



UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
PROGRAMA DE PÓS-GRADUAÇÃO EM EPIDEMIOLOGIA

METODOLOGIA PARA AJUSTE DE INDICADORES
DE DESFECHOS HOSPITALARES POR RISCO
PRÉVIO DO PACIENTE

Aluna: **Mariza Machado Klück**

Orientadora: **Prof^a Jandyra M. Guimarães Fachel**

Porto Alegre, dezembro de 2004



William Farr
(1807-1883)

*What are figures worth if they do
no good to men's bodies or souls?*

William Farr, 1864

Agradecimentos

Nenhuma realização é individual. Sempre é fruto da cooperação direta ou indireta de um grande número de pessoas. A responsabilidade de nomear a todos que me apoiaram, incentivaram e ajudaram nesta etapa, tornando possível sua conclusão, é muito grande. Portanto, quero deixar meu agradecimento a todos aqueles que de uma forma ou de outra estiveram a meu lado e que são, em diferentes graus, co-autores deste trabalho. No entanto, alguns destaques são necessários, mas, nomeando a estes, não estou esquecendo todos os demais. Portanto, meu agradecimento especial para:

Minha amiga e colega Ilda Pereira Silveira, Chefe do SAMIS/HCPA, pelo apoio e incentivo permanentes;

Meu querido colega e amigo Dr. José Ricardo Guimarães, parceiro neste e em todos os demais projetos profissionais nestes últimos 10 anos,

Minha orientadora, Jandyra Fachel, pela orientação segura, pelo incentivo e apoio e, mais que tudo, pela amizade;

À coordenação do PPGEpi, pela oportunidade de realizar este Curso e por acreditar em minha capacidade de fazê-lo;

Minhas colegas do PPGA Alice Zelmanovicz, Ruth Liane Henn e Maria Lúcia Oppermann, companheiras fiéis de toda esta trajetória;

Os demais colegas, professores e colaboradores do PPGEpi, parceiros desta jornada;

O HCPA, especialmente na figura de seu Vice-Presidente Administrativo, Prof. Carlos Alberto Prompt, pela oportunidade de desenvolver este e outros tantos projetos interessantes e desafiadores;

Os acadêmicos de Medicina Lucas Hemb e Maurício Drehmer, pela ajuda inestimável na coleta dos dados;

Meu colega do Departamento de Medicina Social, Paul Fisher, pela ajuda nas traduções e ilustrações, sempre urgentes;

Todos meus amigos do SAMIS/HCPA pela ajuda, apoio e amizade;

O Grupo de Sistemas do HCPA, pela cooperação na obtenção dos dados necessários para esta pesquisa;

Minhas filhas Mariana e Lúcia Klück Stumpf, minhas melhores amigas e apoiadoras, razão maior de todas minhas realizações.

Aos meus pais, **Helmuth e Naura Klück**,
com gratidão e reconhecimento.

SUMÁRIO

Siglas e abreviaturas.....	06
Resumo.....	07
Abstract.....	08
Introdução.....	09
I- Revisão da literatura.....	12
1- Qualidade assistencial e indicadores.....	12
2- Ajuste de risco.....	14
2.1- História.....	14
2.2- Por que ajustar por risco?.....	17
2.3- Como ajustar pelo risco.....	18
3- Escores de risco.....	20
3.1- APACHE.....	21
3.2- POSSUM e P-Possum.....	22
3.3- SAPS.....	24
4- Modelos preditivos de mortalidade cirúrgica.....	26
4.1- Projeto Veteranos.....	27
4.2- Outros modelos.....	29
5- Fatores de risco para mortalidade cirúrgica.....	31
5.1- Classificação ASA.....	31
5.2- Albumina sérica.....	37
5.3- Idade.....	40
5.4- Caráter da cirurgia.....	42
5.5- Outros.....	42
6- Questões metodológicas.....	44
6.1- Fontes de dados para modelos de risco.....	45
6.2- Imputação de dados faltantes.....	49
6.3- Regressão Logística.....	52
Referências bibliográficas.....	55
II- Artigo 1 – <i>Estratégias de modelagem por Regressão Logística para o ajuste de risco</i>	63
III- Artigo 2 - <i>Desenvolvimento e validação de um escore de risco de mortalidade cirúrgica pós laparotomia exploradora</i>	96
IV- Considerações finais.....	128

LISTA DE SIGLAS E ABREVIATURAS

APACHE	Acute Physiology and Chronic Health Evaluation
AR	Ajuste de Risco
ASA	American Society of Anesthesiologists
BUPA	British United Provident Association
DVA	Department of Veterans Affairs
HCPA	Hospital de Clínicas de Porto Alegre
NCEPOD	National Confidential Enquiry into Perioperative Deaths
NSQIP	National Surgical Quality Improvement Program
POSSUM	Physiological and Operative Severity Score for enUmeration of Mortality and morbidity
QI Project	Quality Indicator Project
RIPSA	Rede Interagencial de Informações para a Saúde
RL	Regressão Logística
ROC	Receiver Operation Characteristics
SAPS	Simplified Acute Physiology Score
SIPAGEH	Sistema de Indicadores Padronizados para Gestão Hospitalar
SRS	Surgical Risk Score
SUS	Sistema Único de Saúde

RESUMO

A prática assistencial deve ser apoiada por sistemas de informação clínica adequados, completos e disponíveis, que garantam a segurança do paciente através do suporte à tomada de decisão diagnóstica e terapêutica, com o objetivo de prevenir eventos e desfechos adversos, corrigir rumos e obter dados e indicadores que permitam medir, avaliar e comparar o grau da qualidade assistencial oferecida à população-alvo.

Vários estudos demonstram que as medidas de desfecho da prática assistencial devem ser ajustadas pela gravidade do caso ou por condições próprias do paciente, em um processo conhecido como “ajuste por risco”. O ajuste por risco (ou padronização por severidade) visa a minimizar a influência confundidora de potenciais fatores de risco inerentes ao paciente e/ou a sua patologia, permitindo assim a comparação da qualidade da assistência, ajustando os desfechos encontrados para o risco prévio do paciente.

Inúmeros estudos têm sido publicados, propondo modelos de ajuste de risco para diferentes desfechos de pacientes, mas esta metodologia ainda não está difundida no Brasil. As razões para isso podem ser a falta de conhecimento da metodologia adequada ou a reduzida disponibilidade de bases de dados clínicas adequadas a este tipo de análise.

O objetivo desta tese é o de propor, testar e validar uma metodologia para ajuste de risco de desfechos hospitalares a partir de dados de prontuário eletrônico de pacientes do HCPA. A metodologia sugerida foi testada em uma amostra de dados de pacientes do HCPA visando definir um modelo de risco para óbito após realização de laparotomia exploradora e validada em outra amostra de pacientes do mesmo hospital.

A tese é composta pela revisão da literatura realizada e por dois artigos que descrevem a metodologia proposta e os resultados de sua aplicação na definição de um índice de risco para mortalidade cirúrgica.

A metodologia mostrou-se adequada para seu propósito e permitiu o desenvolvimento de um índice de risco para mortalidade cirúrgica que foi validado na amostra de teste com excelentes resultados. A discriminação do modelo, medida pela área sob a curva ROC, foi de 0,89 na coorte de derivação e de 0,90 na coorte de validação. O projeto permitiu também demonstrar a adequação do prontuário eletrônico de pacientes do HCPA a este propósito, em termos de completude, acurácia e fidedignidade das informações, além do fácil acesso aos dados.

ABSTRACT

Health information systems should provide the data and information required for effective health care delivery. On an individual patient level, the data and information should support diagnostic and therapeutic decision-making that assures patient safety by allowing the decision maker to predict adverse outcomes and correct a course of treatment accordingly. On an population level, the information system should produce process indicators that allow institutions to monitor and compare the quality of their health care processes.

Various studies have shown that the values of health care outcome indicators must be adjusted according to case severity and patient condition, a process known as risk adjustment. The process is an attempt to minimize inter-patient outcome variation by making adjustments for differences in the patients' clinical states that are known to influence process outcomes. Such adjustments allow more meaningful interpretation of indicator data across time, for the purpose of process quality monitoring within a population, and allow valid comparison between populations.

Much work has been published in which risk adjustment models for different patient outcomes have been proposed; however, such models are not widely used in Brazil. This may be due to a lack of expertise in the area of risk modelling or simply the absence of the clinical data bases required to do the modelling. Whatever the reason, the work described herein is an attempt to correct this situation.

The objective of this research was to propose, test and validate a risk assessment model that would allow outcomes based process indicators to be adjusted according to the individual patient risk based on data readily available in the patient record. The model was developed using data on 450 adult exploratory laparotomy patients admitted to the Hospital de Clínicas de Porto Alegre between February of 2000 and January of 2003 using surgical process related mortality as the outcome indicator. The model was validated using a sample of 251 patients at the same hospital.

The thesis is composed of a literature review and two articles: one describing the model development process and the other reporting the results of testing and validating the model on a hospital population using surgical mortality as the process outcome indicator.

The study shows that it is possible to develop a risk assessment model, based on data readily available in the patient record, appropriate for making risk adjustments to process outcome indicators. Further, the study demonstrated the quality of the content, in terms of accessibility, completeness and data fidelity, of the hospital's electronic patient record.

Introdução

A busca da qualidade da atenção dos serviços de saúde deixou de ser uma atitude isolada e tornou-se hoje um imperativo técnico e social. A Sociedade está exigindo cada vez mais a qualidade dos serviços a ela prestados, principalmente por órgãos públicos. Esta exigência torna fundamental a criação de normas e mecanismos de avaliação e controle da qualidade assistencial.

Qualidade assistencial é definida pelo Instituto de Medicina americano (*Institute of Medicine*) como “o grau pelo qual os serviços assistenciais para indivíduos e populações aumentam a probabilidade de desfechos desejados e que sejam consistentes com o conhecimento profissional corrente” (1). Uma vez que a garantia da qualidade exige um maior nível de profissionalismo e, conseqüentemente, um melhor desenvolvimento técnico, a meta final de um programa de busca da qualidade deve ser a busca da “*satisfação do paciente, por intermédio de uma atenção competente, apropriada e oportuna, sem duplicação ou super-utilização dos serviços, com um mínimo de complicações ou seqüelas*” (2).

A prática assistencial deve ser apoiada por sistemas de informação clínica adequados, completos e disponíveis, que garantam a segurança do paciente, através do suporte à tomada de decisão diagnóstica e terapêutica durante o tratamento de pacientes individuais, com o objetivo de prevenir eventos e desfechos adversos, corrigir rumos e melhorar continuamente a qualidade da assistência e a obtenção de dados e indicadores que permitam medir, avaliar e comparar o grau da qualidade assistencial oferecida à população-alvo.

A cultura da qualidade significa planejar sistemas direcionados a prevenir, detectar e minimizar danos e a probabilidade de erros e requer dados e informações que permeiam as várias atividades e procedimentos nas quais o paciente é envolvido (3). De acordo com relatório do Comitê Consultor de Tecnologia da Informação da Presidência dos EUA (PITAC), as ferramentas oferecidas pela tecnologia da informação podem permitir ao setor de assistência à saúde uma produtividade e qualidade sem precedentes se houver uma visão estratégica e pesquisa adequada para garantir seu sucesso (4).

Uma das abordagens para detectar, caracterizar e reportar riscos potencialmente evitáveis é o desenvolvimento de medidas de controle baseadas em dados rotineiramente coletados, que podem ser usados na construção de alertas e de indicadores de problemas potenciais. O consenso geral é o de que os hospitais podem reduzir os danos causados aos pacientes melhorando o meio ambiente para a segurança, implementando

mudanças técnicas tais como a adoção de prontuário eletrônico, que ajudam a melhorar a percepção da equipe assistencial sobre os riscos do paciente (5).

A avaliação e a comparação da qualidade assistencial pode ser feita a partir das medidas de desfecho (indicadores) e ser baseada em padrões definidos em ensaios clínicos ou conferências de consenso, com caráter acadêmico (avaliação absoluta), ser relativa a outras instituições semelhantes (avaliação empírica) ou ser realizada através de auto-comparações no tempo (avaliação institucional). Sistemas que permitam a comparação entre diferentes instituições são os mais amplamente utilizados, mas vários estudos demonstram claramente que os desfechos mais utilizados devem ser ajustados pela gravidade do caso ou por condições próprias do paciente, em um processo conhecido como “ajuste por risco”. O ajuste por risco (ou padronização por severidade) visa minimizar a influência confundidora de potenciais fatores de risco inerentes ao paciente e a sua patologia, permitindo assim a comparação da qualidade do tratamento hospitalar, ajustando os desfechos encontrados ao risco prévio do paciente (6).

O Hospital de Clínicas de Porto Alegre (HCPA) é o hospital-escola da Universidade Federal do Rio Grande do Sul que atende prioritariamente aos pacientes do Sistema Único de Saúde. Com capacidade instalada de 735 leitos, é considerado um Hospital de excelência e de referência para diversas especialidades, tendo sido o primeiro hospital público a obter a Certificação do Programa Nacional de Acreditação Hospitalar. Anualmente são realizados no HCPA cerca de 25.000 internações, 600.000 consultas ambulatoriais, 2.200.000 exames, 4.000 partos e 30.000 cirurgias.

Desde 1996, o Hospital de Clínicas de Porto Alegre vem desenvolvendo seu sistema de indicadores de qualidade assistencial. Este projeto iniciou com a definição do conjunto de 35 indicadores mais relevantes para a instituição (7), que estão sendo progressivamente implementados. Uma das grandes dificuldades encontradas foi a comparação com referenciais externos adequados. O referencial internacional selecionado foi o Projeto de Indicadores de Qualidade (*Quality Indicator Project -QI Project*), criado em Maryland, EUA, em 1989 (8).

No Brasil, o HCPA participa do projeto SIPAGEH (Sistema de Indicadores Padronizados para Gestão Hospitalar), criado em 1998 em Porto Alegre, por 9 hospitais gaúchos (9), e que hoje agrega mais de 50 hospitais públicos e privados de vários estados, que compartilham informações sobre 19 indicadores e utiliza também os dados sobre atenção hospitalar do SUS disponíveis no site do DATASUS (10).

O uso de informações comparativas não se mostrou de utilidade para o HCPA, pois os hospitais participantes apresentam características muito diversas em relação às suas

naturezas, aos seus portes, às suas características de atenção e ao perfil de seus pacientes, o que tem dificultado a utilização adequada de seus dados como critério de comparação. A título de demonstração desta situação, a Figura 1 apresenta os dados relativos às taxas de mortalidade não ajustada de 25 hospitais de diferentes características, participantes do SIPAGEH, e o desempenho relativo do HCPA.

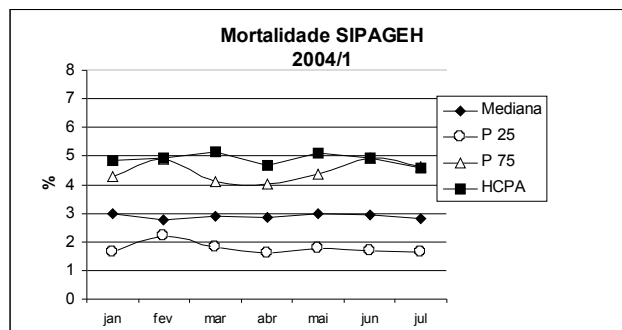


Figura 1: Mortalidade geral- SIPAGEH janeiro-julho 2004

Tornou-se necessário, então, o desenvolvimento de procedimentos que permitam mensurar o grau de risco dos pacientes para diferentes desfechos hospitalares, para ajustar os indicadores e permitir a comparação adequada de indicadores de desempenho de diferentes hospitais e para identificar precocemente aqueles pacientes de risco elevado para determinados desfechos.

O objetivo principal deste projeto de pesquisa foi o de sistematizar, testar e avaliar uma metodologia para o ajuste de indicadores de desfechos pelo risco prévio dos pacientes, permitindo a melhora da qualidade e da segurança da assistência prestada a estes (alerta e prevenção) bem como a comparação adequada do desempenho assistencial entre diferentes hospitais.

São seus objetivos secundários:

- Desenvolver, testar e validar um escore de risco para mortalidade cirúrgica pós laparotomia exploradora em pacientes do HCPA
- Avaliar a disponibilidade, a qualidade e a adequação dos dados de prontuários clínicos do HCPA
- Documentar, a partir da revisão da literatura e da aplicação feita, uma seqüência adequada de etapas necessárias para este tipo de estudo.

I - Revisão da literatura

1- Qualidade, Segurança e Indicadores assistenciais

Segundo Donabedian (11), antes de se pensar em avaliar a *qualidade da assistência*, tanto em termos gerais como em situações específicas, é necessário que haja um acordo sobre como esta qualidade é definida e quais os elementos que a constituem. A definição de qualidade é complexa, pois deve fazer referência a todas as suas dimensões e dependerá do ponto de vista de quem a define. Mas parece haver o consenso de que a qualidade assistencial é um conceito complexo e amplo, cujos componentes podem ser agrupados em sete atributos ou pilares fundamentais (12):

- Eficácia, que é a habilidade da ciência médica em oferecer melhorias na saúde e no bem-estar dos indivíduos.
- Efetividade, que é a relação entre o benefício real oferecido pelo sistema de saúde ou assistência e o resultado potencial de um “sistema ideal”.
- Eficiência,: que é a habilidade de obter ao melhor resultado ao menor custo, isto é, a relação entre o benefício oferecido pelo sistema de saúde ou assistência médica e seu custo econômico.
- Otimização, que é o balanço mais vantajoso entre custo e benefício, ou seja, é o estabelecimento do ponto de equilíbrio relativo, em que o benefício é elevado ao máximo em relação ao seu custo econômico.
- Aceitabilidade: é a adaptação dos cuidados médicos e da assistência à saúde às expectativas, desejos e valores dos pacientes e suas famílias. Este atributo é composto por 5 conceitos: acessibilidade, relação médico-paciente, hotelaria, preferências do paciente quanto aos efeitos da assistência, preferências do paciente quanto aos custos da assistência.
- Legitimidade, que é a conformidade às preferências sociais relativas aos aspectos acima, isto é, a possibilidade de adaptar satisfatoriamente um serviço à comunidade ou à sociedade como um todo. Implica conformidade individual, satisfação e bem-estar da coletividade.
- Equidade, que é a determinação da adequada e justa distribuição dos serviços e benefícios para todos os membros da comunidade, população ou sociedade.

Todas as atividades de avaliação da qualidade dependem da disponibilidade de informações acuradas e adequadas. A fonte fundamental de informações sobre o processo de assistência e seus resultados imediatos é, sem dúvida, o prontuário do

paciente. Mas sabemos que o prontuário é freqüentemente incompleto, omitindo muitas vezes elementos fundamentais a respeito do diagnóstico e tratamento do paciente e, como norma, não incluindo nenhuma referência sobre os processos interpessoais (relação médico-paciente). Além disso, muitas das informações são imprecisas devido a erros de diagnóstico, observações clínicas, avaliação, registro e codificação. Para uma avaliação adequada da qualidade da assistência, portanto, é necessário um esforço no sentido de criar no hospital uma cultura orientada para a valorização do prontuário do paciente, suplementado pela capacidade de coletar informações de outras fontes.

De acordo com definição da Organização Panamericana de Saúde, indicadores são medidas-síntese que contém informação relevante sobre determinados atributos e dimensões do estado de saúde, bem como do desempenho do sistema de saúde(13). Os indicadores proporcionam as informações necessárias e mensuráveis para descrever tanto a realidade como as modificações devidas à presença do serviço ou assistência. É a variável que descreve uma realidade, devendo para isso ter as características de uma medida válida em termos estatísticos.

A construção de um indicador é um processo complexo e sua qualidade depende das propriedades dos componentes utilizados em sua formulação e da precisão dos sistemas de informação empregados (13). O grau de excelência de um indicador deve ser definido por sua validade (capacidade de medir o que se pretende) e confiabilidade (reproduzir os mesmos resultados quando aplicados em condições semelhantes). Em geral, a validade de um indicador é determinada pelas características de sensibilidade (medir as alterações deste fenômeno) e especificidade (medir somente o fenômeno analisado). Outros atributos de qualidade de um indicador são a sua mensurabilidade (basear-se em dados disponíveis ou de fácil obtenção), relevância (responder a prioridades de saúde) e custo-efetividade (os resultados justificam o investimento de tempo e recursos). É desejável que os indicadores possam ser analisados e interpretados com facilidade (simplicidade) e que sejam compreensíveis pelos usuários da informação.

A qualidade e a comparabilidade dos indicadores de saúde podem ser asseguradas pela aplicação sistemática de definições operacionais e de procedimentos padronizados de medição e de cálculo. A seleção do conjunto básico de indicadores e de seus níveis de desagregação (estratos) pode variar em função da disponibilidade de sistemas de informação, fontes de dados, recursos, prioridades e necessidades específicas em cada instituição, região, país etc. A manutenção deste conjunto depende também da simplicidade dos instrumentos e métodos utilizados, de modo a facilitar a operação regular dos sistemas de informação.

Se forem gerados de forma regular e manejados em um sistema dinâmico, os indicadores de saúde constituem ferramenta fundamental para a gestão e avaliação da situação de saúde em todos os níveis. Um conjunto de indicadores tem como propósito produzir evidências sobre a situação de saúde e suas tendências, que devem servir de base empírica para determinar prioridades de ação. Constitui assim insumo para o estabelecimento de políticas e prioridades mais ajustadas às necessidades da população atingida.

Atualmente, no cenário internacional, três iniciativas relevantes se destacam na área de indicadores de saúde: o projeto da Associação de Hospitais de Maryland, que hoje agrega mais de 1000 hospitais e que objetiva avaliar a qualidade da assistência prestada aos pacientes(8), o Sistema de Indicadores Padronizados para Gestão Hospitalar (SIPAGEH) (9) e os sistemas de informação em saúde gerenciados e disponibilizados pelo DATASUS/Ministério da Saúde, através de iniciativas e projetos da Rede Interagencial de Informações para a Saúde (RIPSA) (10).

Vários componentes do sistema de saúde têm considerável interesse em usar medidas de desfecho para avaliar o desempenho do sistema de assistência à saúde. Usar estes indicadores é possível se a informação fornecida é válida para fazer comparações entre instituições. Infelizmente, existe uma variedade de desafios no uso destes indicadores para este fim. Esta avaliação do indicador é uma abordagem para avaliar estas medidas e pode ser uma ferramenta útil nesta área de pesquisa. Medidas de qualidade devem ser vistas não como um simples exercício acadêmico, mas devem estimular o sistema de saúde a prover serviços melhores o que, em troca, resultará em melhor saúde. Para atingir a meta de medir qualidade assistencial, nós devemos ter ferramentas para selecionar objetivamente estas medidas que nos levem o mais rápido possível aos nossos objetivos (14).

2- Ajuste de risco (*Risk Adjustment*)

2.1- História

A discussão sobre ajuste por riscos não é recente. Em 1863, Florence Nightingale (1820-1910) publicou a terceira edição de seu livro "*Notes on Hospitals*"(15), recomendando mudanças fundamentais na configuração, localização e operação dos hospitais para reduzir o número de mortes causadas por condições precárias de higiene e segurança. Sete anos antes, ela havia retornado como heroína da Guerra da Criméia, onde serviu ao Exército inglês em hospitais militares. Em 1855, seis meses após sua chegada ao cenário da guerra na Albânia, a mortalidade nestes hospitais caiu de 42,7% para 2,2%. Ao voltar para Londres, ela continuou seu trabalho em hospitais militares e foi

buscar apoio estatístico com Willian Farr (1807-1883), médico e reformista social famoso, que trabalhava com análises do Registro Geral desde 1838.

Em 1856, ambos começaram um trabalho conjunto, analisando dados e propondo mudanças em hospitais civis e militares. Baseados no 24º Relatório Anual do Registro Geral de Farr, eles analisaram a mortalidade nos 106 principais hospitais da Inglaterra (figura 2) chegando a uma surpreendente taxa de mortalidade de 90.84% nos 24 hospitais de Londres contra 13% em média nos hospitais da área rural. Farr calculou as taxas de mortalidade dividindo o número total de óbitos no ano pelo número de pacientes internados em um determinado dia do ano (8 de abril); se calculado da forma usual atual (número de óbitos dividido pelo número total de pacientes no ano), a taxa de mortalidade nos hospitais londrinos naquela época seria em torno de 10% (Figura 2).

Os números publicados causaram grande polêmica nos círculos médicos da Inglaterra, tendo os médicos de Londres reagido contra a publicação de comparações de taxas de mortalidade entre hospitais ingleses. Segundo estes, Farr não apenas utilizou métodos estatísticos questionáveis como também não levou em conta as diferenças nas características dos pacientes dos diferentes hospitais. “Qualquer comparação que ignore a diferença entre um agricultor que busque alívio para dores musculares em Stoke Pogis (zona rural) e um mecânico realmente doente de Soho ou Southwark que vem a um hospital de Londres, é falaciosa”(15). A própria Florence se convenceu dessa falha e observou que os números absolutos de óbitos em hospitais não expressavam adequadamente a qualidade dos cuidados, pois se deveria considerar “diferenças na idade e no estado do paciente na admissão”.

Nos Estados Unidos, O Dr. Ernest Codman, nos primórdios do século XX, introduziu o conceito de acompanhamento de pacientes e avaliação da efetividade do tratamento, criando o conceito de “desfechos”. Tanto Nightingale quanto Codman viram informações sobre indicadores hospitalares como um meio de melhorar os desfechos dos pacientes e a qualidade da assistência. Seus trabalhos deixaram uma importante lição: medir apenas taxas de eventos não é suficiente. Entender porque estes eventos ocorrem é essencial.

	Number of SPECIAL INMATES on the 8th April, 1861.	Average Number of INMATES in each HOSPITAL.	Number of DEATHS registered in the Year 1861.	MORTALITY per Cent. on INMATES.
IN 106 PRINCIPAL HOSPITALS OF ENGLAND	12709	120	7227	56·87
24 London Hospitals	4214	176	3828	90·84
12 Hospitals in Large Towns	1870	156	1555	83·16
25 County and Important Provincial Hospitals	2248	90	886	39·41
30 Other Hospitals	1136	38	457	40·23
13 Naval and Military Hospitals	3000	231	470	15·67
1 Royal Sea Bathing Infirmary (Margate)	133	133	17	12·78
1 Dane Hill Metropolitan Infirmary (Margate)	108	108	14	12·96

Figura 2- Tabela publicada na 3ª edição de “Notes on Hospitals”, de Florence Nightingale (15)

Atualmente, poucos pesquisadores ou gestores se atrevem a comparar desfechos de pacientes entre hospitais sem um mínimo de ajuste para as diferenças entre os riscos destes pacientes (16). É consenso geral que alguns hospitais tratam doentes mais graves e estes hospitais não podem ser penalizados por aceitar pacientes de alto risco. Por outro lado, sem ajuste de risco, hospitais com baixo desempenho podem argumentar que seus pacientes são mais graves como justificativa para seus indicadores ruins.

Pesquisadores começaram a criar ferramentas para medidas sistemáticas de severidade na década de 70, em geral para atender a necessidades locais. Nesta época, poucos dados estavam disponíveis e os modelos eram construídos através do julgamento de especialistas ou observações clínicas em instituições específicas. Se baseavam no modelo clínico de severidade (diagnóstico, achados fisiológicos, resultados de exames) (17).

A adoção da modalidade de pagamento por grupos de diagnósticos relacionados pelo sistema americano de assistência à saúde *Medicare*, em 1983, elevou as preocupações com severidade a um nível nacional, através da disponibilidade destes diagnósticos nas grandes bases de dados administrativas nacionais, codificados pela Classificação Internacional de Doenças da OMS – 9ª revisão. Apesar das limitações clínicas destas bases de dados, elas apresentam importantes vantagens como a uniformidade, disponibilidade, acesso eletrônico e ampla cobertura. Em 1986, a Administração de Financiamento da Assistência à Saúde americana (*Health Care Financing Administration - HCFA*) foi obrigada a publicar seus dados sobre mortalidade hospitalar, a partir de uma ação movida pelo *New York Times*. Estes dados estavam mal ajustados, com falhas metodológicas importantes, causando um grave prejuízo a muitos hospitais mal

avaliados. Nesta época, alguns estados americanos passaram a exigir informações dos hospitais sobre seus pacientes, o que possibilitou a criação de uma base de dados comparativa pelo “Medicare Group”. Outras iniciativas importantes foram conduzidas, com destaque para os trabalhos do Departamento de Assuntos dos Veteranos (*Veteran’s Affairs Surgical Risk Study (18-23)*). Com estas bases de dados clínicas, os modelos e suas metodologias puderam ser refinados, sendo obtidos através de processos estatísticos mais avançados. Esta nova geração de medidas de severidade parecia ser mais rigorosa cientificamente e com maior credibilidade clínica (16).

2.2- Por que ajustar por risco?

O ajuste de risco (ou padronização por severidade) visa minimizar a influência confundidora de potenciais fatores de risco inerentes ao paciente e/ou a sua patologia, permitindo assim a comparação da qualidade do tratamento hospitalar, ajustando os desfechos encontrados ao risco prévio do paciente (dados demográficos, sócio-econômicos e diagnósticos). O objetivo desse é o de controlar para a influência confundidora da severidade da doença nas comparações que podem ser relacionadas a esta severidade. Ajustamentos por severidade teoricamente minimizam o potencial para falsos julgamentos sobre qualidade assistencial devido a desfechos indesejados cuja prevenção ou controle estão fora do controle do provedor (24).

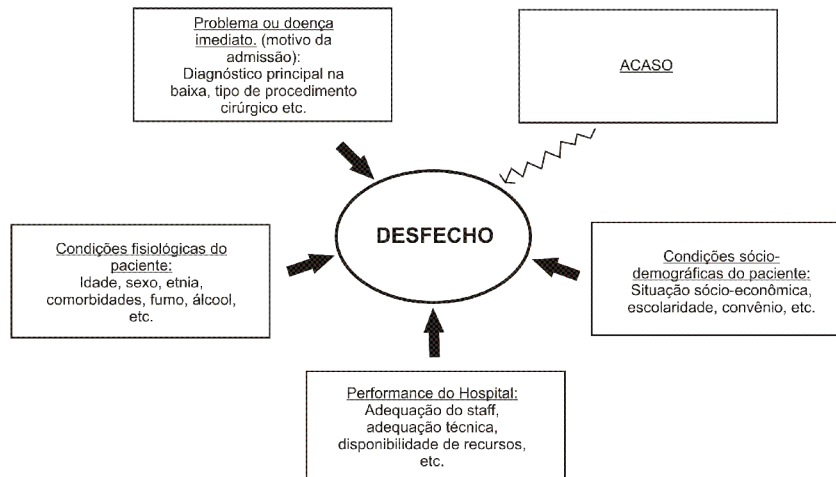
O Conselho de Serviços Médicos da Associação Médica Americana (AMA) define desfecho como o resultado da atenção à saúde que pode englobar alterações biológicas na doença, conforto, habilidade de auto-cuidados, função física e mobilidade, bem estar emocional, desempenho intelectual, satisfação do paciente, conhecimento médico e concordância com a assistência e viabilidade de funcionamento de seu papel social, familiar e de trabalho. O espectro de desfechos tradicionais incluem os assim chamados “5D’s: óbito, incapacidade, doença, desconforto, insatisfação, (*death, disability, diseases, discomfort and dissatisfaction*) (25).

Falando em termos amplos, diferenças de desfechos entre hospitais podem ser vistas como resultado de vários fatores que podem influenciar os resultados da assistência. Segundo lezzoni (6), um desfecho hospitalar pode ser expresso como sendo uma função de fatores de risco inerentes ao paciente, a efetividade da assistência e a sorte:

$$\text{Desfecho} = f (\text{fatores de risco intrínsecos ao paciente, efetividade do tratamento, qualidade da assistência, chance randômica})$$

A figura 3 ilustra esta situação.

Figura 3: Fatores que influenciam os desfechos hospitalares (adaptado de DesHarnais e McLaughlin (24)



2.3- Como ajustar o risco?

Nenhum ajustador de riscos é perfeito. Ajustar para todas as características do paciente não é nem necessário nem possível. No entanto, devem ser feitos esforços no sentido de identificar aqueles fatores que sejam suficientemente válidos para os propósitos explícitos. A meta do ajuste de risco (AR) é considerar as características pertinentes do paciente ao estabelecer inferências sobre a efetividade ou a qualidade da assistência. Ajuste de risco é um termo amplo que está juntando uma série de outros termos mal definidos e mal entendidos mas frequentemente utilizados, tais como Grupos de Diagnósticos Relacionados (*DRGs*), *mix* de casos (*case-mix*), severidade, doença, intensidade, complexidade, comorbidade e situação de saúde (24).

Devido às diferentes definições de risco, planejar estratégias de ajuste é complicado, requerendo no mínimo a identificação de fatores de risco – as características clínicas e outras que refletem a probabilidade do paciente de vir a sofrer o desfecho em estudo. Uma vez identificados, estes fatores de risco devem ser operacionalizados de forma a permitir suas identificações e medidas de forma confiável e sem vieses e precisam ser combinados em formatos normativos ou empíricos para produzir um escore de risco ou uma classificação para cada paciente.

Idealmente devemos usar fatores de risco específicos de cada condição, No entanto, avaliações corretas devem levar em conta apenas as condições existentes no momento da internação do paciente, não utilizando fatores de risco possíveis de serem “iatrogenias” da internação. Por vezes é difícil separar, em registros de alta hospitalar, condições pré-existent das complicações ocorridas durante a internação. De qualquer forma, o uso de informações individuais disponíveis em bases de dados que apoiam sistemas de prontuários eletrônicos são fontes adequadas de dados suficientes

para o desenvolvimento de *proxies* adequadas para fatores de risco. Uma vez controlando por fatores de risco inerentes ao paciente, pode-se obter comparações entre hospitais ou entre diferentes períodos de tempo muito melhores, embora não perfeitas.

O desenvolvimento de um sistema de avaliação de risco que permita o ajuste de indicadores pode ser desenvolvido através de diferentes metodologias. A partir de bases de dados clínicas, os seguintes passos podem ser adotados (24):

- 1- avaliar a qualidade (completude, fidedignidade, acurácia, disponibilidade etc.) e a consistência dos dados disponíveis.
- 2- Coletar os dados adequados (sugeridos pela literatura, dados disponíveis plausíveis para o desfecho em estudo).
- 3- Desenvolver modelos de ajuste por risco, assinalando a probabilidade predita para cada desfecho adverso relevante para cada caso. Esta etapa é, geralmente, realizada através de técnicas de modelagem por regressão logística ou politômica.
- 4- Desenvolver relatórios para cada hospital (ou, dentro do Hospital, para cada equipe ou cada período de tempo) comparando as frequências observadas com aquelas preditas pelo modelo, para cada desfecho.
- 5- Avaliar se as diferenças encontradas são estatisticamente significativas.
- 6- Usando estes relatórios, desenvolver perfis de risco para comparar os hospitais dentro de cada estrato de risco.

Mesmo com estas abordagens não se deve classificar um hospital como bom ou ruim apenas baseados nestes escores de índices de risco, pois devemos reconhecer suas limitações. Além disso, é importante levar em conta que a qualidade dos dados disponíveis não é uniforme entre diferentes hospitais e isso poderá afetar os escores obtidos.

Os dados disponíveis para a avaliação dos fatores de risco e construção dos modelos de ajuste representam um importante componente do sucesso destas estratégias. A maior parte dos estudos publicados utilizou bases de dados administrativas, construídas com o propósito de cobrança (dados secundários), enquanto outros utilizaram dados clínicos adicionais coletados de prontuários de pacientes (dados primários). Os estudos feitos com dados secundários são mais simples, rápidos e baratos, mas com menor credibilidade clínica. Por outro lado, fontes primárias podem não ser completas, acuradas ou válidas, de acordo com a qualidade do registro médico de cada hospital.

Uma vez que o método de ajuste de risco é desenvolvido, este deve ser testado para ver quão bem é capaz de prever o desfecho de interesse de uma perspectiva. É importante salientar que dificilmente um modelo será capaz de prever desfechos

perfeitamente e que sempre existe um potencial para eventos aleatórios imprevisíveis. Considerando as limitações e a infinita variabilidade das características biológicas humanas, não se pode esperar que abordagens de ajuste de risco forneçam uma previsão perfeita de risco para pacientes individuais, pois esta metodologia é uma ferramenta inerentemente imperfeita, embora necessária.

A construção de um sistema de indicadores de qualidade que garanta sua utilização na tomada de decisões na busca da melhoria continuada dos padrões da assistência oferecida aos pacientes e que permita sua avaliação permanente baseada em critérios apropriados é um desafio de grande interesse para a área de Epidemiologia hospitalar. Os referenciais adequados de comparação devem ser buscados em nosso próprio meio e, quando inexistentes, podem e devem ser construídos de forma coletiva.

É importante identificar fatores de risco adequados a este ajuste, que devem ser capazes de compor um modelo de risco adequado em termos de acurácia, calibração e discriminabilidade (6). Criar um método de ajuste por risco que seja clinicamente razoável e ao mesmo tempo logisticamente realizável é desafiador e exige algumas concessões, mas é condição essencial para a correta avaliação da qualidade assistencial.

3- Escores de Risco utilizados para pacientes cirúrgicos

Sistemas de escores de risco tentam quantificar o risco de óbito ou de complicações dos pacientes. O escore é baseado na severidade da doença ou em doenças pré-existentes ou combinação dos dois. O sistema ideal de escore para fins de auditorias cirúrgicas deve considerar tanto a mortalidade quanto a morbidade e deve permitir a avaliação retrospectiva dos atos cirúrgicos. Deve ser de uso rápido e fácil e ser aplicável a todos os procedimentos de cirurgia geral em ambientes de emergência ou não. Deve poder ser usado em qualquer tipo de hospital, prover informações relevantes para o ensino e a pesquisa e deve ser possível de ser integrado a programas pré-existentes sem muito esforço (26).

Existem vários sistemas de escores que prevêm o risco de mortalidade cirúrgica com variados graus de acurácia, mas em geral a morbidade é ignorada. Alguns escores são desenvolvidos para cirurgias específicas ou para avaliar complicações específicas. O mais conhecido destes é o sistema APACHE II (27), ideal para avaliar as condições de pacientes internados em Unidades de Tratamento Intensivo (UTI) e que determina a probabilidade de óbito.

Teoricamente, quanto menor o número de variáveis incluídas, mais simples é o sistema de escore. No entanto, alguns escores reduzem tanto o número de variáveis que

a ausência de uma delas inviabiliza seu cálculo. Outros exigem equações complexas que não podem ser feitas pelos cirurgiões em seu dia-a-dia.

Um modelo ideal para prever complicações em pacientes cirúrgicos deve ser simples e facilmente aplicável para todos os pacientes. Na rotina diária, poucos pacientes necessitam de RX de tórax, ECG e/ou exames de sangue para a avaliação pré-cirúrgica. Portanto, estes parâmetros não são adequados para uso em modelos preditivos de complicações na população geral (28).

3.1- *Acute Physiology and Chronic Health Evaluation - APACHE*

O sistema APACHE original foi proposto nos Estados Unidos por Knaus e colegas em 1981 (29) como um escore composto por pesos relativos de 34 medidas fisiológicas, a soma das quais produzia o “Escore Fisiológico Agudo” (APS). O sistema de atribuição de pesos era baseado em uma escala de 0 a 4 para cada variável e o conjunto de variáveis e pesos relativos foram propostos por um conjunto de especialistas.(29). Todas as variáveis estavam disponíveis no prontuário e o objetivo dos autores foi o de propor um sistema de classificação com bases fisiológicas para medir a severidade das condições clínicas de pacientes admitidos à CTI, permitindo desta forma o ajuste de risco, a comparação de indicadores, a avaliação de novas terapias bem como estudar a utilização de CTIs.

O APACHE II foi proposto pelos mesmos pesquisadores em 1985 (27), usando uma escala de pontos calculada a partir de 12 variáveis fisiológicas medidas rotineiramente, mais a idade e a presença de alguns diagnósticos, para fornecer uma medida geral da severidade do quadro clínico. O objetivo era ainda o de avaliar a utilização adequada de recursos e comparar a eficácia dos cuidados intensivos entre diferentes hospitais, não sendo proposto para avaliar as condições ou probabilidade de sobrevida de pacientes individuais. As variáveis e seus pesos foram determinados por grupos de especialistas, através de julgamento clínico e evidências científicas. As principais mudanças em relação ao modelo original foram a redução do número de variáveis fisiológicas e a inclusão de outras variáveis como idade, problemas crônicos de saúde e caráter da admissão (Clínica, cirúrgica de urgência e cirúrgica eletiva). Esta proposta era mais utilizável, devido à disponibilidade das variáveis necessárias e à facilidade de cálculo.

O sistema APACHE III foi introduzido também por Knaus e colegas em 1991 (30), com o objetivo de refinar o sistema anterior através da reavaliação das variáveis incluídas e de seus pesos relativos. Desta forma, os autores desejavam maximizar o poder explanatório do APACHE II, melhorar sua discriminação e manter a máxima confiabilidade dos dados. O sistema de prognóstico APACHE III, como passou a ser

chamado, é composto por duas opções: o escore APACHE III que permite a estratificação por grupos de risco e uma equação preditiva que usa os escores obtidos e dados sobre diagnósticos e tratamentos imediatamente anteriores à entrada na CTI, para estimar risco individuais. Este sistema foi desenvolvido a partir de processos estatísticos e não mais baseado na opinião de especialistas.

Barie e colegas (31) compararam o desempenho dos sistemas APACHE II e APACHE III para predição de mortalidade em pacientes cirúrgicos criticamente doentes, através de um estudo prospectivo realizado com 844 pacientes de um Hospital *universitário (Cornell Medical Center)*. Os autores concluem que o APACHE III é mais acurado que o APACHE II, mas que sendo um sistema de escore geral para pacientes em situação críticas, sua utilidade específica para pacientes cirúrgicos ainda não foi estabelecida.

3.2- POSSUM – Physiological and Operative Severity Score for enUmeration of Mortality and morbidity

POSSUM foi desenvolvido a partir da identificação, pelo *Royal College of Surgeons of England*, da necessidade de um escore simples que pudesse ser utilizado em cirurgia geral, cujo principal uso fosse o apoio aos processos de auditoria. Inicialmente 62 fatores individuais (48 pré e 14 intra e pós-operatórios) foram submetidos a análise discriminante, que identificou as 35 variáveis mais significativas. O modelo final inclui 12 variáveis (idade, insuficiência cardíaca, doença pulmonar, pressão sistólica, frequência cardíaca, escala de coma Glasgow, hemoglobina, glóbulos brancos, uréia, sódio, potássio e eletrocardiograma) e calcula o risco de mortalidade e morbidade em 4 classes distintas. Análises por regressão logística permitiram a criação de um escore auxiliar, com 6 fatores, para compensar a severidade do procedimento cirúrgico (severidade, procedimentos múltiplos, perda de sangue total, contaminação peritoneal, presença de malignidade, cirurgia de urgência). A partir do modelo foram atribuídos pontos a cada uma, baseados nos coeficientes obtidos.

O sistema foi desenvolvido entre 1988 e 1989, com dados de todos os pacientes internados submetidos à cirurgia no Walton Hospital de Liverpool (1372 pacientes elegíveis de 1440). Todos os pacientes foram classificados antes da cirurgia com o escore fisiológico e no momento da alta, com o escore de gravidade da cirurgia. O valor preditivo dos modelos foi avaliado através da área sob a curva ROC e da comparação entre valores observados e esperados. Este sistema não foi desenvolvido com o objetivo de avaliar pacientes individuais ou apoiar a decisão médica de operar ou não operar um

determinado paciente. Seu objetivo é o de apoiar avaliações comparativas entre equipes, hospitais ou regiões.

Neary e colegas (32) publicaram uma avaliação sistemática do uso do POSSUM como modelo preditivo de mortalidade cirúrgica e concluíram que, quando usado corretamente, POSSUM pode ser utilizado para comparar desfechos entre cirurgiões e entre hospitais. Em cirurgias específicas, equações de regressão logística individuais podem ser necessárias. A busca foi feita em 3 bases de dados (Medline, Cochrane Library e Embase) de 1990-2002. Embora exista uma vasta literatura relatando diferentes experiências com o uso do POSSUM, nenhum estudo randomizado foi localizado.

P-POSSUM

Alguns autores, especialmente de Portsmouth, Reino Unido, não conseguiram reproduzir o poder preditivo da equação original e criticaram a técnica analítica empregada, mas não o conjunto de dados. Whiteley e colegas, em 1996, mostraram como o POSSUM falhava em pacientes desta localidade (33;34).

Estes autores verificaram que o POSSUM superestimava a mortalidade em classes de pacientes de baixo risco, concluindo que era possível usar o conjunto de variáveis, mas que uma nova equação de regressão logística era necessária. Esta nova equação se tornou conhecida como P-POSSUM, que usa diferentes valores para os pontos de corte de cada variável, através da metodologia descrita por Hosmer e Lemeshow (35). Copeland e colegas sustentam que a equação original é superior.

No entanto, Bann e colegas (36) afirmam que há enganos em ambas as versões. Como as variáveis fisiológicas devem ser coletadas o mais próximo possível antes da cirurgia, os cirurgiões poderiam tentar “melhorar” os resultados das análises ou selecionar apenas pacientes em condições adequadas, se recusando a operar pacientes graves.

Dados ignorados representam um problema grande. Em muitos casos, nem todas as análises estão disponíveis, pois os exames não se justificam clinicamente. Também existem problemas com algumas variáveis. O ECG causa confusão ao ser classificado em normal, fibrilação atrial ou outros ritmos anormais. Esta última categoria tem feito com que muitos problemas menores sejam classificados de forma errônea. O escore operatório tem elementos de subjetividade, pois o volume exato da perda sangüínea não pode ser avaliado, bem como a contaminação do peritônio.

POSSUM tem várias limitações: erra em cirurgias ou pacientes de baixo risco, prediz apenas mortalidade em 30 dias, mas mesmo assim é o sistema de risco cirúrgico mais estudado e avaliado internacionalmente e os dados exigidos por este são atualmente coletados rotineiramente em vários hospitais, especialmente na Inglaterra.

Finalmente, os resultados de análises comparativas com POSSUM não podem ser usados para atribuir culpa ou punir pessoas. A assistência a pacientes é de responsabilidade multidisciplinar e se as avaliações mostram uma razão alta entre a mortalidade observada e a esperada, a prática de toda a equipe deve ser revista. POSSUM é simplesmente uma ferramenta para auditorias comparativas e é metodologicamente comparável a outros sistemas sofisticados de análise de *case-mix* tal como o APACHE.

Diferentes resultados foram reportados por diversos autores, avaliando a acurácia preditiva do POSSUM e P-POSSUM. Yii e colegas (37), em uma amostra de 605 pacientes com taxa de óbito de 6,1%, encontraram uma taxa prevista pelo POSSUM de 10,5% e, pelo P-POSSUM, de 4,8%. Copeland e colegas (38) relatam, a partir de uma amostra de 1372 pacientes (4% de mortalidade), que o POSSUM apresentou uma especificidade de 99,3%, sensibilidade de 54,1% e acurácia geral de 96,5%. Bann e Sarin (36) avaliaram 501 pacientes cirúrgicos, dos quais obtiveram os dados necessários para cálculo do POSSUM em apenas 155 destes (29,8%). Bennet-Guerrero e colegas (39) conduziram um projeto conjunto americano e britânico, no qual 1539 pacientes ingleses (9,9% de mortalidade) e 1056 pacientes americanos (2,1% de mortalidade) foram avaliados pelo P-POSSUM. No grupo inglês, o P-POSSUM previu 156 óbitos, tendo sido observados 152. No entanto, entre os pacientes americanos, com baixa taxa de mortalidade, a previsão foi de 82 óbitos, sendo observados apenas 22.

Em conclusão, embora POSSUM e P-POSSUM sejam potencialmente acurados na predição de óbito pós-cirúrgico em amostras com taxa de mortalidade elevadas, uma grande proporção de pacientes são excluídos de sua análise pela exigência de um grande número de variáveis e por considerar variáveis sobre a cirurgia e o cirurgião no modelo final, o que prejudica seu uso para comparação de desempenho entre hospitais ou equipes. A questão a que ele responde é: dado este paciente, esta cirurgia e este cirurgião, qual é a probabilidade de óbito. No entanto, deveria responder a “dado este paciente, qual é a probabilidade de óbito?”

3.3- *Simplified Acute Physiology Score (SAPS II)*

O sistema SAPS, proposto por Le Gal e colaboradores em 1984 (40;41), é composto por 14 variáveis clínicas e biológicas e visa avaliar a probabilidade de óbito de pacientes admitidos a uma UTI. O índice de risco deve ser calculado dentro das primeiras 24 horas da internação, dando tempo suficiente para a coleta dos dados necessários.

A partir do Estudo Europeu/Norte-americano realizado entre 1991 e 1992, foi proposto o SAPS II, criado a partir de um método desenvolvido para converter o score

em probabilidade de óbito hospitalar. Análises de regressão logística foram utilizadas para selecionar as variáveis, decidir o peso relativo das mesmas e converter o escore em probabilidade de óbito hospitalar. Este estudo envolveu 137 CTIs médicas, cirúrgicas e mistas em 12 países. O modelo assim desenvolvido trabalha com 17 variáveis, 12 delas fisiológicas, 3 diagnósticas (AIDS, Câncer metastático e malignidade hematológica) e mais idade e tipo de admissão (cirurgia agendada, cirurgia de emergência ou internação clínica). O modelo foi desenvolvido em uma coorte de 8369 pacientes e validado em outro grupo de 4628 pacientes e testou 37 variáveis coletadas, permanecendo as 17 significativas. Os testes de adequação estatística indicaram que o desempenho do modelo foi bom ($p=0.883$ e $p=0.104$). O poder discriminatório do modelo foi medida pelo Índice-C, que equivale à área sob a curva ROC, que foi de 0.88 e 0.86 respectivamente.

A performance do SAPS II nesta população de pacientes estudada sugere que as probabilidades calculadas refletem os dados observados. Estes resultados demonstram que o SAPS II é um sistema extremamente efetivo para estimar a probabilidade de óbito para pacientes de CTI.

A comparação com o sistema SAPS original, largamente usado na França e em outros países europeus, sugere que embora a maior parte das variáveis sejam as mesmas, modificações significantes foram introduzidas, principalmente em relação a seus pesos relativos.

Os avanços do controle social e governamental dos sistemas de saúde têm aumentado a procura de referenciais adequados para avaliações comparativas entre instituições. Para isso, é essencial que os indicadores selecionados sejam ajustados para o conjunto de pacientes de cada hospital, para agregar na avaliação a gravidade e a dificuldade dos casos tratados.

É possível categorizar previamente os pacientes cirúrgicos em grupos que reflitam o risco de cada procedimento subsequente. Sistemas de escore, embora não consigam evitar os desfechos, podem identificar pacientes nos quais se poderia não considerar a hipótese cirúrgica. Sistemas de escore estão aqui para ficar. Refinamentos posteriores são claramente necessários mas é de interesse de todos os cirurgiões estar envolvidos em tais desenvolvimentos.

4- Modelos preditivos de mortalidade cirúrgica

As análises comparativas de desfechos cirúrgicos estão se tornando cada vez mais importantes dentro da profissão médica e para gestores dos sistemas de saúde, o governo e para o público em geral. Estas análises deveriam permitir a comparação de taxas de morbidade e mortalidade entre diferentes hospitais e entre diferentes equipes e cirurgiões (42) mas, no entanto, as medidas de qualidade podem variar de acordo com o desfecho esperado de cada cirurgia e a severidade do caso antes da intervenção, tornando a comparação entre hospitais e cirurgiões bastante difícil. Diferenças entre os fatores de risco existentes previamente à da cirurgia devem ser levados em conta antes de se atribuir as diferenças entre os desfechos à habilidade do cirurgião ou ao ambiente hospitalar.

Indicadores de desfecho são os mais apropriados para avaliar a qualidade cirúrgica pois em geral estes ocorrem em torno de um evento previsível -a cirurgia- com um desfecho esperado. O desfecho mais avaliado é a mortalidade cirúrgica, por ser facilmente quantificável. Este desfecho pode ser medido de 3 diferentes maneiras (43):

- 1- Mortalidade intra-hospitalar (mais confiável)
- 2- Mortalidade em até 30 dias após o procedimento (mais abrangente)
- 3- Uma combinação de ambas: mortalidade intra-hospitalar em um período de até 30 dias após a cirurgia.

Medidas de desfecho ajustadas por risco são particularmente adequadas para avaliar a qualidade da assistência, especialmente a qualidade da atenção cirúrgica mas a mortalidade cirúrgica é apropriada como medida de qualidade apenas se ocorre com frequência suficiente para permitir estimativas estatísticas estáveis sendo tipicamente utilizada em especialidades cirúrgicas onde são esperadas taxas relativamente altas de óbito, tais como cirurgia cardíaca e outras.

Vários métodos têm sido propostos para padronizar dados de pacientes a fim de permitir comparações diretas da mortalidade cirúrgica mas a adoção desta metodologia ainda oferece muitos desafios e apresenta algumas fragilidades metodológicas.

O pré-requisito mais importante para o uso de dados de prontuário na avaliação comparativa da qualidade assistencial é a aquisição de informações abrangentes, confiáveis, acuradas e coletadas de forma metódica e que tenham credibilidade clínica. Dados administrativos de sumários de alta hospitalar são fontes baratas e fáceis de dados comparativos e usualmente possuem os dados suficientes para o cálculo de taxas de mortalidade geral ou de médias de permanência, mas possuem algumas limitações. O

sistema ideal é aquele que alia a acurácia, confiabilidade e credibilidade dos dados clínicos com o fácil acesso eletrônico, o que é ainda oferecido por poucos hospitais no mundo que possuem um prontuário eletrônico que apresente qualidade de informação aliada a um volume adequado para pesquisa.

4.1- Projeto Veteranos

Muitos consórcios regionais e nacionais americanos coletam dados para cálculo de indicadores de desfechos cirúrgicos ajustados por risco. O maior e mais importantes destes é o NSQIP *National Surgical Quality Improvement Program*, do *Department of Veterans Affairs* (DVA) (20;21). Neste consórcio, formado por 123 centros médicos vinculados ao DVA, enfermeiras treinadas coletam dados pré, intra e pós-cirúrgicos, garantindo a uniformidade e a padronização das definições, dos instrumentos de coleta e dos programas para aquisição dos dados. São coletadas 60 variáveis pré-operatórias, 17 variáveis intra-operatórias e 33 variáveis de desfecho, de todos os pacientes submetidos a procedimentos cirúrgicos. Os dados são transmitidos para uma unidade central, que realiza a limpeza, armazenamento e análise dos mesmos. Até 2000, esta base de dados continha 700.000 casos coletados em 9 especialidades cirúrgicas.

Os desfechos monitorados são a mortalidade, média de permanência e morbidade. A mortalidade é ajustada apenas para aqueles procedimentos de alta taxa de óbito, para garantir a estabilidade dos modelos. Os modelos ajustados são gerados anualmente e os fatores de risco pré-operatórios e seus pesos relativos têm permanecido estáveis desde 1991.

Os relatórios gerados funcionam como um referencial comparativo (*benchmark*) para os hospitais, o que incentiva a busca de melhores práticas, que são disseminadas entre estes (22;44;45). Desde o início do projeto, o volume, a complexidade e o perfil de risco dos pacientes têm se mantido constantes, verificando-se, no entanto, uma redução de 26% na mortalidade cirúrgica e 55% na morbidade, bem como o aumento da satisfação dos pacientes e a redução da média de permanência de 13 para 4 dias. Estas melhoras se devem a avanços nas técnicas cirúrgicas e nos procedimentos diagnósticos mas também à adoção das melhores práticas e à prática da concorrência positiva.

A base de dados do NSQIP, que inclui dados de 310.000 cirurgias, constitui uma fonte de informações clínicas de alto valor para pesquisa. Os modelos foram calibrados para populações restritas (veteranos do exército americano, em geral idosos, do sexo masculino e pobres) e não podem ser generalizados. Os modelos não incluem todas as variáveis potencialmente preditivas e nem todas as utilizadas se referem ao processo

assistencial. Mas usado intra-organizacionalmente, estes desfechos ajustados têm auxiliado na melhoria da qualidade assistencial.

Vários modelos, gerais e específicos, foram desenvolvidos, validados e publicados pelo NSQIP: Em 1997, Khuri e colegas publicaram os resultados do projeto de desenvolvimento e validação de um modelo de risco para pacientes cirúrgicos (21;22). Este projeto utilizou dados de 87.000 cirurgias não cardíacas, sendo o desfecho o óbito ocorrido até 30 dias após a cirurgia. As taxas de mortalidade, para fins de comparação, foram expressas como a razão entre o observado e o esperado. Foram coletadas 67 variáveis pré, intra e pós-cirúrgicas, de acordo com suas relevâncias clínicas, disponibilidade e confiabilidade do dado. Os dados laboratoriais faltantes foram estimados estatisticamente. As variáveis preditoras mais significativas foram a albumina sérica, a classificação pré-anestésica ASA, idade, caráter de emergência da cirurgia e câncer disseminado. A validade preditiva do modelo foi medida pela avaliação da discriminabilidade e da calibração do modelo nas coortes de desenvolvimento e de validação. O modelo validado foi então utilizado para calcular a probabilidade de desfecho de cada paciente individual. As probabilidades dos pacientes de cada hospital foram somadas para obter o número de óbitos esperado, que foi comparado com o número observado. O ajuste de risco teve um impacto apreciável no ranking de performance dos hospitais e permitiu um meio para monitorar e potencialmente melhorar a qualidade cirúrgica. Baseados nos resultados deste estudo, o VA formou o *National Surgical Quality Improvement Program*, que produz relatórios periódicos com dados sobre características pré-operatórias dos pacientes e desfechos ajustados por risco. Estes dados são enviados junto com as médias nacionais, permitindo a identificação de pontos fracos e fortes de cada instituição, estabelecendo políticas de melhoria contínua de qualidade.

Em 2001, Arozullah e colegas, com dados de 13.152 pacientes desta mesma base de dados, desenvolveram e validaram um modelo preditivo de pneumonia pós-operatória (44). As variáveis relevantes e disponíveis foram modeladas por regressão logística, a partir de dados dos pacientes da coorte de desenvolvimento. Utilizando a metodologia proposta por Le Gall e colegas (41), os pesquisadores atribuíram pontos para cada variável, produto da multiplicação por 10 e arredondamento dos coeficientes de regressão obtidos. O total de pontos de cada paciente foi denominado de Índice de Risco para Pneumonia e foi usado em nova equação de regressão logística para converter o índice de risco em probabilidade do desfecho. O poder discriminatório do modelo (área sob a curva ROC) foi de 0,779 para a coorte de desenvolvimento e de 0,792 para a coorte de validação. O índice de risco desenvolvido, segundo os autores, representa um avanço em

relação aos publicados anteriormente por utilizar dados clínicos colhidos prospectivamente, uma grande base de dados nacional e metodologia estatística qualificada. Mesmo assim, os resultados podem não ser generalizáveis para hospitais gerais, porque os hospitais dos Veteranos recebem pacientes graves, com múltiplas comorbidades, homens, idosos e pobres. Este modelo poderá ser utilizado para controlar fatores de risco relativos ao paciente e promover intervenções que visem modificar fatores intra e pós-operatórios. O índice pode ser útil também para ajustes de indicadores em estudos comparativos de equipes, procedimentos ou hospitais.

Em 2003, Webster e colegas (45) publicaram os resultados de uma pesquisa para definir modelos prognósticos de deiscência de sutura após laparotomia. Dados perioperatórios de 17.044 laparotomias (3,4% de deiscência de sutura) foram utilizados para desenvolver o modelo, que foi validado em uma coorte de outros 17.763 pacientes (3,2% de deiscência). O modelo desenvolvido foi utilizado para criar um sistema de escore, chamado de índice de risco para deiscência de sutura abdominal, com o objetivo de identificar pacientes em risco e de orientar o manejo dos pacientes. A metodologia utilizada foi a mesma anteriormente descrita por Arozullah e colegas (44), sendo atribuídos pontos para cada fator de risco multiplicando-se os coeficientes β obtidos na regressão logística por 10 e arredondando para o inteiro mais próximo. Foram calculados os pontos de cada paciente individual, nas duas coortes. Os valores deste índice de risco foram divididos em 4 diferentes grupos. Para cada classe, foi calculada a probabilidade média de acordo com o modelo. Multiplicando estas médias por 100, se produziu a proporção esperada de deiscência que foram comparadas com as proporções observadas em cada classe. O desempenho do modelo foi avaliado através da sua calibração (estatística de Hosmer e Lemeshow) e discriminação (índice-C) nas duas coortes. O modelo resultante apresentou índice-C de 0,731 e 0,737 e estatística de Hosmer e Lemeshow de 0,61 e 0,82 para as coortes de desenvolvimento e de validação, indicando boa discriminação e calibração. Este índice pode ser usado pelas equipes cirúrgicas para apoiar decisões sobre pacientes individuais bem como uma ferramenta de estratificação em estudos que avaliem intervenções para prevenir ou tratar deiscência de sutura cirúrgica ou para comparar o desempenho assistencial de diferentes equipes ou hospitais.

4.2- Outros modelos

Laparotomia de urgência é um procedimento comum e tem uma taxa de mortalidade consideravelmente maior que as eletivas. Pacientes mais idosos têm maior probabilidade de morrer, provavelmente devido às comorbidades. A taxa de mortalidade é alta (10-55%)

mas as comparações entre hospitais ou equipes cirúrgicas é difícil devido aos diferentes *case-mix*. Dos óbitos reportados pelo NCEPOD (*National Confidential Enquiry into Perioperative Deaths*), 86% foram de pacientes acima de 60 anos e 57% em cirurgias de urgência e emergência. Cook e Day (46) estudaram 107 pacientes com 65 anos e mais, submetidos a laparotomia de urgência, para determinar os fatores de risco que influenciam a previsão dos anestesistas, feita logo antes e logo depois da cirurgia. Foi obtido um modelo que responde por 93% da variabilidade da probabilidade de sobrevida. Idade e ASA foram os preditores mais significantes de sobrevida e de predição dos anestesistas. Vários outros fatores foram determinantes significantes de sobrevida mas não da opinião dos anestesistas.

Auditorias cirúrgicas comparativas estão se tornando um procedimento importante na avaliação da qualidade da assistência nos Estados Unidos embora apresente dificuldades devido às análises ajustadas por risco. Vários métodos têm sido propostos para resolver este problema e, em resposta a estas preocupações, Sutton e colegas consideraram um processo alternativo, baseado em dados coletados facilmente independente dos cirurgiões e que podem prover uma predição acurada de mortalidade (42) Foram incorporados os critérios do NCEPOD, a classificação ASA (*American Society of Anesthesiology*) e BUPA (*British United Provident Association*) para a criação de um sistema de score chamado de *Surgical Risk Score* (SRS). O CEPOD classifica as cirurgias em eletivas, agendadas, urgentes e de emergência e seu grau é atribuído pelo médico assistente (de 1 a 4). O BUPA classifica as cirurgias por grau de complexidade, em 5 classes, e seu grau é atribuído pelo cirurgião. A classificação ASA divide os pacientes em 5 classes, baseada na avaliação do anestesista após a anamnese. O score foi desenvolvido com uma coorte de 3144 pacientes (134 óbitos) e validado em outra coorte de 2024 cirurgias (62 óbitos). A análise feita com o modelo de regressão logística utilizando essas 3 variáveis permitiu a definição de uma escala de risco simples, de valor entre 3 e 14 pontos para cada paciente. Esses pontos foram utilizados como variável independente em outra regressão logística para calcular as probabilidades de óbito. A discriminação do modelo foi de 0,93 e a calibração de 0,65. O sistema de score é simples porque utiliza variáveis comumente registradas em prontuários de pacientes cirúrgicos da Inglaterra e pode ser calculado na beira do leito, sem necessidade de exames ou processos elaborados. Este estudo prospectivo demonstrou que o SRS é uma ferramenta efetiva de predição de óbito. É tão acurado quanto outros sistemas mas mais simples. A área sob a curva ROC foi de 0,95 e do POSSUM foi de 0,84.

Wouters e colegas (47) desenvolveram na Holanda um modelo de predição pré-cirúrgica para morbidade e mortalidade imediatas após cirurgia de revascularização do miocárdio, com o objetivo de distinguir pacientes de alto risco baseado em informação pré-operatória global, disponível para a maioria dos pacientes e útil na prática clínica. Os fatores de risco foram identificados por regressão logística, em uma coorte de 563 pacientes, e validado em outra coorte de 969 pacientes. O peso atribuído a cada variável foi obtido a partir do coeficiente β . A área sob a curva ROC foi utilizada para avaliar a discriminabilidade do modelo. Os fatores de risco mais significantes foram o sexo, hipertensão, doença pulmonar, reintervenção, idade, classificação ASA e função ventricular esquerda, com discriminação de 0,81.

5- Fatores de risco para mortalidade cirúrgica

Na literatura há vários fatores de risco descritos como associados à mortalidade cirúrgica, entre estes os descritos a seguir.

5.1- A classificação do estado físico da Sociedade Americana de Anestesiologistas (*American Society of Anesthesiologists - ASA*)

A classificação do Estado Físico da *American Society of Anesthesiology* (ASA) possui uma interessante história de mais de 60 anos. Originalmente criada para prover uma linguagem comum aos anestesiologistas para descrever a situação pré-anestésica de seus pacientes, tem sido também utilizada para prever morbidade e mortalidade cirúrgica, promover ajustes de risco para estudos comparativos e com menos sucesso, estimar o risco anestésico. Embora a classificação ASA seja usada para fins que excedem suas limitações e propósitos, as características essenciais da escala não foram alteradas desde seu início (48).

A classificação ASA foi introduzida em 1941, pelos doutores Meyer Saklad, Ivan Taylor e E. Rovenstein, em um trabalho intitulado "*Grading of patients for surgical procedures*" (49), um sistema que se encontra em uso até hoje com pequenas alterações. Este sistema apresentava 6 classes baseadas nos graus de "distúrbios sistêmicos", classificados como *nenhum*, *definitivo*, *severo* e *extremo*, com as classes 5 e 6 usadas em casos de emergência. O intento desta classificação era o de melhorar a comunicação entre os médicos, não tendo sido proposta como índice de risco. Os autores reconheciam que o risco cirúrgico é alterado pelo procedimento realizado e que a condição pré-cirúrgica não deveria ser utilizada como uma variável preditora independente. Eles

esperavam que os anestesistas de todo o país pudessem adotar sua “terminologia comum” tornando possíveis comparações estatísticas de morbidade e mortalidade (50).

Os autores enfatizaram desde o início que seu sistema não tinha por objetivo estimar o risco operatório, mas simplesmente descrever as condições pré-cirúrgicas dos pacientes, com o propósito de facilitar a “tabulação de dados em anestesia”, concluindo que seria possível correlacionar os desfechos cirúrgicos com esta classificação e com o tipo e condições da cirurgia. Foi uma contribuição de taxonomia, uma classificação que poderia melhorar a comunicação entre os profissionais da área a qual, segundo seus autores, “pode ser difícil no início, para o anestesista, classificar pacientes de acordo com sua situação física. De forma subconsciente, ele tenderá a usar informações sobre o risco do procedimento específico para influenciar sua graduação. Com cuidado, persistência e atenção a detalhes, em pouco tempo ele se limitará a levar em consideração apenas as condições do paciente para sua classificação.” (51).

Em 1961, o sistema foi alterado, substituindo-se o termo “distúrbio sistêmico” por “doença sistêmica”. Foi também acrescentada uma classe para pacientes moribundos cuja morte era esperada em 24 horas, com ou sem cirurgia e as classes separadas para emergências foram eliminadas, com o acréscimo da letra E modificando as 5 classes remanescentes. Estas revisões foram adotadas pela ASA em 1962 (48;50;52).

Esta escala, segundo o *Quality Indicator Project*, é descrita da seguinte forma (8):

ASA I: Paciente saudável, sem doença sistêmica, a ser submetido a cirurgia eletiva. Paciente fora de limites extremos de idade. Exemplo: paciente hígido a ser submetido a cirurgia de correção de hérnia inguinal

ASA II: Indivíduo com uma doença sistêmica bem controlada. A doença não afeta as atividades diárias. Outros riscos anestésicos, incluindo obesidade, alcoolismo e fumo, podem ser incorporados a este nível.

Exemplos: diabete, hipertensão ou anemia

ASA III: Indivíduo com múltiplas doenças sistêmicas ou uma doença sistêmica grave bem controlada. A situação da doença limita a atividade normal. No entanto, não há perigo imediato de morte devido à doença.

Exemplos: diabetes severa com complicações vasculares, insuficiência pulmonar moderada a severa.

ASA IV: Indivíduo com doença incapacitante severa. Normalmente, o estado da doença está mal controlado ou em estágio terminal. Morte devido a falência de órgãos é um risco existente. Exemplos: graus avançados de insuficiência pulmonar, hepática, renal ou endócrina.

ASA V: Paciente em risco iminente de morte. A cirurgia pode ser o último recurso para salvar sua vida. Não se espera que o paciente sobreviva às próximas 24 horas. Em alguns casos, o paciente pode estar em boa situação de saúde antes de algum acidente que o colocou na condição atual. Exemplos: rompimento de aneurisma abdominal com choque profundo, êmbolo pulmonar massivo, trauma cerebral com pressão intracraniana aumentando rapidamente.

A letra “E” é adicionada a qualquer uma das classificações acima para os pacientes submetidos a procedimento de emergência. Um paciente com morte cerebral declarada, doador de órgãos, é classificado como ASA VI. A Figura 4 sumariza esta classificação.

Classe	Estado Físico
I	Paciente hígido, <80 anos de idade
II	Doença sistêmica leve
III	Doença sistêmica severa, não incapacitante
IV	Doença incapacitante que é um risco constante à vida
V	Moribundo, sobreviva de menos de 24 horas, com ou sem cirurgia
E	Sufixo adicionado à classe para Emergência
VI	Doador de órgãos, morte cerebral

Figura 4- Classificação da Sociedade Americana de Anestesiologistas (ASA)

Este sistema foi criado com o propósito de categorizar o estado físico para padronizar análises estatísticas. Nunca teve a pretensão de ser um avaliador de risco para pacientes individuais, visto que outras variáveis como o diagnóstico, tipo de cirurgia, competência da equipe, instalações e outros podem afetar o desfecho. Muitos estudos relacionados a cirurgia e anestesia têm usado este sistema para auditorias e análises estatísticas. Muitos anestesistas também o utilizam para descrever a situação física de seus pacientes para registro em prontuário e para comunicação entre colegas. Como os hospitais estão cada vez mais atentos a controle de custos, a classificação ASA também tem o potencial de apoiar os gestores na distribuição de recursos e na comparação de performance entre hospitais (53).

A classificação do Estado Físico da Sociedade Americana de Anestesiologistas (ASA) tem sido referenciada de forma inapropriada por alguns autores, que a chamam de classes de risco ASA”, valores ASA” ou escores de risco ASA”. Nenhuma destas nomenclaturas é apropriada. Os doutores Meyer Saklad, Ivan Taylor e E.^a Rovenstein originalmente chamaram o sistema como um sistema de categorização para estudos estatísticos, reconhecendo desde o início, em 1941, que o risco cirúrgico não deveria ser incluído nesta classificação porque é alterado pela natureza da intervenção cirúrgica. A *American Society of Anesthesiologists* já modificou o sistema original muitas vezes nos últimos 59 anos, mas o risco não foi incluído em nenhuma das atualizações (54).

Apesar da negativa dos autores sobre o poder preditivo da classificação ASA, através de seu uso por mais de três décadas esta se tornou uma excelente expressão de “risco cirúrgico”, exatamente o que os autores não propunham. Quando utilizado como preditor, estas estimativas se correlacionam com mortalidade cirúrgica geral. Na tabela 1 estão apresentados resultados publicados por diferentes autores, que demonstram a alta discriminabilidade da classificação ASA.

Tabela 1: Taxas de mortalidade nos diferentes grupos de classificação ASA

AUTOR/FONTE	ASA I	ASA II	ASA III	ASA IV	ASA V
Wolters e col. (55),6301 pacientes	0,1%	0,7%	3,5%	18,3%	93,3%
Marx e col.(56),34.145 pacientes	0,06%	0,47%	4,4%	23,5%	50,8%
Vacanti e col. (57),68.388 pacientes	0,08%	0,27%	1,8%	7,8%	9,4%
HCPA, 880 pacientes cirurgia geral, 2004	0% (0/266)	1,4% (6/419)	17,8% (28/157)	53,3% (16/30)	100% (8/8)
Klück e Fachel (58), 651 pacientes laparotomia exploradora 2000-2003	0% (0/22)	5,4% (12/221)	16,8% (38/226)	65,4% (102/156)	96,2% (25/26)

A principal razão para isso é que a situação física é uma expressão unificada da condição pré-operatória geral do paciente que é consistentemente registrada antes da cirurgia. Por definição, qualquer expressão de risco deve ter valor preditivo e para testar qualquer preditor, a estimativa deve ser feita antes do evento. Além do ASA, apenas informações de idade, sexo, raça e doença estão disponíveis rotineiramente antes da cirurgia. Esta classificação fornece um indicador global cujos componentes, embora não definidos especificamente, são claramente multivariados. Portanto, se pode concluir que o anestesiolegista, usando um indicador global e subjetivo, pode distinguir cinco classes de “doença” relacionadas à mortalidade cirúrgica geral (51).

Cullen e colegas (59) analisaram três tipos de cirurgias (substituição do quadril, prostatectomia trans-uretral e colecistectomia). ASA e idade foram usados para ajustar diferenças entre *case-mix* avaliando morbidade cirúrgica e utilização de recursos. As morbidades medidas foram a média de permanência, taxa de complicação e visitas de acompanhamento. Resultados: idade avançada e escores ASA podem prever morbidade cirúrgica específica para cada tipo de cirurgia. Não foi encontrada interação entre idade e ASA. A avaliação destes pacientes pode ser feita de forma simples e sem despesas adicionais.

Prause e colegas (60) realizaram um estudo com o objetivo de estudar a correlação entre mortalidade peri-operatória em pacientes submetidos a cirurgia eletiva e as classificações de Estado Físico da ASA e o risco cardíaco multifatorial de Goldman (61;62) e os dois combinados. Foram analisados dados de 16.227 pacientes de um

período de 5 anos, dos quais 215 morreram em 30 dias. Ambos os índices se correlacionaram bem com a mortalidade, mas ASA foi melhor. Conclusão: Para este grande número de pacientes, a mortalidade pode ser predita pela classe de ASA e em grau menor, pelo índice de Goldman. A combinação dos dois pode aumentar a acurácia de predição de mortalidade peri-operatória.

Hall e Hall (63) avaliaram 2570 pacientes submetidos a cirurgia abdominal. Dados foram coletados prospectivamente. Idade >60 anos e ASA >2 identificaram cerca de 80% dos pacientes com estadia prolongada, sepse, admissão à UTI e óbito. Este efeito foi aditivo o que implica em um grau de independência entre as duas variáveis. ASA e idade são indicadores úteis de adversidade clínica e de utilização de recursos após cirurgia abdominal. Além disso, podem ser usados para comparar os desfechos entre diferentes hospitais.

Anestesiologistas freqüentemente discordam sobre a classificação dos pacientes (8;51;52) No estudo de Haynes e Lawlor (64), dos 10 pacientes hipotéticos os anestesistas consultados alcançaram 80% de concordância de opinião em apenas 3. Nove pacientes foram classificados em 3 classes ou mais, 2 pacientes foram classificados em 4 classes diferentes e 1 paciente, nas 5 classes possíveis. Presumivelmente, muitas destas inconsistências entre observadores acontecem devido às diferenças de opinião sobre o significado de um processo de doença descrito. Por exemplo, angina crônica estável significa uma doença sistêmica leve, doença severa não incapacitante ou doença sistêmica de permanente risco de vida (48)?

Owens e colegas (65) reportaram a inconsistência da classificação ASA através da publicação de resultados de um estudo realizado com 255 anestesistas, solicitados a classificar 10 pacientes hipotéticos. Seis pacientes foram classificados de forma consistente e a idade, obesidade, infarto do miocárdio prévio e anemia causaram controvérsias. Profissionais de diferentes regiões e de diferentes ambientes (acadêmicos e não acadêmicos) não apresentaram diferenças significativas em seus desempenhos. A conclusão dos autores foi a de que “a classificação do estado físico da ASA é útil, mas sofre de uma falta de precisão científica.”

Segundo Mak e colegas (53), muitos estudos têm demonstrado inconsistência entre anestesistas no uso da classificação ASA. Como o primeiro destes foi realizado há mais de 20 anos, é interessante determinar se esta diversidade permanece. O estudo foi desenvolvido com anestesistas de Hong-Kong, que têm diferentes backgrounds de treinamento e trabalham em hospitais públicos ou privados. Métodos: 160 questionários distribuídos, mesmos 10 pacientes hipotéticos usados por Owen e colegas em 1978. Os

resultados foram também comparados aos obtidos por Owen. Resultados: 96 questionários retornaram. Os anestesistas de Hong Kong classificaram de forma diferente dos americanos e o grau de treinamento foi influente. Discussão: a meta inicial foi a criação de uma classificação para pacientes a serem submetidos à cirurgia, descrevendo seus estados físicos e não seus riscos pré-operatórios de complicações. O sistema deveria ser simples e de fácil uso e reprodutível. Esta classificação é hoje adotada em todo o mundo. Os resultados deste estudo mostraram que a variação na classificação ASA é ainda importante após todos estes anos de uso. O percentual de concordância de cada caso esteve entre 40 e 60%. As principais discrepâncias foram entre classes 1 ou 2, classes 4 ou 5, classes 2 ou 3 e classes 3 ou 4. O estudo confirmou que a classificação ASA é bastante subjetiva e que um nível significativo de discrepância existe, mesmo entre especialistas bem treinados. A simplicidade da classificação ASA, mesmo com informações limitadas, tem contribuído para sua ampla adoção e se mantém útil para a comunicação entre colegas. Para análise de casos mais detalhada para fins de auditoria, gestão de riscos e alocação de recursos, é urgentemente necessário um sistema mais preciso e específico.

Várias são as razões para explicar esta variabilidade no uso de um sistema tão difundido, bem documentado e universalmente utilizado. A descrição de cada classe é simples mas imprecisa. Diferentes interpretações podem ser dadas a “limitação funcional” “ameaça permanente à vida” e outras. A prática em hospitais públicos ou privados é diferente, pois hospitais privados teriam maior interesse financeiro na classificação de seus pacientes em um estado de doença mais grave. O treinamento local ou em outro país foi fator de diferença, indicando que a cultura interfere (53).

Não há dúvida de que uma classificação mais acurada teria de ser bem mais complexa. O escore APACHE usado por intensivistas é muito complexo para uso rotineiro dos anestesistas e necessita de dados nem sempre disponíveis. Maior precisão poderia ser obtida apenas com a agregação de mais dados sobre as condições pré-cirúrgicas do paciente(53).

Para se obter a precisão científica reclamada por estes autores, um esforço deve ser feito no sentido de refinar as definições das diferentes classes. Vários estudos recentes sugerem outras variáveis como preditoras de risco para mortalidade cirúrgica e anestésica. É concebível que com esforço suficiente, um índice de risco multifatorial possa ser desenvolvido para cada tipo de cirurgia ou para cada diagnóstico, adicionando fatores tais como a performance prévia do cirurgião, do anestesista e do hospital, que permita prever diferentes desfechos como mortalidade, tempo de permanência e

sobrevida. Segundo Pratt, “esta projeção, com a tecnologia ora disponível, não é mais fantasiosa, mas estes índices, embora mais precisos, estarão longe do subjetivo, intuitivo, vago e indefinido índice global do *Physical Status ASA Classification*, o qual, com sua bela simplicidade e facilidade, nos fornece um preditor útil de mortalidade cirúrgica geral” (50).

5.2- Albumina sérica

A albumina, a maior proteína produzida pelo fígado, é a proteína mais abundante no plasma, representando aproximadamente 60% das proteínas totais. As principais funções da albumina são a manutenção da pressão osmótica do plasma e transportar uma variedade de moléculas incluindo ácidos graxos, bilirrubina, metais, hormônios e drogas (66). A albumina é sintetizada pelo fígado, razão pela qual doenças hepáticas podem levar a hipoalbuminemia, que pode também ser causada por insuficiência renal, que permite que a albumina escape pela urina. Albumina diminuída pode ser explicada também por desnutrição ou dieta pobre em proteína (67).

A concentração de albumina é comumente utilizada como indicador de estado nutricional e o grau de hipoalbuminemia tem indicado o grau de desnutrição de pacientes. O nível de albumina sérica tem também sido correlacionado com desfechos clínicos, sendo a hipoalbuminemia associada a mortalidade e morbidade aumentadas em pacientes hospitalizados. A literatura aponta como valores normais aqueles entre 3,5 e 5,0g/dL, medida pelo método de bromocresol (68).

Desnutrição é o fator número 1 para mortalidade, contribuindo para 12% de todas as mortes, duas vezes a mortalidade por tabaco ou hipertensão. Nos EUA, 50% dos pacientes são desnutridos moderadamente no momento da admissão ao hospital. Vários estudos demonstraram que o estado nutricional se deteriora durante uma hospitalização. Isso é especialmente verdadeiro para pacientes submetidos a cirurgias, nos quais a desnutrição é particularmente deletéria (69).

É importante entender os fatores fisiológicos que determinam a concentração de albumina no plasma e sua relação com o estado nutricional de pacientes hospitalizados. A concentração de albumina representa a soma de uma rede de muitos eventos – a síntese da albumina no fígado, sua degradação, perdas pelo organismo, trocas entre albumina intra e extra-vascular e o volume no qual a albumina está diluída. A albumina é altamente hidrossolúvel e reside no espaço extra-celular. A quantidade normal de albumina em um homem de 70 Kg é de aproximadamente 350g (5g/Kg). Aproximadamente 40% desta (140g) constituem a fração intravascular. Grande parte da porção extra-vascular da albumina está localizada na pele. A concentração da albumina

no sangue é maior do que na linfa ou em outros fluídos extra-celulares e a razão entre a concentração intra e extra-vascular varia de tecido a tecido. Após 30 minutos do início de sua síntese, o hepatócito secreta a proteína na corrente sangüínea, onde tem uma meia-vida de 20 dias. Em condições estáveis, cerca de 14g (200mg/Kg) de albumina são produzidas e degradadas a cada dia. Portanto, aproximadamente 4% da albumina total é degradada e substituída a cada dia (66).

A desnutrição protéico-calórica causa a diminuição da síntese da albumina. A ingestão adequada de nutrientes é importante para a agregação polissomal e manutenção dos níveis celulares de RNA necessários para a síntese protéica. Após 24 horas de jejum, a taxa de síntese de albumina diminui marcadamente mas é rapidamente normalizada com a realimentação. Em situações agudas, a redução de sua síntese tem pouco impacto nos níveis séricos porque o seu *turnover* é lento e sua quantidade total é alta. Na verdade, os níveis séricos de albumina podem até mesmo aumentar em curtos períodos de jejum. Mesmo durante desnutrição crônica, a concentração plasmática da albumina é mantida pela diminuição compensatória de sua degradação e pela transferência de albumina extra-vascular para o compartimento intravascular. Restrições protéicas prolongadas induzidas em voluntários experimentalmente e observadas clinicamente em pacientes com anorexia nervosa causam marcada redução no peso corporal mas pouca mudança na concentração de albumina sérica. No entanto, uma dieta deficiente em proteína mas com calorias adequadas provocam uma redução importante nos seus níveis séricos.

Pacientes hospitalizados podem ter baixos níveis de albumina sérica devido à resposta metabólica ao “*stress* inflamatório”. Processos inflamatórios podem reduzir a síntese protéica, causar aumento em sua degradação e de sua perda transcapilar e, em certas circunstâncias, um aumento nas perdas através das fezes, urina ou tecidos danificados. Durante doenças graves, a permeabilidade vascular aumenta dramaticamente, podendo a perda de albumina do plasma para o espaço extra-vascular dobrar em casos de caquexia por câncer ou ser até 3 vezes maior em pacientes com choque séptico.

Embora a nutrição inadequada possa contribuir para a hipoalbuminemia em pacientes hospitalizados, a resposta metabólica ao *stress* é um fator muito mais importante. O entendimento da fisiologia da albumina explica porque as concentrações de albumina sérica se correlacionam com a severidade do quadro dos pacientes hospitalizados mas são inapropriadas como uma medida do estado nutricional.

Os valores de albumina sérica variam muito porque esta proteína é muito sensível a mudanças no estado de hidratação do organismo. Valores normais: estes valores podem

variar em função da técnica usada e das características nutricionais de diferentes populações. Em geral, se aceita como valores normais aqueles entre 3,4 e 5,4g/dL. Valores abaixo do normal podem indicar ascite, queimaduras extensas, glomerulonefrite, hepatite, cirrose, necrose hepatocelular, síndromes de má-absorção (doença de Crohn ou de Whipple), desnutrição e síndrome nefrótica. Este teste não é acurado se o paciente está recebendo grande quantidade de fluídos intra-venosos, pois isso causará a diluição da albumina. Da mesma forma, pacientes desidratados podem apresentar altas concentrações de albumina. A albumina também diminui na gestação (67).

Cunha (70) classifica os níveis de albumina sérica em 4 grupos:

- <2,1g/dL: hipoalbuminemia grave
- 2,1-2,7g/dL: hipoalbuminemia moderada
- 2,8-3,5g/dL: hipoalbuminemia discreta
- >3,5g/dL: normal

Em cirurgia, a relação entre os níveis séricos de albumina e desfechos adversos são reconhecidos há muitos anos (71). Jones e Eaton, em 1933 (72), já descreveram seus achados neste sentido, em uma série de 26 pacientes cirúrgicos. Mais recentemente, albumina foi descrita como fator preditivo de mortalidade e morbidade pós-cirúrgicas para pacientes submetidos a cirurgia eletiva (73) e morbidade aumentada naqueles submetidos a cirurgia do trato intestinal (74). Gibbs e colegas, analisando dados de 54.215 cirurgias não cardíacas realizadas em 44 hospitais participantes do “*National Veterans Affairs Surgical Risk Study*”, concluíram que a concentração de albumina sérica é um preditor de desfecho cirúrgico melhor do que muitas outras características pré-operatórias dos pacientes (71).

Estudo publicado em 2002, baseado em avaliação de fatores de risco para mortalidade cirúrgica em cirurgia geral (2989 pacientes) concluiu que a albumina sérica e frações do colesterol, rotineiramente medidos em pacientes hospitalizados, são preditores de óbito hospitalar, infecção nosocomial e longa permanência (75). Estudo retrospectivo realizado com consulta a 706 prontuários de 4 hospitais universitários americanos com o objetivo de identificar um indicador pré-operatório simples e clinicamente útil de recuperação pós-operatória de cirurgias de grande porte, concluiu que os valores de albumina sérica, presentes em 74,5% dos prontuários analisados, foi um fator altamente preditivo de mortalidade, cujas taxas variaram de 1-5% para pacientes com albumina maior ou igual a 3,25g/dL até 20-30% para os pacientes com albumina menor ou igual a 1,75g/dL (76).

Vincent e colegas (77) realizaram uma meta-análise para determinar se a hipoalbuminemia é um fator de risco independente para desfechos adversos em pacientes hospitalizados com quadros agudos. Foram analisados 90 estudos de coorte com um total de 291.433 pacientes, concluindo que a hipoalbuminemia é um preditor potente, dose-dependente e independente de maus desfechos. Cada decréscimo de 1g/dL aumenta significativamente a chance de óbito em 137%, morbidade em 89%, longa permanência em 71% e aumento da utilização de recursos em 66%.

Embora aos níveis séricos de albumina possam ser afetados por fatores agudos como trauma e estresse cirúrgico, a concentração de albumina no sangue é um preditor de desfechos cirúrgicos melhor do que muitas outras características pré-operatórias dos pacientes porque é um marcador de doença e de desnutrição bem como possivelmente confira um efeito protetor através de vários mecanismos biológicos (78). É relativamente um teste de baixo custo, que poderia ser utilizado mais freqüentemente como uma ferramenta prognostica para detectar desnutrição e o risco de desfechos adversos, particularmente em populações com alta prevalência de comorbidades.

A albuminemia deveria ser utilizada com maior freqüência como ferramenta para detectar desnutrição e riscos para eventos adversos, particularmente em populações onde condições de comorbidade são relativamente freqüentes por ser de baixo custo e sem riscos para o paciente. Segundo Vincent e colegas, os achados de sua revisão sistemática sugerem que o teste de albumina sérica pré-cirúrgica tem sido sub-utilizado pelos cirurgiões. O custo do teste (US\$2-US\$4) é baixo em relação ao seu valor prognóstico, além de ser um elemento chave na avaliação nutricional acurada e tem sido demonstrado que a terapia nutricional é mais efetiva na redução de morbidade e mortalidade cirúrgicas quando o diagnóstico de desnutrição é baseado em índices nutricionais objetivos que incluam a albumina sérica, mais do que em avaliações subjetivas. Particularmente em populações com alto nível de comorbidades, verifica-se que esse teste deveria ser usado mais freqüentemente como uma ferramenta prognostica e para detectar desnutrição (71).

5.3- Idade

A questão da idade como fator de risco para cirurgia é complexa. Existem vários fatores de risco que têm impacto nas taxas de mortalidade cirúrgica em idosos e os mais importantes destes são as mudanças fisiológicas que ocorrem com o envelhecimento, as condições da doença que originou a necessidade da cirurgia, o tipo de procedimento realizado e o caráter de emergência da cirurgia. Embora existam muitos riscos na realização de cirurgias em idosos, vários destes se beneficiarão da cirurgia e a idade, de

forma isolada, não deve nunca ser usada como critério de decisão de realização de um procedimento cirúrgico.(79).

Finlayson e Birkmeyer (80) analisaram 1.2 milhões de registros de pacientes maiores de 65 anos do *Medicare*, hospitalizados entre 1994 e 1999, para cirurgias eletivas de grande porte, com o objetivo de definir os fatores de risco associados à mortalidade cirúrgica em até 30 dias. Os resultados encontrados indicam que a mortalidade varia de acordo com o procedimento realizado e com a idade dos pacientes. A mortalidade de pacientes com 80 anos e mais é maior que o dobro daqueles com 65-69 anos. Esta avaliação é limitada porque a estimativa de risco cirúrgico deve levar em conta características dos hospitais. Enquanto características do hospital, do cirurgião e do paciente são importantes para estimar o risco de um paciente individual, a mortalidade cirúrgica de uma população pode ser utilizada como ponto de partida para ajudar os pacientes a entender os riscos associados a uma cirurgia, ao decidir por fazer ou não um procedimento eletivo.

Um estudo realizado na Espanha com 710 pacientes com 70 anos ou mais, para identificar os fatores de risco associados a mortalidade após cirurgia abdominal de emergência, comparou os desfechos dos pacientes entre 70 e 79 anos com aqueles com 80 e mais anos. Foram consideradas na análise as variáveis sexo, idade, ASA, tempo entre início dos sintomas e internação e entre a internação e a cirurgia, diagnóstico, tipo de cirurgia, achados operatórios. Os desfechos estudados foram a mortalidade, morbidade e tempo de permanência. Análises de regressão logística mostraram que o ASA, demora do tratamento cirúrgico e tipo de cirurgia (aquelas que permitiam apenas tratamento paliativo) eram os fatores de risco associados a mortalidade, sendo a taxa de mortalidade cirúrgica maior no grupo mais idoso (24% contra 19%). Mais de 50% dos pacientes idosos apresentam doenças associadas, o que explica o maior risco pré-operatório e, de acordo com os achados, os autores recomendam que a decisão de operar ou não um paciente com quadro abdominal agudo de emergência não seja tomada em função de sua idade (81).

Vários estudos publicados sobre escores de risco para mortalidade cirúrgica consideram a idade como uma variável associada de forma independente ao desfecho (21;22;44;46;47;82;83).

5.4- Caráter da cirurgia

De acordo com os critérios do *National Confidential Enquiry into Perioperative Deaths* (NCEPOD), organização britânica que tem por objetivo revisar paráticas cirúrgicas e fazer recomendações para melhorar a qualidade da assistência (84), o caráter de uma intervenção cirúrgica pode ser classificado em 4 grupos:

- 1- Eletiva: cirurgia agendada de acordo com conveniência do paciente e do médico. Ex: cirurgia de hérnia
- 2- Agendada: Cirurgia agendada para as próximas semanas. Ex: cirurgia de câncer de cólon
- 3- Urgente: cirurgia necessária em um lapso de tempo de 24-48 horas. Ex: cirurgia para correção de obstrução de cólon
- 4- Emergência: cirurgia imediata. Ex: correção de ruptura de aneurisma de aorta

O caráter de prioridade de uma cirurgia é associado a desfechos adversos em muitos estudos, pois esta classificação está frequentemente associada à gravidade do quadro clínico e/ou à falta de condições ideais para a realização de um procedimento tanto em termos de estrutura disponível, da equipe e do paciente. A maior parte destes estudos tratam esta variável como dicotômica (eletivas + agendadas e urgentes + emergências) (21;22;44;45;83;85;86), enquanto outros a tratam como variável categórica de 3 (47) ou 4 classes (42). Todos estes estudos, no entanto, desenvolveram modelos de risco nos quais o caráter da cirurgia era um dos preditores significativos.

Para exemplificar, de acordo com dados do sistema de Indicadores Gerenciais do Hospital de Clínicas de Porto Alegre, no período de janeiro de 2002 a outubro de 2004, a taxa de mortalidade cirúrgica geral foi de 4,02% (923/22.986), sendo de 2,56% para cirurgias eletivas (483/18.840) e de 10,61% (440/4.146) para cirurgias de urgência/emergência.

5.5- Outros fatores de risco associados a mortalidade cirúrgica

Outros fatores de risco são associados por diferentes autores a desfechos de mortalidade cirúrgica, sendo a maior parte destas variáveis específicas para um determinado tipo de procedimento e outras aparecem apenas em 1 estudo. O maior destes estudos, realizado pelo Departamento dos Veteranos dos Estados Unidos, analisando dados de mais de 87.000 cirurgias, determinou como variáveis preditoras de óbito após cirurgias não cardíacas, os seguintes fatores de risco: albumina sérica, classificação ASA, câncer disseminado, caráter da cirurgia, ascite e idade (22). O diagnóstico de câncer disseminado não é feito antes da cirurgia na maior parte das vezes, o que impede o seu uso em modelos que buscam avaliar o risco do paciente antes

da realização de uma cirurgia. A presença ou não de ascite, além de ser um dado pouco disponível em bases de dados hospitalares, é associada clinicamente aos níveis séricos de albumina, o que a torna uma variável confundidora. Os demais fatores de risco foram discutidos em capítulos anteriores.

Outros dois fatores aparecem na literatura como possíveis determinantes de mortalidade cirúrgica, embora não tenham sido encontrados em modelos e escores de risco publicados: o volume de cirurgias realizado pela equipe ou pelo hospital e a experiência do cirurgião.

Volume cirúrgico:

Muitos estudos publicados nas três últimas décadas descrevem altas taxas de mortalidade em hospitais com baixo volume cirúrgico em procedimentos selecionados. Algumas revisões recentes afirmam que milhares de mortes cirúrgicas evitáveis ocorrem a cada ano nos EUA porque cirurgias eletivas de alto risco são realizadas em hospitais com experiência inadequada nestes procedimentos. Apesar do recente interesse na variável “volume cirúrgico”, muitos questionam a aplicabilidade destes resultados na prática atual. Primeiro, porque as taxas de mortalidade cirúrgica têm diminuído nos últimos anos graças a avanços de técnicas operatórias, diagnósticas e de prevenção antibiótica. A maior parte dos estudos utilizou dados menos recentes, o que prejudica a comparação. Além disso, muitos estudos se restringiram a regiões geográficas limitadas que possuem poucos centros de alto volume cirúrgico (87).

Birkmeyer e colegas (88) analisaram a relação entre o volume de cirurgias realizadas por hospitais e sua relação com as taxas de mortalidade nos EUA. Este estudo analisou dados do *Medicare* de 1994 a 1999, buscando encontrar a relação entre o volume de cirurgias realizadas e as taxas de mortalidade de cada hospital. Foram analisados 2.5 milhões de procedimentos em 14 tipos de cirurgias. A mortalidade nas 14 cirurgias diminuiu a medida que o volume de cirurgias realizadas pelo hospital aumentou mas a importância relativa deste fator foi diferente para os diferentes procedimentos. A conclusão dos autores foi a de que, na ausência de outras informações sobre a qualidade cirúrgica dos hospitais próximos, os pacientes do sistema *Medicare* que necessitam ser submetidos a cirurgias cardiovasculares ou para remoção de algumas neoplasias podem diminuir seus riscos de forma significativa selecionando um hospital de grande produção cirúrgica.

Experiência do cirurgião

Muitos trabalhos publicados têm demonstrado que hospitais ou médicos com alto volume de produção alcançam melhores resultados para pacientes submetidos a

procedimentos cirúrgicos ou tratamentos clínicos. Em estudos realizados para procedimentos específicos de alto risco, foi determinado que a experiência do cirurgião é um importante determinante do desfecho. A partir desses resultados, é razoável a preocupação sobre o impacto potencial da relativa inexperiência de residentes em hospitais de ensino no início de seu período de treinamento. A cada mês de julho, os hospitais de ensino americanos recebem um grande fluxo de novos médicos recém formados. A educação médica é um componente fundamental da missão de hospitais de ensino e nesses centros, internos, residentes e estagiários respondem por grande parte da assistência. A dinâmica da assistência por médicos menos experientes tem levado a preocupações sobre a qualidade dessa assistência no início de cada ano letivo, conhecido coloquialmente como “*July Phenomenon*”. Enquanto alguns estudos demonstram que os custos da assistência em hospitais de ensino aumentam neste período, outros não relatam nenhum impacto na qualidade da assistência. No entanto, muitos destes estudos são baseados em dados administrativos o que dificulta o ajuste dos dados para potenciais diferenças de categorias de casos e severidade das doenças.

Barry e Rosenthal (89) discutem no artigo “*Is there a July Phenomenon?*” a tese de que a inexperiência dos novos residentes admitidos a hospitais universitários pode piorar os desfechos dos pacientes nos primeiros meses de seus períodos. Analisando uma coorte retrospectiva de 156.136 pacientes consecutivos admitidos em CTIs de 28 hospitais de Ohio, EUA, de 1991 a 1997, os autores não encontraram diferenças entre as chances de óbito no primeiro trimestre em relação aos demais meses em hospitais de grande ou pequeno porte e hospitais de ensino e outros, depois de ajustar os riscos pela classificação APACHE. Uma das limitações citadas é a variação sazonal da mortalidade, que pode mascarar a variação negativa da inexperiência. Como as taxas de mortalidade aumentam nos meses de inverno, esta variável pode dar uma impressão de risco constante durante todo o ano. Não encontrando evidências da existência do “*July Phenomenon*”, os autores recomendam a realização de estudos futuros que examinem fatores organizacionais que permitam aos hospitais com residência médica compensar a inexperiência dos novos médicos.

6- Questões metodológicas

Planejar uma estratégia de ajuste de risco é uma tarefa complexa. No mínimo, requer a definição de fatores de risco – características clínicas e outras que reflitam a probabilidade do paciente vir a sofrer o desfecho de interesse. Uma vez delineados estes fatores de risco devem ser operacionalizados, isto é, definidos de modo a permitir identificação e medidas confiáveis e sem viéses; devem ser combinados de forma

empírica ou normativa para produzir escores de risco ou classificação para cada paciente. A metodologia necessária para estas tarefas é complexa e algumas dimensões-chave merecem ser analisadas (90).

6.1- Fontes de dados para modelos de ajuste de risco

A consideração prática mais importante ao se planejar uma metodologia para ajuste de risco é a exigência dos dados. Em geral, os pesquisadores utilizam dados obtidos de bases de dados eletrônicas geradas pelos hospitais para fins de cobrança ou exigem variáveis clínicas adicionais buscadas em prontuários. As pesquisas em bases de dados administrativas são mais baratas e mais simples mas podem não oferecer a credibilidade clínica de um sistema derivado de dados do prontuário do paciente.

No entanto, o custo para obtenção de bases de dados clínicas mais acuradas continua a ser um impedimento. Foi estimado um custo de US\$17 por caso registrado, o que significava um custo por hospital variando de US\$70.000 para pequenos hospitais rurais até US\$135.000 para grandes hospitais urbanos (91). Muitos estados e instituições americanos consideraram que os ganhos adicionais com estas bases de dados não eram justificados pelos custos e mantiveram seus modelos de ajuste baseados em bases de dados administrativas. No entanto, as preocupações com a confiabilidade e acurácia destas medidas permaneceram.

Bases de dados administrativas

Estas bases são geralmente criadas por provedores a partir de dados da internação do paciente (hospitais ou médicos) para fins de cobrança, tais como a base de dados do *Medicare* americano e do Sistema Único de Saúde brasileiro (92). Essas fontes contêm informação clínica limitada, incluindo idade do paciente, sexo, tipo de saída (alta ou óbito), tipo de admissão (eletiva/urgente), diagnósticos e procedimentos. Visto que estes registros são geralmente derivados de processos de cobrança ou por exigência governamental, são vistos com suspeita por aqueles preocupados com potenciais motivações financeiras por trás dos dados registrados (93). Além dos problemas inerentes à codificação de diagnósticos e procedimentos, existem dúvidas entre os pesquisadores sobre a completude, acurácia e confiabilidade dos dados. Outro problema com o uso de informações oriundas de sumários de alta para o ajuste de risco é o seu caráter retrospectivo: as informações desse tipo de sumário refletem condições diagnosticadas ou tratadas durante toda a admissão, sem especificar quando cada evento ocorreu ou qual a seqüência correta de eventos. Informação sobre o momento do diagnóstico é crucial para a determinação de fatores de risco. A incapacidade de diferenciar o *timing* de complicações potenciais de condições existentes na admissão

usando dados administrativos causa uma séria limitação na sua utilidade para identificação de fatores de risco prévios a uma intervenção. Apesar de suas limitações, dados de sumários de alta administrativos possuem a significativa virtude de serem disponíveis prontamente e serem muito abrangentes em termos de população representada. Levando em conta suas limitações, estas bases de dados podem ser fontes para importantes pesquisas sobre ajuste de risco.

Dados de prontuários de pacientes

Dados primários apresentam suas próprias oportunidades e problemas. Através de revisões de prontuários podem ser coletadas medidas clínicas detalhadas na seqüência em que ocorreram que podem melhorar a significância médica da abordagem de ajuste de risco. No entanto, achados clínicos de exames físicos, testes diagnósticos e mesmo estudos mais sofisticados tecnicamente podem ser vagos e sofre de grandes variações entre observadores. Outra preocupação se refere à completude, acurácia e validade dos dados registrados em prontuário. Os prontuários são cada vez mais utilizados como instrumentos de contenção de custos e propósitos administrativos e legais, o que pode afetar a documentação de registros mais sensíveis do ponto de vista ético. Por outro lado, a linguagem médica pode ser adaptada às expectativas dos possíveis usuários dos prontuários. Portanto, o prontuário pode não refletir um registro completo e objetivo do caso clínico do paciente.

A confiabilidade dos dados primários (prontuário) deriva de uma variedade de fatores, incluindo a clareza e definição dos dados necessários, as regras de interpretação da terminologia e jargões médicos, legibilidade, completude, temporalidade e organização dos prontuários.

No 89º Congresso Clínico do Colégio Americano de Cirurgiões realizado em Chicago de 19 a 23 de outubro de 2003, uma mesa-redonda coordenada pelo Dr. Shukri F. Khuri discutiu a questão da qualidade dos dados publicados na literatura médica (91). Um dos problemas discutidos pelos participantes foi a fonte dos dados utilizados nas pesquisas. Segundo esses, pesquisas que utilizam grandes bases de dados administrativos tais como aquelas formadas por sumários de alta com objetivo de cobrança de sistemas pagadores como o *Medicare*, *Medicaid* e *Veteran's Administration* ou de pagadores privados, apresentam limitações significantes pois refletem os serviços prestados a indivíduos com seguro-saúde, por definição, excluem os pacientes não-segurados que são tipicamente a população mais vulnerável, excluem os serviços que são pagos diretamente pelo paciente e são fortemente influenciados por sistemas de codificação que

são freqüentemente inconsistentes e inaccurados, além de terem um viés determinado pelo interesse de ressarcimento financeiro.

Por outro lado, essas bases de dados possuem qualidades únicas para alguns tipos de pesquisa pois fornecem dados de uma comunidade de pessoas, tais como todos os pacientes hospitalizados em um determinado estado num determinado ano; permitem que pesquisadores rastreiem um paciente determinado (Sistema do VA) através de diferentes hospitais; mascaram informações pessoais para proteger a confiabilidade; já existem, estão disponíveis e são relativamente de uso barato e fácil. Mesmo assim, a falta de especificidade clínica para sistemas de diagnósticos comumente utilizados como a CID (Classificação Internacional de Doenças) limitam a utilidade desses dados para algumas questões de pesquisa. Segundo os painelistas, esses estudos que utilizam um número muito grande de sujeitos continuarão a despertar o interesse de diferentes segmentos, mesmo sendo medíocres do ponto de vista metodológico.(90;94;95)

Romano e colegas (94) analisaram dados administrativos (faturas) comparando-os com dados clínicos (prontuários) de pacientes submetidos a cirurgia cardíaca. Os autores utilizaram dados sobre pacientes de 4 hospitais universitários, 2 com bases de dados clínicas e 2 com bases de dados administrativas. Foram comparadas as prevalências de 14 comorbidades e as taxas de risco associadas a mortalidade. Resultados: condições crônicas ou assintomáticas (infarto prévio, tabaco, hiperlipidemia, insuficiência mitral e outras) foram menos prevalentes nas bases de dados administrativos. A prevalência de informações sobre diabetes, angina e insuficiência cardíaca congestiva foram similares nas duas bases de dados. Estimativas de risco relativo derivadas de dados clínicos igualaram ou ultrapassaram aquelas derivadas de bases de dados administrativos. Os mesmos autores realizaram em 2002, um estudo para verificar se seria possível avaliar complicações pós-operatórias através de bases de dados administrativas (96). Foram selecionados aleatoriamente 991 adultos submetidos a cirurgia eletiva em 30 hospitais da Califórnia, dos quais se comparou as complicações presentes na base de dados com as complicações relatadas por médicos através da análise dos prontuários, todas codificadas pela CID-9. Os médicos identificaram maior número de complicações leves e severas o que, segundo os autores, provoca questionamentos sobre a validade do uso desse tipo de base de dados para avaliar complicações cirúrgicas, mesmo quando codificadas por pessoal altamente treinado.

Best e colegas (97), pesquisadores do “*National Surgical Quality Improvement Program*” (NSQIP) do Departamento de Assuntos dos Veteranos (DVA), realizaram uma comparação entre os dados coletados prospectivamente por enfermeiras treinadas com

as informações administrativo-financeiras disponíveis nos 123 hospitais do sistema, para responder a críticas de que a forma de coleta empregada era dispendiosa e sem custo-efetividade. Foram avaliadas as informações existentes nas duas fontes sobre fatores de risco pré-operatórios e desfechos até 30 dias após a cirurgia, testando seus valores preditivos positivos e sensibilidade. Na CID-9 foram identificados os códigos para 61 características pré-operatórias e 21 eventos adversos. Analisando a presença os não destas informações, os autores concluíram que a sensibilidade e o poder preditivo positivo de bases de dados administrativos são mais pobres do que as bases de dados clínicos, o que impede a recomendação de substituição da fonte de dados.

Pine, Jones e Lou (98) realizaram um estudo de coorte retrospectiva em 22 hospitais de St Louis, Missouri, EUA, com 5.966 pacientes internados em 1995 com diagnóstico de infarto agudo do miocárdio, insuficiência cardíaca congestiva ou pneumonia, para comparar a precisão do ajuste de risco nas medidas de taxas de mortalidade utilizando dados administrativos de sumários de alta hospitalares, esses dados mais dados laboratoriais e dados administrativos mais dados laboratoriais mais dados clínicos. A precisão do modelo de ajuste de risco foi avaliada através da área sob a curva ROC (*Receiver Operating Characteristic Curve*). Os modelos administrativos obtiveram um valor de 0,834 contra 0,875 dos modelos clínicos. Adicionando dados laboratoriais aos dados administrativos, a área sob a Curva ROC passou para 0,86 e a correlação média com os dados obtidos a partir do modelo clínico aumentou de $r=0,86$ para $r=0,95$. Esses resultados indicam que prognósticos de mortalidade hospitalar melhoram consideravelmente com a inclusão de dados laboratoriais (muitas vezes disponíveis eletronicamente) combinados com os dados administrativos disponíveis.

Martins e colegas (99) analisaram o uso da base de dados do Sistema de Informações Hospitalares do Sistema Único de Saúde (SIH/SUS) para construção de modelos de ajuste de risco de indicadores de desempenho, concluindo que faltava nesta base a qualidade da informação diagnóstica necessária, especialmente informações sobre comorbidades.

O poder preditivo de um modelo de ajuste de risco depende da extensão e da acurácia dos dados disponíveis sobre a situação inicial de cada paciente. Tradicionalmente estas informações são obtidas eletronicamente a partir das contas hospitalares dos pacientes (bases de dados administrativas) ou precisam ser trabalhosamente coletadas de prontuários clínicos (dados clínicos). Embora bases de dados administrativos sejam disponíveis em muitos hospitais e seu uso seja de baixo

custo, são criticadas pela falta de detalhamento clínico necessário para permitir um ajuste de risco adequado. Vários estudos já publicados reforçam a convicção de que estas fontes não apresentam os requisitos necessários exigidos para avaliações de risco ou comparações adequadas (98;100). O sistema ideal é aquele que alia a acurácia, confiabilidade e credibilidade dos dados clínicos com o fácil acesso eletrônico, o que é ainda oferecido por poucos hospitais no mundo que possuem um prontuário eletrônico que apresente qualidade de informação aliada a um volume adequado para pesquisa.

6.2 – Imputação de dados

Observações faltantes (*missing data*) são freqüentemente encontradas em pesquisas com populações humanas. Através dos anos, estatísticos desenvolveram um corpo de teoria e métodos para lidar com os mesmos. As abordagens possíveis são a exclusão do caso no momento da análise ou o preenchimento de cada dado com um valor imputado, analisando o conjunto de dados como se fosse completo. Imputação de dados se refere à substituição de dados faltantes por um valor que permita que a análise seja conduzida sem que seja enganosa (101). Alguns conceitos são fundamentais neste processo:

- 1- A imputação não deve introduzir vieses ou alterações na distribuição dos dados ou uma variação significativa dos estimadores.
- 2- O processo de imputação deve basear-se em dados da amostra em vez de utilizar fundamentos externos sobre a natureza dos dados faltantes.
- 3- A imputação não deve levar a situações onde estimativas relevantes sejam baseadas muito pesadamente nos dados imputados.

O tratamento da não-resposta requer um entendimento das características dos não-respondentes e o provável impacto que esses não-respondentes poderiam ter na pesquisa. O consenso é o de que nenhum prejuízo pode ser causado através da imputação de dados para pequenas quantidades de dados faltantes e dessa forma o pesquisador pode produzir um conjunto de dados mais útil. No entanto, havendo uma quantidade significativa de não-resposta (>10%), existe um viés potencial por trás, o que pode ser corrigido por estratégias de imputação adequadas. De qualquer forma, a recomendação é a de se manter no banco de dados duas variáveis distintas, uma com os valores originais e a outra com os valores imputados adicionados.

Os procedimentos de imputação devem ser decididos na etapa de planejamento da pesquisa, para que se minimize a quantidade de dados faltantes e se estabeleça o processo de substituição, obtendo variáveis que possam auxiliar nas estimativas adequadas. Embora a imputação possa melhorar a qualidade dos dados finais, deve-se ter cuidado para escolher a metodologia mais adequada, pois alguns métodos não

preservam a relação entre as variáveis e podem até distorcer essas relações. Existem diferentes abordagens a serem consideradas:

- 1- Substituição: baseia-se na disponibilidade de dados comparáveis. Os dados imputados podem ser extraídos dos registros dos respondentes ou de outras fontes alternativas de dados.
- 2- Estimador: usa informações de outras questões ou de outras respostas e através de operações matemáticas, deriva um valor plausível para o dado faltante. O método estimador mais simples é o da substituição dos valores faltantes pelo valor médio daquela variável. Quando existem valores extremos (*outliers*) na amostra utiliza-se o valor da mediana da variável. Pode-se utilizar a média ou mediana de um sub-grupo de casos com características similares.
- 3- “*Cold deck*”: faz uso de um conjunto fixo de valores, que cobrem todos os itens de dados e que podem ser construídos com o uso de dados históricos, conhecimento especializado etc. Um questionário “perfeito” é criado para responder aos requisitos de imputação parcial ou total.
- 4- “*Hot deck*”: utiliza outros registros como “doadores”, que podem ser selecionados randomicamente de um grupo de casos com o mesmo conjunto de características pré-definidas. O doador também pode ser selecionado através de um método chamado “imputação do vizinho mais próximo”, onde algum critério de classificação é desenvolvido para determinar o registro mais semelhante àquele com o dado faltante e esse registro se torna doador desses dados.
- 5- Regressão múltipla: pode ser usada para imputação de dados simplesmente utilizando uma ou mais variáveis existentes para predizer os valores faltantes de outra variável altamente correlacionada com as anteriores. Este procedimento pode levar a uma adequação estatística extrema dos dados (*over-correct*) introduzindo um baixo nível de ruído nos dados. O SPSS, por exemplo, para corrigir isso, por *default* adiciona o resíduo de um caso escolhido ao acaso a cada valor estimado.
- 6- Estimativa de Máxima Verossimilhança (*Maximum likelihood estimation* (MLE)): Esse método faz referência ao algoritmo EM e é atualmente o método mais comum de imputação, estando disponível no SPSS v.10.
- 7- Imputação Múltipla (IM): é um método que gera valores simulados múltiplos para cada dado incompleto, através do método Monte Carlo. Tem a vantagem de ser simples, sendo aplicável para grandes bases de dados. A eficiência do método é maior quanto menor for a percentagem de dados faltantes a serem imputados.

8- *Aproximate Bayesian Bootstrap* (ABB): utiliza a razão de verossimilhança para estimar a probabilidade de resposta/não-resposta na variável dependente Y , baseado em covariáveis X_i . Os casos são agrupados baseados nessas estimativas de r/nr e dentro de cada categoria são criados os valores para as variáveis faltantes, baseados em amostragens dos dados existentes do mesmo grupo. Este método é apropriado quando o objetivo da pesquisa é analisar Y mas não quando o objetivo seja o de analisar as relações entre as variáveis dependente e independentes, como é freqüentemente o caso em regressão ou modelagem.

Esses métodos podem ser feitos manualmente, automaticamente ou através de combinações. Feito de forma apropriada, a imputação limita os vieses causados pelas não-respostas e apresenta uma forma de auditoria para fins de avaliação e assegura que os registros imputados tenham consistência interna. Um bom procedimento de imputação é automático, objetivo e eficiente.

Como a ausência de dados é praticamente inevitável, na etapa de delineamento do estudo deve-se planejar a investigação de potenciais preditores dos valores faltantes e utilizá-los nas análises. É importante que os dados faltantes não sejam considerados apenas como um problema de análise de dados mas também como uma questão de planejamento da pesquisa e da análise dos resultados.

De qualquer forma, não existe uma decisão simples para a escolha entre imputar dados ou excluir sujeitos de uma pesquisa por falta de dados. O pesquisador deve analisar suas opções e, na fase de análise, certificar-se de que a imputação ou o abandono dos casos incompletos não agregou vieses relevantes aos resultados obtidos. Vários trabalhos publicados que relatam estudos observacionais de populações diversas descrevem a realização de imputação de dados:

Edwards e colegas (86), em pesquisa realizada com dados de mais de 90.000 pacientes para desenvolver modelos estatísticos de risco para mortalidade cirúrgica após cirurgia de substituição de válvulas cardíacas, relatam que para todas as variáveis contínuas, os dados faltantes foram substituídos pela mediana e para as variáveis categóricas, o valor imputado foi aquele mais prevalente na população (a moda).

Khuri e colegas(22), no trabalho que relata os resultados de pesquisa desenvolvida com dados de mais de 87.000 pacientes para o desenvolvimento de um modelo de ajuste de risco para avaliação da qualidade dos serviços cirúrgicos dos hospitais ligados ao *National Veterans Affairs* através das taxas de mortalidade cirúrgica, citam que realizaram um processo de imputação de dados laboratoriais faltantes, pois a existência desses na base de dados era dependente da solicitação dos mesmos pela equipe médica. Um

procedimento de regressão foi utilizado para estimar os valores faltantes. Como o valor da albumina sérica foi identificado como o principal preditor de mortalidade cirúrgica nesse estudo, contribuindo com 0,793 ao índice-C do modelo e este valor estava disponível em apenas 61% dos casos, sendo imputado para os restantes 39%, análises adicionais foram realizadas para confirmar seu valor preditivo. Utilizando apenas os casos de cirurgia geral, os autores desenvolveram outros dois modelos, um incluindo uma variável que indicava a existência ou não da medida de albumina em adição às demais variáveis, que não se mostrou significativa no modelo, e outro restrito aos casos onde esse valor existia. A albumina se mostrou o principal preditor nesse modelo também.

Clark e Altman (102) discutem em seu trabalho a importância da imputação de dados para evitar a perda de poder estatístico e a introdução de vieses de informação. Em uma amostra de 1189 casos (842 óbitos) de câncer de ovário, os casos completos eram apenas 518 (380 óbitos). Através de método de Imputação Múltipla, foram incluídos 3 valores reais para cada imputado, o que permitiu a construção de um modelo com mais significância estatística e com maior habilidade preditiva. Concluem os autores que valores faltantes podem ser imputados em casos onde a razão para o dado estar ausente é conhecida, particularmente quando esse pode ser explicado por variáveis conhecidas. Este procedimento aumenta o poder da análise e produz modelos mais confiáveis e aplicáveis na prática clínica.

Segundo Daley e colegas (103), dados faltantes são um problema comum em estudos clínico-epidemiológicos, principalmente quando são retrospectivos. Não existe uma resposta correta para este problema. O sistema APACHE (30), que avalia gravidade de pacientes admitidos a unidades de tratamento intensivo, assume o valor considerado como limite normal para todas as variáveis laboratoriais faltantes. Mas em outras situações, a ausência de informações pode ser devida a motivos bem diferentes e esta solução não se mostrar aplicável. Principalmente em relação a resultados de exames laboratoriais, devemos entender porque estes estão ausentes, se devido à conduta do médico, necessidade clínica ou preferências do paciente ou mesmo se o paciente morreu antes dos testes serem realizados.

6.3- Regressão Logística

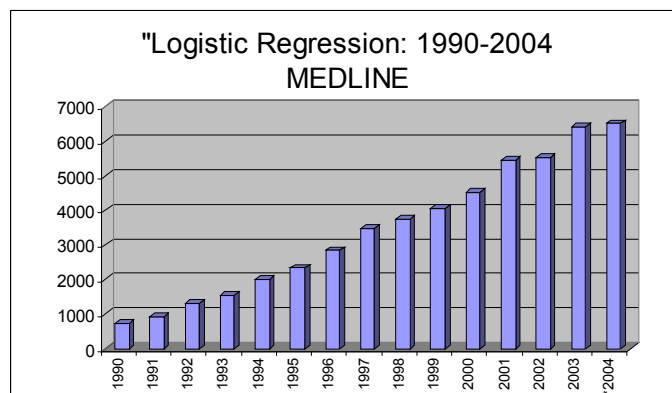
Modelos com múltiplas variáveis estão se tornando rotina em análises estatísticas na literatura médica. Estes métodos analisam a relação entre duas ou mais variáveis predictoras (variável independente, exposição) e uma variável de desfecho (dependente, resposta). Um modelo de regressão serve a dois propósitos: 1- prever a variável desfecho a partir de novos dados de variáveis predictoras; 2- ajudar a responder questões sobre a

área em estudo, porque o coeficiente de cada variável preditora explicitamente descreve a contribuição relativa de cada variável ao desfecho, automaticamente controlando para a influência das outras variáveis do modelo (104).

Muito do entendimento dos efeitos biológicos e seus determinantes são obtidos através de modelos de regressão, termo que teve sua origem no século XIX, com o estatístico inglês Francis Galton, que o utilizou para descrever a relação entre a altura de pais e de filhos. Seu amigo Karl Pearson desenvolveu as bases matemáticas para o que se denominou “análise de regressão”, uma técnica estatística usada para descrever e quantificar a relação entre duas ou mais variáveis (105).

Modelos de regressão logística se tornaram populares na área de epidemiologia clínica a partir do trabalho de Cornfield (1962) e de Truett, Cornfield e Kannel (1967), que realizaram uma análise multivariada com os dados do Estudo Framingham (106), sobre fatores de risco para doença coronariana. A partir deste verdadeiro marco histórico, estes modelos foram generalizados através da introdução de procedimentos de estimativas de máxima verossimilhança (Walter e Duncan, 1967; Day e Kerridge, 1967; Cox, 1970) e se tornaram o método padrão para análise de regressão de dados dicotômicos, especialmente na área da saúde. Referências clássicas para a RL são os trabalhos de Breslow e Day em 1980 (107) e de Kleinbaum, Kupper e Morgenstern em 1982 (108). Em 1989, Hosmer e Lemeshow publicaram seu clássico livro (*Applied Logistic Regression*), com o objetivo de tornar a metodologia de modelagem e a interpretação e relato dos resultados mais compreensível para potenciais usuários (106;107).

Foram encontradas 51.641 referências na MEDLINE com o termo “*logistic regression*”, no período de 1975 até 8 de outubro de 2004, e, como se pode ver na figura 4, é crescente o número absoluto (não ajustado para o número de publicações totais indexadas) de publicações que utilizam esta técnica de modelagem na área médica, especialmente a partir de 2001.



Muitas razões podem ser citadas para a popularidade da regressão logística como uma técnica estatística tanto em estudos epidemiológicos como em pesquisas clínicas. Primeiro, estes modelos ajudam a interpretar e a sumarizar dados, o que permite ao pesquisador estabelecer conclusões sobre seu objeto de estudo. Segundo, regressão logística é útil quando um desfecho dicotômico faz mais sentido do que um desfecho contínuo para responder a uma questão. A terceira razão para usar regressão logística é que os fatores explanatórios podem ser contínuos, categóricos ou binários. Finalmente, modelos de RL podem ser utilizados em análise de dados longitudinais ou transversais (109).

Referências Bibliográficas

- (1) Committee on Quality of Health Care in America. Crossing the Quality Chasm: a new health system for the 21st century. 1st edition ed. Washington, D.C.: National Academy Press, 2001.
- (2) Novaes HP. garantia de qualidade em hospitais da América Latina e Caribe. 1a ed. ignorado, 1992.
- (3) Committee on Quality of Health Care in America IoM. To err is human: building a safer health system. 1st ed. Washington, D.C.: National Academy Press, 2000.
- (4) Ball MJ, Douglas JV. IT, patient safety, and quality care. J Healthc Inf Manag 2002; 16(1):28-33.
- (5) Agency for Health Care Research and Quality. Guide to patient safety indicators. Rockville, MD, 2003.
- (6) Iezzoni LI. Risk adjustment for measuring health care outcomes. 3rd ed. Chicago, Illinois: Health Administration Press, 2003.
- (7) Klück MK, Guimarães JR, Ferreira J, Prompt CA. Definição de indicadores para a gestão da qualidade assistencial hospitalar. Epistula ALASS 2002; 49:10-12.
- (8) Quality Indicator Project. University of Maryland, editor. www.qiproject.org . 2004. Ref Type: Electronic Citation
- (9) Sistema de Indicadores Padronizados para Gestão Hospitalar. SIPAGEH, editor. www.sipageh.unisisnos.br . 2004. UNISINOS, RS. Ref Type: Electronic Citation
- (10) Risi Jr JB. Indicadores e informações em saúde no Brasil: a experiência RIPSA. Organização Panamericana de Saúde, editor. 2004. Ref Type: Unpublished Work
- (11) Donabedian A. The quality of care. How can it be assessed? 1988. Arch Pathol Lab Med 1997; 121(11):1145-1150.
- (12) Donabedian A. The seven pillars of quality. Arch Pathol Lab Med 1990; 114(11):1115-1118.
- (13) Rede Interagencial de Informações para a Saúde. Aspectos gerais. In: RIPSA/Organização Panamericana de Saúde, editor. Indicadores básicos de saúde no Brasil: conceitos e aplicações. Brasília: Organização Panamericana de Saúde, 2002.
- (14) McGlynn EA. The outcomes utility index: will outcomes data tell us what we want to know? Int J Qual Health Care 1998; 10(6):485-490.
- (15) Iezzoni LI. 100 apples divided by 15 red herrings: a cautionary tale from the mid-19th century on comparing hospital mortality rates. Ann Intern Med 1996; 124(12):1079-1085.
- (16) Iezzoni LI. The risks of risk adjustment. JAMA 1997; 278(19):1600-1607.
- (17) Iezzoni LI. Using risk-adjusted outcomes to assess clinical practice: an overview of issues pertaining to risk adjustment. Ann Thorac Surg 1994; 58(6):1822-1826.

- (18) Daley J, Forbes MG, Young GJ, Charns MP, Gibbs JO, Hur K et al. Validating risk-adjusted surgical outcomes: site visit assessment of process and structure. National VA Surgical Risk Study. *J Am Coll Surg* 1997; 185(4):341-351.
- (19) Grover FL, Shroyer AL, Hammermeister KE. Calculating risk and outcome: the Veterans Affairs database. *Ann Thorac Surg* 1996; 62(5 Suppl):S6-11.
- (20) Khuri SF, Daley J, Henderson W, Barbour G, Lowry P, Irvin G et al. The National Veterans Administration Surgical Risk Study: risk adjustment for the comparative assessment of the quality of surgical care. *J Am Coll Surg* 1995; 180(5):519-531.
- (21) Khuri SF, Daley J, Henderson W, Hur K, Demakis J, Aust JB et al. The Department of Veterans Affairs' NSQIP: the first national, validated, outcome-based, risk-adjusted, and peer-controlled program for the measurement and enhancement of the quality of surgical care. National VA Surgical Quality Improvement Program. *Ann Surg* 1998; 228(4):491-507.
- (22) Khuri SF, Daley J, Henderson W, Hur K, Gibbs JO, Barbour G et al. Risk adjustment of the postoperative mortality rate for the comparative assessment of the quality of surgical care: results of the National Veterans Affairs Surgical Risk Study. *J Am Coll Surg* 1997; 185(4):315-327.
- (23) Daley J, Forbes MG, Young GJ, Charns MP, Gibbs JO, Hur K et al. Validating risk-adjusted surgical outcomes: site visit assessment of process and structure. National VA Surgical Risk Study. *J Am Coll Surg* 1997; 185(4):341-351.
- (24) DesHarnais S.I., McLaughlin C.P. The outcome model of quality. In: McLaughlin C.P., Kaluzny AD, editors. *Continuous quality improvement in health care - Theory, implementation and applications*. Maryland, USA: Aspen Publishers, 1999: 59-92.
- (25) Nikas DJ, Freeman JE, Luterman AR, Warnken LC, Nenstiel RD, Farrell M et al. Use of a national data base to assess perioperative risk, morbidity, mortality, and cost savings in coronary artery bypass grafting. *South Med J* 1996; 89(11):1074-1077.
- (26) Dimick JB, Welch HG, Birkmeyer JD. Surgical mortality as an indicator of hospital quality: the problem with small sample size. *JAMA* 2004; 292(7):847-851.
- (27) Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. *Crit Care Med* 1985; 13(10):818-829.
- (28) Veltkamp SC, Kemmeren JM, van der GY, Edlinger M, van der WC. Prediction of serious complications in patients admitted to a surgical ward. *Br J Surg* 2002; 89(1):94-102.
- (29) Knaus WA, Zimmerman JE, Wagner DP, Draper EA, Lawrence DE. APACHE-acute physiology and chronic health evaluation: a physiologically based classification system. *Crit Care Med* 1981; 9(8):591-597.
- (30) Knaus WA, Wagner DP, Draper EA, Zimmerman JE, Bergner M, Bastos PG et al. The APACHE III prognostic system. Risk prediction of hospital mortality for critically ill hospitalized adults. *Chest* 1991; 100(6):1619-1636.
- (31) Barie PS, Hydo LJ, Fischer E. Comparison of APACHE II and III scoring systems for mortality prediction in critical surgical illness. *Arch Surg* 1995; 130(1):77-82.

- (32) Neary WD, Heather BP, Earnshaw JJ. The Physiological and Operative Severity Score for the enUmeration of Mortality and morbidity (POSSUM). *Br J Surg* 2003; 90(2):157-165.
- (33) Prytherch DR, Whiteley MS, Higgins B, Weaver PC, Prout WG, Powell SJ. POSSUM and Portsmouth POSSUM for predicting mortality. Physiological and Operative Severity Score for the enUmeration of Mortality and morbidity. *Br J Surg* 1998; 85(9):1217-1220.
- (34) Whiteley MS, Prytherch DR, Higgins B, Weaver PC, Prout WG. An evaluation of the POSSUM surgical scoring system. *Br J Surg* 1996; 83(6):812-815.
- (35) Hosmer DW, Hosmer T, Le Cessie S, Lemeshow S. A comparison of goodness-of-fit tests for the logistic regression model. *Stat Med* 1997; 16(9):965-980.
- (36) Bann SD, Sarin S. Comparative audit: the trouble with POSSUM. *J R Soc Med* 2001; 94(12):632-634.
- (37) Yii MK, Ng KJ. Risk-adjusted surgical audit with the POSSUM scoring system in a developing country. Physiological and Operative Severity Score for the enUmeration of Mortality and morbidity. *Br J Surg* 2002; 89(1):110-113.
- (38) Copeland GP, Jones D, Walters M. POSSUM: a scoring system for surgical audit. *Br J Surg* 1991; 78(3):355-360.
- (39) Bennett-Guerrero E, Hyam JA, Shaefi S, Prytherch DR, Sutton GL, Weaver PC et al. Comparison of P-POSSUM risk-adjusted mortality rates after surgery between patients in the USA and the UK. *Br J Surg* 2003; 90(12):1593-1598.
- (40) Le Gall JR, Loirat P, Alperovitch A, Glaser P, Granthil C, Mathieu D et al. A simplified acute physiology score for ICU patients. *Crit Care Med* 1984; 12(11):975-977.
- (41) Le Gall JR, Lemeshow S, Saulnier F. A new Simplified Acute Physiology Score (SAPS II) based on a European/North American multicenter study. *JAMA* 1993; 270(24):2957-2963.
- (42) Sutton R, Bann S, Brooks M, Sarin S. The Surgical Risk Scale as an improved tool for risk-adjusted analysis in comparative surgical audit. *Br J Surg* 2002; 89(6):763-768.
- (43) Daley J, Henderson WG, Khuri SF. Risk-adjusted surgical outcomes. *Annu Rev Med* 2001; 52:275-287.
- (44) Arozullah AM, Khuri SF, Henderson WG, Daley J. Development and validation of a multifactorial risk index for predicting postoperative pneumonia after major noncardiac surgery. *Ann Intern Med* 2001; 135(10):847-857.
- (45) Webster C, Neumayer L, Smout R, Horn S, Daley J, Henderson W et al. Prognostic models of abdominal wound dehiscence after laparotomy. *J Surg Res* 2003; 109(2):130-137.
- (46) Cook TM, Day CJ. Hospital mortality after urgent and emergency laparotomy in patients aged 65 yr and over. Risk and prediction of risk using multiple logistic regression analysis. *Br J Anaesth* 1998; 80(6):776-781.
- (47) Wouters SC, Noyez L, Verheugt FW, Brouwer RM. Preoperative prediction of early mortality and morbidity in coronary bypass surgery. *Cardiovasc Surg* 2002; 10(5):500-505.

- (48) Segal S. Women presenting in labor should be classified as ASA E. Harvard Medical School, editor. <http://www.soap.org.newsletter/winter2003/pro.html> . 2003. 27-2-2004. Ref Type: Electronic Citation
- (49) Saklad M. Grading of patients for surgical procedures. *Anesthesiology* 1941; 2:281-284.
- (50) Pratt S.D. Healthy parturients should be classified as ASA I. Harvard Medical School, editor. <http://www.soap.org.newsletters/fall2002/pro.html> . 2004. 25-8-2004. Ref Type: Electronic Citation
- (51) Keats AS. The ASA classification of physical status--a recapitulation. *Anesthesiology* 1978; 49(4):233-236.
- (52) New classification of physical status. *Anesthesiology* 1963; 24:111.
- (53) Mak PH, Campbell RC, Irwin MG. The ASA Physical Status Classification: inter-observer consistency. *American Society of Anesthesiologists. Anaesth Intensive Care* 2002; 30(5):633-640.
- (54) Owens WD. American Society of Anesthesiologists Physical Status Classification System in not a risk classification system. *Anesthesiology* 2001; 94(2):378.
- (55) Wolters U, Wolf T, Stutzer H, Schroder T. ASA classification and perioperative variables as predictors of postoperative outcome. *Br J Anaesth* 1996; 77(2):217-222.
- (56) Marx GF, Mateo CV, Orkin LR. Computer analysis of postanesthetic deaths. *Anesthesiology* 1973; 39(1):54-58.
- (57) Vacanti CJ, VanHouten RJ, Hill RC. A statistical analysis of the relationship of physical status to postoperative mortality in 68,388 cases. *Anesth Analg* 1970; 49(4):564-566.
- (58) Klück MK, Fachel JMG. Desenvolvimento e validação de um modelo de ajuste de risco para mortalidade cirúrgica pós laparotomia exploradora. 2004. Ref Type: Unpublished Work
- (59) Cullen DJ, Apolone G, Greenfield S, Guadagnoli E, Cleary P. ASA Physical Status and age predict morbidity after three surgical procedures. *Ann Surg* 1994; 220(1):3-9.
- (60) Prause G, Ratzenhofer-Comenda B, Pierer G, Smolle-Juttner F, Glanzer H, Smolle J. Can ASA grade or Goldman's cardiac risk index predict peri-operative mortality? A study of 16,227 patients. *Anaesthesia* 1997; 52(3):203-206.
- (61) Goldman L, Caldera DL, Nussbaum SR, Southwick FS, Krogstad D, Murray B et al. Multifactorial index of cardiac risk in noncardiac surgical procedures. *N Engl J Med* 1977; 297(16):845-850.
- (62) Goldman L. Cardiac risk in noncardiac surgery: an update. *Anesth Analg* 1995; 80(4):810-820.
- (63) Hall JC, Hall JL. ASA status and age predict adverse events after abdominal surgery. *J Qual Clin Pract* 1996; 16(2):103-108.
- (64) Haynes SR, Lawler PG. An assessment of the consistency of ASA physical status classification allocation. *Anaesthesia* 1995; 50(3):195-199.

- (65) Owens WD, Felts JA, Spitznagel EL, Jr. ASA physical status classifications: a study of consistency of ratings. *Anesthesiology* 1978; 49(4):239-243.
- (66) Gray GE, Meguid MM. Can total parenteral nutrition reverse hypoalbuminemia in oncology patients? *Nutrition* 1990; 6(3):225-228.
- (67) Muir AJ. Albumina sérica.
<http://www.nlm.nih.gov/medlineplus/ency/article/003480.htm> . 2004. Ref Type: Electronic Citation
- (68) Doumas BT, Watson WA, Biggs HG. Albumin standards and the measurement of serum albumin with bromocresol green. *Clin Chim Acta* 1971; 31(1):87-96.
- (69) Kral JG. Nutrition and metabolism in the surgical patient. 1997. Ref Type: Personal Communication
- (70) Cunha DF. Diretrizes para o emprego adequado de albumina humana. *Rev Bras Clin Terap* 1997; 23:87-92.
- (71) Gibbs J, Cull W, Henderson W, Daley J, Hur K, Khuri SF. Preoperative serum albumin level as a predictor of operative mortality and morbidity: results from the National VA Surgical Risk Study. *Arch Surg* 1999; 134(1):36-42.
- (72) Jones C.M., Eaton F.B. Postoperative nutritional edema. *Arch Surg* 1933; 27:159-177.
- (73) Mullen JL, Gertner MH, Buzby GP, Goodhart GL, Rosato EF. Implications of malnutrition in the surgical patient. *Arch Surg* 1979; 114(2):121-125.
- (74) Buzby GP, Mullen JL, Matthews DC, Hobbs CL, Rosato EF. Prognostic nutritional index in gastrointestinal surgery. *Am J Surg* 1980; 139(1):160-167.
- (75) Delgado-Rodriguez M, Medina-Cuadros M, Gomez-Ortega A, Martinez-Gallego G, Mariscal-Ortiz M, Martinez-Gonzalez MA et al. Cholesterol and serum albumin levels as predictors of cross infection, death, and length of hospital stay. *Arch Surg* 2002; 137(7):805-812.
- (76) Kudsk KA, Tolley EA, DeWitt RC, Janu PG, Blackwell AP, Yeary S et al. Preoperative albumin and surgical site identify surgical risk for major postoperative complications. *JPEN J Parenter Enteral Nutr* 2003; 27(1):1-9.
- (77) Vincent JL, Dubois MJ, Navickis RJ, Wilkes MM. Hypoalbuminemia in acute illness: is there a rationale for intervention? A meta-analysis of cohort studies and controlled trials. *Ann Surg* 2003; 237(3):319-334.
- (78) Goldwasser P, Feldman J. Association of serum albumin and mortality risk. *J Clin Epidemiol* 1997; 50(6):693-703.
- (79) Lubin MF. Is age a risk factor for surgery? *Med Clin North Am* 1993; 77(2):327-333.
- (80) Finlayson EV, Birkmeyer JD. Operative mortality with elective surgery in older adults. *Eff Clin Pract* 2001; 4(4):172-177.
- (81) Arenal JJ, Bengoechea-Beeby M. Mortality associated with emergency abdominal surgery in the elderly. *Can J Surg* 2003; 46(2):111-116.

- (82) Normand ST, Glickman ME, Sharma RG, McNeil BJ. Using admission characteristics to predict short-term mortality from myocardial infarction in elderly patients. Results from the Cooperative Cardiovascular Project. *JAMA* 1996; 275(17):1322-1328.
- (83) Fortescue EB, Kahn K, Bates DW. Development and validation of a clinical prediction rule for major adverse outcomes in coronary bypass grafting. *Am J Cardiol* 2001; 88(11):1251-1258.
- (84) NCEPOD. NCEPOD classification of surgeries. National Confidential Enquiry into Perioperative Deaths, editor. www.ncepod.org.uk . 2004. 25-9-2004. Ref Type: Electronic Citation
- (85) Sadeghi N, Sadeghi S, Mood ZA, Karimi A. Determinants of operative mortality following primary coronary artery bypass surgery. *Eur J Cardiothorac Surg* 2002; 21(2):187-192.
- (86) Edwards FH, Peterson ED, Coombs LP, DeLong ER, Jamieson WR, Shroyer ALW et al. Prediction of operative mortality after valve replacement surgery. *J Am Coll Cardiol* 2001; 37(3):885-892.
- (87) Dudley RA, Johansen KL, Brand R, Rennie DJ, Milstein A. Selective referral to high-volume hospitals: estimating potentially avoidable deaths. *JAMA* 2000; 283(9):1159-1166.
- (88) Birkmeyer JD, Siewers AE, Finlayson EV, Stukel TA, Lucas FL, Batista I et al. Hospital volume and surgical mortality in the United States. *N Engl J Med* 2002; 346(15):1128-1137.
- (89) Barry WA, Rosenthal GE. Is there a July phenomenon? The effect of July admission on intensive care mortality and length of stay in teaching hospitals. *J Gen Intern Med* 2003; 18(8):639-645.
- (90) Romano PS. Can administrative data be used to compare the quality of health care? *Med Care Rev* 1993; 50(4):451-477.
- (91) Tokarski C. Buyer beware: what every surgeon should know about the quality of data in the medical literature. Medscape, editor. <http://www.medscape.com/viewarticle/464285> . 2003. Ref Type: Electronic Citation
- (92) DATASUS. Movimento mensal de internação hospitalar. www.datasus.gov.br/ arquivo de dados-MS BBS . 25-9-2004. Ref Type: Electronic Citation
- (93) Iezzoni LI. Risk adjustment for medical effectiveness research: an overview of conceptual and methodological considerations. *J Investig Med* 1995; 43(2):136-150.
- (94) Romano PS, Roos LL, Luft HS, Jollis JG, Doliszny K. A comparison of administrative versus clinical data: coronary artery bypass surgery as an example. Ischemic Heart Disease Patient Outcomes Research Team. *J Clin Epidemiol* 1994; 47(3):249-260.
- (95) Romano PS, Mark DH. Bias in the coding of hospital discharge data and its implications for quality assessment. *Med Care* 1994; 32(1):81-90.
- (96) Romano PS, Schembri ME, Rainwater JA. Can administrative data be used to ascertain clinically significant postoperative complications? *Am J Med Qual* 2002; 17(4):145-154.

- (97) Best WR, Khuri SF, Phelan M, Hur K, Henderson WG, Demakis JG et al. Identifying patient preoperative risk factors and postoperative adverse events in administrative databases: results from the Department of Veterans Affairs National Surgical Quality Improvement Program. *J Am Coll Surg* 2002; 194(3):257-266.
- (98) Pine M, Jones B, Lou YB. Laboratory values improve predictions of hospital mortality. *Int J Qual Health Care* 1998; 10(6):491-501.
- (99) Martins M, Travassos C, Carvalho dN. [Hospital Information Systems as risk adjustment in performance indicators]. *Rev Saude Publica* 2001; 35(2):185-192.
- (100) Iezzoni LI, Ash AS, Shwartz M, Daley J, Hughes JS, Mackiernan YD. Predicting who dies depends on how severity is measured: implications for evaluating patient outcomes. *Ann Intern Med* 1995; 123(10):763-770.
- (101) Data imputation. Australian Government - Institute of Family Studies . 11-8-2004. Ref Type: Internet Communication
- (102) Clark TG, Altman DG. Developing a prognostic model in the presence of missing data: an ovarian cancer case study. *J Clin Epidemiol* 2003; 56(1):28-37.
- (103) Daley J, Iezzoni LI, Shwartz M. Conceptual and practical issues in developing risk-adjusted methods. In: Iezzoni LI, editor. *Risk adjustment for measuring health care outcomes*. Chicago, Illinois: Health Administration Press, 2003: 179-205.
- (104) Bagley SC, White H, Golomb BA. Logistic regression in the medical literature: standards for use and reporting, with particular attention to one medical domain. *J Clin Epidemiol* 2001; 54(10):979-985.
- (105) Anderson RP, Jin R, Grunkemeier GL. Understanding logistic regression analysis in clinical reports: an introduction. *Ann Thorac Surg* 2003; 75(3):753-757.
- (106) Hosmer DW, Lemeshow S. *Applied logistic regression*. 1st ed. New York, NY, USA: John Wiley & Sons, 1989.
- (107) Breslow NE, Day NE. *Statistical methods in cancer research. Volume I - The analysis of case-control studies*. IARC Sci.Publ., 1980.
- (108) Kleinbaum DG, Morgenstern H, Kupper LL. Selection bias in epidemiologic studies. *Am J Epidemiol* 1981; 113(4):452-463.
- (109) Ostir GV, Uchida T. Logistic regression: a nontechnical review. *Am J Phys Med Rehabil* 2000; 79(6):565-572.

Artigo 1:
Estratégias de modelagem por Regressão Logística
para Ajuste de Risco

Estratégias de modelagem por regressão logística para o ajuste de risco

Mariza Klück^(1,2,3) e Jandyra M. G. Fachel^(1,4)

(1) Programa de Pós-Graduação em Epidemiologia da Universidade Federal do Rio Grande do Sul (UFRGS)

(2) Departamento de Medicina Social da Faculdade de Medicina da UFRGS

(3) Comissão de Prontuários de Pacientes do Hospital de Clínicas de Porto Alegre

(4) Departamento de Estatística do Instituto de Matemática da UFRGS

Resumo

O crescente interesse pela avaliação da qualidade assistencial através de desfechos clínicos relevantes e da utilização de referenciais comparativos tem provocado a discussão sobre a necessidade de ajustar estes indicadores pelo risco inerente ao paciente, o que permitiria a comparação válida dos desempenhos relativos de cada equipe ou hospital. A construção de modelos para ajuste de risco exige uma série de etapas que devem ser conduzidas com o necessário rigor metodológico. Embora já presente na produção científica americana, este tipo de estudo ainda é pouco freqüente no Brasil. Este artigo descreve uma metodologia proposta para o desenvolvimento, validação e aplicação de escores de risco através de modelagem por regressão logística a partir de informações disponíveis em base de dados clínicos. Para exemplificar a metodologia, que é apresentada passo a passo, foram utilizados os dados de uma amostra de 633 pacientes adultos internados no Hospital de Clínicas de Porto Alegre e submetidos a cirurgia com incisão abdominal no período de outubro de 2003 a março de 2004, considerando-se como desfecho o óbito ocorrido na mesma internação.

Palavras-chave: ajuste de risco, regressão logística, gestão da qualidade assistencial, indicadores de desfechos hospitalares

Abstract

There is a growing interest in using the results of outcomes-based, health care process quality evaluation for comparative purposes. This has generated much discussion about how such indicators should be adjusted according to the level of risk inherent to the patients being subjected to the processes. Only by means of such an adjustment can the resulting outcome indicators be used to make valid comparisons. The authors propose, herein, a methodology, based on a logistical regression model, that allows valid risk scores to be calculated from data readily available in the patient record. The model is based on a sample of 633 adult laparotomy patients admitted to the Hospital de Clínicas de Porto Alegre between October of 2003 and March of 2004 using surgical process related mortality as the outcome indicator. All of the independent variables used in the model were available in the patient record.

Key words: risk adjustment, logistic regression, quality indicators in health care

Introdução

A quantificação da severidade do quadro clínico dos pacientes criticamente doentes é controversa e tem sido feita de diferentes maneiras. Os pontos de controvérsia incluem a avaliação interna da qualidade da assistência ou de programas de melhoria, apoio à tomada de decisão clínica em relação a pacientes individuais, a comparação da qualidade assistencial entre hospitais e assegurar equivalência clínica entre grupos para investigação e pesquisa clínicas.

Tema de interesse crescente, a utilização de medidas de desfechos de pacientes para avaliar a qualidade da assistência e a preocupação com o ajuste destas medidas é discutida há mais de um século. Em 1863, Florence Nightingale argumentava que os hospitais londrinos eram perigosos, especialmente quando comparados aos hospitais das zonas rurais. Sua argumentação baseava-se nos dados publicados por ela e William Farr, “*Notes on Hospitals*” (1), que apresentava estatísticas de óbitos de 106 hospitais ingleses referentes ao ano de 1861. A própria Florence, em resposta às críticas recebidas, admitiu posteriormente que os números absolutos de óbitos em hospitais não expressavam adequadamente a qualidade dos cuidados pois se deveriam considerar “diferenças na idade e no estado do paciente na admissão” (2)

Em 1930, o Dr. Ernest Codman publicou um famoso estudo sobre os desfechos de saúde nos Estados Unidos, que enfatizava a importância de avaliar a qualidade assistencial e a necessidade de levar em consideração a severidade ou o estágio da doença, as comorbidades, a atitude do paciente em relação à saúde e à doença e as barreiras econômicas para receber o atendimento (3).

Nos anos a seguir, pesquisadores examinaram várias características dos hospitais para determinar se estas estavam relacionadas com as taxas de óbitos de pacientes. Somente com o surgimento das grandes bases de dados administrativas criadas pela necessidade dos sistemas de cobrança de serviços assistenciais pelos grandes pagadores (*Medicare, Veterans Affairs Department, Sistema Único de Saúde (SUS)* e outros) foi possível o desenvolvimento de modelos de risco baseados em características intrínsecas dos pacientes. Estes sistemas estão se tornando mais acurados e discriminatórios através da utilização de bases de dados clínicas (prontuário eletrônico), recurso cada vez mais freqüente em grandes hospitais universitários.

O ajuste de risco (ou padronização para severidade) é um processo que visa minimizar a influência confundidora de potenciais fatores de risco inerentes ao paciente e/ou a sua patologia, permitindo assim a comparação da qualidade do tratamento hospitalar, ajustando os desfechos encontrados para o risco prévio do paciente (dados

demográficos, sócio-econômicos e diagnósticos). Tais ajustamentos teoricamente minimizam o potencial para falsos julgamentos sobre qualidade assistencial devido a desfechos indesejados cuja prevenção ou controle estão fora do controle do provedor (4).

Planejar estratégias de ajuste é complicado, requerendo a identificação de fatores de risco, isto é, as características clínicas e outras que refletem a probabilidade do paciente de vir a sofrer o desfecho em estudo. Uma vez identificados, estes fatores de risco devem ser operacionalizados e medidos de forma confiável e sem vieses e precisam ser combinados em formatos normativos ou empíricos para produzir um escore de risco ou uma classificação para cada paciente.

Um sistema de avaliação de risco que permita o ajuste de indicadores de desfecho pode ser desenvolvido através de diferentes metodologias (5-8) e vários destes já são utilizados em diferentes locais e para diferentes situações (9-17), com maior ou menor especificidade. Ajuste de risco (*Risk Adjustment*) está se tornando um tema freqüente na literatura médica. Consulta à MEDLINE (17/10/2004) revela a existência de 985 referências de artigos que têm a expressão “risk adjustment” no texto e 577, no título ou no *abstract*, de 1990 até o presente. Em relação ao número total de títulos indexados, percebe-se um interesse crescente neste tema a partir de 1998, mantido até o presente (Figura 1).

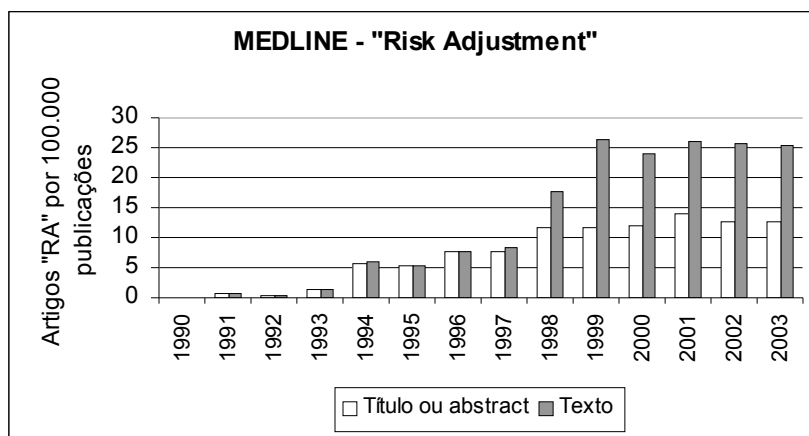


Figura 1: Número de publicações (Medline) sobre Ajuste de Risco por 100.000 artigos

No Brasil esta metodologia ainda não está sendo explorada com a intensidade que seria necessária, razão pela qual a mesma deve ser divulgada e adotada com mais freqüência nas linhas de pesquisa da Epidemiologia. Pesquisa realizada na base LILACS recuperou apenas dois artigos publicados com este termo (18;19) no país.

Criar um método de ajuste de risco (*risk adjustment*) que seja clinicamente razoável e, ao mesmo tempo, logisticamente realizável é desafiador e exige algumas concessões,

mas é condição essencial para a correta avaliação da qualidade assistencial. É importante dominar os conhecimentos necessários para a construção de modelos de risco adequados às diferentes realidades e necessidades.

O objetivo deste trabalho é o de descrever e discutir uma metodologia adaptada à realidade local, proposta para o desenvolvimento, validação e aplicação de escores de risco que permitam a comparação de desfechos entre diferentes hospitais, ajustando estes desfechos às características previamente existentes do paciente.

Ajuste de indicadores por risco

A construção de um modelo de escore de risco acurado, calibrado e discriminatório requer alguns passos fundamentais. Um sistema de avaliação de risco que permita o ajuste de indicadores pode ser desenvolvido através de diferentes metodologias, que, via de regra, utilizam técnicas de regressão logística para análise estatística. A partir da revisão da literatura realizada e experiência prévia das autoras, a seqüência de etapas apresentadas no quadro 1 é proposta para este processo.

As etapas sugeridas serão abordadas detalhadamente e exemplificadas a seguir. Os dados utilizados para geração dos exemplos utilizados são de 633 pacientes internados no Hospital de Clínicas de Porto Alegre e submetidos à cirurgia com incisão abdominal, realizada por equipe de cirurgia geral sob anestesia assistida, no período de outubro de 2003 a março de 2004. O desfecho considerado foi o de óbito durante a internação.

1- PLANEJAMENTO

- Definir o desfecho de relevância
- Selecionar variáveis potencialmente relacionadas ao desfecho
- Avaliar as fontes de dados disponíveis
- Determinar a amostra e planejar amostras para derivação e validação do escore
- Equacionar questões éticas e legais

2- PREPARAÇÃO DO CONJUNTO DE DADOS ANALÍTICO

- Coleta dos dados
- Tratamento dos dados
 - ✓ Tratar variáveis brutas (excluir registros incompletos ou duplos, definir ou obter variáveis a partir de dados primários, criptografar dados sigilosos)
 - ✓ Se indicado, categorizar variáveis quantitativas, reagrupar variáveis categóricas
 - ✓ Tratar dados faltantes (imputação ou exclusão)

3- CONSTRUÇÃO DO MODELO (Amostra de derivação)

- Análise univariável
- Análise multivariável (Regressão Logística - RL)
- Avaliação da consistência do modelo

4- AVALIAÇÃO DA ADEQUAÇÃO DO MODELO

- Acurácia (tabela de classificação, Sensibilidade, Especificidade)
- Calibração (estatística de Hosmer e Lemeshow)
- Discriminação (Índice-C / Curva ROC)

5- CRIAÇÃO DO ESCORE DE RISCO (amostras de Derivação e de Validação)

- Atribuir pesos para cada variável do modelo a partir de coeficientes da RL
- Criar o escore de risco
- Calcular o escore de risco de cada paciente
- Criar classes de risco de acordo com probabilidade de desenvolver o desfecho

6- VALIDAÇÃO

- Nova regressão logística – utilizar escore de cada paciente como variável independente, calculando as probabilidades individuais
- Comparar acurácia, calibração e discriminação obtidas nas duas amostras

7- APLICAÇÃO DO ESCORE DE RISCO PARA COMPARAÇÃO

- Calcular probabilidades de desfecho para cada paciente de todos os hospitais
- Somar probabilidades definidas pela RL (E) e somar desfechos observados de cada hospital (O)
- Calcular a razão (O/E)
- Classificar hospitais de acordo com (O/E)

Quadro 1: Etapas propostas para desenvolvimento de modelos de ajuste de indicadores por risco

1- PLANEJAMENTO

A etapa de planejamento de uma pesquisa é de fundamental importância para o sucesso da mesma. Segundo Daley e colegas (20), nesta fase se deve responder às perguntas: Risco de qual desfecho? Em que período? Para qual população? Para que propósito?

É nesta etapa que o pesquisador deverá definir sua questão de pesquisa, definir sua população de estudo em termos de origem, composição e tamanho, planejar as medidas necessárias, equacionar as questões éticas e legais pertinentes e buscar os recursos necessários para sua realização. Uma pesquisa que objetive permitir o ajuste de um determinado desfecho ou indicador de qualidade assistencial para o risco inerente ao paciente requer algumas etapas de planejamento bem específicas, que serão definidas a seguir.

1.1- Seleção e definição do desfecho de relevância

Os desfechos devem ser definidos bem claramente e, sempre que possível, adotados desfechos já padronizados para que se possa fazer comparações com outros hospitais ou países. O desfecho deve ser freqüente o suficiente para garantir estabilidade em modelos estatísticos. Mesmo em grandes bases de dados não é possível desenvolver um modelo adequado para eventos muito raros sem a utilização de técnicas estatísticas mais complexas (21).

Ao selecionar o desfecho de interesse, o pesquisador deve levar em conta a relevância (magnitude, transcendência e vulnerabilidade) e a utilidade do mesmo (manejo clínico do paciente, avaliação de qualidade, auditorias comparativas etc.). Vários são os desfechos da assistência em ambiente hospitalar que atendem a estes critérios, tais como mortalidade, morbidade, tempo de permanência, reinternação não-programada, tempo de sobrevivência, qualidade de vida, satisfação do paciente, relação custo/efetividade e outros.

A definição precisa do desfecho de interesse não é sempre fácil. O pesquisador deve considerar um balanço entre o indicador desejado e aquele possível de ser medido. É importante também buscar a padronização com trabalhos já publicados, medindo e definindo o desfecho da mesma forma como o é comumente na literatura, para permitir comparações.

Vejamos por exemplo o desfecho de mortalidade cirúrgica. Na literatura especializada encontramos como definição mais correta os óbitos ocorridos em até 30 dias após a cirurgia índice (22). No entanto, após a alta do paciente torna-se muito difícil e dispendioso o acompanhamento deste e a sua avaliação no 30º dia. Para resolver o problema, outros autores definem mortalidade cirúrgica como os óbitos ocorridos durante a internação da cirurgia índice, desfecho de fácil controle e avaliação mesmo em estudos retrospectivos (14;16). Para tornar o indicador mais preciso, os pesquisadores do *National Veterans Affairs Surgical Risk Study* (23) definem este desfecho como uma combinação

das duas definições anteriores: óbitos ocorridos durante a internação, em um período de até 30 dias após a cirurgia-índice (24).

No estudo que exemplifica os conceitos apresentados, foram incluídos 633 pacientes, dos quais 85 (13,4%) morreram durante a internação.

1.2- Identificação das variáveis potencialmente relacionadas ao desfecho

Uma vez selecionado e definido o desfecho, o pesquisador deve identificar as possíveis variáveis associadas a este. A identificação de todos os potenciais fatores de risco e outras variáveis associadas ao desfecho de interesse é etapa de extrema relevância para a condução deste tipo de estudo. Cabe ao pesquisador estabelecer *a priori* um modelo conceitual sobre quais seriam os potenciais fatores de risco que deveriam ser incluídos em um índice para ajuste de um determinado desfecho. Para tanto, o pesquisador deve considerar as variáveis sugeridas por revisão da literatura e opinião de especialistas, atentando para sua plausibilidade biológica, sua relevância clínica bem como para a disponibilidade e confiabilidade dos dados necessários.

As variáveis geralmente selecionadas em estudos para ajuste de risco de desfechos hospitalares são aquelas disponíveis do momento da admissão do paciente até um período adequado de tempo (necessário para o primeiro exame clínico e obtenção de dados laboratoriais iniciais), para evitar que sejam incluídos como fatores de risco condições “iatrogênicas”, isto é, causadas ou agudizadas pelo processo de atenção. Por vezes é difícil separar, em registros de alta hospitalar, condições pré-existentes daquelas complicações ocorridas durante a internação.

Estas variáveis podem ser características relativas ao indivíduo, ao episódio de doença que o levou ao hospital e às condições do paciente no início da hospitalização. Para exemplificar, as variáveis selecionadas para avaliação no estudo sobre fatores de risco associados à mortalidade cirúrgica no HCPA foram as seguintes:

Indivíduo: sexo, idade, etnia, escolaridade, procedência, situação conjugal, convênio

Internação: urgência/eletiva, mês, clínica médica ou cirúrgica

Cirurgia: urgência/eletiva, procedimento, equipe, turno, tipo de anestesia

Paciente: Classificação pré-anestésica ASA ((25), albumina sérica

Um modelo conceitual pré-estabelecido, aliado a uma revisão da literatura e consulta a especialistas bem como a uma avaliação das fontes de dados disponíveis, darão ao pesquisador os subsídios necessários para definir o conjunto de dados que deverão ser coletados a seguir.

1.3- Avaliação das fontes de dados disponíveis e da qualidade e consistência dos dados

O pré-requisito mais importante para o uso de dados de prontuário na avaliação comparativa da qualidade assistencial é a aquisição de informações abrangentes, confiáveis, acuradas e coletadas de forma metódica, que tenham credibilidade clínica (26).

Em geral, os pesquisadores utilizam dados obtidos de bases de dados eletrônicas geradas por hospitais e sistemas de saúde para fins de pagamentos de serviços, complementando com informações clínicas adicionais buscadas em prontuário, quando necessário. As pesquisas em bases de dados administrativas são mais simples, fáceis, rápidas e baratas, mas podem não oferecer a credibilidade clínica necessária para este tipo de modelagem. Bases de dados administrativas contêm informação clínica limitada (em geral, alguns dados sócio-demográficos do indivíduo, tipo de admissão e de saída, diagnóstico principal e procedimentos realizados), que não contempla dados como comorbidades ou cronologia dos fatos e que podem apresentar um viés de codificação diagnóstica decorrente da necessidade de otimização do faturamento.

Prontuários de pacientes e coleta de dados prospectiva oferecem informações mais confiáveis e acuradas, com boa validade clínica, mas estas fontes exigem mais recursos e mais tempo para serem utilizadas.

O pesquisador deve avaliar as fontes de dados disponíveis dentro e fora do hospital, buscando encontrar um balanço adequado entre a qualidade e quantidade de dados necessários com o tempo e os recursos necessários para sua obtenção. Estudos prospectivos geram dados de maior qualidade, pois o pesquisador terá controle sobre suas coletas e registros. Estudos retrospectivos devem se apoiar em fontes de dados adequadas. Estas fontes podem ser de caráter mais administrativo (bases de dados de internações hospitalares (AIH) do SUS, por exemplo) ou caráter clínico (prontuários de pacientes).

Sistemas de prontuários eletrônicos são as fontes mais qualificadas de dados para o desenvolvimento de modelos adequados para fatores de risco, pois são fontes primárias, sem transcrição, legíveis, fornecem os dados em formato eletrônico, são de fácil e rápida aquisição e, se o sistema é de boa qualidade, garante a acurácia dos dados, além do baixo custo. No entanto, poucos hospitais ou instituições oferecem atualmente este recurso de forma adequada, pois o custo desta tecnologia ainda é muito alto e de difícil implementação.

O sistema ideal, portanto, é aquele que alia a acurácia, confiabilidade e credibilidade dos dados clínicos com o fácil acesso eletrônico, o que, no entanto, é oferecido por poucos hospitais no Brasil ou mesmo no resto do mundo.

Para o estudo sobre mortalidade cirúrgica no HCPA, todas as variáveis estavam disponíveis no sistema de prontuário eletrônico da instituição.

1.4- Definição da população de pesquisa, planejamento de sua formação, determinação do tamanho da amostra

A definição da população de pesquisa e sua correta amostragem também deve se ater à adequação da população ao desfecho em análise, em termos de características de composição e tamanho. A prevalência do desfecho, bem como a possibilidade de se obter os dados necessários em tempo hábil devem também ser consideradas.

Quando tratamos com pesquisas para ajuste de riscos de desfechos hospitalares, a seleção da amostra em geral não se configura como um problema pois devemos trabalhar com eventos freqüentes (não raros) experienciados por pacientes identificados e com dados já coletados (estudos retrospectivos) ou possíveis de serem coletados dentro do ambiente hospitalar (estudos prospectivos).

Estimativas prognósticas são mais acuradas quando baseadas em análises cuidadosas de dados de boa qualidade (27). Quando um conjunto de dados adequado é obtido, este deve ser dividido em uma amostra de derivação (desenvolvimento treinamento, aprendizado) e uma amostra de validação (teste). A amostra de validação não deve ser usada no desenvolvimento do modelo e seu propósito é o de permitir a avaliação da acurácia preditiva do modelo. Sem essa amostra não é possível saber se as predições obtidas pelo modelo poderão ser replicadas em outras amostras de pacientes ou se o modelo apenas se ajusta à amostra de derivação. Esta divisão é geralmente feita com proporções que variam de dois terços (5;7;28;29) a um meio (6;8;9;13) da amostra para o desenvolvimento do modelo. A seleção de pacientes das duas amostras pode ser feita aleatoriamente ou por critérios cronológicos, sendo que a segunda opção permite que o modelo seja desenvolvido enquanto ainda se coletam os dados necessários para sua validação.

Um número pequeno de sujeitos e/ou de pacientes que sofreram o desfecho não irão gerar um modelo estatisticamente estável, tornando os resultados pouco generalizáveis. De acordo com o delineamento da pesquisa, pode ser necessário o cálculo do tamanho amostral ou pode-se trabalhar com todos os pacientes submetidos a um episódio em estudo (internação, cirurgia etc.) e que se submeteram ao risco de desenvolver um

determinado desfecho. Todas as técnicas disponíveis exigem um tamanho de amostra adequado, que vai depender do número de variáveis requerido pelo modelo final.

O aspecto crítico do tamanho da amostra é usualmente o número de pacientes que sofreram o desfecho e não o tamanho da amostra. Autores sugerem uma regra prática útil para avaliar a adequação da amostra à metodologia, que é a de 10 desfechos (aquele menos freqüente) por variável preditora do modelo final (30;31) ou de 25 sujeitos por variável (32). O número de preditores se refere ao número de coeficientes do modelo ou parâmetros. Hsieh publicou tabelas para cálculo de tamanho de amostra para regressão logística univariável e multivariável que são relativamente fáceis de serem utilizadas (33).

Em nosso exemplo, a amostra de derivação foi composta por 423 pacientes e a de validação, por outros 210 pacientes, selecionados aleatoriamente entre os 633 registros da base de dados.

1.5- Questões éticas

Ao planejar a pesquisa deve-se considerar as questões éticas e legais envolvidas. Os hospitais devem exigir do pesquisador a aprovação de seu projeto por um Comitê de Ética em Pesquisa credenciado e qualificado, nem sempre do próprio hospital. Para ser aprovado, além da qualidade metodológica e da relevância do tema, o projeto deverá considerar a garantia do sigilo dos dados quanto à identificação dos pacientes ou dos profissionais envolvidos em seu atendimento (através de criptografia, por exemplo) e a necessidade de obtenção de um termo de consentimento informado do paciente, geralmente exigido em pesquisas prospectivas. É recomendável também comunicar a realização da pesquisa ao Serviço Médico responsável pelo atendimento dos pacientes cujos dados serão utilizados, caso o pesquisador não seja membro do mesmo.

2- PREPARAÇÃO DO CONJUNTO DE DADOS ANALÍTICO

A preparação de um banco de dados adequado para a análise é uma etapa de fundamental relevância para o sucesso do projeto e deve ser vista com muita atenção.

2.1- Coleta dos dados

Quando os dados são obtidos de bases de dados eletrônicas (administrativas ou clínicas) é importante que o pesquisador interaja pessoalmente com o analista ou responsável pela extração dos mesmos ou consulte com atenção as informações e instruções existentes nos *sites* de busca, para certificar-se de que os dados recebidos correspondam exatamente às variáveis solicitadas. Muitas vezes estes conceitos não estão claramente definidos para um profissional externo à área assistencial.

A busca em prontuário de suporte em papel deve ser planejada e programada com o Serviço de Arquivo Médico dos hospitais envolvidos, pois em geral exigirá a consulta a

um número muito grande de pastas que devem ser selecionadas e localizadas. A coleta de dados deve ser preferencialmente realizada por pessoas com conhecimento da área em estudo (médicos, profissionais de saúde, estudantes da área) e seu registro pode ser feito em formulários específicos diretamente no computador ou em papel. A entrada direta é mais rápida e evita transcrição mas, por outro lado, o registro em papel permite uma documentação adequada e a possibilidade de correções de eventuais erros no momento da digitação.

A coleta prospectiva exige o treinamento de uma equipe de entrevistadores ou coletadores de informações, para garantir a uniformidade e a padronização de perguntas e respostas.

Seja qual for o método disponível, é necessário que o pesquisador tenha domínio e controle sobre o processo de coleta de dados, garantindo sua fidedignidade.

2.2- Tratamento preliminar dos dados

Depois de coletados e colocados em formato eletrônico, os dados devem ser criteriosamente analisados e tratados, com o objetivo de prepará-los para a análise estatística. É necessário realizar uma avaliação preliminar dos mesmos quanto ao conteúdo, completude, limites aceitáveis para os valores das variáveis e necessidade de tratamento matemático para dados brutos. Outros cuidados incluem a limpeza de registros que não satisfaçam os critérios de inclusão, o tratamento de dados faltantes, decisão sobre manter ou não variáveis quantitativas como tal e proteção do sigilo das informações confidenciais (identificação de pacientes ou médicos, por exemplo) através de procedimentos de criptografia.

Em primeiro lugar, o banco de dados deve ser “limpo”, corrigindo-se a digitação dos dados, eliminando-se os registros duplos, incompletos ou fora dos limites aceitáveis. Após, deve-se calcular e/ou codificar as variáveis que não foram coletadas em seu formato final ou ideal. Bases de dados eletrônicas fornecem resultados de busca nos formatos originais (brutos) das variáveis, exigindo que se calculem informações como intervalos de tempo (idade, tempo de internação, duração da cirurgia etc.) a partir de datas ou horários de nascimento, de internação, de alta ou de realização de um determinado exame ou procedimento.

Também é importante codificar numericamente informações fornecidas em formato de texto, garantindo maior facilidade e agilidade de armazenamento e de análise e, sempre que conceitualmente possível, as categorias existentes deverão ser agrupadas, idealmente em variáveis dicotômicas que expressem condições de risco/não-risco. Variáveis dicotômicas independentes, tais como a presença ou ausência de um fator de

risco, são codificadas como 1 ou 0; se 1, o coeficiente é adicionado e se 0, este não é adicionado à equação de RL.

Variáveis necessárias para a análise que permitam a quebra do nível de sigilo exigido, tais como dados que permitam a identificação do paciente ou do profissional responsável, devem ser criptografadas.

Outro aspecto importante é a verificação da existência de valores extremos, fora dos limites aceitáveis ou esperados (*outliers*). Estes dados devem ser verificados cuidadosamente e, sendo verdadeiros, o pesquisador pode decidir pela exclusão daquela registro ou pela sua manutenção, não esquecendo de comentar esta ocorrência se a mesma repercutir sobre os resultados obtidos.

Na Tabela 1 mostramos alguns exemplos de tratamento de dados da pesquisa de mortalidade cirúrgica no HCPA.

Tabela 1: Exemplos de variáveis utilizadas em estudo de mortalidade cirúrgica e seus valores numéricos

Variável	Formato original	Formato codificado
Sexo	Masculino	1
	Feminino	0
Idade na alta	Data de nascimento (dd/mm/aaaa) Data da alta (dd/mm/aaaa)	(Data da alta) – (data de nascimento)= idade em anos
Duração da cirurgia	Hora de início e hora de fim	Duração em minutos
Turno de realização da cirurgia	Hora de início da cirurgia	1-Início entre 19 e 7 horas 0- Início entre 7 e 19 horas
Situação conjugal	Solteiro	0-com companheiro(a)
	Casado	1-sem companheiro(a)
	Viúvo	
	Separado	
Caráter da cirurgia	Eletiva	0 (Eletiva/agendada)
	Agendada	1 (urgência/emergência)
	Urgência	
	Emergência	
Convênio	SUS	0-Não SUS
	IPE	1-SUS
	IPE PAMES	
	UNIMED	
	
Médico responsável	Privado	
	Fulano	AAY
	Ciclano	ABX
	Beltrano	AZW

Outra questão fundamental é a definição do tratamento para dados faltantes ou perdidos (*missing data*). Cabe ao pesquisador tomar a decisão de excluir o caso devido à falta do dado, mantê-lo como “*missing*” para a utilização destes casos nas análises descritivas, mesmo sabendo que os mesmos não serão incluídos no processo de modelagem, ou imputar algum valor aos mesmos.

Observações faltantes são um problema freqüentemente encontrado em pesquisas com populações humanas. É importante que os dados faltantes não sejam considerados apenas como um problema de análise de dados, mas também um problema de delineamento da pesquisa e da análise dos resultados. Os procedimentos de atribuição

de valores a estas variáveis (imputação) devem ser decididos na etapa de planejamento da pesquisa, para que se minimize a quantidade de dados faltantes e se estabeleça o processo de substituição, obtendo variáveis que possam auxiliar nas estimativas adequadas. Como a ausência de dados é praticamente inevitável, uma etapa prudente, do ponto de vista do delineamento do estudo, é investigar preditores potenciais dessas variáveis faltantes e inclui-los ou utilizá-los nas análises.

Existem diferentes abordagens a serem consideradas (34-38) tais como:

- substituição de uma variável por outra de interpretação semelhante
- imputação de um valor estimado a partir de outras variáveis conhecidas do mesmo indivíduo
- imputação pelo valor da média ou mediana da variável faltante
- substituição aleatória por um valor possível de um conjunto de valores possíveis criados pelo pesquisador
- substituição pelo valor da variável de outro indivíduo com um conjunto de características semelhantes.

Esses métodos podem ser feitos manualmente, automaticamente ou através de combinações. Feito de forma apropriada, a imputação limita os vieses causados pelas não-respostas e assegura que os registros imputados tenham consistência interna. Um bom procedimento de imputação é automático, objetivo e eficiente.

Não existe uma decisão simples para a escolha de imputar dados ou perder sujeitos de uma pesquisa. O pesquisador deve analisar suas opções e, na fase de análise, certificar-se que a imputação ou o abandono dos casos incompletos não agregou vieses relevantes aos resultados obtidos. De qualquer forma, a recomendação é a de se manter no banco de dados duas variáveis distintas, uma com os valores originais e a outra com os valores imputados adicionados. Embora a imputação possa melhorar a qualidade dos dados finais, deve-se ter cuidado para escolher a metodologia mais adequada, pois alguns métodos não preservam a relação entre as variáveis e podem até distorcer essas relações.

3- CONSTRUÇÃO DO MODELO ESTATÍSTICO

Outro desafio para o pesquisador, depois da garantia da acurácia e da confiabilidade dos dados, é a modelagem das variáveis potencialmente preditoras elaboradas e disponíveis. Em geral, quando coletamos dados para definir os fatores de risco associados a determinado desfecho, obtemos mais variáveis do que é possível incluir no modelo e precisamos desenvolver estratégias para selecionar aquelas variáveis que

resultam no “melhor” modelo dentro do contexto científico do problema. O sucesso da modelagem de um conjunto de dados complexos exige conhecimento científico, metodologia estatística, experiência e bom senso, caracterizando-se como um misto de arte e ciência.

A seleção das variáveis é a etapa mais crítica deste processo de busca do modelo mais parcimonioso possível que explique os achados. Quanto menor o número de variáveis incluídas, mais estável e mais facilmente generalizável será o modelo. Mas, por outro lado, não podemos deixar de considerar todas as variáveis que sejam biológica ou clinicamente relevantes para o fenômeno em estudo, não deixando de observar a relação recomendada entre o tamanho das amostras e o número de variáveis incluídas (30-32).

O ajuste por risco é complexo e requer análises estatísticas também complexas. Dos vários métodos disponíveis, o mais comumente utilizado para desfechos dicotômicos como a mortalidade cirúrgica, é a Regressão Logística (RL). A RL fornece uma probabilidade de desfecho positivo estimada para cada paciente bem como o desfecho esperado para uma determinada população, que é a soma das probabilidades individuais. O objetivo principal de um processo de modelagem estatística é o de encontrar o modelo mais adequado e parcimonioso, mas biologicamente razoável, para descrever a relação entre um desfecho (variável dependente ou resposta) e um conjunto de variáveis independentes (preditoras ou explanatórias). É particularmente apropriado para modelos envolvendo estados de saúde (doença/saúde, óbito/sobrevida), para tomada de decisão (sim/não) e portanto, largamente utilizada na área clínica (30).

Para fazer a validação do modelo gerado, nesta etapa de modelagem se trabalha apenas com o conjunto de casos da amostra de derivação. A modelagem estatística é feita em três etapas: análise univariável (com apenas uma variável explicativa), análise multivariável (com várias variáveis explicativas) e avaliação do ajuste do modelo obtido. Vejamos cada uma delas:

3.1- Análise univariável

Cada variável candidata a entrar no modelo deve ser analisada cuidadosamente antes da modelagem, através de análises de frequência, tabelas de contingência e métodos descritivos, buscando-se analisar seu comportamento e sua relação com o desfecho isoladamente. Esta relação pode ser avaliada pelo teste do χ^2 ou análise de RL em variáveis nominais e ordinais e pelo teste-*t* ou análise de RL univariável em variáveis quantitativas. Particular atenção deve ser dada a qualquer tabela de contingência com células vazias. Incluir esta variável no modelo causará resultados numéricos indesejados

e a melhor estratégia para resolver este problema é o agrupamento desta variável em um número menor de categorias, se este agrupamento tiver uma base clínica ou biológica.

As variáveis também devem ser modeladas por RL como preditor isolado, para avaliação de sua contribuição individual ao modelo e para posterior comparação com seu desempenho no modelo multivariável, para identificar potenciais confundidores.

De acordo com os critérios definidos por Hosmer e Lemeshow (39), qualquer variável em cuja análise univariável tenha-se obtido um valor $p < 0,25$ é candidata para o modelo multivariável além de outras com significância clínica conhecida. Mickey e Greenland (40) demonstraram que o uso do nível de significância tradicional de 0,05 freqüentemente elimina variáveis que teriam uma significância adequada no modelo multivariável. O uso de limiar tão alto (0,25) pode provocar a inclusão de variáveis sem importância, mas estas serão descartadas nos passos posteriores.

O valor de $-2\log$ Verossimilhança, também chamado de função-desvio ou *deviance*, é uma medida de quão bem o modelo se ajusta aos dados e quanto menor seu valor, melhor o ajuste. O valor da função-desvio testa a hipótese nula de que os coeficientes dos termos removidos/incluídos do modelo são iguais a zero (41).

Na Tabela 2 podemos ver o desempenho das diferentes variáveis obtidas para ajuste de risco de mortalidade cirúrgica na análise univariável (tabela de contingência (χ^2) e RL (deviance)), com indicação das variáveis já eliminadas nesta etapa e o ordenamento para a modelagem multivariável. A seleção das variáveis para a etapa multivariável foi feita por critérios baseados no quadro conceitual e pelo valor de p ($< 0,25$) sendo seu ordenamento realizado pelos valores de $-2\log$ Verossimilhança ($-2\log L$).

Tabela 2: Avaliação das variáveis para a etapa de modelagem estatística

VARIÁVEL	Tipo	χ^2 (p)	-2log L	Ordenamento	Observações
SEXO	Binária 0=fem, 1=masc	<0,001	318,072	3	
ASA	Categórica, I-II-III-IV-V	<0,001	216,057	-	2 caselas com 0: óbito com ASA I e não óbito em ASA V
ASA	Categórica, I e II-III-IV e V	<0,001	230,594	-	IC95% muito grandes
ASA	Dicotômica:0=IouII,1=>II	<0,001	279,199	1	
IDADE	Contínua	<0,001	308,472	-	-2logL semelhante à idade dicotômica
IDADE >70 anos	Dicotômica (0=não)	<0,001	309,174	2	
IDADE >60 anos	Dicotômica (0=não)	<0,001	313,423	-	-2log Verossimilhança maior que anterior
IDADE até 40, 41-70, +70		<0,001	307,445	-	Categoria "41-70 anos" não significativa
Cor Branca	Dicotômica (0=sim)	0,894		-	P >0,25
SUS	Dicotômica (0=não)	0,974		-	p >0,25
Companheiro	Dicotômica (0=sim)	0,198	332,752	9	
Escolaridade	7 níveis	0,116	322,209	-	Caselas em branco
1º Grau completo	Dicotômica (0=sim)	0,097	333,752	10	
Procedência	1=Porto Alegre, 2=outra	0,406		-	p >0,25
Internação eletiva	0=sim	<0,001	318,960	4	
Cirurgia eletiva	0=sim	0,006	326,876	6	
Clínica cirúrgica	0=sim	0,003	326,904	7	
Turno	0=dia, 1=noite	0,088	331,758	8	
Anestesia geral	0=não	0,008	325,569	5	

3.2- Análise multivariável – Regressão Logística

Modelagem: A abordagem tradicional na modelagem estatística envolve a busca do modelo mais parcimonioso que explique os dados. Minimizando o número de variáveis, o modelo resultante terá mais probabilidade de ser estável numericamente e mais generalizável. No entanto, quanto mais variáveis forem incluídas, maiores se tornam os EP e mais dependente este se torna dos dados observados. Epidemiologistas sugerem incluir todas as variáveis clínicas e intuitivamente relevantes no modelo, desconsiderando suas significâncias estatísticas, pois desta forma teremos um completo controle dos confundimentos.

Várias abordagens são possíveis através do uso de opções dos programas estatísticos (39). A primeira delas é introduzir ou retirar as variáveis de forma “manual”, começando com a mais relevante e incluindo as seguintes ou, se a amostra é suficientemente grande (pelo menos 10 casos com o desfecho menos freqüente por variável), pode-se começar a modelagem com todas as variáveis selecionadas na etapa univariável, retirando-se a cada etapa aquelas não significativas. A cada momento todas as variáveis são reavaliadas, podendo entrar novamente no modelo.

A segunda abordagem é o método *stepwise*, no qual as variáveis são selecionadas automaticamente, sem o controle do pesquisador, tanto para inclusão quanto para exclusão de uma maneira seqüencial baseada apenas em critérios estatísticos. Este método tem duas modalidades: seleção *forward* e seleção *backward*. Na modalidade *forward* as variáveis são adicionadas ao modelo a cada etapa, sendo a primeira variável candidata aquela mais significativa. Candidata-se então a variável seguinte e a cada etapa são retiradas aquelas com valor- $p > 0,05$. Na modalidade *backward*, entram todas as variáveis selecionadas pelo pesquisador ou aquelas disponíveis e a seleção é feita pela retirada da variável menos significativa a cada vez. Esta abordagem é útil e interessante para o pesquisador porque constrói modelos seqüencialmente e permite o exame de uma coleção de modelos que talvez não fossem montados pelo pesquisador.

Uma terceira abordagem é a “*best subset selection*”, um método de seleção pouco utilizado em RL. Um número de modelos contendo uma, duas, três e mais variáveis são examinados para determinar qual é o melhor modelo de acordo com critérios específicos.

Os dois últimos processos de seleção, mecânicos, têm sido criticados porque podem produzir um modelo biologicamente implausível ao selecionar variáveis irrelevantes e devem ser usados com cautela (39). O problema maior não é o fato de tais modelos poderem ser gerados mas sim que o analista falhe em avaliar o modelo cuidadosamente.

A grande disponibilidade destes métodos e a facilidade de seu uso podem reduzir o pesquisador a um mero assessor do computador na seleção dos modelos e por este motivo, apenas quando o analista entende as potencialidades e as limitações destes métodos é que estes podem ser ferramentas úteis no processo de modelagem, especialmente de grandes bases de dados com múltiplas variáveis e grande volume de registros.

Sugere-se que o pesquisador faça o processo de inclusão e exclusão de variáveis através de um método seqüencial manual pois este processo, embora mais laborioso, é útil e intuitivamente interessante pois permite ao pesquisador examinar uma coleção de modelos diferentes, entendendo as razões para a permanência ou a saída de cada variável.

Cada nova variável incluída deve ter sua adequação e importância no modelo verificadas através de sua significância estatística ($p < 0,05$) e pela comparação do valor de $-2\log L$ do novo modelo e do modelo anterior. Variáveis não significativas ou que não contribuem para o refinamento do modelo devem ser eliminadas. Este processo de excluir, recolocar, incluir e verificar continua até que todas as variáveis relevantes tenham sido incluídas no modelo e aquelas excluídas sejam clínica e/ou estatisticamente irrelevantes. Neste ponto, variáveis conceitualmente relevantes, não incluídas inicialmente ou excluídas durante o processo por critérios estatísticos, podem ser incluídas e verificadas, pois poderão ter um comportamento diferente no novo conjunto. O critério geralmente usado para permanência de uma variável no modelo final é $p < 0,05$ ou sua grande relevância clínica ou biológica.

A estatística de Hosmer e Lemeshow (H&L) que mede a adequação do modelo (calibração ou *goodness-of-fit*) deve ser avaliada nesta fase também (39).

A Tabela 3 mostra as diferentes etapas da modelagem dos dados em análise, que iniciou com as duas variáveis mais relevantes às quais se adicionou as demais, em ordem de classificação, excluindo aquelas não significativas estatisticamente.

De acordo com os critérios de significância estatística ($p < 0,05$), de adequação do modelo ($-2\log$ Verossimilhança) e estatística de H&L, dois modelos preliminares foram selecionados, contendo as variáveis ASA (classificação pré-anestésia da American Society of Anesthesiologists), idade contínua ou dicotômica (70 anos), sexo e anestesia.

Tabela 3: Etapas da modelagem: avaliação de diferentes combinações de variáveis

Variáveis	-2log L	Significância estatística (p)	Estatística de Hosmer & Lemeshow	OBSERVAÇÕES
ASA	279,199	<0,001		
ASA + IDADE 70 anos	260,861	<0,05	0,504	
ASA + IDADE + SEXO	253,503	<0,05	0,659	
ASA+IDADE+SEXO+INT. URG	249,601	Int. Urg=0,057	0,399	Não significativa
ASA+IDADE+SEXO+CIR URG	247,311	URG=0,676		Não significativa
ASA+IDADE+SEXO+ANEST+ TURNO	247,484	Turno=0,972		Não significativa
ASA+IDADE+SEXO+ANEST+ INT.URG	244,221	Int.Urg=0,081		Não significativa
ASA+IDADE+SEXO+ANEST. GERAL	247,485	<0,05	0,784	Melhor modelo: boa calibração (H&L alto), significância estatística
ASA+IDADE contínua +SEXO+ANEST GERAL	244,182	<0,05	0,628	Modelo correto, menor valor de -2logL, calibração inferior ao modelo anterior

Linearidade: Tendo obtido o modelo preliminar, deve-se assegurar que as variáveis quantitativas incluídas preencham os critérios de linearidade. A avaliação do critério é feita graficamente. Esta análise gráfica é complexa mas existem processos já implementados em softwares estatísticos, através dos quais se transforma variáveis quantitativas em categóricas, com o grupo mais baixo como valor de referência. Na maior parte das vezes a mera análise visual do pesquisador (“eyeball”) é suficiente para avaliar a forma da curva. Esta verificação pode ser feita graficamente ou através da criação de categorias para esta variável discreta e análise de seus coeficientes.

Interações: Depois de selecionadas as variáveis e de garantir que as variáveis quantitativas estão categorizadas apropriadamente, deve-se verificar as interações relevantes. Uma interação existe quando uma covariável modifica o efeito de um fator de risco, sendo chamada de “modificador de efeito”. Em qualquer modelo, uma interação entre duas variáveis implica que o efeito de uma variável não é constante para todos os níveis da outra. Quando se suspeita de uma interação entre variáveis, deve-se incluir no modelo um termo de interação entre X1 e X2 em uma nova variável X3, a qual é o produto das variáveis X1 e X2 ($X3=X1 \times X2$). A necessidade de incluir termos de interação no modelo é verificada usando o teste da razão de verossimilhança (*log likelihood ratio*) para verificar sua significância, isto é, sua contribuição para a melhor adequação do modelo.

A decisão de inclusão ou não de um termo de interação deve ser baseada em considerações estatísticas e práticas, além de plausibilidade biológica ou clínica. O objetivo é verificar se cada interação melhora o modelo já construído. Para isso, se acrescenta uma interação de cada vez no modelo e se verifica se esta inclusão altera as

estimativas prévias. Deve-se listar todas as interações clínica e biologicamente plausíveis e incluir os termos de interação no modelo, verificando sua contribuição e desempenho. Se a variável de interação melhora o desempenho do modelo ela deve permanecer.

No exemplo utilizado foram testadas no modelo, um a um, os termos de interação IDADEXASA, IDADEXSEXO e ASAXSEXO. Os três termos não melhoraram o ajuste do modelo e não obtiveram significância estatística ($p=0,115$, $0,604$ e $0,507$ respectivamente) não tendo portanto permanecido.

Esta etapa de modelagem é considerada concluída quando permanecem no modelo somente variáveis com significância estatística ($p<0,05$) e variáveis com relevância biológica. A tabela 4 mostra o modelo final obtido no exemplo em uso.

Tabela 4: Modelo final – mortalidade cirúrgica HCPA

VARIÁVEL	Coefficiente β	p	OR	IC 95%	
ASA	2,631	<0,001	13,814	6,173	31,225
IDADE	1,404	<0,001	4,073	2,038	8,139
SEXO	0,998	0,006	2,714	1,327	5,549
ANESTESIA	1,425	0,034	4,159	1,111	15,566
<i>Constante</i>	<i>-4,573</i>				

3.3- Avaliação da consistência do modelo

Para afirmarmos que o modelo está adequado devemos entender o que isto significa. Quando o estágio de construção do modelo está completo, uma série de passos lógicos podem ser tomados para verificar sua adequação aos dados (ajuste).

Inicialmente devemos fazer uma avaliação empírica: o modelo corresponde a um modelo conceitual predeterminado? Está de acordo com a literatura e com as evidências existentes? Se não estiver, pode ser por limitações do banco de dados, erro de modelagem, restrições no número ou seleção e transformação inadequada de variáveis, não inclusão de termos de interação, além da qualidade dos dados, variação aleatória e variação no processo assistencial, incluindo sua qualidade.

Para avaliar o modelo de regressão logística, devemos avaliar a significância estatística de cada variável (valor de p da estatística de Wald), razão de chances (OR) e seus intervalos de confiança (95%), valores dos coeficientes β , a plausibilidade do conjunto de variáveis e interpretação dos coeficientes e respectivas razões de chances (OR) de acordo com o corpo teórico e a adequação do tamanho da amostra em relação ao número de parâmetros do modelo. Para haver a plausibilidade biológica, clínica e/ou epidemiológica do modelo, o conjunto de variáveis deve ser capaz de explicar a ocorrência do desfecho de acordo com modelo conceitual aceito. Havendo discrepâncias, o processo de modelagem deve ser revisto. Os valores dos coeficientes e de razões de

chance também devem estar de acordo com o conhecimento atual. O tamanho da amostra deve ser de pelo menos 10 desfechos menos frequente/variável ou 25 casos/variável. Para haver significância estatística, os valores p de todas as variáveis do modelo final devem ser $<0,05$ e os intervalos de 95% de confiança das razões de chance não devem incluir o valor 1, com exceção das variáveis biologicamente importantes que são fundamentais para o ajuste de outras variáveis do modelo.

No exemplo da Tabela 4 podemos entender que o conjunto de variáveis preditoras é clinicamente plausível, pois a idade avançada, a classificação de pior estado físico e o uso de anestesia geral podem ser facilmente entendidos como fatores de risco para óbito pós-cirúrgico. O fator protetor apresentado pelo sexo feminino deve ser melhor estudado para que possa ser corretamente interpretado. Todas as variáveis são significativas estatisticamente e os valores dos coeficientes correspondem à importância de cada variável, sendo maior para a classificação ASA e menor para o sexo.

Outra avaliação interessante é a análise da lista de casos mal-classificados, gerada pelos programas estatísticos a partir dos resíduos da RL (*casewise list*), buscando entender as razões pelas quais o modelo não classificou corretamente aqueles casos. Muitas vezes, pela análise de cada caso (voltando ao prontuário ou à fonte primária se possível) permite identificar fatores de risco que não foram incluídos ou foram mal categorizados ou ponderados na fase de preparação dos dados.

4- AVALIAÇÃO DA ADEQUAÇÃO DO MODELO

Após estes passos teremos um modelo multivariável preliminar, que deverá ser verificado em sua adequação ou ajuste, conforme veremos nesta etapa. Quando o processo de modelagem está concluído, uma série de passos podem ser dados para esta avaliação. Medidas de adequação são rotineiramente fornecidas pelos programas estatísticos e devem ser analisadas com cuidado. Com o advento de ferramentas poderosas disponíveis em programas estatísticos de uso aparentemente fácil (SPSS, STATA, outros), pesquisadores passaram a tratar os computadores como “caixas-pretas” que ingerem dados e devolvem respostas corretas que não necessitam de avaliações ou julgamentos, o que não corresponde à realidade pois estatísticas, por sua natureza, não fornecem todas as informações necessárias sobre a qualidade do modelo e além disto necessitam ser utilizadas com algum conhecimento técnico.

Segundo Harrel (27), a acurácia preditiva de um modelo tem dois componentes. O valor preditivo do modelo, isto é, o grau pelo qual a equação de RL acuradamente prevê o desfecho de cada paciente, é avaliado através das medidas de calibração e de discriminação destes modelos:

a) Confiabilidade ou calibração: refere-se à quantidade de concordância entre desfechos preditos e observados e pode ser definida como a habilidade do modelo de atribuir riscos apropriados aos pacientes. A calibração também pode ser demonstrada quantitativamente pelo teste de adequação (*goodness of fit*) proposto por Hosmer e Lemeshow (39). Este teste fornece uma estatística χ^2 modificada (estatística de Hosmer e Lemeshow), onde um valor menor é desejado, buscando obter um valor p não significativo para as diferenças entre os desfechos esperados e observados em cada classe de risco.

A estatística de H&L é obtida através do cálculo do χ^2 de uma tabela de contingência $g \times 2$ de frequências esperadas e observadas, onde g é o número de grupos formados pela distribuição de probabilidades estimadas e o 2 se refere aos grupos de desfecho e não-desfecho. A distribuição nos g grupos é feita por probabilidade de desfecho, ficando no grupo 1 aqueles casos de menor probabilidade e no grupo g, os de maior probabilidade. O resultado esperado deve indicar que a diferença não é significativa e o valor-p correspondente deve ter um valor alto. A estatística de H&L é baseada no ordenamento e agrupamento de casos de acordo com as probabilidades estimadas, usualmente com $g=10$, sendo estes grupos chamados de “decis” de risco.

A vantagem desta estatística é a de prover um valor interpretável único que pode ser utilizado para avaliar o modelo. No exemplo em uso o valor obtido para a estatística de H&L foi de 2,451 que, para $g=5$, resultou em um valor-p=0,784, que indica boa calibração ou adequação do modelo aos dados.

b) Discriminação: Uma segunda propriedade é relacionada à acurácia do modelo e é denominada de “poder de discriminação”, definida como a habilidade do modelo de distinguir entre os pacientes que apresentam o desfecho daqueles que não apresentam.

A discriminação é avaliada pelo Índice-C (de Concordância), proposto por Harrel e colegas (42) e seu valor é equivalente à área sob a curva ROC (*Receiver Operating Characteristics*). Esta curva, originária da teoria de detecção de sinal, mostra como o receptor opera a existência de sinal na presença de ruído e plota a probabilidade de detectar um sinal verdadeiro (sensibilidade) e um sinal falso (1-especificidade) para um conjunto de possíveis pontos de corte. A área sob a curva ROC, que varia de 0 a 1, fornece uma medida da habilidade do modelo de discriminar entre aqueles sujeitos que apresentam e os que não apresentam o desfecho de interesse.

A curva é gerada a partir dos valores de probabilidade predita pelo modelo e a área sob a curva (Índice-C) é calculada.

Outro meio mais intuitivo de entender o significado da área sob a curva ROC é o seguinte (39): Consideremos que n_1 é o número de sujeitos com $y=1$ e n_0 , de sujeitos com $y=0$. Então criamos $n_1 \times n_0$ pares: cada sujeito com $y=1$ (desfecho) é pareado com todos os outros com $y=0$ (não desfecho). Destes $n_1 \times n_0$ pares, determinamos a proporção de vezes em que o sujeito com $y=1$ teve a maior das duas probabilidades. Esta proporção é equivalente à área sob a curva ROC.

Por exemplo, na base de dados temos 423 pacientes dos quais 57 foram a óbito após a cirurgia (13,5%). Um total de $(57 \times 366) = 20862$ comparações são feitas e contamos quantas vezes a probabilidade de morrer predita pelo modelo é maior para o sujeito que morreu do que para aquele que não morreu (quando as probabilidades são iguais, se soma $\frac{1}{2}$). Para estes dados, a soma de vezes em que isso aconteceu foi de 17.170. A razão $17.170 / 20.862 = 0,823$, que é igual à área sob a curva ROC obtida.

Um índice-C de 0,5 significa que a probabilidade de identificar corretamente um desfecho em um par escolhido aleatoriamente será de 50%, a mesma que ocorreria devido ao acaso. Uma discriminação perfeita geraria um índice-C de 100% de acurácia. Como regra geral, o resultado é avaliado da seguinte maneira (39):

ROC=0.5:	Não há discriminação, resultado obtido ao acaso
$0.5 \leq \text{ROC} < 0.7$:	Resultado não aceitável, pequena relevância
$0.7 \leq \text{ROC} < 0.8$:	Discriminação aceitável
$0.8 \leq \text{ROC} < 0.9$:	Discriminação excelente
ROC ≥ 0.9 :	Discriminação destacada, quase perfeita

Na prática, modelos típicos de ajuste apresentam índice-C entre 0,7 e 0,8 e é extremamente raro observar valores acima de 0,9. É importante observar que um modelo pode ser bem calibrado e ter uma discriminação inadequada ou vice-versa.

Na Figura 2 observamos a curva ROC gerada a partir do modelo descrito na Tabela 5. A curva tem uma área de 0,823, que indica uma discriminação excelente, de acordo com os critérios encontrados na literatura.

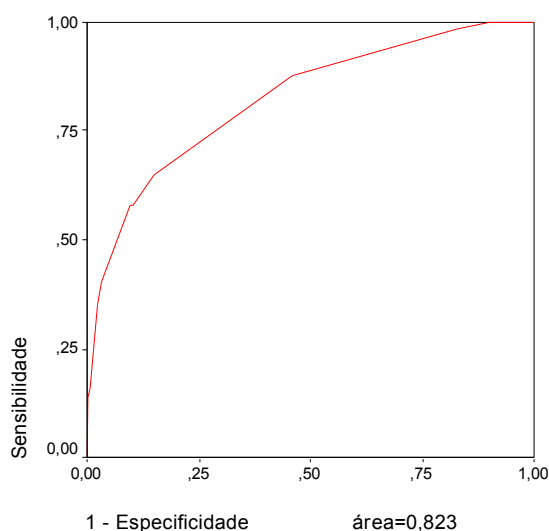


Figura 2: Curva ROC – modelo de risco para mortalidade cirúrgica no HCPA

Outra maneira bastante intuitiva de avaliar a adequação de um modelo de RL é através da análise de sua tabela de classificação, gerada pelos programas estatísticos. Esta tabela é o resultado da classificação da variável de desfecho y com uma variável dicotômica cujo valor é derivado das probabilidades logísticas estimadas (39). Esta tabela é mais apropriada quando a classificação é o objetivo da análise, senão ela apenas deverá complementar métodos de avaliação mais rigorosos.

A tabela 5 mostra que o modelo utilizado para exemplificar os conceitos expostos apresenta uma sensibilidade baixa (31,6%), uma alta especificidade (97,3%) e uma boa acurácia geral (88,4%).

Tabela 5: Tabela de classificação

		Obito predito		% Correta
		Não	Sim	
Obito observado	Não	356	10	97,3
	Sim	39	18	31,6
				88,4

Atenção particular deve ser dada às questões de interpretação. Os coeficientes são clinicamente interpretáveis? A abordagem que seleciona o modelo de melhor ajuste sem considerar as implicações clínicas não é o mais recomendado. Em algumas situações, a inadequação de um modelo pode ser corrigida retornando à etapa de modelagem e verificando novamente a seleção de variáveis e suas escalas. A adequação do modelo é

um processo iterativo. Raramente obtemos um modelo final adequado na primeira passagem pelos dados. No entanto, devemos ter em mente a distinção entre obter um modelo ajustado ou um modelo teoricamente correto.

Quando realizamos uma análise, esperamos que a pesquisa tenha sido planejada cuidadosamente e que todas as potenciais variáveis associadas ao desfecho tenham sido identificadas e coletadas. No entanto, é possível que os fatores clínicos associados com o desfecho não sejam bem conhecidos e, neste caso, uma variável importante não tenha sido coletada. Os vieses potenciais neste caso são enormes. Pouco pode ser feito neste caso, exceto voltar atrás e coletar este dado. Esta abordagem de coleta retroativa de dados, no entanto, é impraticável na maior parte das pesquisas.

5- CRIAÇÃO DO ESCORE DE RISCO

A partir dos valores obtidos para os diferentes parâmetros da regressão logística (razão de chances ou coeficientes β), é possível atribuir pesos para ponderar a gravidade para cada estrato (variável e suas categorias), em número proporcional à “quantidade” de risco que cada paciente apresenta, permitindo a criação de um “escore de risco”. Diferentes abordagens já foram utilizadas por vários autores, todas com os mesmos princípios básicos: atribui-se a cada fator de risco ou a diferentes categorias de um fator, um número de pontos proporcional à importância relativa desta variável no modelo, a partir dos valores dos coeficientes β ou das razões de chance (OR).

Várias metodologias são descritas para esta etapa (5-8), considerando a utilização dos valores do coeficiente β ou das razões de chance (OR) para calcular ou estimar o número de pontos a serem atribuídos a cada fator.

De acordo com o método proposto por Le Gall e colegas (5) e usando o modelo descrito na Tabela 5, poderiam ser atribuídos pontos a cada fator de risco de acordo com os valores dos coeficientes β , multiplicando estes coeficientes por 10 e arredondando para o inteiro mais próximo (5), o que é exemplificado na Tabela 6.

Tabela 6: Ponderação dos fatores de risco a partir dos coeficientes obtidos

Variável	β	Pontos
ASA (classe IV ou V)	2,631	26
IDADE (>70 anos)	1,404	14
ANESTESIA (Geral)	1,425	14
SEXO (Masculino)	0,998	10

A partir destes pontos, é calculado o escore de risco (ER) para cada paciente, de acordo com os fatores de risco presentes, da seguinte forma:

$$ER= 26(ASA) + 14(IDADE) + 14(ANESTESIA) + 10(SEXO)$$

Estes pontos, de acordo com os fatores de risco apresentados por cada paciente, devem gerar um escore de risco cujo valor, neste exemplo, pode variar entre 0 e 65 pontos.

6-VALIDAÇÃO DO MODELO

Validação pode ser definida como sendo a demonstração de que a acurácia preditiva do modelo é similar quando este é aplicado a uma amostra de pacientes diferente daquela usada na construção do modelo. Sugere-se para isto a criação de dois conjuntos de dados, um para o desenvolvimento do modelo e outro para sua validação ou então coletar novos dados depois do modelo pronto. Um meio comum de se fazer esta validação em grandes bancos de dados é separar os pacientes aleatoriamente ou por outros critérios, criando uma amostra de derivação (aprendizado) e outra de validação (teste). O modelo é desenvolvido no conjunto de aprendizado e os resultados são aplicados à amostra de validação. As medidas de calibração e discriminação são avaliadas neste segundo conjunto e se o modelo estiver bem ajustado, o mesmo é dito validado (39;43). Este processo de validação é muito importante quando se quer utilizar o modelo para prever desfechos para futuros pacientes.

Quando o modelo de regressão logística é formulado adequadamente e sua acurácia preditiva é verificada na amostra de validação, este modelo pode ser usado como um sumário conciso das relações entre os preditores e o desfecho. A partir daí, estimativas de desfecho podem ser obtidas para pacientes individuais e para ajustar indicadores para fins de comparação de performance entre hospitais.

No exemplo em uso, os pesos e os escores foram atribuídos e calculados para as duas amostras de pacientes. O escore foi utilizado como variável preditora única em novo modelo de regressão logística com o objetivo de obter as probabilidades de óbito de cada paciente. Na Tabela 7 vemos os valores de calibração, discriminação e acurácia do modelo nas duas amostras, concluindo-se que foi possível repetir o desempenho do modelo obtido na amostra de derivação com a amostra de validação.

Tabela 7: Desempenho do escore de risco nos dois amostras de pacientes

GRUPO	Calibração (H&L)	Discriminação (Índice-C) (curva ROC)	Acurácia
Derivação	0,784	0,823	89,4
Validação	0,866	0,847	87,1

7-APLICAÇÃO DO ESCORE DE RISCO PARA COMPARAÇÃO DE DESEMPENHO

Para se comparar diferentes desempenhos (entre hospitais, equipes, estados, países, períodos de tempo etc.) é necessário calcular os valores esperados (E) de desfecho para todas as amostras, utilizando a variável “ESCORE” em outro modelo de RL executado para se obter a probabilidade predita para cada paciente e compará-los com os valores observados. Para obter estes valores, somam-se as probabilidades calculadas (E) e os desfechos observados (O) de cada grupo a ser comparado. Estes valores podem ser comparados de diferentes maneiras, O-E ou O/E. Diferentes conclusões podem ser tiradas se compararmos de uma maneira ou de outra. A razão O/E é a mais freqüentemente encontrada na literatura. Sendo uma razão, pode variar de 0 ao infinito.

No exemplo aqui utilizado, exemplificamos hipoteticamente a comparação do desempenho entre as equipes cirúrgicas do HCPA, com dados imputados aleatoriamente para fins de demonstração, pois a informação da equipe não foi coletada. Foram calculadas as probabilidades de óbito de todos os pacientes da amostra de derivação a partir de um modelo de RL com a variável “ESCORE”. Após, foram somadas as probabilidades e os casos de óbitos dos pacientes de cada equipe cirúrgica; a seguir, foi calculada a razão O/E de cada equipe. Valores abaixo de 1 indicam que ocorreram menos óbitos do que estava previsto pelo modelo de ajuste de risco, indicando um bom desempenho da equipe. Valores acima de 1 indicariam a ocorrência de um número de óbitos maior do que o esperado, considerando-se o risco dos pacientes atendidos por aquela equipe. Desta forma, o desempenho relativo das equipes pode ser comparado através dos valores da razão O/E, conforme exemplo apresentado na tabela 8.

Tabela 8: Avaliação comparativa das equipes cirúrgicas (dados hipotéticos)

EQUIPE	O = Óbitos observados	E=óbitos estimados	O/E
A	4 (3,4%)	9,25	0,432
B	11 (12,5%)	11,74	0,937
C	26 (24,5%)	18,68	1,392
D	16 (14,2%)	17,33	0,924
Total	57 (13,5%)	57	1

A equipe A apresentou um desempenho bem melhor que o esperado, enquanto que a equipe C teve um número de óbitos 1,4 vezes maior que o previsto. A classificação “bruta” de desempenhos, feita pela taxa de óbitos de cada equipe, indicaria em ordem, as equipes A, B, D e C; a classificação ajustada pelo risco seria diferente: A,D, B e C. Da mesma forma seria possível comparar desempenhos de diferentes hospitais, regiões etc.

Quando se desenvolve um sistema de escore de risco para avaliações comparativas de taxas de desfechos cirúrgicos é importante que, além de acurado, este seja de fácil uso. As variáveis coletadas devem ser em pequeno número, estar disponíveis para todos os pacientes, ser de uso comum em diferentes hospitais e deixar pouco espaço para vieses de observação e de coleta. O escore deve ser fácil de ser calculado a partir dos dados coletados e deve ser possível integrar este sistema de escore em programas de auditoria pré-existentes, com um mínimo de intervenção. Sempre que adequado, o escore deve ser agrupado e convertido em um índice de risco, de 3 a 5 classes, para facilitar sua atribuição e interpretação, bem como limitar os estratos necessários em avaliações comparativas.

Para o escore desenvolvido para ajuste de risco de mortalidade cirúrgica aqui utilizado, são propostas três classes de risco, distribuídas de acordo com pontos de corte identificados na análise, como se vê na tabela 9.

Tabela 9: Criação de classes de risco a partir dos escores

Escore de risco	Fatores de risco presentes	nº Óbitos	Total	% óbitos	Classe de risco
0	ASA=I ou II + Sexo=feminino + anestesia≠geral + idade < 70	0	36	0	I
10	ASA=I ou II + Sexo=masc. + anestesia≠geral + idade < 70	1	28	3,6	I
14	ASA=I ou II + Idade > 70 ou anestesia=geral + Sexo=feminino	6	140	4,3	I
24	ASA=I ou II + Sexo=masc. + idade > 70 ou anestesia=geral	13	126	10,2	II
28	ASA=I ou II + Idade > 70 + Anestesia=geral	4	21	19,0	II
38	ASA=I ou II + Sexo=masc. + Idade > 70 + Anestesia=geral	10	33	30,3	III
41	ASA ≥ III + idade > 70 ou + anestesia=geral	3	7	42,9	III
51	ASA ≥ III + sexo=masc. + idade > 70 ou + anestesia = geral	11	16	68,8	IV
55	ASA ≥ III + idade > 70 + anestesia=geral + sexo=feminino	1	3	33,3	IV
65	ASA ≥ III + idade > 70 + sexo=masc. + anestesia=geral	8	9	88,9	IV

Estas classes fazem sentido tanto do ponto de vista estatístico como biológico, pois a classe I é composta de pacientes que não apresentam nenhum fator de risco ou apenas um fator não grave; a classe II é composta por pacientes que apresentam 2 fatores associados, nenhum destes grave; a classe III é composta por pacientes que apresentam 3 fatores de risco ou uma combinação de 2 fatores, sendo um deles grave; a classe IV

indica aqueles pacientes com 3 fatores de risco onde se inclui o mais grave ou com os 4 fatores juntos.

Graficamente podemos perceber na Figura 3 que as classes de risco estão discriminando adequadamente aqueles pacientes com maior probabilidade de óbito.

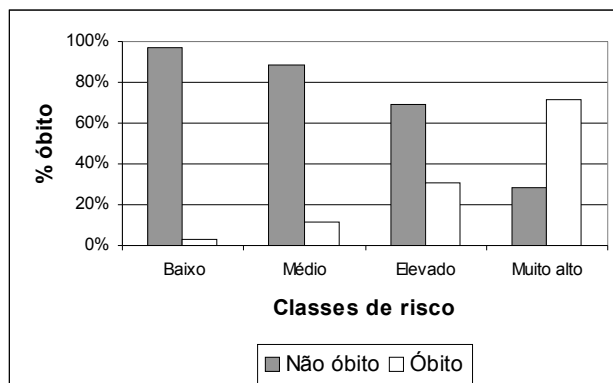


Figura 3: Percentuais de óbitos e não-óbitos em cada classe de risco

Para nos certificar da calibração desta distribuição de classes, podemos comparar os valores observados e esperados de cada classe, concluindo que a distribuição foi corretamente realizada (Tabela 10).

Tabela 10: valores observados e esperados em cada classe de risco

CLASSE	OBSERVADO	ESPERADO*	O/E
I	7	6,99	1,001
II	17	16,39	1,037
III	13	13,68	0,950
IV	20	19,94	1,003

*probabilidade calculada

Considerações finais

Nenhum método de ajuste de riscos é perfeito. Ajustar para todas as características do paciente não é nem necessário nem possível. No entanto, devem ser feitos esforços no sentido de identificar aqueles que sejam suficientemente válidos para os propósitos explícitos. A meta do ajuste de risco (*risk adjustment*) é considerar as características pertinentes do paciente ao estabelecer inferências sobre a efetividade ou a qualidade da assistência. Apesar de suas limitações, os modelos de ajuste de risco se constituem em ferramentas indispensáveis na gestão da qualidade assistencial.

A regressão logística é uma ferramenta poderosa para avaliar a importância relativa dos fatores determinantes de um desfecho. Desta forma, está sendo crescentemente utilizada na Medicina para desenvolvimento de algoritmos diagnósticos e avaliação de prognósticos. Mas esta ferramenta pode ser imperfeita e sujeita a erros e, portanto, seus

resultados devem ser interpretados mais como um “sinal amarelo” do que um veredicto sobre a qualidade ou falta de qualidade de um hospital ou equipe.

O entendimento básico da análise de regressão logística é o primeiro passo para a apreciação da utilidade e das limitações desta técnica. Bagley e colegas (30) sugerem vários critérios para avaliar a qualidade metodológica de um artigo científico que utilize regressão logística como ferramenta de análise, incluindo o número suficiente de casos por variável (10 casos por desfecho menos freqüente ou 25 casos por variável), a verificação da linearidade do efeito das variáveis quantitativas, testes para interações, avaliação de colinearidade, validação do modelo em outra amostra, significância estatística, avaliação da adequação do modelo (Estatística de H&L e Índice-C).

Além destes critérios, os autores também incluem a descrição do processo e critérios de escolha das variáveis preditoras (suas codificações e tratamentos) e do processo de modelagem, explicitando a ordem de entrada e critérios para entrada e saída das variáveis. De 21 artigos analisados, os autores não encontraram nenhum com registro de medidas de adequação do modelo e, na maioria destes, o número de variáveis por evento era insuficiente; nenhum dos estudos realizou uma análise de validação.

A pesquisa em serviços de saúde está sendo chamada de forma crescente a prover orientação metodológica para apoiar a qualidade da assistência na adoção de boas práticas. No meio em que vivemos, de recursos limitados e controle social cada vez mais presente (*accountability*), estudos de efetividade e de desfechos hospitalares podem otimizar custos e qualidade assistencial. Para atingir estas metas, não é suficiente simplesmente conhecer taxas de eventos adversos, mas sim entender porque tais eventos ocorrem. O propósito do ajuste de risco é o de isolar uma causa potencial. Medir riscos é o primeiro passo no uso de estudos e de pesquisa em serviços de saúde. O ajuste de risco é uma ferramenta que pode não ser perfeita, mas é necessária. Outros métodos e abordagens são necessários para se tomar o próximo passo na direção da melhora dos serviços de saúde, procurando o entendimento de como os desfechos podem ser melhor controlados(44).

Bibliografia

- (1) Nightingale F. Looking back. Taken from "Notes on Hospitals" by Florence Nightingale, 1863. *Lancet* 1979; 36(8):39-43.
- (2) Iezzoni LI. 100 apples divided by 15 red herrings: a cautionary tale from the mid-19th century on comparing hospital mortality rates. *Ann Intern Med* 1996; 124(12):1079-1085.
- (3) Graham N. Quality assurance in hospitals. Maryland, USA: Aspen Publishers, 1990.
- (4) DesHarnais S.I., McLaughlin C.P. The outcome model of quality. In: McLaughlin C.P., Kaluzny AD, editors. Continuous quality improvement in health care - Theory, implementation and applications. Maryland, USA: Aspen Publishers, 1999: 59-92.
- (5) Le Gall JR, Lemeshow S, Saulnier F. A new Simplified Acute Physiology Score (SAPS II) based on a European/North American multicenter study. *JAMA* 1993; 270(24):2957-2963.
- (6) Arozullah AM, Khuri SF, Henderson WG, Daley J. Development and validation of a multifactorial risk index for predicting postoperative pneumonia after major noncardiac surgery. *Ann Intern Med* 2001; 135(10):847-857.
- (7) Fortescue EB, Kahn K, Bates DW. Development and validation of a clinical prediction rule for major adverse outcomes in coronary bypass grafting. *Am J Cardiol* 2001; 88(11):1251-1258.
- (8) Polanczyk CA, Rohde LE, Philbin EA, Di Salvo TG. A new casemix adjustment index for hospital mortality among patients with congestive heart failure. *Med Care* 1998; 36(10):1489-1499.
- (9) Webster C, Neumayer L, Smout R, Horn S, Daley J, Henderson W et al. Prognostic models of abdominal wound dehiscence after laparotomy. *J Surg Res* 2003; 109(2):130-137.
- (10) Neary WD, Heather BP, Earnshaw JJ. The Physiological and Operative Severity Score for the enUmeration of Mortality and morbidity (POSSUM). *Br J Surg* 2003; 90(2):157-165.
- (11) Arenal JJ, Bengoechea-Beeby M. Mortality associated with emergency abdominal surgery in the elderly. *Can J Surg* 2003; 46(2):111-116.
- (12) Wouters SC, Noyez L, Verheugt FW, Brouwer RM. Preoperative prediction of early mortality and morbidity in coronary bypass surgery. *Cardiovasc Surg* 2002; 10(5):500-505.
- (13) Sutton R, Bann S, Brooks M, Sarin S. The Surgical Risk Scale as an improved tool for risk-adjusted analysis in comparative surgical audit. *Br J Surg* 2002; 89(6):763-768.
- (14) Edwards FH, Peterson ED, Coombs LP, DeLong ER, Jamieson WR, Shroyer ALW et al. Prediction of operative mortality after valve replacement surgery. *J Am Coll Cardiol* 2001; 37(3):885-892.
- (15) Aoki N, Wall MJ, Demsar J, Zupan B, Granchi T, Schreiber MA et al. Predictive model for survival at the conclusion of a damage control laparotomy. *Am J Surg* 2000; 180(6):540-544.

- (16) Cook TM, Day CJ. Hospital mortality after urgent and emergency laparotomy in patients aged 65 yr and over. Risk and prediction of risk using multiple logistic regression analysis. *Br J Anaesth* 1998; 80(6):776-781.
- (17) Goldman L, Caldera DL, Nussbaum SR, Southwick FS, Krogstad D, Murray B et al. Multifactorial index of cardiac risk in noncardiac surgical procedures. *N Engl J Med* 1977; 297(16):845-850.
- (18) Martins M, Travassos C, Carvalho dN. [Hospital Information Systems as risk adjustment in performance indicators]. *Rev Saude Publica* 2001; 35(2):185-192.
- (19) Silva LK. Validity of the risk adjustment approach to compare outcomes. *Cad Saude Publica* 2003; 19(1):287-295.
- (20) Daley J, Iezzoni LI, Shwartz M. Conceptual and practical issues in developing risk-adjusted methods. In: Iezzoni LI, editor. *Risk adjustment for measuring health care outcomes*. Chicago, Illinois: Health Administration Press, 2003: 179-205.
- (21) Daley J, Henderson WG, Khuri SF. Risk-adjusted surgical outcomes. *Annu Rev Med* 2001; 52:275-287.
- (22) Khuri SF, Daley J, Henderson W, Hur K, Gibbs JO, Barbour G et al. Risk adjustment of the postoperative mortality rate for the comparative assessment of the quality of surgical care: results of the National Veterans Affairs Surgical Risk Study. *J Am Coll Surg* 1997; 185(4):315-327.
- (23) Daley J, Khuri SF, Henderson W, Hur K, Gibbs JO, Barbour G et al. Risk adjustment of the postoperative morbidity rate for the comparative assessment of the quality of surgical care: results of the National Veterans Affairs Surgical Risk Study. *J Am Coll Surg* 1997; 185(4):328-340.
- (24) Greenhouse JB, Bromberg JA, Fromm D. An introduction to logistic regression with an application to the analysis of language recovery following a stroke. *J Commun Disord* 1995; 28(3):229-246.
- (25) Keats AS. The ASA classification of physical status--a recapitulation. *Anesthesiology* 1978; 49(4):233-236.
- (26) Daley J, Ash AS, Iezzoni LI. Validity and reliability of risk adjusters. In: Iezzoni LI, editor. *Risk adjustment for measuring health care outcomes*. Chicago, Illinois: Health Administration Press, 2003: 207-230.
- (27) Harrell FE, Jr., Lee KL, Matchar DB, Reichert TA. Regression models for prognostic prediction: advantages, problems, and suggested solutions. *Cancer Treat Rep* 1985; 69(10):1071-1077.
- (28) Klück MK, Fachel JMG. Desenvolvimento e validação de um modelo de ajuste de risco para mortalidade cirúrgica pós laparotomia exploradora. 2004. Ref Type: Unpublished Work
- (29) Normand ST, Glickman ME, Sharma RG, McNeil BJ. Using admission characteristics to predict short-term mortality from myocardial infarction in elderly patients. Results from the Cooperative Cardiovascular Project. *JAMA* 1996; 275(17):1322-1328.

- (30) Bagley SC, White H, Golomb BA. Logistic regression in the medical literature: standards for use and reporting, with particular attention to one medical domain. *J Clin Epidemiol* 2001; 54(10):979-985.
- (31) Harrell FE, Jr., Lee KL, Califf RM, Pryor DB, Rosati RA. Regression modelling strategies for improved prognostic prediction. *Stat Med* 1984; 3(2):143-152.
- (32) Iverson C. American Medical Association manual of style. Baltimore: Williams & Wilkins, 1998.
- (33) Hsieh FY. Sample size tables for logistic regression. *Stat Med* 1989; 8(7):795-802.
- (34) Data imputation for missing values. NC State University . 11-8-2004. Ref Type: Internet Communication
- (35) Data imputation. Legge JA, editor. www.statcan.ca/english/edu/power/ch3/imputation/imputation.htm . 13-9-2004. Statistics Canada. Ref Type: Electronic Citation
- (36) Davis P, Jenkin G, Coope P, Blakely T, Sporle A, Kiro C. The New Zealand Socio-economic Index of Occupational Status: methodological revision and imputation for missing data. *Aust N Z J Public Health* 2004; 28(2):113-119.
- (37) Heitjan DF. Annotation: what can be done about missing data? Approaches to imputation. *Am J Public Health* 1997; 87(4):548-550.
- (38) www.aifs.gov.au/growingup/pubs/dp3/imputation.pdf. Data imputation. Australian Government - Institute of Family Studies . 11-8-2004. Ref Type: Internet Communication
- (39) Hosmer DW, Lemeshow S. Applied logistic regression. 1st ed. New York, NY, USA: John Wiley & Sons, 1989.
- (40) Mickey RM, Greenland S. The impact of confounder selection criteria on effect estimation. *Am J Epidemiol* 1989; 129(1):125-137.
- (41) Menu de ajuda do SPSS. 2004.
- (42) Harrell FE, Jr., Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. *JAMA* 1982; 247(18):2543-2546.
- (43) Hosmer DW, Lemeshow S. Assessing the fit of the model. Applied Logistic Regression. John Wiley & Sons, 2000: 143-202.
- (44) Iezzoni LI. Risk adjustment for medical effectiveness research: an overview of conceptual and methodological considerations. *J Investig Med* 1995; 43(2):136-150.

**Artigo 2:
Desenvolvimento e validação de um escore de risco
multifatorial para mortalidade cirúrgica pós
laparotomia exploradora**

Desenvolvimento e validação de um escore de risco multifatorial para mortalidade cirúrgica pós laparotomia exploradora

Mariza M. Klück^(1,2,3) e Jandyra M. Guimarães Fachel^(1,4)

(5) Programa de Pós-Graduação em Epidemiologia da Universidade Federal do Rio Grande do Sul (UFRGS)

(6) Departamento de Medicina Social da Faculdade de Medicina da UFRGS

(7) Comissão de Prontuários de Pacientes do Hospital de Clínicas de Porto Alegre

(8) Departamento de Estatística do Instituto de Matemática da UFRGS

Resumo

Justificativa: A definição de modelos de risco para desfechos hospitalares é fundamental para a identificação precoce de pacientes com maior risco permitindo a prevenção ou minimização de seus efeitos bem como para o ajuste dos indicadores destes desfechos, para permitir a correta interpretação e análise comparativa dos mesmos.

Objetivos: desenvolver e validar um escore de risco para mortalidade cirúrgica pós laparotomia exploradora, avaliando a qualidade dos dados disponíveis nos prontuários de pacientes do HCPA.

Delineamento: estudo de coorte retrospectivo baseado em dados de prontuários de pacientes (eletrônico e papel).

Local: Hospital de Clínicas de Porto Alegre, hospital-escola da Universidade Federal do Rio Grande do Sul.

Pacientes: O escore de risco foi desenvolvido em uma coorte de 450 pacientes submetidos a laparotomia exploradora no período de fevereiro de 2000 a dezembro de 2002 e validado em uma coorte de 201 pacientes submetidos ao mesmo procedimento no período de janeiro a dezembro de 2003.

Medidas: Mortalidade cirúrgica, definida como o óbito de paciente durante o episódio de internação, ocorrido em um período de até 30 dias após a realização de laparotomia exploradora.

Resultados: A taxa de óbitos foi de 26,2% na coorte de desenvolvimento e de 24,9% na coorte de validação. O escore de risco foi desenvolvido através de regressão logística tendo o modelo final incluído as variáveis *classificação ASA* (I e II, III, IV e V), *níveis de albumina sérica* (até 2,2, 2,3-3,0 e >3,0 g/dL) e *idade* (<75 anos, >=75 anos). A partir deste escore os pacientes foram divididos em 5 classes de risco, que apresentaram taxa de mortalidade cirúrgica de 4,2%, 11,9%, 37,3%, 56,9% e 78,3% respectivamente ($p < 0,001$). A área sob a curva ROC (estatística-C = discriminabilidade) foi de 0,885 e 0,889 e o índice de Hosmer e Lemeshow (= calibração) de 0,96 e 0,91 para as coortes de derivação e validação.

Conclusões: Foi possível construir um índice de risco com boa calibração e discriminação, baseado em variáveis disponíveis no prontuário de pacientes (papel e eletrônico) do Hospital de Clínicas de Porto Alegre. O modelo final comporta apenas 3 variáveis, todas elas de coleta usual em hospitais brasileiros (ASA, albumina sérica e idade), o que torna sua aplicação muito fácil. O índice de risco para mortalidade cirúrgica identifica pacientes com maior probabilidade de desfecho e pode ser útil na prevenção destes desfechos e na comparação do indicador de mortalidade cirúrgica entre hospitais, ao permitir seu ajuste pelo risco prévio da população atendida.

Palavras-chave:

Mortalidade cirúrgica, ajuste de risco, indicadores de qualidade assistencial, segurança do paciente, modelos estatísticos, regressão logística.

Abstract

Background: The development of risk models for hospital procedures is essential for the identification of individual patient risk in order to reduce or eliminate negative outcomes. As well, patient population risk assessments can be used to adjust outcome indicators permitting more meaningful interpretation in terms of process quality and allowing more meaningful comparisons between hospitals serving different populations.

Objective: Design, develop, verify and validate a surgical mortality risk index.

Design: Retrospective cohort study based on review of electronic and hard copy patient records.

Setting: Hospital de Clínicas de Porto Alegre, the teaching hospital of the Universidade Federal do Rio Grande do Sul.

Patients: : The model was developed using data on 450 exploratory laparotomy patients attended between February, 2002 and January, 2003. The model was verified and validated using a population of 201 exploratory laparotomy patients attended between February, 2003 and January, 2004.

Measurements: Death within 30 days of surgery and within the same admission was considered surgical mortality

Results: In the development cohort, 118 patients (26.2%) died during or after exploratory laparotomy. Risk scores were determined using logistical regression on ASA (I and II, III, IV and V), level of serum albumin (≤ 2.2 , 2.3-3.0 and >3.0 g/dl) and age (<75 years, ≥ 75 years). Based on their scores, this cohort was divided into 3 risk groups: low, medium and high: that represent averages of 4.2%, 30.9% and 78.3% surgical mortality respectively. Using the index developed, risk of surgical death was calculated for each patient in the verification and validation cohort. The degree of association between calculated risk and actual outcome was determined using receiver operating characteristic (ROC) analysis. The areas under the ROC curves, a measure of discriminatory capacity, were 0.885 and 0.889 for the development and the verification and validation cohort, respectively. The Hosmer and Lemeshow index, a measure of model calibration, was 0.96 and 0.91 for the development and the verification and validation cohort, respectively.

Conclusions: It was possible to construct a well calibrated and discriminating surgical mortality risk index based on variable data available in electronic and hard copy patient records at the Hospital de Clínicas de Porto Alegre. Knowing the degree of risk of surgical mortality of a patient can be useful in applying interventions to reduce the probability of surgical death as an outcome. Further the index can be used to adjust surgical mortality statistics according to the degree of risk, thereby permitting valid comparison of surgical mortality, as a process quality indicator, between hospitals.

Key-words:

Pos-operative mortality, Risk Adjustment, quality-of-care indicators, patient safety, statistical models, logistic regression

INTRODUÇÃO

A avaliação da qualidade assistencial é feita, entre outros meios, a partir de medidas de desfecho (indicadores) que podem ser comparados com padrões definidos pela literatura, com resultados de outras instituições semelhantes ou com os indicadores do próprio hospital ao longo do tempo. Sistemas que permitem a comparação entre diferentes instituições são os mais amplamente utilizados, mas vários estudos demonstram que as medidas de desfechos, antes de serem comparadas, devem ser ajustadas pela gravidade dos casos ou por condições próprias dos pacientes, em um processo conhecido como “ajuste de risco” (*risk adjustment*). O ajuste de risco (ou padronização por severidade) visa minimizar a influência confundidora de potenciais fatores de risco inerentes ao paciente e/ou a sua patologia, permitindo assim a comparação da qualidade do tratamento hospitalar. É importante identificar fatores de risco adequados a este ajuste, que devem ser capazes de compor um modelo adequado em termos de acurácia, calibração e discriminabilidade (1).

Idealmente devemos usar fatores de risco específicos de cada condição e avaliações corretas devem levar em conta apenas as condições existentes no momento da internação do paciente, não utilizando fatores de risco possíveis de serem “iatrogenias” do processo assistencial. Por vezes é difícil separar, em registros de alta hospitalar, condições pré-existent das complicações ocorridas durante a internação. De qualquer forma, o uso de informações individuais disponíveis em bases de dados que apoiam sistemas de prontuários eletrônicos são fontes de dados suficientes para o desenvolvimento de *proxies* adequadas para fatores de risco. Através do uso de indicadores ajustados pode-se obter comparações entre hospitais ou entre diferentes períodos de tempo muito melhores, embora não perfeitas.

Vários sistemas de escores foram desenvolvidos e são utilizados para prognóstico de mortalidade e morbidade cirúrgica e para avaliação da qualidade assistencial: o sistema APACHE (*Acute and Chronic Health Evaluation*) (2-8), o escore POSSUM (*Physiological and Operative Severity Score for the enUmeration of Mortality and Morbidity*) (9-15) e o Sistema SAPS (*Simplified Acute Physiology Score*) (16;17), além da classificação ASA (*American Society of Anesthesiology*), que embora não tenha sido criada para prever desfechos cirúrgicos, é de fácil obtenção e de boa acurácia preditiva (18-24).

No entanto, o APACHE é de uso complexo e foi desenvolvido para avaliação de pacientes internados em Unidade de Tratamento Intensivo e consiste, além de dados de idade e doenças crônicas, de dados fisiológicos agudos não adequados para a população cirúrgica. O sistema POSSUM se mostrou útil para avaliações comparativas mas não

permite a avaliação de casos individuais além de requerer variáveis nem sempre disponíveis ou coletadas de rotina em pacientes cirúrgicos tais como RX de tórax, eletrocardiograma e vários exames bioquímicos. O sistema SAPS exige 17 variáveis para sua composição, o que dificulta sua aplicação em larga escala.

Quando se desenvolve um sistema de escore de risco para avaliações comparativas de taxas de desfecho e para suporte à decisão médica em casos individuais, é importante que, além de acurado, este seja de fácil utilização na beira do leito. As variáveis necessárias devem ser em pequeno número, estar disponíveis para todos os pacientes, ser de uso comum em diferentes hospitais e deixar pouco espaço para vieses de observação ou de coleta e o escore deve ser fácil de ser calculado (25). Além destas características, é importante que tais escores tenham validade para o ambiente onde serão aplicados, o que nem sempre é verdadeiro quando se utilizam sistemas desenvolvidos em outras realidades.

Vários estudos específicos têm sido publicados sobre escores de risco desenvolvidos em hospitais ou grupos de hospitais (26-28) sendo o mais importante destes o projeto desenvolvido pelo *National Surgical Quality Improvement Program* (NSQIP) do *Department of Veterans Affairs* dos Estados Unidos, que definiu um modelo de risco para mortalidade cirúrgica não cardíaca composto por 34 variáveis, a partir de uma base de dados de mais de 87.000 pacientes. Este sistema não tem validade externa porque foi desenvolvido para uma população restrita, soldados veteranos do exército americano, em geral homens (93%), idosos, pobres e doentes.(29-37), além de considerar todas as cirurgias em conjunto.

O Hospital de Clínicas de Porto Alegre (HCPA) realiza anualmente cerca de 8.000 internações com realização de cirurgia sob anestesia assistida, com uma taxa de mortalidade cirúrgica geral de 4% (38). Esta taxa é de 2,6% para cirurgias eletivas e de 10,6% para cirurgias de urgência. A cirurgia que apresenta a maior taxa de mortalidade (26%) é a laparotomia exploradora não ginecológica, que é definida como uma cirurgia onde o abdômen é aberto e explorado para fins diagnósticos ou terapêuticos (39).

O HCPA avalia seus indicadores de qualidade assistencial de maneira sistemática e, para fins de comparação externa, participa do Sistema de Indicadores Padronizados para Gestão Hospitalar (SIPAGEH) (40) e busca outros referenciais externos adequados tais como os dados do *Quality Indicator Project*, da Universidade de Maryland (41;42) e a base de dados de internações do Sistema Único de Saúde (43). A grande dificuldade neste processo de comparação é a falta de mecanismos que permitam o ajuste destes indicadores para a gravidade dos pacientes atendidos em cada hospital.

O HCPA se destaca no panorama nacional, entre outras características, por seu sistema clínico informatizado, que gera uma base de dados clínica acurada, completa e disponível que busca garantir a continuidade e a qualidade da assistência aos pacientes bem como apoiar a pesquisa e o ensino, gerar informações e sistemas de indicadores confiáveis e úteis e atender a outros fins legais e administrativos. Esta base contém dados clínicos completos de mais de 100.000 internações realizadas a partir de fevereiro de 2000 (44).

Portanto, visando obter subsídios para promover o ajuste de indicadores por risco prévio do paciente que permita identificar precocemente aqueles pacientes de alto risco bem como realizar avaliações comparativas de indicadores, este projeto de pesquisa teve como objetivo principal o desenvolvimento e a validação de um escore de risco cirúrgico simples e de fácil aplicação, adaptado para uso local, para prever mortalidade pós laparotomia exploradora. Este procedimento foi escolhido porque, segundo Daley e colegas (36), a mortalidade cirúrgica é uma medida apropriada da qualidade assistencial apenas se ocorre com frequência suficiente para permitir estimativas estatísticas estáveis e deve ser usada em procedimentos cirúrgicos onde são esperadas altas taxas de óbito. Foi nosso objetivo secundário avaliar a qualidade e a utilidade da base de dados clínica do Hospital para o desenvolvimento deste tipo de sistema.

MÉTODOS

Delineamento do estudo

Estudo de coorte retrospectivo. O desfecho considerado foi o óbito ocorrido durante a internação, num período de até 30 dias após a realização de cirurgia de laparotomia exploradora. Laparotomia (*lapara*= partes moles do corpo humano entre as costelas e os quadris; *tome*: cortar) exploradora é toda a cirurgia que abre e explora o abdome com fins diagnósticos e de tratamento (45). Quando mais de uma cirurgia foi realizada durante uma mesma internação, a primeira foi definida como sendo a cirurgia-índice.

Local de realização e População de estudo

O estudo foi realizado no Hospital de Clínicas de Porto Alegre (HCPA), hospital-escola da Universidade Federal do Rio Grande do Sul que conta com 735 leitos e realiza em torno de 25.000 internações por ano. Foram arrolados os pacientes internados no HCPA no período de fevereiro de 2000 a dezembro de 2003, submetidos a laparotomia exploradora por equipes de Cirurgia Geral, Proctologia e Urologia. A classificação do procedimento realizado foi verificada através da consulta ao formulário de registro

anestésico existente no prontuário em suporte papel quando da coleta da classificação ASA, não disponível na base de dados.

Definição das coortes

As coortes de derivação e de validação foram definidas por critério temporal, sendo a primeira constituída por pacientes submetidos a procedimento de laparotomia exploradora no período de fevereiro de 2000 (início da disponibilidade do prontuário eletrônico) a dezembro de 2002 e a segunda, por pacientes submetidos ao mesmo procedimento no período de janeiro a dezembro de 2003.

Critérios de inclusão e exclusão

Foram incluídos na amostra todos os pacientes adultos (maiores de 18 anos), internados no Hospital, submetidos a procedimento de laparotomia exploradora no bloco cirúrgico no período de estudo, por equipes das especialidades de cirurgia geral, proctologia e urologia, sob anestesia assistida.

Foram excluídos os pacientes com idade inferior a 18 anos, pacientes submetidos a cirurgia sem uso de anestesia assistida (anestesia local, uso de sedação e/ou analgesia), pacientes submetidos a cirurgia no Centro Cirúrgico Ambulatorial, pacientes ambulatoriais submetidos a procedimento cirúrgico no Bloco, pacientes submetidos a procedimento de laparotomia exploradora por outras especialidades cirúrgica e pacientes cujos prontuários se apresentaram sem a classificação ASA registrada.

Coleta dos dados

Os dados foram obtidos da base de dados de pacientes (Prontuário Eletrônico) do HCPA, com exceção da classificação pré-anestésica ASA, coletada por busca ativa nos prontuários em papel. Foram coletadas variáveis sócio-demográficas, administrativas, clínicas e laboratoriais, de acordo com critérios de relevância clínico-epidemiológica encontrados na literatura, confiabilidade e disponibilidade dos dados. Foram obtidas 45 variáveis originais que deram origem às variáveis relevantes para o estudo. Todas as variáveis utilizadas no processo de modelagem estavam disponíveis antes do procedimento cirúrgico. Os resultados de exames laboratoriais foram considerados aqueles disponíveis no período de internação, na data e horário mais próximos ao procedimento cirúrgico.

Dados faltantes

Com exceção dos dados laboratoriais, todas as demais variáveis estavam integralmente disponíveis nas fontes. A presença de resultados de exames considerados no estudo dependia da solicitação do mesmo pela equipe médica. Em 27,8% dos

pacientes não foi encontrado registro de albumina sérica. Embora alguns autores sugiram o uso de processos de imputação para variáveis ausentes em até 10% dos registros, seguimos o procedimento de Khuri e colegas (46) em seu estudo para ajuste de risco das taxas de mortalidade cirúrgica, que imputaram 39% dos valores de albumina sérica. Em nosso estudo o valor desta variável foi imputado considerando-se o valor da mediana da albumina dos pacientes submetidos a cirurgia eletiva ou de emergência, nas coortes de desenvolvimento (3,1g/dL ou 2,4g/dL) e de validação (3,7g/dL ou 2,6g/dL) (47-50). Análises adicionais foram realizadas para confirmar o acerto deste procedimento, através do desenvolvimento de outros dois modelos, um deles imputando-se o valor do limite inferior da faixa de normalidade da albumina sérica (3,5g/dL) e outro restrito aos casos em que o resultado do exame estava disponível.

Variáveis e medidas

As variáveis não dicotômicas foram analisadas antes da modelagem para definição da melhor forma de codificação. A avaliação pré-anestésica ASA classifica os pacientes em 5 categorias (I a V), de acordo com critérios definidos pela Sociedade Americana de Anestesiologistas (Quadro 1).

<p>ASA I: Paciente saudável, sem doença sistêmica e fora dos extremos de idade.</p> <p>ASA II: Indivíduo com uma doença sistêmica bem controlada, que não afeta sua atividade diária ou paciente com um risco anestésico como tabagismo, obesidade ou alcoolismo.</p> <p>ASA III: Indivíduo com múltiplas doenças sistêmicas ou com uma doença sistêmica grave, que limite sua atividade diária.</p> <p>ASA IV: Indivíduo com doença severa e incapacitante, em estágio terminal ou mal-controlada.</p> <p>ASA V: Paciente em iminente risco de morte, sendo a cirurgia o último recurso possível para preservar a vida ou atenuar o sofrimento.</p>

Quadro 1: Classificação da *American Society of Anesthesiology* (QI Project, 2003)

Apenas 3,4% de todos os pacientes (22 pacientes) foram classificados como ASA I, não havendo nenhum registro de óbito entre estes, e 4,0% (26 pacientes) como ASA V. A classificação ASA foi então agrupada em 3 classes (I e II, III, IV e V), seguindo critérios de Hosmer e Lemeshow (51). Na análise da distribuição dos valores da albumina sérica em grupos de óbito e não-óbito, foram identificados como pontos de corte, para esta amostra de pacientes, os valores de 2,2 e 3,0g/dL, criando-se então 3 categorias (até 2,2; 2,3-3,0; >3,0g/dL).

Sempre que possível, de acordo com critérios de plausibilidade conceitual, as variáveis categóricas foram dicotomizadas (cor, instrução, situação conjugal, procedência, convênio, caráter da internação e da cirurgia, especialidade cirúrgica, tipo de anestesia).

Análise estatística

Foram desenvolvidos modelos estatísticos para prever a mortalidade cirúrgica em um período de até 30 dias, durante a internação do paciente. Os dados foram analisados através dos programas Excell (Microsoft) e SPSS (*Statistical Package for Social Sciences*, versão 10.0). Na etapa de análise descritiva todas as variáveis foram analisadas para identificação e correção de valores fora dos limites (*outliers*) e dados ignorados. Inicialmente foram comparadas as coortes de derivação e de validação e a seguir, dentro da coorte de derivação, os grupos de óbito e não óbito. O processo de modelagem foi realizado apenas com os dados da coorte de derivação.

A relação univariável entre o desfecho e cada possível fator de risco pré-cirúrgico foi testada usando o teste do χ^2 para variáveis categóricas e o teste *t* de Student para variáveis quantitativas. Foram testadas as associações entre as variáveis explicativas para verificar a existência de colinearidade, utilizando-se o coeficiente gama, próprio para variáveis categóricas.

As relações entre o desfecho e as diversas variáveis preditivas consideradas em conjunto foram analisadas através de regressão logística múltipla. Foram considerados apenas os fatores de risco potenciais disponíveis no período pré-operatório e como variável de desfecho, os óbitos ocorridos durante o período de internação, nas coortes de derivação e de validação. Outras variáveis disponíveis também foram analisadas e descritas, mas não foram incluídas no processo de modelagem. Todas as variáveis com plausibilidade clínica e níveis de significância $<0,25$ na análise univariável foram incluídas seqüencialmente na fase de modelagem (51), em ordem de importância, de acordo com os valores da função de verossimilhança ($-2\log likelihood$). Fatores de risco potenciais que não se mostraram estatisticamente significantes no modelo (valor- $p > 0,05$) foram sendo retirados do modelo completo. Variáveis excluídas foram reintroduzidas em estágios subsequentes, em um processo de modelagem que se concluiu quando apenas variáveis significantes permaneceram. Todas as possíveis interações consideradas biológica ou epidemiologicamente plausíveis foram testadas.

Após a obtenção do modelo final, o processo de modelagem foi refeito pelo método automático (*stepwise - forward*) com o objetivo de comparar o desempenho deste método com o método “manual” utilizado.

Desenvolvimento do sistema de escore:

Utilizando o método de Le Gall e colegas (52) a partir do modelo obtido pela RL foram atribuídos pontos para cada fator de risco, multiplicando-se o coeficiente β do modelo de

regressão logística por 10 e arredondando para o inteiro mais próximo. O total de pontos atribuído a cada paciente foi considerado o “índice de risco para mortalidade cirúrgica em laparotomia exploradora”. Este escore foi então utilizado como variável única em uma nova equação de regressão logística aplicada aos pacientes das coortes de derivação e de validação, que calculou as probabilidades do desfecho. O índice de risco foi categorizado em cinco classes de risco, correspondentes a cinco níveis distintos e crescentes de gravidade, para futuras aplicações na prática clínica.

Avaliação da performance do modelo

Para validar o modelo comparou-se sua performance nas coortes de derivação e de validação (51). Seu poder de discriminação foi avaliado pelo índice-C (concordância de probabilidades), sua calibração foi verificada através do teste de Hosmer e Lemeshow e sua acurácia preditiva foi verificada através da tabela de classificação.

O Índice-C é a proporção de todos os possíveis pares de pacientes com e sem desfecho, nos quais a probabilidade de óbito predita pelo modelo para o paciente que morreu é maior do que aquela predita para o paciente que sobreviveu. Para desfechos dicotômicos, o índice-C é igual à área sob a curva ROC. A estatística de Hosmer e Lemeshow avalia a adequação do modelo, comparando os eventos esperados e observados em g grupos de igual tamanho (quantis), definidos por suas probabilidade de óbito predita. A partir do valor do índice-C verifica-se o valor-p para a distribuição χ^2 com $g-2$ graus de liberdade. Quanto maior o valor-p melhor é a adequação do modelo aos dados pois este resultado indica que a diferença entre a frequência esperada e observada dentro de cada grupo não é significativa (51;53-55).

Além dessas análises, os prontuários de todos os casos cujas probabilidades preditas pelo modelo estavam mais de 2 desvios-padrão fora da margem de erro (resíduos padronizados da RL) foram revisados para melhor entendimento de eventuais problemas ou limitações do modelo.

RESULTADOS

Características dos pacientes e taxas de mortalidade cirúrgica

Dos 1042 pacientes internados no HCPA submetidos a procedimento de laparotomia exploradora no HCPA no período de fevereiro de 2000 a dezembro de 2003, foram excluídos 391 pacientes por não atenderem aos critérios estabelecidos (366 pacientes ginecológicos, 16 pacientes de outras especialidades cirúrgicas não contempladas no estudo, 4 menores de 18 anos, 5 pacientes sem ASA informado).

A coorte de derivação foi constituída por 450 pacientes com alta entre o início do estudo até dezembro de 2002 (112 pacientes em 2000, 151 em 2001 e 187 em 2002) e a coorte de validação, por 201 pacientes com alta entre janeiro e dezembro de 2003.

A Tabela 1 apresenta as variáveis analisadas e as codificações adotadas, comparando características dos pacientes das duas coortes, mostrando não haver diferenças significativas entre os mesmos, com exceção dos valores médios da albumina sérica e da duração média da cirurgia, mais altos na coorte de validação. A mortalidade cirúrgica foi de 26,25% na coorte de derivação e de 24,9% na coorte de validação.

A Tabela 2 descreve a coorte de derivação, comparando os pacientes com e sem o desfecho em estudo.

Tabela 1 – Características sócio-demográficas e clínicas dos pacientes das coortes de derivação e de validação

Características dos pacientes	Derivação (n=450)	Validação (n=201)	Valor-p
Variáveis sócio- demográficas:			
Idade (média±DP, anos)	56.95 (±15.4)	57.7 (±17.6)	.582
75 anos e +	50 (11.1%)	32 (15.9%)	.088
Sexo masculino	222 (49.3%)	106 (52.7%)	.422
Cor branca	408 (90.7%)	190 (94.5%)	.096
1o Grau completo = sim	170 (41.4%)	85 (44.5%)	.468
Companheiro/a = sim	281 (62.4%)	117 (58.2%)	.306
SUS = sim	401 (89.1%)	169 (84.1%)	.072
Origem Porto Alegre	202 (44.9%)	103 (51.2%)	.133
Variáveis Internação			
Origem da internação =Clínica médica	101 (22.4%)	56 (27.9%)	.136
Semestre = 1°	241 (53.6%)	96 (47.8%)	.172
Internação urgência	283 (62.9%)	140 (69.7%)	.095
CID principal alta (capítulos mais prevalentes)			.508
Doenças infecciosas (Cap. I)	52 (12.1%)	16 (8.1%)	
Neoplasias (Cap. II)	167 (39.0%)	82 (41.4%)	
Doenças do Ap. Digestivo (Cap. IX)	113 (26.4%)	54 (27.3%)	
Sinais, sintomas, achados anormais	26 (6.1%)	13 (6.6%)	
Permanência (média±DP, em dias)	24.3 (±23.1)	22.7 (±20.3)	.403
Presença de valores de albumina	328 (72.9)	142 (70.6)	.571
Albumina sérica (média±DP, em g/dL)	2.7 (±0.8) (n=328)	2.9(±0.9) (n=142)	.010
Albumina sérica (mediana, em g/dL)	2,6	2,7	
Albumina sérica categórica			.386
1.0-2.2 g/dl	104 (23.1%)	42 (20.9%)	
2.3-3.0 g/dl	166 (36.9%)	67 (33.3%)	
>3.0 g/dl	180 (40.0%)	92 (45.8%)	
Saída por Óbito	124 (27.6%)	53 (26.4%)	.753
Variáveis Cirurgia			
Cirurgia geral = sim	361 (80.2%)	151 (75.1%)	.143
Urgência = sim	230 (51.1%)	112 (55.7%)	.276
Turno = noite	104 (28.9%)	48 (31.8%)	.513
Anestesia geral = sim	423 (94.2%)	192 (95.5%)	.493
ASA			.202
ASA I	12 (2.7%)	10 (5.0%)	
ASA II	163 (36.2%)	58 (28.9%)	
ASA III	150 (33.3%)	76 (37.8%)	
ASA IV	109 (24.2%)	47 (23.4%)	
ASA V	16 (3.6%)	10 (5.0%)	
ASA agrupado			.492
ASA I e II	175 (38,9%)	68 (33.8%)	
ASA III	150 (33.3%)	76 (37.8%)	
ASA IV e V	125 (27.8%)	57 (28.4%)	
Duração da cirurgia (média±DP, minutos)	119,7 (76,89)	141.1 (86.9)	.002
Duração da cirurgia (mediana, minutos)	105,0	125,0	
Dias internação - cirurgia (média±DP)	9.8 (±12.7)	8.6 (±11.1)	.239
Dias entre cirurgia e saída (média±DP)	13.7 (±17.8)	13.5 (±15.1)	.909
Óbito até 30 dias pós-cirurgia	118 (26.2%)	50 (24.9%)	.717
Diagnósticos pós-cirúrgicos			
Isquemia mesentérica	19 (4,2%)	9 (4,5%)	.514
Sepse	49 (10,9%)	18 (9,0%)	.274
Obstrução intestinal	24 (5,3%)	14 (7,0%)	.258

Tabela 2- Características sócio-demográficas e clínicas dos pacientes da coorte de desenvolvimento, com e sem desfecho de óbito em 30 dias

Características dos pacientes	Não-óbito (N=332) n , %	Óbito (N=118) n , %	Valor-p
Variáveis sócio- demográficas:			
Idade (média±DP, anos)	55.42 (±15.1)	61.23 (±15.8)	<.001
75 anos e +	26 (7.8%)	24 (20.3%)	<.001
Sexo masculino	161 (48.5%)	61 (51.7%)	.550
Cor branca	302 (91.0%)	106 (89.8%)	.716
1o Grau completo = não	175 (56.6%)	66 (64.7%)	.151
Companheiro/a = sim	216 (65.1%)	65 (55.1%)	.055
SUS = sim	291 (87.7%)	110 (93.2%)	.095
Origem Porto Alegre	134 (40.4%)	68 (57.6%)	.001
Variáveis Internação			
Clínica de origem = médica	55 (16.6%)	46 (39.0%)	<.001
Semestre = 1º	166 (50.0%)	75 (63.6%)	.011
Internação = urgência	187 (56.3%)	96 (81.4%)	<.001
CID principal alta (capítulos mais prevalentes)			<.001
Doenças infecciosas (Cap. I)	11 (3.4%)	41 (38.7%)	
Neoplasias (Cap. II)	141 (43.8%)	26 (24.5%)	
Doenças do Ap. Digestivo (Cap. IX)	96 (29.8%)	17 (16.0%)	
Sinais, sintomas e achados anormais (Cap. XX)	17 (5.3%)	9 (8.5%)	
Permanência (média±DP, em dias)	25.8 (±25.0)	19.9 (±16.1)	.018
Presença de valores da albumina	219 (66,8)	109 (92,4)	<.001
Albumina sérica (média±DP, em g/dl)	2.94 (±0.7)	2.23 (±0.7)	<.001
Albumina sérica categórica			<.001
1.0-2.2 g/dl	39 (11.7%)	65 (55.1%)	
2.3-3.0 g/dl	126 (38.0%)	40 (33.9%)	
>3.0 g/dl	167 (50.3%)	13 (11.0%)	
Variáveis Cirurgia			
Cirurgia geral = sim	257 (77.4%)	104 (88.1%)	.012
Urgência = sim	163 (49.1%)	97 (82.2%)	<.001
Turno = noite	66 (19.9%)	38 (32.2%)	.006
Anestesia geral = sim	306 (92.4%)	117 (99.2%)	.007
ASA			<.001
ASA I	12 (3.6%)	0 (0,0%)	
ASA II	155 (46.7%)	8 (6.8%)	
ASA III	124 (37.3%)	26 (22.0%)	
ASA IV	38 (11.4%)	71 (60.2%)	
ASA V	3 (.9%)	13 (11.0%)	
ASA agrupado			<.001
ASA I e II	167 (50.3%)	8 (6.8%)	
ASA III	124 (37.3%)	26 (22.0%)	
ASA IV e V	41 (12.3%)	84 (71.2%)	
Duração da cirurgia (média±DP, minutos)	124.3 (±79.5)	106.6 (±67,6)	.032
Dias entre internação e cirurgia (média±DP)	8.9 (±12.8)	12.5 (±12.1)	.009
Dias entre cirurgia e saída (média±DP)	16.1 (±19.6)	6.9 (±7.4)	<.001
Diagnósticos pós-cirúrgicos			
Isquemia mesentérica	6 (1,8%)	13 (11,0%)	<.001
Sepse	6 (1,8%)	43 (36,4%)	<.001
Obstrução intestinal	18 (5,4%)	6 (5,1%)	.553

Verifica-se que os pacientes que foram a óbito apresentam idade média (\pm DP) maior ($61,2\pm 15,7$ e $55,4\pm 15,1$ anos respectivamente), maior proporção de pacientes com procedência local (57,6% e 40,4%), de internação em especialidade não-cirúrgica (39% e 16,6%), de realização da cirurgia no primeiro semestre (63,3% e 50%), de internação de urgência (81,4% e 56,3%), de cirurgia geral (88,1% e 77,4%), de realização de cirurgia em caráter de urgência (82,2% e 49,1%), no primeiro semestre do ano (50% e 63%) e no turno da noite (32,2% e 19,9%) e utilização de anestesia geral (99,2% e 92,4%). Estes pacientes apresentaram um valor médio de albumina sérica menor ($2,23\pm 0,65$ e $2,94\pm 0,73$ g/dL), o mesmo acontecendo em relação à média de permanência (19,9 e 25,8 dias) e à duração da cirurgia (124,3 e 106,6 minutos).

O período entre a internação do paciente e a realização da cirurgia foi maior no grupo de óbito (12,4 e 8,9 dias) enquanto que o período entre a cirurgia e a saída do paciente do hospital foi menor neste grupo (6,9 e 16,1 dias). A ocorrência de sepse e o achado de isquemia mesentérica também se mostraram fatores significativamente associados ao desfecho. A distribuição dos pacientes nas categorias *da American Society of Anesthesiology* (ASA) difere significativamente nos dois grupos, verificando-se uma proporção de 71,2% de pacientes com óbito nas categorias IV e V, contra apenas 12,3% nos demais.

As variáveis sexo, cor, escolaridade, convênio, situação conjugal e duração da cirurgia não apresentaram diferenças significativas entre os dois grupos. A informação de diagnóstico inicial, disponível antes da cirurgia, é de caráter administrativo e por esta razão não foi utilizada. Os diagnósticos médicos, disponíveis no prontuário, são informados apenas no momento da alta ou óbito do paciente, podendo vir a ser utilizados como variáveis preditoras candidatas em estudos com outra abordagem que não preditiva.

A Tabela 3 apresenta o desempenho individual de cada variável selecionada para o processo de modelagem identificando-se nesta fase, como preditores isolados mais relevantes, a classificação ASA, as classes de albumina sérica, grau de instrução, caráter da cirurgia e da internação, idade, procedência e realização de cirurgia anterior.

A variável idade funcionou bem no modelo tanto como variável contínua como dicotômica (até 74 anos e 75 anos e mais). A segunda opção foi escolhida em razão da necessidade de se atribuir pontos a cada variável do modelo para obtenção do escore de risco, o que seria dificultado pela presença de uma variável contínua no modelo (56).

Tabela 3: Análise univariável (coorte de derivação

Variável	-2log V	β	OR (IC95%)	Valor-p
ASA	361,5			<0,001
ASA I e II			Referência	
ASA III		1,48	4,38 (1,92-9,99)	<0,001
ASA IV e V		3,76	42,77 (19,19-95,34)	<0,001
Albumina sérica categórica	414,30			<0,001
Até 2,2		3,06	21,41 (10,74-42,68)	<0,001
2,3-3,0		1,41	4,08 (2,09-7,94)	<0,001
>3,0			Referência	
1o Grau completo=não	458,5	0,34	1,40 (0,88-2,23)	0,152
Cirurgia=urgência	475,6	1,48	4,41 (2,74-7,10)	<0,001
Internação=urgência	492,7	1,22	3,38 (2,03-5,64)	<0,001
Clínica de origem=médica	494,5	1,12	3,22 (2,01-5,15)	<0,001
Turno=noite	510,7	0,65	1,91 (1,20-3,07)	0,007
Idade	505,2	0,03	1,03 (1,01-1,04)	0,001
Idade>=75 anos	505,4	1,10	3,01 (1,65-5,48)	<0,001
Faixa Etária tercis	506,6	0,004		
Até 50 anos			Referência	
50-70 anos		0,51	1,66 (0,99-2,79)	0,056
>70 anos		0,97	2,63 (1,48-4,66)	0,001
Faixa Etária quartis	507,6	0,02		
Até 46 anos			Referência	
47-58 anos		0,35	1,42 (0,75-2,68)	0,277
59-70 anos		0,53	1,71 (0,90-3,24)	0,102
70 anos		0,97	2,66 (1,4204,90)	0,002
Procedência=Porto Alegre	507,4	0,70	2,01(1,31-3,08)	0,001
Cirurgia anterior <15 dias	508,0	0,71	2,04 (1,30-3,20)	0,002
Especialidade=Cirurgia Geral	510,9	0,77	2,17 (1,17-4,01)	0,014
Semestre=1º	511,3	0,56	1,74 (1,13-2,69)	0,012
Dias entre internação e cirurgia	511,5	0,02	1,02 (1,01-1,04)	0,011
Companheiro/a	514,2	0,42	1,52(0,99-2,33)	0,055
Convênio=SUS	514,8	0,66	1,94 (0,88-4,26)	0,100

Análise multivariável

Todas as variáveis identificadas como potencialmente relacionadas ao desfecho que apresentaram valores-p no teste de χ^2 ou no teste-t $\leq 0,25$ foram incluídas no processo de modelagem. As variáveis foram adicionadas uma a uma ao modelo de acordo com o valor da função de Verossimilhança (*-2log likelihood*) medido na análise univariável. (Tabela 3), por processo “manual”. Foram mantidas no modelo final todas as variáveis que apresentaram uma significância $<0,05$, após análise de potenciais fatores de interação. O modelo final obtido após várias etapas e diferentes combinações de variáveis é apresentado na Tabela 4.

De acordo com o modelo selecionado, os fatores de risco associados à mortalidade cirúrgica são a classificação pré-anestésica de alto risco (IV e V) baseada nos critérios da *American Society of Anesthesiology* (ASA), baixos níveis séricos de albumina e idade igual ou superior a 75 anos. Pacientes classificados como ASA III e ASA IV ou V apresentaram respectivamente OR de 3,32 (IC95%:1,41-7,84) e 22,3 (IC95%: 9,46-52,6) em relação

aos pacientes classificados como ASA I e II. Pacientes com 75 anos ou mais têm uma OR de 4,08 (IC95%:1,83-9,1) em relação aos demais. Pacientes com níveis de albumina sérica de até 2,2 g/dl e aqueles com valores entre 2.3 e 3.0 g/dl apresentaram OR de 6,58 (IC95%:2,93-14,81) e de 2,18 (IC95%:1,02-4,67) respectivamente.

Tabela 4- Modelo de Regressão Logística e escores dos fatores de risco para mortalidade cirúrgica (coorte de desenvolvimento)

Variável	OR (IC 95%)	β	Escore
ASA I e II	1.0 (referência)		0
ASA III	3.32 (1.41-7.84)	1.20	12
ASA IV e V	22.30 (9.46-52.60)	3.10	31
Idade \geq 75 anos	4.08 (1.83-9.10)	1.42	14
Idade <75 anos	1.0 (referência)		0
Albumina até 2.2 g/dl	6.58 (2.93-14.81)	1.89	19
Albumina 2.3-3.0 g/dl	2.18 (1.02-4.67)	.79	8
Albumina > 3.0 g/dl	1.0 (referência)		0
Constante		-3,779	

A modelagem realizada pelo método “caixa-preta” (*stepwise* ou passo-a-passo) com os dados da coorte de derivação, resultou em um modelo com 4 variáveis preditivas, das quais 3 estão no modelo final. Além destas, entrou no modelo a variável “1º semestre” (OR=1,913, IC95%:1,079-3,389), que indicaria uma maior probabilidade de óbito no período logo após o ingresso de novos residentes, coincidindo com a época de férias dos professores, descrito na literatura como “July Phenomenon”. (57), No entanto, avaliando o desempenho deste modelo no grupo de validação, verificou-se que esta variável passa a ter um efeito protetor (OR=0,304, IC95%:0,114-0,812) e, modelando com as duas coortes juntas, a mesma não permanece com significância estatística ($p=0,570$). Portanto, a variável não foi incluída no modelo final.

O modelo com a variável “albumina sérica” com imputação de valores normais (3,5g/dL) para os dados faltantes na coorte de desenvolvimento apresentou coeficientes muito semelhantes ao modelo final que considerou os valores de imputação pela mediana dos grupos de cirurgia de urgência e eletiva, conforme podemos verificar na Tabela 5.

Tabela 5: Comparação dos coeficientes dos modelos com diferentes tratamentos dos dados faltantes de albumina sérica (Coorte de derivação)

Variável	β - Medianas (3,1 ou 2,4g/dL)	β - Normal (3,5g/dL)
ASA III	1,201	1,201
ASA IV e V	3,105	3,136
Idade <75 anos	1,406	1,408
Albumina até 2.2 g/dl	1,885	1,756
Albumina 2.3-3.0 g/dl	0,779	0,773
Constante	-3,779	-3,670

As médias e medianas dos valores da albumina nas diferentes estratégias de imputação estão sumarizadas na Tabela 6, cujos resultados demonstram que a imputação pela mediana obteve resultados mais próximos aos valores reais. De qualquer forma, a albumina se manteve como a segunda variável mais importante no modelo, com qualquer estratégia de imputação adotada.

Tabela 6: Comparação dos valores da albumina em diferentes estratégias de imputação

Estratégia de imputação	Médias	DP	Mediana	Área sob a curva ROC
Medianas (3,1 ou 2,4 g/dL)	2,738	0,693	2,700	0,799
Valor normal (3,5g/dL)	2,918	0,756	3,100	0,801
Sem imputação (n=328)	2,702	0,781	2,600	0,790

Analisando cada variável do modelo isoladamente, verifica-se que a área sob a curva ROC (índice-C, discriminabilidade) é de 0,841 para a classificação ASA, de 0,783 para a albumina e de 0,556 para idade ≥ 75 anos, como demonstrado na Figura 1.

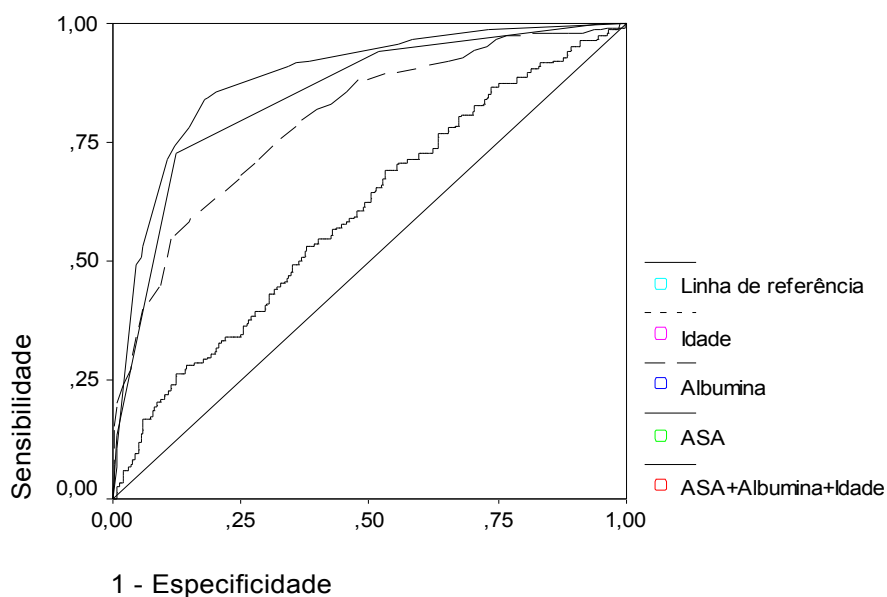


Figura 1: Comparação da discriminabilidade das diferentes variáveis do modelo

Na Tabela 7 são apresentadas os valores do índice-C das diferentes combinações de variáveis, demonstrando que a variável ASA apresenta, isoladamente, o maior poder discriminatório, que melhora com a inclusão da variável idade (0,856), da variável albumina (0,876) ou de ambas (0,866). O poder discriminatório da albumina isolada também melhora com a inclusão da variável idade (0,800). De acordo com os critérios de classificação de Hosmer e Lemeshow, a variável ASA apresenta uma discriminação excelente (0,8-0,9) enquanto a discriminação da albumina seria aceitável (0,7-0,8) isoladamente e excelente quando associada com a idade.

Tabela 7: Discriminação (área sob a curva ROC) das diferentes variáveis e combinações (n=650)

Variáveis	Índice-C
ASA	0,841
Albumina	0,783
Idade	0,556
ASA + Albumina	0,876
ASA + Idade	0,856
Albumina + Idade	0,800
ASA + Albumina + Idade	0,886

Outras combinações de variáveis foram testadas para melhor entendimento dos fatores de risco associados à mortalidade cirúrgica. Na ausência da classificação ASA, o modelo preditivo para a coorte de derivação passa a conter as variáveis albumina sérica, idade ≥ 75 anos, caráter de urgência da cirurgia e procedência local, com calibração de 0,939, discriminação de 0,827 e acurácia de 79,1%. Quando não se inclui a variável albumina, o modelo passa a ser composto pelas variáveis ASA, idade e caráter de urgência da cirurgia (calibração = 0,953, discriminação = 0,865 e acurácia = 83,7%). É interessante notar que o caráter de urgência da cirurgia, fortemente associado ao desfecho quando avaliado isoladamente ($p < 0,001$), é substituído nos modelos tanto pela variável ASA quanto pela variável albumina. A entrada da variável de procedência local na ausência da classificação ASA pode ser explicada pela impossibilidade de remoção de pacientes moradores de outras cidades com ASA IV ou V, pelo risco de óbito no trajeto, embora a associação entre procedência e ASA dicotômica (I,II ou III e IV ou V) não seja estatisticamente significativa ($p = 0,089$).

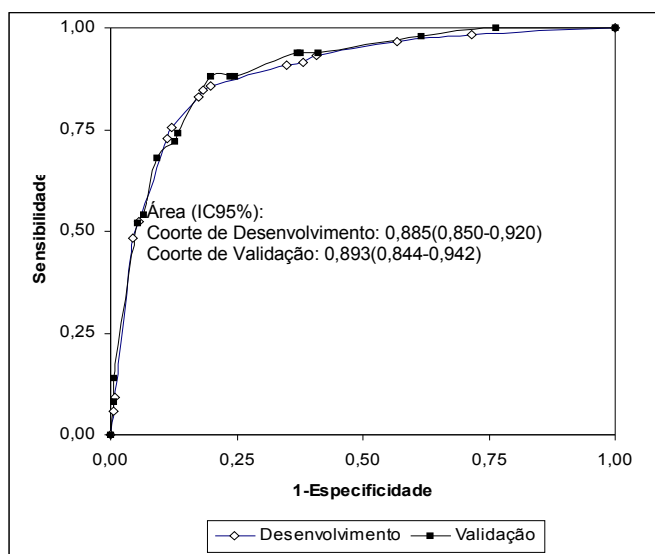
Cálculo do escore

Os escores de risco foram calculados multiplicando-se os valores de β por 10 e arredondando o resultado para o inteiro mais próximo. Desta forma, foram definidos 12 pontos para pacientes de ASA III, 31 pontos para pacientes de ASA IV ou V, 14 pontos para pacientes com 75 anos e mais, 19 pontos para pacientes com albumina inferior a 2.3g/dl, 8 pontos para pacientes com albumina entre 2.3 e 3.0 g/dl e zero pontos para as categorias de referência (Tabela 4).

Atribuição dos escores

Para cada paciente foi calculado um escore de risco, multiplicando-se sua condição em cada variável pelo respectivo peso, conforme indicado na tabela 4. (ASA = 0, 12 ou 31 pontos, albumina = 0, 8 ou 19 pontos e idade = 0 ou 14 pontos).e somando-se estas parcelas ao final. Este escore foi então utilizado como variável independente única em novo modelo de regressão logística, analisando-se as duas coortes separadamente, obtendo-se a probabilidade predita de desfecho de cada paciente. A partir das probabilidades preditas foram geradas as Curvas ROC das duas coortes (Figura 2).

Figura 2: Curva ROC (Receiver Operation Curve) da probabilidade de óbito cirúrgico calculada a partir do escore



de risco atribuído a cada paciente

Avaliação do modelo

Para a coorte de desenvolvimento, o modelo apresentou uma acurácia geral de 85%, sensibilidade de 73% e especificidade de 89% (Tabela 8).

Tabela 8- Análise da acurácia do modelo de risco para mortalidade cirúrgica (coorte de desenvolvimento)

Prognóstico do modelo	Observado		
	Óbito	Não óbito	
Óbito	86	37	<i>Sensibilidade = 73%</i> <i>VPP = 70%</i>
Não Óbito	32	295	
	118	332	<i>Especificidade = 89%</i> <i>VPN = 90%</i> <i>Acurácia = 85%</i>

O modelo aplicado à coorte de derivação resultou em uma estatística-C de 0,89 e um valor de 0,96 no teste de Hosmer e Lemeshow, valores que indicam boa discriminação e calibração respectivamente. Aplicado à coorte de validação, resultou em uma estatística-

C de 0,90 e um valor de 0,92 no teste de Hosmer e Lemeshow, indicando que a discriminação melhorou mas que a calibração foi um pouco mais pobre num conjunto de dados temporalmente independente. Para verificar a performance geral do modelo, o mesmo foi aplicado a todos os pacientes (n=651), obtendo-se valores de calibração e discriminabilidade de 0,94 e 0,89 respectivamente. (Tabela 9).

Tabela 9- Performance do modelo preditivo de regressão logística com a variável “escore total”

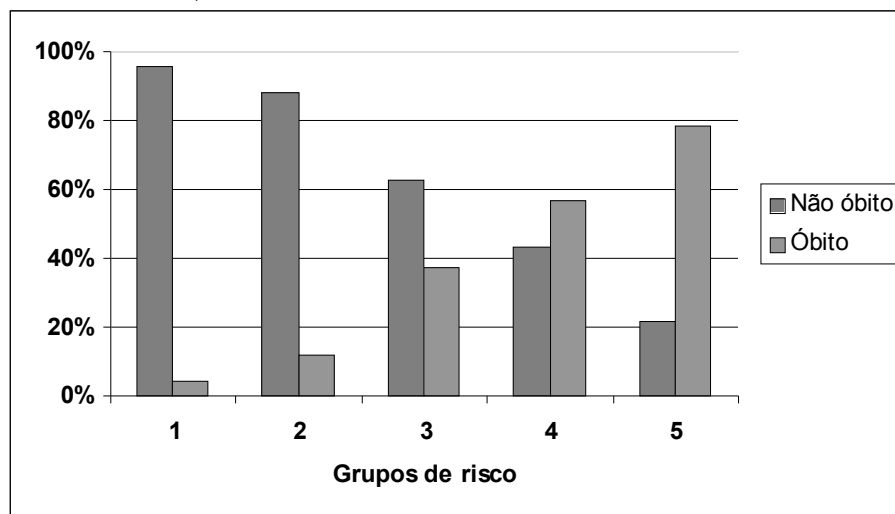
Coorte	Teste de H&L (calibração)	Índice-C (Discriminação)	Acurácia
Derivação	0.961	0.885 (.849-.920)	84.7%
Validação	0.916	0.899 (.852-.946)	83.6%
Todos os pacientes	0.939	0.887 (.859-.916)	84.8%

A análise dos casos com resíduos padronizados maiores que 2,0 dos modelos de RL (*casewise list*) indicou a existência de 13 casos considerados como mal classificados na coorte de derivação e de 5 casos na coorte de validação. Dos 18 casos, 3 foram preditos como óbito, o que não foi observado. Destes, dois pacientes morreram durante a internação, em um período maior que 30 dias (52 e 58 dias) e 1 paciente com mais de 80 anos, albumina baixa e ASA III, sobreviveu por mais de 3 anos, o que deve ser explicado pela variabilidade biológica humana, pela presença de outras variáveis não incluídas no modelo ou por erro experimental. Dos 15 casos em que o modelo não predisse o óbito mas este foi observado, 11 pacientes apresentaram quadro séptico pós-cirúrgico como motivo do óbito (complicação cirúrgica).

Índice de Risco para Mortalidade Cirúrgica

A partir dos escores calculados e mostrados na Tabela 4, os pacientes foram divididos em 5 grupos de risco, de acordo com os escores de risco obtidos (0 a 18 pontos, 19 a 28 pontos., 29 a 38 pontos, 39 a 48 pontos e ≥ 49 pontos). A Figura 3 apresenta a distribuição de todos os pacientes nas diferentes classes de risco, divididos em óbitos e não-óbitos, mostrando que os grupos de risco definidos realmente indicam a probabilidade do desfecho em estudo.

Figura 34. Classes de risco x óbitos: todos os pacientes (n=651)



A tabela 10 apresenta o número de óbitos esperados e observados em cada classe de risco e a razão O/E, que não se mostrou excelente apenas na classe II.

Tabela 10: Distribuição das classes de risco para mortalidade cirúrgica nas coortes de desenvolvimento e de validação

	Classes de risco				
	I (0-18 pontos)	II (19-28 pontos)	III (29-38 pontos)	IV (39-48 pontos)	V (> 49 pontos)
Coorte de derivação e de validação (n,%)	215 (47,8)	79 (17,6)	33 (7,6)	51 (11,3)	72 (16,0)
Óbitos esperados	13,06	15,84	19,74	35,54	83,81
Óbitos observados (25,8%, 168/651)	13	13	22	37	83
Razão O/E	1,00	0,82	1,11	1,04	0,99

Aplicação do índice de risco para ajuste do indicador

Para testar a utilidade do índice desenvolvido no ajuste do indicador de mortalidade cirúrgica, o mesmo foi aplicado para comparação do desempenho das diferentes equipes cirúrgicas do HCPA. Foi calculado o número de óbitos previstos para cada equipe de acordo com a probabilidade obtida pelo modelo e obtida a razão entre os óbitos esperados e observados (Tabela 11).

Tabela 11: Ajuste da mortalidade cirúrgica pelo índice de risco

EQUIPES	Classificação desempenho	OBSERVADO			ESPERADO		Desempenho ajustado
		Óbitos	Cirurgias	% óbitos	Óbitos	O/E	
A	1	6	37	16,2	6,6	0,91	3
B	2	20	114	17,5	25,2	0,79	1
C	3	19	95	20,0	20,3	0,94	4
D	4	8	38	21,1	6,5	1,23	10
E	5	13	48	27,1	10,9	1,19	9

F	6	17	60	28,3	20,3	0,84	2
G	7	38	128	29,7	36,2	1,05	5
H	8	15	47	31,9	13,3	1,13	8
I	9	21	56	37,5	18,9	1,11	6
J	10	11	28	39,3	9,8	1,12	7
Total		168	651	25,8	168	1,00	

Este valor gerou uma classificação de desempenho diferente daquela obtida com os dados não ajustados por risco, conforme podemos ver na Figura 4.

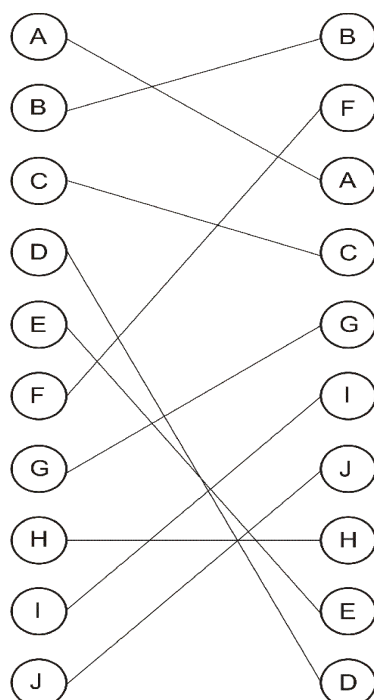


Figura 4: Classificação bruta e ajustada do desempenho das equipes cirúrgicas no indicador “mortalidade cirúrgica”

Avaliação da qualidade das fontes de dados

Com exceção dos valores da classificação ASA, os dados foram obtidos através de *query* na base de dados clínica do HCPA. A classificação ASA foi coletada através de revisão de cerca de 1.000 prontuários em suporte-papel, procedimento que também permitiu a validação das informações sobre o procedimento realizado e sobre a ocorrência de óbito.

Dos prontuários revisados, em apenas cinco a classificação ASA não estava registrada e todos os prontuários solicitados foram localizados no Serviço de Arquivo Médico do HCPA. Na análise dos dados eletrônicos, foi constatado que as variáveis cadastrais data de nascimento, sexo, convênio, profissão, procedência, estado civil e

bairro não apresentaram nenhum valor ignorado e que a informação sobre grau de escolaridade era ignorada em 4,7% dos registros. No entanto, a informação registrada sobre a profissão dos pacientes não permite nenhuma análise de interesse, havendo 180 registros de “aposentados” e 267 registros de “outras”. O registro da raça do paciente não reflete a realidade observada, pois segundo este, 92% dos pacientes do HCPA são brancos. As variáveis clínicas do prontuário eletrônico estavam bastante completas, faltando a informação de diagnóstico inicial em 0,3% dos pacientes e do diagnóstico final em 3,8% destes. Os resultados de exames estavam disponíveis sempre que solicitados pela equipe médica.

DISCUSSÃO

Nossos resultados confirmam os achados de vários outros trabalhos anteriormente publicados sobre fatores de risco para mortalidade cirúrgica não cardíaca considerando a classificação ASA, a idade avançada e a albumina sérica dos pacientes como variáveis preditoras de óbito suficientes para garantir uma acurácia prognóstica adequada.

Existem debates sobre a habilidade preditiva da classificação ASA desde sua primeira descrição (58-62). Esta classificação é feita a partir de dados pouco objetivos ou mensuráveis e visa apenas avaliar as condições do paciente para enfrentar a cirurgia. Pacientes de classes ASA mais altas (IV e V) apresentam maior risco de óbito mas alguns argumentam que o *status* da ASA por si só não discrimina grupos por desfechos porque não leva em conta fatores cirúrgicos ou diagnósticos.

No presente estudo, em que todos os pacientes foram submetidos ao mesmo procedimento cirúrgico, a classificação ASA apresentou um índice-C (discriminação) de 0,84 quando utilizada de forma isolada em equação de regressão logística. Este resultado confirma a importância desta classificação, que foi um fator preditivo relevante de óbito pós-cirúrgico o que demonstra que embora esta classificação seja feita apenas a partir de dados de anamnese e indique uma avaliação pessoal do médico, apresenta excelente acurácia e pode ser utilizada para identificar pacientes de baixo e alto risco bem como para ajustar indicadores, além de ser isenta de custos e de riscos para o paciente (63;64),

Os níveis séricos de albumina são um importante indicador da gravidade das condições gerais do paciente. A albumina sérica é um bom indicador laboratorial do estado nutricional do paciente hospitalizado (65) e a hipoalbuminemia é um marcador de doença severa, encontrada em cerca de 65% dos pacientes de unidades de cuidados intensivos e sua medida faz parte do sistema APACHE III (66), utilizado para o prognóstico de pacientes internados em Unidades de Tratamento Intensivo. A hipoalbuminemia é apontada como fator de risco para desfechos hospitalares em

diversos estudos(67-69) mas definir se os níveis de albumina se devem à desnutrição ou a uma combinação de outros fatores tais como abuso de álcool, doença hepática ou outras etiologias não nutricionais, estava além do escopo destes estudo.

Dados faltantes constituem um grande problema em estudos retrospectivos e nosso estudo não foi uma exceção. Exames laboratoriais, especialmente a albumina sérica, estavam ausentes em grande proporção de pacientes (27,8%). Muitos modelos de avaliação de risco, tais como o APACHE, imputa, para dados laboratoriais faltantes, valores fisiologicamente normais, enquanto outros fazem esta imputação por outros métodos (70-72). É sabido que a substituição de dados faltantes por valores imputados (médias, medianas e outros) pode introduzir vieses nas análises, o que pode ser verificado e corrigido por análises complementares, como foi feito neste trabalho.

Estudos diversos têm mostrado consistentemente que a mortalidade cirúrgica aumenta com a idade, o que poderia ser explicado pelo maior número de comorbidades presentes. Mas na maioria destes estudos, a idade tem sido um fator preditor independente, mesmo controlada por outras variáveis de comorbidades. Embora a idade cronológica não seja um perfeito preditor de fragilidade, pacientes mais velhos claramente possuem menos reserva fisiológica e mais doenças cardíacas, cujas severidades podem ser mascaradas por seus níveis diminuídos de atividade física (73). Neste estudo verificamos que a idade é associada a maior probabilidade de óbito, independente da presença de comorbidades, avaliada subjetivamente pela classificação ASA.

A aferição da acurácia deste modelo foi realizada por meios estatísticos que determinam a confiabilidade e o poder discriminatório do mesmo. Os valores encontrados para o teste de Hosmer e Lemeshow (0,961 para a coorte de desenvolvimento e 0,939 para todos os pacientes) indicam uma calibração muito boa. A área sob a curva ROC foi de 0,885 para a coorte de desenvolvimento, 0,893 para a de validação e de 0,886 para todos os pacientes, o que, de acordo com os critérios da literatura (74), indica uma discriminação excelente.

Quando se desenvolve um sistema de escore de risco para avaliar pacientes individuais e permitir avaliações comparativas de taxas de desfechos cirúrgicos, é importante que, além de acurado, este seja de fácil uso. As variáveis coletadas devem ser em pequeno número, estar disponíveis para todos os pacientes, ser de uso comum em diferentes hospitais e deixar pouco espaço para vieses de observação e de coleta. O escore deve ser fácil de ser calculado a partir dos dados coletados e deve ser possível integrar este sistema de escore em rotinas e sistemas pré-existentes, com um mínimo de

intervenção. Nossos resultados estão de acordo com estas premissas, pois o índice de risco para mortalidade pós laparotomia exploradora foi construído com três variáveis de registro rotineiro em hospitais, sendo de fácil medida e utilização tanto internamente, para identificação de pacientes de alto risco quanto externamente, para comparações de indicadores ajustados entre hospitais.

Nosso modelo apresenta as características de tamanho de amostra exigidas em estudos deste tipo. Herrel (75) e Bagley (76) propõem a regra de 10 desfechos por variável e o manual de estilo da *American Medical Association* apresenta o critério de 25 casos por variável. Considerando que o número de variáveis preditoras referidas acima significa o número de coeficientes do modelo, nosso modelo possui 5 variáveis (2 coeficientes para ASA, 2 para albumina e 1 para idade), o que exigiria pelo menos 50 desfechos ou 125 casos, de acordo com os diferentes critérios. Nosso grupo de desenvolvimento tinha 450 casos e 168 desfechos, enquanto que o grupo de validação apresentou 201 casos e 50 desfechos, atendendo plenamente aos critérios estabelecidos.

A etapa de coleta e tratamento dos dados, bem como os resultados alcançados, demonstraram que a base de dados de pacientes (prontuário eletrônico) do HCPA apresenta critérios de qualidade (completude, fidedignidade, disponibilidade) que a caracterizam como uma fonte de informações clínicas de grande utilidade para pesquisa e perfeitamente adequada para o desenvolvimento de modelos de risco para grande parte dos desfechos hospitalares de interesse epidemiológico. Portanto, foi possível concluir que esta base de dados é adequada para este tipo de pesquisa, o que não foi comprovado por Martins, Travassos e Noronha (77) ao analisarem a base de dados de cobrança de internações hospitalares do SUS para este mesmo fim.

Ao interpretar nossos achados é importante considerar várias limitações potenciais. Primeiro, embora o estudo tenha construído o modelo de ajuste de risco baseado em dados clínicos e tenha incluído fatores de risco já demonstrados como potenciais fatores de risco em estudos anteriores, as variáveis analisadas, embora com um excelente resultado final, estavam limitadas àquelas indicadas pela literatura e pela opinião de especialistas do HCPA e disponíveis no prontuário de pacientes do HCPA (papel e eletrônico) e portanto, potenciais fatores de risco podem ter sido excluídos. Segundo, o estudo foi realizado com grupos de pacientes de apenas um hospital, universitário, público e de alta complexidade, em um período de tempo limitado (2000-2003), e o número de pacientes arrolados, embora suficiente para a metodologia, não é representativo da população geral de Porto Alegre ou do País. Portanto, a generalização

de nossos achados para outros hospitais, especialmente hospitais não universitários e/ou hospitais privados é incerta. O escore de índice foi construído para prever desfechos cirúrgicos pós-laparotomia exploradora, não podendo também ser generalizado para outras cirurgias.

Este estudo apresenta algumas contribuições relevantes para o avanço dos processos de ajuste de risco de indicadores de mortalidade cirúrgica. Demonstramos que é possível desenvolver um modelo estatístico baseado em variáveis clínicas que é capaz de prever o desfecho de mortalidade após realização de laparotomia exploradora, através de um processo simples que envolve avaliação de apenas três variáveis coletadas rotineiramente em nosso meio: ASA, idade e albumina sérica. Este estudo demonstra o valor do desenvolvimento de um conjunto comum de variáveis prognósticas, baseadas em definições padronizadas para condições clínicas específicas, que podem ser incorporados em iniciativas regionais ou nacionais de avaliação comparativa de indicadores, tais como o SIPAGEH (78). Tal padronização facilitaria a comparação dos indicadores de desfecho entre diferentes hospitais e forneceria uma melhor compreensão dos desempenhos relativos.

Concluindo, foi utilizada uma combinação de fatores de risco para desenvolver e validar um índice de risco para prognóstico de mortalidade após laparotomia exploradora. Este modelo poderá vir a ser utilizado pelas equipes médicas do HCPA para controlar fatores de risco relativos ao paciente e intervenções que visem a modificar fatores intra e pós-cirúrgicos, devendo para isso ser validado e ajustado para grupos de procedimentos cirúrgicos mais abrangentes. Para fins de ajuste de indicadores comparativos de mortalidade cirúrgica (SIPAGEH e outros), através da estratificação dos óbitos em grupos de risco, entendemos que seja adequada a utilização isolada da classificação ASA, mais simples e mais disponível, e que apresentou um excelente poder discriminatório em nossos estudos, perfeitamente adequado para este fim.

Acreditamos, portanto, que estamos trazendo uma contribuição importante para a área de epidemiologia hospitalar mas entendemos que novos estudos, especialmente em ambientes multicêntricos, são necessários para garantir a generalização destes achados e determinar o conjunto ótimo de fatores de risco de mortalidade cirúrgica associada ao procedimento de laparotomia exploradora.

O projeto foi aprovado pela Comissão de Bioética do Grupo de Pesquisa e Pós-Graduação do HCPA sob número 431/2002. Todos os dados que permitissem a identificação do paciente ou dos profissionais foram omitidos ou criptografados.

- (1) Iezzoni LI. Risk adjustment for measuring health care outcomes. 3rd ed. Chicago, Illinois: Health Administration Press, 2003.
- (2) Knaus WA, Zimmerman JE, Wagner DP, Draper EA, Lawrence DE. APACHE-acute physiology and chronic health evaluation: a physiologically based classification system. *Crit Care Med* 1981; 9(8):591-597.
- (3) Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. *Crit Care Med* 1985; 13(10):818-829.
- (4) Schein M, Gecelter G, Freinkel Z, Gerding H. APACHE II in emergency operations for perforated ulcers. *Am J Surg* 1990; 159(3):309-313.
- (5) Knaus WA, Wagner DP, Draper EA, Zimmerman JE, Bergner M, Bastos PG et al. The APACHE III prognostic system. Risk prediction of hospital mortality for critically ill hospitalized adults. *Chest* 1991; 100(6):1619-1636.
- (6) Jones HJ, de Cossart L. Risk scoring in surgical patients. *Br J Surg* 1999; 86(2):149-157.
- (7) Cho DY, Wang YC. Comparison of the APACHE III, APACHE II and Glasgow Coma Scale in acute head injury for prediction of mortality and functional outcome. *Intensive Care Med* 1997; 23(1):77-84.
- (8) Barie PS, Hydo LJ, Fischer E. Comparison of APACHE II and III scoring systems for mortality prediction in critical surgical illness. *Arch Surg* 1995; 130(1):77-82.
- (9) Copeland GP, Jones D, Walters M. POSSUM: a scoring system for surgical audit. *Br J Surg* 1991; 78(3):355-360.
- (10) Jones DR, Copeland GP, de Cossart L. Comparison of POSSUM with APACHE II for prediction of outcome from a surgical high-dependency unit. *Br J Surg* 1992; 79(12):1293-1296.
- (11) Prytherch DR, Whiteley MS, Higgins B, Weaver PC, Prout WG, Powell SJ. POSSUM and Portsmouth POSSUM for predicting mortality. Physiological and Operative Severity Score for the enUmeration of Mortality and morbidity. *Br J Surg* 1998; 85(9):1217-1220.
- (12) Whiteley MS, Prytherch DR, Higgins B, Weaver PC, Prout WG. An evaluation of the POSSUM surgical scoring system. *Br J Surg* 1996; 83(6):812-815.
- (13) Bann SD, Sarin S. Comparative audit: the trouble with POSSUM. *J R Soc Med* 2001; 94(12):632-634.
- (14) Yip MK, Ng KJ. Risk-adjusted surgical audit with the POSSUM scoring system in a developing country. Physiological and Operative Severity Score for the enUmeration of Mortality and morbidity. *Br J Surg* 2002; 89(1):110-113.
- (15) Bennett-Guerrero E, Hyam JA, Shaefi S, Prytherch DR, Sutton GL, Weaver PC et al. Comparison of P-POSSUM risk-adjusted mortality rates after surgery between patients in the USA and the UK. *Br J Surg* 2003; 90(12):1593-1598.
- (16) Le Gall JR, Loirat P, Alperovitch A, Glaser P, Granthil C, Mathieu D et al. A simplified acute physiology score for ICU patients. *Crit Care Med* 1984; 12(11):975-977.

- (17) Le Gall JR, Lemeshow S, Saulnier F. A new Simplified Acute Physiology Score (SAPS II) based on a European/North American multicenter study. *JAMA* 1993; 270(24):2957-2963.
- (18) Keats AS. The ASA classification of physical status--a recapitulation. *Anesthesiology* 1978; 49(4):233-236.
- (19) Owens WD, Felts JA, Spitznagel EL, Jr. ASA physical status classifications: a study of consistency of ratings. *Anesthesiology* 1978; 49(4):239-243.
- (20) Cullen DJ, Apolone G, Greenfield S, Guadagnoli E, Cleary P. ASA Physical Status and age predict morbidity after three surgical procedures. *Ann Surg* 1994; 220(1):3-9.
- (21) Haynes SR, Lawler PG. An assessment of the consistency of ASA physical status classification allocation. *Anaesthesia* 1995; 50(3):195-199.
- (22) Hall JC, Hall JL. ASA status and age predict adverse events after abdominal surgery. *J Qual Clin Pract* 1996; 16(2):103-108.
- (23) Prause G, Ratzenhofer-Comenda B, Pierer G, Smolle-Juttner F, Glanzer H, Smolle J. Can ASA grade or Goldman's cardiac risk index predict peri-operative mortality? A study of 16,227 patients. *Anaesthesia* 1997; 52(3):203-206.
- (24) Mak PH, Campbell RC, Irwin MG. The ASA Physical Status Classification: inter-observer consistency. *American Society of Anesthesiologists. Anaesth Intensive Care* 2002; 30(5):633-640.
- (25) Sutton R, Bann S, Brooks M, Sarin S. The Surgical Risk Scale as an improved tool for risk-adjusted analysis in comparative surgical audit. *Br J Surg* 2002; 89(6):763-768.
- (26) Wouters SC, Noyez L, Verheugt FW, Brouwer RM. Preoperative prediction of early mortality and morbidity in coronary bypass surgery. *Cardiovasc Surg* 2002; 10(5):500-505.
- (27) Cook TM, Day CJ. Hospital mortality after urgent and emergency laparotomy in patients aged 65 yr and over. Risk and prediction of risk using multiple logistic regression analysis. *Br J Anaesth* 1998; 80(6):776-781.
- (28) Sutton R, Bann S, Brooks M, Sarin S. The Surgical Risk Scale as an improved tool for risk-adjusted analysis in comparative surgical audit. *Br J Surg* 2002; 89(6):763-768.
- (29) Khuri SF, Daley J, Henderson W, Barbour G, Lowry P, Irvin G et al. The National Veterans Administration Surgical Risk Study: risk adjustment for the comparative assessment of the quality of surgical care. *J Am Coll Surg* 1995; 180(5):519-531.
- (30) Grover FL, Shroyer AL, Hammermeister KE. Calculating risk and outcome: the Veterans Affairs database. *Ann Thorac Surg* 1996; 62(5 Suppl):S6-11.
- (31) Khuri SF, Daley J, Henderson W, Hur K, Gibbs JO, Barbour G et al. Risk adjustment of the postoperative mortality rate for the comparative assessment of the quality of surgical care: results of the National Veterans Affairs Surgical Risk Study. *J Am Coll Surg* 1997; 185(4):315-327.
- (32) Daley J, Forbes MG, Young GJ, Charns MP, Gibbs JO, Hur K et al. Validating risk-adjusted surgical outcomes: site visit assessment of process and structure. *National VA Surgical Risk Study. J Am Coll Surg* 1997; 185(4):341-351.

- (33) Daley J, Forbes MG, Young GJ, Charns MP, Gibbs JO, Hur K et al. Validating risk-adjusted surgical outcomes: site visit assessment of process and structure. National VA Surgical Risk Study. *J Am Coll Surg* 1997; 185(4):341-351.
- (34) Daley J, Khuri SF, Henderson W, Hur K, Gibbs JO, Barbour G et al. Risk adjustment of the postoperative morbidity rate for the comparative assessment of the quality of surgical care: results of the National Veterans Affairs Surgical Risk Study. *J Am Coll Surg* 1997; 185(4):328-340.
- (35) Khuri SF, Daley J, Henderson W, Hur K, Demakis J, Aust JB et al. The Department of Veterans Affairs' NSQIP: the first national, validated, outcome-based, risk-adjusted, and peer-controlled program for the measurement and enhancement of the quality of surgical care. National VA Surgical Quality Improvement Program. *Ann Surg* 1998; 228(4):491-507.
- (36) Daley J, Henderson WG, Khuri SF. Risk-adjusted surgical outcomes. *Annu Rev Med* 2001; 52:275-287.
- (37) Best WR, Khuri SF, Phelan M, Hur K, Henderson WG, Demakis JG et al. Identifying patient preoperative risk factors and postoperative adverse events in administrative databases: results from the Department of Veterans Affairs National Surgical Quality Improvement Program. *J Am Coll Surg* 2002; 194(3):257-266.
- (38) Indicadores de Qualidade Assistencial do HCPA. Hospital de Clínicas de Porto Alegre, editor. Sistema IG -Indicadores de Gestão do HCPA . 26-9-2004. Ref Type: Electronic Citation
- (39) Medlineplus Encyclopedia. Exploratory laparotomy. www.nlm.nih.gov/medlineplus/ency/article/002928.htm . 2004. Ref Type: Electronic Citation
- (40) Sistema de Indicadores Padronizados para Gestão Hospitalar. SIPAGEH, editor. www.sipageh.unisinos.br . 2004. UNISINOS, RS. Ref Type: Electronic Citation
- (41) Kazandjian VA, Lawthers J, Cernak CM, Pipesh FC. Relating outcomes to processes of care: the Maryland Hospital Association's Quality Indicator Project (QI Project). *Jt Comm J Qual Improv* 1993; 19(11):530-538.
- (42) Kazandjian VA, Wood P, Lawthers J. Balancing science and practice in indicator development: the Maryland Hospital Association Quality Indicator (QI) project. *Int J Qual Health Care* 1995; 7(1):39-46.
- (43) DATASUS. Movimento mensal de internação hospitalar. www.datasus.gov.br/ arquivo de dados-MS BBS . 25-9-2004. Ref Type: Electronic Citation
- (44) Kluck M, Prompt CA, Guimaraes JR. The role of hospital information systems in improving patient safety. *Medinfo* 2004; 2004(CD):1683.
- (45) Medlineplus Encyclopedia. Exploratory laparotomy. www.nlm.nih.gov/medlineplus/ency/article/002928.htm . 2004. Ref Type: Electronic Citation
- (46) Khuri SF, Daley J, Henderson W, Hur K, Gibbs JO, Barbour G et al. Risk adjustment of the postoperative mortality rate for the comparative assessment of the quality of surgical care: results of the National Veterans Affairs Surgical Risk Study. *J Am Coll Surg* 1997; 185(4):315-327.

- (47) Davis P, Jenkin G, Coope P, Blakely T, Sporle A, Kiro C. The New Zealand Socio-economic Index of Occupational Status: methodological revision and imputation for missing data. *Aust N Z J Public Health* 2004; 28(2):113-119.
- (48) Garson GD. Data imputation for missing values. www2.chass.ncsu.edu/garson/pa765/missing.htm . 11-8-2004. Ref Type: Electronic Citation
- (49) Heitjan DF. Annotation: what can be done about missing data? Approaches to imputation. *Am J Public Health* 1997; 87(4):548-550.
- (50) www.aifs.gov.au/growingup/pubs/dp3/imputation.pdf. Data imputation. Australian Government - Institute of Family Studies . 11-8-2004. Ref Type: Internet Communication
- (51) Hosmer DW, Lemeshow S. *Applied logistic regression*. 1st ed. New York, NY, USA: John Wiley & Sons, 1989.
- (52) Le Gall JR, Lemeshow S, Saulnier F. A new Simplified Acute Physiology Score (SAPS II) based on a European/North American multicenter study. *JAMA* 1993; 270(24):2957-2963.
- (53) Normand ST, Glickman ME, Sharma RG, McNeil BJ. Using admission characteristics to predict short-term mortality from myocardial infarction in elderly patients. Results from the Cooperative Cardiovascular Project. *JAMA* 1996; 275(17):1322-1328.
- (54) Hosmer DW, Hosmer T, Le Cessie S, Lemeshow S. A comparison of goodness-of-fit tests for the logistic regression model. *Stat Med* 1997; 16(9):965-980.
- (55) Arozullah AM, Khuri SF, Henderson WG, Daley J. Development and validation of a multifactorial risk index for predicting postoperative pneumonia after major noncardiac surgery. *Ann Intern Med* 2001; 135(10):847-857.
- (56) Le Gall JR, Lemeshow S, Saulnier F. A new Simplified Acute Physiology Score (SAPS II) based on a European/North American multicenter study. *JAMA* 1993; 270(24):2957-2963.
- (57) Barry WA, Rosenthal GE. Is there a July phenomenon? The effect of July admission on intensive care mortality and length of stay in teaching hospitals. *J Gen Intern Med* 2003; 18(8):639-645.
- (58) Keats AS. The ASA classification of physical status--a recapitulation. *Anesthesiology* 1978; 49(4):233-236.
- (59) Owens WD, Felts JA, Spitznagel EL, Jr. ASA physical status classifications: a study of consistency of ratings. *Anesthesiology* 1978; 49(4):239-243.
- (60) Haynes SR, Lawler PG. An assessment of the consistency of ASA physical status classification allocation. *Anaesthesia* 1995; 50(3):195-199.
- (61) Prause G, Ratzenhofer-Comenda B, Pierer G, Smolle-Juttner F, Glanzer H, Smolle J. Can ASA grade or Goldman's cardiac risk index predict peri-operative mortality? A study of 16,227 patients. *Anaesthesia* 1997; 52(3):203-206.
- (62) Prause G, Ratzenhofer-Comenda B, Pierer G, Smolle-Juttner F, Glanzer H, Smolle J. Can ASA grade or Goldman's cardiac risk index predict peri-operative mortality? A study of 16,227 patients. *Anaesthesia* 1997; 52(3):203-206.

- (63) Prause G, Ratzenhofer-Comenda B, Pierer G, Smolle-Juttner F, Glanzer H, Smolle J. Can ASA grade or Goldman's cardiac risk index predict peri-operative mortality? A study of 16,227 patients. *Anaesthesia* 1997; 52(3):203-206.
- (64) Mak PH, Campbell RC, Irwin MG. The ASA Physical Status Classification: inter-observer consistency. *American Society of Anesthesiologists. Anaesth Intensive Care* 2002; 30(5):633-640.
- (65) Muir AJ. Albumina sérica. <http://www.nlm.nih.gov/medlineplus/ency/article/003480.htm> . 2004. Ref Type: Electronic Citation
- (66) Vincent JL, Dubois MJ, Navickis RJ, Wilkes MM. Hypoalbuminemia in acute illness: is there a rationale for intervention? A meta-analysis of cohort studies and controlled trials. *Ann Surg* 2003; 237(3):319-334.
- (67) Gibbs J, Cull W, Henderson W, Daley J, Hur K, Khuri SF. Preoperative serum albumin level as a predictor of operative mortality and morbidity: results from the National VA Surgical Risk Study. *Arch Surg* 1999; 134(1):36-42.
- (68) Khuri SF, Daley J, Henderson W, Hur K, Gibbs JO, Barbour G et al. Risk adjustment of the postoperative mortality rate for the comparative assessment of the quality of surgical care: results of the National Veterans Affairs Surgical Risk Study. *J Am Coll Surg* 1997; 185(4):315-327.
- (69) Kudsk KA, Tolley EA, DeWitt RC, Janu PG, Blackwell AP, Yeary S et al. Preoperative albumin and surgical site identify surgical risk for major postoperative complications. *JPEN J Parenter Enteral Nutr* 2003; 27(1):1-9.
- (70) Khuri SF, Daley J, Henderson W, Hur K, Gibbs JO, Barbour G et al. Risk adjustment of the postoperative mortality rate for the comparative assessment of the quality of surgical care: results of the National Veterans Affairs Surgical Risk Study. *J Am Coll Surg* 1997; 185(4):315-327.
- (71) Clark TG, Altman DG. Developing a prognostic model in the presence of missing data: an ovarian cancer case study. *J Clin Epidemiol* 2003; 56(1):28-37.
- (72) Daley J, Iezzoni LI, Shwartz M. Conceptual and practical issues in developing risk-adjusted methods. In: Iezzoni LI, editor. *Risk adjustment for measuring health care outcomes*. Chicago, Illinois: Health Administration Press, 2003: 179-205.
- (73) Goldman L. Cardiac risk in noncardiac surgery: an update. *Anesth Analg* 1995; 80(4):810-820.
- (74) Hosmer DW, Lemeshow S. *Assessing the fit of the model. Applied Logistic Regression*. John Wiley & Sons, 2000: 143-202.
- (75) Harrell FE, Jr., Lee KL, Califf RM, Pryor DB, Rosati RA. Regression modelling strategies for improved prognostic prediction. *Stat Med* 1984; 3(2):143-152.
- (76) Bagley SC, White H, Golomb BA. Logistic regression in the medical literature: standards for use and reporting, with particular attention to one medical domain. *J Clin Epidemiol* 2001; 54(10):979-985.
- (77) Martins M, Travassos C, Noronha JC. Hospital Information Systems as risk adjustment in performance indicators. *Rev Saude Publica* 2001; 35(2):185-192.

- (78) Sistema de Indicadores Padronizados para Gestão Hospitalar. SIPAGEH, editor.
www.sipageh.unisinos.br . 2004. UNISINOS, RS. Ref Type: Electronic Citation

Considerações Finais

Esta tese teve como objetivo principal sistematizar e descrever uma metodologia para modelagem de dados para permitir o ajuste por risco de indicadores de desfechos hospitalares. Seus objetivos secundários foram testar esta metodologia através do desenvolvimento de um índice de risco para mortalidade após realização de laparotomia exploradora em pacientes do Hospital de Clínicas de Porto Alegre, analisando ao mesmo tempo a qualidade dos dados disponíveis no prontuário eletrônico do Hospital.

Vários trabalhos já foram publicados sobre índices de risco, principalmente na área de cirurgia. No entanto, muito pouco se tem feito no Brasil no sentido de aplicar sistemas de ajuste de risco para possibilitar a comparação efetiva da qualidade assistencial através de sistemas de indicadores padronizados, entre hospitais de diferentes características.

Através dos estudos realizados, que incluíram uma extensa revisão da literatura, a coleta e tratamento dos dados necessários, uma etapa cuidadosa e aprofundada de modelagem estatística através de regressão logística e a análise cuidadosa dos resultados obtidos, foi possível concluir que a metodologia proposta é adequada e exeqüível e que os dados disponíveis no prontuário eletrônico do HCPA são confiáveis, completos e abrangem um amplo espectro de informações.

A análise dos dados de pacientes submetidos ao procedimento de laparotomia exploradora permitiu a definição de um índice de risco de mortalidade cirúrgica simples, com apenas três variáveis explicativas, geralmente presentes nos registros hospitalares: a idade do paciente, a classificação pré-anestésica do seu estado físico (ASA) e o valor da albumina sérica antes da cirurgia. A classificação ASA, apesar de não ter caráter prognóstico para mortalidade cirúrgica, se revelou um preditor bastante acurado, demonstrando o valor desta informação (baseada apenas em dados da anamnese), muitas vezes considerada de pouco valor devido a seu caráter subjetivo. O índice de risco foi validado em outra amostra temporalmente independente, com excelentes resultados.

Espera-se que a metodologia proposta seja replicada por outros pesquisadores que atuam nesta linha de pesquisa, difundindo o ajuste de risco como técnica essencial na avaliação de desfechos de saúde. No entanto, os resultados alcançados com o índice de risco cirúrgico foram validados apenas em pacientes do HCPA e é necessário que os mesmos sejam validados com dados de pacientes de outros hospitais para que seu uso possa ser generalizado. Com a disponibilidade de dados existente e o domínio da metodologia, o índice de risco deverá ser também validado em pacientes submetidos a outros procedimentos cirúrgicos para que possa ser utilizado como ajustador de indicadores de mortalidade cirúrgica.