



UNIVERSIDADE FEDERAL DO PARANÁ  
SETOR DE CIÊNCIAS EXATAS  
DEPARTAMENTO DE ESTATÍSTICA  
CURSO DE ESTATÍSTICA

**Joseane Eloise Claudino**  
**Rafaela Helbing de Oliveira**

**MODELO DE FRAÇÃO DE CURA NA ANÁLISE DE SOBREVIDA DE  
PACIENTES COM CÂNCER DE PELE**

**CURITIBA**  
**2013**

**Joseane Eloise Claudino  
Rafaela Helbing de Oliveira**

**MODELO DE MISTURA COM FRAÇÃO DE CURA NA ANÁLISE DE  
SOBREVIDA DE PACIENTES COM CÂNCER DE PELE**

Trabalho de Conclusão de Curso apresentado à disciplina Laboratório de Estatística do Curso de Graduação em Estatística da Universidade Federal do Paraná, como exigência parcial para obtenção do grau de Bacharel em Estatística.

Orientadora: Profa. Dra. Suely Ruiz Giolo

**CURITIBA  
2013**

***Aos nossos pais, que sempre nos deram  
apoio e incentivo para que nossos  
sonhos se tornassem realidade.***

## **AGRADECIMENTOS**

A Deus, por nos manter sempre fortes e confiantes, mesmo nos momentos difíceis.

A nossa orientadora Profa. Dra. Suely Ruiz Giolo, os ensinamentos compartilhados.

Aos nossos amigos, o companheirismo nos bons e maus momentos.

As nossas famílias, o apoio no decorrer desses anos de graduação.

Ao Centro de Tratamento de Câncer de Curitiba, os dados utilizados neste trabalho.

Ao Prof. Dr. César Augusto Taconeli, a sua participação e sugestões como membro da banca deste trabalho.

## RESUMO

### **Modelo de Mistura com Fração de Cura na Análise de Sobrevida de Pacientes com Câncer de Pele**

A pele é o maior órgão do corpo humano e também o maior órgão sensorial. Desempenha funções como proteger o organismo humano e regular a temperatura do corpo. O câncer de pele é dividido em melanoma e não melanoma, sendo o não melanoma o mais incidente. Porém, por apresentar altos índices de cura, sua taxa de mortalidade é uma das mais baixas. Neste trabalho, a sobrevida de 1.505 pacientes diagnosticados com câncer de pele entre os anos de 2000 e 2004 em um centro médico de Curitiba, foi analisada por meio de dois modelos de sobrevivência semiparamétricos. Um deles, o modelo de Cox e, o outro, um modelo de mistura com fração de cura. O primeiro foi escolhido por sua popularidade e, o segundo, pelo fato de o câncer de pele apresentar fração de cura elevada. De modo geral, os dois modelos apresentaram ajustes satisfatórios, com o modelo de fração de cura trazendo informações complementares ao modelo de Cox que podem ser úteis na prática. Em ambos os modelos, as covariáveis que apresentaram associação com o tempo de sobrevida dos pacientes foram: sexo, idade e tratamento feito na instituição. Entre os resultados obtidos, foi possível observar estimativas de probabilidade de cura elevadas em função do sexo, idade e tratamento ao qual o paciente pôde ser submetido em função do estágio da doença no diagnóstico. Contudo, prevenção e diagnóstico precoce são, sem dúvida, primordiais para a redução da incidência do câncer de pele e acréscimo do tempo de sobrevida dos acometidos pelo mesmo.

**Palavras-chave:** Câncer. Fração de Cura. Modelo de Cox. Modelo de Mistura.

## LISTA DE FIGURAS

- Figura 1 – Curvas de sobrevivência obtidas pelo método de Kaplan-Meier para as covariáveis: (a) sexo, (b) idade e (c) avaliação da doença, associadas aos dados de pacientes com câncer de pele.....17
- Figura 2 – Curvas de sobrevivência obtidas pelo método de Kaplan-Meier para as covariáveis: (a) grau de diferenciação da doença, (b) estágio da doença, (c) tratamento feito na instituição e (d) topografia, associadas aos dados de pacientes com câncer de pele. ....18
- Figura 3 – Gráficos dos resíduos padronizados de Schoenfeld *versus* tempo para o modelo de Cox ajustado aos dados de pacientes com câncer de pele.....20
- Figura 4 – Curvas de sobrevivência observada (preta) e estimada (vermelha) pelo modelo de fração de cura para  $S(t|\mathbf{x},\mathbf{z})$  (curvas à esquerda) e  $S(t|U=1,\mathbf{x})$  (curvas à direita) para pacientes do sexo feminino submetidos à cirurgia e idade média (61 anos).. ....21
- Figura 5 – Curvas de sobrevivência observada (preta) e estimada (vermelha) pelo modelo de fração cura para pacientes suscetíveis do sexo masculino submetidos à cirurgia mais quimioterapia e idade média (61 anos).....24

## LISTA DE TABELAS

Tabela 1 – Descrição das covariáveis associadas aos dados de 1.505 pacientes com câncer de pele admitidos em um centro de tratamento em Curitiba (PR) .....	7
Tabela 2 – Frequências, falhas e censuras observadas no estudo de pacientes com câncer de pele tratados em um centro médico de Curitiba (PR) .....	16
Tabela 3 – Resultados do teste <i>log-rank</i> associados às covariáveis registradas para os pacientes com câncer de pele tratados em um centro médico de Curitiba (PR) ..	17
Tabela 4 – Estimativas associadas ao modelo de Cox ajustado aos dados de pacientes com câncer de pele .....	19
Tabela 5 – Coeficientes de correlação de Pearson obtidos entre os resíduos padronizados de Schoenfeld e os tempos para o modelo de Cox ajustado aos dados de pacientes com câncer de pele .....	19
Tabela 6 – Correlação de Pearson entre as probabilidades de sobrevivência estimadas e observadas para combinações das covariáveis e idade média .....	22
Tabela 7 – Estimativas dos parâmetros associados ao componente logístico do modelo de mistura com fração de cura ajustado aos dados de pacientes com câncer de pele .....	23
Tabela 8 – Estimativas dos parâmetros associados ao componente de sobrevivência do modelo de mistura com fração de cura ajustado aos dados de pacientes com câncer de pele .....	23
Tabela 9 – Estimativas da probabilidade de cura obtidas para alguns pacientes de acordo com as covariáveis que apresentaram efeito significativo no modelo de mistura com fração de cura ajustado aos dados de pacientes com câncer de pele ..	24
Tabela 10 – Estimativas obtidas sob o modelo de regressão de Cox e o modelo de mistura com fração de cura nos tempos $t = 30, 60$ e $120$ meses .....	26

## SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	1
<b>2 MATERIAL E MÉTODOS</b> .....	6
2.1 MATERIAL .....	6
2.1.1 Base de dados .....	6
2.2 MÉTODOS .....	6
2.2.1 Estimador de Kaplan-Meier .....	8
2.2.2 Teste <i>log-rank</i> .....	9
2.2.3 Modelo de regressão de Cox .....	9
2.2.4 Modelo de mistura com fração de cura .....	12
<b>3 RESULTADOS</b> .....	15
3.1 ESTATÍSTICA DESCRITIVA .....	15
3.2 RESULTADOS DO MODELO DE COX .....	18
3.3 RESULTADOS DO MODELO DE MISTURA COM FRAÇÃO DE CURA .....	21
3.4 COMPARAÇÕES ENTRE OS RESULTADOS DOS MODELOS .....	25
<b>4 CONCLUSÕES</b> .....	27
<b>REFERÊNCIAS</b> .....	29



## 1 INTRODUÇÃO

Câncer é o termo utilizado para denominar uma doença em que há proliferação descontrolada de células levando a formação de um tecido anormal, o tumor. Estas células tendem a se dividir rapidamente tornando-se agressivas, e quando atingem a corrente sanguínea ou linfática desenvolvem a doença em outros órgãos. Esse fenômeno é denominado metástase.

As causas do câncer são variadas, podendo ser externas ou internas ao organismo, estando ambas inter-relacionadas. As causas externas relacionam-se ao meio ambiente e aos hábitos ou costumes próprios de um ambiente social e cultural. As causas internas são, na maioria das vezes, geneticamente pré-determinadas; estão ligadas à capacidade do organismo de se defender das agressões externas (INCA,1996).

Segundo uma pesquisa publicada em 2012 pelo Hospital A.C.Camargo, localizado em São Paulo, o câncer de pele é o tumor mais frequente na população brasileira. Estima-se que aproximadamente 120 mil pessoas são diagnosticadas anualmente.

A pele, além de ser o maior órgão do corpo humano, é também o maior órgão sensorial. É constituída por duas camadas: a epiderme, na parte externa, e a derme, na parte interna. Desempenha funções como proteger o organismo humano e regular a temperatura do corpo.

O câncer de pele é dividido em melanoma e não melanoma. O não melanoma é o mais incidente no país, sendo responsável por aproximadamente 20% dos casos novos de câncer. Porém, por apresentar altos índices de cura, sua taxa de mortalidade é uma das mais baixas. A maior incidência deste tipo de câncer de pele se dá na região da cabeça e do pescoço, que são justamente os locais de maior exposição direta aos raios solares (INCA, 2002). O melanoma é o mais grave, com grande risco de produzir metástase, apresenta altas taxas de mortalidade nos estágios mais avançados.

Estudos evidenciam existir uma relação entre a exposição excessiva à radiação solar e a deflagração do câncer cutâneo, sendo o principal fator de risco. Fatores genéticos como pessoas de pele, olhos e cabelos claros, contribuem para o desenvolvimento da doença.

Um estudo realizado nos Estados Unidos pela Universidade da Carolina do Norte (UNIAP, 2013) indicou que o câncer de pele é mais perigoso quando localizado no couro cabeludo ou no pescoço do que em qualquer outra parte do corpo. Nesse estudo, a taxa de sobrevivência de cinco anos para pacientes com câncer de pele no couro cabeludo ou pescoço foi de 83%, comparada com 92% para os pacientes que tiveram a doença no rosto, orelhas ou nas extremidades, braços, pernas, mãos e pés. De acordo com os pesquisadores, os pacientes que apresentaram câncer de pele nestas áreas eram um pouco mais velhos (em média, 59 anos), em comparação com a média de 55 anos dos pacientes analisados, além de serem na maioria homens (UNIAP, 2013).

O melanoma maligno é o quinto tipo de câncer mais comum no Reino Unido, sendo responsável por 4% de todos os novos casos (CANCER RESEARCH UK, 2012). Segundo este instituto de pesquisa, a incidência de melanoma maligno está relacionada com a idade. Uma média de 27% dos novos casos de melanoma malignos diagnosticados no Reino Unido entre 2008 e 2010 foi em pessoas com idade inferior a 50 anos, com uma média de 45% dos casos diagnosticados em pessoas com idade igual ou superior a 65 anos. Isso está em contraste com todos os cânceres combinados (excluindo os não melanoma), em que entre 2008 e 2010, 11% dos casos foram diagnosticados em pessoas com idade inferior a 50 anos, e 63% em pessoas com 65 anos ou mais (CANCER RESEARCH UK, 2012).

Segundo informações do *Cancer Council Australia*, organização não governamental da Austrália, a incidência de câncer de pele tem aumentado nas últimas décadas no país. De 1982 a 2007, diagnósticos de melanoma aumentaram cerca de 50%. De 1998 a 2007, o número de consultas para tratar câncer não melanoma aumentou 14%, atingindo 950 mil visitas por ano. Esse tipo de câncer é o mais comum em homens, com quase o dobro da incidência em relação às mulheres. Mais de 434 mil pessoas são tratadas por um ou mais tipos de câncer de pele não melanoma na Austrália a cada ano. Em 2011, 543 pessoas morreram por causa do câncer não melanoma (CANCER COUNCIL AUSTRALIA, 2013).

Desconsiderando o câncer de pele não melanoma, melanoma é o terceiro câncer mais comum em homens e mulheres australianas, e o câncer mais comum em australianos com idades entre 15 e 44 anos. Em 2009, mais de 11.500 pessoas na Austrália foram diagnosticadas com melanoma, sendo que em 2011,

1544 pessoas morreram devido ao melanoma. A taxa de sobrevivência relativa de cinco anos para o melanoma é de 90% para os homens australianos e 94% para as mulheres australianas (CANCER COUNCIL AUSTRALIA, 2013).

As taxas de incidência de melanoma maligno também registraram um aumento global na Grã-Bretanha. Durante um período de 30 anos, essas taxas aumentaram mais rapidamente do que qualquer um dos atuais dez tipos de câncer mais comuns em homens e mulheres. Um estudo publicado em dezembro de 2011 estimou que cerca de 86% dos melanomas malignos diagnosticados no Reino Unido em 2010 estavam ligados à exposição aos raios ultravioletas (RUV) do sol (CANCER RESEARCH UK, 2012). Seguindo um padrão semelhante ao dos homens, o maior aumento global tem sido para as mulheres com idade entre 60 e 79 anos, em que as taxas de incidência têm aumentado mais de cinco vezes; de em torno de 7 casos por 100 mil mulheres em 1975-1977 para cerca de 40 casos por 100 mil mulheres em 2008-2010 (CANCER RESEARCH UK, 2012). Em 2010, no Reino Unido, o risco de desenvolver melanoma maligno foi de 1 em 55 para homens e 1 em 56 para mulheres. Na Grã-Bretanha, o melanoma maligno em homens é mais frequente na região do tronco do corpo, principalmente nas costas, enquanto as pernas é o local mais comum para as mulheres.

O melanoma maligno é o 19o câncer mais comum em todo o mundo; foi estimado ser responsável por cerca de 200.000 novos casos de câncer em 2008 (mais de 1% do total). As taxas de incidência de melanoma maligno são maiores na Austrália e Nova Zelândia e menores na Ásia do Sul e Central (CANCER RESEARCH UK, 2012).

Enquanto estão sendo tratados de um câncer, os indivíduos são avaliados para verificar se o câncer está respondendo à terapia. O tratamento mais eficaz é aquele que produz a cura. A cura é definida como uma remissão completa, na qual desaparece toda evidência de câncer (resposta completa). É usual pesquisadores caracterizarem a cura em termos de um período livre da doença, no qual o câncer desaparece completamente e não recorre dentro deste período pré-definido, frequentemente de 5 ou 10 anos (MSD, 2013). Em uma resposta parcial, o tamanho de um ou mais tumores reduz mais de 50%. Esta resposta pode reduzir os sintomas e pode prolongar a vida, embora o câncer acabe crescendo novamente. O tratamento menos eficaz corresponde ao que não produz resposta.

No que se refere à sobrevida de pacientes com câncer de pele (ou outro tipo), existe um conjunto de técnicas estatísticas que possibilitam estudar a eficiência de tratamentos, homologação de novos medicamentos e, principalmente, estimar o tempo de vida desses pacientes, para que lhes seja proporcionado uma melhor qualidade de vida. Tais técnicas consideram como variável resposta o tempo transcorrido desde um momento específico até a ocorrência do evento de interesse. Esse tempo é denominado tempo de falha, podendo ser o tempo desde o diagnóstico da doença até o óbito do paciente, bem como até a cura ou recidiva da doença (COLOSIMO; GIOLO, 2006).

Nos últimos tempos, tem ocorrido um grande interesse em modelos para dados de sobrevivência com a presença de uma fração de sobreviventes de longa duração (curados ou imunes ao evento). Isto porque, usualmente, os modelos de análise de sobrevivência consideram que todos os indivíduos no estudo são suscetíveis ao evento de interesse, o que pode não ser adequado quando uma fração deles for não suscetível (curados ou imunes). Neste contexto, Corbière e Joly (2007) apresentaram um modelo com fração de cura que assume que a população estudada é uma mistura de indivíduos suscetíveis (não curados), que podem apresentar o evento de interesse, e indivíduos não suscetíveis (curados), os quais nunca apresentarão o evento de interesse.

O objetivo principal desse trabalho é analisar a sobrevida de pacientes diagnosticados com câncer de pele em um centro de tratamento de câncer localizado em Curitiba, Paraná. Devido à alta frequência de indivíduos que não experimentaram o evento de interesse (isto é, não foram a óbito devido ao câncer de pele) até o término do estudo, é utilizado o modelo de sobrevivência com fração de cura apresentado por Corbière e Joly (2007) para analisar a sobrevida dos mesmos. Modelos mais usuais para análise de sobrevivência também são considerados; em particular o modelo de Cox. A variável resposta de interesse nesse estudo é o tempo transcorrido desde o diagnóstico do câncer de pele até o óbito do paciente devido à doença. Para os pacientes que não vieram a óbito, os tempos considerados são àqueles obtidos desde o diagnóstico até a data de encerramento do estudo.

Para ajuste do modelo de Cox é feito uso do pacote *survival* do *software* R (R DEVELOPMENT CORE TEAM, 2011) e, para ajuste do modelo de mistura com fração de cura, a macro PPSMCM do *software* SAS (SAS INSTITUTE INC., 2011), desenvolvida e implementada por Corbière e Joly (2007).

## 2 MATERIAL E MÉTODOS

### 2.1 MATERIAL

#### 2.1.1 Base de dados

A base de dados utilizada neste trabalho foi fornecida por um centro de tratamento de câncer localizado em Curitiba, PR. As informações foram coletadas em prontuários médicos durante 10 anos, tendo início em 2000 e término em 2009.

A base é composta por 1.505 pacientes diagnosticados com câncer de pele, sendo que estes deram entrada no centro de tratamento entre os anos de 2000 e 2004 e foram acompanhados até final de 2009. Destes, 230 (15,28%) foram a óbito. Dos que não vieram a óbito (84,72%), o tempo de acompanhamento variou entre 5 e 10 anos, visto terem entrado no estudo entre 2000 e 2004.

Do total de indivíduos, tem-se aproximadamente 51% do sexo masculino e, os outros 49%, do sexo feminino. As covariáveis avaliadas no decorrer das análises foram: sexo, idade, avaliação da extensão da doença, grau de diferenciação, estadiamento da doença no diagnóstico, ano de diagnóstico, topografia e tratamento realizado na instituição. A Tabela 1 apresenta uma descrição das covariáveis mencionadas.

### 2.2 MÉTODOS

Nos estudos em geral, há uma variável de interesse denominada variável resposta. No estudo em questão, tal resposta refere-se ao tempo decorrido desde o diagnóstico até o óbito do paciente devido ao câncer de pele. Há, então, o interesse em verificar o efeito de fatores de risco sobre o tempo de sobrevivência desses indivíduos.

Técnicas de análise de sobrevivência, que permitem a inclusão dos tempos dos indivíduos que vieram a óbito (tempos de falha) e dos que não vieram a óbito (tempos de censura) nas análises, foram utilizadas nesse trabalho. Tais técnicas

são descritas brevemente a seguir.

Tabela 1 – Descrição das covariáveis associadas aos dados de 1.505 pacientes com câncer de pele admitidos em um centro de tratamento em Curitiba (PR)

Covariável	Categorias	Descrição
<b>Sexo</b>	Feminino e Masculino	Gênero do paciente
<b>Idade</b>	9 a 100	Idade (em anos) do paciente no diagnóstico da doença
<b>Avaliação da extensão da doença (AED)</b>	In situ Localizado Extensão direta Extensão direta envolvendo linfonodos Metástase	Extensão da doença
<b>Grau de diferenciação (GD)</b>	Bem diferenciado Moderadamente diferenciado Pouco diferenciado Sem informação	Quanto mais diferenciado for o tumor, menor será o número de mitoses observada, e menor a agressividade do mesmo
<b>Topografia (TOPO)</b>	Face Ouvido, couro cabeludo e pescoço Troncos e membros superiores Quadril e membros inferiores	Localização do tumor
<b>Estágio</b>	Estágio 0 (restritos a área inicial) Estágio 1 (sem comprometimento linfático) Estágio 2 (espalhado por mais de um tecido) Estágio 3 (causa comprometimento linfático) Estágio 4 (espalhado para outros órgãos) Sem informação	Descreve o quanto o câncer se espalhou pelo corpo, geralmente enumerados de 0 a 4 de acordo com a facilidade de remover em cirurgia.
<b>Tratamento feito na instituição (TFI)</b>	Cirurgia Radioterapia Quimioterapia Cirurgia e Radioterapia Cirurgia e Quimioterapia	Tratamento realizado no combate ao câncer
<b>Ano de diagnóstico</b>	2000 a 2004	Ano em que o paciente entrou na instituição

Fonte: Os autores (2013)

Considerando uma amostra de  $n$  indivíduos, dados de sobrevivência são usualmente representados por  $(t_i, \delta_i, x_i)$ ,  $i = 1, \dots, n$ , em que  $t_i$  denota o tempo de falha ou censura associado ao indivíduo  $i$ ,  $\delta_i$  corresponde a uma variável indicadora

de falha ou censura (1 se falha e 0 se censura) e o vetor  $\mathbf{x}_i = (x_{1i}, \dots, x_{pi})'$  corresponde aos valores observados de  $p$  covariáveis.

Dentre as funções de interesse em estudos com dados de sobrevivência, destacam-se: a) a função de sobrevivência, que é definida como a probabilidade de um indivíduo sobreviver ao tempo  $t$ , isto é,

$$S(t) = P(T \geq t),$$

com  $T$  uma variável aleatória não negativa, representando o tempo de falha e; b) a função de taxa de falha,  $\lambda(t)$ , a qual define a taxa de falha instantânea de um indivíduo no tempo  $t$  condicional à sua sobrevivência até este tempo.

Para estimar a função de sobrevivência  $S(t)$  mencionada, uma das primeiras propostas de estimadores que incorpora observações censuradas foi desenvolvida por Kaplan e Meier em 1958.

### 2.2.1 Estimador de Kaplan-Meier

Esse estimador é denominado não-paramétrico por não especificar nenhuma distribuição de probabilidade para a variável tempo de sobrevivência  $T$ . É também denominado produto-limite, sendo o estimador mais utilizado em estudos clínicos.

Em sua construção, considera tantos intervalos de tempo quantos forem o número de falhas distintas. Os limites dos intervalos de tempo são os tempos de falha da amostra (KAPLAN; MEIER, 1958). A expressão do estimador de Kaplan-Meier é dada da seguinte forma:

$$S(t) = \prod_{j: t_j < t} \left( 1 - \frac{d_j}{n_j} \right),$$

sendo  $t_1 < t_2 < \dots < t_k$  os  $k$  tempos distintos e ordenados de falha,  $d_j$  o número de falhas em  $t_j$ ,  $j = 1, \dots, k$  e  $n_j$  o número de indivíduos sob risco em  $t_j$ , ou seja, os indivíduos que não falharam e não foram censurados até o instante imediatamente anterior a  $t_j$ .



As principais propriedades desse estimador são basicamente as seguintes:

- i) é não-viciado para amostras grandes;
- ii) é fracamente consistente;
- iii) converge assintoticamente para um processo gaussiano e;
- iv) é estimador de máxima verossimilhança de  $S(t)$ .

Este estimador foi utilizado, neste trabalho, para estudar a associação de cada covariável descrita na Tabela 1, com o tempo de sobrevivência dos pacientes.

### 2.2.2 Teste *log-rank*

O teste *log-rank* proposto por Mantel (1966) é um teste utilizado para comparar duas ou mais curvas de sobrevivência. Em ensaios clínicos é frequentemente utilizado para verificar a eficiência de tratamentos.

Se dois grupos estão sendo comparados, o teste é construído com base no número de falhas (denotado por  $O_{1j}$ ) observado em um dos grupos em cada tempo de falha distinto  $t_j$  ( $j = 1, \dots, k$ ) da amostra combinada (junção das duas amostras individuais), e nos respectivos valores esperados de falha (denotados por  $E_{1j}$ ) nos tempos  $t_j$  sob a hipótese nula de não diferenças entre os grupos. Sob esta hipótese, a estatística de teste é expressa por:

$$T = \frac{\left( \sum_{j=1}^k (O_{1j} - E_{1j}) \right)^2}{\sum_{j=1}^k V_j},$$

que segue distribuição qui-quadrado com 1 grau de liberdade, sendo  $V_j$  a variância de  $O_{1j}$  sob  $H_0$ .

### 2.2.3 Modelo de Regressão de Cox

O modelo de regressão de Cox (COX, 1972) é frequentemente utilizado para a análise de dados de estudos que têm por objetivo avaliar o efeito de

diversas covariáveis sobre o tempo de sobrevivência dos indivíduos.

Na presença de  $p$  covariáveis, com  $\mathbf{x}$  denotando o vetor com os componentes  $\mathbf{x} = (x_1, \dots, x_p)'$ , a expressão do modelo de Cox é dada por:

$$\lambda(t | \mathbf{x}) = \lambda_0(t)g(\mathbf{x}'\boldsymbol{\beta}),$$

em que  $g(\mathbf{x}'\boldsymbol{\beta})$  é uma função não-negativa que deve satisfazer  $g(\mathbf{0}) = 1$ .

O modelo de Cox é composto pelo produto de dois componentes, um não-paramétrico e outro paramétrico. O componente não-paramétrico,  $\lambda_0(t)$ , não é especificado e é uma função não-negativa do tempo. Ele é usualmente denominado função de taxa de falha de base, pois  $\lambda(t | \mathbf{x}) = \lambda_0(t)$ , quando  $\mathbf{x} = \mathbf{0}$  (COLOSIMO; GIOLO, 2006). O componente paramétrico é usualmente utilizado da seguinte forma:

$$g(\mathbf{x}'\boldsymbol{\beta}) = \exp(\mathbf{x}'\boldsymbol{\beta}) = \exp(\beta_1 x_1 + \dots + \beta_p x_p),$$

em que  $\boldsymbol{\beta}$  corresponde ao vetor de parâmetros associado às covariáveis.

Esse modelo é também denominado de taxas de falha proporcionais, pois assume que a razão das taxas de falha de dois indivíduos diferentes é constante no tempo (COLOSIMO; GIOLO, 2006). Isto é, a razão das funções de taxas de falha para os indivíduos  $i$  e  $j$ , dada por:

$$\frac{\lambda(t | x_i)}{\lambda(t | x_j)} = \frac{\lambda_0(t) \exp(\mathbf{x}'_i \boldsymbol{\beta})}{\lambda_0(t) \exp(\mathbf{x}'_j \boldsymbol{\beta})} = \exp(\mathbf{x}'_i \boldsymbol{\beta} - \mathbf{x}'_j \boldsymbol{\beta}),$$

não depende do tempo. Esta propriedade é utilizada para interpretar os parâmetros estimados, podendo-se calcular a taxa de falha de um indivíduo em relação a outro.

Para estimar o vetor de parâmetros  $\boldsymbol{\beta}$  é utilizado o método da máxima verossimilhança parcial proposto por Cox (1975). Considerando que existam  $k \leq n$  falhas distintas nos tempos  $t_1 < t_2 < \dots < t_k$ , tal função é expressa por:

$$L(\boldsymbol{\beta}) = \prod_{i=1}^n \left( \frac{\exp(\mathbf{x}'_i \boldsymbol{\beta})}{\sum_{j \in T(t_i)} \exp(\mathbf{x}'_j \boldsymbol{\beta})} \right)^{\delta_i},$$

em que  $\delta_i$  é a variável indicadora de falha.

Os valores de  $\boldsymbol{\beta}$  que maximizam a função  $L(\boldsymbol{\beta})$  são obtidos resolvendo-se o sistema de equações definido por  $U(\boldsymbol{\beta}) = 0$ , em que  $U(\boldsymbol{\beta})$  é o vetor escore de

derivadas de primeira ordem da função  $\ln(L(\beta))$  (COLOSIMO; GIOLO, 2006). Isto é,

$$U(\beta) = \sum_{i=1}^n \delta_i \left[ \mathbf{x}_i - \frac{\sum_{j \in R(t_i)} x_j \exp(\mathbf{x}'_j \hat{\beta})}{\sum_{j \in R(t_i)} \exp(\mathbf{x}'_j \hat{\beta})} \right] = 0.$$

O modelo de Cox, apesar de bastante flexível devido à presença do componente não-paramétrico, não se ajusta bem a todas as situações clínicas. Assim, algumas técnicas estatísticas precisam ser utilizadas para verificar aspectos relacionados à qualidade geral do ajuste do modelo e à suposição de taxas de falha proporcionais, entre outras.

A fim de verificar a suposição de taxas de falha proporcionais no modelo de Cox são utilizados alguns métodos gráficos e testes estatísticos. Neste trabalho, foram analisados os resíduos de Schoenfeld (SCHOENFELD, 1982). Esses resíduos são definidos para cada falha e não são definidos para censuras (COLOSIMO; GIOLO, 2006). Para que a estrutura de correlação dos resíduos seja considerada, os mesmos são padronizados. Detalhes sobre a padronização podem ser encontrados em Colosimo e Giolo (2006). O gráfico dos resíduos de Schoenfeld *versus* os tempos não devem apresentar tendências acentuadas ao longo do tempo para haver evidências a favor da suposição de taxas de falha proporcionais.

Testes sobre a proporcionalidade das taxas de falha, propostos com base nos resíduos padronizados de Schoenfeld, também foram utilizados a fim de verificar a proporcionalidade das taxas de falha sobre todas as covariáveis do modelo. Outra técnica utilizada para verificar esta suposição, foi a de obter, para cada covariável, o coeficiente de correlação de Pearson ( $\rho$ ) entre os resíduos padronizados de Schoenfeld e os tempos  $t$ , em que valores de  $\rho$  próximos de zero evidenciam a não rejeição da suposição de taxas de falha proporcionais.

Quanto ao procedimento de seleção de covariáveis, foi considerado, neste trabalho, o método de seleção *backward*. Este procedimento caracteriza-se por incorporar, inicialmente, todas as covariáveis no modelo e percorrer etapas, nas quais uma variável por vez pode vir a ser excluída. Para auxiliar nesse processo de exclusão de variáveis, foi utilizado o teste de Wald, bem como foi especificado nível de significância de 10% para exclusão. Se em uma dada etapa não houver exclusão de alguma variável, o processo é interrompido e as covariáveis restantes definem o modelo final (CHARNET *et al.*, 1999).

## 2.2.4 Modelo de mistura com fração de cura

Modelos de mistura com fração de cura são usualmente utilizados quando se deseja analisar indivíduos que são acompanhados por um período longo de tempo e observa-se que uma fração razoável deles não irá experimentar o evento de interesse. Tais indivíduos são denominados curados ou imunes ao evento ou, ainda, sobreviventes de longa duração. Como no estudo considerado nesse trabalho há uma fração de indivíduos nesta condição, tal modelo se apresenta como uma possibilidade bastante viável e, por este fato, será considerado para a análise dos tempos de sobrevivência dos pacientes com câncer de pele descritos na Seção 2.1.1.

Para formulação do modelo de mistura com fração de cura, considere a variável  $U$  que denota se um indivíduo é suscetível ( $U = 1$ ) ou não ( $U = 0$ ) ao evento de interesse. Desse modo, para um indivíduo que apresenta o evento de interesse ( $\delta_i = 1$ ) tem-se que  $U = 1$ . Já para um indivíduo que não apresenta o evento ( $\delta_i = 0$ ) pode se ter  $U = 1$  ou  $U = 0$ . Assim, para uma variável aleatória não-negativa  $T$  denotando o tempo até a falha, segue que a probabilidade de um indivíduo qualquer sobreviver ao tempo  $t$  pode ser expressa por:

$$\begin{aligned} S(t | \mathbf{x}, \mathbf{z}) &= P(U = 0 | \mathbf{z})P(T > t | U = 0, \mathbf{x}) + P(U = 1 | \mathbf{z})P(T > t | U = 1, \mathbf{x}) \\ S(t | \mathbf{x}, \mathbf{z}) &= (1 - \pi(\mathbf{z})) + \pi(\mathbf{z})S(t | U = 1, \mathbf{x}), \end{aligned} \quad (1)$$

em que  $S(t | \mathbf{x}, \mathbf{z})$  denota a função de sobrevivência de  $T$  para toda a população,  $S(t | U = 1, \mathbf{x}) = P(T > t | U = 1, \mathbf{x})$  a função de sobrevivência associada aos indivíduos suscetíveis dado o vetor de covariáveis  $\mathbf{x} = (x_1, \dots, x_p)'$  e  $\pi(\mathbf{z}) = P(U = 1 | \mathbf{z})$  a probabilidade de ser suscetível ao evento dado o vetor de covariáveis  $\mathbf{z} = (z_1, \dots, z_q)'$ . Os vetores de covariáveis  $\mathbf{x}$  e  $\mathbf{z}$  podem ou não incluir as mesmas covariáveis.

Neste trabalho, o efeito de  $\mathbf{z}$  sobre a probabilidade  $\pi(\mathbf{z})$  foi modelado por meio de um modelo de regressão para dados binários com ligação *logit*, isto é,

$$\text{logit}(\pi(\mathbf{z})) = \gamma_0 + \gamma_1 z_1 + \dots + \gamma_q z_q = \boldsymbol{\gamma}' \mathbf{z},$$

em que  $\gamma_0$  denota o intercepto e  $\boldsymbol{\gamma}$  o vetor de parâmetros associados a  $\mathbf{z}$ .

Observe que  $S(t | \mathbf{x}, \mathbf{z}) \rightarrow 1 - \pi(\mathbf{z})$  quando  $t \rightarrow \infty$ . Ainda, quando  $\pi(z_i) = 1$  para todo  $z_i$ , isto é, quando não há fração de curados, o modelo de mistura em (1) se reduz ao modelo de sobrevivência padrão. Várias especificações paramétricas e semiparamétricas têm sido propostas para  $S(t | U = 1, \mathbf{x})$ , levando aos modelos de mistura com fração de cura paramétricos e semiparamétricos, respectivamente.

Para a distribuição condicional  $S(t | U = 1, \mathbf{x})$ , foi utilizado, neste trabalho, o modelo semiparamétrico de Cox, tal que:

$$S(t | U = 1, \mathbf{x}) = S_0(t | U = 1)^{\exp(\mathbf{x}'\boldsymbol{\beta})} = \exp(-\exp(\mathbf{x}'\boldsymbol{\beta}) \int_0^t \lambda_0(v | U = 1) dv),$$

em que  $\boldsymbol{\beta}$  representa um vetor de parâmetros de regressão desconhecidos, sendo  $S_0(t | U = 1)$  e  $\lambda_0(t | U = 1)$  denominadas funções condicional de sobrevivência de base e de risco de base, respectivamente. Em consequência, a função condicional de risco acumulada fica expressa como:

$$\Lambda(t | U = 1) = \Lambda_0(t | U = 1) \exp(\mathbf{x}'\boldsymbol{\beta}), \text{ com } \Lambda_0(t | U = 1) = \int_0^t \lambda_0(v | U = 1) dv.$$

Quando não é assumida nenhuma forma paramétrica para  $S_0(t | U = 1)$ , o modelo em (1) é dito modelo de mistura semiparamétrico de riscos proporcionais com fração de cura (CORBIÈRE; JOLY, 2007).

Supondo que os dados estejam na forma  $(t_i, \delta_i, \mathbf{x}_i, \mathbf{z}_i)$ , em que  $\delta_i = 0$  indica censura e  $\delta_i = 1$  falha, para  $i = 1, \dots, n$ , tem-se que a contribuição do indivíduo  $i$  para a função de verossimilhança é dada por  $\pi(\mathbf{z}_i) f(t_i | U = 1, \mathbf{x}_i)$ , se  $\delta_i = 1$ , e  $(1 - \pi(\mathbf{z}_i)) + \pi(\mathbf{z}_i) S(t_i | U = 1, \mathbf{x}_i)$ , se  $\delta_i = 0$ , em que  $f(\cdot) = S(\cdot) \lambda(\cdot)$  denota a função de densidade de probabilidade condicional de T (CORBIÈRE; JOLY, 2007). Desse modo, a função de verossimilhança completa fica expressa por:

$$L(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \prod_{i=1}^n [\pi(\mathbf{z}_i) f(t_i | U = 1, \mathbf{x}_i)]^{\delta_i} [(1 - \pi(\mathbf{z}_i)) S(t_i | U = 1, \mathbf{x}_i)]^{1 - \delta_i}.$$

Estimadores para os vetores de parâmetros  $\boldsymbol{\beta}$ ,  $\boldsymbol{\gamma}$  e para  $S_0(t | U = 1)$  são obtidos maximizando-se a função de verossimilhança  $L(\boldsymbol{\beta}, \boldsymbol{\gamma})$ . O algoritmo EM (*Expectation Maximization*) proporciona uma maneira simples e eficiente para esse procedimento de estimação. Para mais detalhes consultar Corbière e Joly (2007).

Para seleção das covariáveis  $\mathbf{x}$  e  $\mathbf{z}$  associadas ao modelo (1), foi também utilizado o procedimento de seleção *backward* descrito na Seção 2.2.3.

Já quanto à adequação do modelo selecionado, foram comparadas graficamente as curvas de sobrevivência observada (representada pela curva obtida por Kaplan-Meier) e a estimada pelo modelo para algumas combinações das covariáveis. Isso foi feito tanto para a sobrevivência populacional  $S(t | \mathbf{x}, \mathbf{z})$  quanto para a sobrevivência condicional  $S(t | U = 1, \mathbf{x})$ . Adicionalmente, foram calculadas correlações de Pearson entre as probabilidades de sobrevivência observadas (Kaplan-Meier) e preditas pelo modelo para todas as combinações das categorias das covariáveis no modelo. Na presença de covariável contínua, foi considerado o valor médio da mesma. Se as curvas estiverem próximas, e houver alta correlação, pode-se dizer que o modelo se ajusta bem aos dados.

### 3 RESULTADOS

#### 3.1 ESTATÍSTICA DESCRITIVA

Inicialmente foi realizada uma análise descritiva dos dados dos 1.505 pacientes com o auxílio dos *softwares* R e SAS. Assim, de acordo com a Tabela 2, que apresenta as frequências de pacientes, falhas e censuras em função das covariáveis de interesse, podem-se observar frequências similares de pacientes homens e mulheres (51% e 49%, respectivamente), bem como que a faixa etária mais frequente foi a de 60 a 79 anos, compreendendo 50% dos pacientes neste estudo. Ainda, os tumores mais frequentes no diagnóstico foram os *in situ* e localizado (em torno de 80%); assim como aproximadamente 66% estavam nos estágios iniciais (0, 1 ou 2) da doença. Cirurgia foi realizada para em torno de 75%, além de cirurgia mais um tratamento adjuvante (radioterapia ou quimioterapia) em outros 7,2%. A localização do corpo com diagnóstico mais frequente de câncer de pele foi a face (61%), seguido de troncos e membros superiores (20%). Dos 1505 pacientes, em torno de 25% foram diagnosticados em 2000 e 15% em 2004.

A partir da Tabela 2, observa-se também que óbito em razão do câncer de pele foi proporcionalmente maior entre pacientes do sexo masculino (17% contra 13% entre as mulheres), assim como entre os: a) na faixa etária de 80 a 100 anos (39%); b) com metástase (84%); c) tratados com cirurgia e quimioterapia ou somente com quimioterapia (71% e 50%, respectivamente) e; d) localizados no ouvido, couro cabeludo e pescoço (21%) ou quadril e membros inferiores (21%).

O percentual pequeno de óbitos observados entre os pacientes diagnosticados com câncer de pele em 2004 (3%) pode ser consequência, em parte, do menor tempo de acompanhamento desses pacientes, que foi interrompido em 2009. Contudo, outras razões merecem investigação.

Quanto ao grau de diferenciação da doença, foi observada falta de informação para em torno de 85% dos pacientes, o que certamente limita seu uso nas análises estatísticas.

Tabela 2 – Frequências, falhas e censuras observadas no banco de dados de pacientes com câncer de pele tratados em um centro médico de Curitiba (PR)

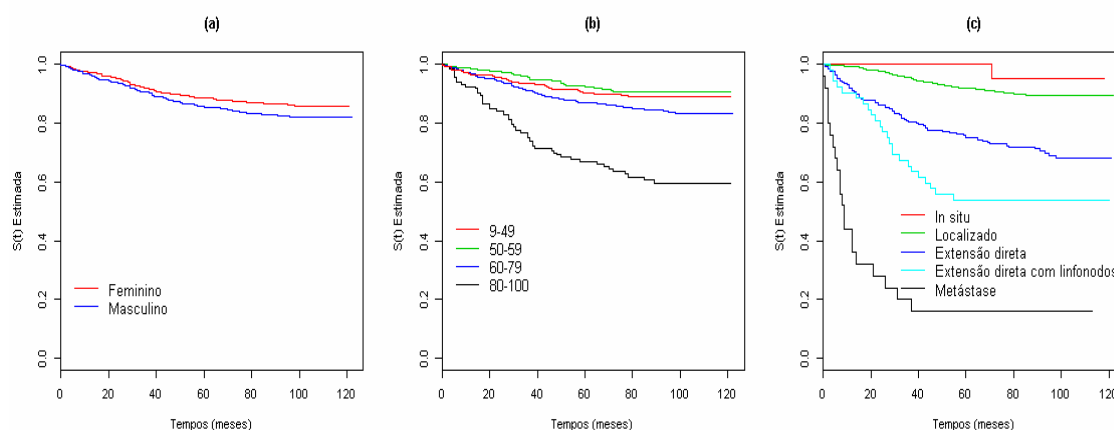
Covariável	Categorias	N	Censura	Falha	% Censura	% Falha
<b>Sexo</b>	Feminino	739	640	99	87%	13%
	Masculino	766	635	131	83%	17%
<b>Idade</b>	9-49	321	287	34	89%	11%
	50-59	296	269	27	91%	9%
	60-79	755	638	117	85%	15%
	80-100	133	81	52	61%	39%
<b>Avaliação da extensão da doença (AED)</b>	In Situ	21	20	1	95%	5%
	Localizado	1193	1074	119	90%	10%
	Extensão Direta	214	149	65	70%	30%
	Extensão Direta com envolvimento Linfonodo	52	28	24	54%	46%
	Metástase	25	4	21	16%	84%
<b>Grau de diferenciação (GD)</b>	Bem Diferenciado	94	74	20	79%	21%
	Moderadamente Diferenciado	123	89	34	72%	28%
	Pouco Diferenciado	12	9	3	75%	25%
	Sem informação	1276	1103	173	86%	14%
<b>Estágio</b>	Estágio 0 (restritos a área inicial)	21	20	1	95%	5%
	Estágio 1 (sem comprometimento linfático)	672	612	60	91%	9%
	Estágio 2 (espalhado por mais de um tecido)	298	246	52	83%	17%
	Estágio 3 (causa comprometimento linfático)	203	135	68	67%	33%
	Estágio 4 (espalhado para outros órgãos)	22	8	14	36%	64%
	Sem informação	289	254	35	88%	12%
<b>Tratamento realizado na instituição (TFI)</b>	Cirurgia	1123	992	131	88%	12%
	Radioterapia	270	208	62	77%	23%
	Quimioterapia	4	2	2	50%	50%
	Cirurgia e Radioterapia	94	69	25	73%	27%
	Cirurgia e Quimioterapia	14	4	10	29%	71%
<b>Topografia (TOPO)</b>	Face	916	797	119	87%	13%
	Ouvido, couro cabeludo e pescoço	191	150	41	79%	21%
	Troncos e membros superiores	300	251	49	84%	16%
	Quadril e membros inferiores	98	77	21	79%	21%
<b>Ano de diagnóstico</b>	2000	371	310	61	84%	16%
	2001	280	217	63	78%	23%
	2002	339	296	43	87%	13%
	2003	286	229	57	80%	20%
	2004	229	223	6	97%	3%

Fonte: Os autores (2013)



Considerando as covariáveis na Tabela 2, foi utilizado, a seguir, o estimador de Kaplan-Meier para cada uma delas, a fim de verificar quais apresentam evidências de associação com o tempo até o óbito, sendo, assim, candidatas a entrarem nos modelos de Cox e de fração de cura. As Figuras 1 e 2 apresentam as curvas de sobrevivência estimadas e sugere que as covariáveis: sexo, idade, avaliação da extensão da doença, estágio da doença e tratamento feito na instituição, apresentam curvas distintas, o que é confirmado pelo teste *log-rank* apresentado na Tabela 3, não somente para estas covariáveis, mas para todas com p-valores inferiores ou próximos de 5%.

Figura 1 – Curvas de sobrevivência obtidas pelo método de Kaplan-Meier para as covariáveis: (a) sexo, (b) idade e (c) avaliação da extensão da doença, associadas aos dados de pacientes com câncer de pele.



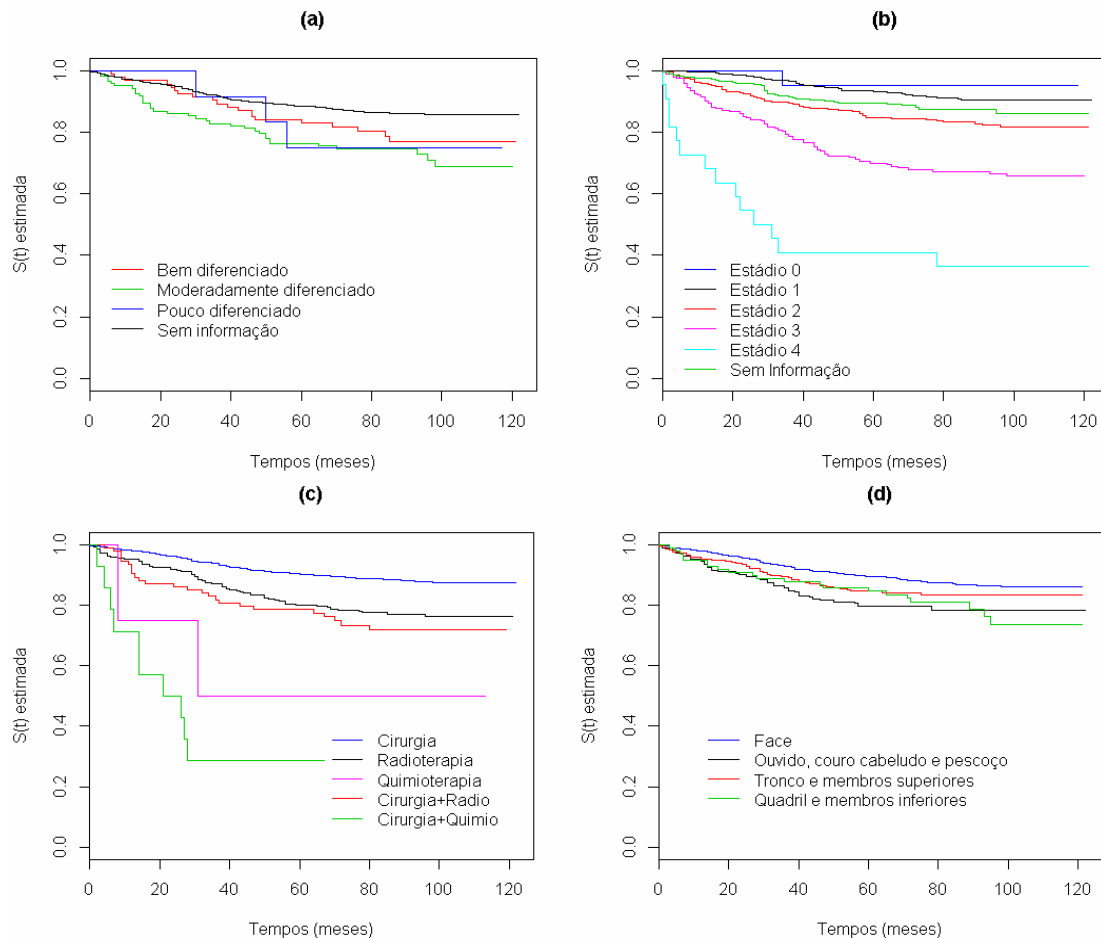
Fonte: Os autores (2013)

Tabela 3 – Resultados do teste log-rank associados às covariáveis registradas no estudo de pacientes com câncer de pele tratados em um centro médico de Curitiba (PR)

Covariável	P-valor associado ao teste <i>log-rank</i>
Sexo	0,0510
Idade	< 0,0001
Avaliação da extensão da doença (AED)	< 0,0001
Grau de Diferenciação (GD)	< 0,0001
Estágio da doença	< 0,0001
Tratamento feito na instituição (TFI)	< 0,0001
Topografia	0,0030
Ano de diagnóstico	< 0,0001

Fonte: Os autores (2013)

**Figura 2** - Curvas de sobrevivência obtidas pelo método de Kaplan-Meier para as covariáveis: (a) grau de diferenciação da doença, (b) estágio da doença, (c) tratamento feito na instituição e (d) topografia, associadas aos dados de pacientes com câncer de pele.



Fonte: Os autores (2013)

### 3.2 RESULTADOS DO MODELO DE COX

Considerando o modelo de Cox descrito na Seção 2.2 e o método de seleção *backward*, foi selecionado o modelo com as covariáveis: sexo, idade e tratamento feito na instituição, cujas estimativas dos parâmetros estão na Tabela 4.

No processo de seleção das covariáveis foi observada uma correlação alta entre tratamento e estágio da doença, o que é esperado, uma vez que pacientes com estágio avançado da doença não são usualmente submetidos à cirurgia; o inverso ocorrendo para aqueles no estágio inicial da doença. Dado que foi observada falta de informação para o estadiamento da doença para em torno de

19% dos pacientes e para a variável tratamento não foi observado dados faltantes, optou-se por manter a variável tratamento no modelo. Além disso, em termos práticos, há também um maior interesse dos médicos em avaliar de que forma o tratamento influencia no tempo de sobrevivência dos pacientes.

Nota-se para as covariáveis categóricas no modelo, que uma das categorias foi utilizada como categoria de referência. Para a covariável tratamento foi cirurgia e, para a covariável sexo, feminino.

Tabela 4 – Estimativas associadas ao modelo de Cox ajustado aos dados de câncer de pele

<b>Covariável</b>	<b>coeficiente</b>	<b>exp(coef)</b>	<b>I.C. 95%</b>	<b>p-valor</b>
Tratamento				
Radioterapia	0,6052	1,83	(1,35; 2,48)	< 0,0001
Quimioterapia	2,5057	12,25	(2,97; 50,41)	0,0005
Cirurgia+Radioterapia	0,8597	2,36	(1,53; 3,63)	< 0,0001
Cirurgia+Quimioterapia	2,9622	19,34	(10,02; 37,3)	< 0,0001
Idade	0,0355	1,03	(1,02; 1,04)	< 0,0001
Sexo: Masculino	0,2521	1,28	(0,98; 1,67)	0,0617

Fonte: Os autores (2013)

Quanto à suposição de taxas de falha proporcionais assumida para este modelo, pode-se observar a partir da Tabela 5 que os valores obtidos para os coeficientes de correlação de Pearson  $\rho$  estão próximos de zero para todas as covariáveis, bem como que os p-valores associados ao teste foram não significativos ao nível de significância de 5%, evidenciando a não rejeição da proporcionalidade das taxas de falha.

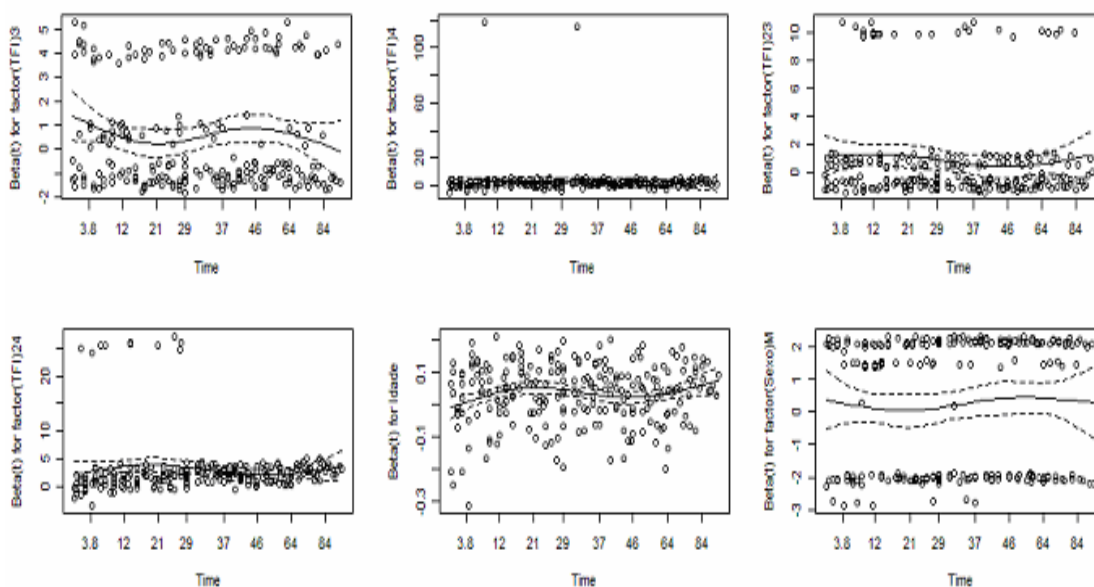
Tabela 5 – Coeficiente de correlação de Pearson entre os resíduos padronizados de Schoenfeld e os tempos para o modelo de Cox ajustado aos dados de pacientes com câncer de pele

<b>Covariável</b>	<b><math>\rho</math></b>	<b>Qui-quadrado</b>	<b>P-valor</b>
Tratamento			
Radioterapia	-0,0442	0,444	0,505
Quimioterapia	-0,0374	0,319	0,572
Cirurgia e Radioterapia	-0,0636	0,928	0,335
Cirurgia e Quimioterapia	-0,0475	0,505	0,478
Idade	0,0751	1,843	0,175
Sexo: Masculino	0,0375	0,315	0,575
GLOBAL	--	4,052	0,670

Fonte: Os autores (2013)

Observa-se também a partir da Figura 3, que os resíduos padronizados de Schoenfeld não apresentam evidências de tendências acentuadas ao longo do tempo, o que reforça a não violação da suposição de taxas de falha proporcionais.

**Figura 3** - Gráficos dos resíduos padronizados de *Schoenfeld* versus tempos para o modelo de Cox ajustado aos dados de pacientes com câncer de pele



Fonte: Os autores (2013)

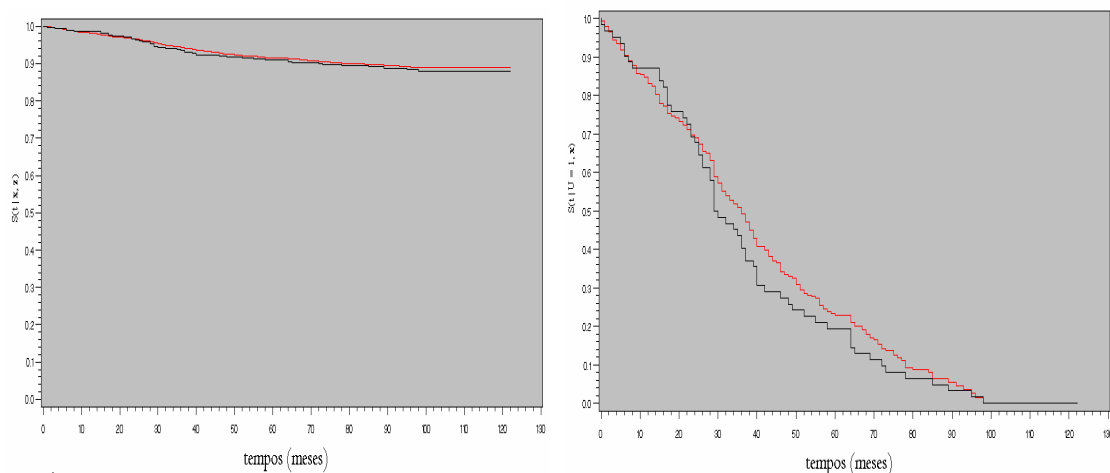
Uma vez que o modelo selecionado apresentou ajuste satisfatório, pode-se concluir a partir das estimativas mostradas na Tabela 4, que o risco de óbito dos pacientes que realizaram radioterapia, foi cerca de 1,83 vez o de pacientes que puderam ser submetidos à cirurgia (por estarem em estágios menos avançados da doença). Já para os pacientes que receberam quimioterapia (e que na sua maioria estavam em um estágio mais avançados da doença), tal risco foi 12,25 vezes o dos tratados com cirurgia. Por outro lado, os que receberam cirurgia mais radioterapia, apresentaram risco de óbito 2,36 vezes o dos que fizeram somente cirurgia. Ainda, os pacientes que receberam cirurgia mais quimioterapia (cuja frequência foi maior no estágio mais avançado da doença) apresentaram risco de óbito 19,34 vezes o dos que receberam cirurgia. Tem-se, também, que os pacientes homens apresentaram risco de óbito 1,28 vez o das mulheres, assim como pacientes com 80 anos apresentaram tal risco em torno de 8 vezes ( $\exp\{0,0355 \cdot (80 - 20)\}$ ) o daqueles com 20 anos de idade.

### 3.3 RESULTADOS DO MODELO DE MISTURA COM FRAÇÃO DE CURA

Para seleção das covariáveis foram ajustados diversos modelos de fração de cura, primeiramente incluindo todas elas, tanto no componente logístico  $\pi(\mathbf{z})$  do modelo quanto no componente de sobrevivência  $S(t | U = 1, \mathbf{x})$ . Em seguida, foram retiradas uma a uma as que não foram significativas. As variáveis que permaneceram no componente logístico foram: idade, sexo e tratamento e, no componente de sobrevivência, somente a variável tratamento. Isso significa que as covariáveis sexo, idade e tratamento influenciaram na fração de cura, enquanto o tratamento no tempo de sobrevida dos pacientes suscetíveis.

Para verificar a adequação do modelo foram obtidas as curvas de sobrevivência observada (representada pela curva obtida pelo estimador de Kaplan-Meier) e a estimada pelo modelo, tanto para a sobrevivência populacional  $S(t | \mathbf{x}, \mathbf{z})$  quanto para a sobrevivência condicional  $S(t | U = 1, \mathbf{x})$ . Estas curvas, para uma das combinações (sexo feminino, cirurgia e idade média), podem ser visualizadas na Figura 4 e mostram que as estimativas produzidas pelo modelo são bastante próximas às obtidas por Kaplan-Meier, evidenciando a adequação do modelo aos dados.

**Figura 4** - Curvas de sobrevivência observadas (preta) e estimadas (vermelha) pelo modelo de mistura com fração de cura para  $S(t | \mathbf{x}, \mathbf{z})$  (curvas à esquerda) e  $S(t | U = 1, \mathbf{x})$  (curvas à direita) para pacientes do sexo feminino submetidos à cirurgia e com idade média.



Fonte: Os autores (2013)

Além disso, foram obtidas as correlações de Pearson entre as probabilidades de sobrevivência observadas (Kaplan-Meier) e previstas pelo modelo para as combinações das categorias das covariáveis categóricas, considerando a idade média em cada combinação. Tais correlações mostradas na Tabela 6 indicam o bom ajuste do modelo, tendo em vista os valores das correlações obtidas (menor 0,8282 e maior 0,9893).

Tabela 6 – Correlação de Pearson entre as probabilidades de sobrevivência estimadas e observadas para as combinações das covariáveis e idade média em cada combinação.

Combinações	Tratamento	Sexo	Correlação de Pearson ( $\rho$ )	$R^2 = (\rho)^2$
1	Cirurgia	Feminino	0,9945	0,9891
2	Cirurgia	Masculino	0,9974	0,9948
3	Cirurgia+quimioterapia	Feminino	0,9459	0,8948
4	Cirurgia+quimioterapia	Masculino	0,9853	0,9708
5	Cirurgia+radioterapia	Feminino	0,9100	0,8282
6	Cirurgia+radioterapia	Masculino	0,9770	0,9546
7	Quimioterapia	Feminino	0,9722	0,9452
8	Radioterapia	Feminino	0,9935	0,9870
9	Radioterapia	Masculino	0,9946	0,9893

Fonte: Os autores (2013)

A Tabela 7 apresenta as estimativas dos parâmetros associados à  $\pi(\mathbf{z})$  obtidas utilizando-se o *software SAS* por meio da macro PSPMCM. Os resultados nesta tabela mostram que o risco de óbito entre os pacientes que receberam como tratamento cirurgia + quimioterapia foi cerca de 27 vezes o risco de pacientes que receberam como tratamento cirurgia. Quanto à variável idade, pode-se dizer que pacientes com idade igual a 60 anos têm risco de óbito aproximadamente 5 vezes o dos pacientes com 20 anos. Ainda, pacientes do sexo masculino têm risco de óbito 1,35 vez o dos pacientes do sexo feminino.

Similarmente, tem-se na Tabela 8 as estimativas dos parâmetros associados à sobrevivência condicional  $S(t|U=1, \mathbf{x})$  do modelo de fração de cura ajustado aos dados sob estudo.

Tabela 7 – Estimativas dos parâmetros associados ao componente logístico do modelo de mistura com fração de cura ajustado aos dados de pacientes com câncer de pele

Covariável	Coeficiente	exp(coef.)	Erro Padrão	Estatística de Wald	P-valor
Intercepto	-4,6057	0,0099	0,4017	131,4299	< 0,0001
Idade	0,0393	1,0400	0,0057	47,5076	< 0,0001
Tratamento					
Radioterapia	0,6375	1,8917	0,1743	13,3804	0,0003
Quimioterapia	2,7298	15,329	1,0566	6,6746	0,0098
Cirurgia e Radioterapia	0,9718	2,6426	0,2545	14,5774	0,0001
Cirurgia e Quimioterapia	3,3099	27,382	0,6292	27,6694	< 0,0001
Sexo: Masculino	0,3009	1,3510	0,1487	4,0938	0,0430

Fonte: Os autores (2013)

Tabela 8 – Estimativas dos parâmetros associados ao componente de sobrevivência condicional do modelo de mistura com fração de cura ajustado aos dados de pacientes com câncer de pele

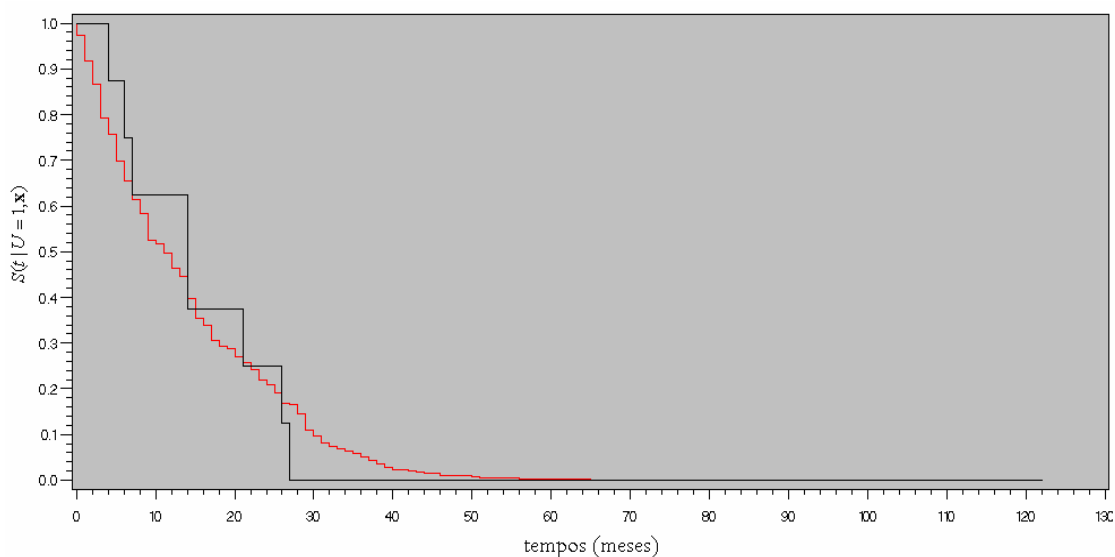
Covariável	Coeficiente	exp(coef.)	Erro Padrão	Estatística de Wald	P-valor
Tratamento					
Radioterapia	0,14507	1,1561	0,15463	0,8803	0,3481
Quimioterapia	1,02656	2,7914	0,71618	2,0546	0,1518
Cirurgia e Radioterapia	0,27409	1,3153	0,21911	1,5648	0,2110
Cirurgia e Quimioterapia	1,43195	4,1868	0,33987	17,7515	< 0,0001

Fonte: Os autores (2013)

Quanto aos resultados na Tabela 8, pode-se dizer que a covariável que apresentou maior influência no tempo de sobrevida dos pacientes que vieram a óbito foi o tratamento realizado. Dentre esses pacientes, os que tiveram menor tempo de sobrevida foram os tratados com cirurgia + quimioterapia (risco de óbito em um determinado tempo  $t$  igual a 4,2 vezes o dos tratados com cirurgia).

Para ilustrar tal fato, a Figura 5 mostra a curva de sobrevivência estimada dos pacientes suscetíveis do sexo masculino, com idade média (61 anos) e que foram submetidos à cirurgia + quimioterapia. Desta curva, pode-se observar que o tempo máximo de sobrevida estimado para esses pacientes foi de aproximadamente 65 meses, bem como que 50% deles apresentaram tempo de sobrevida inferior a 14 meses. Pode-se também observar, que apenas em torno de 10% deles apresentaram tempo de sobrevida superior a 30 meses.

**Figura 5** - Curvas de sobrevivência condicional observada (preta) e estimada (vermelha) pelo modelo de mistura com fração cura para pacientes suscetíveis do sexo masculino submetidos à cirurgia mais quimioterapia e idade média (61 anos).



Fonte: Os autores (2013)

Quanto à probabilidade de cura, esta pode ser estimada para combinações das covariáveis  $\mathbf{z}$  por  $1 - \pi(\mathbf{z})$ . Na Tabela 9, estimativas para tais probabilidades são apresentadas para algumas dessas combinações. Conforme resultados nesta tabela, podem-se notar probabilidades de cura elevadas entre os pacientes. Por exemplo, para o paciente do sexo feminino, com 27 anos de idade que realizou cirurgia, foi estimada uma probabilidade de cura de 99,61%. Já para o paciente com 70 anos de idade, do sexo masculino que realizou cirurgia + radioterapia, tal estimativa decresceu para 92,56%.

Tabela 9 - Estimativas da probabilidade de cura obtidas para alguns pacientes de acordo com as covariáveis que apresentaram efeito significativo no modelo de mistura com fração de cura ajustado aos dados de pacientes com câncer de pele

Pacientes	Idade	Sexo	Tratamento	$\pi(\mathbf{z})$	$1 - \pi(\mathbf{z})$
1	27	Feminino	Cirurgia	0,39%	99,61%
2	41	Feminino	Radioterapia	1,53%	98,47%
3	70	Masculino	Cirurgia+Radioterapia	7,44%	92,56%
4	61	Masculino	Cirurgia+Quimioterapia	0,48%	99,52%

Fonte: Os autores (2013)



### 3.4 COMPARAÇÕES ENTRE OS RESULTADOS DOS MODELOS

Com a finalidade de comparar os modelos de Cox e o de mistura com fração de cura ajustados aos dados, são mostrados na Tabela 10 resultados possíveis de serem obtidos para ambos os modelos considerando 9 combinações das covariáveis remanescentes nos mesmos. A partir desta tabela, fica evidenciado que ambos os modelos fornecem informações sobre os pacientes em quaisquer tempos entre 0 e 120 dias (na tabela foram mostrados para os tempos 30, 60 e 120 meses). Contudo, e diferente do modelo de Cox, o modelo de mistura fornece também informações ao longo do tempo específicas para o subconjunto de pacientes suscetíveis (que vieram a óbito), o que pode auxiliar os médicos na decisão de qual tratamento seria mais adequado para pacientes com perfis similares ao desses pacientes, bem como a terem algum parâmetro que os auxilie a estimar o tempo restante de vida de pacientes em estágios críticos.

Dentre os pacientes suscetíveis, observe a partir da Tabela 10, que o tempo de sobrevida dos submetidos à cirurgia + quimioterapia não foi muito superior a 30 meses para a maioria deles, com menos de 1% sobrevivendo mais de 60 meses. Neste grupo, pode-se também observar que a fração de cura ( $1 - \pi(\mathbf{z})$ ) foi inferior a 30% para os homens e 35% para as mulheres.

Tabela 10 – Estimativas obtidas sob o modelo de regressão de Cox e o modelo de mistura com fração de cura nos tempos  $t = 30, 60$  e  $120$  meses

Combinação das Covariáveis			Modelo de Cox			Modelo de Mistura com Fração de Cura						
			$S(t   \mathbf{x})$			$1 - \pi(\mathbf{z})$	$S(t   \mathbf{x}, \mathbf{z})$			$S(t   U = 1, \mathbf{x})$		
Tratamento	Sexo	Idade média	$t = 30$	$t = 60$	$t = 120$	$t \rightarrow \infty$	$t = 30$	$t = 60$	$t = 120$	$t = 30$	$t = 60$	$t = 120$
Cirurgia	Feminino	61	0,9574	0,9218	0,9001	0,9010	0,9592	0,9236	0,9010	0,5888	0,2283	0,000
Cirurgia	Masculino	60	0,9473	0,9038	0,8775	0,8750	0,9486	0,9036	0,8750	0,5888	0,2283	0,000
Cirurgia+quimioterapia	Feminino	50	0,5655	0,3445	0,2523	0,3386	0,4107	0,3400	0,3386	0,1088	0,0020	0,000
Cirurgia+quimioterapia	Masculino	50	0,4802	0,2538	0,1700	0,2748	0,3538	0,2763	0,2748	0,1088	0,0020	0,000
Cirurgia+radioterapia	Feminino	60	0,9055	0,8305	0,7867	0,7817	0,8905	0,8130	0,7817	0,4982	0,1433	0,000
Cirurgia+radioterapia	Masculino	62	0,8718	0,7738	0,7179	0,7102	0,8545	0,7517	0,7102	0,4982	0,1433	0,000
Quimioterapia	Feminino	48	0,7144	0,5332	0,4437	0,4974	0,6120	0,5055	0,4974	0,2280	0,0162	0,000
Radioterapia	Feminino	66	0,9091	0,8368	0,7944	0,7981	0,9075	0,8347	0,7981	0,5420	0,1813	0,000
Radioterapia	Masculino	64	0,8921	0,8077	0,7589	0,7599	0,8900	0,8034	0,7599	0,5420	0,1813	0,000

Fonte: Os autores (2013)

## 4 CONCLUSÕES

Neste trabalho, dados de 1.505 pacientes diagnosticados com câncer de pele entre 2000 e 2004 e acompanhados até 2009 foram analisados por meio de dois modelos: a) o modelo de regressão de Cox e b) o modelo de mistura semiparamétrico com fração de cura.

Ambos apresentaram ajustes satisfatórios aos dados, o que não muito é usual. Contudo, o modelo de mistura com fração de cura trouxe informações complementares ao do modelo de Cox que podem ser úteis na prática. Enquanto o modelo de Cox forneceu estimativas da probabilidade de um paciente  $i$  com covariáveis  $x_i$  sobreviver a certo tempo  $t$ , o modelo de mistura com fração de cura, além destas estimativas, forneceu também estimativas da probabilidade de pacientes suscetíveis ( $U = 1$ ) sobreviverem a este tempo  $t$ . Ou seja, é possível também obter um perfil dos indivíduos suscetíveis ao longo do período de acompanhamento, bem como saber as covariáveis que mais influenciaram no tempo de sobrevida dos mesmos.

Em ambos os modelos, as covariáveis que apresentaram associação com o tempo de sobrevida dos pacientes foram: sexo, idade e tratamento feito na instituição. Entre os resultados obtidos, foi possível observar estimativas de probabilidade de cura elevadas em função do sexo, idade e tratamento que pôde ser realizado pelo paciente em função do estágio da doença no diagnóstico.

Em relação às variáveis consideradas neste trabalho, algumas delas, como sexo e idade, foram estudadas por outros pesquisadores. Quanto à variável sexo, estudos apresentados pelo *Cancer Council Australia* e pelo UNIAP (Estados Unidos) relataram que homens daqueles países são mais propensos ao câncer de pele. Em nosso estudo, homens e mulheres apresentaram incidências similares; aproximadamente 51% e 49% respectivamente. Contudo, o tempo de sobrevida de homens foi menor em relação ao das mulheres.

Em relação à variável idade, pacientes com idades mais avançadas apresentaram tempo de sobrevida menor. Já quanto à incidência do câncer, esta foi maior entre os indivíduos com idades mais avançadas, concordando com o estudo realizado pela UNIAP, Estados Unidos. Já na Austrália, que tem uma das maiores incidências de câncer de pele do mundo, foi relatado pelo *Cancer Council*

*Australia*, alta incidência em faixas menores de idade.

Embora resultados relacionados à cura do câncer de pele sejam animadores, prevenção e diagnóstico precoce são, sem dúvida, primordiais para a redução da incidência desse câncer e acréscimo do tempo de sobrevivência daqueles acometidos por tal doença.

## REFERÊNCIAS

BONVINO, H.; CHARNET, E.; CHARNET, R.; FREIRE, C. **Análise de modelos de regressão linear**: com aplicações. Campinas, SP: Editora da Unicamp, 2.ed., 2008. 356p.

CANCER RESEARCH UK. **Skin cancer statistics**. Disponível em: <<http://www.cancerresearchuk.org/cancer-info/cancerstats/types/skin/incidence/uk-skin-cancer-incidence-statistics#world>>. Acesso em: 10 Jun. 2013.

CANCER COUNCIL AUSTRALIA. **Skin cancer**. Disponível em: <<http://www.cancer.org.au/about-cancer/types-of-cancer/skin-cancer.html>>. Acesso em: 10 Jun. 2013.

COLOSIMO; E. A., GIOLO, S. R. **Análise de sobrevivência aplicada**. São Paulo: Edgard Blucher, 2006. 370p.

CORBIÈRE, F.; JOLY, P. A SAS macro for parametric and semiparametric mixture cure models. **Computer Methods and Programs in Biomedicine**, v.85, n. 2, p. 173-180, 2007.

COX, D. R. Regression models and life tables. **Journal Royal Statistical Society, Series B**, v. 34, n.2, p. 187-220, 1972.

COX, D. R. Partial Likelihood. **Biometrika**, v. 62, p. 269-276, 1975.

INCA. Instituto Nacional de Câncer. **Prevenção e controle do câncer**. 2002. Disponível em: <[http://www.inca.gov.br/rbc/n\\_49/v04/pdf/norma1.pdf](http://www.inca.gov.br/rbc/n_49/v04/pdf/norma1.pdf)>. Acesso em: 15 Jun. 2013.

INCA. Instituto Nacional de Câncer. **Programa de Epidemiologia e Vigilância do Câncer e seus Fatores de Risco**. 2006. Disponível em: <[http://www.inca.gov.br/conteudo\\_view.asp?id=87](http://www.inca.gov.br/conteudo_view.asp?id=87)>. Acesso em: 15 Jun. 2013.

INCA. Instituto Nacional de Câncer. **Câncer: o que é?** Disponível em: <<http://www2.inca.gov.br/wps/wcm/connect/cancer/site/oquee>>. Acesso em: 10 Fev. 2013.

KAPLAN, E. L.; MEIER, P. Nonparametric estimation from incomplete observations, **Journal of the American Statistical Association**, v. 53, p. 457-481, 1958.

MANTEL, N. Evaluation of survival data and two new rank-order statistics arising in its consideration. **Cancer Chemotherapy Reports**, v. 50, p. 163-170, 1966.

MSD. **Tratamento do câncer**. Disponível em: <[http://mmspf.msdonline.com.br/pacientes/manual\\_merck/secao\\_15/cap\\_166.html](http://mmspf.msdonline.com.br/pacientes/manual_merck/secao_15/cap_166.html)> Acesso em: 25 Jun. 2013.

R DEVELOPMENT CORE TEAM. 2011. **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.

SAS INSTITUTE INC. **The SAS System, release 9.2**. SAS Institute Inc., Cary: NC, 2011.

UNIAP. Unidade de Apoio aos Portadores de Câncer. **Câncer de pele no couro cabeludo é mais perigoso**. Disponível em: <<http://www.uniap.org.br/noticia.asp?id=764&areas=saude>>. Acesso em: 10 Abr. 2013.