1,00
ISS	0,586	0,222	1,00	0,00	1,00
NISS	0,693	0,204	1,00	0,00	1,00

¹ Os resultados de calibração de SOFA, ISS e NISS são medidas de validação aparente, considerando que a ausência de equação previamente descrita impede validação externa completa. A interpretação pode ser prejudicada e os gráficos de calibração podem auxiliar nessa avaliação.

TCE: traumatismo cranioencefálico; E/O: razão esperada por observado; CITL: Calibration in the large; AUROC: area under the receiving operator curve; Brier: escore de Brier; SAPS 3: Simplified Acute Physiology Score 3; SOFA: Sequential Organ Failure Assessment; ISS: Injury Severity Score; NISS: New Injury Severity Score; EG: Equação geral; ACS: América Central/Sul; RC: Equação recalibrada.

Resultados apresentados para análise de casos completos de paciente com TCE grave (429/452).

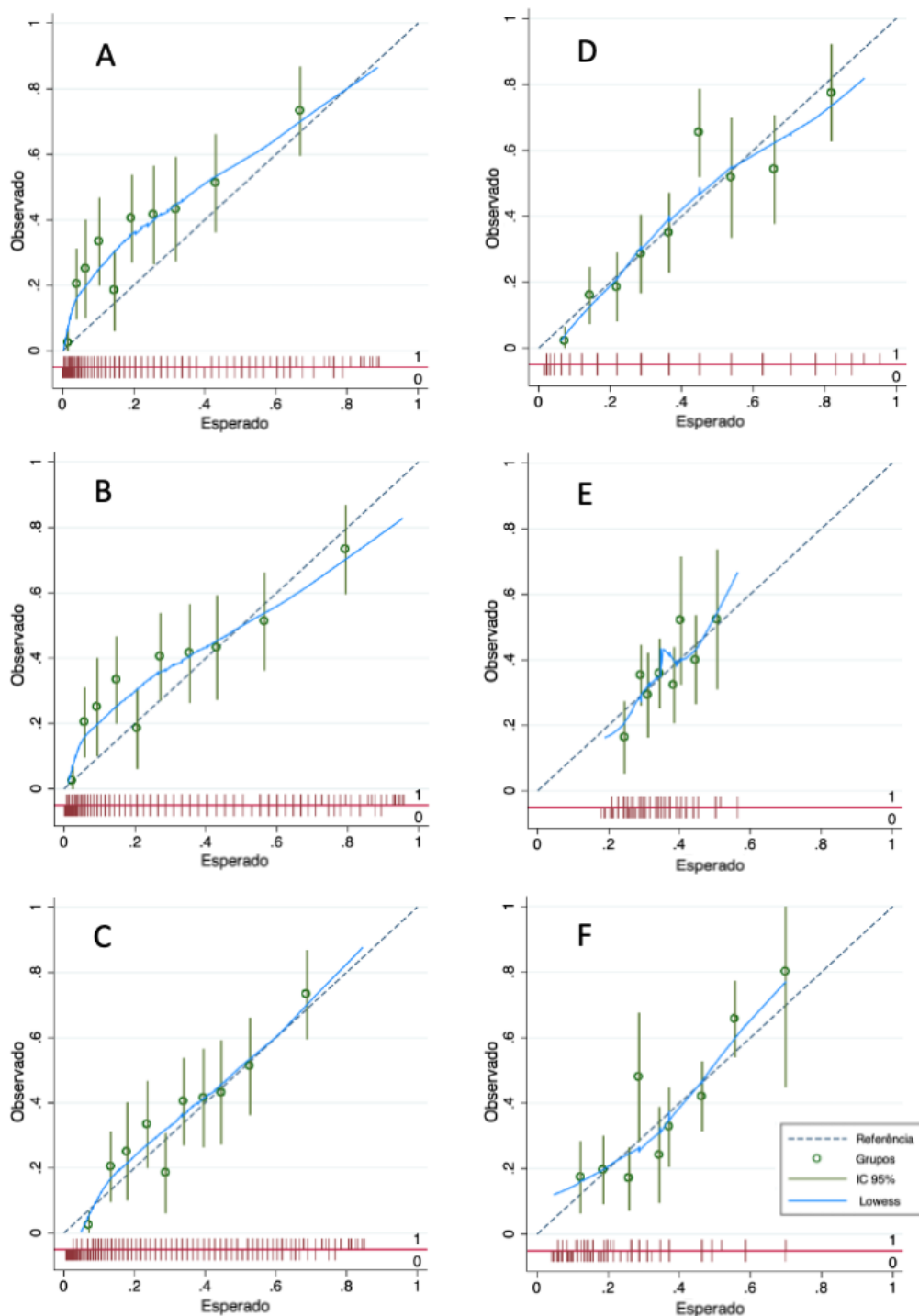


Figura 15 - Gráficos de calibração em pacientes com TCE grave. A: SAPS 3 Equação Geral; B: SAPS 3 Equação América Central/Sul; C: SAPS 3 Equação recalibrada; D: SOFA; E: ISS; F: NISS

5.5 Análise de sensibilidade com imputação múltipla

Após a imputação múltipla de dados faltantes para SAPS 3 e SOFA total máximo em 24 horas de admissão, não houve mudança significativa nos resultados de discriminação, calibração e curva de decisão (Tabela 7).

Tabela 7 - Análise de sensibilidade, após imputação múltipla¹ de dados faltantes nos escores prognósticos

Escore	AUROC	Brier	E / O ²	CITL ²	Slope ²
SAPS 3, EG	0,79	0,157	0,62	0,82	0,69
SAPS 3, ACS	0,79	0,152	0,83	0,35	0,67
SAPS 3, RC	0,79	0,146	1,00	0,01	1,00
SOFA	0,81	0,140	1,00	0,00	1,01
ISS ³	0,61	0,177	1,00	0,00	1,00
NISS ³	0,69	0,165	1,00	0,00	1,00

¹ A imputação múltipla foi realizada com o método de equações encadeadas, incluindo o desfecho, as variáveis exploradas faltantes (SAPS 3 e SOFA) e variáveis auxiliares que poderiam ajudar a prever as variáveis exploradas faltantes (idade, GCS, ISS e NISS, Charlson). Foram realizadas dez imputações, utilizando o método de pareamento preditivo médio para o cálculo das imputações.

² Os resultados de calibração de SOFA, ISS e NISS são medidas de validação aparente, considerando que a ausência de equação previamente descrita impede a validação externa completa. A interpretação pode ser prejudicada e os gráficos de calibração (Figuras 7 a 12) podem auxiliar nessa avaliação.

³ Sem dados faltantes para ISS e NISS. Os resultados representam a coorte completa (n=1.053).

E/O: razão esperada por observado; CITL: Calibration in the large; AUROC: area under the receiving operator curve; Brier: escore de Brier; SAPS 3: Simplified Acute Physiology Score 3; SOFA: Sequential Organ Failure Assessment; ISS: Injury Severity Score; NISS: New Injury Severity Score; EG: Equação geral; ACS: América Central/Sul; RC: Equação recalibrada.

6 Discussão

6 DISCUSSÃO

Neste estudo de validação externa em uma população de pacientes traumatizados críticos, caracterizada sobretudo por pacientes jovens com grande prevalência de uso de suporte orgânico invasivo e TCE associado, os escores SAPS 3 e SOFA total máximo em até 24 horas da admissão na UTI apresentaram boa performance preditiva para mortalidade hospitalar, enquanto os escores anatômicos ISS e NISS tiveram performance ruim. Os escores SAPS 3 equação geral e equação América Central/Sul apresentaram descalibração nesta amostra, que foi corrigida após recalibração do intercepto e *slope*. Quando comparado ao SAPS 3, o SOFA apresentou discriminação semelhante com utilidade clínica discretamente superior ao longo de um amplo intervalo de probabilidade de desfecho.

O escore SAPS 3 apresentou acurácia inferior à descrita em seu modelo original, com AUROC de 0,848 na coorte completa e 0,854 apenas entre as vítimas de trauma ^(2, 3). A perda de calibração do SAPS 3 nas equações geral e América Latina/Sul também já tinha sido demonstrada no subgrupo com trauma na publicação original. Estes resultados não são inesperados, já que é comum que exista alguma deterioração em validações externas ao longo do tempo e quando testadas em subpopulações específicas, sobretudo nesta coorte composta de uma maioria de pacientes jovens e, exclusivamente, vítimas de trauma, penalizados no escore.

Estimativas de risco mal calibradas podem fazer com que este modelo prognóstico seja menos útil para avaliar resultados da UTI, embora

ainda úteis para comparações. Portanto, nós optamos pela recalibração do intercepto e do *slope* do modelo, o que levou a uma melhor performance.

Apesar de sua descrição original em pacientes com sepse ⁽⁵⁴⁾, o escore SOFA tornou-se o escore de disfunção orgânica mais amplamente utilizado em diferentes subpopulações de doentes críticos ^(30, 55), provavelmente facilitado pelo seu cálculo fácil à beira-leito. Originalmente, o escore SOFA não foi descrito para predição de mortalidade, mas avaliações posteriores em diferentes populações demonstraram seu uso para predição de desfechos em UTI ⁽⁵⁵⁾. A extensão e a gravidade da disfunção orgânica medida pelo SOFA também predizem morbimortalidade em pequenas coortes de pacientes com trauma críticos ⁽⁶⁻⁸⁾. A literatura apresenta a discriminação do escore, mas sua calibração não foi adequadamente avaliada e inexistiu equação de predição publicada.

Nesta coorte de pacientes vítimas de trauma grave, o SOFA total máximo em 24 horas de admissão na UTI teve a melhor combinação de discriminação, calibração e utilidade clínica para predição de mortalidade hospitalar, possivelmente, pela alta prevalência de pacientes jovens, com poucas comorbidades e alta gravidade clínica. Isso reforça a importância da carga de disfunção orgânica, como o motor do desfecho clínico de pacientes com trauma críticos, já demonstrada na população com trauma em geral e nos pacientes com TCE ^(56, 57). Além disso, de acordo com as melhores recomendações metodológicas em predição de risco ⁽³⁶⁾, nós apresentamos aqui uma equação de predição de mortalidade, que poderá ser usada para comparações em publicações futuras.

Os modelos anatômicos baseados no AIS foram desenvolvidos para descrever a extensão de lesões sofridas pelos pacientes há 50 anos, e a gravidade de cada lesão foi derivada de opiniões de especialistas ^(1, 9, 10), fato que pode explicar acurácia inferior do que a de modelos derivados de dados ⁽¹⁶⁾. O ISS e NISS diferem por permitirem incluir um grupo diferente de lesões. Enquanto podem ser semelhantes em traumas isolados em um único segmento corporal, o NISS apresenta potencial vantagem em pacientes com trauma contuso multissistêmico, como em nossa coorte ⁽¹⁴⁾. Entretanto, o alfa de Cronbach foi alto entre os escores, demonstrando alta consistência entre eles, provavelmente, porque a maioria das lesões incluídas tinham escore AIS elevado, independente se restritas a um ou mais segmentos corporais.

Em sua descrição inicial, Baker *et al* ⁽¹⁰⁾ demonstraram melhor predição de mortalidade com NISS do que com o ISS, o que foi confirmado em estudos subsequentes e revisões sistemáticas, em diferentes populações, sobretudo em pacientes com traumas contusos ^(13, 14, 58). No entanto, neste estudo ambos os escores apresentaram acurácia inferior às demonstradas em estudos prévios ^(14, 16, 24), provavelmente, por esta ser uma validação em um grupo específico de vítimas de trauma, com pacientes muito graves já admitidos em UTI, com mortalidade superior às prévias. Uma discriminação ruim associada a baixas calibrações e curvas de utilidade clínica sugerem que os escores anatômicos não são adequados para prognosticar pacientes com trauma críticos.

A aplicação dos escores em populações definidas de maneiras diferentes explica parte das diferenças em performances preditivas encontradas na literatura, o que é frequente em estudos em vítimas de trauma. Isso pode explicar os achados de superioridade do SAPS 3 e SOFA nesta validação de pacientes graves, com alto uso de suporte orgânico invasivo no momento da admissão e com altos ISS e NISS. A superioridade de modelos derivados do SAPS 3 e SOFA em relação a escores anatômicos ou combinados específicos para trauma já foi demonstrada em outros estudos ^(7, 8). Contudo, a comparação direta de acurácia é comprometida por critérios de elegibilidade diferentes, o que pode explicar os baixos valores de AUROC aos escores anatômicos demonstrados aqui quando comparados com outros estudos onde variaram entre 0,75 e 0,90 ⁽¹⁴⁾, além de a calibração e utilidade clínica serem raramente reportadas.

Pacientes com TCE grave representam um subgrupo de grande relevância clínica, nos quais os modelos preditivos poderiam ter uma performance diferente dos pacientes com trauma em geral. Nessa coorte, a discriminação piorou para SAPS 3 e SOFA nesta subpopulação, provavelmente, porque este seja um subgrupo heterogêneo de pacientes, com características prognósticas específicas além de GCS, como tamanho e reatividade pupilar, que não são capturadas por estes modelos ⁽⁵⁹⁾. Ainda assim, a disfunção orgânica não-neurológica é frequente e tem papel importante na predição de desfecho, tanto no TCE isolado como nos pacientes com trauma multissistêmico com TCE ^(56, 57, 60). O objetivo deste estudo não foi analisar especificamente pacientes com TCE, mas futuros

estudos prognósticos nessa população devem considerar incluir não apenas variáveis neurológicas, mas também incorporar o escore SOFA, como um marcador de disfunção orgânica nesses doentes.

6.1 Implicações para a prática clínica e avaliação de desfechos da UTI

A utilidade clínica de um modelo preditivo emergiu como uma ferramenta importante de avaliação da performance de um modelo ⁽⁴⁰⁾, mas ainda não tem sido amplamente adotada. Aqui, podemos observar que o escore SOFA pode adicionar maior utilidade ao processo de decisão clínica ao longo de um amplo intervalo de probabilidades do desfecho (em torno de 10% até 60%), diferente dos escores anatômicos. O conceito de curva de decisão clínica é recente e os exemplos mais comuns envolvem decisões de tratamento em oncologia e cardiologia ⁽⁴⁰⁾, sendo desafiadora a aplicação ao contexto de pacientes críticos. Este conceito é importante porque, apesar dessa característica, os resultados da curva de decisão clínica não são adequados para tomada de decisão individual de maneira precoce na trajetória da doença crítica (ex: retirada ou não instalação de suporte orgânico invasivo), já que eles não adicionam benefício no extremo das probabilidades de desfecho porque, individualmente, esses escores não foram desenvolvidos com esta finalidade. Um exemplo de aplicação em tomada de decisão clínica pode envolver, por exemplo, decisões sobre intervenções terapêuticas (cirúrgicas, minimamente invasivas, expectante): o

escore SOFA, marcando disfunção orgânica, pode adicionar mais ao processo de decisão do que escores anatômicos em si, especialmente, nos intervalos intermediários de desfechos clínicos, onde justamente persiste a dúvida clínica e pesam-se riscos e benefícios. Os achados também reforçam mais uma vez que o impacto das perturbações fisiológicas que levam ao desenvolvimento de disfunções orgânicas nas vítimas de trauma é muito maior do que a extensão anatômica das lesões sofridas.

Ainda que o foco da atual pesquisa tenha sido em escores relativos à admissão na UTI, novas pesquisas abordando aspectos dinâmicos do SOFA nesta população de doentes críticos podem adicionar novas perspectivas ao processo de decisão clínica.

Por fim, o SOFA apresenta uma vantagem em relação ao SAPS 3, ISS e NISS em razão da facilidade de seu cálculo para incorporação na tomada de decisão. Embora a fórmula matemática seja simples, a aplicação dos escores anatômicos dependentes de conhecimento da classificação das lesões pelo AIS, não é amplamente difundido entre médicos intensivistas. Apesar de seus cálculos serem realizados da mesma forma, apenas selecionando as lesões de maneira diferente, a modificação subsequente NISS não se popularizou da mesma forma que o antecessor ISS.

Quanto à avaliação de desfechos e *benchmarking*, dado que o SAPS 3 tem uma equação previamente publicada, ele permanece uma ferramenta melhor do que o SOFA. Ao calcular TMP de acordo com SAPS 3, UTIs especializadas em trauma devem reconhecer que existe alguma perda de calibração nessa amostra de pacientes e que objetivar $TMP < 1$ pode não

ser atingível com a equação original, conforme já foi descrito com pacientes com COVID-19 ⁽⁶¹⁾. No entanto, isso não invalida a avaliação da performance da UTI ao longo do tempo.

6.2 Implicações para pesquisa

Modelos prognósticos também têm várias utilidades no campo de pesquisa clínica. Os escores ajudam a definir, por exemplo, subgrupos mais homogêneos com trajetórias clínicas diferentes entre si. Em estudos randomizados, as características prognósticas podem ser usadas para determinar inclusão (enriquecimento prognóstico) e estratificação (heterogeneidade de efeito do tratamento); em estudos observacionais, podem ser usadas para ajuste de variáveis do *baseline*. No caso de pacientes vítimas de trauma, escores como SOFA, que refletem disfunção orgânica e com melhor performance preditiva nessa análise, poderiam ser mais interessantes nesses contextos por oferecer um melhor ajuste da gravidade clínica e consequente morbimortalidade, em detrimento dos ainda tradicionais escores puramente anatômicos ISS e NISS.

6.3 Pontos fortes e limitações

Este estudo de validação externa tem pontos fortes. Em primeiro lugar, a amostra é grande o bastante e com um número de eventos adequados para obter conclusões válidas ⁽⁶²⁾. Os resultados são

representativos de uma população de pacientes com trauma críticos de um país de média renda, onde não existem registros de trauma regionais ou nacionais com o mesmo alcance das iniciativas internacionais. Em segundo lugar, seguimos as mais atualizadas recomendações metodológicas para estudos de validação de modelos preditivos, incluindo medidas de discriminação, calibração e utilidade clínica. A maioria das publicações em modelos preditivos foca excessivamente em acurácia, uma característica de aplicabilidade limitada à beira-leito, e falha ao não apresentar medidas de calibração ⁽⁶³⁾. Em terceiro lugar, nós apresentamos as equações preditivas usadas para validação dos modelos, o que permite futuras validações externas, uma etapa necessária na reprodutibilidade dos resultados de pesquisas nesse campo. Finalmente, a proporção de dados faltantes, uma constante em análises retrospectivas com base em prontuário eletrônico, pode ser considerada pequena nessa amostra ⁽⁵²⁾, com resultados consistentes na análise de sensibilidade e sem dados faltantes em relação a desfechos.

Este estudo apresenta limitações. O desenho retrospectivo com base em banco de dados pode implicar viés de aferição, que poderia levar a conclusões incorretas na análise. No entanto, de modo a mitigar esse viés, a consistência dos dados usados na análise foi avaliada extensamente em um pré-processamento do banco de dados para assegurar a qualidade dos dados e a confiabilidade nos resultados. Para SAPS 3 e SOFA, embora os componentes individuais dos escores tenham sido coletados prospectivamente, eles posteriormente foram corrigidos se necessário. A

coleta dos escores anatômicos foi sim retrospectiva, baseada em resultados de TC e em campos estruturados e não-estruturados do prontuário eletrônico, de maneira semelhante como teriam sido feitos prospectivamente. Para lidar com essa limitação, os pesquisadores foram cegados em relação aos desfechos dos pacientes.

7 Conclusões

7 CONCLUSÕES

Neste estudo de validação externa em uma coorte de pacientes vítimas de trauma admitidos na UTI especializada, os escores SAPS 3 e SOFA total máximo de 24 horas de admissão tiveram melhor performance preditiva para mortalidade hospitalar do que os escores anatômicos ISS e NISS. O SOFA total máximo em 24 horas de admissão apresentou a melhor combinação de discriminação, calibração e utilidade clínica nesta coorte.

Na análise de subgrupo em pacientes com TCE grave, houve piora na discriminação e na calibração dos escores SAPS 3 e SOFA total máximo de 24 horas de admissão.

8 Anexos

8 ANEXOS

Anexo A - *Template* do banco de dados

Variáveis do banco de dados	Apresentação das variáveis
Variáveis de identificação	
Número de identificação	<i>Número de identificação do paciente no banco</i>
Nome completo	<i>Campo aberto</i>
Registro hospitalar	<i>Registro de prontuário hospitalar</i>
Data de nascimento	<i>Formato DD/MM/AAAA</i>
Variáveis demográficas	
Sexo	<i>Variável binária</i>
Comorbidades	<i>Variáveis individuais binárias</i>
Outras comorbidades	<i>Campo aberto</i>
Tabagismo	<i>Variável binária</i>
Etilismo	<i>Variável binária</i>
Variáveis relativas à admissão na UTI	
Data de admissão hospitalar	<i>Formato DD/MM/AAAA</i>
Data de admissão na UTI	<i>Formato DD/MM/AAAA</i>
Passagem prévia por outra UTI	<i>Variável binária</i>
Passagem prévia por outro serviço	<i>Variável binária</i>
Causa de admissão na UTI	<i>Variável categórica</i>
Mecanismo trauma principal e específico	<i>Variável categórica</i>
Descrição de mecanismo trauma	<i>Campo aberto</i>
Segmentos corporais acometidos	<i>Variáveis individuais binárias</i>
Parada cardiorrespiratória	<i>Variável binária</i>
GCS na cena	<i>Variável categórica</i>
PAS na cena	<i>Campo aberto</i>
GCS no pronto socorro	<i>Variável categórica</i>
PAS no pronto socorro	<i>Campo aberto</i>
Data da cirurgia	<i>Formato DD/MM/AAAA</i>
Descrição da intervenção cirúrgica	<i>Campo aberto</i>
Intubação na admissão	<i>Variável binária</i>
Sedativo	<i>Variável categórica</i>
Dose de sedativo	<i>Campo aberto</i>
RASS / GCS	<i>Variável categórica</i>

Continua

Anexo A - Template do banco de dados (conclusão)

Variáveis do banco de dados	Apresentação das variáveis
Droga vasoativa	<i>Variável categórica</i>
Dose de droga vasoativa	<i>Campo aberto</i>
PAS / PAD / PAM	<i>Campo aberto</i>
Suporte ventilatório	<i>Variável categórica</i>
Parâmetros de ventilação mecânica	<i>Campos abertos</i>
Saturação de oxigênio	<i>Campo aberto</i>
Diálise	<i>Variável categórica</i>
Exames laboratoriais	<i>Campos abertos</i>
Diagnóstico inicial	<i>Campo aberto</i>
Impressão	<i>Campo aberto</i>
Conduta inicial	<i>Campo aberto</i>
Variáveis individuais do SAPS 3	<i>Variáveis individuais categóricas</i>
Variáveis relativas à primeira evolução médica (primeiras 24h)	
PAM máxima e mínima nas 24 horas	<i>Campo aberto</i>
Diurese nas 24 horas	<i>Campo aberto</i>
Intubação orotraqueal	<i>Variável binária</i>
Sedativo	<i>Variável categórica</i>
Dose de sedativo	<i>Campo aberto</i>
RASS / GCS	<i>Variável categórica</i>
Droga vasoativa	<i>Variável categórica</i>
Dose de droga vasoativa	<i>Campo aberto</i>
PAS / PAD / PAM	<i>Campo aberto</i>
Suporte ventilatório	<i>Variável categórica</i>
Parâmetros de ventilação mecânica	<i>Campo aberto</i>
Saturação de oxigênio	<i>Campo aberto</i>
Diálise	<i>Variável categórica</i>
Exames laboratoriais	<i>Campos abertos</i>
Diagnóstico e impressão	<i>Campo aberto</i>
Conduta	<i>Campo aberto</i>
Variáveis relativas à alta da UTI	
Data de alta da UTI	<i>Formato DD/MM/AAAA</i>
Destino na alta da UTI	<i>Variável categórica</i>

UTI: unidade de terapia intensiva; GCS: Glasgow Coma Scale; PAS: pressão arterial sistólica; RASS: Richmond Agitation Sedation Scale; PAD: pressão arterial diastólica; PAM: pressão arterial média; SAPS 3: Simplified Acute Physiology Score 3.

Fonte: do autor.

Anexo B - Tabela AIS

Abbreviated Injury Scale – exemplos de classificação de lesões

	1 Leve	2 Moderada	3 Grave	4 Severa	5 Crítica
Cabeça e Pesçoço		Fratura de calota simples, contusão de laringe, fratura ou luxação de processo transversal vertebral, fratura com compressão leve de vértebra	Fratura de base de crânio, fratura de calota com afundamento/cominutiva SEM pneumo/fístula/perda de massa, dissecação/trombose artéria carótica SEM déficit, contusão, HSA traumática, laceração de tireóide, luxação/fratura de lâmina/pedículo/faceta vertebral, fratura com compressão de vértebra	Fratura de base de crânio/calota COM pneumo/fístula/perda de massa, dissecação/trombose artéria carótica COM déficit, hematoma SEM efeito de massa, fratura/laceração/esmagamento traquéia, lesão incompleta de medula	Lesão de tronco cerebral, hematoma COM efeito de massa, LAD, edema cerebral difuso, lesão completa de medula, avulsão/ruptura de traquéia/laringe/esôfago
Face	Fratura nasal/dente/ramo de mandíbula, lesão de conduto auditivo externo	Le Fort I, fratura de malar/órbita/zigomático/corpo ou subcondilar de mandíbula, fratura nasal*/ramo de mandíbula*	Le Fort II, fratura de malar*/órbita*/zigomático*/corpo ou subcondilar de mandíbula*, avulsão ou laceração de olho	Le Fort III	

continua

Anexo B - Tabela AIS (continuação)

	1 Leve	2 Moderada	3 Grave	4 Severa	5 Crítica
Tórax	Fratura de costela	Fratura de costela*/esterno, fratura de >1 costela contínua, fratura ou luxação de processo transversal vertebral, fratura com compressão leve de vértebra	Contusão/laceração pulmonar ≤1 lobo, hemopneumotórax unilateral, fratura de esterno*, fratura de costela >3 associado a hemotórax/pneumotórax/pneumomediastino, luxação/fratura de lâmina/pedículo/faceta vertebral, fratura com compressão de vértebra	Hemopneumotórax bilateral, hemopneumomediastino, tórax instável, lesão inalatória, contusão miocárdica, perfuração parede torácica, contusão/laceração pulmonar multilobar ou bilateral, fratura/laceração/esmagamento traquéia, lesão incompleta de medula, laceração/trombose de artéria subclávia	Laceração/ruptura aorta/brônquio/miocárdio/pericárdio/vasos da base, lesão inalatória em VMI, laceração pulmonar multilobar com hemopneumotórax hipertensivo, hemotórax >1000mL, tamponamento cardíaco, avulsão/ruptura de traquéia/ esôfago, lesão completa de medula
Abdome	Contusão de parede abdominal/períneo	Avulsão/perfuração de parede abdominal, laceração de ureter, fratura ou luxação de processo transversal vertebral, fratura com compressão leve de vértebra	Contusão de órgão, hematoma de retroperitônio, luxação/fratura de lâmina/pedículo/faceta vertebral, fratura com compressão de vértebra	Laceração/perfuração superficial de órgão, ruptura de bexiga/intraperitoneal	Ruptura/avulsão/laceração profunda de órgão

continua

Anexo B - Tabela AIS (conclusão)

	1 Leve	2 Moderada	3 Grave	4 Severa	5 Crítica
Pelve e Extremidades		Luxação acromioclavicular/ esternoclavicular/pat elar, fratura simples exceto fêmur	Fratura em livro aberto da pelve, fratura de fêmur simples, esmagamento/amputaçã o parcial ou completa abaixo do joelho, esmagamento/amputaçã o parcial ou completa de membro superior, outras fraturas*, laceração/trombose de artéria axilar/braquial/femoral/po plíteia	Esmagamento de pelve, esmagamento/amputaçã o parcial ou completa acima do joelho, fratura cominutiva de pelve	
Externo	Abrasão/contusão pequena (<25cm ² face/ <50cm ² corpo), laceração superficial/subcutâ nea (<5 cm face/ <10cm corpo), Queimadura 2º grau profundo/3º grau <6% SCQ	Abrasão/contusão grande (≥25cm ² face/ ≥50cm ² corpo), laceração profunda (≥5 cm face/ ≥10cm corpo), Queimadura 2º grau profundo/3º grau 6-15% SCQ	Queimadura 2º grau profundo/3º grau 16-25% SCQ	Queimadura 2º grau profundo/3º grau 26-35% SCQ	Queimadura 2º grau profundo/3º grau 36-90% SCQ

* exposta, cominutiva ou com desvio, SCQ: superfície corporal queimada.

Fonte: adaptado de *Association for the Advancement of Automotive Medicine*, 2008.

Anexo C - Tabela SOFA

Sequential Organ Failure Assessment					
	0	1	2	3	4
Respiratório PaO ₂ /FiO ₂	≥ 400	< 400	<300	<200	<100
Cardiovascular PAM (mmHg) ou droga em mcg/kg/min	PAM ≥ 70	PAM < 70	Dopamina ≤5 ou dobutamina em qualquer dose	Dopamina >5 ou adrenalina ≤0,1 ou noradrenalina ≤0,1	Dopamina >15 ou adrenalina >0,1 ou noradrenalina >0,1
Renal Creatinina (mg/dL) ou diurese (ml/24h)	< 1,2	1,2-1,9	2,0-3,4	3,5-4,9 ou <500	≥5,0 ou <200
Hepático Bilirrubina (mg/dL)	< 1,2	1,2-1,9	2,0-5,9	6,0-11,9	≥ 12,0
Neurológico GCS	15	13-14	10-12	6-9	3-5
Hematológico Plaquetas (/mm ³)	≥ 150000	< 150000	< 100000	<50000	<20000

PAM: pressão arterial média; GCS: Glasgow Coma Scale.

Fonte: Vincent *et al*, 1996 ⁽⁴⁾.

Anexo D – Charlson Comorbidity Index

Charlson Comorbidity Index	
Comorbidade	Valor
Infarto do miocárdio	1
Insuficiência cardíaca congestiva	1
Doença vascular periférica	1
Demência	1
Doença cerebrovascular (exceto hemiplegia)	1
Doença pulmonar crônica	1
Doença do tecido conjuntivo	1
Diabetes mellitus sem complicação	1
Úlcera péptica	1
Hemiplegia	2
Doença renal moderada ou severa	2
Diabetes mellitus com complicação	2
Tumor maligno	2
Leucemia	2
Linfoma	2
Doença hepática moderada ou severa	3
Tumor maligno, metástase	6
Síndrome de imunodeficiência adquirida humana	6

Fonte: Charlson *et al* ⁽⁴⁶⁾.

Anexo E – Prenchimento do Checklist *TRIPOD*

TRIPOD Checklist: Prediction Model Validation

Title and Abstract			Pages
Title	1	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted	xiii
Abstract	2	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	xiii-xiv
Introduction			
Background and objectives	3a	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	2-4
	3b	Specify the objectives, including whether the study describes the development or validation of the model or both	4, 6
Methods			
Source of data	4a	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	20-22
	4b	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	20
Participants	5a	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	20-21
	5b	Describe eligibility criteria for participants.	21-22
	5c	Give details of treatments received, if relevant.	NA
Outcome	6a	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	22
	6b	Report any actions to blind assessment of the outcome to be predicted.	NA
Predictors	7a	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	22-28
	7b	Report any actions to blind assessment of predictors for the outcome and other predictors.	26
Sample size	8	Explain how the study size was arrived at.	28-29

Continua

Anexo E – Preenchimento do Checklist TRIPOD (continuação)**TRIPOD Checklist: Prediction Model Validation**

Methods		Pages
Missing data	9	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method. 32
Statistical analysis methods	10a	For validation, describe how the predictions were calculated. 26-28
	10b	Specify all measures used to assess model performance and, if relevant, to compare multiple models. 30-32
	10c	Describe any model updating (e.g., recalibration) arising from the validation, if done. 31
Risk groups	11	Provide details on how risk groups were created, if done. NA
Development vs. validation	12	For validation, identify any differences from the development data in setting, eligibility criteria, outcome, and predictors. 20-22, 25-27
Results		
Participants	13a	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful. 34, Figura 1
	13b	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome. 34-35, Figura 1, Tabela 1
	13c	For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors and outcome). 34-35, Tabela 1, Figura 2-5
Model performance	14	Report performance measures (with CIs) for the prediction model. Tabelas 4 e 5
Model-updating	15	If done, report the results from any model updating (i.e., model specification, model performance). 45, Tabela 5, Figura 9

Continua

Anexo E – Preenchimento do Checklist *TRIPOD* (conclusão)

TRIPOD Checklist: Prediction Model Validation

Discussion			Pages
Limitations	16	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	60-62
Interpretation	17a	For validation, discuss the results with reference to performance in the development data, and any other validation data.	54-58
	17b	Give an overall interpretation of the results, considering objectives, limitations, results from similar studies, and other relevant evidence.	54-58
Implications	18	Discuss the potential clinical use of the model and implications for future research.	58-60
Other information			
Supplementary information	19	Provide information about the availability of supplementary resources, such as study protocol, web calculator, and data sets.	NA
Funding	20	Give the source of funding and the role of the funders for the present study.	NA

NA: não aplicável

Fonte: Moons *et al* ⁽³⁶⁾.

9 Referências

9 REFERÊNCIAS

1. Baker SP, O'Neill B, Haddon W, Long WB. The injury severity score: a method for describing patients with multiple injuries and evaluating emergency care. *J Trauma*. 1974;14(3):187-96.
2. Moreno RP, Metnitz PGH, Almeida E, Jordan B, Bauer P, Campos RA, Iapichino G, Edbrooke D, Capuzzo M, Le Gall JR; SAPS 3 Investigators. SAPS 3--From evaluation of the patient to evaluation of the intensive care unit. Part 2: Development of a prognostic model for hospital mortality at ICU admission. *Intensive Care Med*. 2005;31(10):1345-55.
3. Metnitz PGH, Moreno RP, Almeida E, Jordan B, Bauer P, Campos RA, Iapichino G, Edbrooke D, Capuzzo M, Le Gall JR; SAPS 3 Investigators. SAPS 3--From evaluation of the patient to evaluation of the intensive care unit. Part 1: Objectives, methods and cohort description. *Intensive Care Med*. 2005;31(10):1336-44.
4. Vincent JL, Moreno R, Takala J, Willatts S, de Mendonça A, Bruining H, Reinhart CK, Suter PM, Thijs LG. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. *Intensive Care Med*. 1996;22(7):707-10.
5. GBD 2019 Diseases and Injuries Collaborators. Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet*. 2020;396(10258):1204-22.
6. Antonelli M, Moreno R, Vincent JL, Sprung CL, Mendonça A, Passariello M, Riccioni L, Osborn J. Application of SOFA score to trauma

patients. Sequential Organ Failure Assessment. *Intensive Care Med.* 1999;25(4):389-94.

7. Fueglistaler P, Amsler F, Schepp M, Fueglistaler-Montali I, Attenberger C, Pargger H, Jacob AL, Gross T. Prognostic value of Sequential Organ Failure Assessment and Simplified Acute Physiology II Score compared with trauma scores in the outcome of multiple-trauma patients. *Am J Surg.* 2010;200(2):204-14.

8. Hwang SY, Lee JH, Lee YH, Hong CK, Sung AJ, Choi YC. Comparison of the Sequential Organ Failure Assessment, Acute Physiology and Chronic Health Evaluation II scoring system, and Trauma and Injury Severity Score method for predicting the outcomes of intensive care unit trauma patients. *Am J Emerg Med.* 2012;30(5):749-53.

9. Keller WK, Dillihunt RC, Fenner Jr HA, Hobbs NM, Jolley FL, Keeney A, et al. Rating the severity of tissue damage. I. The abbreviated scale. *JAMA.* 1971;215(2):277-80.

10. Osler T, Baker SP, Long W. A modification of the injury severity score that both improves accuracy and simplifies scoring. *J Trauma.* 1997;43(6):922-5.

11. Balogh Z, Offner PJ, Moore EE, Biffl WL. NISS predicts postinjury multiple organ failure better than the ISS. *J Trauma.* 2000;48(4):624-8.

12. Balogh ZJ, Varga E, Tomka J, Süveges G, Tóth L, Simonka JA. The new injury severity score is a better predictor of extended hospitalization and intensive care unit admission than the injury severity score in patients with multiple orthopaedic injuries. *J Orthop Trauma.* 2003;17(7):508-12.

13. Lavoie A, Moore L, LeSage N, Liberman M, Sampalis JS. The Injury Severity Score or the New Injury Severity Score for predicting intensive care unit admission and hospital length of stay? *Injury*. 2005;36(4):477-83.
14. Tohira H, Jacobs I, Mountain D, Gibson N, Yeo A. Systematic review of predictive performance of injury severity scoring tools. *Scand J Trauma Resusc Emerg Med*. 2012;20(1):63.
15. Smith BP, Goldberg AJ, Gaughan JP, Seamon MJ. A comparison of Injury Severity Score and New Injury Severity Score after penetrating trauma: A prospective analysis. *J Trauma Acute Care Surg*. 2015;79(2):269-74.
16. Moore L, Lavoie A, le Sage N, Bergeron E, Emond M, Abdous B. Consensus or data-derived anatomic injury severity scoring? *J Trauma*. 2008;64(2):420-6
17. Dewar D, Moore FA, Moore EE, Balogh Z. Postinjury multiple organ failure. *Injury*. 2009;40(9):912-8.
18. Champion HR, Sacco WJ, Carnazzo AJ, Copes W, Fouty WJ. Trauma score. *Crit Care Med*. 1981;9(9):672-6.
19. Champion HR, Sacco WJ, Copes WS, Gann DS, Gennarelli TA, Flanagan ME. A revision of the Trauma Score. *J Trauma*. 1989;29(5):623-9.
20. Kuhls DA, Malone DL, McCarter RJ, Napolitano LM. Predictors of mortality in adult trauma patients: the Physiologic Trauma Score is equivalent to the Trauma and Injury Severity Score. *J Am Coll Surg*. 2002;194(6):695-704.

21. Boyd CR, Tolson MA, Copes WS. Evaluating trauma care: the TRISS method. Trauma Score and the Injury Severity Score. *J Trauma*. 1987;27(4):370-8.
22. Lefering R. Development and validation of the revised injury severity classification score for severely injured patients. *Eur J Trauma Emerg Surg*. 2009;35(5):437-47.
23. Lefering R, Huber-Wagner S, Nienaber U, Maegele M, Bouillon B. Update of the trauma risk adjustment model of the TraumaRegister DGU™: The Revised Injury Severity Classification, version II. *Crit Care*. 2014;18(5):476.
24. Guzzo JL, Bochicchio G V., Napolitano LM, Malone DL, Meyer W, Scalea TM. Prediction of outcomes in trauma: Anatomic or physiologic parameters? *J Am Coll Surg*. 2005;201(6):891-7.
25. de Munter L, Polinder S, Lansink KWW, Cnossen MC, Steyerberg EW, de Jongh MAC. Mortality prediction models in the general trauma population: A systematic review. *Injury*. 2017;48(2):221-9.
26. Osler T, Nelson LS, Bedrick EJ. Injury severity scoring. *J Intensive Care Med*. 1999;14(1):9-19.
27. Keegan MT, Soares M. What every intensivist should know about prognostic scoring systems and risk-adjusted mortality. *Rev Bras Ter Intensiva*. 2016;28(3):264-9.
28. Zimmerman JE, Kramer AA, McNair DS, Malila FM. Acute Physiology and Chronic Health Evaluation (APACHE) IV: hospital mortality assessment for today's critically ill patients. *Crit Care Med*. 2006;34(5):1297-310.

29. Lemeshow S, Teres D, Klar J, Avrunin JS, Gehlbach SH, Rapoport J. Mortality Probability Models (MPM II) based on an international cohort of intensive care unit patients. *JAMA*. 1993;270(20):2478-86.
30. Minne L, Abu-Hanna A, de Jonge E. Evaluation of SOFA-based models for predicting mortality in the ICU: A systematic review. *Crit Care*. 2008;12(6):R161.
31. Lambden S, Laterre PF, Levy MM, Francois B. The SOFA score—development, utility and challenges of accurate assessment in clinical trials. *Crit Care*. 2019;23(1):374.
32. Vincent JL, de Mendonça A, Cantraine F, Moreno R, Takala J, Suter PM, Sprung CL, Colardyn F, Blecher S. Use of the SOFA score to assess the incidence of organ dysfunction/failure in intensive care units: Results of a multicenter, prospective study. Working group on "sepsis-related problems" of the European Society of Intensive Care Medicine. *Crit Care Med*. 1998;26(11):1793-800.
33. Falcão ALE, Barros AG de A, Bezerra AAM, Ferreira NL, Logato CM, Silva FP, do Monte ABFO, Tonella RM, de Figueiredo LC, Moreno R, Dragosavac D, Andreollo NA. The prognostic accuracy evaluation of SAPS 3, SOFA and APACHE II scores for mortality prediction in the surgical ICU: an external validation study and decision-making analysis. *Ann Intensive Care*. 2019;9(1):18.
34. Moons KGM, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? *BMJ*. 2009;338:b375.
35. Steyerberg EW, Moons KGM, van der Windt DA, Hayden JA, Perel P, Schroter S, Riley RD, Hemingway H, Altman DG, PROGRESS Group.

Guidelines and Guidance Prognosis Research Strategy (PROGRESS) 3: Prognostic Model Research. *PLoS Med* 10(2): e1001381

36. Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW, Vickers AJ, Ransohoff DF, Collins GS. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med*. 2015;162(1):W1-73.

37. Bouwmeester W, A Zuithoff NP, Mallett S, Geerlings MI, Vergouwe Y, Steyerberg EW, Altman DG, Moons KG. Reporting and methods in clinical prediction research: a systematic review. *PLoS Med*. 2012;9(5):1-12.

38. Collins GS, de Groot JA, Dutton S, Omar O, Shanyinde M, Tajar A, Voysey M, Wharton R, Yu LM, Moons KG, Altman DG. External validation of multivariable prediction models: a systematic review of methodological conduct and reporting. *BMC Med Res Methodol*. 2014;14(1):40.

39. Mushkudiani NA, Hukkelhoven CWPM, Hernández A v., Murray GD, Choi SC, Maas AIR, Steyerberg EW. A systematic review finds methodological improvements necessary for prognostic models in determining traumatic brain injury outcomes. *J Clin Epidemiol*. 2008;61(4):331-43.

40. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making*. 2006;26(6):565-74.

41. Riley RD, A Debray TP, Collins GS, Archer L, Ensor J, van Smeden M, Snell KIE. Minimum sample size for external validation of a clinical prediction model with a binary outcome. *Stat Med*. 2021;40(19):4230-51.

42. 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.
43. Vergouwe Y, Steyerberg EW, Eijkemans MJC, Habbema JDF. Substantial effective sample sizes were required for external validation studies of predictive logistic regression models. *J Clin Epidemiol.* 2005;58(5):475-83.
44. Peek N, Arts DGT, Bosman RJ, van der Voort PHJ, de Keizer NF. External validation of prognostic models for critically ill patients required substantial sample sizes. *J Clin Epidemiol.* 2007;60(5):491-501.
45. Steyerberg EW. *Clinical Prediction Models: A Practical Approach to development, Validation, and Updating.* Cham: Springer International Publishing; 2019. (Statistics for Biology and Health).
46. Palazón-Bru A, Folgado-de la Rosa DM, Cortés-Castell E, López-Cascales MT, Gil-Guillén VF. Sample size calculation to externally validate scoring systems based on logistic regression models. Montazeri A, editor. *PLoS One.* 2017;12(5):e0176726.
47. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *Eur Heart J.* 2014;35(29):1925-31.
48. Leisman DE, Harhay MO, Lederer DJ, Abramson M, Adjei AA, Bakker J, Ballas ZK, Barreiro E, Bell SC, Bellomo R, Bernstein JA, Branson RD, Brusasco V, Chalmers JD, Chokroverty S, Citerio G, Collop NA, Cooke CR, Crapo JD, Donaldson G, Fitzgerald DA, Grainger E, Hale L, Herth FJ, Kochanek PM, Marks G, Moorman JR, Ost DE, Schatz M, Sheikh A, Smyth AR, Stewart I, Stewart PW, Swenson ER, Szymusiak R, Teboul JL, Vincent JL, Wedzicha JA, Maslove DM. Development and Reporting of Prediction

Models: Guidance for Authors from Editors of Respiratory, Sleep, and Critical Care Journals. *Crit Care Med.* 2020;48(5):623-33.

49. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, Pencina MJ, Kattan MW. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology.* 2010;21(1):128-38.

50. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics.* 1988;44(3):837-45.

51. StataCorp. *Stata Statistical Software: Release 16.* College Station, Texas, USA: StataCorp LLC; 2019.

52. Vesin A, Azoulay E, Ruckly S, Vignoud L, Rusinová K, Benoit D, Soares M, Azevedo-Maia P, Abroug F, Benbenishty J, Timsit JF. Reporting and handling missing values in clinical studies in intensive care units. *Intensive Care Med.* 2013;39(8):1396-404.

53. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. *Stat Med.* 2011;30(4):377-99.

54. Vincent JL, Moreno R, Takala J, Willatts S, de Mendonça A, Bruining H, Reinhart CK, Suter PM, Thijs LG. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. *Intensive Care Med.* 1996;22(7):707-10.

55. Soo A, Zuege DJ, Fick GH, Niven DJ, Berthiaume LR, Stelfox HT, Doig CJ. Describing organ dysfunction in the intensive care unit: a cohort study of 20,000 patients. *Crit Care.* 2019;23(1):186.

56. Zygun D, Berthiaume L, Laupland K, Kortbeek J, Doig C. SOFA is superior to MOD score for the determination of non-neurologic organ dysfunction in patients with severe traumatic brain injury: A cohort study. *Crit Care*. 2006;10(4):R115.
57. Astarabadi M, Khurram M, Asmar S, Bible L, Chehab M, Castanon L, Ditillo M, Douglas M, Joseph B. The impact of non-neurological organ dysfunction on outcomes in severe isolated traumatic brain injury. *J Trauma Acute Care Surg*. 2020;89(2):405-10.
58. Tay SY, Sloan EP, Zun L, Zaret P. Comparison of the New Injury Severity Score and the Injury Severity Score. *J Trauma*. 2004;56(1):162-4.
59. Steyerberg EW, Mushkudiani N, Perel P, Butcher I, Lu J, Mchugh GS, Murray GD, Marmarou A, Roberts I, Habbema JD, Maas AI. Predicting outcome after traumatic brain injury: development and international validation of prognostic scores based on admission characteristics. *PLoS Med*. 2008;5(8):e165.
60. Martins ET, Linhares MN, Sousa DS, Schroeder HK, Meinerz J, Rigo LA, Bertotti MM, Gullo J, Hohl A, Dal-Pizzol F, Walz R. Mortality in severe traumatic brain injury: A multivariate analysis of 748 Brazilian patients from Florianópolis City. *J Trauma*. 2009;67(1):85-90.
61. Kurtz P, Bastos LSL, Salluh JIF, Bozza FA, Soares M. SAPS-3 performance for hospital mortality prediction in 30,571 patients with COVID-19 admitted to ICUs in Brazil. *Intensive Care Med*. 2021;47(9):1047-9.
62. Collins GS, Ogundimu EO, Altman DG. Sample size considerations for the external validation of a multivariable prognostic model: a resampling study. *Stat Med*. 2016;35(2):214-26.

63. van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW; Topic Group 'Evaluating diagnostic tests and prediction models' of the STRATOS initiative. Calibration: the Achilles heel of predictive analytics. *BMC Med.* 2019;17(1):230.