

## 07.07 Métodos de medição de casuística e ajuste de severidade



O ajuste de riscos é uma metodologia para desenvolver comparações de resultados ou custos entre centros de saúde que atendem pacientes de diferente gravidade (casuística, severidade)

**Autor: Salvador Peiró**

*Área de Pesquisa em Serviços de Saúde, Centro Superior de Pesquisa em Saúde Pública (CSISP), Valência.*

Se recomienda imprimir 2 páginas por hoja

### Citación recomendada:

Peiró S. Métodos de medição de casuística e ajuste de severidade [Internet]. Madrid: Escuela Nacional de Sanidad; Traducción 2015 [consultado día mes año]. Disponible en: direccion url del pdf.



TEXTOS DE ADMINISTRACION SANITARIA Y GESTIÓN CLINICA  
by UNED Y ESCUELA NACIONAL DE SANIDAD  
is licensed under a Creative Commons  
Reconocimiento- No comercial-Sin obra Derivada  
3.0 Unported License.



### Resumo

Comparar serviços de saúde não é tarefa fácil, porque seus resultados clínicos ou custos assistenciais dependem, além da qualidade ou de sua eficiência, das características dos pacientes que atendem. Para realizar estas comparações, os resultados obtidos por cada serviço devem ser **ajustados** segundo as características relevantes dos pacientes atendidos e que determinam seu **risco prévio (severidade, gravidade)** de chegar ao resultado concreto que será avaliado. Ajustar, neste

contexto, significa ponderar cada paciente pelo risco decorrente de suas características clínicas, demográficas entre outras.

## Introdução

### 1. Marco conceitual

#### 1.1. Controle

#### 1.2. Características

### 2. Aspectos metodológicos

#### 2.1 Definição do objetivo do sistema

#### 2.2. Seleção dos fatores de risco

#### 2.3. Construção e desenvolvimento

#### 2.4 Validação

#### 2.5 Rendimento

### 3. Uso prático

### 4. Perspectivas

## Bibliografia

Esta é a base para a construção dos chamados **sistemas de classificação de pacientes** ou **sistemas de ajuste de riscos** que, simplificando, não são mais do que **sistemas que quantificam a probabilidade que os pacientes têm de alcançar um determinado resultado clínico ou de custos**. Alguns destes sistemas, como os Grupos de Diagnósticos Relacionados (DRG, *Diagnosis-Related Group*), são amplamente utilizados para avaliar a eficiência dos hospitais. Outros tentam avaliar sua qualidade ou estabelecer um *ranking* de hospitais em função de diversas dimensões.

## Introdução

A comparação de organizações de saúde proporciona importante oportunidade para identificar possibilidades de melhorias. Embora existam outras possibilidades, na prática, costuma-se comparar:

- **Resultados clínicos:** mortalidade, complicações, reinternações, reintervenções, etc.
- **Processos:** taxa de cesarianas, porcentagem de pacientes que receberam determinados tratamentos etc.
- **Medidas de produtividade,** tanto na forma de processos por unidade de recurso (por exemplo, o índice de rotatividade, a estadia média, a porcentagem de ocupação) como de custos por processo.

A comparação de organizações de saúde é complexa porque, além de sua própria qualidade ou eficiência, ela também depende das características dos pacientes que foram atendidos em cada serviço: um hospital que tenha 10% de internação para cirurgia de catarata e 1% de internação oncológica terá menos mortes e menores custos do que outro que tenha 10% de oncológica e 1% de catarata. Inclusive dentro de um mesmo tipo de pacientes, as diferenças na gravidade podem provocar importantes diferenças nas taxas de mortalidade, custo por processo ou outros indicadores.

A literatura sobre serviços de saúde está repleta de exemplos de comparações mecanicistas realizadas entre organizações (**quadro 1**) que levaram a julgamentos temerários e que, por resultado, colocaram em cheque a utilidade de comparações simplistas para a tomada de decisão na organização do atendimento em saúde.

As características - demográficas, clínicas e outras - dos pacientes definem seu **risco prévio** (*antes de chegar ao serviço de saúde*), **de obter o resultado concreto que será medido**, além da qualidade do atendimento prestado. Este risco prévio é chamado de **gravidade** ou severidade<sup>1</sup>.

#### **Quadro 1. Comparações de mortalidade entre hospitais nos Estados Unidos (1986)**

Entre 1986 e 1993 a Administração Federal dos EUA publicou uma estatística de mortalidade hospitalar com o objetivo de oferecer aos usuários uma orientação sobre a "qualidade" de cada hospital, para que pudessem optar pelos serviços de forma mais informada.

Na estatística publicada em 1986 (a primeira delas, que incluía mais de 5000 hospitais, e já utilizava um sistema de ajuste de riscos), para o hospital com piores resultados a mortalidade foi de 87,6%, enquanto a mortalidade esperada conforme a média ajustada pelo tipo de pacientes atendidos correspondia a 22%.

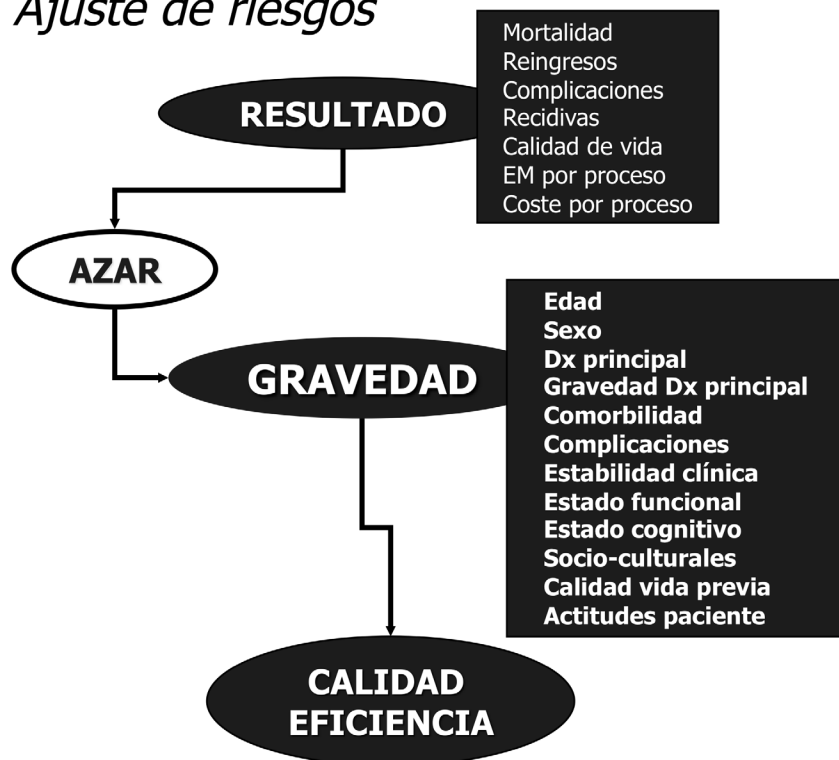
Os dados, tão evidentes, implicariam um grave problema de qualidade, exceto pelo fato de que se tratava de um hospital para pacientes terminais. Todos os pacientes faleciam, o que não significava especificamente um problema de qualidade. O erro evidente é que aqueles que dão entrada em um hospital de pacientes agudos não têm a mesma probabilidade de óbito que os pacientes de hospitais terminais. Embora fossem do mesmo sexo, idade, diagnóstico e tivessem a mesma comorbidade, havia outros fatores que os diferenciavam (uns eram terminais e outros não), fatores que não foram considerados nesta comparação.

<sup>1</sup> Na Espanha a tradução (incorreta) de severity por "severidade" (em vez de "gravidade") é tão popular que neste tema as duas são usadas como sinônimos.

A gravidade, neste contexto, não é (unicamente) equivalente a risco de morte, mas sim à **probabilidade prévia (decorrente das características do paciente e não do atendimento recebido) de obter o resultado que será medido**, que pode ser mortalidade ou qualquer outra ocorrência adversa, mas também qualidade de vida, consumo de recursos, etc. Este marco pode ser representado por meio de um esquema (**figura 1**) no qual o risco decorrente das características prévias do paciente (gravidade), mais a efetividade e a eficiência do provedor, tendo em conta o papel do acaso, levarão a diferentes resultados clínicos ou de custos.

Probabilidade prévia (decorrente das características do paciente e não do atendimento recebido) de obter o resultado que será medido

**Figura 1.**  
**Ajuste de riesgos**



Considerando que o objetivo da avaliação é comparar a qualidade ou eficiência dos serviços de saúde, a abordagem das diferenças de gravidade é fundamental, visto que se deve garantir que as diferenças encontradas nos resultados (incluindo os custos por processo) sejam decorrentes de fato do desempenho apresentado, e não de diferenças entre os pacientes atendidos.

Para isso, a metodologia usada prevê ajustar os resultados obtidos em cada serviço, conforme as *características relevantes* dos pacientes atendidos e que definem seu risco prévio em

relação ao resultado concreto que se deseja medir. Ajustar, neste contexto, significa ponderar cada paciente pelo risco que apresentam conforme suas características clínicas, demográficas e outras.

O pressuposto subjacente a esta abordagem é que se os pacientes tratados fossem iguais (paciente hipotéticos com valores médios em todos os seus fatores de risco), os resultados do atendimento também deveriam ser iguais e as possíveis diferenças entre resultados – se ocorressem – seriam atribuíveis a problemas de qualidade ou de eficiência. Dessa forma, se construirmos grupos de pacientes com riscos similares em relação ao resultado que se deseja medir, estes resultados serão comparáveis para tais grupos e, por sua vez e mediante sua ponderação, permitirão a comparação entre serviços de saúde.

Esta é a base para a construção de **sistemas de classificação de pacientes** (SCP) que, simplificando, não são mais do que **sistemas que quantificam a probabilidade que os pacientes têm de obter um determinado resultado**. A denominação de SCP não é a mais adequada para os métodos de padronização de riscos que não constroem grupos ou classificações, e na literatura vem sendo substituída pela denominação de **sistemas de ajuste de riscos** (SAR), reunindo em um mesmo grupo outros termos com conceitos pouco definidos, como casuística (*case mix*), gravidade (*severity*), doença (*sickness*), intensidade (*intensity*), complexidade (*complexity*), comorbidade (*comorbidity*) ou carga de doença (*burden of disease*).

Desde que em 1983 a *Health Care Financing Administration* (HCFA)<sup>2</sup>, decidiu utilizar um destes sistemas, os **Grupos de Diagnósticos Relacionados** (DRG), como base de um sistema de pagamento prospectivo para o reembolso das hospitalizações (excluídos os honorários médicos e alguns outros itens, como as próteses) os SAR se multiplicaram, tendo a literatura sobre o assunto crescido extraordinariamente. Assim, não é difícil encontrar excelentes revisões sobre o tema, incluindo alguns trabalhos espanhóis.

<sup>2</sup> A HCFA era o organismo encarregado de fazer o reembolso das organizações sanitárias e dos profissionais que prestam serviços ao Medicare, o programa público de saúde para maiores de 65 anos nos EUA.

## 1.- Marco conceitual para o ajuste de riscos

O marco conceitual para o ajuste de riscos inclui diversos elementos: o tipo de resultado que se quer medir, o tipo de serviços a serem comparados, as variáveis que irão definir a gravidade dos pacientes (severidade), as fontes de dados e a modelagem estatística.

### 1.1. Controle dos sistemas de ajuste de riscos

Na **tabela 1** são apresentadas as variáveis utilizadas pelo **Mortality Probability Model II (MPM-II)**, um sistema de ajuste de riscos utilizado em unidades de terapia intensiva (UTI). Os autores do MPM analisaram os dados de milhares de pacientes que deram entrada neste tipo de unidade e avaliaram que, por exemplo, a presença de cirrose tinha o efeito independente de aumentar em até 3 vezes (200%) o risco de morte nos primeiros 30 dias após a admissão, ou que as internações clínicas ou cirúrgicas urgentes tinham 130% mais risco de ir a óbito que as cirúrgicas programadas (veja na **tabela 1** a coluna das *odds ratio*, OR). Mediante uma análise estatística, os autores do MPM-II estimaram os coeficientes de regressão (coeficientes beta) associados a cada variável, de modo que somados a uma constante *y* e calculando o *logit* permitem estimar a probabilidade de morte de cada paciente.

**Tabela 1. VARIÁVEIS, COEFICIENTES E ODDS DO MPM II<sub>24</sub>**

		Beta	OR
Idade ( <i>Odds Ratio</i> cada 10 anos)	10 anos	0,03268	1,4
Cirrose	sim/não	1,08745	3,0
Efeito de massa intracraniana	sim/não	0,91314	2,5
Neoplasia maligna com metástase	sim/não	1,16109	3,2
Internação clínica ou cirúrgica de urgência	sim/não	0,83404	2,3
Coma ou estupor profundo (nas 24 h. após a admissão)	sim/não	1,68790	5,4
Creatinina > 2 mg/dl (nas 24 h. após a admissão)	sim/não	0,72283	2,1
Infecção confirmada (nas 24 h. após a admissão)	sim/não	0,49742	1,6
Ventilação mecânica (nas 24 h. após a admissão)	sim/não	0,80845	2,2
PPO2 < 60 mmHg (nas 24 h. após a admissão)	sim/não	0,46677	1,6
Tempo protrombina > 3 "s/est. (nas 24 h. após a admissão)	sim/não	0,55352	1,7
Urina < 150 ml em 8 h. (nas 24 h. após a admissão)	sim/não	0,82286	2,3
Fármacos vasoativos > 1 hora i.v. (nas 24 h. após a admissão)	sim/não	0,71628	2,0

CONSTANTE

-5,64592

Multiplicar cada coeficiente pelo valor 1 se a variável for afirmativa e 0 se for negativa e por grupos de 10 anos para a idade. Os produtos anteriores são somados para obter o *logit* e a probabilidade de morte é obtida pela fórmula  $e^{\text{logit}}/1+e^{\text{logit}}$ . OR: *Odds ratio*; Baseado em: Rue Monne M et al, 1996.



Suponhamos (**tabela 2**) um paciente de 70 anos, não cirúrgico, que após 24h da admissão na UTI apresenta uma infecção confirmada e está sob ventilação mecânica, sem outras alterações em variáveis do MPM-II<sub>24</sub>.

Para calcular seu risco de morte teríamos que somar a constante  $y$  (-5,64592), os coeficientes da idade ( $7 \cdot 0,03268 = 2,18960$ ; lembremos que este coeficiente é para cada 10 anos de idade, e como o paciente tem 70 anos é necessário multiplicar o coeficiente por 7), do tipo de internação (0,83404, por ser uma internação clínica), a infecção (0,49742) e a ventilação mecânica (0,80845). Na **tabela 2** são mostrados esses dados, e o valor do logit para este paciente.

<b>Tabela 2. COEFICIENTES DE MPM II<sub>24</sub> EM UM PACIENTE HIPOTÉTICO</b>	
70 anos ( $7 \cdot 0,03268$ )	2,18960
Internação cirúrgica urgente ou médica	0,83404
Infecção confirmada (nas 24 h. após a admissão)	0,49742
Ventilação mecânica (nas 24 h. após a admissão)	0,80845
Constante	-5,64592
Logit (somatória de coeficientes e constante)	-1,31640

A probabilidade de morte deste paciente concreto pode ser obtida por meio da fórmula [ $p = e^{\text{logit}} / 1 + e^{\text{logit}}$ ] que, neste caso,  $p(\text{morte}) = e^{-1,3164} / 1 + e^{-1,3164} = 0,21$ . Isto é, o paciente tem uma probabilidade de falecer de 21% ou, melhor dizendo, de cada 100 pacientes como este, espera-se que 21 venham a óbito e 79 apresentem melhora.

Para os intensivistas é útil ter uma ideia do prognóstico de seus pacientes. Mas, além disso, graças a este tipo de sistemas podemos comparar as unidades de terapia intensiva. Por exemplo, se duas UTIs que trataram 100 pacientes cada uma mostram a mesma mortalidade, suponhamos que de 37%, a comparação bruta sugere uma qualidade similar do atendimento. No entanto, verificando a distribuição dos pacientes segundo a probabilidade de morte medida pelo MPM que é mostrada na **tabela 3**, em uma delas seriam esperados 36 óbitos (em vez dos 37 observados) enquanto que na outra, devido à menor gravidade de seus pacientes, seriam esperados no máximo 26 óbitos (11 a menos do que os que realmente ocorreram), aspecto que sugere a possível existência de algum problema de qualidade a ser identificado.

**Tabela 3. COMPARAÇÃO DA MORTALIDADE EM 2 UTIs HIPOTÉTICAS**

Unidade de Terapia Intensiva 1				Unidade de Terapia Intensiva 2			
Admissões UCI 1	Óbitos observados	pMPM	Óbitos esperados	Admissões UCI 2	Óbitos observados	pMPM	Óbitos esperados
10	2	0,15	1,5	10	1	0,11	1,1
20	5	0,24	4,8	20	4	0,18	3,6
30	10	0,33	9,9	30	5	0,22	6,6
30	14	0,45	13,5	30	17	0,34	10,2
10	6	0,60	6,0	10	10	0,48	4,8
Total	100		35,7	100	37		26,3

pMPM: probabilidade de morte prevista pelo *Mortality Probability Model*

Do mesmo modo que utilizamos o MPM para comparar a mortalidade em duas UTIs, podemos utilizar qualquer outro sistema de ajuste de riscos para comparar outros tipos de serviços em relação a um determinado resultado que se pretende medir. Por exemplo, os DRGs como forma de comparar o tempo médio de internação entre dois serviços. Note que este tipo de sistema não é tão diferente da prática clínica de ajustar os pacientes por algum instrumento com a finalidade de comparar ou prever resultados: a escala de coma de Glasgow para comparar a mortalidade pós-traumatismo cranioencefálico entre dois hospitais, ou o Apgar para comparar a mortalidade perinatal entre unidades de neonatologia ou de obstetrícia, ou o sistema ASA para comparar as complicações perioperatórias entre pacientes operados em dois centros ou por duas equipes, etc. Todos estes métodos, uma vez que quantificam o risco de ocorrer um resultado ou desfecho, podem ser considerados também sistemas de ajuste de riscos e a sistemática básica de utilização será ajustar o produto assistencial (internação, partos, cirurgias, etc.) pelo SAR mais adequado ao tipo de comparação que se deseja realizar.

## 1.2. Características dos sistemas de ajuste de riscos

Os SARs mais conhecidos incluem os **Grupos de Diagnósticos Relacionados** (DRGs) e suas variações, o **Disease Staging** (DS) e outros sistemas baseados fundamentalmente na revisão da história clínica, como o **Computerized Severity Index** (CSI) e os sistemas desenhados para pacientes em unidades de



terapia intensiva (**APACHE, MPM, PRISM, TISS, SAPS**), mas existem muitos outros, como POSSUM, COMPLEX, SUPPORT, TISS-28, NEMS, AIM, RUGs em suas diferentes versões etc., bem como sistemas de um determinado procedimento específico ou um diagnóstico, que podem ser de interesse em determinadas áreas.

Uma forma de abordar esses sistemas seria através das seguintes perguntas:

1. Qual produto ou desfecho se busca ajustar? Basicamente, pode ser um **resultado de consumo de recursos**, como a duração média da internação (DRGs) ou a intensidade do uso de recursos (TISS, RUGs), ou um **resultado clínico** como a mortalidade (DS, CSI, APACHE, MPM). O conceito de gravidade em cada caso pode ser muito diferente. Assim, um paciente com uma pontuação MPM muito elevada (elevado risco de morte precoce) pode ser classificado, em um DRG, como sendo de "baixo peso" (baixo custo ou impacto) uma vez que ocorrendo o óbito muito precocemente, seu consumo de recursos será escasso. Por isso, empregar estes sistemas para ajustar resultados diferentes daqueles previstos em sua concepção (por exemplo, DRG para ajustar a mortalidade) pode ser inadequado.
2. Os requisitos de informação de cada sistema também são diferentes? Assim, enquanto que **os sistemas para ajustar resultados de consumo de recursos costumam ser elaborados a partir de bases de dados clínico-administrativos** do tipo Conjunto Mínimo de Dados Básicos (CMBD), **os sistemas para resultados clínicos costumam exigir dados provenientes da história clínica ou outras fontes primárias**, o que tem implicações em seu custo e viabilidade, mas também em sua capacidade de previsão dos riscos e em sua credibilidade clínica.
3. Quanto aos critérios de classificação utilizados, vale destacar, primeiramente, o **papel atribuído ao diagnóstico** (Dx), que permite classificar esses sistemas como *diagnóstico-dependentes* (DRG), em geral vinculados à medição de custos, e *diagnóstico-independentes* (APACHE, MPM), quase sempre

vinculados à medição da mortalidade precoce em unidades de pacientes críticos, nas quais a estabilidade dos sistemas orgânicos costuma ter mais importância do que o próprio diagnóstico. O papel atribuído aos grandes procedimentos cirúrgicos (Px) também é essencial para medir o consumo de recursos.

4. A escolha do momento em que são coletados os dados tem importantes implicações, uma vez que os sistemas que usam o CMBD, que obtém os dados na alta (retrospectivos), não poderão estabelecer a sequência temporal de parte dos eventos ocorridos durante a hospitalização, o que implica em significativas limitações para separar o risco devido à comorbidade prévia do paciente, do risco decorrente das complicações adquiridas no hospital. Este aspecto é de fundamental importância nas comparações entre provedores, pois se for ajustado, por exemplo, o risco decorrente das infecções hospitalares, este não será levado em conta, obtendo-se resultados similares de qualidade ou eficiência em serviços com índices de infecções hospitalares muito diferentes.
5. Por fim, em relação ao tipo de medida que estes sistemas oferecem, alguns utilizam **escalas contínuas**, nas quais é avaliado o risco em relação a uma média obtida (p.ex., no sistema dos DRG, um peso de 1,15 implica um risco de consumo de recursos 15% maior do que a média de todos os pacientes) e têm sentido relativo óbvio, enquanto que outros sistemas utilizam **escalas ordinais** que não implicam tais referências (no CSI uma pontuação "2" não implica o dobro de risco de uma pontuação "1") ou constroem **agrupamentos de base diagnóstica (DRG) estritamente categóricos**.

Na **tabela 4** é mostrada uma generalização das diferenças essenciais entre os sistemas para ajuste de resultados clínicos e de custos.

<b>TABELA 4. Características dos sistemas de ajuste de riscos</b>		
	CLÍNICOS	CUSTOS
Requisitos de informação	++++	+
Fonte usual	Histórico clínico	CMBD
Custo do sistema	++++	+
Capacidade ajuste	+++	+
Credibilidade clínica	+++	+
Desenvolvimento	Julgamentos clínicos	Empírica
Papel do diagnóstico	++	++
Papel do procedimento	+	++++
Coleta de dados	Variável (concorrente)	Na alta (retrospectiva)

## 2. Aspectos metodológicos dos sistemas de ajuste de riscos

Os processos de desenvolvimento e construção dos SAR são essenciais para reconhecer os pontos fortes e fracos destes métodos. Um sistema de ajuste de riscos é elaborado a partir dos seguintes elementos:

- 1) Definição do objetivo do sistema.
- 2) Seleção dos fatores de risco que definirão a gravidade.
- 3) Desenvolvimento do sistema de ajuste (modelos de desenvolvimento).
- 4) Validação do sistema (modelos de validação).

### 2.1 Definição do objetivo do sistema

O primeiro passo na elaboração de um SAR é a definição de seu propósito, isto é, que tipo de serviços se pretende avaliar e qual resultado se deseja isolar da influência dos fatores de risco.

Em relação ao objeto de avaliação, as possibilidades básicas são qualquer tipo de serviços (hospitais gerais, centros de saúde,

centros de longa permanência), unidades de trabalho (serviços clínicos) ou mesmo médicos individuais.

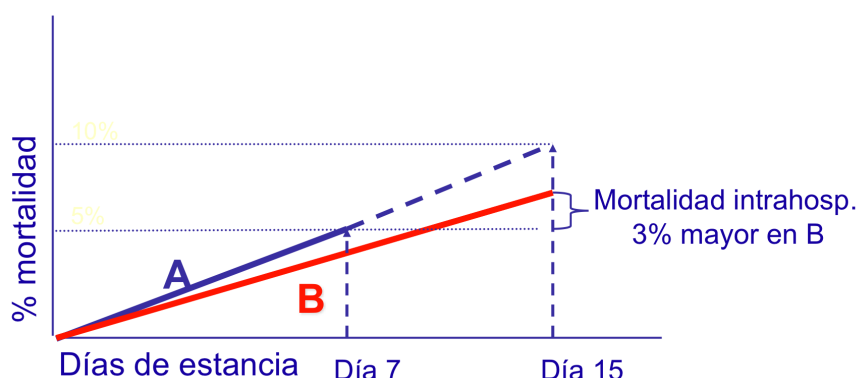
Seleção do resultado e janela de observação. Quando a fonte de informações é o CMBD os **resultados** ficam limitados às variáveis nele disponíveis: mortalidade intra-hospitalar, duração da internação, reinternações, ou a presença de determinados códigos que indicam complicações.

O **tempo de observação** é um elemento chave neste marco. Por um lado a janela temporal para a identificação do resultado precisa ser adequada ao propósito do sistema. Por exemplo, se o objetivo é avaliar a eficácia das UTIs é mais razoável utilizar

## Figura 2. *Efecto de la variabilidad en la ventana temporal de seguimiento de los resultados*

**Hosp A:**  
**Estancia Media = 7 días;**  
**Mortalidad al 7 día= 5%;**  
**Mortalidad al 15 día= 10%**

**Hosp B:**  
**Estancia Media = 15 días;**  
**Mortalidad al 7 día= 4%;**  
**Mortalidad al 15 día= 8%**



a mortalidade em tempos de observação curtos (até 30 dias) do que o retorno por ano, uma vez que a primeira opção terá uma maior capacidade de inferência sobre a qualidade destas unidades, enquanto que a segunda será excessivamente influenciada por todo o tratamento posterior à alta. Em segundo lugar, seria necessário garantir tempos de acompanhamento homogêneos para todos os casos. Assim, a mortalidade intra-

hospitalar pode ser uma medida inadequada, por não identificar a mortalidade no domicílio daqueles pacientes que receberam alta. Isso implica que nos provedores com internações menos prolongadas (por exemplo, os que têm uma política de alta para pacientes terminais) haverá um menor tempo para risco de morte intra-hospitalar e haverá taxas de mortalidade artificialmente menores. Finalmente, outro elemento chave especialmente nos SAR para cuidados intensivos, é o momento em que são coletados os dados, aspecto já comentado em outra seção.

A definição da unidade de análise tem 2 campos de interesse: o campo indivíduo e o campo da análise propriamente dita. Em relação ao campo individual, as possibilidades são variadas: episódios de hospitalização (internação) ou de cuidados (visitas médicas), processos completos que podem incluir um ou mais episódios de hospitalização, com ou sem atendimento ambulatorial, pessoas-ano etc. Este aspecto tem importantes implicações analíticas e sua escolha depende basicamente do propósito da avaliação. Em relação ao campo da análise, as possibilidades são: sistemas gerais, pormenorizados, de todo tipo de pacientes, ou sistemas específicos para uma doença ou um procedimento, ou um grupo de doenças ou procedimentos. Os primeiros, ou seja, aqueles com foco no campo individual, foram elaborados para avaliar a eficácia ou eficiência de hospitais e seu exemplo mais original são os DRG e suas variantes; enquanto que os segundos têm como objetivo avaliar algum tipo mais específico de atendimento prestado.

*Um marco conceitual bem embasado pela literatura deve relacionar as variáveis a ajustar com cada resultado particular*

## 2.2. Seleção dos fatores de risco que definirão a gravidade

Nem todas as características dos pacientes influem de modo significativo em todos os resultados, o que explica a diversidade das dimensões consideradas em cada SAR em função de sua finalidade. Um **marco conceitual bem embasado pela literatura deve relacionar as variáveis a ajustar com cada resultado particular**, aspecto não necessariamente fácil, em função da existência de múltiplas correlações entre muitos dos fatores de risco. As mesmas variáveis podem ser fatores de ajuste de resultados diferentes - como mortalidade e duração da permanência ou custos - mas isto nem sempre acontece, sendo desejável ajustar modelos separadamente para cada resultado

que se deseja medir. Não obstante, em alguns casos, pode ser útil a combinação de sistemas para melhorar a compreensão das variações dos resultados inicialmente obtidos (esta é a base de alguns SAR, como os *Refined-DRG* que criam grupos dentro de cada DRG para aprofundar a capacidade de compreensão sobre os custos).

Para a seleção de fatores de risco, as fórmulas levam em consideração a revisão da literatura clínica e o julgamento dos especialistas, com o desafio de traduzir alguns fatores prognósticos de difícil medição a *proxys* com os quais mantenham uma boa correlação. Em todo caso, considerando-se que o objetivo de todos estes sistemas é a avaliação da qualidade e/ou eficiência do atendimento hospitalar, é importante ***assegurar que as variáveis de ajuste dependam do risco inerente ao paciente e que não sejam uma consequência dos cuidados desenvolvidos pelo próprio serviço.***

Nesta seção será realizado um breve comentário das variáveis mais utilizadas para ajustar a gravidade. Cabe observar que algumas destas variáveis podem ser utilizadas como "resultados" (por exemplo, o estado funcional pode ser utilizado como fator de risco relacionado ao resultado mortalidade, mas também é empregado em outros sistemas como um elemento a justificar outros fatores).

A **idade** é uma característica predeterminada dos pacientes, facilmente disponível, que pode ser um importante previsor de riscos, em parte porque é associada a outras características do paciente (determinados diagnósticos e comorbidades) e em parte porque é um previsor independente de resultados piores (óbito, complicações, duração da internação) uma vez que os idosos podem exigir períodos mais longos de recuperação e ter um risco maior de complicações. Estes efeitos, importantes no âmbito geral, podem ser muito modestos quando são estudados grupos com idades similares ou quando se analisa algum tipo de patologia especialmente prevalente em algum grupo de uma faixa etária específica. Um grupo etário de especial interesse são os maiores de 79 anos, que costumam apresentar mais complicações, gravidade etc., do que outros pacientes, entretanto podem receber um esforço terapêutico menos intensivo do que os pacientes mais jovens com quadros clínicos semelhantes. Isto porque, apesar de representarem um grupo de pacientes que - por sua maior gravidade - seriam considerados como de

*Assegurar que as variáveis de ajuste dependam do risco inerente ao paciente e que não sejam uma consequência dos cuidados desenvolvidos pelo próprio serviço.*



alto custo, acabam por receber dos clínicos um tratamento mais expectante que, por consequência, pode resultar em custos mais baixos. .

O **sexo**, outra variável facilmente disponível e de grande importância nos estudos epidemiológicos, mostrou pouca relevância como predictor de resultados hospitalares de curto prazo, e não está incluído em praticamente nenhum sistema geral de ajuste de riscos.

O **diagnóstico principal** (DxP), a casuística em sentido estrito, oferece poucas dúvidas com relação à sua importância como predictor de riscos, tanto sobre os resultados clínicos como sobre a utilização de recursos. Os critérios essenciais para realizar um Dx (diagnóstico) exato são a etiologia, a localização e as manifestações fisiopatológicas (por exemplo: pneumonia de lóbulo superior esquerdo pelo *Sp. pneumoniae* com insuficiência respiratória). Não obstante, na prática clínica nem sempre é possível ou desejável especificar completamente o Dx e, em muitas ocasiões, este não vai além de um diagnóstico genérico ou sindrômico. Outras vezes não é possível estabelecer a linha divisória entre doença e resultados anormais de um exame, ou, até mesmo, idênticas manifestações provêm de diferentes doenças e vice-versa. A adaptação a esta realidade clínica de alta incerteza e dependência da intensidade diagnóstica levou ao desenvolvimento de classificações diagnósticas não mutuamente excludentes (pacientes iguais podem ser classificados com códigos diferentes: por exemplo, como trombose ou acidente vascular-cerebral ou hemiplegia), o que introduz importantes possibilidades de variabilidade na codificação.

Identificar o DxP pode não ser suficiente para ajustar alguns resultados, e determinar a **gravidade dentro de um mesmo DxP** pode ser essencial em muitos casos (por exemplo, um câncer localizado e um disseminado não têm o mesmo risco de morte nem as mesmas necessidades de recursos). A gravidade do DxP é um conceito que varia em função dos objetivos de cada sistema de medição e que, em geral, é associado ao prognóstico de óbito em uma determinada janela temporal. Em oncologia costuma-se utilizar classificações de gravidade em estados definidos pela clínica ou pela patologia.

A **comorbidade** refere-se à presença de doenças não relacionadas

com o DxP, mas coexistentes no momento da internação. A comorbidade não deve ser confundida com a gravidade do DxP (patologia relacionada) nem com as **complicações** (patologia não presente no momento da internação). Note que se um paciente der entrada com uma pneumonia e tiver uma infecção urinária adicional, ela seria classificada como comorbidade; mas se não a tinha quando foi internado e ela surge durante a hospitalização, será classificada como uma complicação: **a diferença importante é que a primeira é um fator do paciente, e a segunda um possível problema de qualidade que não devemos introduzir no ajuste de riscos.** O caso típico de comorbidade são as patologias crônicas (diabetes, hipertensão arterial, coronariopatia isquêmica, bronquite obstrutiva crônica etc.), embora ocasionalmente possam ser agudas (por exemplo, uma fratura de quadril produzida por queda ao sofrer um acidente vascular-cerebral). Como ocorre com o DxP, não é suficiente identificar a existência de comorbidade, mas será necessário avaliar sua gravidade e extensão em relação ao resultado a medir. Existem índices específicos de comorbidade (índice do Charlson entre outros), mas a maior parte dos SAR inclui diretamente a comorbidade para a construção de grupos de pacientes ou o ajuste de riscos.

*a diferença importante é que a primeira é um fator do paciente, e a segunda um possível problema de qualidade que não devemos introduzir no ajuste de riscos*

A **estabilidade clínica** reflete o estado fisiológico dos sistemas corporais do paciente mediante o exame de seus sinais vitais (frequência cardíaca e respiratória, pressão arterial etc.), bioquímica (K, Na, creatinina,...), parâmetros hematológicos, gases sanguíneos e nível de consciência. É uma das variáveis essenciais nos desfechos de curto prazo, e base de boa parte dos SAR utilizados nas UTIs.

O **estado funcional**, medido pela capacidade de realizar as atividades da vida diária, é outra variável de interesse, não apenas porque tem uma estreita relação com o consumo de recursos de enfermagem, aspecto fundamental nos SAR em centros de longo tratamento, mas porque em alguns casos tem uma forte correlação com o prognóstico.

O **estado psicológico, cognitivo e psicossocial** foi reconhecido em diversos trabalhos como uma variável fundamental para os resultados ou desfechos clínicos; entretanto, se mostra proibitiva para os SAR, pois para sua obtenção exigem entrevistas detalhadas e realizadas por pessoal especializado.

As **características socioculturais e socioeconômicas** também se mostraram como fatores relacionados de modo significativo com os resultados. O baixo nível socioeconômico foi associado a resultados piores em diversos estudos, do mesmo modo que o grupo étnico nos estudos norte-americanos. Mais uma vez, neste caso, antes de ajustar por esses fatores, é necessário definir seu processo de associação com o resultado: seria ou não um fator independente do atendimento em questão? Fatores associados ao nível socioeconômico como o alcoolismo e a drogadicção e os riscos ocupacionais também foram associados a resultados piores; bem como o nível cultural, algumas práticas religiosas e hábitos alimentares restritivos.

Mantendo relação com as características apontadas anteriormente, as **atitudes e preferências do paciente** (por exemplo, recusando cuidados agressivos ou por motivos religiosos, não aderindo a determinados tratamentos, etc.) podem ser um importante previsor de resultados.

---

### **2.3. Elaboração e desenvolvimento de sistemas de ajuste**

---

A necessidade de combinar um bom rendimento estatístico com a aceitabilidade clínica convenceu a maior parte dos pesquisadores de que a combinação de julgamento clínico e modelagem empírica seria melhor do que qualquer das duas aproximações separadamente. Mesmo no caso de modelos estritamente empíricos, o julgamento clínico é útil para decidir os pontos de corte nos valores das variáveis ou para a posterior revisão dos modelos adotados.

A menos que o sistema se baseie exclusivamente em julgamentos clínicos, é necessária uma base de dados para sua construção, e o primeiro passo é a "limpeza" desta base: 1) excluir o tipo de casos que não se deseja considerar na análise, 2) excluir os serviços que não cumpram uma série de critérios de qualidade de informação, 3) depurar os valores não plausíveis ou sem significado, 4) avaliar a especificidade e a congruência das informações, 5) avaliar a possibilidade de construir vínculos entre os registros da mesma base de dados, quando o resultado selecionado assim o exigir.

O problema seguinte é o tratamento dos valores perdidos, desconhecidos, impossíveis ou incongruentes, que vai da eliminação dos casos com algum valor *missing* (inexistente no banco) até sua imputação mediante algum método estatístico. Em ambos os casos, a tendência pode ser importante se os *missing*, como é provável, não forem distribuídos aleatoriamente.

O passo a seguir será a redução de variáveis, especialmente daquelas que apresentarem um alto grau de correlação. O mais comum é o emprego de diversas técnicas de análise, especialmente os modelos multivariados para resultados contínuos ou dicotômicos, e os modelos de redução de dados, existindo diversas revisões que avaliam as limitações técnicas destes modelos. Nesta fase é interessante avaliar a inclusão de variáveis com pouca frequência ou que oferecem resultados ilógicos, e a possibilidade de incorporá-las agrupadas; por exemplo, é possível que algumas comorbidades de baixa frequência não sejam utilizáveis no modelamento, mas possam ser incorporadas em um índice global de comorbidade.

A seguir deve-se explorar a estrutura das variáveis independentes a serem incluídas (fatores de risco) e a forma de sua relação com a dependente (resultado). Isso orientará sobre a possibilidade de introduzir as variáveis como contínuas ou a necessidade de classificá-las (e em quais categorias) ou de incorporar interações. O passo final será utilizar as técnicas estatísticas mais adequadas para efetuar a modelagem, aspecto que será tratado posteriormente como parte do desempenho dos modelos.

---

## 2.4. Validação

---

A noção de medição é a chave dos processos comentados, quer se trate de medições de resultados simples, como a mortalidade, ou de conceitos abstratos, como a gravidade ou a qualidade. Para descrever a qualidade das medições realizadas por um instrumento são empregados dois conceitos: **validade** e **confiabilidade**. A validade denota o grau em que um instrumento de medida mede o que tentava medir, ou visto de outra forma, o grau em que a medição se distancia sistematicamente de seu objetivo. Estritamente falando ***não se valida um instrumento em si, mas sim o propósito para o qual é usado. A validade é, por outro lado, uma questão de gradação, não existe***

**em termos absolutos** e, no mesmo sentido, é um conceito dinâmico.

Em geral, em sua aplicabilidade aos SAR, a medida de resultado funciona como critério (**validade de critério**) sobre o qual se valida o modelamento do risco, consistindo a validação em responder à questão: *até que ponto os indicadores de risco elaborados predizem o resultado que lhes dá sentido?*

Em sua aplicação é importante diferenciar entre os dados usados para desenvolver o modelo e aqueles empregados para sua validação, pois se forem atendidas somente as previsões nos primeiros, a validade do modelo acaba por ser sobrevalorizada. Em geral, a abordagem utilizada ocorre por meio da chamada **validação cruzada** (*cross-validation*), que consiste em dividir aleatoriamente a base de dados e utilizar uma metade para a construção do modelo (base de dados de desenvolvimento) e a outra metade para comprovar que os fatores de risco obtidos na primeira predizem bem os resultados também nesta segunda base (base de dados de validação). A validação pode ser repetida invertendo-se o uso das bases de dados, ou podem ser utilizados outros métodos, como o *bootstrapping* (obter repetidas amostras aleatórias sem substituição, sobre as quais se valida o modelo).

Outros tipos de validação de interesse são a **validade de conteúdo** (até que ponto os fatores de risco incorporados representam o universo de fatores de riscos relevantes? os fatores de risco não incluídos são importantes?), a **validade de construto** (partir da relação teórica entre conceitos para examinar a relação empírica entre suas medições), e a **validade aparente** (os usuários do SAR terão a percepção de sua validade?).

Quando os 'resultados' são utilizados para avaliar a qualidade da assistência, é necessário estabelecer previamente que os resultados podem ser atribuídos à assistência recebida. O termo **validade atribucional** foi empregado para ressaltar este aspecto, e ele examina a questão de verificar se o ajuste de riscos efetuado é suficiente para considerar que as variações em resultados observados não estão relacionadas com as características intrínsecas dos pacientes, isto é, para que possam ser atribuídas a outras causas, como a diferença na efetividade do tratamento ou na qualidade da assistência, sendo um dos aspectos mais discutidos em todos os sistemas de ajuste de riscos.

## 2.5. Desempenho dos sistemas de ajuste de riscos

A **validade final e a utilidade prática dos ajustes são dadas por sua capacidade de captar as diferenças entre pacientes, problema que é abordado em geral comparando as previsões realizadas pelos modelos ajustados com os resultados reais, a citada validade preditiva.** O termo ajuste de riscos sugere que um modelo oferece uma estimativa, para cada paciente, do resultado esperado em função da avaliação de seus fatores de risco. Na prática, nem sempre é assim. Em alguns sistemas, transformar pontuações de risco em resultados esperados é relativamente simplista. Como exemplo, o MPM utiliza uma série de informações para atribuir a cada paciente uma pontuação que é uma estimativa da probabilidade de óbito em 30 dias, e o número de óbitos esperados em um hospital seria a soma de suas probabilidades estimadas. Neste caso, pode-se calcular a probabilidade de óbito para cada paciente a partir de seus dados, e construir uma taxa de mortalidade esperada (somando essas probabilidades e dividindo pelo número de casos) que pode ser comparada com a taxa de mortalidade observada. Em outros sistemas, entretanto, podem ser necessárias diversas mudanças (calibração do modelo e análises posteriores) para permitir esta transformação.

A modelagem multivariada proporciona o marco para a tradução de pontuações de risco, categorias de risco ou variáveis específicas em resultados esperados. As variáveis independentes podem ser qualquer dos fatores de risco comentados anteriormente, e a dependente é o resultado a analisar que, basicamente, pode ser uma variável contínua (dias de internação, custos) ou dicotômica (ocorrência ou não de óbito, complicação, reinternação).

Estatisticamente costuma-se utilizar modelos do tipo **lineares generalizados**, sendo os mais empregados o **modelo de regressão linear múltipla**, o de **regressão logística** e o de **riscos proporcionais de Cox**.

O **coeficiente  $r^2$**  é a medida resumo mais utilizada do rendimento dos modelos multivariados quando a variável dependente é contínua, e pode ser interpretado como a fração da variabilidade total da variável dependente que pode ser explicada por (ou atribuída a) diferenças de risco entre os casos incluídos no

*A validade final e a utilidade prática dos ajustes são dadas por sua capacidade de captar as diferenças entre pacientes, problema que é abordado em geral comparando as previsões realizadas pelos modelos ajustados com os resultados reais, a citada validade preditiva.*



modelo. Teoricamente, o coeficiente  $r^2$  pode alcançar um valor máximo de 1 (previsão perfeita), mas na prática esta expectativa não é realista.

As estratégias publicitárias das empresas que comercializam sistemas de ajuste de riscos tendem a apresentar o  $r^2$  das bases de dados de desenvolvimento como sinônimo de alta capacidade explicativa sobre a variável dependente, assumindo que esta será a mesma ao ser aplicada sobre outros dados, quando mudanças na dispersão da variável dependente terão impacto na capacidade explicativa do modelo. Por isso, **o coeficiente de determinação não é apropriado para comparar modelos desenvolvidos a partir de bases diferentes**. Neste aspecto, seria necessário levar em conta que quanto mais os dados do modelo de desenvolvimento forem diferentes daqueles do serviço de saúde que quer utilizar o sistema, menor será o valor de  $r^2$ . E, em resumo, *os  $r^2$  obtidos em algumas bases de dados não são extrapoláveis a outras bases de dados*.

*O coeficiente de determinação não é apropriado para comparar modelos desenvolvidos a partir de bases diferentes*

*Os modelos tendem a ser tendenciosos, subestimando o risco para os casos de alto risco e superestimando-os de baixo risco.*

É preciso acrescentar que **os modelos tendem a ser tendenciosos, subestimando o risco para os casos de alto risco e superestimando-os de baixo risco**. Por exemplo, os pesos dos DRG são tendenciosos neste sentido, embora eufemisticamente a esta tendência se dê o nome de **"compressão"**. Isto significa que a interpretação dos resultados de um hospital concreto deveria levar em consideração a natureza do padrão, tanto por seu valor relativo como porque os hospitais com casos mais graves terão seus riscos subestimados com base em dados compostos basicamente por casos de baixo risco.

Em conjunto, **o valor de  $r^2$  é função do número de variáveis e interações introduzidas em uma base de dados concreta, do entorno de onde vem a amostra**, especialmente do tipo de pacientes incluídos, da dispersão das variáveis, **da inclusão ou não de outliers e da transformação de variáveis**. Estes aspectos fazem com que, *embora o  $r^2$  seja um valor numérico, determinar sua importância exige a avaliação subjetiva de diversos fatores*. A avaliação destes fatores, entretanto, em muitos sistemas comerciais não poderá ser realizada por se tratar de dados que fazem parte da "caixa preta", considerada segredo comercial e resultará na impossibilidade de avaliar as possíveis distorções do ajuste, e a direção de tais distorções. Em todo caso, **para efeitos práticos, é adequado não acreditar**

*O valor de  $r^2$  é função do número de variáveis e interações introduzidas em uma base de dados concreta, do entorno de onde vem a amostra*

## que os $r^2$ anunciados pelos folhetos comerciais dos SAR sejam um indicador da qualidade do respectivo SAR.

A **regressão logística** é o método mais utilizado para modelar variáveis dependentes dicotômicas. Neste modelo, a variável dependente é o log natural do *Odds Ratio* (OR) da ocorrência. O consenso em relação ao modo de avaliar os modelos que predizem resultados binários é menor. Nestes modelos o que se prediz é a ocorrência ou não de um evento em cada caso, atribuindo a probabilidade de o evento ocorrer. Como no caso dos modelos de regressão linear ordinária ou de mínimos quadrados, as inferências baseadas na modelagem logística serão definitivamente válidas quando cumpridos determinados critérios. Entretanto, tanto um modelo como outro são frequentemente empregados, apesar de não preenchidos todos os critérios. Isto é aceitável parcialmente, uma vez que as técnicas mostraram ser robustas, embora 'distantes' das condições ideais de uso. De fato, a credibilidade dos algoritmos para transformar medidas de risco em previsões levam em consideração principalmente o grau de aproximação entre essas previsões e a realidade.

A prática de avaliar previsões é comum no contexto clínico, por exemplo, na avaliação de testes diagnósticos. Quando o resultado é dicotômico, podemos calcular os índices clássicos de sensibilidade, especificidade, valor preditivo positivo, negativo: proporção - ou probabilidade - de casos (verdadeiros positivos) bem classificados ou detectados pela prova (sensibilidade); a proporção de não casos (verdadeiros negativos) bem classificados (especificidade); proporção de casos entre os positivos (valor preditivo positivo) etc. É importante destacar que a sensibilidade e a especificidade de uma prova com resultado dicotômico não são afetadas por diferenças na proporção de casos existentes na população (prevalência), mas seu valor preditivo sim.

Na avaliação desses modelos são empregados dois critérios: calibração e discriminação. **Um modelo é considerado calibrado em relação a um conjunto de dados quando a média de suas previsões se aproxima da média dos resultados reais.** Por outro lado, **um modelo apresenta boa discriminação, se prevê maiores probabilidades do evento (ex.: óbito) aos pacientes que apresentaram o evento (os que realmente faleceram) do que aos pacientes que não apresentaram o evento.**

*Para efeitos práticos, é adequado não acreditar que os  $r^2$  anunciados pelos folhetos comerciais dos SAR sejam um indicador da qualidade do respectivo SAR.*

*Um modelo é considerado calibrado em relação a um conjunto de dados quando a média de suas previsões se aproxima da média dos resultados reais. Por outro lado, um modelo apresenta boa discriminação, se prevê maiores probabilidades do evento (ex.: óbito) aos pacientes que apresentaram o evento (os que realmente faleceram) do que aos pacientes que não apresentaram o evento*

Tanto a importância relativa de um de outro critério (discriminação x calibração), como os índices para expressá-los, são motivo de controvérsia. Embora seja desejável uma boa calibração, aqueles que destacam a discriminação argumentam que uma deficiência dela não é contornável, diante da possibilidade de recalibração nos casos de uma calibração incorreta. Para ilustrar este argumento imaginemos uma população com 10% de óbitos, em que o modelo atribuisse a cada pessoa uma probabilidade constante de 0,10; a calibração seria perfeita, mas inútil para diferenciar os casos de falecimento daqueles de sobrevivência. Por outro lado, um modelo que atribuisse a esta população uma probabilidade de 0,8 aos pacientes que sobrevivem e de 0,9 aos que falecem, mesmo que em termos numéricos suas previsões sejam errôneas, pode ser empregado de um modo perfeito para prever o resultado. Uma simples mudança (recalibração) pode solucionar o problema deste segundo modelo, mas não existe nenhuma correção possível para o primeiro caso.

Outros autores defendem que se o modelo não alcançar uma boa calibração, não tem valor estudar sua capacidade discriminativa. Se considerarmos a sua finalidade, se o modelo for empregado para fazer **predições individuais**, a capacidade discriminativa assume importância; entretanto, se o modelo pretende determinar a incidência esperada do evento (**abordagem coletiva**) e compará-la com a observada, então a calibração merece maior atenção.

Sobre a medição da discriminação, mesmo que não exista consenso sobre a medida mais apropriada, o **estimador C** é o mais empregado. Diversas definições são possíveis; uma maneira ilustrativa de explicá-lo seria considerar todos os pares possíveis que podem ser formados entre os casos que apresentam o evento e aqueles que não apresentam. O estimador C equivale à proporção de pares para os quais a probabilidade estimada do evento (p. ex. o falecimento) é superior para o membro falecido do par, do que para seu correspondente vivo. Graficamente, o estimador C corresponde à área sob a **curva Receiver Operating Characteristics** (ROC), um gráfico no qual são observados todos os pares sensibilidade/especificidade decorrentes da aplicação de diferentes pontos de corte em toda a categoria de resultados observados. No eixo das ordenadas situa-se a fração de verdadeiros positivos e no das abscissas a fração de falsos positivos (1-sensibilidade). Cada ponto

constitutivo da curva corresponde a uma decisão sobre o ponto de corte para a previsão. A curva, em uma prova ou modelo com discriminação perfeita, sem sobreposição de resultados ou previsões entre as populações com e sem o evento, alcançará o extremo superior esquerdo, onde sensibilidade e especificidade assumem seu valor máximo em termos de estimador C. **O estimador C assume valores entre 0,5 (o modelo não funciona melhor que o acaso para prever o resultado) e 1 (o modelo prevê perfeitamente).**

No que diz respeito às suas características, vale destacar que o estimador C depende apenas das categorias de previsões, não de seus valores. Assim, um modelo que atribuir probabilidades sem nenhuma sobreposição entre quem desenvolve e não desenvolve o evento, por exemplo, 0,20 para os que falecem e 0,19 para os que sobrevivem, terá um estimador C de valor 1.

A **curva de calibração** compara graficamente a probabilidade estimada com o resultado real. Para construí-la, os casos são divididos em grupos segundo a probabilidade estimada. A média da probabilidade estimada em cada grupo é comparada com a frequência real de evento no grupo. Lemeshow e Hosmer propuseram aplicar a prova de qui-quadrado aos dados organizados de modo similar à curva de calibração. Basicamente consiste na divisão dos dados em décimos, de acordo com o nível crescente de risco, e a comparação em cada décimo da frequência observada e esperada. O estimador, obtido a partir das separações dos décimos e uma estimativa de sua variação, é distribuído como um qui-quadrado com 8 graus de liberdade.

Todas as questões comentadas relativas ao coeficiente de determinação e seu uso na avaliação do rendimento de modelos com variáveis contínuas são aplicáveis quando se planeja sua aplicação em modelos com variável resposta dicotômica. Entretanto, quando a variável é dicotômica e foi modelada por um procedimento diferente do procedimento de mínimos quadrados, existem poucos precedentes que suportem o uso e interpretação do  $r^2$ . Os pesquisadores concluíram **que os modelos que predizem resultados do tipo 0/1 raramente alcançam um  $r^2$  de 0,30**. Além do mais, foi observado que seu tamanho é afetado pela frequência do resultado. Como vantagem do  $r^2$  observou-se que ele se mostra mais sensível do que o estimador C quando se deseja comparar a mudança no rendimento de sucessivos modelos sobre a mesma população.

*O estimador C assume valores entre 0,5 (o modelo não funciona melhor que o acaso para prever o resultado) e 1 (o modelo prevê perfeitamente)*

Os cuidados aplicáveis a estes modelos e índices de rendimento são todos aqueles apresentados anteriormente, especialmente em relação às comparações dos valores de  $r^2$  ou C, quando são obtidos de diferentes grupos de dados, ou quando existem diferentes critérios de definição da população, protocolos na depuração de dados, ou quando casos difíceis estão sobre representados, em situações definitivas que fazem com que se altere significativamente a variabilidade que o modelo precisa explicar.

### 3. Sistemas de ajuste de riscos: uso prático

O único sistema de ajuste amplamente utilizado na gestão hospitalar na Espanha corresponde ao DRG<sup>3</sup>. No Sistema Nacional de Saúde (SNS) são utilizados especialmente os AP-DRG (3M<sup>©</sup>) e os CMS-DRG, e seu uso mais difundido é para ajustar a comparação da internação média (EM) entre hospitais. A EM de um hospital depende do tipo de pacientes que trata, e quando as EM de diversos hospitais são avaliadas comparativamente, as diferenças decorrerão tanto de uma maior ou menor eficiência dos serviços quanto de sua diferente casuística, aspecto que é controlado ponderando a duração da estadia pelo peso de cada DRG atendido. Este uso permite o desenvolvimento de alguns indicadores. A partir desses indicadores é possível revisar os processos nos DRGs em que o hospital se mostra ineficiente para que sejam melhorados.

Outros usos incluem os reembolsos cruzados entre hospitais ou comunidades autônomas pelos pacientes que são atendidos em outra comunidade ou em outra área, se funcionarem sob orçamento per capita. Na Catalunha, o DRG foi empregado parcialmente para o reembolso dos hospitais da *Xarxa Hospitalaria d'Utilització Pública*.

As formas típicas de utilização incluem:

1. Cálculo da **internação média ajustada por funcionamento** (EMf), isto é, a EM que seria apresentada por um hospital

<sup>3</sup> As unidades de terapia intensiva utilizam diversos sistemas (APACHE, SAPS, MPM) com finalidade prognóstica. Do mesmo modo, em algumas especialidades são utilizadas diversas escalas habituais na prática clínica: o ASA pelos anestesistas, o POSSUM em cirurgia, a escala de coma de Glasgow e outras, são exemplo dessas escalas.



com os pacientes (DRG) que atendeu se mantivesse em cada grupo a EM padrão de referência, geralmente a média de uma base de dados com diversos hospitais similares.

2. Uma segunda opção é calcular a **internação média ajustada por casuística** (EMc), aquela que um hospital apresentaria se usasse a mesma casuística que o padrão de referência, mas com suas EM em cada DRG. No primeiro caso é oferecida uma ideia da complexidade do hospital, enquanto que no segundo a casuística é controlada enquanto fator de confusão no prolongamento da internação média.

O produto mais típico em gestão hospitalar é uma listagem que inclui todos os DRGs, ordenados por volume, a EMc do hospital e a EMP, e a estadia a mais ou a menos que teve um hospital em relação ao padrão de referência. Isto permite atuar na gestão daqueles DRG com maior estadia ajustada do que o padrão da base de dados.

Note que **o DRG considera o contexto hospitalar como um todo - não é específico para um determinado serviço - avaliando assim o conjunto de cuidados do hospital e sua coordenação**. Assim, um paciente com um enfarte agudo de miocárdio pode ter passado pela urgência, UTI, cardiologia, cirurgia cardiovascular e reanimação, e ter recebido serviços de radiodiagnóstico, análises clínicas e outros. Considerar que um possível excesso de permanência depende do serviço que lhe deu alta (cardiologia) e não do conjunto do atendimento é um erro muito comum na gestão hospitalar do SNS e que deve ser evitado.

Deve-se observar que os DRG são uma boa aproximação para a eficiência produtiva (respondem à pergunta de se “produzir” uma alta hospitalar se mostra mais ou menos dispendioso em um ou outro serviço) e têm sua lógica nos sistemas de saúde que reembolsam seus hospitais por alta (pagamento prospectivo, como é o caso dos EUA). Os DRGs não dão informações sobre outros aspectos importantes da eficiência, como se um serviço opera mais ou menos do que outros, ou se obtém mais ou menos receita ou intervenções desnecessárias. Uma vez que existem relações entre o volume produzido e os custos unitários (os custos fixos são divididos entre uma maior quantidade de altas e cada uma delas é mais barata, embora o agregado implique

*O DRG considera o contexto hospitalar como um todo - não é específico para um determinado serviço - avaliando assim o conjunto de cuidados do hospital e sua coordenação*



um gasto maior), o reembolso pelo DRG incentiva a utilização de serviços e a eficiência dos provedores não deve ser avaliada somente com este indicador.

#### **4. Sistemas de ajuste de riscos: perspectivas**

Embora os sistemas de ajustes tenham importantes limitações, a comparação *ajustada* de organizações pode atender aos objetivos básicos de:

- ✓ Proporcionar a gestores, compradores, usuários e responsáveis pela tomada de decisão informações, ajustadas por gravidade, para comparar o custo, a utilização e a qualidade dos provedores de serviços de saúde,
- ✓ Proporcionar aos provedores de serviços uma referência de seus custos e qualidade em relação a outros provedores, ajustada de acordo com a gravidade de seus pacientes.
- ✓ Proporcionar aos clínicos uma referência da eficácia de um tratamento/exame, e sobre quais grupos de pacientes têm a maior probabilidade de obter um benefício ou de acarretar um dano concreto.

A utilidade destas comparações para os serviços de saúde é evidente. Se uma UTI tem uma mortalidade intra-hospitalar que é o dobro da de outras unidades depois de seu ajuste pelo APACHE III ou o MPM, este serviço deve revisar seus processos de atendimento para identificar possíveis problemas de qualidade. Este caso é extensivo a muitas outras situações (infecções hospitalares, complicações cirúrgicas, mortalidade intraoperatória, reinternações por determinadas causas etc.). De outro ponto de vista, se os pacientes submetidos à reparação de hérnia inguinal têm uma taxa de recidivas ao ano cinco vezes maior do que das intervenções com reforço de parede, valeria a pena realizar um estudo causal? Estes resultados ocorrem em todos os grupos de pacientes? Há grupos em que não há diferenças?

Para os compradores de serviços de saúde as questões de fundo são do tipo: se um hospital tiver maiores custos para os mesmos processos, decorrentes do fato de que sua taxa de infecções

hospitalares é o triplo de outros hospitais, por que reembolsar estes custos e financiar os custos de não qualidade? E os maiores custos decorrentes do uso desnecessário de estadias ou testes diagnósticos? Por que a população para a qual os serviços são comprados deve estar sujeita a um maior risco de mortalidade? Embora a utilidade das comparações para os compradores de serviços decorra - em boa parte - de sua capacidade real de compra (em função da existência de um mercado de atendimento de saúde e de usuários que escolhem um ou outro comprador segundo o custo das prestações e qualidade do atendimento - aspectos não tão evidentes no contexto espanhol), o acúmulo das funções de comprador de serviços e administrador público, aliado à exigência de gestão eficaz e atendimento de qualidade, amplia a utilidade destes ajustes no Sistema Nacional de Saúde.

As principais limitações desta aproximação decorrem da disponibilidade das variáveis essenciais para o ajuste, da qualidade das fontes de dados e do desempenho do sistema de ajuste de riscos empregado. Estas limitações não são facilmente evidenciadas (etampouco a tentação de desenvolver comparações mecanicistas). Não se deve esquecer que os resultados a partir de bases de dados são sempre retrospectivos, e é pouco provável que, inclusive nas bases com maior quantidade de informação, possam ser ajustadas todas as variáveis relevantes e tratá-las "como se não tivesse existido seleção".

O desenvolvimento - e utilização - de perfis de prática a partir do CMBD é um dos desafios enfrentados pelo Sistema Nacional de Saúde na medida em que a monitoração de resultados ajustados por gravidade é uma necessidade para a prática clínica, a gestão sanitária e qualquer política de incentivos.

Diante de qualquer de suas limitações, o ajuste de gravidade - incluindo custos - a partir do CMBD, tem as vantagens de seu baixo custo relativo e de estar disponível ou facilmente disponível. Utilizados de forma criativa e com visão crítica, os sistemas orientados para os resultados constituem um instrumento fundamental e de indiscutível utilidade para a tomada de decisão em gestão sanitária, para a avaliação da qualidade assistencial e para a monitorização do comportamento dos fornecedores de serviços, o que não significa que sejam o único instrumento, que vão substituir os indicadores de produtividade clássicos ou que não possam ser usados combinados com outros indicadores.

Em muitos casos, as aproximações realizadas a partir de seus dados podem não ser suficientes para julgar de forma definitiva a qualidade ou a eficiência de um hospital ou um serviço médico, mas quase sempre podem ser úteis para identificar problemas que exijam um estudo posterior. A combinação de CMBD e revisão de históricos clínicos (*audit* médico) pode ser a chave das melhorias na qualidade e na eficiência do atendimento hospitalar.

## Referências bibliográficas

1. Blumberg MS. Risk adjusting health care outcomes: a methodologic review. **Med Care Rev.** **1986**; 43:351-393.
2. Casas M. Los grupos relacionados de diagnóstico: experiencia y perspectivas de utilización. Barcelona: Masson y SG editores, 1991.
3. Hornbrook MC. Techniques for assessing hospital case mix. **Annu Rev Public Health.** **1985**; 6:295-324.
4. 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**;123:763-70. [<http://www.annals.org/content/123/10/763.full.pdf+html?sid=2507d74d-1026-491d-a728-3dc57bea2e63>]
5. Iezzoni LI, Ed. Risk adjustment for measuring health care outcomes (2nd edition). Ann Arbor, Michigan: Health Administration Press, 1996.
6. Iezzoni LI. The risks of risk adjustment. **JAMA.** **1997**;278:1600-7. [<http://jama.ama-assn.org/content/278/19/1600.abstract>]
7. Librero J, Peiró S, Ordiñana R. Comparación de resultados, calidad y costes usando bases de datos. Valencia: Instituto Valenciano de Estudios en Salud Pública, 1998.
8. Peiró S, Casas M. Análisis comparado de la actividad y resultados de los hospitales. Situación en España y perspectivas. En: Cabasés JM, Villalbí JR, Aibar C, eds. Invertir en Salud. Prioridades para la salud pública en España-Informe SESPAS, 2002. Valencia: SESPAS y Escuela Valenciana de Estudios para la Salud, 2002; 511-29. [<http://www.sespas.es/informe2002/cap24.pdf>]

9. Peiró S, Librero J. Evaluación de calidad a partir del conjunto mínimo de datos básicos al alta hospitalaria. **Rev Neurol.** **1999**;29:651-61. [<http://www.revneurol.com/sec/resumen.php?id=99330#>]
  
10. Peiró S. Los mejores hospitales. Entre la necesidad de información comparativa y la confusión. **Rev Calidad Asistencial.** **2001**;16: 119-30. [<http://www.elsevier.es/sites/default/files/elsevier/pdf/256/256v16n02a13028294pdf001.pdf>]