

Ajustamento pelo risco: da conceptualização à operacionalização

CARLOS COSTA

As organizações de saúde, em geral, e os hospitais, em particular, são frequentemente reconhecidos por terem particularidades e especificidades que conferem uma especial complexidade ao seu processo produtivo e à sua gestão (Jacobs, 1974; Butler, 1995).

Neste sentido, na literatura hospitalar emergem alguns temas como prioritários tanto na investigação como na avaliação do seu funcionamento, nomeadamente os relacionados com a produção, com o financiamento, com a qualidade, com a eficiência e com a avaliação do seu desempenho.

O estado da arte da avaliação do desempenho das organizações de saúde parece seguir a trilogia definida por Donabedian (1985) — Estrutura, Processo e Resultados. Existem diversas perspectivas para a avaliação do desempenho na óptica dos Resultados — efectividade, eficiência ou desempenho financeiro. No entanto, qualquer que seja a utilizada, o ajustamento pelo risco é necessário para se avaliar a actividade das organizações de saúde, como forma de medir as características dos doentes que podem influenciar os resultados de saúde.

Como possíveis indicadores de resultados, existem a mortalidade (resultados finais), as complicações e as readmissões (resultados intermédios). Com excepção dos estudos realizados por Thomas (1996) e Thomas e Hofer (1998 e 1999), praticamente ninguém contesta a relação entre estes indica-

dores e a efectividade dos cuidados. Chamando, no entanto, a atenção para a necessidade de se definirem modelos de ajustamento pelo risco e ainda para algumas dificuldades conceptuais e operacionais para se atingir este objectivo.

Em relação à eficiência técnica dos hospitais, os indicadores tradicionalmente mais utilizados para a sua avaliação são os custos médios e a demora média. Também neste domínio, a grande maioria dos estudos aponta para que a gravidade aumenta o poder justificativo do consumo de recursos e que o ajustamento pelo risco é útil para avaliar a eficiência dos hospitais.

Em relação aos sistemas usados para medir a severidade e, consequentemente, ajustar pelo risco, o seu desenvolvimento apresenta, na generalidade, dois tipos de preocupações: a definição dos suportes de recolha da informação e a definição dos momentos de medição.

Em última instância, o dilema que se coloca reside na definição de prioridades e daquilo que se pretende sacrificar. Quando se entende que os aspectos financeiros são determinantes, então será natural que se privilegie o recurso quase exclusivo a elementos dos resumos de alta como suporte de recolha da informação. Quando se defende que a validade de construção e de conteúdo é um aspecto a preservar, então o recurso aos elementos dos processos clínicos é inevitável.

A definição dos momentos de medição dos dados tem repercussões em dois níveis de análise: na neutralidade económica do sistema e na prospectividade do sistema.

O impacto destas questões na avaliação da efectividade e da eficiência dos hospitais não é uma questão pacífica, visto que existem autores que defendem a utilização de modelos baseados nos resumos de alta, enquanto outros defendem a supremacia dos modelos baseados nos dados dos processos

□

Carlos Costa é professor auxiliar da Escola Nacional de Saúde Pública — UNL.

Entregue em Março de 2006.

clínicos, para finalmente outros argumentarem que a utilização de uns ou outros é indiferente, pelo que o processo de escolha deve obedecer a critérios mais pragmáticos, como a sua exequibilidade e os respectivos custos de implementação e de exploração.

Em relação às possibilidades que neste momento se colocam em Portugal para a utilização e aplicação de sistemas de ajustamento pelo risco, verifica-se que é praticamente impossível a curto prazo aplicar modelos com base em dados clínicos. Esta opção não deve impedir que a médio prazo se altere o sistema de informação dos hospitais, de forma a considerar a eventualidade de se utilizarem estes modelos.

Existem diversos problemas quando se pretendem aplicar sistemas de ajustamento de risco a populações diferentes ou a subgrupos distintos das populações donde o sistema foi originalmente construído, existindo a necessidade de verificar o ajustamento do modelo à população em questão, em função da sua calibração e discriminação.

Palavras-chave: ajustamento pelo risco; avaliação de cuidados de saúde; resultados; severidade; sistemas de informação; hospitais; administração hospitalar.

1. Introdução

As organizações de saúde, em geral, e os hospitais, em particular, são frequentemente reconhecidos por terem particularidades e especificidades que conferem uma especial complexidade ao seu processo produtivo e à sua gestão (Jacobs, 1974; Butler, 1995). Estas questões estão associadas às condições próprias do mercado onde operam, onde avultam os aspectos relacionados com a assimetria de informação, com a limitação na soberania do consumidor, com a existência de uma integração vertical incompleta, com a procura derivada, com a relação de agência e com a irrelevância do factor preço para racionalizar a tomada de decisões (Evans, 1984).

Por outro lado, avultam ainda alguns aspectos internos, como a existência de uma dupla linha de autoridade, uma decisão sectorizada, o carácter multiproducto da sua actividade e a existência de um critério de sucesso, por vezes formal e pouco explícito (Costa e Reis, 1993), que tornam ainda os hospitais mais diferentes das tradicionais organizações que operam no mercado em geral.

Neste sentido, na literatura hospitalar emergem alguns temas como prioritários tanto na investigação como na avaliação do seu funcionamento, nomeadamente os relacionados com a produção, com o financiamento, com a qualidade, com a eficiência e com a avaliação do seu desempenho.

A medição da produção hospitalar assume especial importância tanto por questões estruturais, associadas

à natureza intrinsecamente social e delicada do seu produto, como por questões operacionais, que passam desde o simples conhecimento e comparação da sua actividade a questões mais técnicas, como a avaliação dos cuidados prestados.

Além disso, a medição da produção hospitalar ganha também especial relevo porque existem várias razões, tanto associadas à procura como à oferta, que dificultam a identificação de produtos e, naturalmente, a sua comparação (Hornbrook, 1982).

O conhecimento e eventual publicitação do desempenho dos hospitais é importante não só para se poderem comparar estas organizações de saúde, mas igualmente para se poder reduzir a actual assimetria de informação (Griffith, Alexander e Jelinek, 2002). Assim, as avaliações da efectividade e da eficiência dos hospitais constituem-se como dimensões relevantes para uma perspectiva mais ampla — a avaliação do desempenho dos hospitais [Iezzoni, 1995; Hughes *et al.*, 1996; Healthcare Information Agency (HCIA), 1999].

A dificuldade em avaliar o desempenho das organizações de saúde tem sido evidenciada em diversos estudos nacionais e internacionais, sendo de referir os aspectos associados com a multiplicidade de perspectivas e de agentes, a existência de uma dupla linha de autoridade e a especificidade na garantia da qualidade dos cuidados prestados (Jacobs, 1974; Donabedian, 1980).

Na grande maioria das organizações, o seu desempenho é avaliado pelos resultados (Costa e Reis, 1993). Nas organizações não lucrativas em geral e nas organizações de saúde este axioma tem sido posto em causa (Evans, 1981; McGuire, Henderson e Money, 1988).

O estado da arte da avaliação do desempenho das organizações de saúde parece seguir a trilogia definida por Donabedian (1985) — Estrutura, Processo e Resultados.

Em que estrutura pode ser definida pelas características da oferta (organizações de saúde); processo por aquilo que é feito ao doente e resultados pela forma como o doente responde aos cuidados que lhe são prestados (QMAS, 1997).

Na realidade, embora esta trilogia tivesse originalmente sido concebida para aplicação na garantia da qualidade dos cuidados prestados, parece igualmente legítimo afirmar que esta perspectiva é aplicável a outras realidades da actividade hospitalar, designadamente da eficiência.

Nas experiências dos EUA, por exemplo, deve referir-se que, enquanto na metodologia *Best Hospitals* (Hill, Winfrey e Rudolph, 1997; O'Muircheartaigh, Murphy e Moore, 2002) e no modelo preconizado por DesHarnais e outros (1997 e 2000) se privilegia

essencialmente a dimensão qualidade (com a particularidade de a primeira encerrar as três dimensões da qualidade, enquanto a última foca essencialmente os resultados), o modelo desenvolvido pela HCIA (1999) e por Griffith, Alexander e Jelinek (2002), pelo contrário, envolve as perspectivas da qualidade, eficiência e rendibilidade dos capitais, embora considere somente uma dimensão — os resultados.

No Reino Unido, embora existam basicamente duas grandes abordagens, a primeira baseada na definição de indicadores do Serviço Nacional de Saúde (Chang *et al.*, 2002) e a segunda interessada na concretização de uma metodologia para estabelecer *hospital ratings* (NHS, 2002; Snelling, 2003), ambas englobam indicadores de estrutura, processo e resultados.

Na Austrália, embora os modelos utilizados sejam completamente diferentes, oscilam entre indicadores de qualidade, com primazia no processo, e outros mais gerais, com indicadores sobre financiamento, acesso, eficiência e qualidade, nos quais a perspectiva privilegiada é o processo (Ibrahim *et al.*, 1998; Degeling *et al.*, 2000; NHPC, 2001; ACHS, 2002). Esta discussão internacional, para além de não se revelar pacífica, reflecte na generalidade uma filosofia básica: enquanto nos EUA se pensa que, se os resultados das organizações de saúde forem razoáveis, então é natural que a estrutura e o processo apresentem o mesmo sentido, no Reino Unido e na Austrália, bem como noutros países europeus, pensa-se que, se a estrutura e o processo apresentarem bons indicadores, então é natural esperar que os resultados das organizações de saúde sejam igualmente bons. Finalmente, deve referir-se que, independentemente da perspectiva que se utilize para avaliação do desempenho — efectividade, eficiência ou desempenho financeiro —, o ajustamento pelo risco é necessário para se avaliar a actividade das organizações de saúde, visto que é necessário medir as características dos doentes que podem influenciar os resultados de saúde.

2. Ajustamento pelo risco — razão de ser

Em termos genéricos, a principal finalidade do ajustamento pelo risco é a de controlar os factores que os doentes apresentam ao contactarem uma determinada organização de saúde que podem afectar a sua probabilidade de obterem um bom ou um mau resultado (Iezzoni, 1996a). Assim, o *rationale* para o ajustamento pelo risco consiste na eliminação dos factores que podem conduzir a diferentes resultados de saúde, essencialmente aqueles que estão associados aos doentes (Iezzoni, 1997b e 1997c). Contudo, algumas questões permanecem ainda em grande debate

(Iezzoni, 1996b e 1997c; Thomas e Hofer, 1998 e 1999):

- O que é que significa, exactamente, risco?
- Como é que o risco deve ser medido?

A primeira questão que se deve colocar respeita ao risco de quê? Para esta interrogação pode existir uma grande variedade de respostas que vão desde o consumo de recursos para a probabilidade de morte iminente, para a existência de complicações ou ainda para o desempenho funcional dos doentes e para a satisfação do cliente.

Este aspecto conduz a duas situações. A primeira é que não existe uma única abordagem de ajustamento pelo risco que apresente o mesmo nível de significância para todos os diferentes resultados de saúde (Iezzoni, 1996b).

A segunda é que existe uma grande variedade de abordagens que pretendem estabelecer medidas de ajustamento pelo risco (Costa, 1991; Iezzoni, 1997c). Por outro lado, o ajustamento pelo risco tem sido frequentemente associado à questão da severidade do estado do doente.

Este assunto é ainda mais importante quando se tem presente que não existe uma definição única sobre severidade/gravidade.

Na realidade, Gertman e Lowenstein (1984) sintetizam esta dificuldade na definição de severidade, avançando que severidade é aquilo que os sociólogos designam, como uma expressão da sabedoria popular, por satisfação ou felicidade, operacionalmente indefinível, mas facilmente aceite por todos os intervenientes.

Ou ainda, como referem Thomas e Longo (1989), que, ao executarem uma extensa revisão de literatura sobre a severidade, encontraram definições tão distintas como probabilidade de morte, diminuição das funções, «carga da doença», integridade fisiológica ou dificuldade de repor o nível de saúde dos doentes. Tudo isto tem levado a uma grande proliferação de sistemas de classificação de doentes que pretendem medir a severidade, cada qual com a sua definição (Costa, 1991; Iezzoni, 1997b e 1997c). Ou a que existam diversas perspectivas, conforme os agentes, variando por especialidades médicas e entre médicos e gestores, estando os primeiros mais interessados nos resultados de saúde e os segundos com a necessidade de recursos e com a intensidade do consumo de recursos (Iezzoni, 1995).

Assim, a primeira questão que se coloca quando se pretende definir o sistema de ajustamento pelo risco reside no seu âmbito de aplicação:

«A definição de produtos deve incidir essencialmente sobre a capacidade de um sistema definir produtos

homogéneos em relação ao consumo de recursos, em relação à significância clínica ou ainda em relação a ambos?»

A resposta a esta questão parece, numa primeira abordagem, óbvia e simples: o sistema de classificação de doentes deve estabelecer categorias que simultaneamente proporcionem a sua homogeneidade clínica e económica.

Neste sentido, a criação de um sistema misto possibilita não só a sua múltipla utilização, conforme as finalidades a que se destina, mas também uma potencial minimização dos conflitos entre as duas linhas de autoridade presentes no hospital, visto que o plano de acção e de análise se centra no mesmo instrumento. Contudo, a operacionalização deste princípio tem suscitado diversos problemas e contradições.

Em primeiro lugar, porque, conforme referem Hornbrook (1982) e Aronow (1988), entre outros, a definição prévia da finalidade principal a que um sistema se destina constitui quase exclusivamente a única metodologia objectiva para se avaliar a sua efectividade.

Em segundo lugar, atendendo a aspectos de uma natureza mais prática, essencialmente os que se referem às necessidades imediatas de utilização do sistema, para financiamento ou para garantia de qualidade, por exemplo, e ainda aos próprios custos de implementação e de exploração do sistema, têm originado uma proliferação de sistemas de classificação dos doentes para se medir exactamente a mesma realidade.

Tudo isto tem conduzido a uma primazia do princípio da definição da finalidade, remetendo-se para um momento posterior a investigação sobre a adequação do sistema de classificação de doentes a outras finalidades.

Assim, não é de estranhar que a grande discussão que ainda existe no panorama internacional se situe sobre a predominância de uma perspectiva em relação à outra, com sistemas que privilegiam a utilização de recursos, como, por exemplo, os Diagnosis Related Groups (DRGs) ou o Acuity Index Method (AIM), outros a severidade, como o Acute Physiologic and Chronic Health Evaluation (APACHE), o System of Acute Physiologic Score (SAPS), o Mortality Probability Model (MPM), o Disease Staging ou os MedisGroups, por exemplo, enquanto outros afirmam que substanciam os dois princípios, como os Patient Management Categories (PMCs), por exemplo (Iezzoni, 1997b).

A definição e proliferação de sistemas de classificação de doentes para a medição da severidade do doente têm suscitado diversas questões, dentro das quais avultam os aspectos relacionados com as dimensões do risco, com a existência de associação

entre os resultados de saúde (mortalidade, complicações e readmissões) e a qualidade dos cuidados prestados ou com a eficiência, a escolha de modelos de ajustamento pelo risco e as propriedades estatísticas dos modelos de ajustamento pelo risco.

Atendendo, à sua importância conceptual e operacional, são estes os aspectos que se passarão a analisar com maior profundidade.

3. Dimensões do risco

Actualmente, é quase pacificamente aceite que as dimensões de risco são as seguintes (Knaus *et al.*, 1985 e 1991; Blumberg, 1986; Horn, 1988; Murphy e Cluff, 1990; Le Gall, Lemeshow e Saulnier, 1993; Lemeshow *et al.*, 1993; Iezzoni, 1997d):

- Idade;
- Sexo;
- Estado fisiológico do doente;
- Diagnóstico principal;
- Gravidade do diagnóstico principal;
- Dimensão e gravidade das comorbidades;
- Situação/estado (*status*) funcional;
- Situação psicológica e cognitiva dos doentes;
- Atributos culturais, éticos e sócio-económicos;
- Atitudes e preferências dos consumidores.

A conjugação destes factores permite igualmente afirmar que um determinado sistema de classificação de doentes tem significância clínica, definindo-se significância clínica (Wood, Ament e Kobrinsky, 1981) pela extensão e compreensão do conhecimento da situação de saúde de cada doente proporcionada pelo sistema de classificação, sendo estas adaptáveis às expectativas dos médicos, permitindo simultaneamente, sem outros elementos adicionais, a troca de informações sobre essas mesmas expectativas.

3.1. Idade

A idade é uma característica intrínseca dos doentes que não depende da efectividade dos cuidados prestados. Por outro lado, é igualmente aceite que, para a grande maioria das situações clínicas, os idosos apresentam maiores probabilidades de desenvolverem situações adversas do que a população jovem (Iezzoni, 1997d).

Neste sentido, tem sido defendida a inclusão da idade como factor de risco, independentemente das outras características consideradas. Esta posição é bem patente na grande maioria dos sistemas desenvolvidos até ao momento (APACHE, MPM, ou SAPS, por

exemplo), tendo, inclusivamente, sido evidenciada a sua ausência como justificativa de desempenhos inferiores de alguns sistemas de medição da severidade, como, por exemplo, é patente no estudo de Blumberg (1991), em que, utilizando os MedisGroups, ficou provado que a consideração da idade aumentava o poder explicativo deste sistema.

O aumento da idade não só aumenta o risco de morte, mas também a probabilidade de se desenvolverem complicações (Goldman *et al.*, 1977; Weintraub *et al.*, 1991; Forman *et al.*, 1992).

Poderão ainda ocorrer alguns problemas na consideração da efectividade média dos hospitais para a categoria dos muito idosos. Na realidade, embora subsista alguma polémica, alguns estudos evidenciam que os idosos são tratados com menos intensidade para as mesmas situações comparativamente com populações mais jovens (Greenfield *et al.*, 1987; Bennet *et al.*, 1991).

Outro aspecto eventualmente a considerar no factor idade é a distinção entre recém-nascidos, crianças, adolescentes e adultos. De facto, para além da consideração de um diferente conjunto de variáveis explicativas para o risco de morte, podem ainda existir diferentes níveis de «normalidade» e de «anormalidade» para as mesmas variáveis em função da idade (Pollack, Ruttiman e Getson, 1988; Pollack, Patel e Ruttiman, 1996).

Para obviar a esta situação alguns estudos utilizam somente as populações adultas (com 18 ou mais anos) para se apurar o risco de morte, como, por exemplo, se pode constatar nos artigos de Iezzoni e outros (1995a) e de Thomas, Ashcraft e Zimmerman (1986).

3.2. Sexo

As diferenças entre os sexos podem igualmente constituir um factor suplementar para se avaliar o risco de morte tanto pelas diferenças anatómicas, fisiológicas e hormonais existentes como pelo facto de existirem algumas situações em que a resposta à terapêutica é diferente entre os dois sexos (Iezzoni, 1997d).

Ainda neste sentido, um estudo realizado por Gordon e Rosenthal (1999), para além de identificar diversos estudos com diferentes riscos entre o sexo masculino e o sexo feminino, acaba por concluir pela existência de um risco superior de morte para os homens em quatro doenças com tratamento médico (doença cerebrovascular, doenças respiratórias das vias superiores, hemorragia gastrointestinal e pneumonia), embora para o enfarte agudo do miocárdio e para a insuficiência cardíaca os riscos de morte sejam praticamente idênticos entre os dois sexos.

Contudo, a grande maioria dos sistemas desenvolvidos até ao momento não inclui o sexo como factor de risco da morte, visto que o seu poder explicativo não tem sido significativo (Francks, Gold e McClancy, 1996; Iezzoni, 1997d).

3.3. Estado fisiológico do doente

A avaliação da estabilidade clínica reflecte o funcionamento fisiológico em termos de equilíbrio do estado geral de saúde (Iezzoni, 1997d).

Esta avaliação do estado fisiológico é considerada de importância crucial na previsão da mortalidade a curto prazo, visto que estas variáveis representam as funções orgânicas que mantêm os doentes vivos (Iezzoni, 1997d).

É devido a esta natureza global que frequentemente se considera que o resultado de cada variável é independente do diagnóstico/doença, sendo, consequentemente, um bom indicador do risco de cada doente (Knaus, Wagner e Draper, 1984; Knaus *et al.*, 1985 e 1993; Le Gall, Lemeshow e Saulnier, 1993; Lemeshow *et al.*, 1993).

Estes mesmos autores referem ainda que com poucas variáveis, as quais, por sua vez, constituem rotinas dos hospitais, se conseguem alcançar sistemas de classificação de doentes com significância clínica e igualmente com bons níveis de previsão do risco.

Apesar destas vantagens, é igualmente referido que estas variáveis fisiológicas podem não constituir o único vector de previsão da mortalidade hospitalar. É essencialmente atendendo a este facto que sistemas como o APACHE ou MPM, por exemplo, incluem outro tipo de variáveis para complementar o modelo para previsão da mortalidade, como é o caso das comorbilidades presentes no momento de admissão ou, inclusivamente, o tipo de contacto com o hospital (Knaus, Wagner e Draper, 1984; Knaus *et al.*, 1985 e 1993; Le Gall, Lemeshow e Saulnier, 1993; Lemeshow *et al.*, 1993).

Por outro lado, pese embora o facto de a medição do estado fisiológico do doente poder ser global, existem algumas correntes que defendem a inclusão de escalões de variação por variáveis distintas em função do diagnóstico principal (Horn, 1986, 1988 e 1997) ou ainda a consideração de pesos específicos por diagnóstico (Daley *et al.*, 1988).

Finalmente, deverá ser tomada em consideração a metodologia para escolha das variáveis representativas do estado fisiológico do doente.

Basicamente, três alternativas se colocam (Gustafson *et al.*, 1986; Iezzoni, 1997b):

- Utilização de painéis de peritos;

- Utilização de painéis de peritos, com posterior investigação/modelização empírica;
- Investigação/modelização empírica.

Na prática, as três alternativas têm sido utilizadas. Sistemas como o Disease Staging, PMCs, MedisGroups e APACHE I e II, por exemplo, recorreram a painéis de peritos para escolher e atribuir *scores* às variáveis. Sistemas como o APACHE III utilizaram painéis de peritos para escolha das variáveis e modelização empírica para atribuição dos *scores* por variável. Sistemas como o MPM utilizaram a modelização empírica para escolha e atribuição de *scores* por variável, tendo, num segundo momento, recorrido a painéis de peritos para discutir a pertinência de todas as variáveis incluídas e/ou a necessidade de incluir outras variáveis não incluídas no modelo.

O principal aspecto a salientar tem residido na transição histórica destas formulações. De facto, os primeiros sistemas a serem desenvolvidos foram baseados quase exclusivamente nas opiniões de painéis de peritos, existindo, inclusivamente, quem defendesse (Gustafson *et al.*, 1986) que este tipo de metodologia era o único possível para a definição de sistemas de medição da severidade.

Mais recentemente, o recurso à modelização empírica tem vindo a ganhar relevo e apoiantes, pelo que parece a lógica mais válida para o desenvolvimento de um modelo de previsão da mortalidade hospitalar.

3.4. Diagnóstico principal

Segundo Hornbrook (1982), o diagnóstico é uma hipótese em relação à natureza da doença, aplicando-se de seguida esta hipótese de diagnóstico no tratamento e na recuperação do doente. Neste sentido, os diagnósticos condicionam a tecnologia de tratamento e, conseqüentemente, o tipo e intensidade de cuidados necessários para tratar cada doença.

Contudo, a identificação do diagnóstico poderá somente constituir o primeiro passo na abordagem do doente, visto que o estabelecimento do prognóstico é igualmente importante na determinação do processo de tratamento de cada doente (Gonnella, Hornbrook e Louis, 1984).

Atendendo a que o prognóstico de cada situação pode ser perspectivado a curto, a médio e longo prazos e a que, frequentemente, este mesmo prognóstico em função do tempo poderá ser diferente e concomitantemente operacionalizado de outra forma, interessa delimitar o âmbito de aplicação deste conceito (Hughes *et al.*, 1996).

Como o principal objectivo desta metodologia consiste na identificação dos produtos do hospital para

ajustamento pelo risco, nesta fase, delimitada ao internamento, interessa muito mais considerar o prognóstico a curto prazo. Assim, a manutenção da vida tem constituído a maior preocupação nesta situação, pelo que o prognóstico é frequentemente considerado em termos de morte (Blumberg, 1986). No entanto, a morte nem sempre constitui o episódio terminal para todos os diagnósticos, pelo que dificuldades adicionais emergem quando se considera o diagnóstico como uma das dimensões do risco (Hill, Winfrey e Rudolph, 1997).

Por outro lado, na comparação de risco entre diferentes diagnósticos têm igualmente subsistido diversos problemas tanto quando se comparam diagnósticos em que a morte é um *outcome* frequente como quando a comparação é exercida entre diagnósticos com probabilidades de morte distintas.

De facto, a comparação da severidade entre doenças tem sido frequentemente referida como uma impossibilidade prática, traduzida, por vezes, em expressões como comparar laranjas com maçãs (Iezzoni, 1996c), ou ainda, de forma mais sistematizada, na impossibilidade de se compararem estádios de severidade no sistema de classificação Disease Staging (Plomman, 1982; Hornbrook, 1985; Aronow, 1988).

Estes aspectos têm suscitado a dúvida sobre a utilização do diagnóstico como um factor de risco de morte.

Como foi referido, alguns sistemas defendem a criação de metodologias de medição da severidade específicas, ou seja, por diagnóstico, como é o caso do Computerized Severity Index (CSI) (Horn *et al.*, 1991). Outros defendem a aplicação de metodologias genéricas, como é o caso da grande maioria dos sistemas desenvolvidos para doentes tratados em unidades de cuidados intensivos (APACHE e SAPS, por exemplo) (Knaus, Wagner e Draper, 1984; Knaus *et al.*, 1985 e 1993; Le Gall, Lemeshow e Saulnier, 1993; Lemeshow *et al.*, 1993). Existem ainda situações mistas, como é o caso do sistema Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatment (SUPPORT) (Murphy e Cluff, 1990), em que, pese embora o facto de o sistema ser genérico na perspectiva da medição de cada uma das variáveis, é frequentemente utilizado na análise isolada por diagnóstico.

3.5. Comorbilidades

Por comorbilidades entende-se a existência de outros diagnósticos que não estão etiologicamente relacionados com o diagnóstico principal (Iezzoni, 1997d). Na grande maioria das situações, a estas comorbilidades está associada a ideia de doenças manifestadas

pelos doentes no momento de admissão do hospital. Estas comorbidades têm sido frequentemente consideradas indicadores válidos para se estabelecerem previsões sobre a morte a médio e longo prazos (Charlson *et al.*, 1987) e a curto prazo (Knaus *et al.*, 1981 e 1985; Hornbrook, 1982; Brewster *et al.*, 1985; Blumberg, 1985; Aronow, 1988; Horn, 1988; Philiphs e Knaus, 1990; Greenfield *et al.*, 1993; Le Gall, Lemeshow e Saulnier, 1993; Lemeshow *et al.*, 1993; Ruttiman, 1994; Brailer *et al.*, 1996; Elixhauser *et al.*, 1998; Gijzen *et al.*, 2001) e, inclusivamente, podem afectar a tolerância dos doentes para receberem determinado tratamento ou mesmo na respectiva resposta à terapêutica (Iezzoni, 1997d; Bleichrodt, Crainich e Eeckhoudt, 2003).

3.6. Situação/estado funcional

A situação funcional é frequentemente considerada outro dos factores de risco de morte. Isto porque, por exemplo, a dificuldade em andar pode resultar de outros factores, como anomalias congénitas, doenças neurológicas, etc. (Iezzoni, 1997d).

Neste sentido, a situação funcional respeita a áreas limitadas do funcionamento físico, mental e social nas actividades diárias. É, em síntese, o resultado do estado de saúde das pessoas (ausência de doença), do seu bem-estar (capacidade para participar activamente na «vida») e da capacidade para suportar e ultrapassar os problemas de saúde (Rubenstein *et al.*, 1989).

Contudo, atendendo a diversos problemas operacionais e conceptuais, a situação funcional não tem sido muito utilizada para a definição de modelos de ajustamento pelo risco.

3.7. Outras dimensões de risco

As restantes dimensões do risco — a situação psicológica e cognitiva dos doentes, os atributos culturais, étnicos e sócio-económicos dos doentes e as atitudes e preferências dos consumidores — são igualmente consideradas relevantes no apuramento da probabilidade de morte (Kaplan, Greenfield e Ware, 1989; Applegate, Blass e Williams, 1990; Epstein *et al.*, 1991; Iezzoni, 1991; Case *et al.*, 1992; Bindman e Grumbach, 1992; Burstin, Lipsitz e Brennan, 1992).

Contudo, limitações de natureza prática, essencialmente as relacionadas com a dificuldade e fiabilidade na recolha dos dados, têm conduzido a que não sejam dimensões muito utilizadas nos sistemas de ajustamento pelo risco.

4. Resultados de saúde e efectividade dos cuidados prestados

Conforme foi referido, a escolha das características dos doentes para o modelo de ajustamento pelo risco depende essencialmente do tipo de resultado que se pretende analisar.

A título perfeitamente exemplificativo, indicam-se alguns tipos de resultados (Iezzoni, 1995):

- Longevidade, morte;
- Estabilidade fisiológica;
- Complicações;
- Doenças crónicas e disfunções fisiológicas (morbilidade);
- Estatuto funcional, incapacidade;
- Desempenho funcional, desvantagem;
- Qualidade de vida;
- Custos de tratamento;
- Utilização de serviços (internamentos, readmissões, consultas);
- Demora média.

Neste momento serão somente discutidas as questões relacionadas com a efectividade dos cuidados prestados, nomeadamente a mortalidade, as complicações e as readmissões (Thomas, Holloway e Guire, 1993; DesHarnais *et al.*, 1997; Hill, Winfrey e Rudolph, 1997; HCIA, 1999; Chang, Lin e Northcott, 2002; NHS, 2002; Griffith, Alexander e Jelinek, 2002; O'Muircheartaigh, Murphy e Moore, 2002; Selim *et al.*, 2002). Dentro destas deve-se ainda evidenciar que existem abordagens que representam resultados intermédios (complicações e readmissões) e um indicador de resultados finais — a mortalidade.

Depois, serão analisadas as questões associadas com a relevância da utilização deste tipo de indicadores para a avaliação da efectividade e da qualidade dos hospitais.

Começando por discutir os aspectos respeitantes à mortalidade, deve referir-se que a grande maioria dos autores considera que existe uma relação entre mortalidade e qualidade dos cuidados prestados (Blumberg, 1987; Dubois *et al.*, 1987; Fink, Yano e Brook, 1989; Iezzoni *et al.*, 1992a, 1992b e 1992c; Krakauer *et al.*, 1992; Landon *et al.*, 1996; DesHarnais *et al.*, 1997; Hill, Winfrey e Rudolph, 1997; Mennemeyer, Morrissey e Howard, 1997; HCIA, 1999; Chang, Lin e Northcott, 2002; Griffith, Alexander e Jelinek, 2002; O'Muircheartaigh, Murphy e Moore, 2002).

Contudo, têm sido discutidos alguns aspectos, como a escolha do tipo de dados que devem ser utilizados e do sistema de classificação de doentes (Iezzoni, 1993), sobre a janela de observação da mortalidade

(Chassin *et al.*, 1989; Garnick, DeLong e Luft, 1995; Hughes *et al.*, 1996) e se a mortalidade é ou não um resultado indesejável (Iezzoni, 1995).

Na situação contrária, isto é, que não existe qualquer relação entre mortalidade e qualidade dos cuidados prestados, encontram-se autores como Thomas e Hofer (1998 e 1999), os quais defendem que a taxa de mortalidade não é um bom indicador daquele atributo.

De facto, Thomas e Hofer (1999) afirmam que a taxa de mortalidade ajustada pelo risco não é um bom indicador de qualidade, defendendo ainda que indicadores de processo são mais adequados para se medir a efectividade dos cuidados prestados.

Estes mesmos autores (Thomas e Hofer, 1998) afirmam que, apesar de todos os esforços dedicados à análise da mortalidade hospitalar, tem sido dada pouca atenção à sua utilidade como indicador da qualidade dos cuidados prestados, podendo, inclusivamente, a publicação de dados deste tipo contribuir para a desinformação do público sobre o desempenho dos hospitais.

Ainda no mesmo artigo, estes autores referem que, pese embora o facto de desde 1991 serem publicados dados nacionais (EUA) sobre o desempenho dos hospitais, sendo a taxa de mortalidade ajustada pelo risco o indicador mais utilizado, estas conclusões têm sido contestadas pelos hospitais com o principal argumento de que dão indicações erróneas e enviesadas sobre a actividade hospitalar (Thomas e Hofer, 1998).

Para sustentarem as suas informações Thomas e Hofer (1998) formulam três hipóteses de investigação:

- Será que os hospitais que têm pior qualidade de cuidados prestados apresentam taxas de mortalidade ajustadas pelo risco mais elevadas?
- Podem os consumidores acreditar que os hospitais com taxas de mortalidade ajustadas pelo risco mais baixas são na realidade prestadores com qualidade?
- Para os hospitais será que as diferenças encontradas nas taxas de mortalidade ajustadas pelo risco estão correlacionadas com as diferenças observadas na qualidade prestada?

Em relação à primeira interrogação, a resposta é positiva, enquanto para as restantes a resposta encontrada pelos autores é negativa.

Assim, estes autores concluem que a sua revisão de literatura sugere que os doentes que receberam cuidados com qualidade mais baixa tiveram taxas de mortalidade mais elevadas. Encontraram igualmente que, em média, os doentes tratados em hospitais que

foram identificados como *outliers* em termos de mortalidade elevada estão em maior risco de receberem cuidados de baixa qualidade. Contudo, algumas reservas e cautelas devem ser adoptadas quando se utiliza a mortalidade como indicador de qualidade.

Num estudo realizado por Bravo e outros (2002) pretendia-se analisar as consequências da regulação e da qualidade de cuidados na mortalidade no Canadá. As principais conclusões apontam para que a sobrevivência dos doentes não foi significativamente influenciada pelo tipo de regulação e que, pelo contrário, os doentes tratados com baixa qualidade apresentaram uma sobrevivência menor do que aqueles que tiveram acesso a cuidados de boa qualidade.

Por outro lado, Fink, Yano e Brook (1989) referem que, apesar das suas limitações, a metodologia da Health Care Financing Administration (HCFA) tem assumido um papel importante na avaliação da qualidade dos cuidados prestados. Referem ainda que uma ligação directa entre a qualidade dos cuidados prestados e as taxas de mortalidade foi estabelecida por Dubois, Brook e Rogers (1987).

A este aspecto acrescem os argumentos apresentados por Iezzoni e outros (1992b) e Iezzoni (1995), nos quais se defende que as principais vantagens da utilização de dados sobre a mortalidade hospitalar são a sua acessibilidade e a sua validade teórica como indicador da qualidade dos cuidados prestados e ainda que a grande maioria dos estudos sobre a efectividade consideram a mortalidade, essencialmente, porque é facilmente definida e medida e apresenta um valor claro para os doentes — na grande maioria das situações a morte é um resultado indesejável.

Em 1997 um estudo efectuado por Mennemeyer, Morrissey e Howard, embora não pondo em causa a utilização da taxa de mortalidade como indicador de qualidade, refere que os principais problemas na metodologia utilizada pela HCFA são os seguintes:

- Os dados dos hospitais apresentam erros e diferenças substanciais na codificação. Estes resultados podem condicionar os valores previsionais dos modelos (posteriormente esta temática será discutida);
- Os sistemas de classificação de doentes não medem com precisão a severidade do estado do doente;
- As medidas de severidade não distinguem entre problemas de saúde presentes no momento de admissão e aqueles adquiridos durante a hospitalização, tendo como principal consequência que os hospitais com pior qualidade parecem ter doentes mais graves;
- Como os hospitais têm diferentes práticas na admissão de doentes, tal pode conduzir a enviesa-

mentos no tipo de doentes tratados e, concomitantemente, nos modelos de ajustamento pelo risco.

Convém ainda ter presente o defendido por Rosen e Green (1987) — as altas taxas de mortalidade não são um problema em si mesmas, devendo a principal preocupação residir na proporção de mortes evitáveis — e ainda por Griffith, Alexander e Jelinek (2002), no qual é afirmado que os hospitais que apresentam melhores resultados tanto na taxa de mortalidade ajustada pelo risco como na taxa de complicações ajustada pelo risco parecem demonstrar um melhor desempenho em termos de uma maior flexibilidade e de uma melhor adaptação às preferências dos consumidores.

Green, Wintfeld e Passman (1990) defendem ainda que as taxas de mortalidade hospitalar podem ser úteis para os doentes e médicos se estes utilizarem estas indicações para escolherem o hospital que oferece a melhor perspectiva de sobrevivência. Contudo, referem igualmente que estes valores de mortalidade podem ser enviesados se a idade, a casuística e as necessidades de cuidados prolongados não forem consideradas para a avaliação do desempenho hospitalar.

Por outro lado, Jencks e outros (1988) referem ainda que a utilização de taxas de mortalidade ajustadas pelo risco poderá revelar grande utilidade para a avaliação da efectividade dos cuidados prestados, desde que a gravidade do doente seja devidamente avaliada no momento de admissão.

Outro aspecto complexo na definição de resultados em saúde respeita à janela de observação, isto é, ao período de tempo no qual o doente é seguido (Iezzoni, 1995). Neste contexto, para além de ser necessário distinguir entre resultados intermédios e resultados finais, é necessário que o período em análise seja indicado, pois podem ocorrer algumas situações em que a morte possa não ser um bom indicador de resultados.

Tudo isto tem suscitado um grande debate sobre o período de observação — mortalidade intra-hospitalar, mortalidade a 30 dias após a admissão ou mortalidade a 180 dias após a admissão (Iezzoni *et al.*, 1995a)?

Em 1995, Garnick, DeLong e Luft concluem que a utilização de dados respeitantes à mortalidade a 180 dias após a admissão conduz a resultados semelhantes aos observados com a mortalidade a 30 dias, referindo ainda que estas conclusões são consistentes com o facto de a mortalidade ocorrida após 30 dias ter pouco a ver com as consequências de um episódio de internamento hospitalar.

Num estudo realizado em 1989, Chassin e outros encontram algumas diferenças entre resultados ava-

liados pela mortalidade intra-hospitalar e pela mortalidade após 30 dias da admissão, afirmando que tal se pode dever essencialmente a diferentes práticas dos hospitais na alta dos doentes.

No campo contrário encontram-se os estudos de Knaus e outros (1985), Le Gall, Lemeshow e Saulnier (1993), Lemeshow e outros (1993) e Iezzoni (1997b e 1997c), nos quais se defende que a grande maioria dos sistemas de classificação de doentes contém algumas variáveis que são caracterizadoras e explicativas da mortalidade intra-hospitalar, embora não contenham as mesmas potencialidades quando se considera um prazo mais dilatado, a 30, 60 ou 180 dias. A este respeito, Iezzoni e outros (1996a e 1996b) referem que apesar de a consideração da mortalidade a 30 dias após a admissão ser importante, essencialmente devido ao facto de existirem políticas diferentes entre hospitais na alta dos doentes, não existem razões para acreditar nem suficiente evidência científica de que os resultados da avaliação dos hospitais em função da mortalidade sejam diferentes quando se utiliza somente a mortalidade intra-hospitalar.

Atendendo a estes aspectos, é de referir que essencialmente por razões operacionais, nomeadamente a disponibilidade e a fiabilidade dos dados, que, a grande maioria dos estudos efectuados utilizam a mortalidade intra-hospitalar como medida de resultados de saúde. Este aspecto parece ainda ser mais pertinente em Portugal, visto que é praticamente impossível seguir a mortalidade dos doentes a 30 ou a 180 dias, visto que estes dados não são recolhidos de uma forma consistente e rotineira pelas organizações de saúde.

Em termos de síntese, no que se refere à mortalidade, pode acompanhar-se o defendido por Iezzoni (1993) — «Embora a credibilidade dos dados hospitalares fosse inicialmente contestada, vários estudos (Steen *et al.*, 1993; Krakauer e Jacoby, 1993) sugerem que a utilização de modelos de ajustamento pelo risco melhora o valor preditivo, o rigor estatístico da análise» — ou ainda por Krakauer e outros (1992) — «Atendendo a que o resultado do tratamento é fortemente dependente da doença e da severidade do estado do doente, a validade da inferência sobre o desempenho dos hospitais através da taxa de mortalidade ajustada pelo risco está fortemente dependente da adequação do sistema de informação e do modelo utilizado para o ajustamento pelo risco» —, pelo que, para além de ser pertinente utilizar a mortalidade como indicador de qualidade, é igualmente necessário escolher um sistema de ajustamento pelo risco.

A relação entre os indicadores de resultados intermédios — complicações e readmissões — e a qualidade

dos cuidados prestados não tem merecido tanta atenção internacional como a existente para a mortalidade.

Se, por um lado, se encontram autores em que esta associação é pacífica (Iezzoni *et al.*, 1992d, 1994b, 1994c e 1997; DesHarnais *et al.*, 1997 e 2000; HCIA, 1999; Arozullah *et al.*, 2003), também tem sido referenciado por outros (Thomas, 1996) que tal associação não está presente.

Este último autor não encontrou nenhuma evidência que sustentasse que os doentes que foram tratados com piores níveis de qualidade tenham apresentado uma maior probabilidade de serem readmitidos. Estes resultados de investigação são consistentes na grande maioria das doze condições estudadas, tendo as exceções, ou seja, os casos com níveis de readmissão distintos, sido encontradas para os casos médicos do foro cardíaco (readmissões a 60 dias) e para as doenças respiratórias (readmissões a 90 dias), mas com um sentido diferente do esperado, ou seja, uma relação directa entre cuidados de qualidade e maiores readmissões (Thomas, 1996).

Num estudo realizado por Iezzoni e colegas em 1994b sobre as complicações foram estabelecidas as seguintes hipóteses de investigação:

- Será que o desempenho relativo dos hospitais, medido pelas diferenças entre taxas de complicações observadas e esperadas, é semelhante entre os serviços de cada hospital?

Os autores concluem que os hospitais onde se observam maiores diferenças entre valores observados e esperados para um determinado serviço também são aqueles onde se observa uma maior diferença na globalidade do hospital.

- Quais as características dos hospitais que estão associadas com as taxas de complicações?

O estudo refere que a dimensão, a função ensino e a oferta de serviços em cirurgia cardiotorácica e de ressonância magnética são as características que conduzem a maiores diferenças entre valores observados e esperados.

- Estas características variam de serviço para serviço?

Os resultados apresentam comportamentos distintos entre casos médicos e cirúrgicos.

- Como é que se compara o desempenho dos hospitais quando se considera a taxa de complicações ou a taxa de mortalidade?

O estudo refere que o desempenho dos hospitais, quando avaliado pelas complicações, não está correlacionado com os valores da mortalidade, apresentando, inclusivamente, uma relação negativa nos casos médicos.

Num estudo realizado por Lawthers e outros (2000) tentou-se validar um instrumento, o «Complications Screening Program» (CSP), ao testar-se se as codificações derivadas da Classificação Internacional das Doenças, 9.^a revisão, «Modificações clínicas» (CID-9-MC), são completas e precisas e ainda se o algoritmo do CSP permite identificar as condições presentes no momento de admissão daquelas adquiridas durante o tratamento. O CSP estabelece seis grupos de risco: grandes cirurgias, pequenas cirurgias, procedimentos cardíacos invasivos, endoscopia, casos médicos e total de casos. Os principais resultados referem que o CSP tem um bom desempenho na identificação de situações que podem ser consideradas complicações, embora com comportamentos distintos entre casos cirúrgicos e médicos, com estes últimos a apresentarem piores valores. No que respeita à distinção entre complicações presentes na admissão e adquiridas durante o episódio de internamento, a situação é problemática, sugerindo-se que a melhor solução passa pela criação de um código especial que permita identificar as situações presentes no momento de admissão.

Outro estudo realizado por Weingart e outros (2000) pretendia igualmente validar o CSP na identificação de complicações e de problemas potenciais na qualidade dos cuidados prestados. Os principais resultados referem que a probabilidade de ocorrer uma situação de baixa qualidade nas situações identificadas pelo CSP foi de 29,5% para os casos cirúrgicos e de 15,7% nos casos médicos. Pelo contrário, a probabilidade de encontrar uma situação com baixa qualidade nas situações não identificadas pelo CSP foi de 2,1% tanto para casos cirúrgicos como médicos. Os autores, à semelhança do referido no estudo de Lamers *et al.* (2000), concluem que o CSP apresenta desempenhos distintos para situações cirúrgicas e médicas, embora defendam que o CSP é eficiente para identificar problemas de qualidade, sugerindo ainda que mais investigação é necessária para se avaliarem as percepções dos médicos sobre a qualidade.

Estes resultados merecem uma série de reflexões suplementares, designadamente as associadas com os sistemas de informação existentes e com as diferentes práticas dos hospitais, tanto em termos de codificação como em termos de tratamento.

No que se refere aos sistemas de informação existentes, deve referir-se que os mesmos não disponibili-

zam dados sobre as complicações presentes no momento de admissão e as adquiridas durante o processo de tratamento. Este facto, só por si, pode justificar que existam situações em que doentes com mais complicações adquiridas durante o processo de tratamento apresentem resultados diferentes em termos de mortalidade devido a diferenciais das condições preexistentes (Iezzoni, 1997b).

As diferentes práticas na codificação, com as necessárias repercussões nos diagnósticos secundários e concomitantemente nas complicações, podem levar a que os hospitais apresentem piores relações entre os valores esperados e os observados nas complicações, embora o mesmo não tenha forçosamente de se passar quando se analisa a mortalidade (Kalish *et al.*, 1995; Weingart *et al.*, 2000).

No que se refere às práticas de tratamento dos hospitais, pode acontecer que alguns hospitais tratem os doentes com maior agressividade e que, consequentemente, aumentem as complicações, embora simultaneamente se observe uma melhor qualidade e uma diminuição das mortes (Weingart *et al.*, 2000).

Neste sentido, deve evidenciar-se que, com excepção dos estudos realizados por Thomas (1996) e por Thomas e Hofer (1998 e 1999), praticamente ninguém contesta a relação entre os indicadores de resultados intermédios e finais e a efectividade dos cuidados, chamando-se, no entanto, a atenção para a necessidade de se definirem modelos de ajustamento pelo risco e ainda para algumas dificuldades conceptuais e operacionais para se atingir este objectivo.

5. Resultados de saúde e eficiência dos cuidados prestados

A eficiência técnica é uma medida que traduz uma relação entre os recursos utilizados e os resultados obtidos (Pereira, 1993).

Tradicionalmente, os indicadores mais utilizados para avaliação da eficiência técnica dos hospitais são os custos médios e a demora média (Evans, 1981; McGuire, Henderson e Money, 1988; Butler, 1995). Existe igualmente uma grande discussão a nível internacional sobre a validade e comparabilidade destes dois indicadores, sendo de evidenciar algumas diferenças entre os valores teóricos e os resultados de estudos empíricos.

Na realidade, atendendo essencialmente à intensidade dos recursos, admite-se que os custos e a duração de internamento tenham comportamentos distintos, visto que, por um lado, a intensidade dos recursos diminui à medida que aumenta a duração de internamento e que, por outro lado, os custos aumentam à medida

que diminui a duração de internamento (Butler, 1995; Evans, Hwang e Nagarajan, 1995).

Por intensidade na utilização de recursos entende-se a quantidade de recursos utilizados num determinado período de tempo, por exemplo, por dia de internamento (Luke, 1979).

No entanto, vários estudos empíricos têm encontrado uma associação forte entre custos ou *charges* e a demora média (Lave e Leinhardt, 1976; Jones, 1985; Coffey e Goldfarb, 1986; Horn *et al.*, 1991; Knaus *et al.*, 1993; Bradbury, Golec e Steen, 1994; Chen e Naylor, 1994; Mozes *et al.*, 1994; Becker *et al.*, 1995; MEDSTAT, 2001; Clark e Ryan, 2002; Polverejan *et al.*, 2003).

Por outro lado, embora existam diversas críticas à utilização da duração de internamento, dentro das quais avultam as questões da intensidade de utilização de recursos e a prática distinta dos hospitais em relação à alta dos doentes (Knaus *et al.*, 1993; Costa, 1994; Butler, 1995), existem igualmente argumentos contra a utilização dos custos médios, designadamente os relacionados com as diferentes práticas contabilísticas dos hospitais (Ament *et al.*, 1982; Berki, Ashcraft e Newbrander, 1984; Brewster *et al.*, 1985; Hornbrook e Monheit, 1985; Thomas, Ashcraft e Zimmerman, 1986; Vladeck *et al.*, 1988).

A este propósito, Mozes *et al.* (1994) referem, inclusivamente, que a utilização da demora média é preferível à dos custos, visto que estes são muito mais sensíveis a alterações nos procedimentos de apuramento tanto em termos históricos como geográficos. Quando se considera a realidade portuguesa, esta polémica é mais fácil de ultrapassar, dado que na sua grande maioria os hospitais portugueses apuram custos por serviço, utilizando ainda critérios de distribuição e imputação de rubricas de despesa distintos. Neste sentido, de seguida serão discutidos os aspectos associados com a duração de internamento ou com a demora média em relação:

- Ao *rationale* ou suporte teórico do ajustamento pelo risco;
- À evidência prática do ajustamento pelo risco para avaliação da demora média;
- À pesquisa de eventuais contradições entre a efectividade dos cuidados prestados e a eficiência técnica.

A discussão sobre a eficiência dos hospitais, mais concretamente sobre a demora média, ganhou maior relevo após a introdução do pagamento prospectivo por caso nos EUA em 1983 (Iezzoni *et al.*, 1996c), sendo frequentemente discutido se a gravidade dos doentes afecta ou não o consumo de recursos (Vladeck, 1984; Coffey e Goldfarb, 1986; Jencks e

Dobson, 1987; Lave e Franck, 1990; Horn *et al.*, 1991; Burns, Wholey e Abeln, 1993; Kelleher, 1993; Chen e Naylor, 1994; Mozes *et al.*, 1994; Voss *et al.*, 1994; Shwartz *et al.*, 1996; Marazzi *et al.*, 1998; Clark e Ryan, 2002).

Praticamente, todos os estudos anteriormente referidos apontam para grandes discrepâncias na demora média entre os hospitais, sendo frequentemente referido que as principais razões justificativas podem ser encontradas nas características dos doentes (idade, doença, gravidade e situação sócio-económica), nas características dos hospitais (dimensão, desenvolvimento tecnológico, função ensino e preferências de tratamento dos prestadores) e nas condições do mercado concorrencial (localização, extensão da oferta e modalidades de pagamento) (Goldfarb, Hornbrook e Higgins, 1983; Coffey e Goldfarb, 1986; Lave e Franck, 1990; Thomas *et al.*, 1998; Rapoport *et al.*, 2003).

Na sua grande maioria, os estudos apontam para a necessidade de se ajustar pelo risco, controlando desta forma algumas das características dos doentes, para melhor se comparar a demora média entre os hospitais, sendo ainda referido que os sistemas de classificação de doentes que apresentam um bom desempenho para prevenir a mortalidade podem não ter um resultado idêntico para estabelecerem previsões sobre custos ou duração de internamento (Voss *et al.*, 1994).

Tal pode ser devido a questões estruturais, como a existência de contradições entre gravidade e consumo de recursos, bem patente na situação dos mortos ou dos cuidados paliativos (Coffey e Goldfarb, 1986; Murphy e Cluff, 1990; Knaus *et al.*, 1993), ou ainda a diferentes relações entre as mesmas variáveis para explicar a mortalidade ou o consumo de recursos (Goldfarb, Hornbrook e Higgins, 1983; Shwartz *et al.*, 1996; Iezzoni *et al.*, 1996c; MEDSTAT, 2001; Houchens, 2002).

Num estudo realizado por Shwartz e outros (1996) para episódios de internamento com fractura da anca, os autores pretenderam investigar a importância da gravidade para explicar variações na duração de internamento ao nível de cada doente e ainda a contribuição das medidas de gravidade para explicar diferenças na demora média entre hospitais, tendo, na generalidade, chegado às seguintes conclusões:

- A gravidade tem um poder explicativo reduzido ao nível da duração de internamento por doente;
- Praticamente todos os sistemas de classificação de doentes utilizados (14 sistemas diferentes) apresentam desempenhos semelhantes na explicação individual da duração de internamento, embora o valor explicativo medido pelo coefi-

ciente de determinação seja mais elevado com os All Patient Refined Diagnosis Related Groups (APRDRGs) e com o Disease Staging (numa versão que inclui as comorbilidades desenvolvida por Naessens *et al.*, 1992);

- A gravidade tem igualmente um bom poder explicativo quando se analisam as diferenças na demora média entre hospitais, tendo ainda todos os sistemas de classificação de doentes apresentado uma boa concordância na ordenação dos hospitais;
- Referem também que o ajustamento pela gravidade revelou pouca importância na ordenação dos hospitais, visto que esta não é significativamente diferente da obtida quando se utiliza a demora média;
- No entanto, afirmam que, apesar dos resultados encontrados, o ajustamento pelo risco revela alguma utilidade para a revisão de utilização e para se melhorar o desempenho na duração do internamento a nível individual, visto que as diferenças encontradas na demora média entre estádios de gravidade são importantes (entre 50% e 100% superiores nos doentes mais graves).

Outro estudo realizado em 1996 por Iezzoni e outros (1996c) pretende estudar as mesmas questões, mas agora com aplicação aos episódios de internamento com pneumonia. Na generalidade, as conclusões são idênticas às referidas no estudo anterior, o que pode ser considerado normal, visto que os autores são os mesmos.

Num estudo realizado por Becker e outros (1995) estudou-se a importância da gravidade (medida pelo APACHE II) para avaliar o desempenho das unidades de cuidados intensivos no que se refere à demora média, consumo de recursos e mortalidade nos doentes que tiveram um intervenção cirúrgica cardíaca. Neste estudo o ajustamento pelo risco demonstrou um bom poder explicativo sobre a demora média no que se refere a grupos de doentes, embora com poder explicativo reduzido a nível individual (Becker *et al.*, 1995).

No entanto, referem ainda que estes resultados são normais, dado que o ajustamento pelo risco não se destina a aplicações individuais (doente a doente), referindo ainda que a comparação entre valores observados e esperados para a demora média para grupos de doentes permite aos médicos melhorar a eficiência dos cuidados prestados (Becker *et al.*, 1995).

Num estudo realizado por Knaus e outros (1993) sobre as variações na demora média e na mortalidade em unidades de cuidados intensivos, os autores referem a importância do ajustamento pelo risco para se avaliarem os resultados.

Referem ainda que a duração de internamento é frequentemente utilizada como uma medida dos custos, apesar dos problemas decorrentes da diferente intensidade de cuidados prestados, que a contribuição da gravidade é mais relevante para explicar diferenças na mortalidade e que para a duração de internamento devem ser consideradas conjuntamente a gravidade e a doença principal para se aumentar o respectivo poder explicativo (Knaus *et al.*, 1993).

Finalmente, estes autores referem ainda a existência de potenciais problemas entre a gravidade e a demora média, essencialmente pelo diferente comportamento apresentado pelos doentes falecidos, recomendando ainda que se devem realizar estudos mais aprofundados sobre a duração de internamento, utilizando interações entre doenças e gravidade (Knaus *et al.*, 1993).

Num estudo realizado por Bradbury, Golec e Steen (1994) foi encontrada uma associação positiva entre a mortalidade ajustada pelo risco e os custos ajustados pelo risco e a demora média ajustada pelo risco, embora com significância estatística somente para a associação mortalidade/demora média. Neste estudo é igualmente referido que as características dos hospitais analisadas (dimensão, recursos humanos, estatuto universitário e localização em áreas urbanas) não estão relacionadas com a demora média.

Num outro estudo realizado por Coffey e Goldfarb (1986) analisaram-se as diferenças entre os DRGs e o Disease Staging para analisar a demora média e o financiamento dos hospitais, sendo de evidenciar os seguintes aspectos:

- Os dois sistemas de classificação de doentes são conceptual e empiricamente diferentes, nomeadamente no facto de os DRGs classificarem doentes por procedimentos, enquanto tal não ocorre no Disease Staging;
- Apesar destas diferenças entre os dois sistemas de classificação de doentes, os mesmos apresentam um desempenho semelhante para explicar as variações na demora média entre hospitais;
- No entanto, existem diferenças significativas nos pagamentos aos hospitais, com as organizações de saúde com maior dimensão a serem beneficiadas pelos DRGs;
- Referem ainda que os hospitais com maior dimensão tratam doentes menos graves, embora com índices de *casemix* dos DRGs mais elevados, essencialmente porque estas organizações aceitam maiores proporções de doentes menos graves para optimizarem a sua ocupação.

Westert e outros (1993) analisaram a variação na demora média entre hospitais e entre médicos no

mesmo hospital. Após ajustarem a demora média pelo risco, os autores referem que, por doença, as diferenças encontradas na demora média por médicos no mesmo hospital são insignificantes quando comparadas com as diferenças na demora média encontradas entre hospitais e que existe uma grande variação na prática médica quando o mesmo médico trabalha em dois hospitais diferentes.

Chen e Naylor (1994) analisaram a variação na demora média dos hospitais para o enfarte agudo do miocárdio. Os autores referem que existe uma grande variação na demora média entre os hospitais, a qual é, no entanto, reduzida quando se retiram os óbitos, as transferências e os doentes com revascularização. Referem ainda que a duração de internamento é uma componente importante dos custos por doente e que as características dos doentes, designadamente a gravidade, contribuem para explicar 12,6% da demora média. Finalmente, referem que as características dos hospitais não revelam utilidade para prever a demora média.

Num estudo realizado por Voss e outros (1994) é referido que a gravidade medida por critérios clínicos produziu melhores previsões para a demora média do que os DRGs, embora uma parte substancial na variação do consumo de recursos fique por explicar. Neste estudo é ainda evidenciado que o ajustamento pelo risco para se avaliar a demora média é mais relevante para os casos médicos, em comparação com os casos cirúrgicos.

Deve ainda referir-se que existe uma variedade de outros estudos (Jones, 1985; Lave e Franck, 1990; Cleary *et al.*, 1991; Horn *et al.*, 1991; Burns, Wholey e Abeln, 1993; Kelleher, 1993; Mozes *et al.*, 1994; Yuan *et al.*, 2000; Kaboli *et al.*, 2001; Clark e Ryan, 2002), os quais, pese embora o facto de terem sido aplicados a populações distintas, evidenciam praticamente os mesmos resultados, nomeadamente que a gravidade aumenta o poder justificativo do consumo de recursos e que o ajustamento pelo risco é útil para avaliar a eficiência dos hospitais.

Finalmente, devem ser discutidas as eventuais contradições entre a optimização da efectividade dos cuidados prestados e a optimização da respectiva eficiência.

Bradbury, Golec e Steen (1994) começam por situar o problema. Por um lado, face à teoria da produção (*input/output*), é de esperar uma relação negativa, ou seja, quanto mais recursos são aplicados, menor é a mortalidade; no entanto, por outro lado, referem outro tipo de argumentos que contrariam esta relação, designadamente o facto de a melhoria dos resultados de saúde estar associada a uma maior eficiência na utilização de recursos (por exemplo, a diminuição das complicações pode implicar uma redução nos

custos) e de uma maior efectividade no diagnóstico e tratamento dos doentes implicar igualmente uma redução no consumo de recursos.

No referido estudo os autores acabam por concluir que existe uma associação positiva entre o consumo de recursos e a mortalidade, embora somente com significância quando se utiliza a duração de internamento para se medir o consumo de recursos (Bradbury, Golec e Steen, 1994).

Outros estudos internacionais apresentam resultados diferentes. Enquanto Stearns (1991) refere que não existem contradições entre a mortalidade e o consumo de recursos, Chen e Naylor (1994) afirmam que não existe qualquer relação entre a mortalidade e a demora média.

Por outro lado, Voss e outros (1994) e Yuan e outros (2000) apresentam diferenças no comportamento dos hospitais para optimizarem a efectividade e a eficiência, enquanto Knaus e outros (1993) e Dismuke e Sena (1998) vão ainda mais longe, visto que avançam com contradições na prossecução dos dois objectivos, embora os primeiros utilizem predominantemente argumentos teóricos e os segundos apresentem resultados empíricos de hospitais portugueses.

Na mesma linha de pensamento, ou seja, comportamento distinto entre a gravidade e a demora média, surgem os artigos de Iezzoni e outros (1996c) e Shwartz e outros (1996), os quais utilizam a probabilidade de morte e a probabilidade de morte ao quadrado para pesquisar se a duração de internamento é menor em estadios de gravidade menos e mais elevados.

Em síntese, pode afirmar-se que subsiste ainda uma grande polémica sobre a eventual existência de contradições para se optimizar a efectividade e a eficiência dos hospitais tanto à luz dos comportamentos teóricos esperados como em função dos resultados empíricos, pelo que se torna um assunto com necessidade de aprofundamento.

6. Modelos de ajustamento pelo risco

A recolha de dados para o desenvolvimento de um sistema de severidade e, conseqüentemente, de ajustamento pelo risco apresenta, na generalidade, dois tipos de preocupações:

- A definição dos suportes de recolha da informação;
- A definição dos momentos de medição.

Existe uma grande discussão sobre os suportes de recolha da informação que devem ser utilizados para o desenvolvimento de um sistema de medição da severidade.

Na realidade, enquanto alguns autores (Fetter *et al.*, 1980; Young, Swinkola e Zorn, 1982; Gonnella, Hornbrook e Louis, 1984; Gonnella *et al.*, 1990; Young, Kohler e Kowalsky, 1994, por exemplo) defendem a utilização dos resumos informatizados de alta, baseados no Uniforme Hospital Discharge Data Abstract (UHDDA), outros (Horn, 1986 e 1988; Knaus *et al.*, 1981, 1985 e 1986; Brewster *et al.*, 1995; Le Gall, Lemeshow e Saulnier, 1993; Lemeshow *et al.*, 1993; Moreno, Apolone e Miranda, 1998, por exemplo) argumentam que os elementos dos resumos informatizados de alta devem ser complementados com outros presentes nos processos clínicos dos doentes.

As principais razões para estas discrepâncias podem ser sintetizadas pelo seguinte balanço: privilegiar a operacionalidade e economia dos sistemas de classificação de doentes ou privilegiar a validade de construção e de conteúdo dos sistemas de classificação de doentes.

Na validade de construção está-se a determinar a correspondência entre os valores apresentados pelo sistema e os valores reais, ou seja, será que o sistema está a medir aquilo que pretende (Hornbrook, 1982; Thomas, Ashcraft e Zimmerman, 1986).

E na validade de conteúdo está-se a apurar a lógica interna dos sistemas para definir categorias e o seu grau de credibilidade e de compreensão por parte dos profissionais de saúde (Hornbrook, 1982; Thomas, Ashcraft e Zimmerman, 1986).

Os princípios de operacionalidade e de economia, embora se encontrem aplicados a planos conceptuais diferentes, apelam para a maior acessibilidade e disponibilidade dos dados proporcionada pelos resumos informatizados de alta.

Por outro lado, a necessidade de recurso suplementar a dados constantes nos processos clínicos dos doentes, para além de dificultar essa mesma acessibilidade e disponibilidade, apela para uma maior compreensão e extensão dos dados recolhidos, o que implicará uma maior consistência e validade dos próprios sistemas de classificação.

Esta distinção tem sido crucial para o desenvolvimento dos próprios sistemas de classificação. Na realidade, os sistemas que utilizam preferencialmente os resumos de alta definem e medem o risco em termos da quantidade dos órgãos afectados. Pelo contrário, os sistemas que suplementarmente recorrem aos elementos dos processos clínicos definem e medem esse risco, não pela quantidade dos órgãos afectados, mas pela extensão da gravidade das insuficiências de um ou mais órgãos.

Em última instância, o dilema que se coloca reside na definição de prioridades e daquilo que se pretende sacrificar.

Quando se entende que os aspectos financeiros são determinantes, então será natural que se privilegie o recurso quase exclusivo a elementos dos resumos de alta. Quando se defende que a validade de construção e de conteúdo de um sistema de classificação de doentes para medição da severidade é um aspecto a preservar, então o recurso aos elementos dos processos clínicos é inevitável.

Um outro plano que simultaneamente deverá ser considerado para a recolha de dados diz respeito à fiabilidade. Em relação ao conceito de fiabilidade, estão associados dois tipos de questões: a possibilidade de o sistema ser utilizado da mesma forma por parte de diferentes classificadores ou pelo mesmo classificador em diferentes períodos de tempo; a possibilidade de o sistema proporcionar aos hospitais a classificação de doentes em situações que lhes são mais favoráveis para efeitos de financiamento ou para otimizarem resultados, por exemplo (Hornbrook, 1982).

Estas questões, bem como outras relacionadas com as próprias medidas de fiabilidade, são discutidas num artigo de Costa e Nogueira (1994). As principais conclusões referem o facto de os sistemas de classificação que utilizam os resumos de alta (basicamente informação sobre diagnóstico principal e diagnósticos secundários) apresentarem na generalidade níveis de fiabilidade mais baixos do que os sistemas que recorrem aos processos clínicos (basicamente informação sobre sinais vitais e resultados de exames complementares de diagnóstico).

As principais razões estão associadas ao maior grau de subjectividade que os sistemas baseados na identificação de diagnóstico apresentam, inclusivamente devido à sobreposição de categorias existente no CID-9-MC (McMahon e Smits, 1986), sistema utilizado para codificação dos diagnósticos nos hospitais, em oposição à objectividade dos elementos proporcionados pelos sinais vitais e resultados dos exames complementares de diagnóstico, em que praticamente os problemas que podem existir dizem respeito somente ou a perdas de informação ou a erros de transcrição, o que confere uma maior objectividade aos sistemas que recorrem a estes elementos (Iezzoni *et al.*, 1995b, 1996d e 1996e; Iezzoni, 1997b).

A definição dos momentos de medição dos dados tem repercussões em dois níveis de análise:

- Na neutralidade económica do sistema;
- Na prospectividade do sistema.

Com a neutralidade económica discute-se se o sistema de classificação deve ou não deve ser indiferente em relação ao desempenho dos hospitais

(Jencks *et al.*, 1984). Ou seja, se a multiplicidade de tratamentos que existem ou estão disponíveis em determinado hospital deve ou não ser considerada mais um factor para desagregar ou definir novos produtos. Por exemplo, um sistema com neutralidade económica considera produtos diferentes situações clínicas com o mesmo diagnóstico e prognóstico mas com tratamento médico ou cirúrgico.

Quando se defende que um sistema de classificação deve ter neutralidade económica, então na sua definição e operacionalização está implícita a incorporação da eficiência de tratamento, da qualidade dos cuidados prestados, da disponibilidade de *inputs*, bem como da natureza e manifestação da doença (Costa, 1994).

Por outro lado, resulta lógico que para a obtenção de um sistema que tenha neutralidade económica devem ser considerados elementos respeitantes a todos os momentos do episódio de internamento. Ou seja, elementos representantes da admissão, da estada (máximos ou médios) e da alta.

Contudo, tem frequentemente sido referido (Thomas *et al.*, 1998; Thomas e Longo, 1990; Iezzoni *et al.*, 1988, 1991a e 1991b; Greenfield *et al.*, 1988) que a inclusão de elementos respeitantes a todo o episódio de internamento, para além de diminuir o risco da actividade das organizações de saúde, pode originar problemas suplementares derivados do registo de doenças iatrogénicas, o que, em última instância, pode limitar a fiabilidade e a validade do sistema de classificação, já que idênticas situações de saúde, à partida, podem conduzir a produtos distintos, dependendo da disponibilidade de *inputs*, da eficiência e da efectividade dos próprios hospitais.

A prospectividade do sistema de classificação constitui um factor relativamente fácil de descrever. A mesma só poderá ser assegurada quando um sistema se baseia nos elementos de admissão, visto que a consideração de elementos da estada e/ou da alta condicionam a respectiva classificação para um período imediatamente posterior à alta do doente (Green, Passman e Wintfeld, 1991).

Atendendo a que no processo de decisão clínica existe a necessidade de decidir a «quente», ou seja, imediatamente a seguir à admissão dos doentes, resulta natural admitir que um sistema de classificação prospectivo será aquele que sintetiza melhor o processo de produção nos hospitais.

Dos argumentos atrás referidos fácil será inferir que um sistema de classificação ideal não deve ter neutralidade económica e deve ser prospectivo.

A forma de contacto com as organizações de saúde, designadamente a explicitação da dicotomia entre admissões programadas ou urgentes, bem como o tipo de tratamento, expresso pela identificação de

tratamentos médicos ou cirúrgicos, têm sido igualmente apontados como factores que podem justificar riscos de morte distintos.

Na realidade, sistemas como o APACHE III ou o SAPS, por exemplo, têm sustentado que a forma de admissão (programada ou urgente) é um factor preditivo de morte (Knaus *et al.*, 1985; Le Gall, Lemeshow e Saulnier, 1993; Carneiro, 1994), com riscos superiores nas admissões urgentes, essencialmente devido ao facto de os doentes admitidos por esta via estarem menos estudados e/ou eventualmente serem mais graves.

Esta mesma conclusão foi retirada por Costa, Delgado e Carvalho (1989) num estudo efectuado num hospital português, o qual, para além de identificar níveis de severidade mais elevados em doentes admitidos pela urgência, referia igualmente um maior consumo de recursos neste tipo de doentes.

Para o tipo de tratamento, médico ou cirúrgico, existe alguma diversidade nos sistemas desenvolvidos até ao presente momento.

Sistemas como o APACHE I, CSI, Disease Staging, MPM e MedisGroups, por exemplo, não consideram que a presença de intervenção cirúrgica seja um factor adicional explicativo da morte (Iezzoni, 1997b). Pelo contrário, os sistemas APACHE II e III, DRGs, AIM, SAPS e PMCs, por exemplo, entendem que a presença de intervenção cirúrgica é mais um factor explicativo da morte (Iezzoni, 1997b).

Atendendo a esta contradição, bem como ao facto de a presença ou ausência de intervenção cirúrgica constituir um dos factores que potencialmente podem explicar níveis distintos de utilização de recursos (Hornbrook, 1982 e 1985; Hornbrook e Monheit, 1985; McMahan e Billi, 1988; Ahicart, 1988; Charbonneau *et al.*, 1988; Rosko, 1988), a presença/ausência de tratamento cirúrgico poderá constituir um elemento importante para a criação de modelos de ajustamento pelo risco.

Os principais problemas referentes à explicitação final do sistema de classificação dizem respeito a dois aspectos (Knaus *et al.*, 1986; Ruttiman, 1994; Iezzoni, 1997c):

- Ao tipo de escala apresentado pelo sistema de classificação;
- À formulação final do sistema de medição.

Existem quatro tipos de escalas, de níveis hierarquizados (Melo, 1985): nominais, ordinais, de intervalos e de razões.

Segundo o mesmo autor (Melo, 1985):

- As escalas nominais são constituídas por um conjunto de classes de equivalência, sendo esta esta-

belecida por um critério empírico ligado à realidade a «medir»;

- As escalas ordinais estabelecem uma relação de maior ou menor em relação às categorias adjacentes, não permitindo relacionar aritmeticamente as diferenças entre as diversas categorias;
- As escalas de intervalos são aquelas em que é definida uma igualdade de intervalos;
- As escalas de razão permitem a comparação não só da amplitude de intervalos, mas também da respectiva razão.

Esta distinção não é meramente estatística, visto que para efeitos de gestão ou de financiamento, por exemplo, interessa não só identificar os produtos (definir categorias), mas também comparar esses mesmos produtos/categorias.

Com escalas nominais ou ordinais, somente a primeira finalidade (identificação de produtos) poderá ser atingida, necessitando-se no mínimo de uma escala de intervalos para se proceder às posteriores comparações e avaliações.

Esta problemática tem sido evidenciada em quase todos os sistemas desenvolvidos até ao momento.

Na realidade, os sistemas DRGs e PMCs apresentam escalas do tipo nominal, isto é, identificam somente categorias. Os sistemas AIM, Disease Staging, CSI e MedisGroups apresentam escalas do tipo ordinal. Os sistemas APACHE, MPM, SAPS, apresentam escalas de intervalos ou de razão. Por outro lado, algumas versões do Disease Staging, dos PMCs ou dos MedisGroups, por exemplo, apresentam escalas de intervalos. O mesmo acontece quando os DRGs são aplicados para efeitos de financiamento, em que a atribuição de pesos relativos para este efeito confere à escala potencialidades do tipo de escala de razão (Iezzoni, 1997b).

Do atrás referido resulta que se verificam preocupações crescentes para a adopção de sistemas de classificação que apresentem, no mínimo, escalas de intervalos.

Contudo, atendendo a que a maior compreensão e aceitabilidade por parte dos médicos constitui igualmente um aspecto da maior importância, será natural que noutra fase se possam igualmente apurar escalas distintas para cada variável considerada no modelo final (Lemeshow *et al.*, 1993; Ruttiman, 1994).

Mas será que estas questões conceptuais e operacionais sobre a definição de sistemas de classificação de doentes para ajustamento pelo risco têm implicações na avaliação da efectividade e da eficiência dos hospitais?

A resposta a esta questão não é pacífica, visto que, como anteriormente foi referido, existem autores que defendem a utilização de modelos baseados nos resu-

mos de alta, enquanto outros defendem a supremacia dos modelos baseados nos dados dos processos clínicos, para finalmente outros argumentarem que a utilização de uns ou outros é indiferente, pelo que o processo de decisão sobre a escolha de determinado sistema de classificação de doentes deve obedecer a critérios mais pragmáticos, como a sua exequibilidade e os respectivos custos de implementação e de exploração.

Na realidade, Krakauer e outros (1992), embora começando por afirmar que a avaliação dos resultados do tratamento dos hospitais é fortemente dependente da «doença», do doente e da sua gravidade e que a validade da inferência sobre o desempenho dos hospitais através da mortalidade ajustada pelo risco está fortemente dependente da adequação do sistema de informação e do sistema utilizado para ajustamento pelo risco, acabam por concluir que existe uma concordância significativa entre a mortalidade prevista tanto quanto se utilizam dados dos resumos de alta como os respeitantes aos processos clínicos. Isto porque nos estudos que efectuaram constataram que os modelos de ajustamento pelo risco apresentam valores moderados para a sensibilidade (81%) e para a especificidade (79%). Por outro lado, o seu valor preditivo negativo é alto (90%), pelo que concluem que, se um hospital não é *outlier* no modelo administrativo, também é provável que o não seja no modelo clínico. Contudo, o valor preditivo positivo é baixo (64%), pelo que afirmam que, se um hospital é *outlier* no modelo administrativo, existe ainda uma probabilidade forte de o não ser no modelo clínico (36%) (para melhor perceber os conceitos estatísticos referidos, v. Ruttimann, 1994, e Ash e Shwartz, 1997). No entanto, atendendo a que os resultados de investigação parecem demonstrar que a classificação dos hospitais é muito mais estável quando se utilizam modelos baseados nos dados administrativos, os autores acabam por concluir que estes modelos de ajustamento pelo risco contêm informação suficiente para se avaliarem as variações nos resultados de saúde e, conseqüentemente, no desempenho dos hospitais.

Na mesma linha de pensamento encontram-se os argumentos de DesHarnais e colegas (2000), os quais, para além de referirem as vantagens económicas da utilização de modelos de ajustamento pelo risco baseados nos resumos de alta, referem ainda que os resultados de aplicação são, no mínimo, similares aos disponibilizados pelos modelos que recorrem aos processos clínicos.

Numa perspectiva distinta pode referir-se o estudo de Hughes e outros (1996), no qual os autores defendem que a avaliação da precisão do modelo de ajustamento pelo risco, para além de depender do tipo de risco que está a ser avaliado, poderá ainda ser

influenciada por algumas perspectivas contraditórias. A este respeito, os autores afirmam que, embora se espere que o consumo de recursos, a duração de internamento e o risco de morte apresentem uma associação forte, a existência de um volume importante de mortes prematuras pode contrariar este comportamento esperado.

Numa atitude mais céptica encontra-se o estudo de Manheim e outros (1992), os quais defendem que a variação encontrada nas taxas de mortalidade ajustadas pelo risco é devida a diferentes políticas e práticas dos hospitais na admissão de doentes. Assim, acabam por concluir que sem a existência de dados clínicos mais detalhados e acessíveis não é possível estabelecer uma relação determinística entre o processo e a mortalidade.

Este tipo de argumento vem na sequência das conclusões retiradas por Greenfield e outros (1988) e por Green, Wintfeld e Passman (1990). No primeiro artigo, os autores concluem que as comorbilidades devem ser consideradas em qualquer metodologia de avaliação da qualidade, pelo que, se tal não for considerado, as variações nas práticas e nas políticas de admissão de cada hospital podem contribuir para interpretações erróneas sobre a qualidade hospitalar. O estudo de Green, Wintfeld e Passman (1990) refere a necessidade de a taxa de mortalidade hospitalar ser corrigida pela severidade do estado do doente presente no momento de admissão para desta forma se estabelecerem inferências válidas sobre a qualidade ou efectividade dos cuidados prestados.

Num artigo publicado por Iezzoni (1997c) é descrito um estudo no qual são comparados os desempenhos estatísticos de cinco sistemas de classificação — MedisGroups, APACHE (*score* fisiológico), Disease Staging, Patient Management Categories e All Patient Refined Diagnosis Related Groups (APRDRGs) — e são utilizados dois indicadores para avaliar o desempenho estatístico, a estatística c e o coeficiente de determinação (r^2), tendo chegado à conclusão de que as medidas baseadas nos resumos clínicos de alta (Disease Staging, Patient Management Categories e All Patient Refined Diagnosis Related Groups) apresentam, na generalidade, melhores resultados.

No mesmo artigo, e quando se comparam as taxas de mortalidade esperadas em cada modelo, é referido que os sistemas de classificação utilizados apresentam resultados diferentes, com maiores discrepâncias entre as medidas baseadas nos resumos de alta e as que utilizam os processos clínicos. Contudo, tal também está presente entre medidas baseadas somente nos resumos de alta (PMCs e APRDRGs, por exemplo) e nos processos clínicos (entre o APACHE e os MedisGroups, para as pneumonias).

Neste sentido, Iezzoni (1997c) acaba por concluir que os hospitais devem estar preocupados com as comparações entre as taxas de mortalidade observadas e as esperadas, visto que os resultados são diferentes em função do sistema de classificação de doentes utilizado.

Refere ainda que, apesar do seu poder explicativo, a utilização de sistemas de classificação de doentes baseados nos resumos informatizados de alta pode suscitar alguns problemas para se avaliar a qualidade, essencialmente porque estas medidas estão fortemente dependentes dos problemas decorrentes da CID-9-CM (McMahon e Smits, 1986), do nível de codificação (Jencks, Williams e Kay, 1988; Iezzoni e Moskowitz, 1986; Iezzoni *et al.*, 1992d; Green e Wintfeld, 1993) e dos incentivos financeiros para se proceder à codificação (Simborg, 1981).

Finalmente, afirma que, sempre que possível, deve ser feita uma distinção entre as condições presentes no momento da admissão e as adquiridas durante o internamento para que desta forma se evite o *death code creep*, como forma de otimizar os resultados de saúde de cada hospital (Iezzoni, 1997c).

Noutro estudo realizado por Iezzoni e outros (1995a), os autores pretendem comparar a qualidade dos hospitais, medida pela mortalidade, em função de diversos sistemas de classificação de doentes. Os modelos de ajustamento pelo risco utilizados foram os MedisGroups, APACHE (*score* fisiológico), Disease Staging e All Patient Refined Diagnosis Related Groups (APRDRGs). Os resultados são similares aos apresentados no estudo de 1997 (Iezzoni, 1997c), com as diferentes medidas a apresentarem valores distintos para as taxas de mortalidade observada, como traduzem os resultados do teste de Hosmer-Lemeshow, e com as medidas baseadas nos resumos de alta a apresentarem um poder explicativo mais elevado.

Para este último aspecto avançam igualmente com uma explicação. O melhor valor explicativo das medidas baseadas nos resumos de alta pode depender da política de codificação, o que pode originar que, frequentemente, os doentes possam ser considerados mais «doentes» através desta modalidade do que na realidade se observa quando se consultam os processos clínicos (Iezzoni, 1995).

Ainda a respeito da utilização de modelos de ajustamento pelo risco baseados nos dados administrativos, Iezzoni e outros (1992d) referem a importância da codificação. Neste artigo, a autora defende que o preenchimento incompleto dos diagnósticos secundários nas bases de dados administrativas pode conduzir a enviesamentos na atribuição de risco, isto também porque muitas bases de dados somente permitem a consideração de cinco diagnósticos.

O estudo pretende avaliar o impacto de se ter mais diagnósticos nas bases de dados, apresentando os seguintes principais resultados e conclusões: alguns problemas de saúde, como a diabetes, enfarte do miocárdio antigo e angina, os quais seria de esperar que aumentassem o risco de morte, estão associados a riscos de morte baixos; o enviesamento ou diferentes práticas de codificação nas bases de dados administrativas podem contribuir para estes resultados, pelo que a intenção de se melhorar o registo dos diagnósticos secundários baseada somente no aumento dos campos preenchidos pode ser insuficiente.

Tendo em atenção estes aspectos, Iezzoni (1997b) refere alguns requisitos para os resumos informatizados de alta no futuro, nomeadamente os respeitantes aos elementos/variáveis que devem estar presentes, às metodologias e procedimentos de codificação, à necessidade de organizações independentes avaliar todo este processo e ainda sobre o impacto previsional da alteração das modalidades de pagamento dos hospitais.

Assim, a escolha de um modelo de ajustamento pelo risco deve ter em conta os aspectos associados (Iezzoni, 1990 e 1997e; Griffith, Alexander e Jelinek, 2002):

- À significância clínica do sistema de classificação de doentes;
- À sua adaptabilidade às características de cada hospital;
- À credibilidade por parte dos médicos em relação ao seu desempenho para avaliar a qualidade dos cuidados prestados;
- À acessibilidade de dados em função do sistema de informação do hospital;
- À fiabilidade do sistema;
- Às exigências em termos da utilização de recursos por parte do hospital;
- Aos custos de exploração do sistema;
- À motivação e ao momento de escolha do sistema de classificação de doentes.

Neste sentido, interessa agora apresentar e discutir quais as possibilidades que neste momento se colocam em Portugal para a utilização e aplicação de sistemas de ajustamento pelo risco.

Existe um intenso debate em termos internacionais sobre a escolha dos melhores sistemas de ajustamento de risco, sendo de evidenciar os seguintes aspectos:

- Quais os melhores modelos? Os baseados em dados administrativos? Ou os baseados em modelos clínicos?
- Deve escolher-se somente um modelo de ajustamento de risco para todas as finalidades? Ou deve

escolher-se o melhor modelo para a respectiva finalidade?

Em relação ao primeiro aspecto, a dicotomia entre modelos administrativos e modelos clínicos, é conveniente apresentar previamente a sua caracterização. Modelos administrativos são, em termos internacionais, aqueles que utilizam os dados constantes dos resumos de alta. Na generalidade, este tipo de suporte de dados contém elementos respeitantes à identificação do doente, à natureza da admissão, às transferências internas, ao destino após a alta, ao diagnóstico de admissão, à morfologia tumoral, ao peso à nascença, aos diagnósticos (até 20 e com indicação do diagnóstico principal), às causas externas de lesão/efeito adverso (até 20), aos procedimentos (até 20), à data da primeira intervenção cirúrgica, ao número de dias em unidades de cuidados intensivos e à identificação do médico responsável pelo tratamento e do médico codificador.

Os modelos clínicos são aqueles que incluem os dados dos modelos administrativos, acrescidos de alguns elementos constantes dos processos clínicos, essencialmente aqueles que permitem caracterizar a história e o exame objectivo do doente, os resultados dos meios complementares de diagnóstico, tudo isto para se estabelecer um diagnóstico, estimar um prognóstico e prescrever o tratamento adequado.

Como foi referido, o balanço que deve presidir a esta discussão deve ter em conta aspectos relacionados com a precisão e a validade do sistema, com os custos de implementação e de exploração destes sistemas e com o seu nível de operacionalização.

Enquanto os aspectos associados à precisão e à validade dos sistemas de ajustamento pelo risco apontam para uma ligeira supremacia dos modelos clínicos, as questões relacionadas com os custos do sistema e a sua operacionalização referem que os modelos administrativos devem ser privilegiados.

Para Portugal, independentemente do valor intrínseco da precisão e da validade dos modelos, é praticamente impossível aplicar, a curto prazo, os modelos clínicos, pelo que se sugere a utilização exclusiva de modelos administrativos.

Esta opção não deve impedir que a médio prazo se altere o sistema de informação dos hospitais, de forma a considerar a eventualidade de se utilizarem os modelos clínicos.

No que se refere ao segundo aspecto, escolha de um ou mais modelos para ajustamento pelo risco em função da finalidade a que estes se destinam, a resposta pode ser mais flexível.

Assim, atendendo somente a este princípio, deveria recomendar-se a utilização de diferentes modelos de ajustamento pelo risco em função da finalidade a que

o sistema se destina. O principal obstáculo consistirá então nos respectivos custos de implementação e de exploração dos referidos sistemas.

A este respeito, a situação é mais pacífica. De facto, tendo em conta que se vão utilizar modelos administrativos, para os quais já existe actualmente colheita de dados e rotinas (que já envolvem, portanto, despesas), a utilização de mais do que um modelo representa apenas custos marginais insignificantes, pouco mais afinal do que aqueles decorrentes da aquisição de licenças para a utilização de programas informáticos.

Ou, ainda numa perspectiva mais positiva, a utilização de pelo menos dois sistemas permite rentabilizar o investimento feito na recolha de dados, visto que permite a introdução de metodologias mais correctas e amplas para se avaliar o desempenho dos hospitais. Neste sentido, é importante identificar e caracterizar os principais sistemas de ajustamento pelo risco existentes para finalmente se formularem recomendações sobre quais devem ser aplicados em Portugal (v. *Quadro I*).

Em Portugal, enquanto não for alterada a folha de admissão e alta e, concomitantemente, o resumo informatizado de alta, forma de recolha de dados que configura o que internacionalmente é conhecido por «dados administrativos», torna-se impossível utilizar de forma sistemática os sistemas de severidade que necessitem de dados clínicos, ou seja, elementos referentes a sintomas, sinais vitais e de disfunção fisiológica, os quais somente podem ser obtidos com consulta retrospectiva e exaustiva aos processos clínicos.

Por esta razão, sistemas como os MedisGroups, APACHE e Computerized Severity of Illness não podem ser utilizados.

Dos restantes, embora por razões diferentes, não devem ser utilizados os sistemas Patient Management Categories e o Índice de Comorbilidade:

- Os Patient Management Categories, porque, atendendo às suas características específicas, onde avulta a criação de um valor escalar que traduz o consumo de recursos necessários para cada categoria de doença, necessitam de uma adaptação específica ao estado da arte e à cultura portuguesa para o tratamento dos doentes;
- O Índice de Comorbilidade, porque considera a mortalidade após um ano de contacto com os hospitais, o que, para além de suscitar alguns problemas técnicos (essencialmente a causalidade entre esta mortalidade e o tratamento hospitalar), é praticamente impossível de utilizar em Portugal, visto que as estatísti-

cas de rotina não disponibilizam estes indicadores.

Assim, restam os sistemas de classificação Disease Staging e as versões Severity (SDRGs), Refined (RDRGs), All Patient Refined (APRDRGs) e International Refined dos Diagnosis Related Groups (IRDRGs).

Tendo em atenção as características destes sistemas de classificação de doentes, os quais, para além de recorrerem a metodologias distintas para a identificação da severidade e de apresentarem produtos e resultados igualmente diferentes, apresentam, no entanto, uma característica comum — utilizam os mesmos dados para identificarem produtos e estabelecerem medidas de ajustamento pelo risco —, será interessante, no futuro, considerar a utilização con-

junta de uma das versões dos DRGs (IRDRGs) e o Disease Staging para se avaliar o desempenho dos hospitais portugueses.

Os Diagnosis Related Groups começaram a ser concebidos em finais da década de 60 na Universidade de Yale com a principal finalidade de associarem o tipo de doentes tratados (*casemix*) com os respectivos custos (Fetter *et al.*, 1980).

A descrição genérica dos DRGs, bem como as suas principais vantagens e inconvenientes, encontram-se bem identificados noutros artigos (Fetter *et al.*, 1980; Hornbrook, 1982; Aronow, 1988; Averill *et al.*, 1988; Urbano e Bentes, 1990; Casas, 1991; Costa e Nogueira, 1994; Costa, 1994; Bentes *et al.*, 1996; Mullin, Averill e Boucher, 2002; Rodrigues, 2002), pelo que neste momento somente se evidencia que o desenvolvimento deste sistema de classificação de

Quadro I
Principais sistemas de ajustamento pelo risco

Sistema de severidade	Dados necessários	Definição de severidade	Forma de classificação
MedisGroups (versão original)	Dados administrativos e processo clínico	Instabilidade clínica indicada pela mortalidade	Score de admissão, entre 0 e 4
MedisGroups (versão empírica)	Dados administrativos e processo clínico	Mortalidade hospitalar	Probabilidade de morte, entre 0 e 1
APACHE II (<i>score</i> fisiológico)	Dados administrativos e processo clínico	Mortalidade hospitalar para doentes das UCIs	Score entre 0 e 60
APACHE III (<i>score</i> fisiológico)	Dados administrativos e processo clínico	Mortalidade hospitalar para doentes das UCIs	Score entre 0 e 252
Computerized Severity of Illness	Dados administrativos e processo clínico	Dificuldades no tratamento encontradas pelos médicos	Score entre 1 e 4 por doente e por cada código da CID-9-MC
Disease Staging (versão original)	Dados administrativos	Estadio da doença baseado no risco de morte ou de incapacidade funcional	Três estadios, 1, 2 e 3, com subestadios dentro de cada estadio
Disease Staging (probabilidade de morte)	Dados administrativos	Probabilidade de morte hospitalar	Probabilidade de morte, entre 0 e 1
Patient Management Categories	Dados administrativos	Morbilidade e mortalidade hospitalar	Score entre 1 e 7
Índice de Comorbilidade (Índice de Charlson)	Dados administrativos	Risco de morte dentro de um ano após a hospitalização	Valor único derivado de uma escala aditiva, representando o número e a severidade das comorbilidades
Refined Diagnosis Related Groups	Dados administrativos	Total da facturação hospitalar	Quatro escalões de severidade dentro de cada DRG adjacente
All Patient Refined Diagnosis Related Groups	Dados administrativos	Total da duração de internamento. Total da facturação hospitalar	Três classes de severidade dentro dos DRGs médicos; as mortes precoces são consideradas no grupo de severidade mais baixo

Fonte: Iezzoni (1997b).

doentes obedeceu a uma série de requisitos impostos pelos seus autores (Fetter *et al.*, 1980; Costa, 1994):

- O sistema de classificação teria de ser clinicamente interpretável com grupos de diagnóstico homogêneos. Ou seja, os médicos deveriam estar aptos a relacionar os doentes de cada grupo com um determinado padrão de tratamento;
- A classificação deveria ser obtida mediante o recurso a informação generalizada e disponível nos hospitais;
- Os grupos constituídos pelo sistema de classificação deveriam ter um número finito, preferencialmente na ordem das centenas, e ser exaustivos e mutuamente exclusivos;
- Cada grupo deveria incluir doentes com um consumo previsível de recursos similar;
- A definição de grupos deveria ser comparável entre os diversos sistemas de codificação.

Por sua vez, o Disease Staging é um sistema de classificação de doentes que utiliza os elementos constantes do resumo de alta (dados administrativos), designadamente o diagnóstico principal e os diagnósticos secundários, para produzir grupos de doentes que requerem tratamentos idênticos e com idênticos resultados esperados (Garg *et al.*, 1978; Gonnella, Hornbrook e Louis, 1984; Gonnella, Louis e Gozum, 1999).

A descrição genérica, bem como as suas principais vantagens e inconvenientes, encontram-se bem identificados noutros artigos (Garg *et al.*, 1978; Hornbrook, 1982; Gonnella, Hornbrook e Louis, 1984; Thomas, Ashcraft e Zimmerman, 1986; Aronow, 1988; Costa, 1991; Gonnella, Louis e Gozum, 1999; Gonnella e Louis, 2006), pelo que neste momento somente se destaca a existência de uma versão automatizada do Coded Staging Criteria desenvolvida a partir dos dados incluídos nos resumos de alta de cada hospital (Gonnella, Louis e Gozum, 1999; MEDSTAT, 2001).

Esta versão automatizada (MEDSTAT, 2001) permite identificar os estadios e subestadios referidos na versão clínica e original, apresenta previsões, numa escala de intervalos, para a duração de internamento, para os custos, para a mortalidade, para as complicações e para as readmissões.

O *software* do Disease Staging, atentos os problemas da escala ordinal e de comparação entre doenças referidos para a versão clínica, produz, na generalidade três tipos de indicadores: indicadores de recursos, indicadores de demora média e indicadores de mortalidade (MEDSTAT, 2001). Estes indicadores, para além de obviarem aos problemas de comparação presentes na versão clínica do Disease Staging, per-

mitiram o aparecimento de metodologias para se avaliar a efectividade e a eficiência das organizações de saúde.

Finalmente, deve ser referido um aspecto ainda não debatido neste artigo: as características dos hospitais devem ou não ser consideradas em modelos para ajustamento pelo risco?

A grande maioria dos autores defende a sua exclusão (Fetter *et al.*, 1980; Young, Swinkola e Zorn, 1982; Knaus *et al.*, 1986; Horn, 1988; Gonnella *et al.*, 1990; Le Gall, Lemeshow e Saulnier, 1993; Lemeshow *et al.*, 1993; Brewster *et al.*, 1995).

Os principais argumentos para esta exclusão encontram-se bem sintetizados no artigo de DesHarnais e outros (1997). Neste artigo defende-se que as características dos hospitais (o tipo de propriedade, a dimensão, a função de ensino e a existência de programas de formação profissional, bem como a localização, urbana ou rural, e a taxa de ocupação dos hospitais) não devem ser utilizadas no ajustamento pelo risco, essencialmente porque estas dimensões não qualificam o risco clínico de cada doente. Antes pelo contrário, estas características representam o risco institucional associado ao facto de um doente ser admitido num determinado hospital. Enfatizam ainda a importância de se excluirmos as características dos hospitais simplesmente pelo facto de poderem ser atribuídos níveis de risco diferentes a doentes com situações clínicas idênticas, em função do tipo de hospital onde este recebeu tratamento.

7. Propriedades dos modelos de ajustamento pelo risco

Na sua grande maioria, os modelos de ajustamento pelo risco têm recorrido a técnicas de regressão logística multivariada para se estabelecerem previsões sobre o resultado em análise (Iezzoni, 1997b).

Em geral, os modelos de regressão constituem um método importante para a análise multivariada de dados, visto que permitem traduzir a relação existente entre uma variável dependente ou resultado e várias variáveis explicativas.

Neste particular, os modelos de regressão logística representam um método comum para a análise de relações em que a variável dependente não é contínua (Hosmer e Lemeshow, 1989).

Esta é a situação mais frequente quando se pretende analisar a mortalidade hospitalar, em que a variável dependente é dicotómica, assumindo os valores 0 (alta vivo) e 1 (alta falecido) em função de um conjunto de variáveis independentes, frequentemente designadas por covariáveis (Hosmer e Lemeshow, 1989; Ruttiman, 1994).

As características técnicas, bem como os procedimentos necessários para se realizarem regressões logísticas, encontram-se bem documentados em diversos estudos e publicações (Greenland, 1984; McNeil e Hanley, 1984; Hosmer e Lemeshow, 1989; Hosmer, Jovanovic e Lemeshow, 1989; Pryor e Lee, 1991; Ruttimann, 1994; Smith *et al.*, 1996), pelo que somente se evidenciarão os aspectos relacionados com o ajustamento do modelo.

Segundo Ruttiman (1994), a regressão logística apresenta as seguintes vantagens: (1) não é necessário introduzir muitas restrições sobre a distribuição das covariáveis; (2) o modelo logístico disponibiliza estimativas directas sobre a probabilidade de ocorrência do *outcome*.

Para tal, e após a estimação dos coeficientes de regressão logística para as covariáveis com significância estatística, deve ter-se em atenção o seguinte (Hosmer e Lemeshow, 1989):

- Pode estimar-se o respectivo *odds-ratio* para cada uma das covariáveis;
- Deve proceder-se à avaliação do ajustamento do modelo, o que deverá somente ocorrer quando existir a convicção de que o modelo final contém todas as covariáveis, interacções incluídas, pertinentes para a previsão da variável dependente.

O ajustamento do modelo deve ser avaliado em função de dois aspectos:

1. A calibração;
2. A discriminação.

Na calibração analisam-se os desvios entre a mortalidade observada e a mortalidade prevista pelo modelo; por outras palavras, comparam-se os valores médios entre a mortalidade observada e os valores médios para a probabilidade prevista pelo modelo (Ash e Shwartz, 1997).

Neste sentido, a afirmação de que um modelo está bem calibrado baseia-se no pressuposto de que todos os parâmetros incluídos no modelo derivam do respectivo ajustamento do modelo aos dados. Quando se utiliza o método dos mínimos quadrados, é natural que o erro de calibração seja mínimo. Contudo, quando se utiliza outra abordagem, como, por exemplo, a da máxima verosimilhança, é natural que existam diferenças entre os valores médios observados e os valores médios previstos (Ash e Shwartz, 1997). A discriminação analisa o facto de o sistema prever taxas de mortalidade mais elevadas para os doentes que efectivamente morrem, comparativamente com os doentes que efectivamente não morrem (Ash e Shwartz, 1997).

Existe um intenso debate sobre qual a característica mais importante para determinar o nível de ajustamento de um modelo.

Enquanto Lemeshow e Hosmer (1982) defendem que, se um modelo não está bem calibrado, é inútil avaliar a sua discriminação, Harrel e outros (1984) referem que a discriminação é a característica mais importante de um modelo, visto que a calibração pode ser atingida posteriormente com alguns ajustamentos subjectivos, derivados da integração no modelo das opiniões dos peritos.

Isto porque, se um modelo tem boa discriminação, a calibração pode ser alcançada sem prejuízo daquele atributo, enquanto a ausência de poder de discriminação nunca pode ser corrigida com melhorias na sua calibração (Harrel *et al.*, 1984).

Por outro lado, segundo Ash e Shwartz (1997), quando se pretende diferenciar somente o risco de morte, ou seja, a distinção entre mortos e sobreviventes, a calibração não é um aspecto importante. O mesmo já não deve ser afirmado quando se pretendem comparar os valores da mortalidade prevista com a observada, essencialmente para efeitos de garantia de qualidade com o estabelecimento de «casos-sentinelas», em que a calibração assume um carácter muito importante.

8. Validação dos modelos de ajustamento pelo risco

Existem diversos problemas quando se pretendem aplicar sistemas de ajustamento de risco a populações diferentes ou a subgrupos distintos das populações donde o sistema foi originalmente construído (Iezzoni, 1999).

Num estudo realizado por Justice, Covinsky e Berlin (1999) são referidos alguns aspectos conceptuais e metodológicos respeitantes à utilização de modelos de ajustamento pelo risco (v. *Quadro II*).

Atendendo a que existem problemas na aplicação indiscriminada de modelos a diferentes populações, os autores descrevem uma abordagem baseada na precisão dos modelos (calibração e discriminação) e na generalização dos modelos (reprodutibilidade e transportabilidade).

Como subsistem alguns problemas semânticos, os autores avançam com definições de precisão e de generalização (v. *Quadro III*), as quais são consideradas úteis para a validação de modelos de ajustamento pelo risco a populações distintas (Justice, Covinsky e Berlin, 1999).

Os autores (Justice, Covinsky e Berlin, 1999) propõem ainda uma hierarquia para a validação externa de sistemas de ajustamento pelo risco (v. *Quadro III*).

Num estudo realizado por Flanders e outros (1999) pretende-se avaliar a discriminação e a calibração do *Pneumonia Severity-of-Illness* (PSI), uma medida de prognóstico específica por doença, quando aplicado aos doentes com *Community-acquired pneumonia* (CAP).

Os principais resultados referem que o PSI apresenta uma boa discriminação entre doentes com alto risco de morte e aqueles que apresentam baixo risco de morte.

Contudo, a calibração do modelo foi fraca, tendo o PSI previsto 2,4 vezes mais mortes do que aquelas que ocorreram nos doentes com CAP.

Para tal são avançadas algumas explicações, dentro das quais avulta a questão de o modelo não apresentar boa calibração porque existem alguns factores que não estão a ser devidamente valorizados e medidos.

Estes factores poderão ser as diferenças organizacionais e regionais nos critérios de admissão de doentes, as diferenças no tratamento e na disponibilidade tecnológica por parte dos hospitais e as diferenças nos próprios factores de prognóstico.

Neste sentido, os autores apresentam uma metodologia para a recalibração do modelo, a qual se baseou na regressão logística, de acordo com o recomendado nos estudos de Iezzoni (1993 e 1996b), referindo ainda que a recalibração é necessária sempre que se aplicam resultados a uma população diferente.

Finalmente, deve ser referido um artigo de Iezzoni (1999) no qual a autora, baseando-se nos resultados de Flanders e outros (1999) e de Justice, Covinsky e Berlin (1999), refere que, sempre que se transportam resultados de medidas estatísticas para outras populações tanto em termos espaciais como temporais, os

Quadro II Aspectos conceptuais e metodológicos dos modelos de ajustamento pelo risco

Termo	Definição ou critério
Precisão	O nível segundo o qual os valores previstos estão em conformidade com os observados
Calibração	A probabilidade prevista nunca é demasiado alta nem demasiado baixa
Discriminação	A ordenação dos riscos individuais está correcta
Generalização	Capacidade de o modelo de ajustamento pelo risco estabelecer previsões precisas numa população diferente
Reprodutibilidade	O sistema é preciso em doentes não incluídos no seu desenvolvimento, mas que são de uma população idêntica
Transportabilidade	O sistema é preciso em doentes de uma população diferente, embora relacionada com a original
Histórica	A precisão mantém-se em períodos de tempo diferentes
Geográfica	A precisão mantém-se quando o sistema é testado em localizações diferentes
Metodológica	A precisão mantém-se quando o sistema é testado em métodos de recolha de dados distintos
Espectro	A precisão mantém-se em populações que, em média, estão mais ou menos avançadas na progressão da doença ou com uma trajectória diferente de doença
<i>Follow-up</i>	A precisão mantém-se quando o sistema é testado em períodos curtos ou longos

Quadro III Hierarquia para a validação externa de sistemas de ajustamento pelo risco

Nível de validação	Avaliação cumulativa da generalização
0. Interna	Reprodutibilidade
1. Prospectiva	Reprodutibilidade; transportabilidade histórica
2. Independente	Reprodutibilidade; transportabilidades histórica, geográfica, metodológica e de espectro
3. Localização múltipla	Reprodutibilidade; transportabilidades histórica, geográfica, metodológica e de espectro
4. Independente e múltipla	Reprodutibilidade; transportabilidades histórica, geográfica, metodológica e de espectro
5. Independente e múltipla para tabelas de sobrevivência	Reprodutibilidade; transportabilidades histórica, geográfica, metodológica, de espectro e de <i>follow-up</i>

analistas devem observar o princípio *Caveat emptor*. Ou seja, muito dificilmente o modelo aplicado a diferentes populações apresentará resultados precisos. A autora refere ainda que estes problemas de calibração podem ser solucionados com recurso a técnicas estatísticas, como a regressão logística.

Para melhor se compreender a dimensão dos fenómenos e se definir uma metodologia para a sua solução, a autora sugere os seguintes princípios (Iezzoni, 1999):

- Identificação precisa da versão dos sistemas de classificação de doentes que está a ser utilizada no estudo;
- Sempre que for conveniente, os autores devem utilizar os mecanismos mais adequados de diagnóstico do modelo. No caso dos sistemas de ajustamento pelo risco, deve sempre avaliar-se a discriminação e a calibração do modelo;
- Quando se realizarem processos de recalibração, os autores devem referir detalhadamente todos os passos que foram utilizados na técnica.

Mesmo que se cumpram estas fases, a autora chama a atenção para o facto de que a recalibração é sempre um fenómeno único, visto que o novo modelo somente está ajustado aos dados e à população que serviu de base para o estudo.

Assim, pode concluir-se que, embora exista uma enorme variedade de modelos para a avaliação do desempenho, os quais configuram diversas dimensões e perspectivas, existem igualmente alguns elementos comuns, designadamente a necessidade de combinar as características dos doentes com diversos procedimentos estatísticos, tendo em vista o ajustamento pelo risco.

Por sua vez, o ajustamento pelo risco é considerado um procedimento inevitável e incontornável de uma abordagem correcta para avaliar o desempenho hospitalar.

9. Conclusões

As organizações de saúde, em geral, e os hospitais, em particular, são frequentemente reconhecidos por terem particularidades e especificidades que conferem uma especial complexidade ao seu processo produtivo e à sua gestão (Jacobs, 1974; Butler, 1995). Neste sentido, na literatura hospitalar emergem alguns temas como prioritários tanto na investigação como na avaliação do seu funcionamento, nomeadamente os relacionados com a produção, com o financiamento, com a qualidade, com a eficiência e com a avaliação do seu desempenho.

A medição da produção hospitalar assume especial importância tanto por questões estruturais, associadas à natureza intrinsecamente social e delicada do seu produto, como por questões operacionais, que passam desde o simples conhecimento e comparação da sua actividade a questões mais técnicas, como a avaliação dos cuidados prestados.

O conhecimento e eventual publicitação do desempenho dos hospitais é importante não só para se poderem comparar estas organizações de saúde, mas igualmente para se poder reduzir a actual assimetria de informação (Griffith, Alexander e Jelinek, 2002). O estado da arte da avaliação do desempenho das organizações de saúde parece seguir a trilogia definida por Donabedian (1985) — Estrutura, Processo e Resultados. Existem diversas perspectivas para a avaliação do desempenho na óptica dos Resultados — efectividade, eficiência ou desempenho financeiro. No entanto, qualquer que seja a utilizada, o ajustamento pelo risco é necessário para se avaliar a actividade das organizações de saúde, como forma de medir as características dos doentes que podem influenciar os resultados de saúde.

Em termos genéricos, a principal finalidade do ajustamento pelo risco é a de controlar os factores que os doentes apresentam ao contactarem uma determinada organização de saúde que podem afectar a sua probabilidade de obterem um bom ou um mau resultado (Iezzoni, 1996a). Contudo, algumas questões permanecem ainda em grande debate. A primeira questão respeita ao risco de quê, com respostas que vão desde o consumo de recursos à probabilidade de morte iminente, à existência de complicações ou ainda ao desempenho funcional dos doentes e à satisfação do cliente. Este aspecto conduz a duas situações. A primeira é que não existe uma única abordagem de ajustamento pelo risco que apresente o mesmo nível de significância para todos os diferentes resultados de saúde. A segunda é que existe uma grande variedade de abordagens que pretendem estabelecer medidas de ajustamento pelo risco.

Por outro lado, o ajustamento pelo risco tem sido frequentemente associado à questão da severidade do estado do doente. Este assunto é ainda mais importante quando se tem presente que não existe uma definição única sobre severidade/gravidade. Assim, não é de estranhar que a grande discussão que ainda existe no panorama internacional se situe sobre a predominância de uma de duas perspectivas, com sistemas que privilegiam a utilização de recursos, outros a severidade, outros ambos.

Como possíveis indicadores de resultados, existem a mortalidade (resultados finais), as complicações e as readmissões (resultados intermédios). Com excepção dos estudos realizados por Thomas (1996) e Thomas

e Hofer (1998 e 1999), praticamente ninguém contesta a relação entre estes indicadores e a efectividade dos cuidados. Chamando, no entanto, a atenção para a necessidade de se definirem modelos de ajustamento pelo risco e ainda para algumas dificuldades conceptuais e operacionais para se atingir este objectivo. Situam-se aqui aspectos como a escolha do tipo de dados que deve ser utilizados (clínicos ou administrativos) e do sistema de classificação de doentes, sobre a janela de observação da mortalidade e se a mortalidade é ou não um resultado indesejável.

Em relação à eficiência técnica dos hospitais, os indicadores tradicionalmente mais utilizados para avaliação da eficiência técnica dos hospitais são os custos médios e a demora média, admitindo-se que ambos tenham comportamento distinto, atendendo essencialmente à intensidade dos recursos. Também neste domínio, a grande maioria dos estudos aponta para que a gravidade aumenta o poder justificativo do consumo de recursos e que o ajustamento pelo risco é útil para avaliar a eficiência dos hospitais.

Em relação aos sistemas usados para medir a severidade e, consequentemente, ajustar pelo risco, o seu desenvolvimento apresenta, na generalidade, dois tipos de preocupações: a definição dos suportes de recolha da informação e a definição dos momentos de medição.

Em última instância, o dilema que se coloca reside na definição de prioridades e daquilo que se pretende sacrificar. Quando se entende que os aspectos financeiros são determinantes, então será natural que se privilegie o recurso quase exclusivo a elementos dos resumos de alta como suporte de recolha da informação. Quando se defende que a validade de construção e de conteúdo de um sistema de classificação de doentes para medição da severidade é um aspecto a preservar, então o recurso aos elementos dos processos clínicos é inevitável.

A definição dos momentos de medição dos dados tem repercussões em dois níveis de análise: na neutralidade económica do sistema e na prospectividade do sistema.

Com a neutralidade económica discute-se se o sistema de classificação deve ou não ser indiferente aos tratamentos disponíveis ou utilizados no hospital, isto é, se estes devem ser considerados mais um factor para se desagregarem ou definirem novos produtos. Contudo, este procedimento pode limitar a fiabilidade e a validade do sistema de classificação, derivado do registo de doenças iatrogénicas.

A prospectividade do sistema de classificação constitui um factor relativamente fácil de descrever: a mesma só poderá ser assegurada quando um sistema se baseia nos elementos de admissão (Green, Passman e Wintfeld, 1991). Atendendo a que no processo

de decisão clínica existe a necessidade de decidir imediatamente a seguir à admissão dos doentes, resulta natural admitir que um sistema de classificação prospectivo será aquele que sintetiza melhor o processo de produção nos hospitais.

Dos argumentos atrás referidos fácil será inferir que um sistema de classificação ideal não deve ter neutralidade económica e deve ser prospectivo.

Mas será que estas questões conceptuais e operacionais sobre a definição de sistemas de classificação de doentes para ajustamento pelo risco têm implicações na avaliação da efectividade e da eficiência dos hospitais?

A resposta a esta questão não é pacífica, visto que, como anteriormente foi referido, existem autores que defendem a utilização de modelos baseados nos resumos de alta, enquanto outros defendem a supremacia dos modelos baseados nos dados dos processos clínicos, para, finalmente, outros argumentarem que a utilização de uns ou outros é indiferente, pelo que o processo de decisão sobre a escolha de determinado sistema de classificação de doentes deve obedecer a critérios mais pragmáticos, como a sua exequibilidade e os respectivos custos de implementação e de exploração.

Em relação às possibilidades que neste momento se colocam em Portugal para a utilização e aplicação de sistemas de ajustamento pelo risco, verifica-se que, independentemente do valor intrínseco da precisão e da validade dos modelos clínicos, é praticamente impossível a curto prazo aplicar estes modelos, pelo que se sugere a utilização exclusiva de modelos administrativos. Esta opção não deve impedir que a médio prazo se altere o sistema de informação dos hospitais, de forma a considerar a eventualidade de se utilizarem os modelos clínicos.

Existem diversos problemas quando se pretendem aplicar sistemas de ajustamento de risco a populações diferentes ou a subgrupos distintos das populações donde o sistema foi originalmente construído. Ou seja, muito dificilmente o modelo aplicado a diferentes populações apresentará resultados precisos. Assim, só depois de verificado o ajustamento do modelo à população em questão, avaliado em função da sua calibração e discriminação, este pode ser aplicado.

Resumindo, embora exista uma enorme variedade de modelos para a avaliação do desempenho, os quais configuram diversas dimensões e perspectivas, existem igualmente alguns elementos comuns, designadamente a necessidade de combinar as características dos doentes com diversos procedimentos estatísticos, tendo em vista o ajustamento pelo risco.

Por sua vez, o ajustamento pelo risco é considerado um procedimento inevitável e incontornável de uma abordagem correcta para avaliar o desempenho hospitalar.

□ Bibliografia

- Australian Council on Health Care Standards — Determining the Potential to Improve Quality of Care : ACHS Indicator Results for Australia and New Zealand 1998-2002. Ultimo, New South Wales: The ACHS, Health Services Research Group, University of Newcastle, 2002.
- Ahicart, C. — Técnicas de Medicion del Case-Mix Hospitalario. *Hospital 2000*. 1988; Suplementos 1, 2, 3 e 4.
- Ament, R. P.; Dreachslin, J. L.; Kobrinsky, E. J. e Wood, W. R. — Three Case-Type Classifications: Suitability for the Use in Reimbursing Hospitals. *Medical Care*. 1982; 20 (5), 460-467.
- Applegate, W. B.; Blass, J. P. e Williams, T. F. — Instruments for the Functional Assessment of the Older Patients. *New England Journal of Medicine*. 1990; 322 (17), 1207-1214.
- Aronow, D. B. — Severity of Illness Measurement: Applications in Quality Assurance and Utilization Review. *Medical Care Review*. 1988; 45 (2), 339-366.
- Arozullah, A. M.; Henderson, W. G.; Khuri, S. F. e Daley, J. — Postoperative Mortality and Pulmonary Complication Rankings: How Well Do They Correlate at the Hospital Level? *Medical Care*. 2003; 41 (8), 979-991.
- Ash, A. A.; Shwartz M — Evaluating the Performance of Risk-Adjustment Methods: Dichotomous Variables. In Iezzoni LI (editor), «Risk Adjustment for Measuring Healthcare Outcomes». 2nd Edition. Chicago: Health Administration Press, 1997. 427-469.
- Averill, R. F.; Muldoon, J. H.; Vertrees, J. C.; Goldfield, N. I.; Mullin, R. L.; Fineran, E. C.; Zhang, M. Z.; Steinbeck B. e Grant T. — The Evolution of Casemix Measurement Using Diagnosis Related Groups (DRGs). Salt Lake City: 3M Health Information Systems, 1998. 5-98.
- Becker, R. B.; Zimmerman, J. E.; Knaus, W. A.; Wagner, D. P.; Seneff, M. G.; Draper, E. A.; Higgins, T. L.; Estafanous, F. G. e Loop, F. D. — The Use of APACHE III to Evaluate ICU Length of Stay, Resource Use, and Mortality after Coronary Artery Bypass Surgery. *Journal of Cardiovascular Surgery*. 1995; 36 (1), 1-11.
- Bennet, C. L.; Greenfield, H.; Aronow, P.; Ganz, P. e Vogelzang, N. J. — Elashoff R. M. Patterns of Care Related to Age of Men with Prostate Cancer. *Cancer*. 1991; 67 (10), 2633-2641.
- Bentes, M.; Gonçalves, M.; Tranquada, S. e Urbano J. — A Utilização de GDH's como Instrumento de Financiamento Hospitalar. *Gestão Hospitalar*. 1996; 33 (9), 33-40.
- Berki, S. E.; Ashcraft, M. L. e Newbrander, W. C. — Length-of-Stay Variations within ICDA-8 Diagnosis Related Groups. *Medical Care*. 1984; 22 (2), 126-142.
- Bindman, A. B. e Grumbach K. — America's Safety Net : The Wrong Place at the Wrong Time. *JAMA*. 1992; 268 (17), 2426-2427.
- Bleichrodt, H.; Crainich, D. e Eeckhoudt L. — The Effect of Comorbidities on Treatment Decisions. *Journal of Health Economics*. 2003; 22 (5), 805-820.
- Blumberg, M. S. — Risk Adjusting Health Care Outcomes : A Methodologic Review. *Medical Care Review*. 1986; 43 (2), 351-393.
- Blumberg, M. S. — Comments on HCFA Hospital Death Rate Statistical Outliers. *Health Services Research*. 1987; 21 (6), 715-739.
- Bradbury, R. C.; Golec, J. H. e Steen P. M. — Relating Hospital Health Outcomes and Resource Expenditures. *Inquiry*. 1994; 31 (1), 56-65.
- Brailer, D. J.; Kroch, E.; Pauly, M. V. e Huang J. — Comorbidity-Adjusted Complication Risk : A New Outcome Quality Measure. *Medical Care*. 1996; 34 (5), 490-505.
- Bravo, G.; Dubois, M. F.; Wals, P.; Herbert, R. e Messier L. — Relationship between Regulatory Status, Quality of Care, and Three-Year Mortality in Canadian Residential Care Facilities : A Longitudinal Study. *Health Services Research*. 2002; 37 (5), 1181-1196.
- Brewster, A. C.; Jacobs, C. M. e Bradbury, R. C. — Classifying Severity of Illness by Using Clinical Findings. *Health Care Financing Administration*. 1984; Annual Supplement, 107-108.
- Brewster, A. C.; Karlin, B. G.; Hyde, L. A.; Jacobs, C. M.; Bradbury, R. C. e Chae Y. M. — MEDISGROUPS: A Clinically Based Approach to Classifying Patients at Hospital Admission. *Inquiry*. 1985; 22 (4), 377-387.
- Burns, L. R.; Wholey, D. R. e Abeln, M. O. — Hospital Utilization and Mortality Levels for Patients in the Arizona Health Care Cost Containment System. *Inquiry*. 1993; 30 (2), 142-156.
- Burstin, H. R.; Lipsitz, S. R. e Brennan, T. A. — Socioeconomic Status and Risk for Substandard Medical Care. *JAMA*. 1992; 268 (17), 2383-2387.
- Butler, J. R. — Hospital Cost Analysis. Dordrecht, The Netherlands: Kluwer Academic Publishers, 1995.
- Carneiro, A. V. — Factores de Prognóstico em Medicina Intensiva : aplicação de Sistemas Gerais de Prognóstico em Três Unidades de Cuidados Intensivos de Um Hospital Universitário. Lisboa: Faculdade de Medicina de Lisboa. Universidade de Lisboa, 1994. Tese de Doutoramento.
- Casas, M. — Clasificación de Pacientes y Producción Hospitalaria : los Grupos Relacionados con el Diagnóstico (GRD). In Casas M (editor) — «Los Grupos Relacionados con el Diagnóstico (GRD): Experiencia y Perspectiva de Utilización». Barcelona: Masson, 1991. 23-43.
- Case, R. B.; Moss, A. J.; Case, N.; McDermott, M. e Eberly S. — Living Alone After Myocardial Infarction : Impact on Prognosis. *JAMA*. 1992; 267 (4), 515-519.
- Chang, L. C.; Lin, S. W.; Northcott, D. N. — The NHS Performance Assessment Framework : a «Balanced Scorecard» Approach? *Journal of Management in Medicine*. 2002; 16 (5), 345-358.
- Charbonneau, C.; Ostrowsky, C.; Phoener, E. T.; Lindsay, P.; Panniers, T. L.; Houghton, P. e Albright J. — Validity and Reliability Issues in Alternative Patient Classification Systems. *Medical Care*. 1988; 26 (8), 800-813.
- Charlson, M. E.; Pompei, P.; Ales K. L. e MacKenzie, C. R. — A New Method of Classifying Prognostic Comorbidity in Longitudinal Studies : Development and Validation. *Journal of Chronic Diseases*. 1987; 40 (5), 373-383.
- Chassin, M. R.; Park, R. E.; Lohr, K. L.; Keesy, J. e Brook, R. H. — Differences among Hospitals in Medicare Patient Mortality. *Health Services Research*. 1989; 24 (1), 1-31.
- Chen, E.; Naylor, D. — Variation in Hospital Length of Stay for Acute Myocardial Infarction in Ontario, Canada. *Medical Care*. 1994; 32 (5), 420-435.
- Clark, D. E. e Ryan, L. M. — Concurrent Prediction of Hospital Mortality and Length of Stay from Risk Factors on Admission. *Health Services Research*. 2002; 37 (3), 631-645.
- Cleary, P. D.; Greenfield, S.; Mulley, A. G.; Pauker, S. G.; Schroeder, S. A.; Wexler, L. e McNeil, B. J. — Variations in Length of Stay and Outcomes for Six Medical and Surgical Conditions in Massachusetts and California. *JAMA*. 1991; 266 (1), 73-79.

- Coffey, R. M. e Goldfarb, M. G. — DRGs and Disease Staging for Reimbursing Medicare Patients. *Medical Care*. 1986; 24 (9), 814-829.
- Costa, C.; Delgado, M. e Carvalho, R. — Relação entre as Características da Procura e a Utilização de Recursos num Hospital. *Revista Portuguesa de Saúde Pública*. 1989; 7 (4), 27-40.
- Costa, C. e Reis, V. — O Sucesso nas Organizações de Saúde. *Revista Portuguesa de Saúde Pública*. 1993; 11 (3), 59-68.
- Costa, C. e Nogueira, P. — Produção Hospitalar e Fiabilidade. *Revista Portuguesa de Saúde Pública*. 1994; 12 (2), 31-40.
- Costa, C. — A Severidade da Doença : Identificação e Caracterização de Alguns Sistemas de Classificação. *Revista Portuguesa de Saúde Pública*. 1991; 9 (1), 37-43.
- Costa, C. — Os DRGs (Diagnosis Related Groups) e a Gestão do Hospital. *Revista Portuguesa de Gestão*. 1994; III/IV, 47-65.
- Daley, J.; Jencks, S.; Draper, D.; Lenhart, G.; Thomas, N. e Walker, J. — Predicting Hospital-Associated Mortality for Medicare Patients : A Method for Patients with Stroke, Pneumonia, Acute Myocardial Infarction and Congestive Heart Failure. *JAMA*. 1988; 260 (24), 3617-3624.
- Degeling, P.; Sorensen, R.; Maxwell, S.; Aisbet, C.; Zhang, K. e Coyle, B. — The Organization of Hospital Care and its Effects. Sydney: Centre for Clinical Governance Research. University of New South Wales, 2000.
- DesHarnais, S. I.; Forthman, M. T.; Lowry, J. M. e Wooster, L. D. — Risk-Adjusted Quality Outcome Measures : Indexes for Benchmarking Rates of Mortality, Complications and Readmissions. *Quality Management in Health Care*. 1997; 5 (2), 80-87.
- DesHarnais, S. I.; Forthman, M. T.; Lowry, J. M. e Wooster, L. D. — Risk-Adjusted Clinical Quality Indicators : Indexes for Measuring and Monitoring Rates of Mortality, Complications, and Readmissions. *Quality Management in Health Care*. 2000; 9 (1), 14-22.
- Dismuke, C. E. e Sena V. — Hospital Productivity and Efficiency Measurement in the Presence of Undesirable Output. Lisboa: Associação Portuguesa de Economia da Saúde, 1998 (Documento de Trabalho; 2/98).
- Donabedian, A. — Methods for Deriving Criteria for Assessing the Quality of Medical Care. *Medical Care Review*. 1980; 37 (7), 653-698.
- Donabedian, A. — The Epidemiology of Quality. *Inquiry*. 1985; 22 (3), 282-292.
- Dubois, R. W.; Brook, R. H. e Rogers, W. H. — Adjusted Hospital Death Rates : A Potential Screen for Quality of Medical Care. *American Journal of Public Health*. 1987; 77 (9), 1162-1166.
- Dubois, R. W.; Rogers, W. H.; Moxley, J. H.; Draper, D. e Brook, R. H. — Hospital Inpatient Mortality : Is it a Predictor of Quality? *The New England Journal of Medicine*. 1987; 317 (26), 1674-1680.
- Elixhauser, A.; Steiner, C.; Harris, D. R. e Coffey, R. M. — Comorbidity Measures for Use with Administrative Data. *Medical Care*. 1998; 36 (1), 8-27.
- Epstein, A. M.; Boge, J.; Dreyer, P. e Thorpe, K. E. — Trends in Length of Stay and Rates of Readmission in Massachusetts: Implications for Monitoring Quality of Care. *Inquiry*. 1991; 28 (1), 19-28.
- Evans, J. H.; Hwang, Y. e Nagarajan, N. — Physicians' Response to Length-of-Stay Profiling. *Medical Care*. 1995; 33 (11), 1106-1119.
- Evans, R. G. — Incomplete Vertical Integration : The Distinctive Structure of Health Care Industry. In Van der Gaag J, Perlman M (editors) — «Health, Economics and Health Economics», Amsterdam: North Holland, 1981, 329-354.
- Evans, R. G. — Strained Mercy : The Economics of Canadian Health Care. Toronto: Butterworths, 1984.
- Fetter, R. B.; Shin, Y.; Freeman, J. L.; Averill, R. F. e Thompson, J. D. — Case Mix Definition by Diagnosis-Related Groups. *Medical Care*. 1980; 18, Supplement, 1-53.
- Fink, A.; Yano, E. M. e Brook, R. H. — The Condition of the Literature on Differences in Hospital Mortality. *Medical Care*. 1989; 27 (4), 315-336.
- Flanders, W. D.; Tucker, G.; Krishnadasan, A.; Martin, D.; Honig, E. e McClellan, W. M. — Validation of Pneumonia Severity Index : Importance of Study-Specific Recalibration. *Journal of General Internal Medicine*. 1999; 14 (6), 333-340.
- Forman, D. E.; Berman, A. D.; McCabe, C. H.; Baim, D. S. e Wey, J. Y. — PTCA in the Elderly : The Young-Old versus the Old-Old. *Journal of the American Geriatric Society*. 1992; 40 (1), 19-22.
- Francks, P.; Gold, M. R. e Clancy, C. M. — Use of Care and Subsequent Mortality : The Importance of Gender. *Health Services Research*. 1996; 31 (3), 347-363.
- Garg, M. L.; Louis, D. Z.; Gliebe, W. A.; Spirka, C. S.; Skipper, C. S. e Parekh, R. R. — Evaluating Inpatient Costs : The Staging Mechanism. *Medical Care*. 1978; 16 (3), 191-201.
- Garnick, D. W.; DeLong, E. R. e Luft, H. S. — Measuring Hospital Mortality Rates : Are 30-Day Data Enough? *Health Services Research*. 1995; 29 (6), 679-695.
- Gertman P. M. e Lowenstein S. — A Research Paradigm for Severity of Illness : Issues for the Diagnosis-Related Group System. *Health Care Financing Administration*. 1984; Annual Supplement, 79-90.
- Gijsen, R.; Hoeymans, N.; Schellevis, F. G.; Ruwaard, D.; Satariano, W. A. e Bos, G. A. — Causes and Consequences of Comorbidity : A Review. *Journal of Clinical Epidemiology*. 2001; 54 (7), 661-674.
- Goldfarb, M. G.; Hornbrook, M. C. e Higgins, C. S. — Determinants of Hospital Use : A Cross Diagnostic Analysis. *Medical Care*. 1983; 21 (1), 48-66.
- Goldman, L.; Caldera, D. L.; Nussbaum, S. R.; Southwick, F. S.; Krogstad, D.; Murray, B.; Burke, D. S.; O'Malley, T. A.; Goroll, A. H.; Caplan, C. H.; Nola, J.; Carabello, B. e Slater, E. E. — Multifactorial Index of Cardiac Risk in Noncardiac Surgical Procedures. *New England Journal of Medicine*. 1977; 297 (16), 845-850.
- Gonnella, J. S.; Hornbrook, M. C. e Louis, D. Z. — Staging of Disease : A Case-Mix Measurement. *JAMA*. 1984; 251 (5), 637-644.
- Gonnella, J. S.; Louis, D. Z.; Zeleznik, C. e Turner, B. J. — The Problem of Late Hospitalization : A Quality and Cost Issue. *Academic Medicine*. 1990; 65 (5), 314-319.
- Gonnella, J. S.; Louis, D. Z. e Gozum, M. E.; 1999 (editors) — Disease Staging : Clinical Criteria (Version 17). Santa Barbara, CA: MEDSTAT Group, 1999.
- Gordon, H. S. e Rosenthal, G. E. — The Relationship of Gender and In-Hospital Death : Increased Risk of Death in Men. *Medical Care*. 1999; 37 (3), 318-324.
- Green, J.; Wintfeld, N. e Passman, L. J. — The Importance of Severity of Illness in Assessing Hospital Mortality. *JAMA*. 1990; 263 (2), 241-246.

- Green, J.; Passman, L. J. e Wintfeld, N. — Analyzing Hospital Mortality : The Consequences of Diversity in Patient Mix. *JAMA*. 1991; 265 (14), 1849-1853.
- Greenfield, S.; Blanco, D. M.; Elashoff, R. M. e Ganz, P. A. — Patterns of Care Related to Age of Breast Cancer Patients. *JAMA*. 1987; 257 (20), 2766-2770.
- Greenfield, S.; Aronow, H. U.; Elashoff, R. M. e Watanabe, D. — Flaws in Mortality Data : The Hazards of Ignoring Comorbid Disease. *JAMA*. 1988; 260 (15), 2253-2255.
- Greenland, S. — Bias Methods for Deriving Standardized Morbidity Ratio and Attributable Fraction Estimates. *Statistics in Medicine*. 1984; 3, 131-141.
- Griffith, J. R.; Alexander, J. A. e Jelinek, R. C. — Measuring Comparative Hospital Performance. *Journal of Healthcare Management*. 2000; 47 (1), 41-56.
- Gustafson, D. H.; Fryback, D. G.; Rose, J. H.; Yick, V.; Prokop, C. T.; Detmer, D. E. e Moore, J. — A Decision Theoretic Methodology for Severity Index Development. *Medical Decision Making*. 1986; 6 (1), 27-35.
- Harrel, F. A.; Lee, K. L.; Califf, R. M.; Pryor, D. B. e Rosati, R. A. — Regression Modelling Strategies for Improved Prognostic Prediction. *Statistics in Medicine*. 1984; 3, 143-152.
- HCIA — One Hundred Top Hospitals : Benchmarks for Success. Evanston, Illinois: HCIA. L.L.C.; 1999.
- Hill, C. A.; Winfrey, K. L. e Rudolph, B. A. — «Best Hospitals» : A Description of the Methodology for the Index of Hospital Quality. *Inquiry*. 1997; 34 (1), 80-90.
- Horn, S. D.; Sharkey, P. D.; Bucle, J. M.; Backofen, J. E.; Averill, R. F. e Horn R. A. — The Relationship Between Severity of Illness and Hospital Length of Stay and Mortality. *Medical Care*. 1991; 29 (4), 305-317.
- Horn, S. D. — Measuring Severity : How Sick Is Sick? How Well Is Well? *Healthcare Financing Management*. 1986; 40 (10) 21, 24-32.
- Horn, S. D. — Severity of Illness and the Adverse Patient Occurrence Index : A Reliability Study and Policy Implications. *Medical Care*. 1988; 26 (7), 736-738.
- Horn, S. D. — Physician Profiling : How Can it Be Misleading and What to Do. *Consultant*. 1997; 27 (12), 86-88; 93-94.
- Hornbrook, M. C. e Monheit, A. C. — The Contribution of Case-Mix Severity to the Hospital Cost-Output Relation. *Inquiry*. 1985; 22 (3), 259-271.
- Hornbrook, M. C. — Hospital Case Mix : Its Definition Measurement and Use : Part I. The Conceptual Framework. Part II. Review of Alternative Measures. *Medical Care Review*. 1982; 39 (1), 1-43. *Medical Care Review*. 1982; 39 (2), 73-123.
- Hornbrook, M. C. — Techniques for Assessing Hospital Case Mix. *Annual Review of Public Health*. 1985; 6, 295-324.
- Hosmer, D. W. e Jovanovic, B.; Lemeshow S. — Best Subsets Logistic Regression. *Biometrics*. 1989; 45, 1265-1270.
- Hosmer, D. W. e Lemeshow, S. — Applied Logistic Regression. New York: John Wiley & Sons, 1989.
- Houchens, R. L. — A General Method for Measuring Inpatient Comorbidity Burden on Inpatient Mortality from a Large Administrative Database. Santa Barbara, CA: MEDSTAT, 2002.
- Hughes, J. S.; Iezzoni, L. I.; Daley, J. e Greenberg, L. — How Severity Measures Rate Hospitalized Patients. *Journal of General Internal Medicine*. 1996; 11 (5), 303-311.
- Ibrahim, J. E.; Majoor, J. W.; Boyce, M. W. e McNeil, J. J. — Pilot Hospital-Wide Clinical Indicators Project : Final Report. Canberra: Commonwealth of Australia, 1998.
- Iezzoni, L. I. e Moskowitz, M. A. — Clinical Overlap among Medical Diagnosis-Related Groups. *JAMA*. 1986; 255 (7), 927-929.
- Iezzoni, L. I., Ash, A. A.; Cobb, J. L. e Moskowitz, M. A. — Admission MedisGroups Score and the Cost of Hospitalization. *Medical Care*. 1988; 26 (11), 1068-1080.
- Iezzoni, L. I.; Ash, A. A.; Coffman, G. e Moskowitz, M. A. — Admission and Mid-Stay MedisGroups Scores as Predictors of Death within 30 Days of Hospital Admission. *American Journal of Public Health*. 1991a; 81 (1), 74-78.
- Iezzoni, L. I.; Henderson, M. G.; Bergman, A. e Drews, R. E. — Purpose of Admission and Resource Use During Cancer Hospitalizations. *Health Care Financing Review*. 1991b; 13 (2), 29-40.
- Iezzoni, L. I. e Daley, J. — A Description and Clinical Assessment of the Computerized Severity Index. *Quality Review Bulletin*. 1992; February, 44-52.
- Iezzoni, L. I.; Ash, A. S.; Coffman, G. A. e Moskowitz, M. A. — Predicting In-Hospital Mortality : A Comparison of Severity Measurement Approaches. *Medical Care*. 1992a; 30 (4), 347-359.
- Iezzoni, L. I.; Restuccia, J. D.; Shwartz, M.; Schaumburg, D.; Coffman, G. A.; Kreger, B. E.; Butterly, J. R. e Selker, H. P. — The Utility of Severity of Illness Information in Assessing the Quality of Hospital Care : The Role of Clinical Trajectory. *Medical Care*. 1992b; 30 (5), 428-444.
- Iezzoni, L. I.; Foley, S. M.; Heeren, T.; Daley, J.; Duncan, C.; Fisher, E. S. e Hughes, J. — A Method for Screening the Quality of Hospital Care Using Administrative Data : Preliminary Validation Results. *Quality Research Bulletin*. 1992c; November, 361-371.
- Iezzoni, L. I.; Foley, S. M.; Daley, J.; Hughes, J.; Fisher, E. S. e Heeren, T. — Comorbidities, Complications and Coding Bias : Does the Number of Diagnosis Codes Matter in Predicting In-Hospital Mortality? *JAMA*. 1992d; 267 (16), 2197-2203.
- Iezzoni, L. I.; Hotchin, E. K.; Ash, A. A.; Shwartz, M. e Mackiernan, Y. — MedisGroups Data Bases : The Impact of Data Collection Guidelines on Predicting In-Hospital Mortality. *Medical Care*. 1993; 31 (3), 277-283.
- Iezzoni, L. I.; Daley, J.; Heeren, T.; Foley, S. M.; Hughes, J. S.; Fisher, E. S.; Duncan, C. C. e Coffman, G. A. — Using Administrative Data to Screen Hospitals for High Complications Rates. *Inquiry*. 1994b; 31 (1), 40-55.
- Iezzoni, L. I.; Daley, J.; Heeren, T.; Foley, S. M.; Fisher, E. S.; Duncan, C.; Hughes, J. S. e Coffman, G. A. — Identifying Complications of Care Using Administrative Data. *Medical Care*. 1994c; 32 (7), 700-715.
- Iezzoni, L. I.; Ash, A. A.; Shwartz, M.; Daley, J.; Hughes, J. e MacKiernan, Y. D. — Predicting Who Dies Depends on How Severity is Measured : Implications for Evaluating Patient Outcomes. *Annals of Internal Medicine*. 1995a; 123 (10), 763-770.
- Iezzoni, L. I.; Shwartz, M.; Ash, A. A.; Hughes, J. S.; Daley, J. e Mackiernan, Y. D. — Using Severity-Adjusted Stroke Mortality Rates to Judge Hospitals. *International Journal for Quality in Health Care*. 1995b; 7 (2), 81-94.
- Iezzoni, L. I.; Shwartz, M.; Ash, A. A. e Mackiernan, Y. D. — Using Severity Measures to Predict the Likelihood of Death for Pneumonia Inpatients. *Journal of General Internal Medicine*. 1996a; 11 (1), 23-31.
- Iezzoni, L. I.; Shwartz, M.; Ash, A. A.; Hughes, J. S.; Daley, J. e Mackiernan, Y. D. — Severity Measurement Methods and

- Judging Hospital Death Rates for Pneumonia. *Medical Care*. 1996b; 34 (1), 11-28.
- Iezzoni, L. I.; Shwartz, M.; Ash, A. A. e Mackiernan, Y. D. — Does Severity Explain Differences in Hospital Length of Stay for Pneumonia Patients? *Journal of Health Services Research*. 1996c; 1 (2), 65-76.
- Iezzoni, L. I.; Shwartz M.; Ash A. A. e Mackiernan, Y. D. — Predicting In-Hospital Mortality for Stroke Patients : Results Differ across Severity-Measurement Methods. *Medical Decision Making*. 1996d ; 16 (4), 248-256.
- Iezzoni, L. I.; Ash, A. A.; Shwartz, M.; Daley, J.; Hughes, J. e Mackiernan, Y. D. — Judging Hospitals by Severity-Adjusted Mortality Rates : The Influence of the Severity-Adjustment Method. *American Journal of Public Health*. 1996e; 86 (10), 1379-1387.
- Iezzoni, L. I.; Ash, A. A.; Shwartz, M. e Mackiernan, Y. D. — Differences in Procedure Use, In-Hospital Mortality, and Severity by Gender for Acute Myocardial Infarction Patients : Are Answers Affected by Data Source and Severity Measure? *Medical Care*. 1997; 35 (2), 158-171.
- Iezzoni, L. I. — Using Administrative Diagnostic Data to Assess the Quality of Hospital Care. *International Journal of Technology Assessment in Health Care*. 1990; 6 (2), 272-281.
- Iezzoni, L. I. — Monitoring Quality of Care : What Do We Need to Know? *Inquiry*. 1993; 30 (2), 112-114.
- Iezzoni, L. I. — Risk Adjustment for Medical Effectiveness Research : An Overview of Conceptual and Methodological Considerations. *Journal of Investigate Medicine*. 1995; 43 (2), 136-150.
- Iezzoni, L. I. — An Introduction to Risk Adjustment. *American Journal of Medical Quality*. 1996a; 11 (1), S8-S11.
- Iezzoni, L. I. — On Opening «Black Boxes» and Looking Inside. *International Journal of Quality in Health Care*. 1996b; 8 (3), 209-210.
- Iezzoni, L. I. — Risks and Outcomes. In Iezzoni LI (editor) — «Risk Adjustment for Measuring Healthcare Outcomes». 2nd Edition. Chicago: Health Administration Press, 1997b. 1-40.
- Iezzoni, L. I. — The Risks of Risk Adjustment. *JAMA*. 1997c; 278 (19), 1600-1607.
- Iezzoni, L. I. — Dimensions of Risk. In Iezzoni LI (editor) — «Risk Adjustment for Measuring Healthcare Outcomes». 2nd Edition. Chicago: Health Administration Press. 1997d. 43-167.
- Iezzoni, L. I. — Assessing Quality Using Administrative Data. *Annals of Internal Medicine*. 1997e; 127 (8, Part 2), 666-674.
- Iezzoni, L. I. — Data Sources and Implications : Administrative Databases. In Iezzoni LI (editor) — «Risk Adjustment for Measuring Healthcare Outcomes». 2nd Edition. Chicago: Health Administration Press, 1997f. 169-242.
- Iezzoni LI — Statistically Derived Predictive Models : Caveat Emptor. *Journal of General Internal Medicine*. 1999; 14 (6), 388-389.
- Jacobs, P. — A Survey of Economic Models of Hospitals. *Inquiry*. 1974; 11 (2), 83-97.
- Jencks, S. F. e Dobson, A. — Refining Case-Mix Adjustment : The Research Evidence. *New England Journal of Medicine*. 1987; 317 (11), 679-686.
- Jencks, S. F.; Daley, J.; Draper, D.; Thomas, N.; Lenhart, G. e Walker J. — Interpreting Hospital Mortality Data : The Role of Clinical Risk Adjustment. *JAMA*. 1988; 260 (24), 3611-3616.
- Jencks, S. F.; Dobson, A.; Willis, P. e Feinstein, P. H. — Evaluating and Improving the Measurement of Hospital Case Mix. *Health Care Financing Administration*. 1984; Annual Supplement, 1-11.
- Jencks, S. F.; Williams, D. K. e Kay, T. L. — Assessing Hospital-Associated Deaths from Discharge Data : The Role of Length of Stay and Comorbidities. *JAMA*. 1988; 260 (15), 2240-2246.
- Jones, K. R. — Predicting Hospital Charge and Stay Variation : The Role of Patient Teaching Status, Controlling for Diagnosis Related Groups, Demographic Characteristics, and Severity of Illness. *Medical Care*. 1985; 23 (3), 220-235.
- Justice, A. C.; Covinsky, K. E. e Berlin, J. A. — Assessing the Generability of Prognostic Information. *Annals of Internal Medicine*. 1999; 130 (5), 515-524.
- Kaboli, P. J.; Barnett, M. J.; Fuehrer, S. M. e Rosenthal, G. E. — Length of Stay as a Source of Bias in Comparing Performance in VA and Private Sector Facilities. *Medical Care*. 2001; 39 (9), 1014-1024.
- Kalish, R. L.; Daley, J.; Duncan, C.; Davis, R. B.; Coffma, G. A. e Iezzoni, L. I. — Costs of Potential Complications of Care for Major Surgery Patients. *American Journal of Medical Quality*. 1995; 10 (1), 48-54.
- Kaplan, S. H.; Greenfield, S. e Ware, J. R. — Assessing the Effects of Physician-Patient Interactions on the Outcomes of a Chronic Disease. *Medical Care*. 1989; 27 (3), supplement, S110-S127.
- Kazis, L. E. — Health Outcome Assessments in Medicine : History, Applications, and New Directions. *Advanced Internal Medicine*. 1991; 36, 109-130.
- Kelleher, C. — Relationship of Physician Ratings of Severity and Difficulty of Clinical Management to Length of Stay. *Health Services Research*. 1993; 27 (6), 841-855.
- Knaus, W. A.; Zimmerman, J. E.; Wagner, D. P. e Lawrence, D. E. — APACHE : Acute Physiology and Chronic Health Evaluation : A Physiologically Based Classification System. *Critical Care Medicine*. 1981; 9 (8), 591-597.
- Knaus, W. A.; Wagner, D. P. e Draper, E. A. — The Value of Measuring Severity of Disease in Clinical Research on Acutely Ill Patients. *Journal of Chronic Diseases*. 1984; 37 (6), 455-463.
- Knaus, W. A.; Draper, E. A.; Wagner, D. P. e Zimmerman, J. E. — APACHE II : A Severity of Disease Classification System. *Critical Care Medicine*. 1985; 9 (8), 591-597.
- Knaus, W. A.; Draper, E. A.; Wagner, D. P. e Zimmerman, J. E. — An Evaluation of Outcome from Intensive Care in Major Medical Centers. *Annals of Internal Medicine*. 1986; 104 (3), 410-418.
- Knaus, W. A.; Wagner, D. P.; Draper, E. A.; Zimmerman, J. E.; Bergner, M.; Bastos, P. G.; Sirio, C. A.; Murphy, D. J.; Lotring, T.; Damiano, A. e Harrel, F. E. — The APACHE III Prognostic System : Risk Prediction of Mortality for Critically Ill Hospitalized Adults. *Chest*. 1991; 100 (6), 1619-1636.
- Knaus, W. A.; Wagner, D. P.; Zimmerman, J. E. e Draper, E. A. — Variations in Mortality and Length of Stay in Intensive Care Units. *Annals of Internal Medicine*. 1993; 118 (10), 753-761.
- Krakauer, H. e Jacoby, I. — Predicting the Course of Disease. *Inquiry*. 1993; 30 (2), 115-127.
- Krakauer, H.; Bailey, R. C.; Skellan, K. J.; Stewart, J. D.; Hartz, A. J.; Kuhn, E. M. e Rimm, A. A. — Evaluation of the HCFA Model for the Analysis of Mortality Following Hospitalization. *Health Services Research*. 1992; 27 (3), 317-335.
- Landon, B.; Iezzoni, L. I.; Ash, A. A.; Shwartz, M.; Daley, J.; Hughes, J. S. e Mackiernan, Y. D. — Judging Hospitals by

- Severity-Adjusted Mortality Rates : The Case of CABG Surgery. *Inquiry*. 1996; 33 (2), 155-166.
- Lave, J. R. e Leinhardt, S. — An Evaluation of a Hospital Stay Regulatory Mechanism. *Inquiry*. 1976; 13 (4), 327-342.
- Lave, J. R. e Franck, R. G. — Effect of the Structure of Hospital Payment on Length of Stay. *Health Services Research*. 1990; 25 (2), 327-347.
- Lawthers, A. G.; McCarthy, E. P.; Davis, R. B.; Peterson, L. E.; Palmer, R. H. e Iezzoni, L. I. — Identification of In-Hospital Complications From Claims Data : Is It Valid? *Medical Care*. 2000; 38 (8), 785-795.
- Le Gall, J. R.; Lemeshow, S. e Saulnier, F. — A New Simplified Acute Physiology Score (SAPSII) Based on a European/North American Multicenter Study. *JAMA*. 1993; 270 (24), 2957-2963.
- Lemeshow, S. e Hosmer, D. W. — A Review of Goodness of Fit Statistics for Use in the Development of Logistic Regression Models. *American Journal of Epidemiology*. 1982; 115 (1), 92-106.
- Lemeshow, S.; Teres, D.; Klar, J.; Avrunin, J. S.; Gehlbach, S. H. e Rapoport, J. — Mortality Probability Models (MPM II) Based on an International Cohort of Intensive Care Units Patients. *JAMA*. 1993; 270 (20), 2478-2486.
- Luke, R. D. — Dimensions in Hospital Case Mix Measurement. *Inquiry*. 1979; 16 (1), 38-49.
- Manheim, L. M.; Feinglass, J.; Shortell, S. M. e Hughes, E. F. — Regional Variation in Medicare Hospital Mortality. *Inquiry*. 1992; 29 (1), 55-66.
- Marazzi, A.; Paccaud, F.; Ruffieux, C. e Begun, C. — Fitting the Distributions of Length of Stay by Parametric Models. *Medical Care*. 1998; 36 (6), 915-927.
- McGuire, A.; Henderson, J. e Mooney, G. — The Economics of Health Care : An Introductory Text. London: Routledge, 1988.
- McMahon, L. F. e Billi, J. E. — Measurement of Severity of Illness and the Medicare Prospective Payment System: State of the Art and Future Directions. *Journal of General Internal Medicine*. 1988; 3 (5), 482-490.
- McMahon, L. F. e Smiths, H. L. — Can Medicare Prospective Payment Survive the ICD-9-CM Disease Classification System? *Annals of Internal Medicine*. 1986; 104 (4), 562-566.
- McNeil, B. K. e Haley, J. A. — Statistical Approaches to the Analysis of Receiver Operating Characteristic (ROC) Curves. *Medical Decision Making*. 1984; 4 (2), 137-150.
- MEDSTAT — Disease Staging Software, Version 4.12 : User Guide. Ann Arbor, Michigan: The MEDSTAT Group, Inc.; 2001.
- MEDSTAT — Disease Staging — Calibration and Recalibration Procedures. The MEDSTAT Group, Inc.; Ann Arbor, Michigan, 2002.
- Melo, F. G. — Métodos Estatísticos em Estudos Comparativos : Comparação de Tratamentos. Lisboa: Escola Nacional de Saúde Pública, 1985.
- Menemeyer, S. T.; Morrissey, M. A. e Howard, L. Z. — Death and Reputation : How Consumers Acted Upon HCFA Mortality Information. *Inquiry*. 1997; 34 (2), 117-128.
- Moreno, R.; Apolone, G. e Miranda, D. R. — Evaluation of the Uniformity of Fit of General Outcome Prediction Models. *Intensive Care Medicine*. 1998; 24 (1), 40-47.
- Mozes, B.; Easterling, M. J.; Sheiner, L. B.; Melmon, K. L.; Kline, R.; Goldman, E. S. e Brown, A. N. — Case-Mix Adjustment Using Objective Measures of Severity: The Case for Laboratory Data. *Health Services Research*. 1994; 28 (6), 689-712.
- Mullin, R. L.; Averill, R. F. e Boucher, S. A. — A Comparison of American and Australian DRG Systems. In 18th PCS/E, Innsbruck, Austria, 2002 — Patient Classification Systems Europe, 2-5 October 2002.
- Murphy D. J. e Cluff L. E. (editors) — SUPPORT : Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatments : Study Design. *Journal of Clinical Epidemiology*. 1990; 43 (Supplement). 1S-123S
- Naessens, J. M.; Leibson, C. L.; Krishan, I. e Ballard, D. J. — Contribution of a Measure of Disease Complexity (COMPLEX) to Prediction of Outcome and Charges Among Hospitalized Patients. *Mayo Clinic Process*. 1992; 67, 1140-1149.
- NHPC — National Health Performance Framework Report. Brisbane: National Health Performance Committee. Queensland Health, 2001.
- NHS — NHS Performance Indicators : Acute NHS Hospital Trusts: February, 2002. London, United Kingdom: National Health Service, 2002.
- O'Muircheartaigh, C.; Murphy, J. e Moore, W. — The 2002 Index of Hospital Quality. Chicago: NORC, University of Chicago, 2002.
- Pereira, J. — Economia da Saúde : Glossário de Termos e de Conceitos. Lisboa: Associação Portuguesa de Economia da Saúde, 1993 (Documento de Trabalho; 1/93)
- Plomman, M. P. — Choosing a Patient Classification System to Describe the Hospital Product. *Hospital & Health Services Administration*. 1985; 30 (3), 106-117.
- Polverejan, E.; Gardiner, J. C.; Bradley, C. J. e Rovner, M. H. — Estimation of Mean Hospital Cost as a Function of Length of Stay and Patient Characteristics. *Health Economics*. 2003; 12 (11), 935-947.
- Pollack, M. M.; Ruttiman, U. E. e Getson, P. R. — Pediatric Risk of Mortality (PRISM) Score. *Critical Care Medicine*. 1988; 16 (11), 1110-1116.
- Pollack, M. M.; Patel, K. M. e Ruttiman, U. E. — PRISM III : An Updated Pediatric Risk of Mortality. *Critical Care Medicine*. 1996; 24 (5), 743-752.
- Pryor, D. B. e Lee, K. L. — Methods for the Analysis and Assessment of Clinical Databases : The Clinician's Perspective. *Statistics in Medicine*. 1991; 10, 617-628.
- Quality Measurement Advisory Service — Quality Measurement Tools : Assessing Hospital Performance. Washington: QMAS, 1997.
- Rapoport, J.; Teres, D.; Zhao, Y. e Lemeshow, S. — Length of Stay as a Guide to Hospital Economic Performance for ICU Patients. *Medical Care*. 2003; 41 (3), 386-397.
- Rodrigues, S. — Análise da Produção do Internamento do Hospital de São Teotónio – Viseu: Os GDHs enquanto Instrumento de Gestão. Lisboa: Escola Nacional de Saúde Pública, 2003. Dissertação do XXXI Curso de Especialização em Administração Hospitalar.
- Rosen, H. M. e Gree, B. A. — The HCFA Excess Mortality Lists : A Methodological Critique. *Hospital & Health Services Administration*. February 1987; 119-127.
- Rosko, M. D. — DRGs and Severity of Illness Measures : An Analysis of Patient Classification Systems. *Journal of Medical Systems*. 1988; 12 (4), 257-274.
- Rubenstein, L. V.; Calkins, D. R.; Young, R. T.; Cleary, P. D.; Fink, A.; Kosekoff, J.; Jette, A. M.; Davies, A. R.; Delbano, T. L. e Brook, R. H. — Improving Patient Function : A Randomized Trial of Functional Disability Screening. *Annals of Internal Medicine*. 1989; 111 (10), 836-842.

- Ruttiman, U. E. — Statistical Approaches to Development and Validation of Predictive Instruments. *Critical Care Clinics*. 1994; 10 (1), 19-35.
- Selim, A. J.; Berlowitz, D. R.; Fincke, G.; Rosen, A.; Ren, X. S.; Christiansen, C. L.; Cong, Z.; Lee, A. e Kazis, L. — Risk-Adjusted Mortality Rates as a Potential Outcome Indicator for Outpatient Quality Assessments. *Medical Care*. 2002; 40 (3), 237-245.
- Shwartz, M.; Ash, A. A. e Iezzoni, L. I. — Comparing Outcomes Across Providers. In Iezzoni LI (editor) — «Risk Adjustment for Measuring Healthcare Outcomes». 2nd Edition. Chicago: Health Administration Press, 1997. 471-516.
- Shwartz, M.; Iezzoni, L. I.; Ash, A. A. e MacKiernan, Y. D. — Do Severity Measures Explain Differences in Length of Hospital Stay? The Case of Hip Fracture. *Health Services Research*. 1996; 31 (4), 365-385.
- Simborg, D. W. — DRG Creep : A New Hospital-Acquired Disease. *New England Journal of Medicine*. 1981; 304 (26), 1602-1604.
- Smith, P. J.; Thompson, T. J.; Engelgau, M. M. e Herman, W. H. — A Generalized Linear Model for Analyzing Receiver Operator Characteristic Curves. *Statistics in Medicine*. 1996; 15, 323-333.
- Snelling, I. — Do Star Ratings Really Reflect Hospital Performance? *Journal of Health Organization and Management*. 2003; 17 (3), 210-223.
- Stearns, S. C. — Hospital Discharge Decisions, Health Outcomes, and the Use of Unobserved Information on Case-Mix Severity. *Health Services Research*. 1991; 26 (1), 27-51.
- Steen, P. M.; Brewster, A. C.; Bradbury, R. C.; Estabrook, E. e Young, J. A. — Predicted Probabilities of Hospital Death as a Measure of Admission Severity of Illness. *Inquiry*. 1993; 30 (2), 128-141.
- Thomas, J. W.; Ashcraft, M. L. e Zimmerman, J. — An Evaluation of Alternative Severity of Illness Measures for Use by University Hospitals. Ann Arbor, Michigan: Department of Health Services Management and Policy, The University of Michigan, 1986.
- Thomas, J. W. e Longo D. R. — Application of Severity Measurement Systems for Hospital Quality Measurement. *Hospital & Health Services Administration*. 1990; 35 (2), 221-243.
- Thomas, J. W. e Ashcraft, M. L. — Measuring Severity of Illness : Six Severity Systems and Their Ability to Explain Cost Variations. *Inquiry*. 1991; 28 (1), 39-55.
- Thomas, J. W.; Holloway, J. J. e Guire, K. E. — Validating Risk-Adjusted Mortality as an Indicator for Quality of Care. *Inquiry*. 1993; 30 (1), 6-22.
- Thomas, J. W.; Bates, E. W.; Hofer, T.; Perkins, A.; Murphy, N. F. e Webb, C. — Interpreting Risk-Adjusted Length of Stay Patterns for VA Hospitals. *Medical Care*. 1998; 36 (12), 1660-1675.
- Thomas, J. W. e Hofer, T. P. — Research Evidence on the Validity of Risk-Adjusted Mortality Rate as a Measure of Hospital Quality of Care. *Medical Care Research and Review*. 1998; 55 (4), 371-404.
- Thomas, J. W. e Hofer, T. P. — Accuracy of Risk-Adjusted Mortality Rate as a Measure of Hospital Quality of Care. *Medical Care*. 1999; 37 (1), 83-92.
- Thomas, J. W. — Does Risk-Adjusted Readmission Rate Provide Valid Information on Hospital Quality? *Inquiry*. 1996; 33(3), 258-270.
- Tully, L. e Rulon, V. — Evolution of the Uses of ICD-9-CM Coding : Medicare Risk Adjustment Methodology for Managed Care Plans. *Topics in Health Information Management*. 2000; 21 (2), 62-67.
- Urbano, J. e Bentes, M. — Definição da Produção Hospitalar : Os Grupos de Diagnósticos Homogêneos. *Revista Portuguesa de Saúde Pública*. 1990; 8 (1), 49-60.
- Vladeck, B. C.; Goodwin, E. J.; Myers, L. P. e Sinisi, M. — Consumers and Hospital Use : the HCFA «death list». *Health Affairs*. 1988; 7 (1), 122-125.
- Vladeck, B. C. — Medicare Hospital Payment by Diagnosis-Related Groups. *Annals of Internal Medicine*. 1984; 100 (4), 576-591.
- Voss, B. W.; Hasman, A.; Rutten, F.; Zwaan, C. e Carpay, J. J. — Explaining Cost Variations in DRGs «Acute Myocardial Infarction» by Severity of Illness. *Health Policy*. 1994; 28 (1), 37-50.
- Weingart, S. N.; Iezzoni, L. I.; Davis, R. B.; Palmer, R. H.; Cahalane, M.; Hamel, M. B.; Mukamal, K.; Philips, R. S.; Davies, D. T. e Banks, N. J. — Use of Administrative Data to Find Substandard Care : Validation of the Complications Screening Program. *Medical Care*. 2000; 38 (8), 796-806.
- Weintraub, W. S.; Craver, J. M.; Cohen, C. L.; Jones, E. L. e Guyton, R. A. — Influence of Age on Results of Coronary Artery Surgery. *Circulation*. 1991; 84 (5), Supplement III, III226-III235.
- Wood, W. R.; Ament, R. P. e Kobrinsky, E. J. — A Foundation for Hospital Case Mix Measurement. *Inquiry*. 1981; 18 (3), 247-254.
- Young, W. W.; Kohler, S. e Kowalski, J. — PMC Patient Severity Scale : Derivation and Validation. *Health Services Research*. 1994; 29 (3), 367-390.
- Young, W. W.; Swinkola, R. B. e Zorn, D. M. — The Measurement of Hospital Case Mix. *Medical Care*. 1982; 20 (5), 501-512.
- Yuan, Z.; Cooper, G. S.; Einstadter, D.; Cebul, R. D. e Rimm, A. A. — The Association between Hospital Type and Mortality and Length of Stay. *Medical Care*. 2000; 38 (2), 231-245.

□ Abstract

RISK ADJUSTMENT: FROM CONCEPTUALIZATION TO OPERACIONALIZATION

Health organizations and hospitals in particular are frequently acknowledged for presenting specificities which make their productive process and management particularly complex.

In this context, the literature highlights some priorities, both in terms of research and of assessment, namely those related to production, payment, quality, efficiency, and performance evaluation.

The state of the art of performance evaluation of health organizations is driven by the trilogy conceived by Donabedian (1985) – Structure, Process and Outcomes. There are several perspectives to evaluate Outcomes – effectiveness, efficiency or financial performance. However, in order to consider the characteristics of patients that can influence those outcomes, risk adjustment methodologies are always necessary.

The list of outcome indicators is large, but one can mention the mortality (final outcomes), the complications and the readmissions (intermediate outcomes). With the exception of Thomas (1996) and Thomas and Hofer (1998 and 1999), there is evidence on the existence of a relation between these indicators and effectiveness of care.

To assess the technical efficiency of hospitals, the indicators usually used are costs and average length of stay. Also in this situation, the literature shows that severity increases the fitness model to explain resource consumption and that risk adjustment is also a useful tool to evaluate efficiency.

The development of measures of severity used for risk adjustment includes two central issues: the definition of the source of data and timing of its collection.

However, the choice between alternatives depends mainly on the priorities defined. If financial aspects are central, it is natural that administrative data will be preferred. When the focus is on construct validity and content validity, then the use of clinical data is inevitable.

Considering the timing of data collection two main issues should be addressed: the economic neutrality and the prospectiveness of the indicator. There is a large discussion concerning those aspects, since some authors argue that administrative models are preferable and others support the use of clinical models, while others defend that it is indifferent to use either type of measures, and that the choice should be made considering pragmatic issues, such as feasibility and costs of implementation and use.

In the short term, it is almost impossible to use clinical models in Portugal. However, this does not mean that in the medium term the hospital information system should not be adapted in order to include this type of data.

Finally, it should be mentioned that when a risk adjustment measure is applied to a different population from where it was developed, either geographical or historical, some procedures should be made in order to recalibrate the measure to this new population.

Keywords: risk adjustment; health care evaluation; outcomes; severity; information systems; hospitals; hospital administration.