

Efectividade e volume: uma aplicação a hospitais portugueses

SÍLVIA LOPES

Contexto: a avaliação da qualidade como tema potencialmente importante para utentes e prestadores de cuidados de saúde. A taxa de mortalidade como medida de resultados com um adequado ajustamento do risco. A existência de determinadas características estruturais do hospital às quais está associada uma menor mortalidade.

Objectivos: identificar diferenças no desempenho e na taxa de mortalidade dos hospitais e investigar que características estruturais justificam essas diferenças.

Metodologia: foram seleccionados os episódios de internamento das doenças de maior mortalidade hospitalar. A medida de desempenho considerada foi a comparação entre a mortalidade observada e a mortalidade esperada, calculada a partir da escala preditiva da mortalidade do Disease Staging, recalibrada para Portugal. A medida de desempenho foi analisada por hospital, doença e grupo de doenças. A ordenação dos hospitais pelo desempenho foi comparada com a ordenação dos hospitais pela taxa de mortalidade observada. O desempenho dentro de cada hospital foi analisado para um grupo de doenças seleccionadas. A relação entre o valor da medida de desempenho e as variáveis «número de episódios», «índice tecnológico» e «gravidade dos doentes tratados» foi analisada através da regressão linear para o conjunto dos episódios e para cada doença e grupo de doenças.

□

Sílvia Lopes é assistente da Escola Nacional de Saúde Pública — UNL.

Entregue em Março de 2006.

Resultados: foram incluídos 379 074 episódios, agrupados em 21 doenças e 8 grupos de doenças e tratados em 81 hospitais. A taxa de mortalidade observada foi de 12%. Existiam diferenças no desempenho por hospital, alguns dos quais se destacam pelo seu melhor/pior nível de desempenho. Foram observadas as limitações da taxa de mortalidade bruta como instrumento de análise do desempenho, no contexto de hospitais com diferentes níveis de risco dos doentes tratados. Para além disso, evidenciou-se que a análise do hospital como um todo ou em cada uma das partes tem resultados distintos, dada a existência de diferentes níveis de desempenho dentro do hospital. Finalmente, verificou-se que a relação entre volume e desempenho, quando existe, é, na quase totalidade dos casos, não linear e inversa à referida na literatura.

Palavras-chave: resultados; avaliação da qualidade; mortalidade; Disease Staging; ajustamento pelo risco.

1. Introdução

A qualidade da prestação de cuidados de saúde é um objectivo reconhecido quer por parte dos profissionais, directa ou indirectamente ligados à prestação, quer por parte dos doentes (Flood *et al.*, 1984). No entender de Donabedian (1988b), não existe apenas um, mas vários conceitos de qualidade, pois este varia consoante a posição em relação ao sistema de saúde e consoante aquilo que se entenda por «saúde». Assim, o conceito de qualidade pode ser entendido apenas no âmbito do desempenho dos pro-

fissionais ou ser progressivamente alargado ao contexto onde decorre a prestação, ao impacto da prestação na vida dos doentes e das suas famílias ou à distribuição social dos níveis de qualidade na comunidade.

O conceito de qualidade pode ser definido como a capacidade de atingir objectivos pretendidos, usando meios legítimos. Na maioria dos casos, os objectivos consistem num determinado estado de saúde atingível, com o conceito de saúde a ser usado de forma mais lata ou mais restrita (Donabedian, 1988b).

Assim se compreende que se apontem três razões para a avaliação da qualidade, nenhuma delas mutuamente exclusiva, mas antes em diferentes perspectivas. A «garantia de qualidade», uma abordagem na perspectiva da regulação, visa o cumprimento dos requisitos mínimos estabelecidos pelo regulador. A «competição pela qualidade», também numa abordagem de regulação, visa a recolha e divulgação de informações com vista a informar o mercado acerca da qualidade dos diferentes prestadores. A «melhoria da qualidade» visa recolher a evidência necessária à implementação das mudanças que permitirão melhorar a qualidade (Scanlon *et al.*, 2001).

Donabedian (1988b) definiu três dimensões de avaliação da qualidade: estrutura, processo e resultados. A estrutura de um hospital pode definir-se como as características do contexto em que decorre a prestação. Assim, a dimensão «estrutura» define as ferramentas e o ambiente disponíveis para os prestadores de cuidados (O'Muircheartaigh *et al.*, 2004). A estrutura inclui, pois, os recursos materiais, tais como instalações e equipamentos, os recursos financeiros, os recursos humanos, o número de profissionais e as suas qualificações, e a estrutura organizacional, tais como o modo de financiamento, organização do pessoal médico ou os protocolos de revisão (Donabedian, 1988b). Quanto ao conceito de resultados, trata-se de mudanças no estado de saúde que podem, com garantias, ser atribuídas à prestação que a antecedeu (Donabedian, 1988a). O processo consiste no que é efectivamente feito e inclui a procura de cuidados por parte do doente, o diagnóstico e a implementação do tratamento por parte do médico (Donabedian, 1988b). Estas três dimensões pressupõem que uma boa estrutura aumenta a probabilidade de um bom processo e um bom processo aumenta a probabilidade de bons resultados (Donabedian, 1988b). A hipótese sobre esta relação tem de ser comprovada e, se não for válida, a avaliação da qualidade também não o será (Donabedian, 1988a).

Sendo fundamental nas diversas actividades que compõem a prestação de cuidados (prestação de cuidados urgentes, em ambulatório ou de reabilitação), a qualidade adquire no internamento uma destacada

importância. Isto porque a prestação de cuidados em internamento é, sem prejuízo das restantes, uma função essencial para o hospital.

Neste trabalho procura-se identificar que prestadores apresentam melhor desempenho, no domínio da efectividade, através da taxa de mortalidade no internamento. Cumprido este primeiro objectivo, o segundo é a investigação de que razões, a nível das características estruturais, justificarão esse melhor desempenho, com especial destaque para o volume.

2. A relação entre volume e resultados

2.1. Evidências

A dimensão «estrutura» motiva e encoraja a escolha de acções eficazes, adequadas e custo-efectivas (Flood *et al.*, 2000). Assim, pela natureza da associação entre estrutura e resultados, a avaliação da estrutura dá informação sobre se as condições existentes potenciam ou não uma boa prestação (Donabedian, 1988b).

Entre as componentes da estrutura cuja associação com os resultados tem sido mais estudada contam-se as características dos recursos humanos, a existência de fim lucrativo, a existência da função ensino, a disponibilidade de tecnologia, a localização geográfica, o número de camas, o volume produtivo e outras características estruturais (nomeadamente a comunicação, coordenação, continuidade, *accountability*, autoridade, visibilidade da prática para outros, ambiente) (Donabedian, 1988a).

A relação entre o volume e os resultados tem sido alvo de inúmeras análises. Em 1996 tinham sido publicados 220 estudos nos quinze anos anteriores (excluindo aqueles que se debruçaram sobre bases de dados sensivelmente iguais) (Sowden e Sheldon, 1998). A maioria dos estudos indica a existência de uma associação, pelo menos, para alguns diagnósticos (Mitchell e Shortell, 1997). Burns e Wholey (1991) acrescentam que, enquanto para alguns diagnósticos a relação é positiva, para outros é negativa ou inexistente.

Estes resultados são geralmente interpretados à luz da trilogia proposta por Donabedian: estrutura, processo e resultados. Assume-se que hospitais de elevado volume têm características associadas a maior qualidade e julga-se que os prestadores nestes hospitais melhoram os seus processos através da experiência (Urbach e Baxter, 2004). Importa, no entanto, ter em conta que a probabilidade de fazer «a coisa certa» aumenta com o número de procedimentos realizados, mas este não é um requisito obrigatório (Ko *et al.*, 2002).

Luft, Hunt e Maerki (1987) propõem duas hipóteses para a explicação da relação entre volume e resultados. Uma delas foi já apontada: os resultados melhoraram devido à acumulação de experiência (*practice makes perfect*). A outra explicação é que os hospitais com melhores resultados atraem um maior número de doentes (*selective referral*).

Recentemente, Urbach e Baxter sustentaram ter, pela primeira vez, investigado a relação entre a mortalidade a trinta dias após um dado procedimento e o volume de outros procedimentos do hospital. Os resultados encontrados sugerem a existência de melhores resultados em hospitais de elevado volume, independentemente do procedimento no qual esse hospital é considerado de elevado volume (Urbach e Baxter, 2004).

Consoante a medida de resultados utilizada, a associação entre volume e resultados parece variar. Para as cirurgias por neoplasia do pulmão, a mortalidade no internamento mostrou-se negativamente associada ao volume, enquanto para a mortalidade a trinta dias não se encontrou associação (Bach *et al.*, 2001). Num outro estudo, os autores sugerem que, com o alargamento do período de análise, as variáveis de volume ganham importância (Ko *et al.*, 2002).

A existência de menores taxas de mortalidade ajustadas em hospitais de maior volume não significa que num hospital em que o volume aumenta ao longo do tempo a taxa de mortalidade ajustada diminua (Farley e Ozminkowski, 1992). Boles, cit. por Sowden e Sheldon (1998), defende também que, ainda que a mortalidade seja maior em hospitais de menor volume, o aumento do volume ao longo do tempo não diminui a mortalidade. Num outro estudo, os resultados indicam que a variação do volume ao longo do tempo não influencia os resultados, existindo indícios de que a relação entre volume e resultados reflecte diferenças fixas na qualidade entre hospitais de baixo e elevado volume (Hamilton e Ho, 1997).

2.2. Avaliação pelos resultados — o ajustamento do risco

Contudo, importa salientar que, ao analisar a relação entre a estrutura e os resultados, existem diversas questões a ter em conta, em especial no que concerne à forma de considerar os resultados. Antes de mais, existe a imperiosa necessidade de, antes de avaliar os resultados de uma instituição, ajustar as características dos doentes tratados e que influenciam o risco desse resultado. Assim, é necessário proceder ao ajustamento do risco, cujo objectivo é levar em conta as características pertinentes dos doentes, antes de

inferir acerca da efectividade dos cuidados prestados (Iezzoni, 1997a). As características dos doentes que influenciam a probabilidade de ocorrência de um determinado resultado variam entre os possíveis resultados e consoante o horizonte temporal considerado. No entanto, podem apresentar-se as seguintes dimensões gerais do risco: idade, sexo, raça/etnia, estado fisiológico do doente, diagnóstico principal, extensão e gravidade das comorbilidades, limitações funcionais, estado psicológico, cognitivo e psicossocial, características e comportamentos culturais e sócio-económicos, estado de saúde e qualidade de vida, atitudes e comportamentos face aos resultados (Iezzoni, 1997a).

Para além das características dos doentes apontadas, as variáveis de utilização dos serviços podem também ser usadas como indicadores de severidade (tipo de admissão, procedimentos realizados, internamentos prévios). No entanto, a introdução de variáveis controláveis pelo hospital não é desejável (Thomas e Longo, 1990; Wray *et al.*, 1997). O uso de procedimentos como medida de ajustamento pode induzir a sua utilização e, para além disso, se os procedimentos forem efectuados como resposta a uma complicação, a possibilidade de uma menor qualidade é interiorizada pela medida (Wray *et al.*, 1997).

Da literatura analisada verifica-se a existência de uma certa tendência no que toca às características relevantes dos dados administrativos e dos dados clínicos para o ajustamento do risco. No entanto, as opiniões sobre as potencialidades de cada um são bastante mais diversas.

Quanto às desvantagens dos dados administrativos, pode ser apontado o facto de estarem sujeitos a todas as potencialidades e limitações da ICD-9-CM (Wray *et al.*, 1997; Iezzoni, 1997a). Para além disso, a exaustividade da codificação das comorbilidades varia entre hospitais (Wray *et al.*, 1997), ao que acresce o facto de a codificação dos diagnósticos levantar mais questões do que a interpretação dos resultados das análises clínicas e dos sinais vitais (Costa e Nogueira, 1994). No entanto, ainda que não haja problemas na codificação, o resumo de alta não possui informação numa série de domínios, nomeadamente os resultados de análises laboratoriais e radiológicas (regularmente incluídas no processo clínico), ou as limitações funcionais (que também não são regularmente registadas no processo clínico), tendo estas demonstrado ser determinantes dos resultados (Wray *et al.*, 1997). Apesar da importância das desvantagens apresentadas, a natureza retrospectiva da codificação constitui, provavelmente, uma das principais desvantagens do uso de dados administrativos, pois não permite distinguir quais os diagnósticos adicionais presentes no momento da admissão

(Jencks *et al.*, 1988). Assim, a capacidade de previsão de resultados de sistemas baseados em informação administrativa deriva, em parte, da codificação de acontecimentos «próximos da morte», avaliações *post hoc* que melhoram a capacidade de previsão, mas comprometem a validade para o uso em estudos de qualidade (Iezzoni, 1997b). Por último, a possibilidade de manipulação da codificação dos dados, quer a nível da sobrecodificação em caso de óbito (*death-code creep*), quer a nível da codificação não sustentada pelo processo clínico (para efeitos de financiamento — *DRG creep*), constitui uma desvantagem no uso deste tipo de dados (Iezzoni, 1997b). As vantagens do uso de dados administrativos no ajustamento do risco podem ser sintetizadas nas que a seguir se enumeram (Iezzoni *et al.*, 1996a; Pine *et al.*, 1995; Thomas e Longo, 1990). São dados regularmente recolhidos pelos hospitais, pelo que estão mais facilmente disponíveis e a menor custo. Também a sua facilidade de tratamento é maior do que a dos dados clínicos, pelo facto de os dados administrativos estarem, desde logo, disponíveis informaticamente.

Ainda que com todas estas limitações, os dados administrativos são atractivos para alguns para o ajustamento do risco (Iezzoni, 1996a). Apesar da progressiva informatização do processo clínico e da crescente facilidade de tratamento de dados clínicos, a recolha e o tratamento de dados administrativos tendem a ser menos dispendiosos.

No que respeita a este tema em particular, verificou-se que a tendência da influência das características estruturais dos hospitais sobre a mortalidade era consistente, se comparados os resultados do modelo administrativo com o clínico (Krakauer *et al.*, 1992). O tipo de dados utilizado no ajustamento do risco é determinante, mas não pode ser dissociado do sistema de classificação de doentes/medida de gravidade utilizada. Neste trabalho foram usados dois sistemas de classificação de doentes: os Grupos de Diagnósticos Homogéneos (GDHs) e o Disease Staging.

Citando Fetter *et al.*, Hornbrook (1985) aponta como objectivo dos GDHs: relacionar as características demográficas, do diagnóstico e da terapêutica dos doentes com os cuidados que lhes são prestados, de forma a permitir que os doentes sejam diferentes apenas nas variáveis relacionadas com a sua condição (idade, diagnóstico principal, por exemplo) e processo de tratamento (procedimento, por exemplo) que afectam a sua utilização dos serviços do hospital. Na construção dos GDHs, a ênfase foi posta nos padrões de tratamento (Hornbrook, 1985).

O outro sistema utilizado foi o Disease Staging. Foi construído com base no princípio da não considera-

ção dos padrões de utilização dos recursos nem da resposta esperada à terapia; por base tem apenas o próprio modelo do processo da doença. O conceito de gravidade usado no Disease Staging é a probabilidade de morte como resultado da doença, sem consideração do tratamento. Foi desenvolvido para aplicação manual ao processo clínico, bem como para aplicação informática a um resumo de alta, sendo os resultados obtidos da primeira forma mais válidos, pela abundância relativa de informação.

As doenças são divididas em categorias de gravidade crescente (Gonella *et al.*, 1984):

Estadio 1: doença sem complicações;

Estadio 2: doença com complicações locais;

Estadio 3: doença que compreende diferentes localizações ou complicações sistémicas.

Os estadios e subestadios são atribuídos ao diagnóstico principal e às comorbilidades de cada episódio, tendo em conta os restantes diagnósticos presentes (Gonella *et al.*, 1984). Os estadios são ordinais dentro de cada doença principal: o estadio 1 numa dada doença principal pode ter implicações no risco de morte diferentes do mesmo estadio noutra doença principal; o estadio 2 não representa o dobro da gravidade do estadio 1 (Louis, 2004).

O sistema inclui também escalas preditivas (demora média, consumo de recursos, mortalidade, complicações, readmissões) que permitem comparar as doenças principais e em que se consideram a doença principal e o estadio, o número e a gravidade das comorbilidades, a idade, o sexo, o procedimento e o tipo de admissão (Louis, 2004).

A existência de diversos sistemas de classificação de doentes levanta a questão de saber até que ponto os seus resultados são distintos. Num estudo de 14 métodos distintos de ajustamento pelo risco em episódios de pneumonia, os resultados encontrados por Iezzoni *et al.* indicam que, a nível do episódio, as diferentes medidas (MedisGroups, Disease Staging, APACHE III, Patient Management Categories Severity Scale, All-Patient Refined Diagnosis Related Groups) discordavam com frequência em relação à probabilidade de morte. Por outro lado, a variação dos valores da comparação entre o observado e o esperado, a nível dos prestadores, mostra alguma concordância entre as diferentes medidas. Os resultados são semelhantes num estudo com episódios de enfarte agudo do miocárdio (Iezzoni, 1996b e 1996c).

Apesar de permanecer a dúvida sobre qual dos métodos será melhor para isolar as diferenças de qualidade, o ajustamento justifica-se como forma de promover o diálogo e melhorias na qualidade, bem

como de evitar a penalização dos hospitais em que são tratados doentes com maior risco e a hipótese de que, na sequência da divulgação de resultados, esses doentes sejam rejeitados pelos prestadores (Iezzoni, 1996c).

2.3. Validade da taxa de mortalidade como medida de resultados

Para saber se a mortalidade é apropriada como medida de qualidade deve procurar-se a existência de uma relação estreita entre esta e um critério cuja relação com a qualidade seja conhecida ou presumida (Park *et al.*, 1990; Thomas *et al.*, 1993). A existência de associação entre duas supostas medidas de qualidade, uma de processo e uma de resultados, a que se reconhecem imperfeições, confirma a validade de ambas. Quando as análises não mostram associação, fica por saber qual delas não é uma medida válida de qualidade (Thomas *et al.*, 1993). Outra proposta sugere que se recorra a modelos de simulação para conhecer o peso das variações aleatórias nas taxas de mortalidade (Park *et al.*, 1990)

As tentativas de validação da taxa de mortalidade como medida de resultados têm dado indicações distintas. Em alguns casos, a mortalidade mostrou ser válida como medida de resultados para toda a população considerada. Este foi o caso para a associação encontrada entre a taxa de mortalidade a trinta dias e a qualidade medida por revisão implícita (Rubenstein *et al.*, 1990); entre a mortalidade e a revisão implícita e explícita (Keeler *et al.*, 1992) e entre a taxa de mortalidade e a taxa de episódios com problemas de qualidade medida por revisão (Manheim *et al.*, 1992).

No entanto, por vezes, a validade da taxa de mortalidade mostrou variar consoante a doença. Enquanto existe uma associação significativa entre a qualidade avaliada por revisão explícita e a taxa de mortalidade ajustada a trinta dias para os episódios de insuficiência cardíaca congestiva, enfarte agudo do miocárdio, pneumonia e acidente cerebrovascular, esta não existe para os casos de fractura da anca (Kahn *et al.*, 1990). Num outro estudo, os resultados encontrados por Park *et al.* (1990) indicam que existe associação entre a taxa de mortalidade ajustada e a qualidade medida por revisão para episódios de insuficiência cardíaca congestiva, mas não para episódios de enfarte agudo do miocárdio. Noutro estudo, os resultados para o enfarte agudo do miocárdio, pneumonia, arritmia e alterações da condução os resultados são inconclusivos. Apesar disso, a taxa de mortalidade ajustada variava entre hospitais com diferentes níveis de qualidade medida por revisão para os episódios de

angina e cirurgia cardíaca e fracturas do fémur/pélicas. Para episódios de sepse, doença cerebrovascular, insuficiência cardíaca e choque e procedimentos no intestino, a taxa de mortalidade ajustada não diferia (Thomas *et al.*, 1993). Outros autores estabelecem uma separação distinta: a taxa de mortalidade ajustada não é uma boa medida da qualidade para episódios médicos, podendo sê-lo em alguns tipos de episódios cirúrgicos (Hofer e Hayward, 1996).

Assim, e apesar de não existir consenso a nível dos resultados encontrados, pode dizer-se que existe uma tendência para a validade da taxa de mortalidade ajustada como medida de resultados.

No entanto, a própria definição da medida levanta questões, nomeadamente no que diz respeito ao período de análise. Este período, no caso da taxa de mortalidade, pode estar limitado ao período do internamento ou pode incluir os trinta, sessenta ou noventa dias após a admissão ou, em certos casos, após a realização de determinado procedimento.

O uso das taxas de mortalidade no internamento deriva de diversos factores, um deles sendo a maior disponibilidade dos dados, face aos outros indicadores (Kahn *et al.*, 1988). Verifica-se também que a maioria dos hospitais não possui informação sobre os óbitos após a alta (Jencks *et al.*, 1988) e que se trata de um indicador de fácil interpretação e com relevo para o público (Kahn *et al.*, 1988).

Apesar disso, a consideração da taxa de mortalidade no internamento implica que se tenha em mente que, no contexto de uma pressão crescente para a redução da demora média, diversos autores apontem como inconveniente a sua sensibilidade à política de altas do hospital (Jarman et al., 1999; Sowden e Sheldon, 1998). Assim, a consideração de um limite preciso e igual para todos os hospitais, por exemplo trinta dias após a admissão, eliminaria esse efeito (Kahn et al., 1988). Também outros autores defendem esta posição, argumentando que, do ponto de vista da epidemiologia, é mais informativa a observação da mortalidade ao longo de um período uniforme do que ao longo de um período determinado pelas práticas médica e administrativa (Jencks et al., 1988).

3. Objectivos

Neste estudo procurou-se dar resposta aos seguintes objectivos:

- Identificar diferenças no desempenho ao nível da efectividade;
- Investigar que características estruturais justificam essas diferenças.

4. Metodologia

Recorreu-se a diversas fontes de dados, nomeadamente a base de dados dos resumos de alta, cedida pela Direcção-Geral da Saúde (DGS). Os dados para a construção do índice tecnológico foram obtidos tanto junto da DGS como através de dados divulgados pela Sociedade Portuguesa de Cuidados Intensivos.

Da base de dados dos anos de 2000 e 2001 foram excluídos os episódios cujo destino após a alta é diferente de «Para o domicílio» ou «Falecido» e os episódios de neonatologia. Seguidamente, seleccionaram-se os grupos de doenças do Disease Staging, responsáveis por 80% dos óbitos, e, dentro destes, as doenças principais, que representavam 80% dos óbitos. Os grupos de doenças foram a primeira unidade de análise, e só depois as doenças principais, para que estas fossem mais homogéneas entre si.

Foram depois acrescentadas a doença da artéria coronária com cirurgia de revascularização coronária (CV15) e todas as doenças principais «pneumonia» não incluídas na selecção inicial.

Deste valor foram excluídos: episódios admitidos em hospitais com menos de 200 episódios, admitidos em hospitais *outliers*¹ ao nível da taxa de mortalidade com idade registada superior a 100 anos e sem informação completa. A população em estudo ficou assim definida, com um total de 379 074 observações. Esta população, com alta em 81 hospitais, representa 20% dos internamentos e 62% dos óbitos nos hospitais presentes na base de dados dos resumos de alta nos anos de 2000 e 2001.

Para este trabalho estavam disponíveis dois sistemas de classificação de doentes: os GDHs e o Disease Staging. A opção pelo Disease Staging foi motivada pelo facto de o critério de distinção entre as categorias ser a diferente probabilidade de morte, pelo que pareceu mais adequado que, numa análise da mortalidade, os episódios fossem divididos em grupos relativamente homogéneos em relação à mortalidade e não à demora média.

Foram analisadas as seguintes variáveis dependentes: taxa de mortalidade observada, taxa de mortalidade esperada e desempenho.

A taxa de mortalidade esperada deriva da escala de mortalidade do Disease Staging, que utiliza os resumos de alta dos hospitais (a mesma base que serve para a classificação dos GDHs), embora com *outputs* diferentes: doença principal, comorbilidades, estadios e subestadios da doença principal (para estes são igualmente consideradas as complicações), estadios e subestadios das comorbilidades, mortalidade prevista

e duração de internamento prevista. Como tanto a mortalidade como a duração de internamento previstas estão calibradas para os dados de origem (EUA), as mesmas foram recalibradas para os dados portugueses (Costa *et al.*, 2003).

A medida de efectividade do hospital deriva da comparação entre valores esperados e observados do número de óbitos. Para obter o número de óbitos esperados de um hospital foi calculada a média da probabilidade de morte dos episódios do hospital. O valor obtido foi depois multiplicado pelo número de episódios desse hospital. A comparação foi feita através de um *z-score*, calculado da seguinte forma (Iezzoni, 1997b): desempenho = (número de óbitos observado – número de óbitos esperado)/desvio-padrão dos óbitos observados. Deste modo, um hospital com um valor do *z-score* negativo tem um número de óbitos observados inferior ao número de óbitos esperados. Na opção pela subtracção procurou-se evitar que os resultados fossem sensíveis a pequenas variações de pequenos valores (Shwartz *et al.*, 1997). Por outro lado, distinguiram-se hospitalares que, com zero óbitos observados, mostravam distintos óbitos esperados (Luft *et al.*, 1987).

As variáveis independentes aqui consideradas são as seguintes: gravidade, índice tecnológico e volume. Para medir a gravidade dos doentes tratados (GRAV) utilizou-se a percentagem de doentes saídos no estadios 3 do Disease Staging no total de doentes saídos em cada hospital. Este estadios foi seleccionado por ser aquele que, dos três possíveis, melhor distinguiu os hospitais. O índice tecnológico (IT) de cada hospital foi avaliado pela metodologia usada no relatório «The 2003 index of hospital quality» (O’Muircheartaigh *et al.*, 2004). O volume é dado pelo número de episódios de internamento. De forma a considerar eventuais efeitos não lineares, este foi considerado de duas formas, termo simples (DS) e ao quadrado (DS²) (Hamilton e Ho, 1998)².

Os modelos foram considerados não significativos nos casos em que a significância do teste *F* indicava a não validade do modelo a nível global (< 0,05), ou em que todos os coeficientes não são significativamente diferentes de zero, ou em que existe multicolinearidade³, ou em que os resíduos são não normais.

² Inicialmente, foram incluídas variáveis relativas à lotação, pessoal médico e pessoal de enfermagem, mas foram excluídas por se encontrarem correlacionadas entre si, sendo o volume usado como a única variável de dimensão. A partir do código de residência registado na base de dados dos resumos de alta procurou-se distinguir os hospitais quanto ao nível sócio-económico dos doentes, mas esta variável não é incluída no modelo, por introduzir multicolinearidade.

³ Os modelos em que o *condition index* se situa entre 15 e 30 estão assinalados.

¹ Os *outliers* foram identificados pelo método [$Q3 + 1,5*(Q3 - Q1)$ < X < $Q1 - 1,5*(Q3 - Q1)$] (Hoaglin, 1993).

Os grupos de doenças/doenças principais com menos de 30 episódios não foram analisados individualmente.

5. Resultados

5.1. Caracterização da população

Depois de seleccionados os episódios a analisar, o número total de observações cifrou-se nas 379 074 e a taxa de mortalidade nos 12,01%. A idade dos doentes é, em média, de 64,51 (DP: 21,86) anos. Ao nível dos grupos de doenças, são as doenças respiratórias (28%) e as doenças cardiovasculares e do coração (23%) as mais frequentes.

A base de dados por hospital contém um total de 81 observações, o valor da taxa de mortalidade observada é de 12% (intervalo de confiança: 11,91%; 12,12%) e coincide com a taxa de mortalidade esperada (12%; intervalo de confiança: 11,91; 12,00%). Por hospital, a taxa de mortalidade observada varia entre 3,79% (3,30; 4,29) e 21,64% (20,56; 22,71) e a taxa de mortalidade esperada varia entre 3,72%

(3,48%; 3,96%) e 19,28% (18,47%; 20,09%). O hospital com a menor taxa de mortalidade observada (5) é também o hospital com a menor taxa de mortalidade esperada. O hospital com a maior taxa da mortalidade observada (12) não coincide com o hospital com a maior taxa de mortalidade esperada (87).

Em algumas das variáveis independentes existem diferenças consideráveis entre hospitais: a média de doentes saídos no grupo de hospitais com menor número de episódios (isto é, até ao primeiro quartil) é de 1139, sendo de 10 187 no grupo de hospitais com maior número de episódios. O mesmo acontece no caso do índice tecnológico: os hospitais com menor nível tecnológico apresentam um índice de 2, tendo os hospitais com maior nível tecnológico um índice de 19.

5.2. Análise do desempenho por hospital

O indicador de desempenho, construído a partir da comparação entre os valores observados e esperados do número de óbitos por hospital, está representado

Quadro I
Características dos hospitais

(n = 81)	Média	DP	Média até Q ₁	Média depois Q ₃	Mínimo	Máximo
TM (%)	12,01	–	8,95	17,36	3,79	21,64
TM esperada (%)	11,96	–	9,71	15,51	3,72	19,28
Volume	4 680	3962	1 139	10 187	362	20 017
Gravidade	32,40	–	24,78	39,75	8,32	52,09
Tecnologia	9,40	6,76	2	19	0	29

TM = taxa de mortalidade.

na *Figura 1*. Apenas num hospital (47) os valores esperados e observados coincidem. O hospital 12 mostra o maior valor do *z-score* do conjunto dos hospitais: 0,80. Neste hospital, o número de óbitos esperado é de 867 e o observado é de 1220 (mais 353 óbitos). O hospital 61 mostra o menor valor do *z-score*: -0,62. Neste hospital, o número de óbitos esperado é de 1384 e o observado é de 1111 (menos 273 óbitos).

Existem 38 hospitais com valores negativos do *z-score*, e o número de episódios com alta nestes hospitais representa 56% do total. Os hospitais com um número de óbitos superior ao esperado são 43 e representam 44% do total de episódios. Assim, verifica-se que um pouco mais de metade dos episódios em análise foi tratado nos hospitais com bom desempenho (número de óbitos inferior ao esperado). O facto de estes hospitais representarem cerca de metade das observações verifica-se por construção, dadas as características do *z-score*.

Por forma a distinguir entre os hospitais com bom desempenho, aqueles que se destacam pelo seu elevado desempenho, foram calculados os *outliers* da medida de desempenho. Identificaram-se, deste modo, dois tipos de hospitais: hospitais com elevado

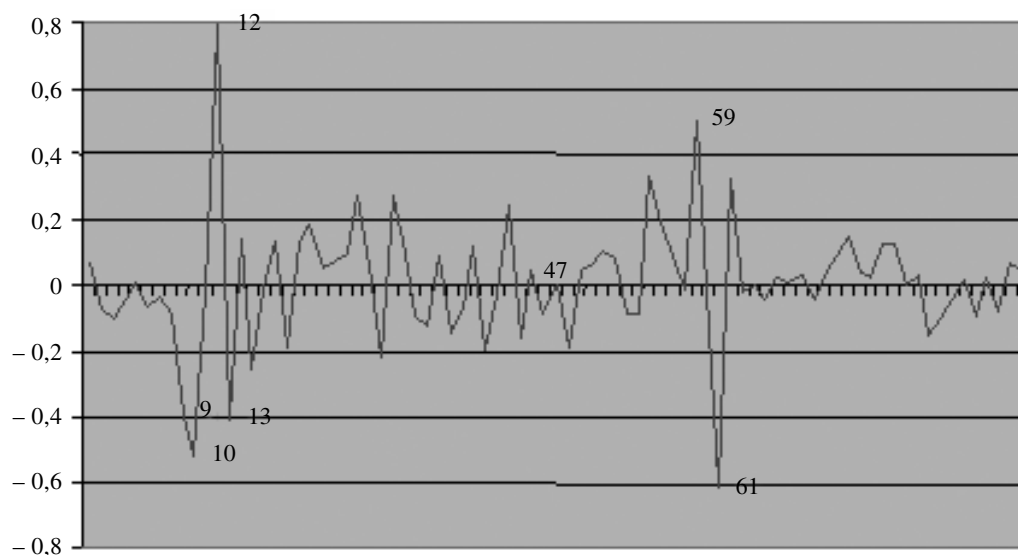
desempenho (hospitais 9, 10, 13 e 61) e hospitais com fraco desempenho (hospitais 12 e 59).

Ao analisar os 81 hospitais, nas diferentes perspectivas⁴, 46 não se destacam em nenhuma delas pelo seu elevado/fraco desempenho. Dos 35 que se destacam pelo menos numa perspectiva, 14 destacam-se pelo seu elevado desempenho numas perspectivas e fraco desempenho noutras, 8 apenas pelo seu elevado desempenho e 13 apenas pelo seu fraco desempenho. Analisando por hospital, destacam-se os hospitais 13 e 61 pela frequência com que surgem entre os hospitais com elevado desempenho (12). Contudo, o hospital 13 figura entre os hospitais de fraco desempenho quando as perspectivas de análise são o traumatismo craniano (CN08), a doença coronária (CORO) e a fractura do colo ou cabeça do fémur (MS09).

Os hospitais que surgem com mais frequência no grupo de fraco desempenho são os hospitais 12 e 59. No entanto, nos episódios de bronquite e bronquiolite

⁴ A designação «perspectiva» é usada ao longo do texto para fazer referência aos grupos de episódios analisados separadamente: episódios médicos, episódios cirúrgicos, grupos de doenças e doenças principais.

Figura 1
Medida de desempenho por hospital



aguda (RS73), o hospital 12 está entre os melhores hospitais.

Em suma, verifica-se que o desempenho varia entre hospitais, existindo hospitais que se destacam pelo seu pior/melhor desempenho.

5.3. Análise comparada do desempenho e da taxa de mortalidade observada

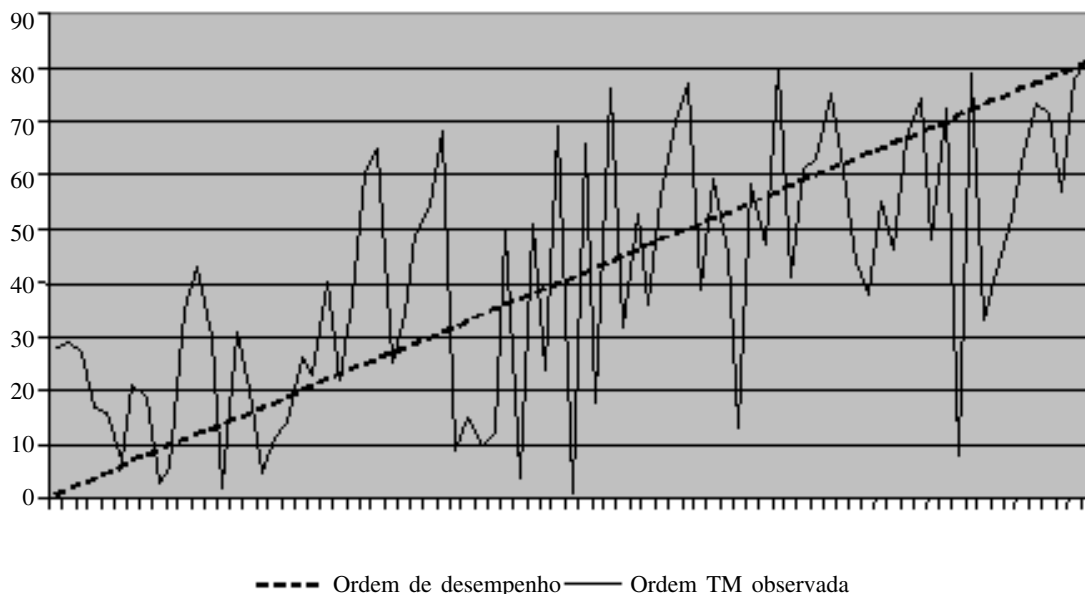
Até aqui pode verificar-se que existem diferenças no desempenho entre hospitais, surgindo a questão sobre a coincidência entre estas diferenças e as diferenças da taxa de mortalidade observada. Assim, os hospitais foram ordenados pelo desempenho e pela taxa de mortalidade observada e as ordenações foram comparadas através do valor Kappa. Quando este é inferior a 0,40, a concordância entre a ordenação pela taxa de mortalidade observada e a ordenação pelo desempenho é fraca (Pestana e Gageiro, 2003).

Na *Figura 2*, os hospitais estão ordenados por ordem crescente de desempenho: os hospitais mais à

esquerda têm bom desempenho (observado inferior ao esperado); os hospitais mais à direita têm mau desempenho. Da análise da figura verifica-se que a posição relativa de cada hospital varia consoante se analise a taxa de mortalidade observada ou o desempenho, não se vislumbrando um padrão. O valor de Kappa observado (0,01) indica uma fraca concordância entre os critérios de ordenação. Apesar disso, as ordens estão positiva e moderadamente correlacionadas (0,62).

Se analisarmos a taxa de mortalidade observada em cada quartil da medida de desempenho, verifica-se que no quartil onde se situam os 20 melhores hospitais a taxa de mortalidade observada varia entre 5,2% e 12,7%. Verifica-se ainda que o valor mínimo da taxa de mortalidade observada (3,79%) se encontra no grupo de hospitais entre as posições 20.^a e 40.^a na ordenação por desempenho. O valor máximo da taxa de mortalidade observada (21,64%) encontra-se depois do terceiro quartil. Tal como foi observado na *Figura 1*, o hospital com o valor máximo da taxa de mortalidade encontra-se na 81.^a posição do desempenho.

Figura 2
Ordem relativa dos hospitais por taxa de mortalidade (TM) observada e por desempenho



Em geral, à medida que se avança na medida de desempenho para hospitais com excesso de óbitos observados, a taxa de mortalidade tende a ser maior, o que explica a existência de correlação positiva.

Por outro lado, não há uma distinção nítida na taxa de mortalidade observada nos diferentes quartis de desempenho, uma vez que a taxa de mortalidade média de um determinado quartil está compreendida entre o valor máximo e mínimo de todos os quartis. Para todas as perspectivas analisadas, o Kappa indica que não existe concordância entre a ordenação pelo desempenho e a ordenação pela taxa de mortalidade observada; no máximo, ele toma o valor de 0,078. Existem perspectivas em que, ao contrário do que se verifica no total de episódios, existe uma baixa correlação (inferior a 0,40) entre o desempenho e a taxa de mortalidade observada: doenças respiratórias (RS), neoplasia do pulmão, brônquios ou mediastino (RS15) e pneumonia (PNEU). Este facto poderá ser explicado pela existência de grandes diferenças, por hospital, entre a ordenação pelos dois critérios.

Em suma:

- A nível global, não existe qualquer concordância entre a taxa de mortalidade e o desempenho. No entanto, existe correlação entre os dois;
- Em todas as perspectivas não existe concordância, para além de que a ordenação pela taxa de mortalidade nem sempre está correlacionada com a ordenação pelo desempenho.

5.4. Análise das diferenças de desempenho por hospital

Depois de analisadas as diferenças de desempenho entre os diversos hospitais, procurou-se analisar o

desempenho do mesmo hospital em cada uma das diferentes perspectivas, o que se designou por «variabilidade interna».

A existência de variabilidade interna pode já ser observada em análises anteriores. Recorde-se que foi identificada a existência de 14 hospitais que em algumas perspectivas estavam no grupo dos melhores hospitais (*outliers* negativos), enquanto noutras estavam no grupo dos piores hospitais (*outliers* positivos).

Verifica-se que, pelo teste *t* para amostras emparelhadas, as médias dos desempenhos nos episódios cirúrgicos e médicos são significativamente diferentes da média do desempenho no total de episódios. Quanto aos 10 hospitais com maiores diferenças no desempenho entre episódios médicos e cirúrgicos, 5 deles estão entre os 20 hospitais com melhor desempenho a nível global.

Para uma análise por doença principal, seleccionaram-se as cinco doenças principais com maior taxa de mortalidade observada e as cinco doenças principais com o maior número de óbitos. Estas doenças principais representam 56% do total de doentes saídos e o desempenho total do hospital está altamente correlacionado com o desempenho nestas doenças (coeficiente de Pearson = 0,82). A existência de correlação entre o coeficiente de variação do desempenho dentro do hospital e o número de doenças tratadas (coeficiente de Pearson = 0,47) levou a que se considerassem apenas os 26 hospitais que tratavam as nove doenças escolhidas.

A variação do desempenho dentro de um hospital nas diferentes doenças principais pode ser analisada de duas formas: ou comparando o valor da medida de desempenho nas doenças principais, ou comparando a posição relativa do hospital nas doenças principais, estando as duas formas de medir a variabilidade

Quadro II

Média, desvio-padrão, mínimo e máximo da taxa de mortalidade e da medida de desempenho, por quartil da medida de desempenho

Quartil	Desempenho				Taxa de mortalidade (%)			
	Média	DP	Mínimo	Máximo	Média	DP	Mínimo	Máximo
1	-0,21	0,15	-0,62	-0,08	9,65	1,87	5,15	12,66
2	-0,03	0,03	-0,07	0,01	11,51	3,17	3,79	16,35
3	0,05	0,02	0,01	0,09	14,32	2,98	9,49	20,99
4	0,22	0,17	0,09	0,80	15,00	3,19	8,49	21,64

interna correlacionadas entre si (coeficiente de Pearson = 0,77).

Cada hospital, em média, apresenta um coeficiente de variação da medida de desempenho de 0,15, ou seja, no limite máximo da dispersão fraca. Já quanto ao coeficiente de variação da posição relativa dá indicação de uma dispersão média ou elevada do desempenho em cada hospital (0,55).

Formaram-se quatro grupos de hospitais, com base no coeficiente de variação da posição relativa⁵ e identificou-se qual o nível de desempenho mais frequente em cada quartil. Os hospitais mais frequentes até ao primeiro quartil do coeficiente de variação são os de pior desempenho, enquanto os hospitais mais frequentes depois do terceiro quartil do coeficiente

⁵ Apresentam-se apenas os resultados para os quartis do coeficiente de variação da posição relativa, uma vez que os resultados obtidos para o coeficiente de variação do desempenho são relativamente semelhantes.

de variação são aqueles que mostram um melhor desempenho, o que significa que, à medida que o coeficiente de variação aumenta, o nível de desempenho também aumenta.

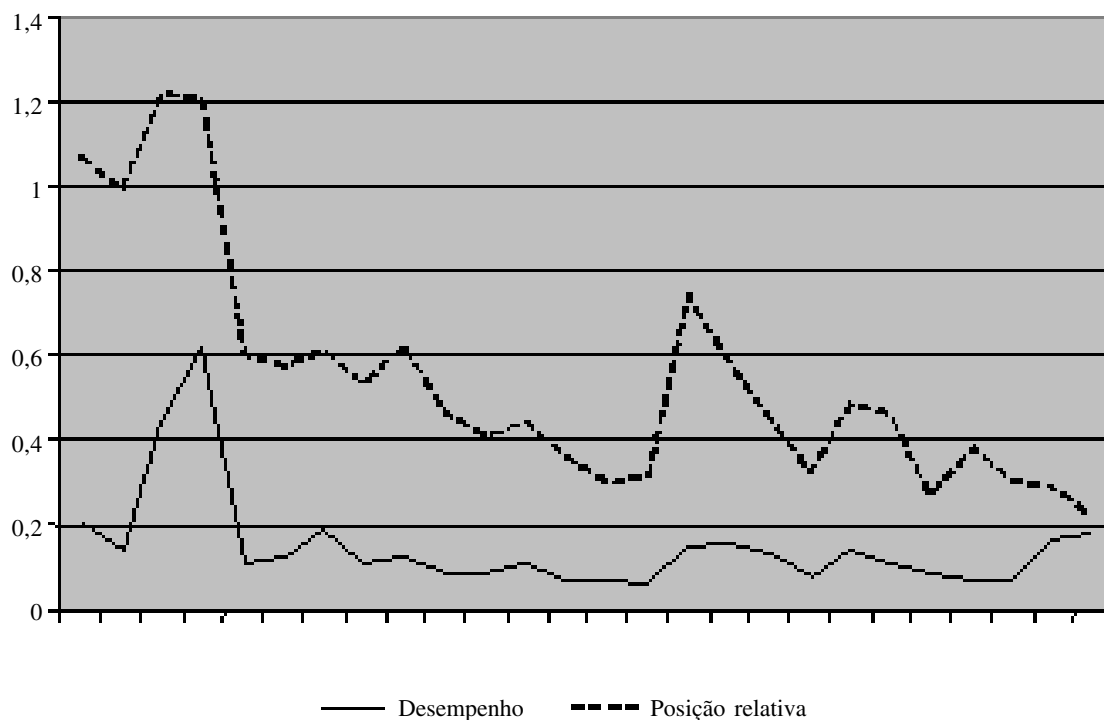
Em suma:

- Existem diferenças ao nível do desempenho dentro de cada hospital;
- Os hospitais de melhor desempenho tendem a apresentar as maiores diferenças.

5.5. Relação entre características dos hospitais e desempenho

Para identificar possíveis influências das características dos hospitais no seu desempenho ao nível da efectividade recorreu-se à análise de regressão, relacionando o desempenho com o número de episódios (termo simples e ao quadrado), o índice tecnológico e a gravidade.

Figura 3
Coeficiente de variação do desempenho e da posição relativa por hospital (ordenados por ordem crescente de desempenho)



Começando pelo total de episódios, verifica-se que existe uma associação entre o volume (tanto na forma linear como quadrática) e o desempenho. O modelo explica 20,7% da variação do desempenho apenas com estas variáveis. A probabilidade de os coeficientes da gravidade e do índice tecnológico serem zero (isto é, não estarem relacionados com o desempenho), analisados pelo teste *t*, leva à sua não inclusão no modelo explicativo.

Na representação dos valores previstos pelo modelo de desempenho e do número de episódios (*Figura 4*) observa-se que existe uma relação não linear, o que era já visível pelo facto de a variável DS2 ser significativa. Assim, o modelo indica que o desempenho é melhor (menor valor do *z-score*) nos hospitais com poucos e nos hospitais com muitos doentes saídos. Ou seja, até um certo ponto, o aumento do volume implica que o desempenho piore. A partir desse ponto, aquele onde a inclinação da curva se inverte, o aumento do volume implica que o desempenho melhore. Pelo modelo, os valores mínimos do *z-score* verificam-se nos hospitais de maior volume.

Analisando os resíduos deste modelo, verifica-se que estes seguem uma distribuição normal quando a nor-

malidade é testada pelo teste de Kolmogorov-Smirnov (K-S).

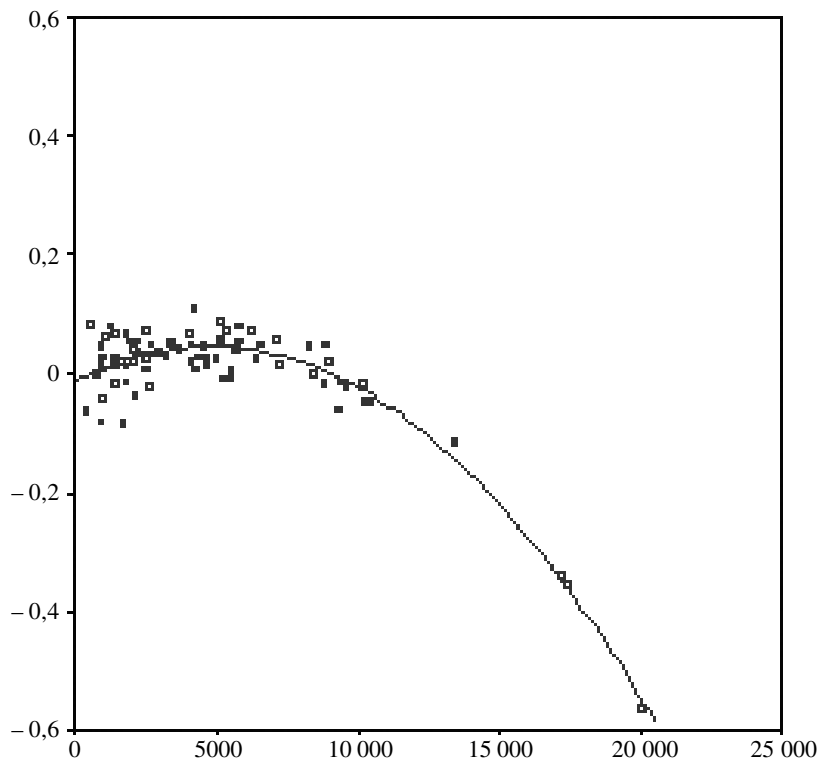
A mesma análise foi realizada para cada uma das perspectivas (30), sendo o modelo significativo em apenas 13 delas.

O modelo não é significativo para os episódios cirúrgicos, enquanto, para os episódios médicos, apenas a variável DS2 é significativa.

Existem perspectivas que mostram um comportamento semelhante ao total dos episódios, ou seja, até um determinado ponto, o desempenho piora com o aumento do volume e, a partir desse ponto, o desempenho melhora com o aumento do volume. Também nestas perspectivas os valores mínimos do *z-score* se encontram nos hospitais de maior volume. Este grupo é composto pelas doenças do sistema nervoso central (CN), doença cerebrovascular (CN17), doenças gastrointestinais (GI), neoplasia do cólon e recto (GI24), doenças hepatobiliares (HB), neoplasia do pâncreas (HB09), outra neoplasia maligna do sistema hepatobiliar e pancreático (HB70) e diabetes *mellitus* (ND12).

O índice tecnológico mostra-se significativo, em conjunto com as variáveis DS e DS2, para a infecção das

Figura 4
Valores da medida de desempenho previstos pelo modelo por número de episódios



vias urinárias (RN01), em que o valor da medida de desempenho diminui quando aumenta o valor do índice tecnológico. Em dois casos, a doença pulmonar obstrutiva crónica (RS13) e a insuficiência renal (RN02), a única variável independente com coeficiente significativo é a gravidade. Na primeira (RS13), os hospitais com episódios com maior gravidade têm um pior desempenho, verificando-se o inverso na segunda (RN02). Estes valores indiciam que uma aplicação do modelo aos diferentes estádios tem resultados distintos.

O valor máximo do R^2a (0,775) encontra-se nos episódios de outra neoplasia maligna do sistema hepatobiliar e pancreático (HB70), enquanto o valor mínimo (0,099) se encontra nas doenças do sistema nervoso central (CN).

O modelo não é significativo para o traumatismo craniano (CN08), doenças cardiovasculares (CV), arritmia e alterações da condução (CV09), doença coronária (CORO), outra doença cardíaca (CV73), neoplasia do estômago (GI26), úlcera péptica (GI30), outras doenças do intestino e peritoneu (GI75), fratura do colo ou cabeça do fémur (MS09), doenças renais (RN), doenças respiratórias (RS), neoplasia do pulmão, brônquios ou mediastino (RS15), pneumonia (PNEU), bronquite e bronquiolite aguda (RS73), outras doenças agudas do pulmão (RS76) e outras doenças do sistema respiratório, excepto doenças respiratórias altas (RS80).

Em suma:

- Na maioria das perspectivas (17 em 30), o modelo de desempenho não é significativo;
- A variável que, com maior frequência, explica as diferenças de desempenho é o volume. A relação entre volume e desempenho não é linear;
- Em algumas perspectivas, o desempenho varia com a gravidade. Apenas numa das perspectivas varia com o índice tecnológico.

6. Discussão

6.1. Discussão metodológica

Em primeiro lugar, os resultados são condicionados pelo uso de dados administrativos para o ajustamento do risco. Tal como apontado no enquadramento teórico, estes dados não permitem a distinção entre os diagnósticos presentes no momento da alta e aqueles que se desenvolveram depois da admissão, permitindo «interiorizar» potenciais diferenças de qualidade. Para minorar esta característica dos dados administrativos foi implementada no estado da Califórnia, desde 1996, uma medida com vista a que os

hospitais, para cada diagnóstico secundário registado, registem também se estava ou não presente na admissão (Averill *et al.*, 1998). No entanto, ainda que com essas melhorias, os dados administrativos mantêm-se condicionados pelas potencialidades e limitações da codificação com a ICD-9-CM. É apontada na literatura a existência de diferenças nas práticas de codificação dos diferentes hospitais. Não são conhecidos estudos sobre a fiabilidade da codificação em Portugal, pelo que não foi possível abordar esse tema. No entanto, e uma vez que é aceite que, potencialmente, existem menos problemas na fiabilidade dentro de cada hospital — os procedimentos são mais homogêneos neste plano — do que entre hospitais (Costa e Nogueira, 1994), os resultados da análise à variabilidade interna serão pouco influenciados por este facto. Um estudo em que fossem incluídos dados clínicos não teria estas desvantagens. Alguma da informação que poderia ser utilizada está já, com relativa frequência, disponível informaticamente, nomeadamente ao nível das análises laboratoriais. A sistematização dessa informação no resumo de alta traria vantagens: a avaliação de desempenho que considerasse essas medidas, desde que devidamente assegurada a protecção dos dados, seria uma ferramenta com maior potencial para identificar áreas de melhoria nos hospitais. Outra vantagem é a possibilidade de análise dos dados clínicos de uma forma não retrospectiva: podem ser analisados em distintos momentos do episódio e podem ser estabelecidas comparações entre esses momentos. Para além disso, essa comparação, se feita ainda durante o internamento, serve de base à identificação de necessidades de revisão e, se necessário, de implementação de mudanças com efeitos mais imediatos. A utilização de dados clínicos num trabalho deste tipo traria, muito provavelmente, resultados diferentes, pois, à partida, estaria em condições de identificar hospitais com maior frequência de complicações evitáveis/não expectáveis.

Por outro lado, existem dimensões do risco não consideradas: os dados não incluem a existência de limitações funcionais. Existem outras variáveis apontadas como dimensões do risco pelas quais não foi possível ajustar: a existência de diferentes atitudes e comportamentos face aos resultados, características e comportamentos sócio-culturais e económicos, estado de saúde e qualidade de vida. Não existiam dados disponíveis para considerar estas questões, pelo que, a existirem diferenças sistemáticas entre hospitais nestes domínios, existirão alguns cujo risco estará sub/sobreavaliado.

Está estudado o facto de diferentes sistemas de classificação de doentes atribuírem diferentes níveis de gravidade ao mesmo episódio (Iezzoni, 1996b e 1996c).

Ainda que essas diferenças sejam menores quando se analisa o desempenho de cada hospital, elas existem, o que permite perguntar se os resultados seriam diferentes se outro sistema que estivesse disponível tivesse sido usado. A resposta a essa pergunta é, provavelmente, sim. As dúvidas são maiores no que toca à dimensão dessa diferença.

A medida de resultados usada foi o número de óbitos hospitalares, que pode ser sensível à política de altas do hospital. Seria interessante, caso estivessem disponíveis dados fiáveis, a comparação entre os resultados obtidos pelo método da mortalidade hospitalar e pelo método da mortalidade num período fixo. Este é um problema que se coloca de outra forma quando se usam dados clínicos, já que, nesse caso, a análise das diferenças na gravidade do doente no momento da alta permite identificar diferentes políticas de altas dos hospitais.

Analisando as variáveis incluídas no estudo, é de referir que a consideração da tecnologia ao nível do hospital e não especificamente por cada doença pode ter consequências sobre a frequência com que esta está associada às diferenças de desempenho. Aliás, o mesmo relatório a cuja metodologia se recorreu para a construção do índice tecnológico (American Best Hospitals) recorre também a índices de tecnologia específicos. A análise da relação entre um índice de tecnologia adaptado a cada doença e o desempenho pode produzir resultados distintos dos encontrados e ser, eventualmente, encontrada com maior frequência. No entanto, uma vez que se pretendia comparar entre doenças, foi usado o índice geral. Por outro lado, existe também a eventualidade de que a percentagem de doentes saídos no estadió 3 não seja aquela que, ao nível de cada doença, melhor distinga a gravidade dos episódios em cada hospital.

Com a inclusão de variáveis aqui não consideradas, como a coordenação, comunicação ou a liderança, é possível que o poder explicativo do modelo de desempenho aumentasse. A disponibilidade dos recursos humanos e materiais não é suficiente para um bom desempenho, pois estes podem ser subaproveitados. No caso dos recursos humanos, um grupo de bons profissionais não faz necessariamente um bom quadro de recursos humanos: a sua distribuição pelas áreas do hospital, acompanhamento e *feedback* disponibilizados e possibilidade de formação contínua são aspectos decisivos para a efectividade. Para além de que a retenção dos profissionais e o acumular de experiência, com o tempo, tendem a melhorar os resultados. As características da comunicação e da coordenação no hospital são também importantes e têm, potencialmente, efeito sobre os resultados para os doentes. Todas estas características não foram incluídas neste estudo, sendo de esperar que possam

explicar variações do desempenho. Uma análise deste tipo é complexa, em particular pela operacionalização das variáveis, mas também pela necessidade da recolha de dados primários. No entanto, é uma área a merecer investigação para que, uma vez identificadas, se possam diminuir as más práticas e disseminar as boas práticas.

6.2. Discussão dos resultados

O primeiro dos objectivos deste trabalho — identificar diferenças no desempenho — foi concretizado. Efectivamente, quando se mede o desempenho de um hospital a partir da mortalidade, existem hospitais com melhor desempenho do que outros, existindo alguns que se destacam pelo seu melhor/pior desempenho.

Verificou-se ainda que a análise da taxa de mortalidade observada dá indicações distintas da análise da efectividade. Este resultado é compreensível, pois os hospitais diferem no nível de risco dos doentes tratados: a taxa de mortalidade esperada varia entre 3,7% e 19,3%. Assim, uma avaliação com base na taxa de mortalidade «prejudicaria» os hospitais que tratam estes doentes mais graves, ou seja, com maior risco. A análise à taxa de mortalidade bruta é uma hipótese cujas limitações têm sido frequentemente apontadas na literatura.

Uma análise ao nível do hospital permitiu observar que os hospitais que, ao nível global, têm melhor desempenho são os que apresentam maior variabilidade interna. Este facto parece indiciar que esse melhor desempenho foi alcançado a partir de «ilhas de excelência» dentro do hospital, mais do que a partir de uma situação global. Ou seja, parece existir uma compartimentação dentro de cada hospital que permite que um hospital esteja entre os melhores do país numa dada doença e entre os piores do país numa outra⁶. Poderia apontar-se que, com o aumento da dimensão do hospital, este tipo de situações é mais frequente, o que foi confirmado pela existência de uma correlação positiva entre o número de doenças e a medida de variabilidade interna. No entanto, quando se analisaram apenas os hospitais que tratavam todas as doenças, este resultado manteve-se: foram os hospitais que mostravam melhor desempenho a nível global aqueles que mostraram maior variabilidade interna. Este facto poderá estar relacionado com a diversidade da oferta do hospital e com a existência de objectivos distintos e nem sempre interdependentes entre os diferentes grupos/departa-

⁶ Tal como descrito nos resultados, existem 14 hospitais nestas circunstâncias, de um total de 81.

mentos. É importante salientar que nesta diversidade de situações reside uma diversidade de oportunidades de melhoria (Flood *et al.*, 2000).

Quanto ao segundo objectivo — investigar que características estruturais justificam diferenças de desempenho —, constatou-se que, para o total de episódios, o desempenho tem uma relação não linear com o volume: os hospitais com melhor desempenho são aqueles que se encontram nos valores extremos de volume. A associação do volume aos resultados é apontada por diversos estudos (Bach, 2001; Birkmeyer, 2002; Flood *et al.*, 1984; Hamilton e Ho, 1998; Urbach e Baxter, 2004; Ko *et al.*, 2002). Na neoplasia do pâncreas, uma das perspectivas em que o volume apresenta maior poder explicativo das diferenças de desempenho, o desempenho mostrou-se associado ao volume com frequência (Urbach e Baxter, 2004).

Na literatura, na maioria dos casos, verifica-se uma relação linear: o desempenho melhora à medida que o volume aumenta. A hipótese de uma relação não linear é também compreensível: até um determinado nível de volume o desempenho melhora com o aumento do volume; a partir desse nível, a dimensão do hospital é tal que ultrapassa a escala óptima e o desempenho piora cada vez mais à medida que o volume aumenta. A relação entre volume e resultados aqui obtida difere da encontrada na literatura. No entanto, da metodologia usada neste estudo, por comparação com a usada noutros estudos, não são evidentes diferenças que possam explicar esse facto. O número de observações ao nível de episódios e de hospitais é considerável e a medida de ajustamento do risco usada é internacionalmente reconhecida. Daí que se entenda que a diferença deriva, provavelmente, das características dos hospitais analisados neste estudo.

Neste ponto colocam-se duas hipóteses: as diferenças resultam de práticas na codificação ou resultam de diferenças efectivas dos hospitais. Caso se trate da primeira hipótese, será necessário actuar sobre as características do registo no processo clínico, bem como aumentar o investimento na formação dos codificadores, de forma a aumentar a fiabilidade dos dados. A fiabilidade da informação presente nos resumos de alta é decisiva para a actividade do hospital, por servir, para além do financiamento, para caracterizar a oferta do hospital e ser um mapa de possibilidades de melhoria no hospital. A segunda hipótese, caso se verifique, tem implicações distintas. Analisando os níveis de volume em que o seu aumento implica um pior desempenho, podem apontar-se algumas explicações possíveis. Uma delas é que nesses hospitais existem outras variáveis de estrutura associadas ao volume e aqui não incluídas. Em particular, dimensões como a coordenação entre

as diferentes partes do hospital ou as características da comunicação podem ter características menos desejáveis nos hospitais de maior volume e influenciar, dessa forma, a relação entre volume e resultados na primeira parte da curva. Este é um facto que urge estudar com maior profundidade. Em segundo lugar, as características do processo desempenham aqui um papel fundamental, pois a relação entre estrutura e resultados é «mediada» pelo processo. Assim, nesses casos é também possível que a influência das características da estrutura sobre o processo possa não ser a esperada. Seria relevante realizar um estudo mais profundo, com maior número de variáveis e inclusão de variáveis de processo, que permitisse a rejeição ou aceitação das hipóteses aqui avançadas.

Ao contrário do que se aponta na literatura (Flood, 1984), o desempenho nos episódios cirúrgicos não varia com o volume, enquanto isso se verifica nos episódios médicos. Investigou-se se este facto poderia estar relacionado com a distribuição dos valores do desempenho, tendo-se verificado que, nos episódios cirúrgicos, a amplitude interquartis da medida de desempenho é superior em cerca de metade dos casos e que, à excepção da doença coronária, existem diferenças consideráveis ao nível da percentagem de *outliers* no total das observações. A existência de níveis de desempenho mais «dispersos» pode dificultar a identificação de associações com as variáveis incluídas. Esta grande diferença nos resultados entre hospitais em episódios cirúrgicos é uma questão que merece aprofundamento, de forma a encontrar possíveis razões para a sua existência.

Enquanto existem estudos em que os resultados, a nível global, são influenciados pela tecnologia (Manheim *et al.*, 1992; Kuhn *et al.*, 1991), neste estudo, as infecções das vias urinárias são a única doença principal em que o desempenho dos hospitais é influenciado pela tecnologia disponível. O facto de o desempenho nas diversas neoplasias não variar com o índice tecnológico, onde se inclui a existência de radioterapia, e o facto de o desempenho nas doenças cardiovasculares não variar com o mesmo índice, onde se inclui a realização de angioplastias, cateterismos, cirurgias a céu aberto e disponibilidade de unidade de cuidados intensivos de cardiologia, levantam a suspeita de que aquela se trata de uma relação espúria. Uma das hipóteses é que os doentes de cada nível de risco, por via da referência, sejam tratados em hospitais de tecnologia semelhante, e a similitude de níveis de tecnologia explicaria a ausência de associação. Esta é uma das várias questões levantadas por este trabalho que justificam uma investigação mais profunda.

Algumas doenças destacaram-se pelo facto de o desempenho estar associado à gravidade. Uma delas

é a insuficiência renal, em que, quanto menor a gravidade, melhor o desempenho. Neste caso, a taxa de mortalidade esperada por subestadio varia entre 4% e 51%. Assim, é de esperar que dentro desta doença estejam incluídos episódios de características muito distintas, pelo que as variáveis que influenciam o desempenho serão distintas. Também para a doença pulmonar obstrutiva crónica, o desempenho varia com a gravidade, mas os hospitais com maior percentagem de episódios graves têm melhor desempenho. No entanto, os níveis de risco dos episódios são relativamente semelhantes entre si, pelo que só uma análise mais aprofundada permitiria explicar estas diferenças.

7. Síntese final

O objectivo da medição da qualidade é, evidentemente, estabelecer o ponto de partida para a sua melhoria (Blumenthal, 1996, 891). Nesta linha, este trabalho tinha dois objectivos centrais. Primeiro, a identificação da existência de diferenças de desempenho. Segundo, a investigação das características estruturais que podem justificar estas diferenças.

Abordaram-se na primeira parte as questões relativas às potencialidades da análise da mortalidade no internamento como medida de qualidade e preocupações que devem estar presentes na análise. As dimensões da estrutura do hospital, que são frequentemente apontadas como influenciando os resultados, em particular o volume, foram abordadas pela perspectiva de diferentes autores. Depois de apresentadas as hipóteses que orientariam a análise e qual a metodologia a adoptar, apresentaram-se os resultados obtidos, entre os quais se destacaram os que se seguem. O desempenho varia entre hospitais, existindo hospitais que se destacam pelo seu pior/melhor desempenho.

A análise da taxa de mortalidade, a nível global, não mostra qualquer concordância com a análise do desempenho, embora estejam correlacionadas. No entanto, para algumas doenças, esta correlação não existe.

Pode observar-se ainda que existem diferenças a nível do desempenho dentro de cada hospital, tendendo os hospitais com melhor desempenho a mostrar as maiores diferenças internas.

No que se refere aos resultados da análise da regressão, verifica-se que, na maioria das perspectivas (17 em 30), o modelo de desempenho não é significativo. Nos que são significativos, a variável que, com maior frequência, explica as diferenças de desempenho é o volume. A relação entre volume e desempenho não é linear. Para além disso, em algumas perspectivas, o

desempenho varia com a gravidade. Apenas numa das perspectivas varia com o índice tecnológico.

Estes resultados foram influenciados pela metodologia adoptada, destacando-se o uso de dados administrativos para o ajustamento do risco. Também neste domínio, mas no que toca às variáveis incluídas, se considera a possibilidade de que a inclusão de outras variáveis de recursos humanos, que não o número, bem como de variáveis de estrutura no domínio da coordenação e comunicação, aumentaria o poder explicativo do modelo do desempenho.

Os resultados aqui encontrados a nível das variáveis explicativas do desempenho diferem dos mais frequentes na literatura, pelo que se sugere uma análise mais aprofundada destas matérias que complemente e melhore a já efectuada. De todas as linhas de investigação futura que foram sendo apontadas ao longo do trabalho, esta é certamente a mais relevante.

□ Bibliografia

AVERILL, R. F., *et al.* — The evolution of case-mix measurement using Diagnosis Related Groups (DRGs). *3M HIS Research Report*. 5 : 98 (1998) 1-40.

BACH, P. B., *et al.* — The influence of hospital volume on survival after resection for lung cancer. *New England Journal of Medicine*. 345 : 3 (2001) 181-188.

BIRKMEYER, J. D., *et al.* — Hospital volume and surgical mortality in the United States. *New England Journal of Medicine*. 346 : 15 (2002) 1128-1137.

BLUMENTHAL, D. — Quality of health care : what is it? *New England Journal of Medicine*. 335 : 12 (1996) 891-894.

BURNS, L. R.; WHOLEY, D. R. — The effects of patient, hospital and physician characteristics on length of stay and mortality. *Medical Care*. 29 : 3 (1991) 251-271.

COSTA, C.; LOUIS, D.; GONELLA, J. — An assessment of Portuguese hospital performance. 19th PCSE Conference, Washington, 2003.

COSTA, C.; NOGUEIRA, P. — A medição da produção do hospital : a importância da fiabilidade. *Revista Portuguesa de Saúde Pública*. 12 : 2 (1994) 31-40.

DONABEDIAN, A. — Quality assessment and assurance : unity of purpose, diversity of means. *Inquiry*. 25 : 1 (1988) 173-192.

DONABEDIAN, A. — The quality of care : how can it be assessed? *Journal of American Medical Association*. 260 : 12 (1988) 1743-1748.

FARLEY, D. E.; OZMINKOWSKI, R. J. — Volume-outcome relationships and in hospital mortality : the effect of changes in volume over time. *Medical Care*. 30 : 1 (1992) 77-94.

- FLOOD, A. B., *et al.* — Organizational performance : managing for efficiency and effectiveness. In SHORTELL, S. M.; KALUZNY, A. D., ed. lit. — Managing health care organizations : organization theory and behaviour. 3rd ed. New York : Delmar, 1994.
- FLOOD, A. B.; SCOTT, W. R.; EWY, W. — Does practice make perfect? Part I : The relation between volume and outcomes for selected diagnosis categories. *Medical Care*. 22 : 2 (1984) 98-114.
- FLOOD, A. B.; SCOTT, W. R.; EWY, W. — Does practice make perfect? Part II : The relation between volume and outcomes and other hospital characteristics. *Medical Care*. 22 : 2 (1984) 115-125.
- GONNELLA, J. S.; HORN BROOK, M. C.; LOUIS, D. Z. — Staging of disease : a case-mix measurement. *Journal of American Medical Association*. 251 : 5 (1984) 637-644.
- HAMILTON, B. H.; HO, V. — Does practice make perfect? Examining the relationship between hospital volume and outcomes for hip fracture patients in Quebec. *Medical Care*. 36 : 6 (1998) 892-903.
- HOFER, T. P.; HAYWARD, R. A. — Identify poor-quality hospitals : can hospital mortality rates detect quality problems for medical diagnoses? *Medical Care*. 34 : 8 (1996) 737-753.
- HORN BROOK, M. C. — Techniques for assessing hospital case mix. *Annual Review of Public Health*. 6 (1985) 295-324.
- IEZZONI, L. I. — Data sources and implications : administrative databases. In IEZZONI, L. I., ed. lit. — Risk adjustment for measuring health outcomes. 2nd ed. Chicago : Health Administration Press, 1997a.
- IEZZONI, L. I. — Dimensions of risk. In IEZZONI, L. I., ed. lit. — Risk adjustment for measuring health outcomes. 2nd ed. Chicago : Health Administration Press, 1997a.
- IEZZONI, L. I. — Risk and outcomes. In IEZZONI, L. I., ed. lit. — Risk adjustment for measuring health outcomes. 2nd ed. Chicago : Health Administration Press, 1997a.
- IEZZONI, L. I. — The risks of risk adjustment. *Journal of American Medical Association*. 278 : 19 (1997b) 1600-1607.
- IEZZONI, L. I., *et al.* — Judging hospitals by severity-adjusted mortality rates : the influence of the severity-adjustment method. *American Journal of Public Health*. 86 : 10 (1996a) 1379-1387.
- IEZZONI, L. I., *et al.* — Predicting in-hospital mortality for stroke patients : results differ across severity-measurement methods. *Medical Decision Making*. 16 : 4 (1996b) 346-356.
- IEZZONI, L. I., *et al.* — Severity measurement methods and judging hospital death rates for pneumonia. *Medical Care*. 34 : 11 (1996c) 11-28.
- JARMAN, B., *et al.* — Explaining differences in English hospital death rates using routinely collected data. *British Medical Journal*. 318 : 7197 (1999) 1515-1520.
- JENCKS, S. F.; WILLIAMS, D. K.; KAY, T. L. — Assessing hospital-associated deaths from discharge data : the role of length of stay and co-morbidities. *Journal of American Medical Association*. 260 : 15 (1988) 2240-2246.
- KAHN, K. L., *et al.* — Interpreting hospital mortality data. *Journal of American Medical Association*. 260 : 24 (1988) 3625-3628.
- KAHN, K. L., *et al.* — Measuring quality of care with explicit process criteria before and after implementation of the DRG-based prospective payment system. *Journal of American Medical Association*. 264 : 15 (1990) 1969-1973.
- KAHN, K. L., *et al.* — The effects of the DRG-based prospective payment system on quality of care for hospitalized Medicare patients. *Journal of American Medical Association*. 264 : 15 (1990) 1953-1955.
- KEELER, E. B., *et al.* — Hospital characteristics and quality of care. *Journal of American Medical Association*. 268 : 13 (1992) 1709-1714.
- KO, C. Y., *et al.* — Are high-volume surgeons and hospitals the most important predictors of in-hospital outcome for colon cancer resection? *Surgery*. 312 : 2 (2002) 268-273.
- KRAKAUER, H., *et al.* — Evaluation of the HCFA model for the analysis of mortality following hospitalization. *Health Services Research*. 27 : 3 (1992) 317-335.
- KUHN, E. M., *et al.* — The relationship of hospital characteristics and the results of peer review in six large states. *Medical Care*. 29 : 10 (1991) 1028-1037.
- LOPES, S. — Mortalidade hospitalar e estrutura dos hospitais. Dissertação do XXXII Curso de Especialização em Administração Hospitalar (2002-2004).
- LOUIS, D. Z. — Severity of illness and the evaluation of health care. In Workshop, 1, Lisboa, 21-22 de Janeiro de 2004 — Gestão em Saúde. Lisboa: [s. n.], 2004.
- LUFT, H. S.; HUNT, S. S.; MAERKI, S. C. — The volume-outcome relationship : practice makes perfect or selective referral patterns? *Health Services Research*. 22 : 2 (1987) 157-182.
- MANHEIM, L. M., *et al.* — Regional variations in Medicare hospital mortality. *Inquiry*. 29 : 1 (1992) 55-66.
- MITCHELL, P. H.; SHORTELL, S. M. — Adverse outcomes and variations in organization of care delivery. *Medical Care*. 35 : 11 (1997) 19-32.
- O'MUIRCHARTAIGH, C.; BURKE, A.; MURPHY, W. — The 2003 index of hospital quality. Washington, DC: NORC, [2004] (U.S. News & World Report's «America's Best Hospitals») http://www.norc.uchicago.edu/new/2003_Index_of_Hospital_Quality.pdf (2004-04-29).
- PARK, R. E., *et al.* — Explaining variations in hospital death rates : randomness, severity of illness, quality of care. *Journal of American Medical Association*. 264 : 4 (1990) 484-490.
- PESTANA, M.; GAGEIRO, J. — Análise de dados para ciências sociais : a complementaridade do SPSS. 3.ª ed. rev. e aumentada. Lisboa : Edições Sílabo, 2003.
- PINE, M., *et al.* — Predictions of hospital mortality rates: a comparison of data sources. *Annals of Internal Medicine*. 126 : 5 (1997) 347-354.
- RUBENSTEIN, L. V., *et al.* — Changes in quality of care for five diseases measured by implicit review. *Journal of American Medical Association*. 264 : 15 (1990) 1974-1979.
- SCANLON, D. P., *et al.* — Options for assessing PPO quality : accreditation and profiling as accountability strategies. *Medical Care Research and Review*. 58 : Suppl. 1 (2001) 70-100.
- SHWARTZ, M.; ASH, A. S.; IEZZONI, L. I. — Comparing outcomes across providers. In IEZZONI, L. I., ed. lit. — Risk adjustment for measuring health outcomes. 2nd ed. Chicago : Health Administration Press, 1997b.
- SOWDEN, A. J.; SHELDON, T. A. — Does volume really affect outcome? Lessons from the evidence. *Journal of Health Services Research and Policy*. 3 : 3 (1998) 187-190.
- THOMAS, J. W.; HOLLOWAY, J. J.; GUIRE, K. E. — Validating risk-adjusted mortality as an indicator of quality of care. *Inquiry*. 30 : 1 (1993) 6-22.
- THOMAS, J. W.; LONGO, D. R. — Application of severity measurement systems for hospital quality management. *Hospital and Health Services Administration*. 35 : 2 (1990) 221-243.

URBACH, D. R.; BAXTER, N. — Does it matter what a hospital is «high volume» for? Specificity of hospital volume-outcome associations for surgical procedures : analysis of administrative data. *British Medical Journal*. 328 : 7442 (2004) 737-740.

WRAY, N. P., *et al.* — Case-mix adjustment using administrative databases : a paradigm to guide future research. *Medical Care Research and Review*. 54 : 3 (1997) 326-356.

□ Abstract

EFFECTIVENESS AND VOLUME : A STUDY ON PORTUGUESE HOSPITALS

Context: quality is an important matter, both for patients and caregivers, being the mortality rate one of the possible indicators of results to be monitored. However, it is fundamental to adjust for the differences in the patients' risk. There is an extensive body of literature concerning the impact of the hospital structure in its results.

Objectives: to identify differences in hospitals' performance, in terms of their effectiveness and to investigate which structural characteristics might be related to those results.

Methods: the study includes diseases with high impact in-hospital mortality. The performance is evaluated through the comparison between observed mortality and predicted mortality (from Disease Staging predictive scales). Analyses of the indicator of performance for each hospital, disease and group of diseases were made. The ranking of mortality was compared with the ranking of performance for each hospital. The differences in performance per disease within the hospital were investigated. Performance was regressed against a group of variables (volume, technology and severity), for all admissions, for each group of diseases and for each disease.

Results: a database with 379 074 admissions (21 diseases, 8 groups) was considered, with a mortality rate of 12%. There were differences in performance among the group of hospitals, and some outliers were found. The limitations of not adjusting mortality rate for the patients' risk when evaluating performance were shown. Hospitals have a different performance, depending on the disease. Performance was found to be non-linearly related with volume, but this relation is opposite to that described in the literature.

Keywords: outcomes; quality assessment; mortality; Disease Staging; risk adjustment.