

Adrien Francais<sup>1</sup>, Aurélien Vesin<sup>2</sup>,  
Jean-François Timsit<sup>3</sup>

## Como realizar ensaios clínicos em terapia intensiva utilizando base de dados de alta qualidade

*How to conduct clinical research studies using high quality-clinical databases in the critical care*

1. INSERM U823/ Equipe 11 (Desfecho de Cancer e Doenças Graves); Centro de Coordenação de base de dados OUTCOMEREA do Instituto Albert Bonniot, La Tronche, França.  
2. INSERM U823/ Equipe 11 (Desfecho de Cancer e Doenças Graves); Centro de Coordenação de base de dados OUTCOMEREA do Instituto Albert Bonniot, La Tronche, França.  
3. INSERM U823/ Equipe 11 (Desfecho de Cancer e Doenças Graves); Centro de Coordenação de base de dados OUTCOMEREA do Instituto Albert Bonniot, La Tronche, França e Unidade de Terapia Intensiva do Hospital Universitário Albert Michallon, Grenoble, França

Recebido do Albert Bonniot Institute, La Tronche, France.

Submetido em 16 de maio de 2008  
Aceito em 10 de julho de 2008

### Endereço para correspondência:

Prof Jean-François Timsit  
INSERM U823/ Team 11 (Outcome of Cancer and Critical Illnesses),  
Coordinating Center of the OUTCOMEREA Database  
Albert Bonniot Institute, La Tronche  
38706 CEDEX; France  
Tel : 33(0)4 76 76 87 79  
Fax : 33(0)4 76 76 55 69  
Email: jf.timsit@outcomerea.org.

### RESUMO

As fontes de informação relacionadas à terapia intensiva e os meios de comunicá-la aumentam rapidamente. Neste artigo, apresentamos um panorama geral sobre o que deve ser feito para obter dados de alta qualidade nas unidades de terapia intensiva. Os princípios para a escolha da questão clínica da pesquisa, os desfechos,

as variáveis explanatórias e os métodos estatísticos a serem utilizados para abordar a questão são comentados de forma geral, com ênfase nos erros e ciladas mais frequentes que devem ser evitados.

**Descritores:** Sistemas de informação; Base de dados factuais; Pesquisa biomédica; Estudos de avaliação; Coleta de dados/métodos; Controle de qualidade

### INTRODUÇÃO

Hoje, a necessidade por informação exata útil não só na clínica e nos serviços de saúde mas também em auditorias clínicas e pesquisa de avaliação é sobejamente reconhecida. Obviamente, os que propõem cada uma destas atividades deveriam agregar seus recursos para obter os dados de que precisam. De fato, há um reconhecimento crescente sobre o valor em potencial dessas bases de dados, não só para auditorias clínicas mas também para empreender estudos clínicos e avaliar tecnologias de saúde.<sup>1</sup>

Para atingir estes objetivos, médicos, administradores, consumidores e pesquisadores devem definir as variáveis necessárias e propor definições completas e exatas com base nas definições padrão de distúrbios clínicos, intervenções e desfechos.

Bases de dados de alta qualidade ajudariam a fornecer estimativas atualizadas das probabilidades de desfechos diferentes nas situações clínicas típicas da unidade de terapia intensiva (UTI). Elas deveriam ajudar a determinar fatores de risco e modelos prognósticos para vários eventos e poderiam ser usadas para alocação de recursos entre regiões ou hospitais.

O grande número de pacientes incluídos nas bases de dados permite análises de subgrupos, que, por sua vez, podem dar aos médicos informações sobre categorias específicas de pacientes. Da mesma forma poderiam ser usadas para estudar doenças ou intervenções incomuns.

A pesquisa usando bases de dados clínicos de alta qualidade é relativamente barata, pois o corpo do sistema de coleta de dados é definido, seus custos estão distribuídos entre muitas pesquisas e são compartilhados com outras aplicações como gestão clínica, auditoria e usos administrativos.

Pesquisas numa base de dados clínicos não deveriam substituir ensaios clínicos

controlados e randomizados, mas podem, em nossa opinião, desempenhar um papel importante na demonstração de evidências.<sup>2</sup> Primeiro, poderiam gerar hipóteses sobre fatores de risco e formas de prevenir uma doença em particular. Segundo, ao aumentar o grau de incerteza entre os médicos, poderiam aumentar a probabilidade de médicos participarem de um ensaio clínico. Terceiro, fornecem uma infra-estrutura permanente para desenhar ensaios multicêntricos. Quarto, poderiam ser usadas para confirmar, numa população geral não selecionada de pacientes, as conclusões dos ensaios clínicos, muitas vezes feitos em UTI selecionadas, altamente especializadas e de ensino.

Neste texto, iremos descrever como construir uma base de dados de alta qualidade a ser usada para fins de pesquisa e para esquematizar as melhores formas de explorar os eventos e desfechos da UTI.

### **COMO OBTER UMA BASE DE DADOS DE ALTA QUALIDADE?**

O desfecho é medido por mortalidade, morbidade (eventos iatrogênicos, infecções nosocomiais), estado funcional, qualidade de vida, custo de atendimento, tempo de internação, tempo para retorno ao trabalho, satisfação do paciente e da família. Fatores que podem influenciar a qualidade dos dados do desfecho são os métodos de coleta de dados, padronização das definições, atualização dos dados e número adequado de pacientes e desfecho.

As grandes bases de dados têm numerosas falhas, inclusive não captar os fatores de risco pouco usuais e assim subestimar o risco de um dado paciente, falta de homogeneidade das diferentes populações de pacientes e a exatidão da aquisição de dados.

Propusemos algumas regras importantes<sup>3</sup> usada em bases de dados de UTI mais reconhecidas<sup>4-7</sup>.

- No início da base de dados

Os dados precisam ser exatamente definidos e as características dos dados devem estar num dicionário de dados. Deve ser montado um conjunto mínimo de dados.

O protocolo para a coleta de dados deve ser feito por escrito, em que conste a definição das possíveis falhas na coleta de dados, bem como a criação de um controle manual ou automatizado dos dados.

Por fim, o centro de coordenação central deveria criar um formulário "amigável" para registro dos caso usando de preferência um formato eletrônico.

De forma ideal, como a coorte deve representar a população geral da UTI, cada paciente deve ser registrado na base de dados.

As UTI que aceitarem participar no registro de dados

devem estar cientes de um plano para assegurar a qualidade e aceitar algumas regras:

Novos centros participantes devem ser visitados e treinados.<sup>8</sup> Além disso, os novos centros participantes e, periodicamente, os demais centros devem ser visitados para monitorar os dados e realizar auditorias da qualidade, usando para esta finalidade, os registros de casos ou fontes externas de dados.

Procedimentos de controle e checagem devem ser realizados ao importar os dados para a base de dados central.

- Comunicação com os centros

É importante que seja criada uma comunicação consistente e uniforme com aqueles envolvidos na coleta de dados.<sup>9</sup> Os dados são de propriedade coletiva e é importante que haja oportunidades de introduzir novas questões ou idéias. Também é importante reconhecer o trabalho dos profissionais (agradecimentos, autoria) e organizar uma realimentação dos dados. Regras de publicação referentes à autoria deveriam ser definidas logo no início da criação da base de dados e aceitas pelas novas UTI participantes.

- Programa de melhoria da qualidade

Relatórios da qualidade dos dados são fornecidos para os centros. Os centros deveriam verificar os erros detectados, corrigir dados inadequados e preencher dados incompletos. Após auditoria dos dados ou teste de variabilidade é importante dar um retorno e recomendações aos centros para solucionar as causas dos erros.

Mesmo para participantes bem treinados, deve ser organizado periodicamente um curso de treinamento para computar os mesmos registros de casos e identificar novas causas de discrepâncias. A causa e as conseqüências de erros deveriam ser discutidas e levar a uma exatidão maior das definições. É interessante que a motivação dos participantes aumenta após os cursos.

O centro coordenador central é responsável pelo ajuste dos questionários, software, dicionário, protocolos e material de treinamento para incluir novas variáveis ou resolver possíveis erros ou discrepâncias.

### **COMO ESCOLHER O DESENHO DO ESTUDO**

A base de dados de alta qualidade pode ser usada para fazer estudos de coorte. Poderia também ser o cerne de caso-controle aninhados ou estudos expostos - não-expostos. Um estudo de coorte é o desenho mais apropriado para conduzir estudos em UTI<sup>10</sup> porque o acompanhamento do paciente é fácil de coletar, pois a grande maioria dos eventos e fatores de risco acontecem durante a internação hospitalar. A inclu-

são sistemática de cada paciente que chega à UTI favorece a computação da prevalência dos desfechos.

Estudos de caso-controle são uma alternativa ao desenho de coorte comparando pacientes com um evento (por exemplo, morte) e outros sem, mas são raramente feitos na UTI.<sup>11,12</sup> A correlação de um caso com vários controles é possível para a prevenção contra fatores de confusão. Estudos de caso-controle em geral são mais rápidos e menos dispendiosos do que estudos de coorte, mas são muito sensíveis a um viés (como viés de memória e viés de seleção).<sup>10</sup>

## ESCOPO DO ESTUDO

O escopo de estudos embasados em bases de dados de UTI está em geral associado ao levantamento de fatores de risco de morte ou eventos adversos na unidade. Em casos mais raros, os estudos feitos em UTI visam prever os eventos que dependem de tempo na UTI<sup>13</sup> (isto é, número de infecções por ano) ou para modular a duração (isto é, tempo de internação).<sup>14</sup>

A primeira abordagem é acessar a viabilidade de um estudo garantindo que a frequência do evento de interesse seja suficientemente alta. Depois, é preciso calcular o poder estatístico. Este consiste em quantificar a capacidade de detectar associação de qualquer fator de risco com o evento de interesse.<sup>15</sup> Este dado pode ser computado usando softwares específicos.

## VARIÁVEL DE DESFECHO

Em estudo clínico as variáveis qualitativas mais frequentes são resultados binários (isto é, morto ou vivo). Sobrevida precoce (dia 14, dia 28), na maioria das vezes reflete o efeito de doença aguda, porém sobrevida tardia (3 meses ou um ano) reflete mais precisamente o impacto da doença para o paciente e a família.<sup>16</sup> Desfechos baseados em variáveis de tempo, como sobrevida em UTI ou hospitalar, devem ser desencorajados, pois levariam a vieses nas avaliações.

A sobrevida pode ser analisada em sua forma original (por regressão logística) ou completada por uma variável descrevendo o tempo-até-evento (análise de sobrevida). A primeira pode ser estendida para desenho de caso-controle, mas não considera tempo de exposição, ao contrário de outros métodos que permitem controlar o tempo passado antes da ocorrência (ou não) do evento. Nas análises de sobrevida, os pacientes são censurados se não sofrem o evento até deixarem o estudo ou até o término do estudo. Também os modelos padrão supõem a independência entre tempo censurado e tempo de evento (censura não informativa). Todavia, este poderosa pressuposto é frequentemente viola-

do, sobretudo em estudos na UTI, onde o tempo até alta e tempo até morte são totalmente dependentes.<sup>17,18</sup>

No caso de uma variável de interesse quantitativa, é aconselhável simplesmente traçar um histograma para tecer suposições sobre o padrão geral de distribuição (Figura 1). Medidas biológicas clássicas são muitas vezes distribuídas conforme a distribuição Gaussiana (normal) (p.ex. idade, *Simplified Acute Physiology Score* (SAPS II)) que é admitida em muitos testes estatísticos e permite usar estatísticas descritivas padrão (média e desvio padrão). Porém, em alguns casos as variáveis são distribuídas de outra forma (p.ex. tempo de internação ou número de eventos adversos). A média torna-se realmente diferente da mediana. Assim, são preferidos o primeiro e último quartil à média e desvio padrão que se tornam inadequados para a descrição dos dados com um intervalo de confiança absurdo (valores negativos). Caso contrário, a transformação da variável pode ser realizada para que a sua forma de distribuição se aproxime da normalidade (p.ex. transformação do log). No caso de distribuições atípicas e, sobretudo, quando há um padrão de “múltiplos picos”, sugere-se a divisão da variável em um número de classes iguais ao número de picos.<sup>19</sup>

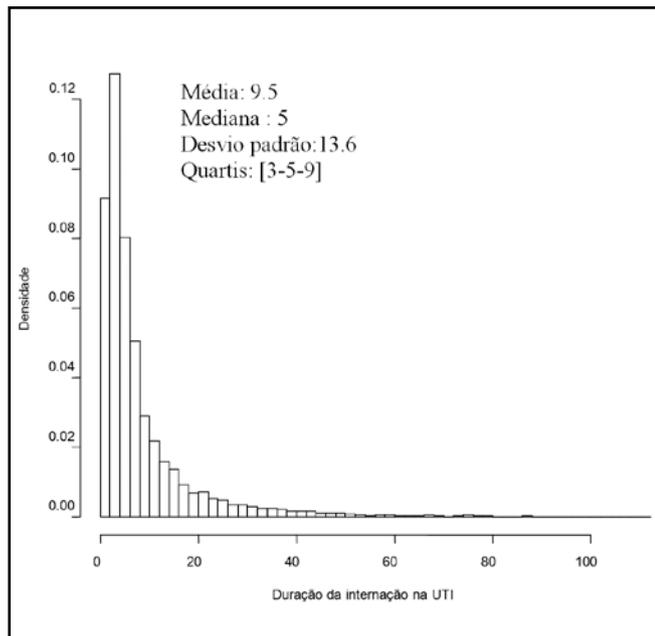
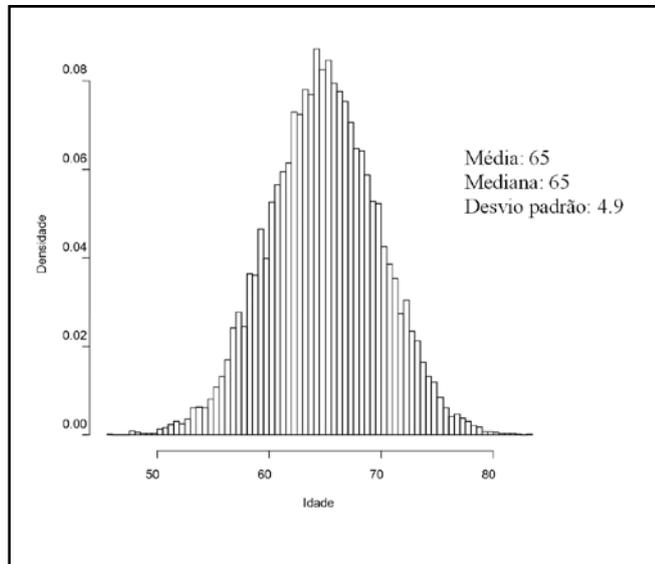
Em casos especiais, a variável de interesse pode ser uma medida subjetiva de um fenômeno (p.ex. o escore Mac Cabe de qualidade de vida).<sup>20</sup> É mais prudente fazer várias medições por sujeito para assegurar sua reprodutibilidade e, portanto, confiabilidade da análise.<sup>21,22</sup>

## VARIÁVEIS EXPLANATÓRIAS

A descrição dos dados de coorte é um trabalho essencial que permite verificar a consistência dos dados, frequência dos valores faltantes e mostrar as primeiras pistas interessantes a serem investigadas.

Técnicas de descrição de dados variam conforme os diferentes tipos de variáveis: quantitativas (p.ex. escore de gravidade como SAPS), qualitativas binárias (p.ex. gênero) ou nominais (p.ex. principais sintomas na UTI) e variáveis ordinárias (p.ex. códigos da American Society of Anesthesiology -ASA). Isto pode ser feito por tabelas de contingência (frequências e porcentagens) para variáveis categóricas ou média mediana, desvio padrão, quartis e histograma para as contínuas.

A contagem dos dados faltantes (“missing variable”) requer atenção especial para identificar as variáveis que estão mal relatadas. A multiplicidade dos valores faltantes nos dados implica deletar observações na análise estatística padrão. Caso haja mais que 5% de dados faltantes é aconselhável deixar a variável fora da análise.<sup>10</sup> Todavia, para tratar desta questão existem alguns métodos eficientes para completar os dados.<sup>23</sup>



**Figura 1: Distribuição das variáveis quantitativas.** O histograma de idade parece ter distribuição normal: a média e a mediana são idênticas e correspondem ao eixo simétrico de distribuição. Neste caso, tanto a média  $\pm$  desvio padrão ou mediana e quartis podem ser usados para descrever corretamente dados e intervalos de confiança podem ser facilmente computados. Ao contrário, a distribuição do tempo de internação na UTI é assimétrica e mesmo que a maioria das internações na UTI seja curtas, valores extremos podem ser identificados (p.ex. 90 dias). Neste caso, a média que está fortemente influenciada por valores extremos é inapropriada para descrever tempo de internação. Assim, as medianas e os quartis deveriam ser preferidos para descrever esta variável.

## RELAÇÃO ENTRE A VARIÁVEL DE DESFECHO E AS VARIÁVEIS EXPLANATÓRIAS

Ao investigar sobre associações entre variáveis explanatórias e variável de desfecho é preciso destacar os fatores de risco do evento de desfecho. Esta metodologia chamada “análise univariada” está descrita no Quadro 1. A escolha entre testes paramétricos e não paramétricos é possível e supõe (ou não) uma lei probabilística subjacente. O uso de testes paramétricos é sempre mais poderoso, porém nem sempre se aplica (Quadro 1). As variáveis assinaladas pela literatura, pela experiência e pelo bom senso precisam ser verificadas.

Avaliar o efeito de uma variável quantitativa num desfecho binário, em geral supõe que o aumento no risco do evento de interesse é constante seja qual for seu grau. Por exemplo, o aumento no risco de mortalidade no 28º dia é obviamente menor entre os com 15-25 anos do que entre os com 75-85 anos de idade. Essa suposição de linearidade do aumento de risco deve ser verificada por testes estatísticos (funções spline, log de risco).<sup>19</sup> Se o teste desta suposição for rejeitado para uma determinada variável, ela deverá ser recodificada em várias classes (p.ex. quartis ou pontos de corte de risco). Caso contrário, deve-se preferir a variável contínua para guardar toda a informação.

## O MÁGICO VALOR P

Mas qual limiar do valor p temos que escolher? A significância estatística se baseia numa escolha um tanto quanto arbitrária do grau (muitas vezes 5%). Este limiar psicológico de 5%, também chamado erro 5% tipo I (risco de concluir para uma diferença significativa entre dois grupos enquanto isto for de causa aleatória) nem sempre é relevante. Por exemplo, valores de p de 0,048 e 0,052 estão muito próximos, porém as conclusões são completamente diferentes. No primeiro caso concluímos para uma diferença significativa entre dois grupos e no segundo, para ausência de diferença entre grupos. Assim para cada caso crítico, vale mais a pena pensar no interesse médico do que apenas levar em conta a significância estatística.

Quando muitas variáveis são testadas, o risco de se concluir por uma associação significativa quando ela é aleatória aumenta com o número de testes feitos (i.e. múltiplas comparações).

O risco de erro comumente usado (valor p) é 5% para um teste, mas se forem feitos 4 testes, o risco chega a 18,5%. Portanto, um limiar mais baixo de valor p deve ser considerado. Isto pode ser feito usando métodos apropriados tais como a correção de Bonferroni.<sup>19,24</sup>

## COMO ESCOLHER UM MODELO APROPRIADO?

O grande número de ferramentas estatísticas disponíveis permite analisar qualquer tipo de variável e qualquer desenho do estudo. O conjunto de modelos estatísticos usados em pesquisa médica limita-se a alguns mais simples porque não demandam um conhecimento maior de estatística para compreendê-los e são considerados robustos.

Médicos precisam conhecer a condição de validade dos modelos e devem ser capazes de discutir com especialistas em estatística, contudo, nesta fase recomenda-se a ajuda de um

epidemiologista experiente ou de estatísticos.

Assim, apresentamos no Quadro 2 uma seleção de modelos que representam a vasta maioria do que é usado para analisar dados de UTI.<sup>25</sup>

Além dos já mencionados, existem outros modelos (ou extensões) que correspondem a desenhos de estudo mais complexos ou menos comuns. A modelagem em múltiplos níveis permite considerar a repetição de medidas em um mesmo assunto ou uma estrutura hierárquica nos dados (um paciente na UTI, muitas UTI em um país, muitos países no mundo...).<sup>26,27</sup> Previsão e séries temporais consistem em analisar a evolução de

**Quadro 1 - Metodologia para comparações das variáveis**

Tipo de variável de interesse	Tipo de variável testada	Metodologia	Teste estatístico	Principais pressupostos
Qualitativa (2 ou mais grupos)	Quantitativa	Estatística descritiva para cada grupo (media, erro padrão, mediana, quartis...)	2 grupos : teste T * Wilcoxon Grupos k I: ANOVA * Kruskall-Wallis	Para ANOVA, as variáveis dependentes deveriam ser distribuídas normalmente dentro de grupos e variâncias são idênticas.
	Qualitativa	Tabela 2-lados (tabulação cruzada) das frequências e porcentagens	Qui quadrado Exato de Fisher	Para o teste Qui-quadrado as frequências esperadas em cada célula da tabela devem ser maiores que 5.
Quantitativa	Quantitativa	Coeficiente de correlação	Correlação de Pearson* Correlação de Spearman	Identificam uma relação linear entre as variáveis de interesse e explanatórias.
	Qualitativa	Idêntica à primeira parte desta tabela (relação entre variáveis qualitativas e quantitativas)		

\*Teste paramétrico. ANOVA – análise de variância

**Quadro 2 - Diferentes modelos estatísticos**

Tipos de variáveis de interesse	Modelo a ser usado	Principais suposições estatísticas a serem verificadas	Outros métodos possíveis (não exaustivo)
Binário (p.ex. morto vs. vivo; doente vs. saudável)	Regressão logística	Há uma relação log-linear entre as variáveis independentes e as probabilidades da resposta. As variáveis preditivas são independentes entre si.	
Qualitativo não estruturado (p.ex. diagnóstico)	Politômico regressão logística	Há uma relação log-linear entre as variáveis independentes e as probabilidades da resposta. As variáveis preditivas são independentes entre si..	Árvores de decisão Análise discriminante
Qualitativo estruturado (p.ex. score ADL de qualidade de vida)	Regressão de Poisson	Média da variável resposta igual à variância. Variáveis preditivas são independentes.	Regressão binomial negativa Regressão logística politômica ordinal
Contínuo (p.ex. taxa de plaquetas no sangue)	Regressão linear	Relação linear entre as variáveis explanatórias e a variável resposta. Os residuais (valores preditos menos observados) são normalmente distribuídos. Variáveis preditivas são independentes.	
Dados censurados (p.ex. ocorrência de morte na UTI)	Modelo de risco proporcional de Cox	Suposições de risco proporcional. Há uma relação log-linear entre as variáveis independentes e a função de risco subjacente O censor deve ser não-informativo (veja texto). Frequência de eventos deve ser >5%	

ADL – Atividades da vida diária, UTI – unidade de terapia intensiva

uma série de valores correlacionados no tempo (p.ex. modelos ARIMA).<sup>28</sup> Modelos de sobrevivência mais amplos podem integrar covariáveis tempo-dependentes ou contemplar informações sobre a censura.<sup>29,30</sup>

## SELEÇÃO DE VARIÁVEIS AO CONSTRUIR UM MODELO

Na análise univariada, nós assinalamos todas as associações que existem com a variável de desfecho sem considerar se algumas variáveis explanatórias estavam relacionadas entre si. No extremo oposto, a análise multivariada supõe independência dos preditores, assim não devem ser inseridas variáveis contendo informação semelhante (p.ex. dois escores de disfunção de órgãos como *Logistic Organ Dysfunction score* (LODS) e *Sequential Organ Failure Assessment* (SOFA)).<sup>19</sup> Existem alguns algoritmos de seleção (para trás, para frente, e passo a passo) para ajudar na escolha das variáveis a serem incluídas no modelo. Elas selecionam automaticamente o melhor subconjunto de variáveis que estão independentemente associadas a variável de interesse. Estes procedimentos estão implementados na maioria dos softwares de análise estatística (SAS, Stata, SPSS, R...). Todavia, é bem melhor selecionar as variáveis que devem ser introduzidas no modelo com base em regras médicas e não estatísticas, como a publicação prévia na literatura, a facilidade e reprodutibilidade da coleta de dados e ausência de dados faltantes.

A construção do modelo será diferente de acordo com seu intuito. Um modelo preditivo requer a inclusão de numerosas variáveis explanatórias para aumentar a capacidade preditiva (em geral limiares do valor p de 5 a 10%). Ao contrário, modelos visando buscar fatores de risco independentes precisam de uma seleção mais rígida de preditores (valores p mais baixos para inclusão e permanência no modelo: em geral limiares de 1 a 5%). De fato, as variáveis podem ser forçadas num modelo se forem consideradas essenciais para a interpretação do modelo (p.ex. ajuste do número de eventos adversos ao tempo de internação em uma regressão logística). Há que ter muito cuidado ao introduzir variáveis de modalidades atípicas ou informação subjetiva. Isto aumenta a variabilidade e leva a avaliações não confiáveis do risco associado.<sup>31</sup>

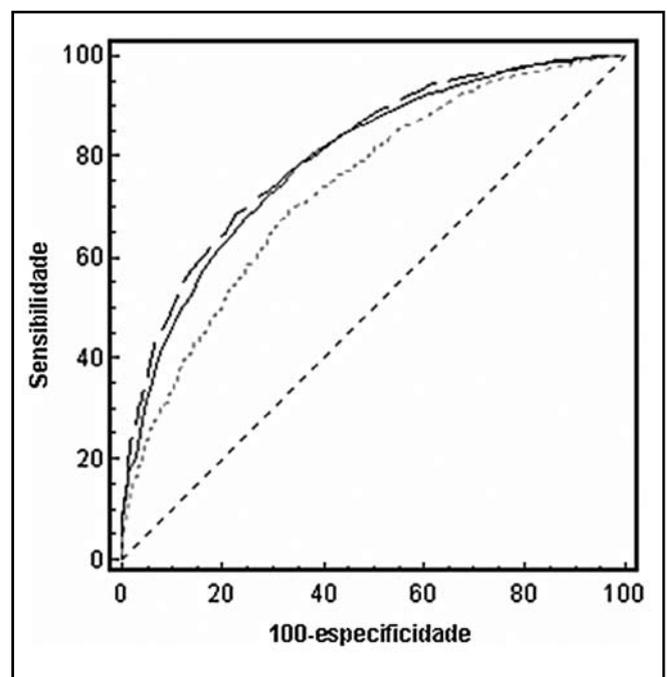
Por fim, introdução de termos de interação com motivação clínica deveria ser testada (como interação entre tratamento e gravidade da doença sendo tratada) pela adição de termos de produto cruzado.<sup>19</sup> Termos de interação aumentam rapidamente se o número de variáveis introduzidas no modelo aumentar. Para 3 variáveis há  $2^3-3=5$  termos de interação, para n variáveis há  $2^n-n$  termos de interação. Portanto, recomenda-se testar apenas termos de interação plausíveis e compreensíveis (bicaudados).

Finalmente, não há um modelo universal melhor. Assim, a escolha do modelo definitivo deve ser motivada pela facilidade de uso por médicos (dados rapidamente disponíveis na unidade, simplicidade de cálculo e interpretação) e a capacidade dos dados se adequarem ao modelo.

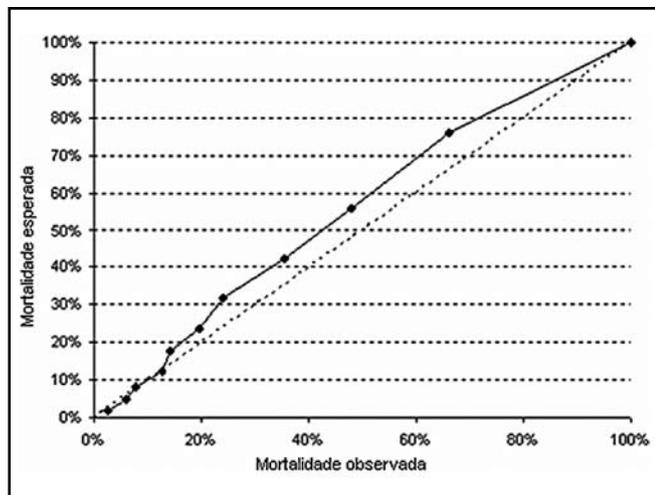
## AJUSTE E VALIDAÇÃO DO MODELO

A validação do modelo é uma parte essencial do processo de desenvolvimento do modelo, destina-se a confirmar sua eficácia e, assim, sua utilidade no processo decisório. A qualidade do modelo prognóstico padrão pode ser avaliada por discriminação e calibração.<sup>19,32</sup>

Num caso de desfecho binário, a área sob uma curva “*receiver operating characteristics*” (ASC-ROC) avalia a sensibilidade e a especificidade do modelo para prever o desfecho seja qual for o limiar de positividade. O valor de 0,5 indica que não há discriminação preditiva (isto é, ao acaso) e um valor de 1 indica a discriminação perfeita de pacientes com desfechos diferentes. Uma discriminação está adequada a partir de 0,80.<sup>19</sup> Um exemplo de curva ROC é descrita na Figura 2.



**Figura 2 – Exemplos de curvas Receiver Operating Curves (ROC) para três escores de gravidade (MPM II 0, MPM III 0, SAPS 3) para desfecho de mortalidade, do estudo OUTCOMEREA.** As três curvas ROC representam o gráfico de sensibilidade versus 100-especificidade enquanto se muda o valor do limiar de discriminação nos escores. A área sob a curva (ASC) representa a capacidade de discriminação dos escores de gravidade que é comparada à área sob a linha diagonal (capacidade de discriminação do aleatório = 0,5).



**Figura 3 – Calibração do escore SAPS 3 (do estudo OUTCOMERIA).** O gráfico de calibração representa a taxa mortalidade predita, estratificada por intervalos de 10% de risco de mortalidade pela taxa de mortalidade observada. Em um modelo bem calibrado a taxa esperada de mortalidade num grupo deveria corresponder à de mortalidade observada (linha diagonal). Esta hipótese pode ser avaliada pelo teste de Hosmer e Lemeshow.

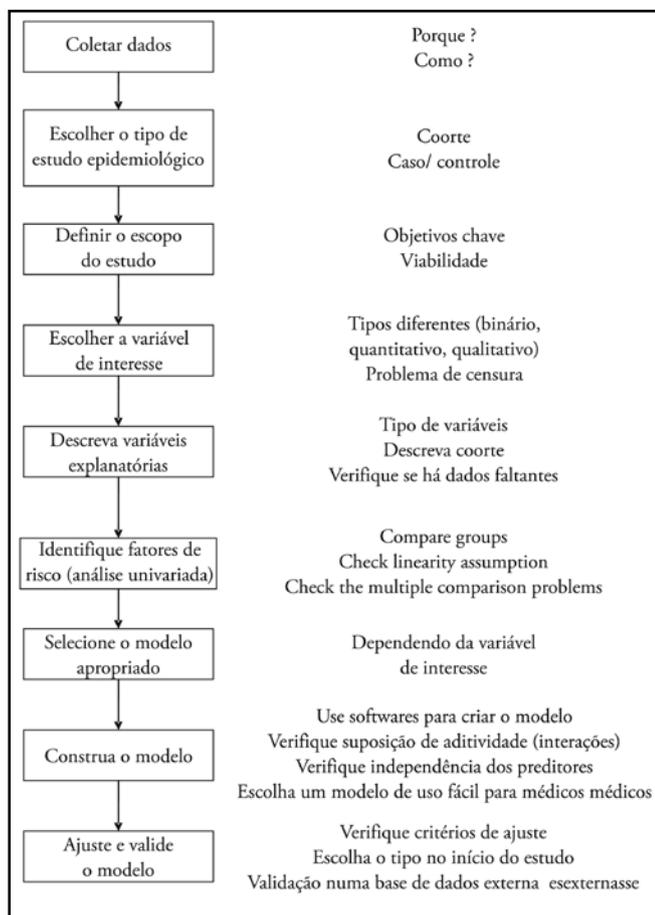
O componente de calibração (ou ajuste) descreve globalmente a exatidão preditiva ao controlar o viés entre respostas preditas e observadas (veja detalhes do teste Qui quadrado Hosmer-Lemeshow).<sup>32</sup> Por exemplo, se a mortalidade média predita para um grupo de pacientes semelhantes for 0,3 e a proporção real morrendo for 0,3 as previsões estão bem calibradas. Pode ser traçado um gráfico como na Figura 3.

Validação de um modelo prognóstico pode ser feita usando a mesma coorte (interna: “bootstrap” de separação de dados e validação cruzada) ou usando pacientes de outra coorte (externa). O segundo método é mais confiável, porém mais difícil de ser feito na prática. Detalhes sobre esta técnica estão disponíveis na literatura<sup>19,33</sup> e esta escolha deve ser feita no início do estudo.

### CONCLUSÃO

Um fluxograma resumindo as diferentes etapas desenvolvidas neste documento e que deveriam ser seguidas para conduzir uma análise multivariada em UTI pode ser visto na Figura 4.

Hoje mais e mais dados, ferramentas estatísticas, revistas ou informações científicas estão disponíveis na Internet. Para chegar a uma conclusão adequada a partir de um estudo, deve-se decidir criteriosamente a metodologia exata a ser usada antes de empreender qualquer análise de dados, para reduzir o risco de conclusões aleatórias ou vieses. A colaboração entre médicos de UTI e epidemiologistas e/ou bioestatísticos é imprescindível para evitar comparações excessivas e inúmeras ciladas.



**Figura 4 – Etapas a serem seguidas para conduzir uma análise multivariada**

### ABSTRACT

The sources of intensive care-related information and the means of communication increase rapidly. We presented here an overview of what should be done to collect high quality database. In a second part, the principle of the choice of the research question, the outcome, the explanatory variables and the statistical methods to address the question are overviewed, emphasizing major and frequent pitfalls which should be avoided.

**Keywords:** Information systems; Database, factual; Biomedical research; Evaluation studies; Data collection/methods; Quality control

## REFERENCES

01. Black N. High-quality clinical databases: breaking down barriers. *Lancet*. 1999; 353(9160):1205-6. Comment in: *Lancet*. 1999; 354(9172):75.
02. McKee M, Britton A, Black N, McPherson K, Sanderson C, Bain C. Methods in health services research. Interpreting the evidence: choosing between randomised and non-randomised studies. *BMJ*. 1999; 319(7205):312-5.
03. Arts DG, De Keizer NF, Scheffer GJ. Defining and improving data quality in medical registries: a literature review, case study, and generic framework. *J Am Med Inform Assoc*. 2002; 9(6): 600-11. Review.
04. Aegerter P, Auvert B, Buonamico G, Sznajder M, Beauchet A, Guidet B, et al. [Organization and quality control of a clinical database on intensive care medicine in central and suburban Paris]. *Rev Epidemiol Sante Publique*. 1998; 46(3): 226-37. French.
05. Harrison DA, Brady AR, Rowan K. Case mix, outcome and length of stay for admissions to adult, general critical care units in England, Wales and Northern Ireland: the Intensive Care National Audit & Research Centre Case Mix Programme Database. *Crit Care*. 2004; 8(2): R99-111.
06. Stow PJ, Hart GK, Higlett T, George C, Herkes R, McWilliam D, Bellomo R; for the ANZICS Database Management Committee. Development and implementation of a high-quality clinical database: the Australian and New Zealand Intensive Care Society Adult Patient Database. *J Crit Care*. 2006; 21(2): 133-41.
07. Timsit JF, Fosse JP, Troché G, De Lassece A, Alberti C, Garrouste-Orgeas M, Bornstain C, Adrie C, Cheval C, Chevret S; For the OUTCOMEREA Study Group, France. Calibration and discrimination by daily logistic organ dysfunction scoring comparatively with daily sequential organ failure assessment scoring for predicting hospital mortality in critically ill patients. *Crit Care Med*. 2002; 30(9): 2003-13. Comment in: *Crit Care Med*. 2002; 30(9):2151.
08. Arts DG, Bosman RJ, de Jonge E, Joore JC, de Keizer NF. Training in data definitions improves quality of intensive care data. *Crit Care*. 2003; 7(2):179-84.
09. Proctor SJ, Taylor PR. A practical guide to continuous population-based data collection (PACE): a process facilitating uniformity of care and research into practice. *QJM*. 2000; 93(2):67-73.
10. Bowers D. *Medical Statistics from Scratch*. New York: John Wiley & Sons; 2003.
11. Adrie C, Azoulay E, Francois A, Clec'h C, Darques L, Schwebel C, Nakache D, Jamali S, Goldgran-Toledano D, Garrouste-Orgeas M, Timsit JF; OutcomeRea Study Group. Influence of gender on the outcome of severe sepsis: a reappraisal. *Chest*. 2007; 132(6):1786-93.
12. de Lassece A, Timsit JF, Tafflet M, Azoulay E, Jamali S, Vincent F, Cohen Y, Garrouste-Orgeas M, Alberti C, Dreyfuss D; OUTCOMEREA Study Group. Pneumothorax in the intensive care unit: incidence, risk factors, and outcome. *Anesthesiology*. 2006; 104(1):5-13.
13. Moine P, Timsit JF, De Lassece A, Troché G, Fosse JP, Alberti C, Cohen Y; OUTCOMEREA Study Group. Mortality associated with late-onset pneumonia in the intensive care unit: results of a multi-center cohort study. *Intensive Care Med*. 2002; 28(2):154-63.
14. Perez A, Chan W, Dennis RJ. Predicting the length of stay of patients admitted for intensive care using a first step analysis. *Health Serv Outcomes Res Methodol*. 2007; 6(3-4):127-38.
15. Machin D, Campbell MJ, Fayers PM, Pinol APY. *Sample size tables for clinical studies*. 2a ed. Oxford: Blackwell Science; 1997.
16. Angus DC, Laterre PF, Helterbrand J, Ely EW, Ball DE, Garg R, Weissfeld LA, Bernard GR; PROWESS Investigators. The effect of drotrecogin alfa (activated) on long-term survival after severe sepsis. *Crit Care Med*. 2004; 32(11):2199-206. Comment in: *Crit Care Med*. 2004; 32(11):2347. *Crit Care Med*. 2004; 32(11):2348-9.
17. Chevret S. Logistic or Cox model to identify risk factors of nosocomial infection: still a controversial issue. *Intensive Care Med*. 2001; 27(10):1559-60. Comment on: *Intensive Care Med*. 2001; 27(8):1254-62.
18. Cox D. Regression models and life-tables. *J Royal Stat Soc*. 1972; 34(2):187-220.
19. Harrell FE, Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*. 1996; 15(4):361-87.
20. Afessa B, Keegan MT. Predicting mortality in intensive care unit survivors using a subjective scoring system. *Crit Care*. 2007; 11(1): 109. Comment on: *Crit Care*. 2006;10(6):R179.
21. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas*. 1960; 20(1):37-46.
22. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. 1979; 86(2):420-8.
23. Molenberghs G, Kenward M. *Missing data in clinical studies*. New York: Wiley InterScience; 2007.
24. Westfall PH, Tobias RD. *Multiple comparisons and multiple tests using SAS system*. SAS Publishing; 2003.
25. Katz MH. *Multivariable analysis: a practical guide for clinicians*. 2nd ed. Cambridge; New York: Cambridge University Press; 2006.
26. Twisk JWR. *Applied multilevel analysis: a practical guide*.

- New York: Cambridge University Press; 2006.
27. Goldstein H, Browne W, Rasbash J. Multilevel modelling in medical data. *Stat Med.* 2002; 21(21):3291-315.
  28. Charbonneau P, Parienti JJ, Thibon P, Ramakers M, Daubin C, du Cheyron D, Lebouvier G, Le Coutour X, Lelercq R; French Fluoroquinolone Free (3F) Study Group. Fluoroquinolone use and methicillin-resistant *Staphylococcus aureus* isolation rates in hospitalized patients: a quasi experimental study. *Clin Infect Dis.* 2006; 42(6):778-84. Comment in: *Clin Infect Dis.* 2006; 42(6):785-7.
  29. Kim HT. Cumulative incidence in competing risks data and competing risks regression analysis. *Clin Cancer Res.* 2007;13(2 Pt 1): 559-65.
  30. Resche-Rigon M, Azoulay E, Chevret S. Evaluating mortality in intensive care units: contribution of competing risks analyses. *Crit Care.* 2005; 10(1):R5. Comment in: *Crit Care.* 2006; 10(1):103.
  31. Spiegelman D, Carroll RJ, Kipnis V. Efficient regression calibration for logistic regression in main study/internal validation study designs with an imperfect reference instrument. *Stat Med.* 2001;20(1):139-60.
  32. Hosmer DW, Lemeshow S. *Applied logistic regression.* New York: Wiley; 1989.
  33. Altman DG, Royston P. What do we mean by validating a prognostic model? *Stat Med.* 2000; 19(4):453-73.