

Técnicas de análise de sobrevida

Survival analysis techniques

Maria Teresa Bustamante-Teixeira ¹

Eduardo Faerstein ²

Maria do Rosário Latorre ³

¹ Núcleo de Assessoria
Treinamento e Estudos
em Saúde, Universidade
Federal de Juiz de Fora.
Rua Antônio Carlos Pereira
328, Juiz de Fora, MG
36071-110, Brasil.
teita@nates.ufjf.br

² Departamento de
Epidemiologia, Instituto
de Medicina Social,
Universidade Estadual
do Rio de Janeiro.
Rua São Francisco Xavier
524, Bloco E,
Rio de Janeiro, RJ
20559-900, Brasil.
eduardof@uerj.br

³ Departamento de
Epidemiologia, Faculdade
de Saúde Pública,
Universidade de São Paulo.
Av. Dr. Arnaldo 715,
São Paulo, SP
01246-904, Brasil.
mdrddola@usp.br

Abstract *Statistical methods known as survival analyses are useful for analyzing time-related events, in which time from a benchmark event to an endpoint is the focus of interest. Survival analysis describes not only patient survival statistics (as suggested by the name), but also other dichotomous outcomes such as time of remission, time of breastfeeding, etc. This paper discusses survival analysis techniques, commenting and comparing their utilization, especially in the field of oncology. It also presents and discusses types of epidemiological studies and data sources to which this type of analysis is applied. The authors take into account the difference between hospital-based or clinical series and population-based approaches. Interpretation of results is also discussed.*

Key words *Survivors; Survival Analysis; Epidemiologic Studies*

Resumo *As técnicas estatísticas conhecidas como análise de sobrevida são utilizadas quando se pretende analisar um fenômeno em relação a um período de tempo, isto é, ao tempo transcorrido entre um evento inicial, no qual um sujeito ou um objeto entra em um estado particular e um evento final, que modifica este estado. Assim, descrevem não só, como sugerido por seu nome, se os pacientes vivem ou morrem, mas também outros desfechos dicotômicos tais como recaída da doença, desmame do lactente etc. Neste trabalho são abordadas técnicas de análise de sobrevida, comparando-as e comentando sua utilização, especialmente na área oncológica. São ainda apresentados e discutidos os tipos de estudos epidemiológicos e de fontes de dados sujeitos a este tipo de análise. Faz-se a diferença entre estudos com base hospitalar ou de séries clínicas e estudos de base populacional, além de se discutir a interpretação dos resultados obtidos.*

Palavras-chave *Sobreviventes; Análise de Sobrevivência; Estudos Epidemiológicos*

Análise de sobrevida

Qualquer que seja o tipo de estudo epidemiológico, geralmente há uma variável de interesse, também chamada de variável *dependente* ou *resposta*. Essa variável pode ser o número de casos de determinada doença, ou a sua incidência, ou a sua probabilidade de ocorrência, ou outra medida que vise descrever a frequência com que a doença ocorre. Às vezes, a variável dependente de interesse é o tempo decorrido até o aparecimento de algum evento, e aí se incluem os estudos de análise de sobrevida. Há, ainda, uma ou mais variáveis, denominadas *independentes*, *preditoras* ou *covariáveis*, cujo relacionamento com a variável dependente é o objetivo do estudo epidemiológico. E, nesse contexto, a análise quantitativa é imprescindível, pois os modelos estatísticos expressam a variável dependente como uma função matemática conhecida das variáveis independentes. Há, então, o interesse em se verificar o efeito de fatores de risco ou de fatores prognósticos (sejam eles quantitativos ou qualitativos) no tempo de sobrevida de um indivíduo ou de um grupo, bem como definir as probabilidades de sobrevida em diversos momentos no seguimento do grupo. Considera-se *sobrevida*, no presente trabalho, o *tempo* desde a *entrada* do indivíduo no estudo (data do diagnóstico) até a *ocorrência* do evento de interesse (falha) ou até a *censura* (perda por tempo de observação incompleto) na observação (Kleinbaum, 1995).

Cabe ressaltar que, se neste texto é enfocada a análise de sobrevida em oncologia, as técnicas de análise propostas, evidentemente, não se restringem a esta área. Podem e devem ser utilizadas e aplicadas a outros grupos de patologias como doenças cardiovasculares, diabetes, AIDS (Fonseca et al., 1999; Mocroft et al., 1999) etc.; de condutas tais como aleitamento materno (Howard et al., 2000), transplantes (Sudan et al., 2000), aderência a programas (Bishop et al., 1998; Zibners et al., 1999) etc.; ou mesmo em outras áreas de conhecimento. Aliás, os campos de aplicação deste tipo de análise são vastos, podendo ser utilizada na indústria, onde começou a ser desenvolvida, e ainda no comércio, na sociologia, na demografia etc. (Lee, 1992).

Em estudos de sobrevida, as pessoas são acompanhadas por meio da ocorrência de um evento. Esse evento pode ser, por exemplo, o diagnóstico da doença, ou a realização de cirurgia, ou o nascimento. Geralmente, as pessoas são incluídas no estudo em diferentes tempos do ano calendário; porém, na análise, todos os indivíduos têm seu tempo de sobrevi-

vência contado a partir da entrada no estudo (que é considerado como tempo zero). Os inícios são, portanto, truncados à esquerda, ou seja, a observação de cada indivíduo começa a partir de determinado momento, sem levar em conta o que aconteceu no passado (Cox & Oakes, 1984). O evento final corresponde geralmente ao óbito, ou à variação de um parâmetro biológico, ou ainda a um determinado evento que indique a modificação do estado inicial (cura, recorrência, retorno ao trabalho etc.). O evento final é denominado falha (*failure*), por referir-se mais freqüentemente a eventos indesejáveis como o aparecimento de doença ou morte (Kleinbaum, 1995).

Em estudos em que há seguimento, pode ocorrer que alguns indivíduos não sejam observados até a ocorrência da falha, ou seja, tenham seu tempo de observação incompleto. Esse tipo de perda no tempo de observação é denominado censura. Isso pode ocorrer quando os indivíduos permanecem sem mudança de estado ao término do estudo, ou falecem por causas não relacionadas com a doença de interesse, ou abandonam o estudo, ou fogem à observação. Por vezes, a cura e/ou recuperação também podem ser consideradas como censura na observação. Os estudos em que existe censura são denominados *com observações incompletas*. Uma suposição importante é a de que os *indivíduos censurados* em determinado tempo t são *representativos* de todos os indivíduos que estavam sujeitos ao risco de *ter falha* em t (Szklo & Nieto, 2000).

Tipos de estudos em oncologia que utilizam análise de sobrevida

Há dois tipos de estudos que podem utilizar o tempo como variável de interesse. Um deles é o estudo experimental (ensaios clínicos controlados aleatorizados), indicado para avaliar formas de tratamento. Outro tipo são os estudos de coorte observacionais, cujos dados podem ser obtidos pela coleta direta em prontuários médicos ou em bases de dados já existentes (dados secundários). Essas fontes de dados secundários podem ser de base hospitalar (registros hospitalares de câncer) ou populacional (registros de câncer de base populacional).

Registros de base hospitalar são aqueles que se referem a todos os casos tratados e acompanhados em uma instituição. Fornecem informações tanto para a administração do hospital quanto para pesquisadores interessados em informações sobre os resultados do tratamento nos diferentes grupos e fatores de risco ou fa-

tores prognósticos. Contribuem ainda na atenção ao paciente individualmente, uma vez que asseguram o seguimento destes pacientes (Young, 1991).

Já os registros de base populacional (Jensen et al., 1991) coletam dados que se referem a todos os casos de câncer verificados em uma população de uma área geográfica definida, tendo como principal objetivo produzir indicadores como a incidência, a prevalência e a sobrevida para sua área de abrangência. Nas análises de sobrevida de base populacional o período de tempo analisado é, via de regra, aquele decorrido entre o diagnóstico da doença e o óbito. Nesses casos, a sobrevida exprime a probabilidade de que um paciente esteja vivo após um período determinado (por exemplo, após 1, 3, 5 anos do diagnóstico) (Micheli et al., 1992) e refletem, além da história natural da doença, as atividades de controle do câncer, incluindo rastreamento (*screening*) e organização e qualidade dos serviços de saúde (Black et al., 1998).

Embora utilizem os mesmos métodos, as taxas de sobrevida calculadas com base em casos incidentes na população devem ser distinguidas daquela calculada ao se valer de uma série de pacientes hospitalares ou de um ensaio clínico. Os pacientes avaliados em ensaios clínicos e séries hospitalares são submetidos a numerosos critérios de seleção (reconhecidos ou não) e raramente representam a diversidade presente na população de casos (Stiller, 1994). De um modo geral, séries hospitalares apresentam resultados de sobrevida mais otimistas (Sue-Ling et al., 1993). Por outro lado, a sobrevida com base populacional não permite, quase sempre, a avaliação de variáveis como o estadiamento dos tumores e o tratamento, que são importantes para o clínico.

Métodos estatísticos para análise de sobrevida

Na análise de sobrevida, os parâmetros mais importantes são a probabilidade de sobrevivência no curso de cada um dos intervalos considerados e a probabilidade de sobrevida acumulada (tratada correntemente como taxa de sobrevida), isto é, a probabilidade de sobreviver do tempo zero até o tempo final considerado. Esta última equivale à probabilidade de sobreviver em todos os intervalos anteriores ao momento considerado e, usualmente, é denominada $S(t)$. A escolha do modelo estatístico mais apropriado dependerá do tipo do delineamento do estudo epidemiológico, de seus objetivos, das variáveis estudadas e da maneira

pela qual foram coletados e categorizados os dados.

A estimativa da probabilidade de sobrevida é, com certeza, mais válida e mais precisa para o período inicial do seguimento, no qual estão disponíveis informações sobre a maioria dos pacientes. Nos períodos posteriores, as informações podem ficar limitadas devido às perdas de seguimento e ao pequeno número de eventos (Fletcher et al., 1996).

Somente nas décadas de 1950 e de 1960 apareceram as primeiras propostas de estimadores das probabilidades de sobrevida que incorporavam a censura, vale dizer, modelos para observações incompletas. As principais técnicas são o método atuarial e o método do produto-limite de Kaplan-Meier.

Método atuarial ou tábua de vida

O método atuarial para dados incompletos (Lee, 1992; Selvin, 1996) calcula as probabilidades de sobrevida em intervalos fixados previamente, e o número dos expostos a risco corresponde aos pacientes vivos ao início de cada intervalo x . O número de expostos (l_x) é ajustado de acordo com o número de censuras que ocorreram neste período, na suposição de que as censuras ocorreram uniformemente durante o período x e que a experiência subsequente dos casos censurados é a mesma daqueles que permanecem em observação (Kahn & Sempos, 1989). Assim, na presença de censura, é feito um ajuste no número de pessoas expostas ao risco no início do período x , subtraindo-se metade das censuras do total de expostos ao risco no início do período, supondo-se que estes indivíduos estiveram, em média, expostos ao risco apenas metade do intervalo de seguimento (Szklo & Nieto, 2000). Nessa tábua de vida, o tempo também é dividido em intervalos fixos e a probabilidade de óbito (q_x) e a de sobrevida (p_j) são calculadas para cada um dos intervalos. Tem-se então:

$$(q_x = d_x/l_x^*), (p_x = 1 - q_x) \text{ e } l_x^* = l_x - \frac{w_x}{2}, \text{ onde:}$$

q_x = probabilidade condicional de morte no intervalo x

p_x = probabilidade condicional de sobrevida no intervalo x

l_x^* = o número de expostos ao risco, no intervalo x , corrigidos de acordo com a censura

l_x = número de pessoas expostas ao risco no início do período

w_x = número de pessoas perdidas de observação no intervalo x

Uma das fórmulas das funções de sobrevivência é a da probabilidade de sobrevivência acumulada até o tempo x , ou $S(t)$ é dada por:

$$S(t) = \prod_{x=0}^{t-1} p_x^* = \prod_{x=0}^{t-1} (1 - q^*)$$

Há ainda a fórmula da função de riscos (*hazard function*), ou $h(t)$, também conhecida como força instantânea de mortalidade ou taxa instantânea de óbito em um período curto de tempo, dado que um indivíduo estava vivo até o instante $t-1$:

$$h(t) = \frac{d_x}{1_x - \frac{d_x}{2}}, \text{ dado que } T > t_{x-1}$$

Uma demonstração do método atuarial de cálculo dos elementos da sobrevivência pode ser acompanhada na Tabela 1, em que se observa que, no intervalo entre 1 e 2 anos após o diagnóstico (segunda linha), havia 98 pacientes vivos no início do intervalo, ocorreram 46 óbitos e quatro pacientes foram perdidos do seguimento. Assim, 96 ($98 - (4/2)$) estavam expostos ao risco de morrer, resultando na probabilidade de morte de 0,47927 ($46/96$) e, conseqüentemente, na probabilidade de sobreviver de 0,52083 ($1 - 0,47927$). A sobrevivência acumulada a dois anos do diagnóstico é de 0,16676, resultado da multiplicação da sobrevivência nos intervalos do primeiro (0,32018) e do segundo ano (0,52083). Desta forma, tem-se que a probabilidade de um paciente com câncer de estômago estar vivo após o segundo ano do diagnóstico é de 16,7%. Esses resultados podem ainda ser visualizados no Figura 1.

Um esquema diferente é aquele em que o desfecho de interesse pode ser classificado em diversas categorias, geralmente diferentes causas de óbito. Para esse tipo de análise é introduzida a metodologia de riscos competitivos e são modelos que envolvem a teoria de cadeias markovianas (Andersen et al., 1993; Becker, 1989). Essa tábua de vida é conhecida como tábua de vida com múltiplos decrementos. Quando se trabalha com os riscos competitivos os tempos de sobrevivência para cada uma das causas formam um sistema em que cada um deles é uma variável aleatória

Método de Kaplan-Meier

Na análise de sobrevivência pelo método de Kaplan-Meier (Kaplan & Meier, 1958; Lee, 1992; Kleinbaum, 1995) os intervalos de tempo não são fixos, mas determinados pelo aparecimento de uma falha (por exemplo, o óbito). Nessa situação, o número de óbitos em cada intervalo deve ser um. Esse é um método não paramétrico, ou seja, que independe da distribuição de probabilidade (Colton, 1979), e para calcular os estimadores, primeiramente, deve-se ordenar os tempos de sobrevivência em ordem crescente ($t_1 \leq t_2 \leq \dots \leq t_n$). Os sobreviventes ao tempo t (l_t) são ajustados pela censura, ou seja, os pacientes censurados entram no cálculo da função de probabilidade de sobrevivência acumulada até o momento de serem considerados como perda. Isto propicia o uso mais eficiente das informações disponíveis (Szklo & Nieto, 2000).

Define-se a função $S(t)$ por um estimador conhecido como *estimador produto limite de*

Tabela 1

Cálculo da probabilidade de sobrevivência observada pelo método atuarial (casos incidentes de câncer de estômago em Campinas, São Paulo, entre 1991 e 1994).

Intervalo (anos) (i)	Vivo ao início do intervalo (l_i)	Número de óbitos no intervalo (d_i)	Número de pessoas perdidas de observação no intervalo (w_i)	Efetivo exposto ao risco de morrer (l_i^*)	Probabilidade condicional de morte (q_i)	Probabilidade condicional de sobrevivência ($p_{x,i}^*$)	Probabilidade acumulada de sobrevivência ao final do intervalo $S(t_x)$
0-1	337	224	15	329,5	0,67982	0,32018	0,32018
1-2	98	46	4	96	0,47927	0,52083	0,16676
2-3	48	13	0	48	0,27083	0,72917	0,12160
3-4	35	6	2	34	0,17647	0,82353	0,10014
4-5	27	4	2	26	0,15385	0,84615	0,08473

Onde: ($l_x^* = l_x - \frac{w_x}{2}$), ($q_i = d_i/l_i^*$) e ($p_j = 1 - q_j$)

Kaplan-Meier, pois é o limite do produto dos termos até o tempo t :

$$S(t) = \prod_{i=0}^j \frac{I_j - i}{I_j}, \text{ onde } \begin{cases} i = 1, & \text{se for falha} \\ i = 0, & \text{se for censura} \end{cases}$$

e I_j = número de expostos ao risco no início do período.

No caso de haver empate, utiliza-se na fórmula o maior valor de i . Por exemplo:

$$\text{se } t_2 = t_3 = t_4 \Rightarrow p(t_2) = p(t_3) = p(t_4) = \frac{I_2 - 4}{I_2}$$

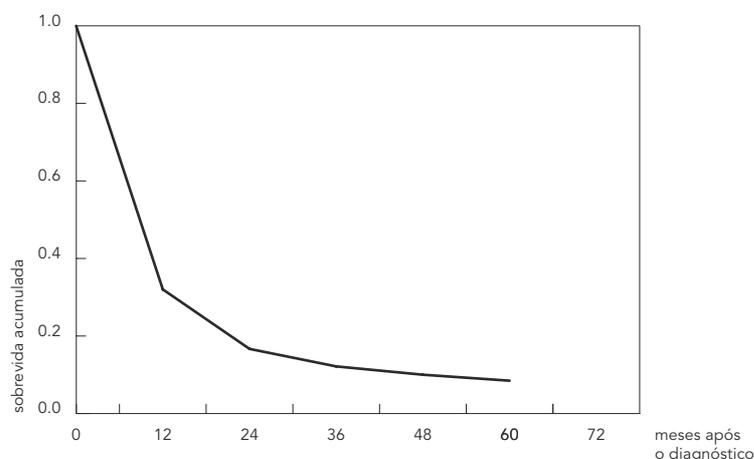
Uma demonstração do cálculo da taxa de sobrevivida observada pelo método de Kaplan-Meier pode ser vista na Tabela 2. Observando a linha referente a 11 meses após o diagnóstico (oitava linha), verifica-se a ocorrência de 1 óbito entre os 12 pacientes vivos no início do mês. Assim, a probabilidade condicional de morte foi de 0,08333 (1/12) e a probabilidade condicional de sobrevivida de 0,91667 (1 - 0,08333) naquele mês, resultando numa sobrevivida acumulada aos 11 meses de 0,42780, resultado da multiplicação das probabilidades de sobrevivida dos períodos anteriores. Desta forma, deduz-se que, para os pacientes com câncer de estômago do sexo masculino e com idade entre 45 e 54 anos, diagnosticados em Campinas, São Paulo, entre 1991 e 1994, a probabilidade de sobrevivida após 11 meses foi de 42,8%. A Figura 2 apresenta esta curva.

A distinção essencial entre o método atuarial e o método de Kaplan-Meier é que este último elimina a necessidade de assumir que as censuras das observações ocorram uniformemente durante este intervalo. Assume-se apenas que as observações censuradas teriam a mesma experiência futura do que aquelas que continuam sendo observadas (Kahn & Sempos, 1989). O método de Kaplan-Meier pode ser utilizado para qualquer tamanho de estudo, mas é especialmente útil naqueles estudos com um número pequeno de observações, enquanto o método atuarial é mais apropriado para grandes amostras (Lee, 1992).

Os dois métodos referidos assumem como premissa que as observações censuradas têm a mesma probabilidade de sofrerem o evento que aquelas que permanecem em observação, isto é, as censuras devem ser independentes da sobrevivida. Nos estudos que contemplam períodos extensos de observação é necessário assegurar que não tenham existido mudanças importantes nas características destes indivíduos e no diagnóstico, exposição ou tratamento da doença em estudo ao longo deste período. Tais mudanças poderiam introduzir viés nas esti-

Figura 1

Probabilidade de sobrevivida acumulada para câncer de estômago, em Campinas, São Paulo, diagnosticados entre 1991-1994 (método atuarial).



mativas de sobrevivida, cuja direção depende das características da coorte e do período estudado (Kahn & Sempos, 1989; Szklo & Nieto, 2000).

A aplicação desses modelos permite comparar o conjunto de curvas de sobrevivida das diversas categorias de uma única variável independente. Para comparar as curvas de sobrevivida acumulada entre diferentes categorias de uma mesma variável, recomenda-se utilizar o teste *log-rank* (Cox & Oakes, 1984; Kleinbaum, 1995), que se baseia no confronto entre os óbitos observados nos dois grupos e aqueles esperados. A diferença entre óbitos observados e esperados é avaliada por meio do teste do qui-quadrado.

Métodos de cálculo para estimar a variância e os intervalos de confiança da probabilidade de sobrevivida estão disponíveis e são bem descritos por Kleinbaum (1995), Lee (1992), Parkin & Hakulinen (1991), Selvin (1996), e Szklo & Nieto (2000). Esta estimativa enfatiza o tamanho do efeito, indica a faixa de valores plausíveis para a sobrevivida e fornece informações sobre poder estatístico.

Há controvérsias sobre como tratar os óbitos por outra causa que não a doença de interesse ou os óbitos por causa desconhecida. Há autores que analisam estes pacientes como falha e, neste caso, a taxa de sobrevivida reflete a mortalidade geral para este grupo de pacientes (sobrevivida global). A Agência Internacional de Pesquisa em Câncer (*International Agency for*

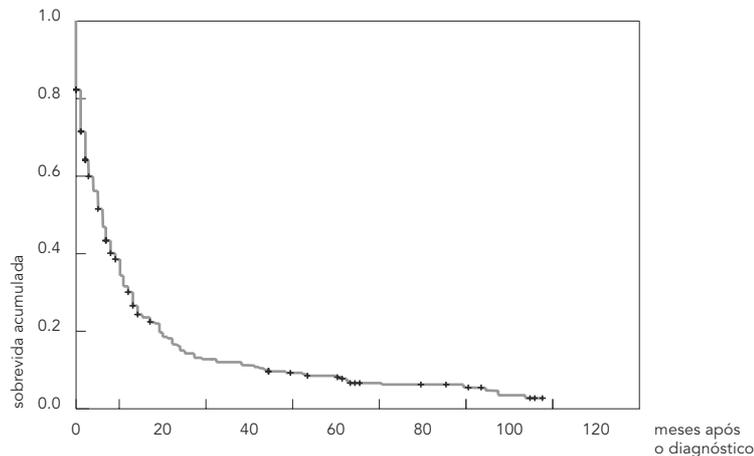
Tabela 2

Cálculo da probabilidade de sobrevivida observada pelo método de Kaplan-Meier (casos incidentes de câncer de estômago em homens com idade entre 45 e 54 anos, Campinas, São Paulo, entre 1991 e 1994).

Meses após o diagnóstico (i)	Número de vivos no início do mês (l_i)	Óbitos (d_i)	Censuras (w_i)	Probabilidade condicional de morte ($q_i = d_i/l_i$)	Probabilidade condicional de sobrevivida ($p_i = 1 - q_i$)	Probabilidade acumulada de sobrevivida $S(t_i)$
0	29	5	0	0,17241	0,82759	0,82759
1	24	2	0	0,08333	0,91667	0,75863
2	22	2	1	0,09091	0,90909	0,68966
3	19	2	0	0,10526	0,89474	0,61707
5	17	1	0	0,05882	0,94118	0,58077
6	16	1	1	0,06250	0,93750	0,54447
10	14	2	0	0,14286	0,85714	0,46669
11	12	1	0	0,08333	0,91667	0,42780
13	11	3	0	0,27273	0,72727	0,31113
14	8	1	0	0,12500	0,87500	0,27224
18	7	1	0	0,14286	0,85714	0,23335
19	6	1	0	0,16667	0,83333	0,19446
38	5	1	0	0,20000	0,80000	0,15557
48	4	1	0	0,25000	0,75000	0,11668
53	3	1	0	0,33333	0,66667	0,07779

Figura 2

Probabilidade de sobrevivida acumulada para câncer de estômago em Campinas, São Paulo, diagnosticados no período 1991-1994 (método de Kaplan-Meier).



Research on Cancer – IARC), em suas publicações (Berrino et al., 1995; Black & Swaminathan, 1998; Estevè et al., 1994; Parkin & Hakulinen, 1991), utiliza esta última opção, sugerindo que, quando as informações sobre a causa de óbito são confiáveis, se podem também obter probabilidades de sobrevivida corrigidas, estas sim, considerando como falha apenas os óbitos pela doença em questão. Vale lembrar que o objetivo maior é o de descrever a mortalidade atribuída à doença em estudo e com esta finalidade podemos utilizar as *taxas de sobrevivida corrigidas*, que são calculadas da mesma maneira (método atuarial ou Kaplan-Meier); contudo, consideram como óbito apenas aqueles atribuídos à doença em estudo. Neste caso os pacientes que morreram por outra causa ou por causa desconhecida devem ser tratados como pacientes com tempo de observação incompleto (censura).

Sobrevivida relativa

Nos registros de base populacional, o objetivo maior do seguimento é definir o estado vital (*status*) do paciente e não a causa do óbito. Assim, as taxas de sobrevivida observadas com base populacional podem não refletir a real força de mortalidade pela doença em questão, já que

a força da mortalidade por outras causas a que estão submetidos os pacientes com câncer tendem a reduzir a taxa de sobrevivida em relação àquela que ocorreria se a doença do paciente fosse a única causa de morte considerada. A doença pode ter o mesmo efeito biológico sobre a mortalidade dos pacientes em todos os grupos de idade, mas a mortalidade devida a outras causas é mais alta nos grupos de faixa etária mais alta. Assim, por razões não necessariamente ligadas à doença dos pacientes, as taxas de sobrevivida observadas nos pacientes mais velhos e mais jovens não são comparáveis.

Além disto, a informação sobre a causa de morte dos pacientes muitas vezes é vaga, não disponível ou pouco confiável. A distinção entre morte pela doença, morte devida à toxicidade do tratamento e morte devida à causa independente muitas vezes é difícil, particularmente em doenças crônicas como as neoplasias.

Em vista destes dois problemas é preferível considerar na análise as mortes por todas as causas e corrigi-las, utilizando a sobrevivida relativa (*relative survival rate* – RSR) (Berkson & Cage, 1950; Estevè et al., 1994; Verdecchia et al., 1995), que é definida como a razão entre a sobrevivida observada no grupo de pacientes ($S(t)$, baseada em tábuas de vida para dados incompletos) e aquela esperada na população a que estes pacientes pertencem, que são, na realidade, as estimadas tomando-se por base a tábua de vida atuarial para dados completos (que estima a esperança de vida) (Ederer et al., 1961; Hakulinen, 1982). A fórmula, a seguir, define a RSR:

$$RSR = \frac{\text{taxa de sobrevivida observada}}{\text{taxa de sobrevivida esperada}} = \frac{S(t, \text{grupo em estudo})}{S(t, \text{tábua de vida atuarial para dados completos})} \times 100$$

As variáveis usualmente levadas em consideração são as demográficas, como sexo, idade e ano de calendário para uma área geográfica definida. A sobrevivida relativa é, portanto, uma estimativa de como seria a sobrevivida dos pacientes com câncer na ausência de outras causas de morte que não o câncer.

Para o cálculo da taxa de RSR, é fundamental que as tábuas de vida da população sejam fidedignas e sincronizadas com o período da análise (Verdecchia et al., 1995).

Uma demonstração do método para o cálculo da taxa de sobrevivida relativa se encontra na Tabela 3. Observa-se que, ao final do primeiro ano, a sobrevivida relativa dos pacientes com câncer de estômago em Campinas, São Paulo, foi de 33,4% (0,32018/0,95800).

Sobrevivida relativa padronizada por idade

Assim como a maioria dos fenômenos biológicos, as taxas de sobrevivida são afetadas pela idade. A sobrevivida relativa, como leva em consideração a idade e a probabilidade de morrer por todas as causas, ajusta por idade somente em parte. Desta maneira, quando se comparam populações com diferentes estruturas etárias, seria recomendado restringir-se às comparações dentro dos mesmos grupos etários. Entretanto, na prática, o que se recomenda é utilizar a sobrevivida padronizada por idade. Há duas populações-padrão propostas, ambas baseadas na localização do tumor. Uma é a apresentada pelo estudo *European Cancer Registry-based Study Cancer Patients' Survival and Care* (EUROCARE), que representa os dados dos registros europeus de câncer (Berrino et al., 1995), e a outra propõe uma população mundial padrão de pacientes com câncer construída com

Tabela 3

Taxas de sobrevivida observada, esperada e relativa de câncer de estômago em Campinas, São Paulo (casos incidentes entre 1991 e 1994).

Tempo após o diagnóstico (em anos)	Número de pessoas vivas no início do período	Sobrevivida observada ^a	Sobrevivida esperada ^b	Sobrevivida relativa
0-1	337	0,32018	0,95800	0,33422
1-2	98	0,16676	0,95060	0,17543
2-3	48	0,12160	0,94222	0,12906
3-4	35	0,10014	0,94222	0,10628
4-5	27	0,08473	0,93057	0,09105

^a Probabilidade de sobrevivida atuarial por câncer de estômago em Campinas, São Paulo.

^b Tábua de vida de Campinas, São Paulo.

base nas estimativas das taxas de incidência mundiais (Black & Bashir, 1998) (Tabela 4 e 5).

A padronização de taxas é feita multiplicando-se as taxas encontradas em cada faixa etária, pela população-padrão (relativa ou absoluta) daquela mesma faixa etária e, ao final, a taxa padronizada corresponde à soma dos valores encontrados para cada grupo de idade (dividido por cem ou pelo total da população-padrão). Pode-se observar na Tabela 6, que tomando as taxas de sobrevida relativa por câncer de estômago em Campinas após um ano do diagnóstico (2ª coluna), e multiplicando-se pela população-padrão européia (3ª coluna), obtêm-se os valores especificados na 4ª coluna, que somados ($92,4 + 391,8 + 646,4 + 1172,3 + 867,2 = 3170,1$) e posteriormente divididos por 100, resultam em 31,7% que é o valor da taxa de sobrevida relativa padronizada por idade após um ano do diagnóstico, dos pacientes com câncer de estômago do Município de Campinas.

Modelo de Cox

A análise de regressão múltipla também pode ser feita na análise de sobrevida, quando se deseja avaliar o efeito conjunto de algumas variáveis independentes, sejam as observações incompletas ou não. Os primeiros modelos de regressão para análise de sobrevida foram desenvolvidos na década de 1960 (Harris & Albert, 1991) e eram totalmente paramétricos, ou seja, baseados nas premissas de validade da estatística tradicional. Em 1972, Cox desenvolveu um modelo de regressão semi-paramétrico, também conhecido como *modelo de riscos proporcionais de Cox*, *modelo de Cox*, ou *regressão de Cox* (Cox, 1972). Essa técnica é indicada quando se deseja estudar sobrevivência sob o prisma de causalidade ou da predição, pois fornece as estimativas das razões de risco dos fatores estudados, podendo-se avaliar o impacto que alguns fatores de risco ou fatores prognósticos têm no tempo até a ocorrência do evento de interesse. A função de riscos (*hazard function* - $h(t)$), no modelo de Cox (Andersen et al., 1993; Cox, 1972; Cox & Oakes, 1984; Crowley & Breslow, 1994; Harris & Albert, 1991; Lee, 1992; Miller Jr., 1981), é considerada como variável dependente, e os riscos de morte por uma determinada causa são o produto de uma função não especificada de tempo (que é comum a todos os indivíduos) e uma função conhecida (que é a combinação linear das covariáveis X_i , sendo $i = 1, 2, \dots, k$). Nele, a função de riscos $h(t)$ é escrita em termos das covariáveis:

$$h(t / X_1, X_2, \dots, X_k) = h_0(t) \cdot \exp(\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k)$$

onde $h_0(t)$ é a parte não paramétrica do modelo, e, em estudos em que o objetivo é estimar fatores prognósticos, não há interesse em defini-la (pois é comum a todos os indivíduos). Os coeficientes de regressão (β_i) são estimados pelo método da máxima verossimilhança parcial.

Ao se fazer a divisão dos dois lados da equação por $h_0(t)$, obtêm-se:

$$\frac{h(t / X_1, X_2, \dots, X_k)}{h_0(t)} = \exp(\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k)$$

$$\text{O quociente } \frac{h(t / X_1, X_2, \dots, X_k)}{h_0(t)}$$

é chamado de *função das razões de riscos* (Andersen, 1991; Cox & Oakes, 1984) - *HR(i)* - (*relative hazard function*) ou de *índice prognóstico* - *prognostic index* - (Altman & Andersen, 1989).

$$HR(i) = HR_i = \exp(\beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik})$$

Esta fórmula também é útil para estimar a razão entre as funções de riscos (HR) para cada uma das variáveis independentes (X_i), supondo todas as outras $X_{j \neq i}$ como constantes.

$$HR(X_i) = \exp(\beta_i)$$

As suposições feitas são as de que diferentes indivíduos têm funções de riscos proporcionais entre si, e que a razão entre essas funções de risco não varia no tempo. Quando, durante o período de seguimento, a probabilidade de sobrevida de um grupo de indivíduos expostos a determinado fator não for proporcional à dos não expostos, isto é, os riscos não são constantes e proporcionais durante o período, deve-se fazer modificação no modelo que Cox propôs inicialmente. Esse último modelo é conhecido como *modelo de Cox com variável tempo-dependente* (Cox & Oakes, 1984). A descrição detalhada dos modelos de Cox (convencional e com variável tempo-dependente) pode ser visto em Andersen et al. (1993), Cox (1972), Cox & Oakes (1984), Crowley & Breslow (1994), Harris & Albert (1991), Kleinbaum (1995), Latorre (1996), Lee (1992) e Miller Jr. (1981).

Na Tabela 5, são apresentados os resultados da análise múltipla, utilizando o modelo de Cox, também com os dados do Registro de Câncer de Base Populacional (RCBP) de Campinas. Em relação à variável estágio, por exemplo, verifica-se que os estádios metastáticos apresentaram um risco de morrer (HR_{br}) 20,26 vezes maior do que os tumores localizados, com um intervalo de confiança de 2,67-153,68. Quando

Tabela 4

População-padrão européia de pacientes de câncer (percentuais)*.

Localização	Idade					Total
	15-44	45-54	55-64	65-74	75-99	
Casos de câncer em adultos (%)						
Língua	7,4	16,3	26,4	26,6	23,3	100
Cavidade oral	6,0	15,5	26,7	27,5	24,3	100
Orofaringe	6,4	21,0	29,7	26,7	16,2	100
Nasofaringe	17,6	17,5	26,7	24,2	13,9	100
Hipofaringe	4,6	16,4	30,1	28,6	20,4	100
Cabeça e pescoço	7,1	17,0	27,7	27,1	21,1	100
Esôfago	2,2	8,8	22,8	32,9	33,3	100
Estômago	3,1	7,8	18,5	34,5	36,1	100
Cólon	3,5	7,7	18,4	32,4	38,0	100
Reto	3,0	8,5	20,9	34,3	33,3	100
Pâncreas	2,2	7,2	20,3	35,1	35,3	100
Laringe	4,1	16,2	33,1	31,1	15,4	100
Pulmão	1,9	9,2	27,3	38,7	23,0	100
Osso	31,8	11,9	17,1	21,3	17,9	100
Mama	13,1	19,5	23,0	23,8	20,6	100
Colo uterino	28,8	17,3	23,5	19,5	10,8	100
Corpo uterino	3,4	16,2	32,9	28,4	19,1	100
Ovário	10,3	18,2	26,5	26,0	19,0	100
Vagina	4,5	7,1	15,3	30,3	42,8	100
Testículo	78,3	10,8	5,8	3,2	2,0	100
Pênis	8,2	11,1	20,8	33,2	26,7	100
Rim	5,9	13,7	27,1	32,2	21,1	100
Cérebro	24,3	18,4	26,7	23,5	7,1	100
Doença de Hodgkin	54,9	11,0	12,6	13,1	8,4	100
Leucemia linfática aguda	48,4	6,9	12,8	16,9	15,0	100
Leucemia linfática crônica	1,8	7,1	18,9	34,8	37,4	100
Leucemia mielóide aguda	17,4	11,2	17,6	27,9	26,0	100
Leucemia mielóide crônica	16,3	12,5	17,6	26,4	27,2	100
Leucemia	12,7	9,2	17,3	29,4	31,4	100

Localização	Idade				Total
	0-1	2-4	5-9	10-14	
Casos de câncer em crianças (%)					
Cérebro	10,8	22,2	34,2	32,8	100
Leucemia	11,6	36,8	27,9	23,7	100

* adaptado de Verdecchia et al. (1995).

ajustado para as outras variáveis, o risco (HR_{aj}) foi 16,81 com um intervalo de confiança de 2,16-131,09. Não foi observada diferença significativa do risco entre os sexos, faixas etárias, nem entre os anos de diagnóstico.

Interpretação dos dados de sobrevida

A interpretação dos dados de sobrevida e de sua tendência não é uma tarefa fácil, e compa-

rar a sobrevida entre diferentes estudos, sendo estes provenientes de séries clínicas ou de populações definidas geograficamente, se torna uma tarefa ainda mais difícil.

Problemas metodológicos que afetam as estimativas de sobrevida

Segundo Berrino et al. (1997), os problemas metodológicos que afetam as estimativas de sobrevida podem ser classificados em dois grupos:

Tabela 5

População-padrão mundial de pacientes de câncer (percentuais)*.

Localização CID-9	Idade								Total
	0-4	5-9	10-14	15-44	45-54	55-64	65-74	75+	
140-208	0,5	0,6	0,7	14,6	14,3	20,2	20,8	28,3	100
140-149	0,1	0,2	0,4	18,0	19,1	23,1	18,9	20,2	100
150	0,0	0,0	0,0	10,6	16,6	22,7	22,8	27,4	100
151	0,0	0,0	0,0	12,0	15,5	19,4	21,9	31,2	100
153-154	0,0	0,0	0,1	8,6	10,8	17,0	22,3	41,2	100
155	0,1	0,3	0,4	18,6	18,3	22,7	19,2	20,4	100
157	0,0	0,0	0,0	10,3	13,7	17,4	21,8	36,8	100
161	0,0	0,0	0,1	11,3	18,4	26,1	21,6	22,5	100
162	0,0	0,0	0,0	7,6	13,6	23,0	24,2	31,6	100
172	0,0	0,1	0,4	25,3	17,7	18,0	17,0	21,5	100
174	0,0	0,0	0,1	22,2	20,5	19,9	17,2	20,1	100
180	0,0	0,0	0,1	28,3	24,1	20,7	14,5	12,3	100
182	0,0	0,0	0,1	11,1	18,9	27,2	21,6	21,1	100
183	0,2	0,4	0,7	23,4	19,2	21,3	17,2	17,6	100
185	0,0	0,0	0,0	0,6	2,4	9,5	20,8	66,7	100
186	0,6	0,1	0,1	79,8	10,2	4,7	2,4	2,1	100
188	0,0	0,0	0,1	6,9	10,2	17,9	23,3	41,6	100
189	4,5	2,1	1,3	10,7	13,0	21,5	20,7	26,2	100
200-203	2,8	2,7	2,7	23,4	13,2	16,2	16,6	22,4	100
204-208	8,1	6,3	5,3	26,0	10,0	11,8	13,0	19,5	100

* adaptado de Black & Bashir (1998).

CID-9 = Classificação Internacional de Doenças 9ª revisão.

Tabela 6

Padronização por idade das taxas de sobrevida relativa (RSR) a 1 e 5 anos do diagnóstico por câncer de estômago em Campinas, São Paulo, 1991 a 1994, utilizando as populações-padrão européia (Peuro) e mundial (Pmundial).

	RSR		Peuro %	Pmundial		%		
	1 ano	5 anos		1 ano	5 anos		1 ano	5 anos
15-44	30	10	3,08	92,4	30,8	12	360	120
45-54	50	13	7,84	391,8	101,9	15,5	775	201,5
55-64	35	12	18,47	646,4	221,6	19,4	679	232,8
65-74	34	9	34,48	1172,3	310,3	21,9	744,6	197,1
75 e +	24	5	36,13	867,2	180,7	30,2	724,8	151
Todos	32	9						
ASRS*				31,7	8,5		32,8	9,0

* ASRS = taxas de sobrevida relativa padronizadas por idade.

Tabela 7

Modelo de Cox para casos incidentes de câncer de estômago no Município de Campinas, São Paulo, entre 1991 e 1994.

Variável	HR _{br}	IC 95% (HR _b)	p (HR _b)	HR _{aj}	IC 95% (HR _{aj})	p (HR _{aj})
Faixa etária						
15-44	1,00	referência		1,00	referência	
45-54	0,73	0,40-1,32	0,2926	0,90	0,35-2,28	0,8195
55-64	0,90	0,54-1,52	0,7059	1,06	0,43-2,63	0,8935
65-74	1,01	0,61-1,68	0,9568	1,40	0,63-3,15	0,4119
>75	1,35	0,81-2,23	0,2520	1,41	0,58-3,42	0,4424
Sexo						
Masculino	1,00	referência		1,00	referência	
Feminino	0,97	0,74-1,25	0,7897	1,16	0,73-1,85	0,5410
Ano						
1994	1,00	referência		1,00	referência	
1993	1,12	0,76-1,66	0,5646	1,33	0,68-2,60	0,4062
1992	1,13	0,78-1,63	0,5083	1,16	0,63-2,14	0,9979
1991	1,03	0,71-1,48	0,8918	1,22	0,53-2,85 0,6384	0,6399
Estádio						
Localizado	1,00	referência		1,00	referência	
Regional	9,77	1,31-73,02	0,0264	7,99	1,02-63,00	0,0483
Metastático	20,26	2,67-153,68	0,0036	16,81	2,16-131,09	0,0071

(1) fatores que podem ser controlados na análise (ao menos parcialmente) e (2) fatores que dependem da validade dos dados.

Alguns fatores que podem ser controlados na análise

- Mortalidade por outras causas: pode ser controlada utilizando-se a sobrevida corrigida e/ou a sobrevida relativa.
- Fatores demográficos, como idade ao diagnóstico, sexo dos pacientes e características sócio-econômicas: recomenda-se utilizar a sobrevida padronizada para idade ou incorporar estas variáveis na análise da regressão múltipla de Cox.
- Época do diagnóstico, centro de tratamento, recursos diagnósticos e outras variáveis relacionadas aos cuidados de saúde: se disponíveis, devem ser incorporadas ao modelo de regressão múltipla de Cox ou realizar a análise estratificada (Kaplan-Meier ou sobrevida relativa).
- Estadiamento, morfologia e outras variáveis relativas à doença: se disponíveis, devem ser incorporados ao modelo de regressão múltipla de Cox ou realizar a análise estratificada (Kaplan-Meier ou sobrevida relativa).

- Diferentes metodologias estatísticas: deve-se comparar taxas de sobrevida estimadas pelo mesmo método.

Fatores que dependem da validade dos dados

Dentre os fatores que dependem da validade dos dados e que devem ser levados em conta na sua interpretação, destacam-se:

- **Definição da doença e sua evolução no tempo**

Quando as probabilidades de sobrevida são comparadas entre diferentes populações ou em períodos de tempo sucessivos, é fundamental que a definição da doença seja a mesma em todos os grupos. Assim, quando se estudam as neoplasias malignas (Códigos 140 a 208 da Classificação Internacional de Doenças 9ª revisão (OMS, 1985) e códigos C 00 a C 97 da CID 10ª revisão; OMS, 1995), devem ser cuidadosamente excluídos os tumores classificados como benignos, incertos se benignos ou malignos e aqueles *in situ*. Para alguns tipos de tumor esta diferenciação é difícil e pode impossibili-

tar as comparações da sobrevida. Os mielomas múltiplos e os cânceres de bexiga e próstata são um exemplo bastante ilustrativo, e para estudos de sobrevida nestes tipos tumorais é fundamental o estabelecimento de critérios padronizados para a inclusão dos pacientes nas séries a serem comparadas (Berrino et al., 1995).

Um outro fator que pode influenciar e distorcer as comparações de sobrevida é a definição da sede do tumor. Analisar, por exemplo, o câncer de faringe, tomando em conjunto orofaringe (CID-9 146), nasofaringe (CID-9 147), hipofaringe (CID-9 148) e ainda faringe e localizações mal definidas do lábio, da boca e da faringe (CID-9 149), mistura diferentes entidades clínicas com etiologias e prognósticos diferentes. Desta forma, diferenças de sobrevida observada entre populações diversas podem, na realidade, refletir apenas diferentes incidências destas localizações. Isto ocorre ainda dentro de uma mesma localização, tornando importante que a localização da sede tumoral seja bem definida e que especifique a sub-localização do tumor, utilizando o quarto dígito proposto pelo CID.

A morfologia do tumor é também um fator muito importante, pois sabe-se que entre os vários tipos histológicos observam-se comportamentos muito variáveis tanto nas neoplasias diagnosticadas em um mesmo estadiamento, quanto naquelas que afetam um mesmo órgão.

O estadiamento da doença ao diagnóstico é geralmente o principal fator determinante da sobrevida dos pacientes. Isto ocorre porque certos tratamentos são disponíveis apenas para estágios iniciais da doença e porque os tratamentos disponíveis são mais eficazes se iniciados antes que as metástases tenham ocorrido. No entanto, estes dados de estadiamento nem sempre estão disponíveis, especialmente quando se analisam os dados de registros de base populacional; além disto, utilizam várias classificações, o que pode comprometer a comparabilidade (MacLennan, 1991). Mesmo quando estes dados estão disponíveis, variações na disponibilidade de tecnologias diagnósticas podem gerar erros nesta classificação.

- **Sensibilidade (cobertura) e especificidade (qualidade) do registro**

A qualidade dos dados utilizados é um fator determinante na validade das comparações de probabilidades de sobrevida. Comparados aos estudos baseados em dados hospitalares, a maior vantagem dos estudos com base populacional é o de evitar viés de seleção dos pacientes, uma vez que contemplaria todos os casos

incidentes naquela região geográfica. Estes estudos constituem-se, portanto, nos mais indicados para comparações internacionais de sobrevida. Neste sentido, a cobertura destes registros é fundamental, pois se a captação de casos incidentes não é completa pode também gerar viés de seleção.

Algumas vezes são registrados casos somente mediante a declaração de óbito e são chamados de DCO (*death certificate only*). Eles são incluídos como casos incidentes, mas não entram nas análises de sobrevida, pois a data do diagnóstico da doença é desconhecida. Se a proporção de casos de DCO é relativamente baixa, ou seja, menor do que 10%, esta exclusão não altera muito os resultados finais da estimativa de sobrevida. Porém, quando há um percentual alto de casos de DCO, sua exclusão retira das estimativas, talvez, aqueles de pior prognóstico, o que pode levar a uma sobrestimativa da sobrevida, em virtude deste viés de seleção. Alguns casos podem ser identificados no registro de óbitos; todavia, outros dados referentes ao caso podem ser obtidos em outras fontes como os prontuários médicos, por exemplo. Estes casos, tratados como DCI (*death certificate initiated*), isto é, identificados com base na declaração de óbito, entram nas estatísticas de incidência e de sobrevida. A inclusão de muitos destes casos, ditos DCI, subestimarão os resultados da análise de sobrevida, visto que estes casos podem representar aqueles com pior prognóstico. A avaliação destes percentuais de casos registrados nas declarações de óbito constituem-se num bom indicador da qualidade do registro, e quanto menores os valores encontrados, melhor a qualidade.

- **Qualidade do seguimento**

Para permitirem estimar taxas de sobrevida, os casos incluídos no estudo ou registrados devem ser seguidos para avaliar se os pacientes estão vivos ou mortos. Nos ensaios clínicos, o seguimento destes pacientes ao longo do estudo está implícito no desenho, e, em geral, o tempo de acompanhamento é relativamente curto, gerando perdas de seguimento mínimas e possibilitando, portanto, resultados bastante confiáveis. Nos registros de câncer, o método adotado para o seguimento dos pacientes pode ser classificado como ativo ou passivo. O seguimento ativo implica uma comparação sistemática dos casos registrados com outros bancos de dados da população e contato direto com os hospitais, médicos ou parentes dos pacientes. O seguimento passivo baseia-se em grande parte na notificação dos óbitos pelas autoridades

des administrativas ou sistemas de mortalidade (Parkin & Hakulinen, 1991). No estudo EU-ROCARE (Berrino et al., 1995), que envolveu trinta registros de vários países europeus, 19 realizavam seguimento ativo, seis passivo, e cinco adotavam uma forma mista de seguimento, a saber, faziam o seguimento passivo complementado por uma investigação ativa do estado vital do paciente, caso a declaração de óbito não chegasse depois de um dado período. Swaminathan et al. (1998), analisando a sobrevida com base nos registros de câncer de dez países em desenvolvimento, referem que em seis deles o seguimento era predominantemente ativo e nos quatro outros predominantemente passivo.

Os registros com seguimento passivo assumem que um indivíduo esteja vivo até que uma notificação de morte seja encontrada. Logo, perdas no seguimento podem levar a superestimar as taxas reais de sobrevida porque os casos considerados perdidos no seguimento podem estar mortos. Estas perdas poderiam também ser devido à emigração destes casos e, nestas condições, não há como avaliar a direção deste erro.

Quando se adota o seguimento ativo, sempre há um número de pacientes não localizados. Para estes pacientes, considerados como perdidos de seguimento (censuras), o tempo de sobrevida será considerado até a data de seu último contato. Entretanto, é provável que a maioria deles esteja viva, pois se tivessem ido a óbito seu atestado de óbito teria chegado ao registro. Assim, as taxas de sobrevida encontradas tenderiam a subestimar as taxas verdadeiras.

Vale lembrar que nas estimativas das probabilidades de sobrevida está implícito assumir que os casos censurados têm a mesma probabilidade de morte do que aqueles que permanecem em observação. Assim, é necessário que as censuras sejam independentes da sobrevida, caso contrário teríamos um viés em razão das perdas diferenciais no seguimento (Szklo & Nieto, 2000).

- **Definição da data de diagnóstico**

A data de diagnóstico é a data inicial a partir da qual é calculada a sobrevida. Nos registros de base populacional, esta data da incidência pode ser a do primeiro diagnóstico clínico, a da primeira consulta ou admissão no hospital, ou a da primeira confirmação histológica (MacLennan, 1991). Para os registros hospitalares, a data de admissão no hospital é a que deve ser usada. Nos ensaios clínicos, em que os resultados do tratamento são comparados, deve-se usar a data de início no protocolo (Peto et al., 1976, 1977).

O intervalo de tempo entre estas datas pode variar de poucos dias a poucos meses, mas é improvável que ele possa afetar as estimativas e comparações de sobrevida de longa duração.

- **Relacionados ao cuidado de saúde**

A sobrevida é influenciada pela disponibilidade de serviços de saúde de várias maneiras. O acesso aos serviços, os programas de rastreamento (*screening*), a disponibilidade de meios diagnósticos e de tratamento interferem na história natural da doença, resultando em variações da sobrevida, que pode aumentar ou diminuir de acordo com a eficácia destas intervenções. A interpretação destes achados deve ser cuidadosa, pois podem refletir mudanças que, na realidade, não contribuem para o aumento real do tempo de vida destes pacientes.

O diagnóstico precoce pode ser fruto de uma identificação clínica propiciada pela maior atenção dada a estes sintomas pelo médico e mesmo pelo paciente, ou a um diagnóstico pré-clínico feito com programas de rastreamento ou mesmo com a descoberta ocasional no curso de uma avaliação clínica com outras finalidades. Esta antecipação diagnóstica pode ou não significar um aumento real da sobrevida, dependendo dos recursos terapêuticos disponíveis e de seus melhores resultados nos tumores em estágios iniciais. É importante ter isto em mente, pois diagnósticos precoces em doenças que não dispõem de tratamentos efetivos podem trazer desvantagens para o paciente.

Os programas de rastreamento (*screening*) objetivam identificar os tumores em estágios iniciais ou mesmo pré-malignos propiciando o tratamento da doença ainda em sua fase inicial, o que o torna, em diversas situações, mais efetivo. No entanto, interpretar as estatísticas de sobrevida em termos de benefícios para o paciente diagnosticado no rastreamento é problemático, uma vez que o rastreamento antecipa o diagnóstico, podendo ou não reduzir o risco de morte pela doença em questão. A simples antecipação do diagnóstico, considerando igual a eficácia da intervenção terapêutica, alonga apenas aparentemente a sobrevida: a morte do paciente não é adiada, mas o período de observação de sua doença é mais longo. Esta antecipação do diagnóstico gera o que é chamado de viés de tempo de avanço (*lead-time bias*). Tabàr et al. (1992), analisando os dados da Suécia relativos ao programa de rastreamento de câncer de mama, afirmam encontrar uma redução substancial na taxa de mortalidade das mulheres diagnosticadas pelo rastreamento, e atribuem este aumento da sobrevida a dois fato-

res: aumento do tempo de seguimento pela antecipação diagnóstica e adiamento da morte. Há, porém, outras análises em que o efeito do rastreamento mostra um aumento da sobrevida atribuído apenas a um aumento do tempo de seguimento, sem efeito na duração da vida, como mostrado nos resultados do estudo canadense de rastreamento de câncer de mama (Miller et al., 1992). Ressalte-se que parte dos pacientes diagnosticados e tratados em fases precoces de qualquer tipo de tumor poderiam não desenvolver sintomatologia por muitos anos e até durante toda a sua vida, embora sem tratamento.

Desta forma, precisa-se adotar uma visão crítica em relação aos ditos sucessos terapêuticos e ao crescimento da sobrevida a médio e a longo prazo observada, atualmente, em muitos trabalhos com pacientes oncológicos.

Os avanços nas técnicas de diagnóstico e a disponibilidade de meios diagnósticos podem afetar as estimativas de sobrevida, pois o mesmo câncer pode ser diagnosticado como metastático em um hospital e localizado em outro, dependendo do equipamento diagnóstico disponível para identificar metástases não observadas clinicamente. Este fenômeno, tratado como "fenômeno de migração entre estágios" (*stage-migration phenomenon*) (Feinstein et al., 1985), no qual casos submetidos à análise de sobrevida "migram" de um estágio localizado para o avançado graças a meios diagnósticos mais potentes, e aparentemente melhoram o prognóstico dos casos nas duas categorias. Este fenômeno pode ocorrer em relação a grupos populacionais diversos e também em relação a variações temporais, uma vez que a incorporação de avanços tecnológicos podem ocorrer rapidamente no decorrer dos anos. Portanto, comparações de sobrevida considerando o estágio da doença ao diagnóstico, entre populações com diferentes disponibilidades de meios diagnósticos, podem ser prejudicadas. Para proceder a análises ajustadas por estágio ou análises específicas para cada estágio, é necessário garantir que as categorias utilizadas sejam comparáveis e, ainda, que tenham sido definidas com base nos mesmos procedimentos diagnósticos. Análises que comparam probabilidades de sobrevida entre diferentes grupos, incluindo pacientes com tumor de todos os estágios, não estão sujeitos a este problema.

Vários estudos mostram uma sobrevida maior nos pacientes participantes de ensaios clínicos ou tratados em hospitais especializados em câncer em relação a pacientes tratados em serviços não especializados (Gillis & Hole, 1996; Selby et al., 1996; Stiller, 1994). Estas diferenças podem ser por causa da melhor capacitação dos

centros especializados na definição do diagnóstico e na indicação da terapia mais adequada, ou ser em virtude do viés de seleção, considerando que os pacientes em estágios mais avançados da doença ou que estão sujeitos a maiores fatores de co-morbidade podem não ser referenciados para serviços especializados e tendem a ser excluídos dos ensaios clínicos (Berrino et al., 1997). Desta forma é necessário muita cautela ao se compararem estatísticas de sobrevida provenientes de instituições especializadas ou gerais.

Considerações finais

Cabe uma reflexão sobre as medidas de ocorrência das doenças utilizadas em epidemiologia, tais como a incidência, a mortalidade e a probabilidade de sobrevida. A análise da tendência dos coeficientes padronizados de mortalidade por todos os cânceres, segundo Bailar III & Smith (1986), constitui-se no melhor indicador para avaliar os resultados dos programas de controle de câncer. A incidência reflete com mais acurácia o risco de doença e permite maior controle do impacto das intervenções e de seus efeitos no controle dos fatores de risco. As análises de tendência da incidência e as comparações geográficas permitem mapear riscos e estabelecer hipóteses etiológicas. Auxiliam ainda no planejamento de programas de prevenção e tratamento. As probabilidades de sobrevida, por sua vez, refletem e avaliam os avanços diagnósticos e terapêuticos e, quando desenvolvida com base populacional, contribuem na descrição do comportamento da doença e dos fatores prognósticos a ela relacionados. As taxas de sobrevida podem ainda ser utilizadas para estimar a eficiência global do sistema de saúde, que depende não só da qualidade do cuidado prestado, mas também da acessibilidade ao sistema de saúde e, conseqüentemente, à probabilidade de um diagnóstico e tratamento precoces (Estève et al., 1994).

Cabe ainda refletir sobre a utilidade destes indicadores no âmbito da atenção ao indivíduo. Os estudos etiológicos identificam fatores de risco e são úteis para esclarecer e incentivar hábitos saudáveis que contribuam para a queda da incidência do câncer. Os estudos terapêuticos, prognósticos e de sobrevida permitem ao profissional de saúde e ao paciente oncológico conhecer o comportamento desta doença e obter respostas para as dúvidas que os mobilizam, possibilitando uma abordagem realista e que proporcione maior qualidade de vida aos indivíduos acometidos por esta doença.

Agradecimentos

Os autores agradecem a Gulnar A. S. Mendonça, a José Maria P. Souza, a Maria Lúcia Penna as valiosas sugestões. A Dra. Bustamante-Teixeira agradece a A. Verdecchia, R. Capocaccia, A. Mariotto, G. de Angelis e D. Grecco, a calorosa acolhida e assistência no período em que permaneceu no Istituto Superiore di Sanità – Roma. A Dra. Latorre agradece ao Conselho Nacional de Desenvolvimento Científico e Tecnológico a bolsa de pesquisador (processo nº 30.0318-97/9).

Referências

- ALTMAN, D. G. & ANDERSEN, P. K., 1989. Bootstrap investigation of the stability of a Cox regression model. *Statistics in Medicine*, 8:771-783.
- ANDERSEN, P. K., 1991. Survival analysis 1982-1991: The second decade of the proportional hazards regression model. *Statistics in Medicine*, 10:1931-1944.
- ANDERSEN, P. K.; BORGAN, O.; GILL, R. D. & KEIDING, N., 1993. *Statistical Models Based on Counting Process*. New York: Springer-Verlag.
- BAILAR III, J. C. & SMITH, E. M., 1986. Progress against cancer? *New England Journal of Medicine*, 314:1226-1232.
- BECKER, N. G., 1989. *Analysis of Infectious Disease Data*. London: Chapman and Hall.
- BERKSON, J. & CAGE, R. P., 1950. Calculation of survival rates for cancer. *Proceedings of the Staff Meeting of the Mayo Clinic*, 25:270-286.
- BERRINO, F.; SANT, M.; VERDECCHIA, A.; CAPOCACCIA, R.; HAKULINEN, T. & ESTÈVE, J. (ed.), 1995. *Survival of cancer patients in Europe: The EURO-CARE study*. IARC Scientific Publications 132. Lyon: International Agency for Research on Cancer.
- BERRINO, F.; MICHELI, A.; SANT, M. & CAPOCACCIA, R., 1997. Interpreting survival differences and trends. *Tumori*, 83:9-16.
- BLACK, R. J. & BASHIR, B. A., 1998. World standard cancer patient populations: A resource for comparative analysis of survival data. In: *Cancer Survival in Developing Countries* (R. Sankaranarayanan, R. J. Black & D. M. Parkin, ed.), IARC Scientific Publications 145. Lyon: International Agency for Research on Cancer.
- BLACK, R. J.; SANKARANARAYANAN, R. & PARKIN, D. M., 1998. Interpretation of population-based cancer survival data. In: *Cancer Survival in Developing Countries* (R. Sankaranarayanan, R. J. Black & D. M. Parkin, ed.), IARC Scientific Publications 145. Lyon: International Agency for Research on Cancer.
- BLACK, R. J. & SWAMINATHAN, R., 1998. Statistical methods for the analysis of cancer survival data. In: *Cancer Survival in Developing Countries* (R. Sankaranarayanan, R. J. Black & D. M. Parkin, ed.), IARC Scientific Publications 145. Lyon: International Agency for Research on Cancer.
- COLTON, T., 1979. *Statistica in Medicina*. Padova: Piccin Editore.
- COX, D. R., 1972. Regression models and life tables (with discussion). *Journal of the Royal Statistical Society: Series B*, 34:187-220.
- COX, D. R. & OAKES, D., 1984. *Analysis of Survival Data*. London: Chapman & Hall.
- CROWLEY, J. & BRESLOW, N., 1994. Statistical analysis of survival data. *Annual Review of Public Health*, 5:385-411.
- EDERER, E.; AXTELL, L. M. & CUTLER, S. J., 1961. The relative survival rate: A statistical methodology. *Journal of the National Cancer Institute*, 6:101-121.
- ESTÈVE, J.; BENHAMOU, E. & RAYMOND, L., 1994. Statistical Methods in Cancer Research. In: *Descriptive Epidemiology*, IARC Scientific Publication 128, pp. 213-279, v. 4, Lyon: International Agency for Research on Cancer.
- FEINSTEIN, A. R.; DANIEL, M.; SOSIN, M. D. & WELLS, C. K., 1985. The will Rogers Phenomenon: Stage migration and new diagnostic techniques as a source of misleading statistics for survival cancer. *New England Journal of Medicine*, 312:1604-1608.
- FLETCHER, R. H.; FLETCHER, S. W. & WAGNER, E. H., 1996. *Epidemiologia Clínica: Elementos Essenciais*, 3ª Ed. Porto Alegre: Artes Médicas.
- FONSECA, L. A.; REINGOLD, A. L.; CASSEB, J. R.; BRIGIDO, L. F. & DUARTE, A. J., 1999. AIDS incidence and survival in hospital-based cohort of asymptomatic HIV seropositive patients in São Paulo, Brazil. *International Journal of Epidemiology*, 28:1156-1160.
- GILLIS, C. R. & HOLE, D. J., 1996. Survival outcome of care by specialist surgeons in breast cancer: A study of 3786 patients in the west of Scotland. *BMJ*, 312:145-148.
- HAKULINEN, T., 1982. Cancer survival corrected for heterogeneity in patient withdrawal. *Biometrics*, 38:933-942.
- HARRIS, E. K. & ALBERT, A., 1991. *Survivorship Analysis for Clinical Studies*. New York: Marcel Dekker Incorporation.
- HOWARD, C.; HOWARD, F.; LAWRENCE, R.; ANDRESEN, E.; DEBLIECK, E. & WEITZMAN, M., 2000. Office prenatal formula advertising and its

- effect on breast-feeding patterns. *Obstetrics and Gynecology*, 95:296-303.
- JENSEN, O. M.; PARKIN, D. M.; MACLENNAN, R.; MUIR, C. S. & SKEET, R. G., 1991. *Cancer Registration Principles and Methods*. IARC Scientific Publications 95. Lyon: International Agency for Research on Cancer.
- KAHN, H. A. & SEMPOS, C. T., 1989. *Statistical Methods in Epidemiology*. New York/Oxford: Oxford University Press.
- KAPLAN, E. L. & MEIER, P., 1958. Non parametric estimation from incomplete observation. *Journal of the American Statistics Association*, 53:457-481.
- KLEINBAUM, D. G., 1995. *Survival Analysis: A Self-Learning Text*. New York: Springer.
- LATORRE, M. R. D. O., 1996. *Comparação entre Alguns Métodos Estatísticos em Análise de Sobrevida: Aplicação em uma Coorte de Pacientes com Câncer de Pênis*. Tese de Doutorado, São Paulo: Faculdade de Saúde Pública, Universidade de São Paulo.
- LEE, E. T., 1992. *Statistical Methods for Survival Data Analysis*. 2nd Ed. New York: John Wiley & Sons.
- MACLENNAN, R., 1991. Items of patient information which may be collected by registries. In: *Cancer Registration Principles and Methods* (O. M. Jensen, D. M. Parkin, R. M. Maclennan, C. S. Muir & R. G. Skeet, ed.), IARC Scientific Publications 95, Lyon: International Agency for Research on Cancer.
- MICHELI, A.; BALDASSERONI, A.; BRUZZI, P.; FAGGIANO, F.; GATTA, G.; IVALDI, C.; MAGNANI, C.; MERLETTI, F.; NINU, B. & SANTI, M., 1992. La sopravvivenza per tumori dell'apparato respiratorio: Dati italiani su base di popolazione e confronti internazionali. *Annali dell'Istituto Superiore di Sanità*, 28:71-90.
- MILLER, A.; BAINES, C. J.; TO, T. & WALL, C., 1992. Canadian national breast screening study: 2. Breast cancer detection and death rates among women aged 50 to 59 years. *Canadian Medical Association Journal*, 147:1477-1488.
- MILLER Jr., R. G., 1981. *Survival Analysis*. New York: John Wiley & Sons.
- MOCROFT, A.; MADGE, S.; JOHNSON, A. M.; LAZZARIN, A.; CLUMECK, N.; VIARD, J. P.; GATELL, J.; BLAXHULT, A. & LUNDGREN, J. D., 1999. A comparison of exposure groups in the EuroSIDA study: Starting highly active antiretroviral therapy (HAART), response to HAART, and survival. *Journal of Acquired Immune Deficiency Syndromes*, 22:369-378.
- OMS (Organização Mundial da Saúde), 1985. *Manual da Classificação Estatística Internacional de Doenças, Lesões e Causas de Óbitos - Nona Conferência de Revisão*. São Paulo: Centro Brasileiro de Classificação de Doenças em Português.
- OMS (Organização Mundial da Saúde), 1995. *Classificação Estatística Internacional de Doenças e Problemas Relacionados à Saúde, Décima Revisão*. São Paulo: Centro Brasileiro de Classificação de Doenças em Português.
- PARKIN, D. M. & HAKULINEN, T., 1991. Analysis of survival. In: *Cancer Registration Principles and Methods* (O. M. Jensen, D. M. Parkin, R. Maclennan, C. S. Muir & R. G. Skeet, ed.), IARC Scientific Publications 95, pp. 159-176. Lyon: International Agency for Research on Cancer.
- PETO, R.; PIKE, M. C.; ARMITAGE, P.; BRESLOW, N. E.; COX, D. R.; HOWARD, S. V.; MANTEL, N.; MCPHERSON, K.; PETO, J. & SMITH, P. G., 1976. Design and analysis of randomized clinical trials requiring prolonged observation of each patient. I. Introduction and design. *British Journal of Cancer*, 34:585-612.
- PETO, R.; PIKE, M. C.; ARMITAGE, P.; BRESLOW, N. E.; COX, D. R.; HOWARD, S. V.; MANTEL, N.; MCPHERSON, K.; PETO, J. & SMITH, P. G., 1977. Design and analysis of randomized clinical trials requiring prolonged observation of each patient. II Analysis and examples. *British Journal of Cancer*, 35:1-39.
- SELBY, P.; GILLIS, C. & HAWARD, R., 1996. Benefits from specialised cancer care. *Lancet*, 348:313-318.
- SELVIN, S., 1996. *Statistical Analysis of Epidemiologic Data*. 2nd Ed. New York/Oxford: Oxford University Press.
- STILLER, C. A., 1994. Centralised treatment entry to trials and survival. *British Journal of Cancer*, 70: 352-362.
- SWAMINATHAN, R.; BLACK, R. J. & SANKARANARAYANAN, R., 1998. Database on cancer survival from developing countries. In: *Cancer Survival in Developing Countries* (R. Sankaranarayanan, R. J. Black & D. M. Parkin, ed.), IARC Scientific Publications 145, pp. 19-26. Lyon: International Agency for Research on Cancer.
- SUDAN, D.; SUDAN, R. & STRATTA, R., 2000. Long-term outcome of simultaneous kidney-pancreas transplantation: Analysis of 61 patients with more than 5 years follow-up. *Transplantation*, 69:550-555.
- SUE-LING, H. M.; JOHNSTON, D.; MARTIN, I. G.; DIXON, M. F.; LANSDOWN, M. R. J.; MCMAHOM, M. J. & AXON, A. T. R., 1993. Gastric cancer: A curable disease in Britain. *BMI*, 307:187-210.
- SZKLO, M. & NIETO, F. J., 2000. *Epidemiology: Beyond the Basics*. Annapolis: Aspen Publishers.
- TABÀR, L.; FAGERBERG, G.; DUFFY, S. W.; DAY, N. E.; GAD, A. & GRÖNTOFT, O., 1992. Update of the Swedish Two-County program of mammographic screening for breast cancer. *Radiologic Clinics of North America*, 30:187-210.
- VERDECCHIA, A.; CAPOCACCIA, R. & HAKULINEN, T., 1995. Methods of data analysis. In: *Survival of Cancer Patients in Europe: The Eurocare Study* (F. Berrino, M. Sant, A. Verdecchia, R. Capocaccia, T. Hakulinen & J. Estève, ed.), IARC Scientific Publications 132, pp. 32-37, Lyon: International Agency for Research on Cancer.
- YOUNG, J. L., 1991. The hospital-based cancer registry. In: *Cancer Registration: Principles and Methods* (O. M. Jensen, D. M. Parkin, R. Maclennan, C. S. Muir, R. G. Skeet, ed.), IARC Scientific Publication 95, pp. 177-184, Lyon: International Agency for Research on Cancer.
- ZIBNERS, A.; CROMER, B. A. & HAYES, J., 1999. Comparison of continuation rates for hormonal contraception among adolescents. *Journal of Pediatric and Adolescent Gynecology*, 12:90-94.

Recebido em 24 de agosto de 2001

Versão final reapresentada em 5 de novembro de 2001

Aprovado em 3 de dezembro de 2001