**COMPARAÇÃO ENTRE ALGUNS MÉTODOS ESTATÍSTICOS EM ANÁLISE
DE SOBREVIVÊNCIA: APLICAÇÃO EM UMA COORTE DE PACIENTES
COM CÂNCER DE PÊNIS**

MARIA DO ROSARIO DIAS DE OLIVEIRA LATORRE

Tese apresentada à Faculdade de Saúde Pública
da Universidade de São Paulo para a obtenção
do título de Doutor em Saúde Pública

ORIENTADOR: EDUARDO MASSAD

**São Paulo
1996**

Quando a atitude de viver
é uma extensão do coração
é muito mais que um prazer,
é toda a carga de emoção
que era um encontro com o sonho,
que só pintava no horizonte,
e de repente diz presente,
sorri, e beija nossa fronte,
e abraça e arrebatou a gente,
é bom dizer: viver valeu!
ah! já não é nem mais alegria,
já não é nem felicidade
é tudo aquilo num sol riso,
é tudo aquilo que é preciso
é tudo aquilo paraíso,
não há palavra que explique
é só dizer: viver valeu!
ah! eu me ofereço esse momento,
que não tem paga nem tem preço
essa magia eu reconheço,
aqui está a minha sorte,
me descobrir tão fraca e forte,
me descobrir tão sal e doce
e o que era amargo acabou-se,
é bom dizer: viver valeu,
é bom dizer: amar valeu, amar valeu!

Gonzaguinha

AGRADECIMENTOS

À **Profa. Sabina Léa Davidson Gotlieb**, na qualidade de Presidente da Comissão de Pós Graduação da Faculdade de Saúde Pública, por todo o apoio na concessão da bolsa de estudos. E, como amiga, pelos comentários da tese. Agradecimento especial pelo carinho e incentivo.

À **Dirce**, pela ajuda inestimável na elaboração dos modelos de processos de contagem.

Aos amigos do Departamento de Epidemiologia. Em particular à **Denise**, pela ajuda em diversos momentos; ao **Jair**, pelo sorteio da amostra; e ao **Fernão** pela ajuda na instalação dos *softwares*.

À **Dna. Neusa, Cidinha, Carminha, Junior e Sr. José**, funcionários da Faculdade de Saúde Pública, pela ajuda em momentos críticos no final deste trabalho. Muito obrigada!

À **Fundação Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)** pela concessão de bolsa de estudos para realizar estágio na Divisão de Epidemiologia do Departamento de Oncologia da *McGill University*.

Ao **Prof. Eduardo Franco**, pela oportunidade de meu estágio na Divisão de Epidemiologia do Departamento de Oncologia da *McGill University* e pelo “empréstimo” do banco de dados. Obrigada pela atenção e carinho com que me recebeu em Montreal.

Às pessoas que coletaram os dados e prepararam a base de dados no computador: **Dr. Serrano, Sra. Raimunda Pereira e Sra. Valéria Lombardo**. E ao **Dr. Ademar Lopes** pelas sugestões neste trabalho.

RESUMO

O objetivo deste trabalho foi comparar o desempenho do modelo de riscos proporcionais de Cox convencional, modelo de Cox modificado quando os riscos não são proporcionais e o modelo de análise de sobrevida baseado na teoria de processos de contagem. Para tanto utilizou-se uma coorte de 648 pacientes portadores de câncer de pênis, atendidos no Departamento de Cirurgia Pélvica do Hospital A. C. Camargo, no período de 1953 a 1985. Dessa coorte foram selecionadas três amostras com o objetivo de validar internamente os resultados da análise de sobrevida do banco de dados original.

Os resultados do modelo de riscos proporcionais de Cox, no banco de dados original, foram confirmados por uma das amostras desse conjunto de dados. Apenas o estadiamento N foi confirmado como fator prognóstico também nas outras duas amostras.

O modelo de riscos proporcionais de Cox e o modelo de análise de sobrevida baseado na teoria de processos de contagem apresentaram resultados semelhantes, na definição dos fatores prognósticos dessa coorte de pacientes com câncer de pênis. O modelo utilizando processos de contagem é mais sofisticado, do ponto de vista matemático. Porém o modelo de Cox está disponível em grande número de pacotes estatísticos e a interpretação de seus coeficientes se faz com maior facilidade. Por isso, talvez, continue a ser a técnica estatística mais utilizada quando o objetivo do estudo é definir fatores prognósticos e grupos de risco.

Os fatores prognósticos para a sobrevida de pacientes com câncer de pênis foram os estadiamentos T e N e o grau de diferenciação do tumor. Esses resultados foram ajustados pelo ano de início de tratamento no Hospital A.C. Camargo. Os pacientes com prognóstico favorável foram os que apresentaram tumor pequeno, sem presença de linfonodos clinicamente positivos, e tumor bem diferenciado.

ABSTRACT

The aim of this study was to compare the performance of the Cox proportional hazards model, the Cox model with time-dependent covariates and the survival model using the counting process theory. These methods were applied in a cohort of 648 patients with penile cancer treated at the Department of Pelvic Surgery, Hospital A.C. Camargo (São Paulo-Brazil), between 1953 and 1985. Three samples were selected from the total database in order to check the internal validity.

The prognostic factors selected using the Cox proportional hazards model were the same in one sample. The only prognostic factor selected in all samples was the N stage.

The T and N stages, and the grade of differentiation were independent prognostic factors of survival using both the Cox proportional hazards model and the survival model using the counting process theory. The statistical significance was the same and even the values of estimation of the coefficients were very close. The survival model using the counting process is more sophisticated from the mathematical point of view, but the Cox model is more available in statistical software, and, probably because of this, is more applied in survival analysis than the model using the counting process.

Patients with small tumors, clinically negatives nodes and well differentiated tumors showed a favorable prognosis. These results were adjusted by year of the beginning in the study.

ÍNDICE

	página
I. Introdução	1
II. Métodos estatísticos em análise de sobrevida	20
III. O câncer de pênis	38
IV. Hipótese do trabalho	55
V. Objetivos	57
VI. Material e Métodos	60
VII. Resultados	71
VIII. Discussão	125
IX. Conclusões	137
X. Referências bibliográficas	139
ANEXOS	150

I. INTRODUÇÃO

Os principais objetivos da pesquisa epidemiológica^(43,68) são: descrever a ocorrência da doença em uma população (a frequência e sua tendência no tempo e no espaço) e explicar esse fato, identificando quais os fatores e circunstâncias que mais estariam contribuindo para isso. Além disso, há a possibilidade de predizer a ocorrência da doença em uma população (ou sub-população) com base em suas características ou em determinada exposição e, dessa forma, prevenir e/ou controlar o agravo à saúde.

Para atender aos objetivos da Epidemiologia pode-se trabalhar com modelos que tentem reproduzir fatos reais, procurando descrever determinada doença de uma maneira mais simples e/ou conveniente, através de fórmulas matemáticas^(14,31).

No início do século havia a preocupação em tentar "modelar" as informações a respeito da mortalidade e sobrevivência das pessoas portadoras de algumas doenças, descobrindo uma função matemática ou distribuição de probabilidades conhecida que pudesse explicar os dados existentes⁽⁴⁶⁾. No entanto esse procedimento era extremamente complexo, visto que quanto mais parâmetros eram utilizados, menos operacional ficava o modelo. Somente com o surgimento do computador é que a metodologia quantitativa pôde ser melhor utilizada e foi possível dar maior liberdade às "fantasias matemáticas"⁽⁵²⁾. Com essa ferramenta é possível trabalhar-se com grandes bancos de dados, utilizando cálculos intensivos e interativos que antes estavam fora de questão^(14,29).

Os objetivos principais de um modelo são estudar o comportamento de uma população sob o ponto de vista de determinada doença, predizer esse comportamento, simular diversas situações futuras e descobrir relações não aparentes⁽⁴⁶⁾. Observa-se que esses objetivos adaptam-se perfeitamente às

finalidades essenciais da Epidemiologia. Para tanto, como já comentado, o modelo deve procurar retratar, o melhor possível, a realidade. E é aí que aparece o primeiro (talvez o mais difícil) problema quando se trabalha com modelagem: quanto mais real é o modelo, do ponto de vista biológico, mais difícil é o seu tratamento matemático⁽⁵²⁾, tornando-se, às vezes, tão complexo que perde sua utilidade.

Os modelos dividem-se em dois grandes grupos: os **determinísticos** e os **estocásticos**.

Um modelo determinístico é aquele constituído por uma equação (ou sistema de equações), que descreve (ou procura descrever) a história natural da doença e/ou o relacionamento das populações envolvidas (suscetíveis, agente etiológico, vetor e outras). São os que melhor se afiguram para representar a história natural da doença, pelo fato de serem estruturais e não funcionais como os modelos estatísticos. Além disso, somente com essa classe de modelos é possível estudar a ação e o mecanismo biológico dos fatores causais. Os modelos matemáticos não são simplesmente exercícios teóricos, mas fornecem subsídios importantes para estratégias de controle das doenças. No caso das doenças transmissíveis⁽¹⁰⁾, por exemplo, podem ser utilizados para avaliar o comportamento de certa doença em uma população com determinadas características (número de suscetíveis, infectados, doentes, imunes e outras) e avaliar o impacto que terão as possíveis medidas de prevenção e controle (vacinação, eliminação do vetor e outras). Simulações podem ser feitas, sem que sejam utilizadas populações humanas. Porém essa classe de modelos exige que o pesquisador tenha amplos conhecimentos de cálculo, equações diferenciais e outras ferramentas matemáticas não triviais, o que faz com que poucos se aventurem nessa área.

Os modelos estocásticos são funções matemáticas definidas, que incorporam o termo de erro aleatório em suas equações^(29,31,44). Quando se estuda causalidade

por meio de modelos estatísticos, a relação causal pode ser analisada avaliando-se o relacionamento entre as variáveis de estudo, considerando-se o possível erro aleatório existente. A maior desvantagem de sua utilização são as suposições paramétricas de um modelo estatístico. Algumas vezes o mecanismo biológico da doença pode ser melhor escrito através de uma específica fórmula matemática, que não é exatamente a que é assumida pelo modelo estatístico. A sua grande vantagem é a possibilidade de se avaliarem múltiplos fatores em estudo, utilizando-se fórmulas relativamente simples.

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 sua probabilidade de ocorrência, ou outra medida que vise descrever a frequência com que a mesma ocorre. Às vezes, a variável dependente de interesse é o tempo decorrido até o aparecimento de algum evento. 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.

As aplicações dos modelos estatísticos são diversas e podem ser citadas, dentre as mais importantes (ou mais frequentes)⁽⁴⁴⁾:

- caracterizar a associação entre a variável dependente e a(s) independente(s). Essa associação é descrita por meio do valor do parâmetro estimado (extensão), seu sinal (direção) e sua significância estatística (força);
- determinar a melhor fórmula que descreva a variável dependente como uma função da(s) variável(eis) independente(s);

- controlar o efeito de confusão e avaliar o possível efeito interativo entre duas ou mais variáveis.

A escolha do modelo estatístico mais apropriado dependerá do tipo do delineamento do estudo epidemiológico, seus objetivos, as variáveis estudadas e a maneira pela qual foram coletados e categorizados os dados.

É importante salientar aqui o grande mérito da análise descritiva no início de qualquer trabalho de modelagem estatística. A análise das tabelas, gráficos, medidas de tendência central e de dispersão, e do relacionamento entre as variáveis - associação ou correlação - fornecem as indicações de qual modelo deverá ser utilizado, quais as variáveis mais relevantes do estudo e como devem ser categorizadas, se for esse o caso. Essa fase preliminar, que muitas vezes é deixada de lado pelos pesquisadores, é imprescindível para que os dados forneçam as indicações da relação funcional entre a variável dependente e as covariáveis.

Os avanços na informática ocorridos a partir da década de 70 permitiram trabalhar com modelos estatísticos mais complexos e amplos. Uma pessoa consegue utilizar técnicas estatísticas razoavelmente sofisticadas, facilmente, e em curto espaço de tempo graças ao desenvolvimento de *softwares* cada vez mais potentes e fáceis de manusear, sem que para isso seja necessário conhecer a técnica, suas suposições básicas e restrições.

Contudo, utilizar a técnica é a parte mais simples de todo o processo de modelagem. Questões do tipo **qual modelo utilizar, quando, porquê e para quê** exigem que o pesquisador tenha sólidos conhecimentos de Estatística e Epidemiologia, criatividade e bom senso. Análises estatísticas razoavelmente sofisticadas passaram a ser muito empregadas e são comuns em artigos da literatura médica, porém muitas vezes a sua utilização é discutível^(5,17,31,35). Isso enfatiza a

necessidade de que as escolas de graduação na área da saúde, em geral, se dediquem, também, ao treinamento em relação às técnicas estatísticas aplicadas à Epidemiologia. Os alunos de graduação, de pós-graduação e pesquisadores, devem aprender, também, a ler criticamente, as seções de material e métodos e de resultados de qualquer trabalho científico, melhorando, assim, a qualidade de suas pesquisas, no que se refere à análise estatística.

CONCATO e colaboradores⁽¹⁷⁾ avaliaram diversos problemas em artigos científicos cujo objetivo era determinar risco por meio de modelos multivariados. Mostram a ampla utilização dessa metodologia em trabalhos no *Lancet* e no *New England Journal of Medicine*, onde em 15% dos artigos publicados no período de 1985 a 1989 os autores utilizaram análise multivariada. Referem que esses modelos são úteis não só em estudos de causalidade, mas, também, na determinação de escores de risco e na quantificação de riscos individuais. Chamam a atenção para a pouca referência sobre as suposições básicas e limitações das técnicas utilizadas. Os problemas mais comuns encontrados pelos autores na análise dos trabalhos publicados foram:

- resultados pouco confiáveis devido à grande amplitude dos intervalos de confiança associados às estimativas de riscos individuais. Isso ocorre, principalmente, quando há pequeno número de observações ou de eventos de interesse; muitas vezes o tamanho da população (ou da amostra) é adequado, porém o evento em estudo é raro, o que leva a um pequeno número de óbitos ou de casos;
- não verificação da suposição de linearidade;
- violação da suposição de que os riscos para uma variável independente, no modelo de Cox, são constantemente proporcionais;
- não avaliação da possível interação entre as variáveis independentes;
- não especificação da maneira pela qual as variáveis foram categorizadas;

- não menção da metodologia utilizada para a seleção das variáveis que compuseram o modelo final;
- não análise da colinearidade entre as variáveis independentes;
- não há menção de estudo a respeito dos valores aberrantes; e
- os modelos não foram validados.

As críticas feitas aos trabalhos que vêm sendo publicados servem de alerta aos pesquisadores sobre a importância que deve ser dada à metodologia empregada na análise dos dados, pois o sucesso de um estudo não é garantido apenas na fase de planejamento e coleta. Por mais que a máquina facilite a parte referente ao armazenamento e análise dos dados, ela não substitue o homem no que se refere à criatividade e ao bom senso. Além disso, sólidos conhecimentos da metodologia empregada e de epidemiologia são imprescindíveis.

Ao se elaborar um modelo estatístico deve-se ter o cuidado de validar os achados^(13,25,34,35,77,78). A maioria dos trabalhos que utilizam análise multivariada têm por objetivo a definição de um modelo estatístico para estimar o efeito conjunto de um grupo de variáveis em relação à uma variável dependente. O valor de um grupo de fatores de risco ou prognósticos definido a partir de um específico conjunto de dados vai depender do quanto esses resultados refletem o que realmente acontece na população em estudo (validade interna) e o quanto as conclusões podem ser generalizadas (validade externa).

Para avaliar a validade interna de um modelo, pode-se extrair amostras do banco de dados original e refazer a análise. Se as variáveis do modelo final permanecerem as mesmas, com a mesma significância estatística, o modelo pode ser considerado como válido internamente, ou seja, os resultados obtidos refletem os fatores de risco daquele específico banco de dados. O inconveniente desse procedimento é que algumas vezes não se dispõe de um banco de dados com número

de observações ou de falhas grande o suficiente para que se sorteie amostras. FEINSTEIN e colaboradores^(25,77,78) e HARREL e colaboradores⁽³⁴⁾ argumentam que esse procedimento pode ser muito influenciado pela composição da amostra, ou seja, se permanecem as mesmas porcentagens de ocorrência das falhas e a mesma correlação entre as variáveis independentes.

Um outro aspecto da validação de um modelo estatístico é a generalização dos resultados, também chamada de validade externa. Isto é, quanto os resultados obtidos em um particular estudo podem ser reproduzidos em outro banco de dados. Estes dois (ou mais) bancos de dados devem ser independentes. Alguns autores^(25,34,35,77,78) recomendam que se divida o banco de dados original em dois, fazendo o modelo de regressão em cada um deles separadamente, como se fossem dois bancos de dados originais, isto é, de populações independentes. Vale a pena salientar que isso só é possível se houver um banco de dados grande e com a ocorrência de muitos eventos.

Há técnicas de amostragem⁽³⁴⁾, como o *bootstrap*⁽⁴⁾, que são recomendadas para sortear amostras com o objetivo de validar (interna e externamente) modelos estatísticos.

Um outro aspecto a ser abordado em modelagem estatística é a avaliação da acurácia do modelo^(13,25,34,35,77,78). Isso pode ser analisado por dois aspectos: um é a capacidade de predição do modelo, ou seja, no caso de análise de sobrevivência, identificar como falha o indivíduo em que realmente ocorreu o evento. Outro é o seu poder discriminatório, ou seja, a habilidade em separar os diversos grupos de risco. Quando o objetivo do estudo é a predição, se não for avaliada a acurácia do modelo, o pesquisador pode não perceber ou avaliar que algumas das associações encontradas são espúrias.

Um aspecto de interesse em estudos epidemiológicos é a variável **tempo**. Um indivíduo pode apresentar diversos "estados de doença" (ou "de saúde", como preferem alguns) entre o completo bem estar físico, mental e social e o óbito. Em algumas pesquisas há o acompanhamento de indivíduos em uma população (ou grupo) durante um período de tempo até a ocorrência de um evento de interesse (geralmente aparecimento da doença, recorrência ou óbito). O tempo é definido como função das variáveis independentes e há a possibilidade de se avaliar a influência que determinadas covariáveis têm no comportamento do tempo decorrido entre o início da observação e o evento.

Há diversos delineamentos em epidemiologia, geralmente prospectivos como os de coorte e os ensaios clínicos, que podem utilizar o tempo como variável dependente. Nesses estudos, normalmente, há o interesse em se verificar o efeito de fatores de risco ou de fatores prognósticos (sejam eles quantitativos ou qualitativos) no tempo de sobrevivência de um indivíduo de um grupo, bem como definir as probabilidades de sobrevida em diversos momentos no seguimento do grupo. Sobrevivência, no presente trabalho, é considerado o tempo desde a entrada do indivíduo no estudo, até a ocorrência do evento de interesse (falha) ou até a censura (perda) na observação.

Em estudos de sobrevida as pessoas são acompanhadas a partir 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 sobrevivência 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⁽¹⁹⁾.

O indivíduo é seguido, então, até uma (ou mais) mudança(s) de estado (também denominado compartimento). O esquema mais simples de estudo é aquele onde o indivíduo é acompanhado até a ocorrência de apenas um evento de interesse. Nesse esquema são analisados apenas dois compartimentos, como, por exemplo, ausência e presença de uma doença:

No entanto pode haver o interesse em diversas mudanças de estado, como cura e/ou recuperação, recorrência ou óbito. Os compartimentos que descrevem os diversos "estados de saúde" (ou "de doença") obedecem a um processo de Markov finito, onde os estados são considerados etapas transitórias, com exceção do óbito que é considerado um estado absorvente, pois uma vez atingido esse compartimento a pessoa ali permanece^(8,10).

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^(8,10). 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.

À medida que os modelos se tornam mais complexos, ou seja, com os indivíduos podendo ocupar diversos compartimentos no tempo (processo de Markov de nascimento-imigração-morte), é necessário se aplicar o modelo de intensidade multiplicativa que foi introduzido por AALEN⁽⁸⁾. Esse autor discutiu o caso não paramétrico.

Em estudos onde há seguimento pode ocorrer que alguns indivíduos não são observados até a ocorrência da falha, ou seja, têm 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 (falha) ao término do estudo, ou falecem por causas não relacionadas com a doença de interesse, ou abandonam o estudo ou são perdidos de observação. Por vezes a cura e/ou recuperação também podem ser consideradas como censura na observação. Os estudos em que existem perdas 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 .

Somente nas décadas de 50 e de 60 é que apareceram as primeiras propostas de estimadores das probabilidades de sobrevivência incorporando a censura⁽⁷⁰⁾, ou seja, modelos para observações incompletas. Um deles foi um modelo não paramétrico e foi proposto por Kaplan e Meier⁽⁴¹⁾. Na análise de sobrevivência pelo método de Kaplan e Meier a divisão de tempo não é arbitrária, mas determinada sempre que aparece uma falha (por exemplo, o óbito). Através desse modelo pode-se comparar o conjunto de curvas de sobrevida das diversas categorias de uma única variável independente. Quando um dos objetivos do estudo é a determinação de fatores de risco ou prognóstico de uma doença, deve-se realizar, primeiramente, a análise univariada pelo método de Kaplan-Meier, que está descrito com mais detalhes no capítulo II desta tese (“Métodos estatísticos em Análise de Sobrevida”).

A análise de regressão multivariada também pode ser feita na análise de sobrevivência, 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 60⁽³⁶⁾ e eram totalmente paramétricos. Em 1972, D.R. Cox⁽¹⁸⁾ desenvolveu um modelo de

regressão semi-paramétrico, também conhecido como modelo de riscos proporcionais de Cox, modelo de Cox, ou regressão de Cox. Essa técnica é indicada quando se deseja estudar sobrevivência sob o prisma de causalidade, pois fornece as estimativas das razões de risco dos fatores estudados. Também pode-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*), no modelo de Cox^(8,18,19,20,36,47,56,57), 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). Os coeficientes de regressão são estimados pelo método da verossimilhança parcial.

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 pequena modificação no modelo que Cox propôs inicialmente. Esse último modelo é conhecido como modelo de Cox com variável tempo-dependente^(8,19,36,56,57). A descrição detalhada dos modelos de Cox (convencional e com variável tempo-dependente) está no capítulo II desta tese (“Métodos Estatísticos em Análise de Sobrevida”).

O modelo de riscos proporcionais de Cox é uma técnica que está sendo cada vez mais utilizada, em diversas doenças, principalmente devido à sua disponibilidade em vários pacotes estatísticos. Como cita ANDERSEN⁽⁷⁾, em seu artigo de revisão sobre análise de sobrevivência de 1982 a 1990, no início da década de 80 a média anual de citações do trabalho de Cox era de 230. No final da década essa média passou para 540, mostrando o impacto causado por esse modelo em

trabalhos científicos da área biomédica. Nessa mesma publicação o autor chama a atenção para o cuidado que se deve ter com as análises estatísticas, pois em grande parte dos trabalhos analisados não é explicitada a metodologia, como e porquê a análise foi feita e como foram tratadas as perdas. Além disso, em poucas publicações é expressa a maneira pela qual foi feito o ajuste do modelo, sugerindo que, naqueles trabalhos onde não há essa descrição, pouco tempo foi dedicado à modelagem. Outro ponto relevante é a explicação do propósito e da estratégia de análise, mostrando quais as variáveis que foram examinadas, quais as que foram consideradas de confusão ou de interação e como foram categorizadas. Isso propicia ao leitor a oportunidade de analisar criticamente o trabalho e/ou consultar alguém que possa lhe explicar a metodologia utilizada. Vale a pena salientar que poucos autores que utilizaram o modelo de Cox na análise de seus dados fazem referência sobre a verificação se os riscos eram, realmente, proporcionais⁽⁶⁹⁾. AALEN⁽²⁾ comenta que não há garantias de que o modelo de riscos proporcionais de Cox forneça boas estimativas quando as suas suposições não estão satisfeitas.

Mais recentemente têm sido desenvolvidas diversas técnicas, sejam elas de regressão ou não, com o objetivo de se liberarem das suposições paramétricas do modelo de Cox. Esses novos métodos, geralmente, necessitam de cálculos computacionais intensivos e requerem grandes conjuntos de dados⁽⁶⁹⁾. Durante a década de 70^(7,8), alguns autores propuseram novos estimadores e testes estatísticos, tanto para a clássica análise de sobrevivência quanto para a metodologia desenvolvida por Kaplan e Meier, utilizando, fundamentalmente, a teoria de tempo contínuo de *martingale*, processos estocásticos e a teoria de processos de contagem. Durante a década de 80, esse mesmo conjunto de conhecimentos serviu para sofisticar, do ponto de vista matemático, a teoria exposta por Cox^(7,8).

Quando consultam-se livros textos de análise de sobrevida, aparecem, sempre, os métodos clássicos de análise, quais sejam, o método de Kaplan-Meier e o

modelo de riscos proporcionais, bem como a descrição de algumas distribuições que podem ser utilizadas para modelar a função de probabilidades acumuladas, como a exponencial, Weibull, log-normal e Gama. No entanto, ao se fazer a revisão de literatura em periódicos internacionais, verificou-se que a cada momento surgem novos estimadores que tentam corrigir ou fugir às suposições, ou “melhorar” o desempenho desses métodos. Porém poucos são os trabalhos que comparam as diversas metodologias em análise multivariada de sobrevida. Em que pese a sua importância, em termos de precisão matemática, essa profusão de estimadores às vezes pode mais confundir, do que ajudar a resolver os problemas de violação das suposições dos modelos. Surge a indagação: “Como resolver qual o estimador mais apropriado para um específico conjunto de dados?”. FEINSTEIN e colaboradores^(25,77,78) chamam a atenção para algumas questões:

- quando diferentes procedimentos de análise produziram mesmos resultados para os dados referentes ao mesmo fenômeno biológico?
- seriam as mesmas variáveis independentes escolhidas como preditoras, isto é, significantes estatisticamente se fossem utilizados diferentes técnicas de análise?
- as variáveis preditoras escolhidas teriam similares pesos e/ou a mesma ordenação de importância, em termos de fatores prognósticos ou de risco?
- os resultados produzidos por dois bancos de dados diferentes, referentes ao mesmo fenômeno biológico, teriam acurácia similares?

Essas questões ainda são um desafio para os que trabalham com modelagem estatística. Especificamente em relação à análise de sobrevida, há diversas propostas^(1,6,7,8,15,16,24,69,76,80), utilizando modificações no modelo de Cox, ou novas teorias como processos de contagem, *trees* e *network*. As que se referem à teoria de processos de contagem são as mais publicadas e discutidas; e a maioria dos trabalhos encontrados nessa área^(1,6,7,8,76) tentam tornar essa metodologia mais acessível ao público em geral, pois envolve uma teoria avançada de cálculo de probabilidades,

baseada na escola francesa. Essa parece ser uma metodologia promissora que incorpora não só a censura nas observações, mas o componente tempo-dependente no modelo multivariado.

Na revisão de literatura efetuada pela autora, encontrou-se apenas um estudo, o qual foi publicado em dois artigos^(1,2), que comparou essas duas técnicas: análise multivariada através do modelo de riscos proporcionais de Cox e modelo de sobrevida utilizando processos de contagem. Os trabalhos que aplicaram diversas técnicas de análise de sobrevida compararam métodos tradicionais em estatística, como se verá a seguir.

GOPE e colaboradores⁽³²⁾ compararam a tábua de vida atuarial, o modelo de Cox utilizando variável tempo-dependente ou não, a distribuição de Weibull e o modelo log-logístico aplicando-os em uma coorte de pacientes com câncer de mama. Concluem que as estimativas da probabilidade de sobrevida acumulada fornecidas pelo modelo de Cox univariado foram semelhantes às da tábua de vida atuarial. Tanto a distribuição de Weibull, como o modelo log-logístico superestimaram o número de óbitos da coorte e, por isso, não forneceram boas estimativas. O modelo de Cox utilizando variável tempo-dependente levou às melhores estimativas da função de sobrevida acumulada, sendo, portanto, recomendado em análises em que há um tempo de seguimento longo.

HARREL e colaboradores⁽³⁴⁾ compararam os resultados obtidos pelo modelo de Cox e pela análise por conglomerados (*cluster*) na determinação de um conjunto de fatores prognósticos para uma coorte de pacientes com angina que se submeteram a cateterização cardíaca por suspeita de doença coronariana. Embora a análise por conglomerados tenha se mostrado útil em uma fase inicial da análise dos dados, ela foi recomendada para estudos de sobrevida apenas quando o número de variáveis é muito grande e o tamanho da amostra pequeno. O modelo de Cox mostrou-se

adequado para estimar os fatores prognósticos e forneceu os mesmos resultados em várias sub-amostras do banco de dados original.

FEINSTEIN e colaboradores, em um trabalho publicado em três artigos consecutivos^(25,77,78), compararam diversas técnicas de análise multivariada: o modelo de Cox, análise discriminante, regressão linear e regressão logística, sempre utilizando o tempo como variável dependente. Desenvolveram esses modelos em uma coorte de pacientes com câncer de pulmão. Mostraram que todos os modelos concordaram na escolha das duas primeiras variáveis, mas na seqüência da modelagem, o modelo final de cada técnica foi diferente. Não recomendam nenhuma técnica em especial, porém consideram que análises multivariadas são úteis para detectar apenas os fatores prognósticos mais “poderosos”, ou seja, aqueles que realmente influenciam na sobrevida de qualquer paciente portador da doença em estudo. Sugerem que as associações encontradas apenas pela significância estatística, sem respaldo da literatura, não devem ser consideradas no modelo final.

ESDAILE e colaboradores⁽²²⁾, analisando fatores prognósticos em uma coorte de pacientes com nefrite lúpica, compararam a regressão de Cox em dois modelos: incorporando ou não a variável como tempo-dependente. Quando o tempo de seguimento foi curto (5 anos) ambos os modelos tiveram desempenho similares. Porém, em seguimentos longos, o modelo considerando as variáveis tempo-dependentes foi melhor para estimar a função de sobrevida acumulada. Embora os modelos não tivessem sido validados em outro conjunto de dados, o modelo de Cox convencional é bom para estimar fatores de risco para eventos de curta evolução ou seguimentos não longos. Em outras situações é preferível o modelo assumindo as variáveis tempo-dependente.

No caso específico da teoria envolvendo processos de contagem, há três trabalhos publicados que utilizaram essa classe de modelos em dois bancos de

dados. Um estudo foi o de uma coorte de mulheres que deram à luz em 1975 e o evento de interesse foi a internação em um hospital psiquiátrico em algum momento próximo ao nascimento do bebê⁽⁶⁾. A outra coorte foi de pacientes com carcinoma de oro-faringe^(1,2). Na primeira coorte, foi feito um modelo com três variáveis independentes, porém não foram comparados os resultados obtidos com aqueles do modelo de Cox. Na segunda, apenas foi feita a comparação visual entre as funções de sobrevida acumulada estimadas pelo modelo de Cox e pelo modelo incorporando a técnica de processos de contagem, não se comentando as vantagens e desvantagens de cada técnica. Na comparação visual as estimativas foram equivalentes.

É importante, pois, que se discuta e avalie a melhora que os novos estimadores fornecem à modelagem estatística quando o tempo é variável dependente, pois a utilização de modelos estatísticos em estudos epidemiológicos pode ser realizada de uma maneira sensata e acertada sempre que a metodologia empregada for adequada ao delineamento e objetivos do estudo e às perdas existentes nas observações.

Surge, então, a idéia de se estudar com mais profundidade a aplicação da teoria de processos de contagem na análise de sobrevida, pois o pesquisador se depara com muitas opções e fica difícil distinguir aquelas que são puras sofisticções matemáticas daquelas que, realmente, têm utilidade na pesquisa epidemiológica. Um dos requisitos necessários para se compararem diversos modelos estatísticos em análise de sobrevida é a existência de um conjunto de dados grande, com muitos eventos de interesse, que, assim, permita que sejam feitas diversas simulações.

A oportunidade de analisar um banco de dados com essas características surgiu do contato feito pela autora deste trabalho e o Prof. Eduardo L. Franco, que é chefe da Divisão de Epidemiologia, do Departamento de Oncologia da Universidade

McGill (Montreal - Canadá). Ele participou de uma pesquisa retrospectiva de 648 pacientes portadores de câncer de pênis atendidos no Departamento de Cirurgia Pélvica do Hospital A.C. Camargo da Fundação Antonio Prudente, em São Paulo, de 1953 à 1985, sendo essa a maior coorte de pacientes com essa localização de câncer. O questionário da pesquisa encontra-se no anexo 1. Esse banco de dados foi analisado parcialmente⁽⁴⁹⁾ - 145 casos - tendo sido feita análise da sobrevida livre da doença e total apenas para os pacientes que se submeteram a amputação peniana e linfadenectomia, com ênfase nas variáveis histopatológicas. No entanto não foi feita a análise de sobrevida para os 648 pacientes da coorte e, além disso, há a possibilidade de existência de tempo-dependência em algumas variáveis. Isso permite uma série de experimentos em modelagem estatística, exigindo a utilização de técnicas mais sofisticadas de análise. Esse é um desafio para o pesquisador que trabalha em modelagem estatística.

Especificamente em relação ao câncer de pênis, há poucos estudos de sobrevivência sobre essa patologia, geralmente com pequeno número de casos e, muitos deles, não avaliam o efeito conjunto das covariáveis, através de modelos multivariados. Isso reforça a importância de se trabalhar com um banco de dados com grande número de pacientes. Ainda há controvérsias sobre condutas e tratamento adequados. Há muitos efeitos colaterais na linfadenectomia profilática; e existem problemas psico-sexuais importantes, seja quando se realiza a penectomia parcial ou a total, principalmente em pacientes jovens. A controvérsia persiste, principalmente, pelo fato de não existirem protocolos padronizados e, até o momento, não haver consenso sobre o grupo de maior risco para o óbito em que deveria se recomendar uma terapia mais agressiva. A análise multivariada de sobrevida, através da determinação dos fatores prognósticos significantes, muito contribuiria para a delimitação destes grupos. Na revisão de literatura realizada pela autora foram encontrados apenas dois trabalhos^(38,49) que utilizaram a análise multivariada de sobrevida para definir os fatores prognósticos do câncer de pênis.

Através da análise dessa coorte pretendeu-se, fundamentalmente, comparar o modelo de riscos proporcionais de Cox (com variável tempo-dependente e não) e o modelo utilizando a teoria de processos de contagem na análise de sobrevida. Ao lado disso, o conhecimento dos fatores prognósticos para o câncer de pênis, definidos através de um modelo multivariado, forneceria subsídios para a execução de medidas de prevenção e controle da doença e promoção da saúde.

II. MÉTODOS ESTATÍSTICOS EM ANÁLISE DE SOBREVIDA

Neste capítulo são apresentados alguns métodos estatísticos em análise de sobrevida, a saber: tábua de vida atuarial para dados completos e para dados incompletos, método de Kaplan e Meier, modelo de riscos proporcionais de Cox (convencional e com variável tempo-dependente) e modelo utilizando processos de contagem. Primeiramente serão descritas as funções de sobrevida e, a seguir, como podem ser estimadas a partir da conceituação teórica de cada método, seja ele paramétrico ou não. Para facilidade, neste capítulo, o evento de interesse será considerado o óbito, porém vale a pena lembrar que o evento de interesse pode ser qualquer tipo de falha (recorrência, cura, etc).

O tempo, em análises de sobrevida, é uma variável aleatória contínua (T) não negativa ($T \geq 0$). As funções que definem a distribuição do tempo de sobrevida são: a função densidade de probabilidade [$f(t)$] - e sua respectiva função de distribuição acumulada [$F(t)$] - a função distribuição de probabilidade de sobrevida acumulada [$S(t)$] e a função de riscos [$h(t)$] - com sua função de riscos acumulada [$H(t)$].

A função densidade de probabilidade [$f(t)$], definida como:

$$f(t) = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \cdot P[t \leq T \leq t + \Delta t] \quad (1)$$

A sua função acumulada é a função distribuição acumulada [$F(t)$], definida como:

$$F(t) = P[T \leq t] = \int_0^t f(s) ds \quad (2)$$

A partir da função distribuição acumulada, define-se a função de probabilidades de sobrevida acumulada [$S(t)$]:

$$S(t) = P[T > t] = 1 - F(t) = \int_t^{\infty} f(s) ds \quad (3)$$

A função de riscos (*hazard function*), $[h(t)]$, também conhecida como força instantânea de mortalidade ou taxa instantânea de óbito em um período curto de tempo, dado que o indivíduo estava vivo até o instante t , é definida por:

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} P[t \leq T \leq t + \Delta t / T \geq t] = \frac{f(t)}{1 - F(t)} = \frac{f(t)}{S(t)} = -\frac{dS(t)}{dt} \cdot \frac{1}{S(t)} \quad (4)$$

Por último, deriva-se a função de riscos acumulada $[H(t)]$:

$$H(t) = \int_0^t h(s) ds = -\ln[S(t)] \quad (5)$$

$$\Rightarrow S(t) = \exp[-H(t)] = \exp\left(-\int_0^t h(s) ds\right) \quad (6)$$

Observa-se que todas as funções apresentadas relacionam-se entre si. Portanto basta que se conheça apenas uma, para que todas as outras sejam derivadas. É importante lembrar que essas funções são, conceitualmente, medidas instantâneas no tempo. Isso implica que, em determinado momento, apenas um óbito pode ocorrer, ou seja, não há empates. No caso de existirem muitos empates, isto é, mais de um evento em um pequeno intervalo de tempo, há pequenas variações nos estimadores e os dados são trabalhados não individualmente, mas em grupos.

Normalmente, em estudos de sobrevivência, as funções de maior interesse são a de probabilidade de sobrevida acumulada $[S(t)]$ e a de riscos $[h(t)]$ e há diversas maneiras de se estimá-las.

ALGUNS ESTIMADORES DAS FUNÇÕES DE SOBREVIDA

1. Uma suposição simples é a de que a função de riscos é constante no tempo. Nesse caso, pode-se presumir que $h(t)$ é uma constante h e tem-se, por (6), que $S(t) = \exp(-h \cdot t)$.

2. Pode-se supor que $f(t)$, ou qualquer outras das funções, tenha uma distribuição conhecida. Essa suposição pode basear-se na aparência do gráfico ou por razões teóricas. Duas das distribuições mais utilizadas em análises de sobrevida são a exponencial e a Weibull^(8,18,19,36,47,56,57,70):

• exponencial

$$f(t) = \begin{cases} \lambda e^{-\lambda t}, & \text{se } t \geq 0 \text{ e } \lambda > 0 \\ 0, & \text{se } t < 0 \end{cases} \quad (7)$$

$$\xrightarrow{2} F(t) = 1 - e^{-\lambda t}, \text{ se } t \geq 0 \quad (8)$$

$$\xrightarrow{3} S(t) = e^{-\lambda t}, \text{ se } t \geq 0 \quad (9)$$

$$\xrightarrow{4} h(t) = \lambda, \text{ se } t \geq 0 \quad (10)$$

• Weibull

$$f(t) = \begin{cases} \lambda \gamma (\lambda.t)^{\gamma-1} e^{-(\lambda.t)^\gamma}, & \text{se } t \geq 0 \text{ e } \gamma, \lambda > 0 \\ 0, & \text{se } t < 0 \end{cases} \quad (11)$$

$$\xrightarrow{2} F(t) = 1 - e^{-(\lambda.t)^\gamma}, \text{ se } t \geq 0 \quad (12)$$

$$\xrightarrow{3} S(t) = e^{-(\lambda.t)^\gamma}, \text{ se } t \geq 0 \quad (13)$$

$$\xrightarrow{4} h(t) = \lambda \gamma (\lambda.t)^{\gamma-1}, \text{ se } t \geq 0 \quad (14)$$

Por outro lado, ao se trabalhar com um método estatístico em análise de sobrevida, propõem-se estimadores para cada uma das funções de sobrevida, que serão computados a partir de um conjunto de dados. Há diversos desses métodos e não se pretende esgotá-los neste capítulo. Serão apresentados, a seguir, apenas as funções de análise de sobrevivência para a tábua de vida atuarial (com e sem censura), para o método de Kaplan-Meier, do modelo de riscos proporcionais de Cox (com e sem variável tempo-dependente) e do modelo multivariado utilizando a teoria de processos de contagem.

3. tábua de vida atuarial (dados completos) ^(36,47,56,57,70)

Os estudos envolvendo a tábua de vida demográfica iniciaram-se no século XVII com Edmund Halley^(20,70). Nesse modelo todas as observações são completas (isto é, não existe censura) e seu objetivo é fornecer o padrão de sobrevivência de indivíduos de um grupo, sujeito a coeficientes de mortalidade específicos por idade

(ou faixa etária). Através dessa técnica o tempo de seguimento é dividido em intervalos de tempo (ou de idade) fixos. Em cada intervalo x ($x=t+\Delta t$), tem-se:

d_x : número de óbitos que ocorreram no intervalo x ,

l_x : número de pessoas expostas ao risco de morrer no início do intervalo x ,

q_x : probabilidade de morrer no intervalo x , e

p_x : probabilidade de sobreviver no intervalo x .

$$\text{onde: } \begin{cases} q_x = \frac{d_x}{l_x} & (15) \\ p_x = 1 - q_x = \frac{l_x - d_x}{l_x} & (16) \end{cases}$$

A partir dessas definições estimam-se as funções de sobrevida:

$$F(T=t) = \prod_{x=0}^{t-1} q_x = \prod_{x=0}^{t-1} \frac{d_x}{l_x} \quad (17)$$

$$S(T=t) = \prod_{x=0}^{t-1} p_x = \prod_{x=0}^{t-1} \frac{l_x - d_x}{l_x} \quad (18)$$

$$h(t) = \frac{d_x}{l_x - \frac{d_x}{2}}, \text{ dado que } T > t_{x-1} \quad (19)$$

4. tábua de vida atuarial (dados incompletos) ^(36,47,56,57,70)

Somente mais recentemente é que o conceito de censura foi incorporado na tábua de vida atuarial. Nessa tábua de vida o tempo também é dividido em intervalos fixos e a probabilidade de óbito é calculada para cada um dos intervalos. Na tábua de vida atuarial com dados incompletos, isto é, na presença de censura, é feito um pequeno ajuste no número de pessoas expostas ao risco no início do período x , diminuindo-se metade das censuras do total de expostos ao risco no início do período. Nesse caso supõe-se que as censuras ocorreram uniformemente durante

o período x . Tem-se que: $l_x^* = l_x - \frac{w_x + u_x}{2}$, onde

w_x = número de pessoas que terminaram vivas o estudo, no intervalo x

u_x = número de pessoas perdidas de observação

As fórmulas das funções de sobrevivência permanecem as mesmas, porém o número de expostos ao risco no início do período x é l_x^* :

$$F(T = t) = \prod_{x=0}^{t-1} q_x^* = \prod_{x=0}^{t-1} \frac{d_x}{l_x^*} \quad (20)$$

$$S(T = t) = \prod_{x=0}^{t-1} p_x^* = \prod_{x=0}^{t-1} \frac{l_x^* - d_x}{l_x^*} \quad (21)$$

$$h(t) = \frac{d_x}{l_x^* - \frac{d_x}{2}}, \text{ dado que } T > t_{x-1} \quad (22)$$

5. método de Kaplan-Meier^(36,41,47,56,57,70)

O método de Kaplan-Meier é um caso particular da tábua de vida onde a divisão de tempo não é arbitrária, mas determinada sempre que aparece uma falha (por exemplo, o óbito). Nessa situação, o número de óbitos em cada intervalo deve ser 1.

Esse é um método não paramétrico e para calcularem-se os estimadores, primeiramente, deve-se ordenar os tempos de sobrevida em ordem crescente ($t_1 \leq t_2 \leq \dots \leq t_n$). Os sobreviventes ao tempo t (l_t) são estimados por:

$$l_t = l_0 \cdot p_t = l_0 \cdot \frac{n_j^* - 1}{n_j^*} \quad (23) \quad , \text{ onde}$$

n_j^* são os sobreviventes no tempo t_j , ajustados pela censura, ou seja, os pacientes censurados entram no cálculo da função de probabilidade de sobrevida até o momento de serem considerados como perda.

A suposição importante nas censuras é que os indivíduos que foram censurados no tempo t são representativos de todos os indivíduos que estavam sujeitos ao risco de morrer em t .

Define-se a função $S(t)$, através de um estimador que é conhecido como o estimador produto limite de Kaplan-Meier, pois é o limite do produto dos termos até o tempo t :

$$p_j = P[T > t_j / T > t_{j-1}] = \frac{l_j - i}{l_j} \quad , \text{ onde } \begin{cases} i = 1, & \text{se for falha} \\ i = 0, & \text{se for censura} \end{cases}$$

$$S(t) = \prod_{i=0}^j \frac{l_j - i}{l_j}, \text{ onde } \begin{cases} i = 1, & \text{se for falha} \\ i = 0, & \text{se for censura} \end{cases} \quad (24)$$

No caso de haver empate, o maior valor de i é utilizado na fórmula, como por exemplo:

$$\text{se } t_2 = t_3 = t_4 \Rightarrow p(t_2) = p(t_3) = p(t_4) = \frac{l_2 - 4}{l_2}$$

6. modelo de riscos proporcionais de Cox (convencional) ^(8,18,19,36,47,56,57,70)

COX⁽¹⁸⁾ propôs um modelo de regressão semi-paramétrico para estimar o efeito conjunto das covariáveis X_i ($i=1,2, \dots, k$), também conhecido como modelo de riscos proporcionais, regressão de Cox ou modelo de Cox. Esse é um modelo muito útil em estudos de fatores prognósticos, pois avalia o efeito conjunto das variáveis independentes na sobrevivência de um indivíduo. Nele a função de riscos $h(t)$ é escrita em termos das covariáveis:

$$h(t / X_1, X_2, \dots, X_k) = h_0(t) \cdot \exp(\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k) \quad (25)$$

$h_0(t)$ é a parte não paramétrica do modelo e em estudos onde 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 dividirem-se os dois lados da equação (25) por $h_0(t)$, obtém-se a equação:

$$\frac{h(t / X_1, X_2, \dots, X_k)}{h_0(t)} = \exp(\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k) \quad (26)$$

O quociente $\frac{h(t / X_1, X_2, \dots, X_k)}{h_0(t)}$ é chamado função das razões de riscos^(7,19,36) - HR(i) - (*relative hazard function*) ou de índice prognóstico⁽⁴⁾ (*prognostic index*). Essa função fornece a relação entre a *hazard* de um indivíduo do estudo comparada à de um indivíduo cuja *hazard* é $h_0(t)$ (ou seja, um indivíduo para quem $X_1=X_2=\dots=X_k=0$):

$$HR(i) = HR_i = \exp(\beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik}) \quad (27)$$

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

$$HR(X_i) = \exp(\beta_i) \quad (28)$$

Através do modelo de riscos proporcionais de Cox escrevem-se as funções de sobrevida em termos das covariáveis. Tem, então, que:

$$H(t / X_1, X_2, \dots, X_k) = H_0(t) \cdot \exp(\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k) \quad (29)$$

Como $S(t) = \exp[-H(t)]$ (6), deriva-se:

$$\begin{aligned}
 S_0(t) &= \exp(-H_0(t)) \\
 e \\
 S(t/X_1, X_2, \dots, X_k) &= (S_0(t))^{\exp(\beta_1 X_1 + \dots + \beta_k X_k)}
 \end{aligned}
 \tag{30}$$

Essas duas relações são úteis na avaliação da tempo dependência de uma variável.

Para se ter uma estimativa das probabilidades em análise de sobrevivência, é preciso que se defina uma das funções: $h_0(t)$ ou $S_0(t)$ ou $H_0(t)$. Uma estimativa não paramétrica para $H_0(t)$, baseada no modelo de Cox, é dada pela fórmula^(7,19,36):

$$H_0(t) = \sum_{t_j \leq t} \frac{d_j}{\sum_{j \in R_i} \exp(\beta_i X_{ij})} \tag{31},$$

onde R_i é o conjunto de indivíduos em risco em t_i .

A partir da determinação de $H_0(t)$ por essa fórmula, calcula-se $S(t)$ através da relação já apresentada (30).

PRENTICE & GLOECKLER⁽⁶⁵⁾ propuseram uma versão do modelo de Cox para dados agrupados, quando houver muitos empates no banco de dados em que se está trabalhando. Isso porque, também no modelo de Cox, supõe-se que, em cada momento, apenas um óbito ocorre. No entanto, quando há poucos empates o modelo convencional pode ser utilizado sem problemas.

7. modelo de Cox com riscos não proporcionais^(8,18,19,36,47,56,57,70) (modelo de Cox modificado)

Uma das suposições do modelo de Cox convencional é a de que a função de razão de riscos para uma específica variável $[HR(X_i)]$ é constante durante o tempo de seguimento. Porém em alguns estudos, principalmente em seguimentos longos, isso não acontece. Por exemplo, uma terapêutica, mesmo que constante no tempo, pode ser perigosa apenas nos primeiros meses de tratamento.

Para se verificar se a razão de riscos de uma específica variável é proporcional no tempo, deve-se estratificar o grupo de pacientes de acordo com as suas categorias e, a seguir, calcular-se $S(t)$ para cada um deles.

Como já visto anteriormente pelas fórmulas (5), (25) e (30), sabe-se que:

$$H(t / X_1, X_2, \dots, X_k) = H_0(t) \cdot \exp(\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k)$$

$$\Rightarrow \ln(H(t / X_1, X_2, \dots, X_k)) - \ln(H_0(t)) = \exp(\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k)$$

$$\Rightarrow \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k = \ln(-\ln(S(t))) \quad (32)$$

Sob a hipótese de que os riscos são proporcionais, os gráficos de $\ln(-\ln(S(t)))$ em relação ao tempo, para cada uma das categorias de uma específica variável (X_i), devem ser paralelos e a distância entre as curvas é β_i . Alguns autores^(37,54) propõem métodos para comparação dessas curvas. Porém outros^(20,36,56) comentam que essa comparação pode ser feita visualmente. Na prática

a comparação visual é muito utilizada, a não ser que se disponha de um *software* que faça este teste estatístico.

Há várias formas de se analisar, no modelo proposto por Cox, uma variável cujos riscos não são proporcionais no tempo. Deve-se, para isso, incorporar uma função do tempo associado à essa variável. As equações mais utilizadas são que incorporam o tempo, ou seu logaritmo, como uma função linear da covariável no modelo de Cox convencional:

$$h(t / X_1) = h_0(t) \cdot \exp[(\alpha_1 + \alpha_2 t)X_1] \quad (33)$$

ou

$$h(t / X_1) = h_0(t) \cdot \exp[(\alpha_1 + \alpha_2 \ln(t))X_1] \quad (34)$$

A extensão do modelo de Cox dada pela fórmula (33) é, talvez, a mais utilizada^(7,19,22,36,36) para se analisar variáveis que não têm suas funções de riscos proporcionais no tempo, provavelmente pela menor dificuldade de sua interpretação.

Alguns autores^(7,19,22,36,36), e o próprio Cox⁽¹⁹⁾ está entre eles, chamam este modelo (33) de modelo de Cox com variável tempo-dependente. Porém nesta tese ele será denominado modelo de Cox modificado.

8. variável tempo-dependente

Uma variável é considerada tempo-dependente quando seu valor muda durante o período do estudo, como por exemplo, a idade do paciente, a dose do medicamento ou o valor da pressão arterial. Isso é particularmente importante em seguimentos longos, como o que ocorre na maioria das doenças crônicas.

Embora o tratamento estatístico não seja extremamente complexo, há algumas precauções que se deve ter. Se se deseja utilizar o modelo de Cox modificado, toda variável tempo-dependente deve ser tratada como categórica, sempre. Às vezes isso não é trivial nas variáveis quantitativas. Outra situação problemática ocorre quando a variável tem várias medidas ao longo do seguimento, em intervalos irregulares de tempo. Como incorporar esse acontecimento no modelo? A análise descritiva, nesse caso, deve ser mais criteriosa, pois só dessa maneira haverá indicações da melhor forma dessa variável ser modelada.

Há várias propostas para se tratar a variável tempo-dependente e as questões levantadas em cada modelo vão depender do tipo de doença e da natureza da covariável. HARRIS & ALBERT(36) apresentam 3 soluções já discutidas em literatura: o modelo de Cox modificado, modelo de cadeias de Markov e um último modelo que combina regressão de Cox e cadeias de Markov. Cox⁽¹⁹⁾ denomina as variáveis tempo-dependentes de “evolutivas” (*evolutionary covariates*) e as analisa através do modelo de Cox modificado. Propõe que a incorporação do tempo associado à covariável pode ser feita por t , ou pela variável tempo centralizada, ou seja $(t-\mu_t)$.

Já outros autores que trabalham com processos de contagem^(1,2,6,8) sugerem que o modelo de processos de contagem para análise de sobrevida é o que melhor analisa esse tipo de variável. Argumentam que, para se utilizar o modelo de Cox para variável tempo-dependente, é necessário que se conheça o valor dessa covariável para todos os indivíduos ao risco em cada tempo de falha e há um aumento considerável do tempo de processamento dos dados. Porém a principal crítica é a dificuldade de interpretação dos resultados ao se utilizar o modelo de Cox modificado.

9. modelo utilizando processos de contagem^(1,2,8,37,54,76)

Processo de contagem⁽⁸⁾ é um processo estocástico onde há o registro de ocorrências no tempo de eventos disjuntos, isto é, discretos: $\mathbf{N} = (N_1, N_2, \dots, N_k)$, onde todos os N_i valem zero no tempo zero e vão assumindo incrementos constantes e não decrescentes de tamanho 1 (isto é, podem assumir os valores 1, 2, ...) ao longo do tempo. Dois componentes não mudam de valor simultaneamente e, em poucas situações, são infinitos. Supondo que existam n indivíduos em observação, define-se o processo de contagem $N^{(n)}$:

$N_i^{(n)} \rightarrow$ conta o número de eventos observados no i -ésimo indivíduo

e

$N^{(n)} = (N_1^{(n)}, \dots, N_n^{(n)}) \rightarrow$ um processo de contagem

Define-se uma outra variável aleatória $\mathbf{D}$ (indicador de censura) que assume o valor 1 quando ocorre falha na observação e valor 0, quando censura. Na teoria de processos de contagem é utilizado o conceito de **filtração** ($\mathcal{f}$) que é uma família de σ -álgebras definidas no espaço amostral (γ). Quando se escreve $\mathcal{f}_-$ indica que se dispõe de dados apenas até o instante t . A probabilidade do tempo de sobrevivência é definida como:

$$P(X[t, t+dt], D_i = 1 / \mathcal{f}_-) = \begin{cases} h(t) dt, & \text{se } X \geq t \\ 0, & \text{se } X < t \end{cases}$$

A partir desses conceitos são desenvolvidos estimadores para a esperança matemática e variância do número de falhas observadas em determinado período de tempo, condicionados ao que ocorreu no passado.

AALEN, primeiramente, propôs um modelo de intensidade multiplicativa, não paramétrico utilizando processos de contagem, e, depois, o generalizou incorporando as covariáveis^(1,2,8). Esse é um modelo linear que é assim definido:

Supõe-se que cada $N^{(n)}$ tem um processo de intensidade aleatório $\lambda^{(n)} = (\lambda_1^{(n)}, \dots, \lambda_n^{(n)})$. Esse processo aleatório equivale ao $h(t)$, que vem sendo apresentado até o presente momento. Optou-se, nessa descrição, por seguir a notação adotada nos livros/artigos sobre processos de contagem para evitar fazer-se adaptações.

No modelo de análise de sobrevida utilizando processos de contagem, define-se o processo de intensidade como:

$$\boxed{\lambda_i^n(t) = Y_i^n(t) \cdot \lambda_0(t) \cdot \exp(\beta_0^* X_i^n(t))} \quad (35), \text{ onde:}$$

$\lambda_i^n(t) \rightarrow$ processo de intensidade (equivalente ao $h(t)$)

$$Y_i^n(t) = \begin{cases} 1, & \text{se o individuo e obito em } t \\ 0, & \text{se o individuo e censura em } t \text{ (ie, nao obito em } t) \end{cases}$$

$X_i^n(t) = (X_{i1}^{(n)}, \dots, X_{ik}^{(n)})$, $\rightarrow$ é um vetor coluna com as k covariáveis do processo de contagem para o i -ésimo indivíduo.

β_0^* $\rightarrow$ é um vetor coluna dos k coeficientes das k covariáveis do processo.

Esse é o modelo que será testado nesta tese (35).

É interessante ressaltar que o componente $\exp(\beta_0^* X_i^n(t))$ da equação (35) varia de acordo com o tempo. E essa é a parte atraente do modelo de análise de sobrevida que se baseia na teoria de processos de contagem, pois, automaticamente, fica incorporada a tempo-dependência no modelo.

Não serão dados detalhes do processo de estimação dos coeficientes de regressão, mas dizer que $N^{(n)}$ é um processo de contagem com processo de intensidade $\lambda^{(n)}$, significa que é necessário se definir um processo $M_i^{(n)}$ (que é considerado localmente *martingale*), tal que:

$$M_i^{(n)}(t) = N_i^{(n)}(t) - \int_0^t \lambda_i^{(n)}(u) du, i = 1, \dots, n \text{ e } t \in [0, 1] \quad (36)$$

Essa equação permite que se faça uma generalização da equação do modelo de Cox (25) utilizando processos de contagem :

$$\lambda_i^{(n)}(t_+) = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \cdot P\left[N_i^{(n)}(t + \Delta t) - N_i^{(n)}(t) = 1 / f_t^{(n)}\right] \quad (37)$$

$$\Rightarrow \lambda_i^n(t) = Y_i^n(t) \cdot \lambda_0(t) \cdot \exp(\beta_0^* X_i^n(t)) \quad (35) \text{ , já definida}$$

- componente de suscetibilidade individual (*frailty*)^(7,8,53)

Há duas suposições implícitas à todos os modelos/métodos apresentados até agora: independência entre os indivíduos e homogeneidade à todos os indivíduos para um específico valor de uma variável independente. Porém o risco de óbito de cada indivíduo pode ser decomposto em duas partes. Uma é um componente comum a todos indivíduos do mesmo grupo de risco e outro é um componente intrínseco, desconhecido, que é devido à suscetibilidade individual, conhecido por *frailty*. Essa “fragilidade” é uma característica de cada pessoa e pode explicar o comportamento diferente de um grupo de pessoas sujeitos a um mesmo grupo de fatores de risco, diante das diversas causas de morte.

Esse é um conceito relativamente recente e só foi desenvolvido teoricamente. No momento estão começando os estudos⁽⁵³⁾ que incorporam esse conjunto teórico em doenças transmissíveis. Parece interessante que esse componente de heterogeneidade individual seja incorporado aos estimadores de análise de sobrevida. No entanto este assunto não será abordado nesta tese.

III. O CÂNCER DE PÊNIS

EPIDEMIOLOGIA DESCRITIVA

O câncer de pênis é uma doença rara, representando cerca de 2 a 5% de todos os cânceres urogenitais⁽²³⁾. Nos Estados Unidos o coeficiente de incidência anual é de 0,2 por 100.000 homens⁽⁴⁸⁾, representando de 0,4 a 0,6% de todos os casos de câncer do sexo masculino⁽¹²⁾. Pelas estimativas feitas pela *American Cancer Society*⁽¹³⁾ esperavam-se 1100 casos novos e 210 óbitos de câncer de pênis nos Estados Unidos, em 1995. Essa localização de câncer passa a ter maior importância onde não há boas condições de higiene ou em comunidades onde a circuncisão não é feita de rotina. Em algumas áreas da Ásia, América do Sul e África, essa localização de câncer representa cerca de 10 a 20% dos cânceres do sexo masculino⁽¹²⁾. No Paraguai é o mais freqüente câncer do trato geniturinário (45 a 76% dos casos)⁽²³⁾. Em Uganda o coeficiente de incidência (1964/1968) é de 2,2 por 100.000 homens, sendo essa a neoplasia maligna mais comum no sexo masculino (12%)⁽⁷³⁾. No Brasil a porcentagem de óbitos por câncer de pênis está entre as dez primeiras localizações dentre o total de neoplasias malignas, nas regiões Norte, Nordeste e Centro-Oeste, o mesmo não ocorrendo nas regiões Sul e Sudeste, onde o óbito por essa doença é raro⁽⁵⁸⁾.

A tabela 1 e o gráfico 1 apresentam os coeficientes padronizados de incidência anual para algumas localidades, que foram publicados no sexto volume do *Cancer Incidence in Five Continents*⁽⁶⁴⁾, com exceção de algumas localidades brasileiras que são de outras fontes.

Os maiores coeficientes padronizados de incidência referem-se à Recife, Belém e Fortaleza, seguidos de Porto Alegre e Porto Rico. Outros coeficientes também altos são os de Mandras (Índia), Chiang Mai (Tailândia) e São Paulo. Comparando-se com a maioria das populações européias, Estados Unidos, Canadá, comunidades japonesas e da Oceania, onde os coeficientes padronizados estão

sempre abaixo de 1 por 100.000 homens, verifica-se que as taxas encontradas nas populações brasileiras citadas são cerca de 10 vezes superiores às dos países mais desenvolvidos. Vale salientar que Recife possui a maior incidência de câncer de pênis no mundo. O menor coeficiente de incidência anual padronizado de câncer de pênis é o dos judeus em Israel (0,1 por cem mil homens).

Tabela 1 - Coeficientes padronizados* de incidência anual (por 100000 homens) de câncer de pênis, segundo área e período.

Área	Período	Coeficiente padronizado de incidência anual (por 100.000 homens)
ÁFRICA		
• Gambia	1987/1989	0,9
AMÉRICA		
• Goiânia (Brasil)	1988/1989	2,2
• Porto Alegre (Brasil)	1987	3,4
• Belém (Brasil) ⁽⁵⁸⁾	1987	5,2
• Recife (Brasil) ⁽⁵⁸⁾	1980	8,3
• Fortaleza (Brasil) ⁽⁵⁸⁾	1983	5,7
• São Paulo (Brasil) ⁽⁵⁹⁾	1978	2,5
• Porto Rico (EUA)	1983/1987	3,0
• SEER Program brancos	1983/1987	0,6
negros		0,5
• Canadá	1983/1987	0,7
ÁSIA		
• Shanghai (China)	1983/1987	0,3
• Madras (Índia)	1983/1987	2,7
• Israel (todos os judeus)	1982/1986	0,1
• Osaka (Japão)	1983/1987	0,3
• Chiang Mai (Tailândia)	1983/1987	2,8
EUROPA		
• Saarland (Alemanha)	1983/1987	0,6
• Inglaterra e País de Gales	1983/1986	0,7
• Torino (Itália)	1985/1987	0,8
• Cracóvia (Polônia)	1983/1986	0,8
• Genebra (Suíça)	1983/1987	0,6
• Vila Nova de Gaia (Portugal)	1983/1987	1,5
OCEANIA		
• New South Wales (Austrália)	1983/1987	0,5

Fonte: referência 64

* população padrão utilizada é a população mundial de 1960.

Gráfico 1 - Coeficientes padronizados de incidência anual de câncer de pênis (por 100000 homens), segundo área e período

A tabela 2 e o gráfico 2 apresentam os coeficientes de mortalidade padronizados de câncer de pênis para algumas localidades do Brasil. As cidades do Norte e Nordeste apresentam os maiores coeficientes de mortalidade, muito embora as taxas de mortalidade por câncer peniano sejam bem menores do que as taxas de incidência. Novamente Recife é a cidade que apresenta o maior coeficiente padronizado. Não se encontrou referência de mortalidade por câncer de pênis em países desenvolvidos para se comparar com nossas taxas, muito provavelmente por ser essa uma causa de morte muito rara nessas localidades.

Tabela 2 - Coeficientes padronizados* de mortalidade anual (por 100000 homens) de câncer de pênis, segundo área e período.

Área	Período	Coeficiente padronizado de mortalidade anual (por 100.000 homens)
• Belém ⁽⁵⁸⁾	1987	0,7
• Fortaleza ⁽⁵⁸⁾	1983	1,3
• Recife ⁽⁵⁸⁾	1980	2,2
• São Paulo ⁽⁶⁰⁾	1978	0,5
• Goiânia ⁽⁵⁸⁾	1988	0,2
• Porto Alegre ⁽⁵⁸⁾	1987	0,1

* população padrão utilizada é a população mundial de 1960.

Gráfico 2 - Coeficientes padronizados de mortalidade anual de câncer de pênis (por 100000 homens), segundo área e período

ETIOLOGIA

O aparecimento de câncer de pênis está associado à ausência de circuncisão^(12,39,50,79), maus hábitos de higiene^(11,23,40,50,61) e classe social mais baixa⁽³⁹⁾.

O efeito irritativo do esmegma também está relacionado a um maior risco de câncer de pênis⁽¹¹⁾.

Essa localização de câncer é rara em judeus e em muçulmanos (que praticam a circuncisão entre 3 e 13 anos)⁽³⁰⁾. WISWELL⁽⁷⁹⁾ recomenda a circuncisão profilática, pois diminui o risco para diversas doenças, tais como infecções do trato urinário, doenças sexualmente transmissíveis (DST), infecções penianas e câncer de pênis e de útero. Embora muitos tenham receio dos efeitos colaterais da circuncisão, esse autor comenta que desde o início da década de 50 ocorreram apenas três óbitos por problemas relacionados à esse procedimento contra mais de 7000 óbitos por câncer de pênis. Desses óbitos por neoplasia, 95% poderiam ter sido evitados se o homem fosse circuncisado. Por outro lado, como o carcinoma de pênis é muito raro, haveria a necessidade de serem feitas muitas circuncisões, comparativamente ao número de casos. Por exemplo, JENSEN⁽³⁹⁾ cita que na Dinamarca seriam necessárias 1000 circuncisões para prevenir um homem de adquirir câncer de pênis antes dos 64 anos.

BURGERS e colaboradores⁽¹²⁾ e LUCIA & MILLER⁽⁵⁰⁾ referem que a idade média dos pacientes está sempre entre as quinta e sexta décadas, apesar de existirem casos de jovens com esse tipo de neoplasia. Cerca de 22% dos pacientes têm menos de 40 anos e aqueles com menos de 30 anos não chegam a 10%. FRALEY e colaboradores⁽²⁷⁾, em estudo nos Estados Unidos, citam que a idade de seus pacientes variou de 53 a 89 anos (média de 71 anos). Em estudo na Índia, KAMAT e colaboradores⁽⁴⁰⁾, encontraram pacientes mais jovens (idades entre 19 e 84 anos), sendo que 84% deles tinham entre 40 e 70 anos.

Nos Estados Unidos, alguns estudos^(50,73) apresentam razão entre raças (negro:branco) de 2 para 1, porém acredita-se que as condições sócio-econômicas em que vivem os negros é que favorecem uma maior incidência nesse grupo étnico.

DALING e colaboradores⁽²¹⁾ referem que o cigarro pode atuar como um fator promotor na malignidade do tumor e mostraram que o fumo pode ser um fator de risco para o câncer de pênis. Em estudo caso-controle realizado por esses autores encontrou-se que fumantes, a partir dos 60 anos, tinham 4,2 vezes a chance de ter câncer de pênis quando comparado com os não fumantes ($IC_{95\%} (OR)=[1,4 ; 12,2]$), risco esse ajustado por idade, número de parceiros e procedência do paciente. Este foi o único estudo em que se encontrou referência ao fumo como fator de risco para câncer de pênis.

A associação entre o câncer de pênis e doenças sexualmente transmissíveis tem sido estudada^(11,39,61), porém a relação etiológica ainda não é consistente.

A etiologia viral do carcinoma peniano tem sido sugerida⁽³³⁾. Existem alguns estudos que procuram associar o câncer de pênis com o câncer de colo uterino, mostrando tanto a alta correlação positiva entre as frequências relativas de câncer de pênis e o de câncer de colo uterino dentre o total de neoplasias⁽²⁸⁾, bem como o aumento de risco de desenvolvimento de câncer de colo uterino em esposas de pacientes com câncer de pênis⁽⁷³⁾. Isso pode sugerir que as duas localizações de câncer possam ter um agente comum, no que se refere à doença sexualmente transmissível, provavelmente o vírus papiloma humano (HPV).

Alguns estudos mencionam a presença do HPV ou a infecção por herpes em pacientes com câncer de pênis, porém os resultados ainda são inconclusivos⁽²³⁾. LUCIA & MILLER⁽⁵⁰⁾ citam que em, aproximadamente, 50% dos carcinomas epidermóides de pênis foi possível identificar a presença de HPV 16 ou 18. VILLA & LOPES⁽⁷⁵⁾, no Brasil, encontraram a presença de HPV 18 em 7 dos 18 pacientes analisados. Já TORNESELLO e colaboradores⁽⁷³⁾, em Uganda, acharam presença

de sequência de HPV em todos os 13 pacientes com câncer de pênis, sendo possível identificar HPV 16 em apenas um.

HISTOLOGIA

A grande maioria dos cânceres de pênis é constituída por carcinoma de células escamosas (cerca de 95% dos casos)^(12,23,30,50), bem diferenciado. Outros tipos histológicos como os sarcomas (54% dentre os de células não escamosas)⁽¹²⁾, melanoma maligno e neoplasmas de tecidos moles também são descritos.

As lesões pré-cancerosas descritas são^(23,33,39): leucoplasia, eritroplasia de Queryat, doença de Bowen, balanitis xerótica obliterante e tumor de Buschke-Lowenstein. JENSEN⁽³⁹⁾ refere que os pacientes registrados no Registro de Câncer da Dinamarca e que apresentavam lesões pré-cancerosas, possuíam essas lesões, em média, sete anos antes de serem diagnosticados.

Com o advento da AIDS, houve um aumento do número indivíduos com Sarcoma de Kaposi. Cerca de 20% dos pacientes aidéticos com Sarcoma de Kaposi apresentam lesão no pênis^(12,50).

A metástase geralmente ocorre nos linfonodos inguinais, embora raramente possam ocorrer metástases à distância, como no cérebro, pulmão, fígado, pele e osso^(12,40).

ESTADIAMENTO

O correto estadiamento da doença é difícil de ser feito através do exame clínico^(9,12). Estudos mostram que o exame de palpação dos linfonodos tem de 70 a 85% de sensibilidade e 50% de especificidade na capacidade de detectar a presença de metástases nos gânglios linfáticos inguinais. SRINIVAS e colaboradores⁽⁷¹⁾

encontraram acurácia de 85% para o exame clínico. LOPES⁽⁴⁹⁾, analisando pacientes que foram submetidos a amputação peniana e linfadenectomia, encontrou sensibilidade de 66,7% e 52,3% de especificidade. Como há muita dificuldade em se distinguir os pacientes em estadiamento I e II, alguns advogam a linfadenectomia profilática para aumentar a especificidade. ABI-AAD e deKERNION⁽³⁾ fizeram uma revisão de literatura mostrando a controvérsia existente sobre esse procedimento.

Existem diversas classificações utilizadas para se definir o estadiamento, porém não há padronização entre os estudos. Por isso a comparação das séries, no que diz respeito às análises de sobrevida ou entre tratamentos, é árdua. As mais utilizadas são a classificação de Jackson e o TNM da União Internacional Contra o Câncer (UICC), que possui duas versões (1978 e 1987).

A classificação de Jackson é a mais antiga (1966), por isso, talvez, a mais utilizada. No entanto caracteriza, com detalhes, apenas o tumor primário e a presença de metástase e não o envolvimento de linfonodos,. De acordo com essa classificação os estadiamentos de tumor são assim definidos^(23,42,61):

- estadio I: doença limitada à glândula e/ou prepúcio.
- estadio II: tumor invadindo o corpo do pênis sem metástases nos linfonodos inguinais ou à distância.
- estadio III: metástases ressecáveis para os linfonodos regionais.
- estadio IV: tumor invadindo estruturas adjacentes ou metástases nos linfonodos regionais ou metástases à distância.

Na classificação TNM⁽⁷⁴⁾ de 1978, adotada pela União Internacional Contra o Câncer (UICC), levam-se em conta três características: a extensão do tumor primário (T), a presença ou ausência de metástases em linfonodos regionais e a extensão

deste comprometimento (N) e a presença ou ausência de metástases à distância (M). A classificação é a seguinte (anexo2):

T: tumor primário

- T₀: sem evidências de tumor primário;
- T₁: tumor de 2 cm ou menos em sua maior dimensão, estritamente superficial ou exofítico;
- T₂: tumor com mais de 2 cm e até 5 cm em sua maior dimensão ou tumor com extensão mínima;
- T₃: tumor com mais de 5 cm em sua maior dimensão, ou tumor com extensão profunda incluindo a uretra;
- T₄: tumor infiltrando estruturas vizinhas; e
- T_X: o tumor primário não pode ser avaliado por falta de condições mínimas.

N: linfonodos regionais

- N₀: sem evidências de envolvimento de linfonodos regionais;
- N₁: evidências de envolvimento unilateral de linfonodos regionais móveis;
- N₂: evidências de envolvimento bilateral de linfonodos regionais móveis;
- N₃: evidências de envolvimento de linfonodos regionais fixos; e
- N_X: os linfonodos regionais não puderam ser avaliados por falta de condições mínimas.

M: metástases à distância

- M₀: sem evidências de metástases à distância;
- M₁: evidências de metástases à distância; e
- M_X: a metástase à distância não pode ser avaliada por falta de condições mínimas.

Não se recomenda que sejam feitos agrupamentos entre as categorias T, N e M.

A versão de 1978 do TNM é mais utilizada do que a de 1987, talvez pelo fato das séries existentes serem antigas. HORENBLAS & van TINTEREN⁽³⁸⁾, comparando as duas versões do TNM mostram que a nova versão (1987) é essencialmente anátomo patológica e não auxilia a caracterizar melhor o estadiamento clínico.

MAYCHE & PYRHÖNEN⁽⁵¹⁾ compararam as classificações de 1978 e de 1987 da UICC, a de Jackson e uma terceira proposta por eles e mostraram que o resultado em relação à sobrevida é diferente, de acordo com o sistema de classificação utilizado. No que se refere à comparação entre as duas versões do TNM, esses autores concordaram com HORENBLAS & van TINTEREN⁽³⁸⁾, ou seja, a nova versão não privilegia o estadiamento clínico. Além disso, em cerca de 35% dos pacientes registrados no Registro de Câncer da Finlândia não se conseguiu fazer a classificação pelo TNM. Os autores finlandeses não recomendam a classificação de Jackson, pois essa não leva em conta o tamanho do tumor, importante fator prognóstico para a sobrevida de pacientes portadores de carcinoma de pênis.

TRATAMENTO

Os fatores que definem a conduta no tratamento do carcinoma peniano são o tamanho e localização do tumor, extensão da infiltração e grau de destruição tumoral do tecido normal⁽³³⁾. O tratamento dos tumores infiltrativos do pênis é, essencialmente, cirúrgico. Existem muitas controvérsias sobre o tratamento dos linfonodos devido à dificuldade de definir o estadiamento.

Devido à raridade do câncer de pênis, as séries existentes são pequenas, ou reportam pacientes diagnosticados há mais de 20 anos e não existem ensaios clínicos padronizados comparando métodos de tratamento. A ausência de dor e os preconceitos oriundos de uma doença em órgãos genitais, fazem com que os pacientes demorem a procurar atendimento médico (geralmente mais de um ano após o aparecimento dos sintomas^(12,23,40)). Quando o fazem, já se encontram em estadio avançado da doença. NATH e colaboradores⁽⁶¹⁾, descrevendo pacientes atendidos no Hospital Universitário de Lasaka, Zambia, referem que muitos procuram tratamentos alternativos antes de ir ao médico, devido aos preconceitos oriundos de uma doença em órgão genital. Pacientes em estadio III ou IV apresentam metástases nos gânglios linfáticos inguinais, o que diminui sensivelmente as chances de sobrevivência.

Outro fator para a demora para o início do tratamento é a dificuldade, às vezes, em se fazer o diagnóstico inicial pela presença de diversas doenças venéreas concomitantemente. Isso dificulta a visualização do câncer de pênis^(12,66).

O adequado controle da doença primária pode levar a cura, porém se este câncer não for diagnosticado precocemente sua letalidade é alta. A amputação peniana com 2 cm de margem é o método adequado para o controle do tumor primário infiltrativo^(23,26,49,55). Tumores localizados no prepúcio podem ser tratados através da circuncisão⁽²⁶⁾. A cirurgia é o principal método de tratamento dos linfonodos regionais.

No entanto, se o paciente não está motivado, provavelmente não fará o seguimento adequado, e, por isso, alguns autores^(45,63) recomendam a penectomia parcial ou total mesmo em estadios não avançados da doença.

Em estadiamentos mais avançados da doença sempre se recomenda cirurgia mais agressiva⁽⁴⁸⁾, porém muitos dos pacientes se recusam a fazê-la⁽⁶¹⁾. Esse tipo de

tratamento pode acarretar problemas psico-sexuais importantes, principalmente no primeiro ano após o tratamento e em pacientes jovens.

Segundo FAIR e colaboradores⁽²³⁾ de 40 a 50% dos pacientes com linfonodos positivos podem ficar livres da doença após a linfadenectomia. No estadiamento I da doença, observa-se cerca de 20% de falsos-negativos⁽²³⁾, não se recomendando a linfadenectomia profilática. Já no estadiamento II, cerca de 60% dos pacientes com nódulo positivo apresentam comprometimento bilateral pouco tempo depois.

O emprego da linfadenectomia profilática é assunto de discussão^(3,9,23,26,27,30,40,48,50,62,63,81). A controvérsia existente diz respeito à decisão a ser tomada na presença de linfonodos clinicamente negativos: linfadenectomia profilática ou conduta expectante. Embora possa controlar a doença nos linfonodos, a morbidade decorrente da cirurgia é muito alta⁽²⁷⁾ e cerca de 10% dos pacientes morrem devido a problemas com a cirurgia⁽³⁾. O problema com a conduta expectante é que, conforme já citado, a especificidade do exame clínico é muito baixa (muito falso negativo) e quando aparece a metástase a sobrevida é menor. ABI-AAD & deKERNION⁽³⁾, após revisão dos estudos em literatura, recomendam a linfadenectomia profilática quando:

- o tumor primário está invadindo o corpo cavernoso ou é histologicamente indiferenciado;
- linfonodos permanecem palpáveis mesmo depois do controle do tumor primário e de antibioticoterapia; e/ou
- não ha confiabilidade de que o seguimento do paciente será adequado.

É importante salientar que a maioria das séries disponíveis são muito antigas. Os estudos em coortes mais recentes mostram que o ganho na sobrevida com a

linfadenectomia profilática não é significativa, ao contrário do que indicavam as séries mais antigas.

Nas lesões infiltrativas a cirurgia é o tratamento de eleição. A radioterapia pode ser utilizada em lesões superficiais, com a vantagem, óbvia, de manutenção do pênis. Embora haja benefícios do ponto de vista psicológico do paciente, há muitos efeitos adversos, como a estenose da uretra, efeito tóxico, bem como maior chance de recorrência da doença⁽³⁰⁾. Cerca de 70 a 80% dos pacientes de estádios I a III tratados com radioterapia obtêm sucesso com a terapia, mas esse resultado deve se avaliado com cautela. Pelo fato de existirem diversos critérios de estadiamento, há muitas dificuldades na comparação dos resultados entre as séries⁽²³⁾. FAIR e colaboradores⁽²³⁾ e GERBAULET & LAPIN⁽³⁰⁾ comentam que 20% dos pacientes podem ter recorrência da doença após cinco anos da radioterapia. KAMAT e colaboradores⁽⁴²⁾ recomendam o emprego da radioterapia associada com múltiplas drogas, principalmente em lesões superficiais em pacientes jovens⁽⁴⁰⁾.

O efeito da quimioterapia isoladamente tem sido pouco avaliado e é necessário que se realizem protocolos padronizados para recomendar uma terapia mais adequada^(23,42).

SOBREVIDA

O fator prognóstico mais importante, segundo a literatura^(12,23,27,30,38,39,40,62,67,71,72,81), é o estadiamento da doença, principalmente no que se refere à presença ou não de metástases nos linfonodos regionais. Conforme já salientado, a dificuldade que se depara na comparação dos estudos que analisam a sobrevida de pacientes portadores de câncer de pênis é a falta de um critério único para definir estadiamento.

A tabela 3 apresenta os resultados de diversos estudos que avaliaram a sobrevida global de pacientes com câncer de pênis.

Tabela 3. Taxas de sobrevida de diversas coortes de pacientes com câncer de pênis, segundo características do estudo.

Autor(es)	local	ano	nº de pacientes	condição do paciente	taxa de sobrevida (%)	
					5 anos	10 anos
Fraley e col ⁽²⁷⁾	EUA	1971/1985	11	estadio I, sem linfad.	100	...
			4	estadio I, com linfad.	91	...
			12	estadio II, com demora na linfad.	8	...
			9	estadio II, com linfad.	100	...
			8	estadio II, sem linfad.	25	...
			9	estadio III, com linfad.	75	...
			5	estadio IV	0	...
Gerbaulet e col ⁽³⁰⁾	França	...	109	total	74	52
				linfonodos negat.	82	59
				linfonodos posit.	36	18
Horenblas & van Tinteren ⁽³⁸⁾	Holanda	1956/1989	118	estadio I	93	...
				estadio II+III	55	...
				estadio IV	30	...
Jensen ⁽³⁹⁾	Dinamarca	1942/1962	511	total	68	67
Kamat e col ⁽⁴⁰⁾	Índia	1984/1987	150	T ₁	83,3	...
				T ₂	60,4	...
				T ₃	39,5	...
				N ₀	75,4	...
				N ₁	40,0*	...
				N ₂	38,8*	...
				Grau ₁	90,3	...
				Grau ₂	58,6	...
Lopes ⁽⁴⁹⁾	São Paulo	1953/1985	145	Grau ₃	16,7	...
				total	54,3	54,3
				T ₁ +T ₂	61,5	61,5
				T ₃	46,5	46,7
Ravi ⁽⁶⁷⁾	India	1962/1986	201	T ₄	85,7	85,7
				linfonodos negat.	95	...
				linfonodos posit.	76	...
				metást. unilat.	86	...
				metást. bilat.	60	...
Srinivas e col. ⁽⁷¹⁾	New York	1950/1980	199	gâng. pélvico pos.	0	...
				linfonodos negat.	74	...
				1 nódulo positivo	82	...
				2 a 6 nódulos posit.	54	...
				> 6 nódulos posit.	40	...
				nódulos pos. bilat.	12	...
				doença extra nodal	4,5	...
Sufrin & Huben ⁽⁷²⁾	EUA	...	...	total	52	...
				linfonodos negat.	66	...
				linfonodos posit.	27	...
Young e col. ⁽⁸¹⁾	Illinois	1958/1983	34	estadio I	77	...

* sobrevida após 3 anos do diagnóstico

Observa-se que os pacientes com taxas de sobrevida após cinco anos de diagnóstico maiores que 80% são aqueles que se apresentam em estadio I da doença ou sem a presença de linfonodos bilaterais. Apenas LOPES⁽⁴⁹⁾ encontrou taxa de sobrevida global próxima a 86% para pacientes em estadiamento T₄, porcentagem esta superior aos outros grupos de estadiamento em seu próprio estudo; no entanto nesse grupo havia apenas 7 pacientes. No estudo de SUFRIN & HUBEN⁽⁷²⁾, as taxas de sobrevida após 5 anos foram bem abaixo do que se encontrou descrito na literatura.

SRINIVAS e colaboradores⁽⁷¹⁾, em estudo retrospectivo de 199 pacientes (1950/1980) com câncer de pênis na Índia, encontraram uma taxa de sobrevida, após 5 anos de diagnóstico da doença, de 56% para os pacientes com envolvimento unilateral de nódulos inguinais, decaindo para 9% quando o comprometimento era bilateral. Em outro estudo na Índia (1962/1986), RAVI⁽⁶⁷⁾ verificou que os fatores prognósticos adversos eram o envolvimento de três ou mais linfonodos, ruptura capsular e comprometimento de nódulos pélvicos.

YOUNG e colaboradores⁽⁸¹⁾ não encontraram diferença estatisticamente significativa entre as curvas de sobrevida de pacientes em estadio I e II (critério de M.D. Anderson), mesmo quando submetidos à linfadenectomia profilática. Todos os pacientes de estadiamento II morreram após 24 meses de ressecção do linfonodo.

Poucos são os estudos que fazem a análise de sobrevivência multivariada. Na revisão de literatura foram encontrados apenas dois trabalhos.

HORENBLAS & van TINTEREN⁽³⁸⁾, analisando os casos de câncer de pênis do *The Netherlands Cancer Institute* no período de 1956 a 1989, identificaram que pacientes com pequenos tumores, sem linfonodos metastáticos e tumor bem

diferenciado mostram um prognóstico favorável. Na análise multivariada apenas a categoria N e o grau de diferenciação do tumor se mostraram importantes fatores prognósticos. Embora refiram ter utilizado o modelo de riscos proporcionais de Cox na análise multivariada, nos resultados só apresentam os coeficientes estimados: 1,0992 para N ($p < 0,0001$) e 0,9467 para o grau de diferenciação do tumor ($p = 0,0001$). Aparentemente essas variáveis foram analisadas como contínuas, no modelo de Cox, não permitindo avaliar, corretamente, as estimativas das razões de risco para cada uma das categorias.

LOPES⁽⁴⁹⁾, estudando apenas pacientes que foram submetidos à amputação e linfadenectomia, encontrou na análise multivariada pelo modelo de Cox, que os fatores importantes para o risco de óbito foram: presença de linfonodos metastáticos, estadio clínico T (apenas a categoria T₃ foi de risco) e a infiltração eosinofílica (a categoria “moderado” foi significativa). A HR da presença de linfonodos metastáticos foi de 2,7 {IC(HR)=[1,5;4,7]}, quando comparado com a ausência de linfonodos metastáticos. Um paciente em T₃ tem HR de 2 {IC(HR)=[1,1;3,6]}, quando comparado com pacientes em T₁+T₂. As HR's dos pacientes em T₄ e T_x não foram estatisticamente significantes. Seus valores foram, respectivamente, 0,5 {IC(HR)=[0,1;3,5]} e 0,8 {IC(HR)=[0,2;2,9]}. As presenças de infiltrado eosinofílico leve, moderado, acentuado e ignorado tiveram HR de 0,6 {IC(HR)=[0,3;1,1]}, 2,1 {IC(HR)=[1,0;4,3]}, 0,2 {IC(HR)=[0,1;0,9]} e 1,2 {IC(HR)=[0,3;5,0]}, quando comparadas com a ausência de infiltrado eosinofílico.

IV. HIPÓTESE DO TRABALHO

- O modelo de riscos proporcionais de Cox é uma técnica de análise multivariada útil para estudos de fatores prognósticos. A inclusão da variável tempo-dependente e a observância dos pré-requisitos técnicos do modelo aprimoram suas estimativas.

V. OBJETIVOS

GERAL

Comparar o desempenho do modelo de riscos proporcionais de Cox convencional, com incorporação de variáveis cujas razões de risco não são proporcionais no seguimento e o modelo multivariado de sobrevida que utiliza a teoria de processos de contagem.

ESPECÍFICOS

1. Comparar as estimativas fornecidas pelo método de Kaplan-Meier, modelo de riscos proporcionais de Cox convencional, modelo de riscos proporcionais de Cox modificado e o modelo multivariado de sobrevida que utiliza a teoria de processos de contagem, aplicando estas metodologias em uma coorte de pacientes com câncer de pênis.
2. Validar os resultados obtidos no modelo de riscos proporcionais de Cox convencional em três amostras do banco de dados original.
3. Estudar os fatores prognósticos do câncer de pênis.

VI. MATERIAL E MÉTODOS

VI. 1. População de estudo

A comparação entre os diversos métodos estatísticos de análise de sobrevivência foi feita utilizando-se uma coorte de 648 pacientes de carcinoma espinocelular primário de pênis, atendidos no Departamento de Cirurgia Pélvica do Hospital A.C. Camargo da Fundação Antônio Prudente, em São Paulo, de 1953 a 1985. Os coordenadores desta pesquisa foram o Prof. Eduardo L. Franco, que atualmente é Professor Titular na Divisão Epidemiologia do Departamento de Oncologia da McGill University (Montreal-Canadá) e o Dr. Ademar Lopes que, atualmente, é chefe do Departamento de Cirurgia Pélvica do Hospital A.C. Camargo da Fundação Antônio Prudente. Esse banco de dados foi parcialmente analisado, conforme já referido. Em 1995 o Dr. Ademar Lopes⁽⁴⁹⁾ defendeu sua tese de Doutorado na Faculdade de Medicina da Universidade de São Paulo analisando a sobrevida livre da doença e total dos pacientes que se submeteram à amputação peniana e linfadenectomia (145 casos), com ênfase nas variáveis histopatológicas.

O questionário utilizado para essa pesquisa encontra-se no Anexo 1.

VI. 2. variáveis em estudo

variável dependente

Foi o tempo decorrido entre a data da primeira consulta e o óbito do paciente.

variáveis independentes e respectivas categorias

- variáveis referentes às características sócio-demográficas

- * idade: analisada em três faixas etárias (18 a 39 anos, 40 a 59 e 60 a 87 anos).

- * raça: branco e não branco.
- * estado civil: casado e não casado.
- * zona de residência: rural e urbana.
- * procedência: regiões sudeste (SE), sul (S), nordeste (NE) e norte e centro-oeste (N+CO). As regiões norte e centro-oeste constituíram-se em uma só categoria pelo pequeno número de casos.
- * ano de início de tratamento (ANO): dividido em décadas (1953 a 1959, 1960 a 1969, 1970 a 1979 e 1980 a 1985).

- variáveis clínicas e sintomas pré-tratamento

- * doença venérea: antecedentes de doença venérea (sim/não).
- * tratamento não oncológico prévio: sim/não.
- * tratamento oncológico prévio: sim/não.
- * sintoma - ferida: sim/não.
- * sintoma - dor: sim/não.
- * sintoma - inflamação: sim/não.
- * sintoma - disúria: sim/não.
- * sintoma - prurido: sim/não.
- * tempo de evolução da doença, em meses: analisado em quatro categorias (1 a 6 meses, de 7 a 12 meses, de 13 a 23 meses e 24 meses ou mais).

- variáveis relacionadas ao tumor

- * presença de lesão associada: sim/não.
- * presença de tumor concomitante: sim/não.
- * acometimento de glândula: sim/não. (categoria 187.2 da CID-9ª revisão).
- * acometimento de escroto: sim/não. (categoria 187.7 da CID-9ª revisão).
- * acometimento de prepúcio: sim/não. (categoria 187.1 da CID-9ª revisão).

- * acometimento de corpo: sim/não. (categoria 187.3 da CID-9ª revisão).
- * acometimento de uretra: sim/não. (categoria 189.3 da CID-9ª revisão).
- * acometimento de raiz: sim/não.
- * acometimento de outros locais do pênis (OUTRO): sim/não. Outros locais da região peniana são pele do pênis (categoria 187.4 da CID-9ª revisão), epidídimo (categoria 187.5 da CID-9ª revisão), cordão espermático e vesícula seminal (categoria 187.6 da CID-9ª revisão), e pênis SOE (categoria 187.9 da CID-9ª revisão).
- * total de gânglios positivos: analisado em quatro categorias (1 a 6 meses, de 7 a 12 meses, de 13 a 23 meses e 24 meses ou mais).
- * estadiamento T: (0+I), II, III e IV.
- * estadiamento N: 0, I, II e III.
- * estadiamento M: 0 e I.
- * grau de diferenciação do tumor: foi analisado em três categorias (bem, moderado e pouco, e sem outra especificação - SOE).

• variáveis relacionadas aos procedimentos e terapia

- * cirurgia: amputação total, outros tipos de cirurgia e não realizada (NR).
- * linfadenectomia: profilática, necessária e NR.
- * esvaziamento ganglionar direito: NR, inguinal e (inguinal+ilíaco).
- * esvaziamento ganglionar esquerdo: NR, inguinal e (inguinal+ilíaco).
- * motivo do esvaziamento ganglionar direito: NR, profilático e necessário.
- * motivo do esvaziamento ganglionar esquerdo: NR, profilático e necessário.
- * tempo para a realização do esvaziamento: NR, (lado direito ou esquerdo em menos de 7 meses) e demais.
- * quimioterapia: sim/não.
- * radioterapia: sim/não.

Em todas as variáveis que possuíam informações ignoradas, foi definida uma nova categoria “ignorado” e a análise de sobrevida incluiu todos os casos, inclusive os com alguma(s) informação (ões) ignorada(s).

VI. 3. As amostras

Foram selecionadas três amostras do banco de dados original(B1), com o objetivo de validar o conjunto de fatores prognósticos definidos pelo modelo de Cox convencional. Quando o objetivo é a validação externa dos resultados, é recomendado⁽³⁴⁾ que se utilizem diferentes estratégias de amostragem para definirem-se os sub-grupos ou, então, divida-se o banco de dados original em 2 outros.

Porém neste estudo decidiu-se por trabalhar com três amostras do banco de dados original, objetivando-se validar internamente os resultados da análise do banco de dados original utilizando o modelo de Cox convencional.

- A primeira amostra (A1) foi feita selecionando-se 25% dos 648 pacientes do banco de dados original, através de amostragem casual simples por meio de sorteio por dígitos aleatórios(162 casos).
- A segunda amostra (A2) não foi sorteada, mas foram selecionados apenas os pacientes que dispunham de todas as informações para as seguintes variáveis independentes: tempo de evolução da doença, local de acometimento (glândula, prepúcio, corpo, escroto, uretra, raiz e outro), total de gânglios positivos, estadiamento do tumor (classificação T, N e M) e grau de diferenciação do tumor. Portanto, nesse banco de dados, constituído de 391 pacientes, não havia valores ignorados (*missing values*).

- Do banco de dados em que não havia valores ignorados (A2), foi sorteada uma amostra (A3) através de amostragem casual simples de 25% desse conjunto de dados. O sorteio foi por meio de dígitos aleatórios. Essa terceira amostra constituiu-se de 98 pacientes.

Fica claro que as três amostras são dependentes do banco de dados original.

VI. 4. A análise de sobrevivência

Neste estudo foi analisada a sobrevida total dos pacientes até cinco anos de acompanhamento. Apesar de se dispor de seguimento até dez anos, nesse período 44% (285/648) dos pacientes foram considerados perdidos de vista. Por isso optou-se por trabalhar com a sobrevida até 5 anos onde esse tipo de perda foi de 33% (217/648).

A análise de sobrevida foi realizada considerando como evento de interesse (falha) o óbito do paciente. Considerou-se como paciente censurado aquele que morreu por outra causa, ou foi perdido de vista (isto é, aquele paciente que não voltou ao serviço durante um tempo duas vezes maior do que o tempo que estava marcada a última consulta) ou terminou o estudo vivo. Nesta análise todos os pacientes que tiveram tempo de acompanhamento superior a 5 anos foram truncados à direita, ou seja, o tempo de sobrevida foi de 5 anos e foram considerados censura.

VI.4.1.Método de Kaplan-Meier

Primeiramente foram calculadas as funções de probabilidade de sobrevida $[S(t)]$, bem como elaborados os respectivos gráficos de $S(t)$ para cada uma das categorias das variáveis independentes. O objetivo desta etapa foi verificar o comportamento das curvas de $S(t)$ para cada uma das categorias das variáveis de

estudo. Essa análise descritiva auxilia na elaboração do modelo de Cox, pois já se conhece o comportamento de $S(t)$. Utilizou-se o teste *log rank* (Cox-Mantel)⁽¹⁹⁾ para avaliar a diferença entre as curvas de sobrevida das diversas categorias de cada variável. Há vários testes para se compararem as curvas de sobrevida calculadas pelo método de Kaplan-Meier. A que se utilizou nesse trabalho é a conhecida por Mantel log rank statistic ou Cox-Mantel log rank statistic ou, simplesmente, teste log rank. Ao se analisar uma específica variável, foram excluídos todos os pacientes com informação “ignorada” para a mesma.

VI.4.2. Modelo de riscos proporcionais de Cox multivariado

- convencional (M1)

Foram elaborados modelos multivariados para a sobrevida total (global) em cada um dos quatro conjuntos de dados (B1, A1, A2 e A3) com o objetivo de se verificar quais os fatores prognósticos que cada conjunto definiria.

Para tanto construíram-se, em primeiro lugar, modelos univariados com cada uma das covariáveis. O objetivo dessa etapa foi selecionar as variáveis para o modelo multivariado. Foram selecionadas todas as variáveis independentes cujo modelo univariado tivesse nível descritivo menor que 10% ($p < 0,10$). Caso o modelo tivesse $p < 0,10$, porém a única categoria significativa fosse a “ignorado”, a variável era desconsiderada. Também não foram incluídas, na análise multivariada, as variáveis em que não tivesse óbito em alguma das categorias. As variáveis idade e ano de início no estudo foram testadas no modelo multivariado, independente da sua significância estatística na análise univariada.

O processo de modelagem foi o *stepwise forward*, recomendado por FEINSTEIN e colaboradores^(25,77,78) e HARREL e colaboradores⁽³⁴⁾, quando se deseja

Para se avaliar se uma variável independente tinha razão de riscos proporcionais ao longo do seguimento, foram construídos gráficos de $\ln(-\ln(S(t)))$ em relação ao tempo, para cada uma delas, utilizando o estimador produto-limite de Kaplan-Meier (24) para estimar $S(t)$. O banco de dados utilizado nesta etapa foi o A2, pois nele as variáveis de estudo não tinham categoria “ignorado”. A comparação entre as curvas de $S(t)$ para as diversas categorias foi visual. Sempre que as curvas fossem paralelas, as razões de risco foram consideradas proporcionais.

A seguir elaborou-se o modelo de Cox multivariado, pelo *stepwise forward*, incorporando o tempo no modelo e assumindo o tempo como função linear da covariável. Esse modelo foi denominado **M2**.

VI.4.3. Modelo utilizando processos de contagem (M3)

Foi realizado esse modelo utilizando a mesma metodologia exposta para o modelo de Cox, qual seja, primeiramente a análise univariada e, posteriormente, a análise multivariada pelo *stepwise forward*. Foram consideradas nessa etapa apenas as variáveis significantes na análise univariada do modelo de Cox convencional realizada no banco de dados original. Esse modelo foi denominado **M3**.

VI.5. Validação dos resultados obtidos na determinação dos fatores prognósticos pelo modelo de Cox convencional (Modelo M1)

O objetivo dessa etapa foi validar os fatores prognósticos definidos através do modelo de Cox multivariado na análise do banco de dados original. Para tanto

foram comparados os resultados obtidos com esse banco de dados (B1) com os resultados dos outros bancos de dados (A1,A2 e A3). Essa análise não pode ser feita diretamente, através das funções de verossimilhança, pois o tamanho dos bancos de dados era diferente.

Foram comparados os resultados das razões de riscos para cada indivíduo (HR_i) calculadas a partir das fórmulas (27) definidas em cada um dos modelos de Cox convencional dos diferentes bancos de dados. As quatro fórmulas foram aplicadas nos 391 pacientes do banco de dados A2. Nessa comparação foram utilizados três diferentes métodos:

- foi calculado o coeficiente de correlação de Pearson entre as HR_i .
- foi calculado o coeficiente de confiabilidade alfa de Chronbach (*alpha Chronback reliability coefficient*)
- foi calculada a estatística Kappa para verificar o grau de concordância entre grupos de razões de risco (0-1, 2-5 e 6 e mais).

VI.6. Comparação entre o modelo de Cox convencional e o de processos de contagem

A comparação entre os modelos foi feita, primeiramente, verificando-se quais as variáveis estatisticamente significantes na análise univariada e a ordem de entrada no modelo multivariado. A seguir cotejaram-se os modelos multivariados finais, verificando quais as variáveis e sua significância estatística, e o valor da razão de verossimilhança.

VI. 7. *SOFTWARES*

- KMSURV: no método de Kaplan-Meier.
- COXSURV: na elaboração dos modelos de Cox.
- EPI-INFO (versão 5.1.b.): na realização das tabelas de frequências.
- HARVARD GRAPHICS for WINDOWS 2.0: na elaboração de alguns gráficos.
- SPSS (versão 5.0.1.) para WINDOWS: na elaboração dos modelos com a variável tempo-dependente, cálculos do coeficiente de correlação de Pearson, da estatística Kappa, e dos coeficientes de confiabilidade alfa de Chronbach; e os gráficos da função de probabilidade $S(t)$.
- S Plus para Unix (versão 3.2): na elaboração do modelo de processos de contagem.
- WORD FOR WINDOWS (versão 6.0): como processador de textos e para alguns gráficos.

VII. RESULTADOS

VII. 1. Características dos pacientes com câncer de pênis

Foram acompanhados 648 pacientes com câncer de pênis, durante o período de 1953 a 1985. Desses, 164 faleceram (25,3%) por essa causa. A tabela 4 e os gráficos de 3 a 8 apresentam os resultados referentes às características sócio-demográficas desses pacientes. Cerca de um quinto dos pacientes (21%) iniciaram no estudo na década de 50. Essa porcentagem aumentou para 33% na década de 60, passando a 30% e 16% nas décadas seguintes. A porcentagem de óbitos foi decrescendo durante o período de estudo. Na década de 50 ocorreram 34% dos óbitos, enquanto que na década de 80 houve 11 óbitos (10,5%).

A idade mínima foi 18 e a máxima 87 anos, com idade média de 54,3 anos (desvio padrão=13,2 anos) e mediana de 54 anos. Apenas 14% dos indivíduos tinham menos que 40 anos, mas foi nessa faixa etária que se verificou a maior porcentagem de óbitos, 33,3%, contra pouco mais de 20% nas demais faixas etárias.

Tabela 4. Características sócio-demográficas dos pacientes com câncer de pênis atendidos no Hospital A.C. Camargo. 1953 - 1985.

variável	categoria	total		óbito	
		Nº	%*	Nº	%**
idade (anos)	18 - 39	93	14,4	31	33,3
	40 - 59	331	51,1	82	24,8
	60 - 87	224	34,5	51	22,8
raça	branco	499	77,0	131	26,3
	não branco	149	23,0	33	22,1
estado civil	casado	379	95,0	91	24,0
	não casado	20	5,0	6	30,0
	ignorado	249	38,4	67	26,9
zona	rural	282	45,0	72	25,5
	urbana	345	55,0	86	24,9
	ignorado	21	3,2	6	28,6
procedência	SE	484	77,5	129	26,7
	S	40	6,4	6	15,0
	NE	62	9,9	12	19,4
	N+CO	39	6,2	11	28,2
	ignorado	23	3,5	6	26,1
ano de início	1953 - 1959	137	21,1	47	34,3
	1960 - 1969	214	33,0	62	29,0
	1970 - 1979	192	29,7	44	22,9
	1980 - 1985	105	16,2	11	10,5
Total		648	100,0	164	25,3

*: porcentagem calculada excluindo-se pacientes sem informação. com referência à cada variável. A porcentagem da categoria "ignorada" foi calculada em relação aos 648 pacientes

**: porcentagem calculada em relação ao total de pacientes da categoria

A maioria dos pacientes era branco (77%, excluindo-se os pacientes com informação ignorada) e casado (95%, excluindo-se os pacientes com informação

ignorada), no entanto a porcentagem de óbitos foi semelhante para brancos e não brancos (26,3% / 22,1%) e casados e não casados (24,0% / 30,0%). É importante ressaltar que em 38,4% (249/648) dos pacientes o estado civil foi ignorado.

Houve, aproximadamente, a mesma porcentagem de pacientes provenientes da zona urbana (55%) e rural (45%) e em ambas as categorias a porcentagem de óbitos foi próxima a 25%. A maioria morava na região Sudeste (77%).

A tabela 5 e os gráficos de 9 a 17 apresentam os resultados referentes às variáveis clínicas e sintomas pré-tratamento dos pacientes com câncer de pênis.

Tabela 5. Variáveis clínicas e sintomas pré-tratamento dos pacientes com câncer de pênis atendidos no Hospital A.C. Camargo. 1953 - 1985.

variável	categoria	total		óbito	
		Nº	%*	Nº	%**
doença venérea	sim	165	56,3	35	21,2
	não	128	43,7	28	21,9
	ignorado	355	54,8	101	28,5
tratamento não oncológico prévio	sim	245	82,5	58	23,7
	não	52	17,5	10	19,2
	ignorado	351	54,2	96	27,4
tratamento oncológico prévio	sim	88	13,6	36	40,9
	não	560	86,4	128	22,9
sintoma: ferida	sim	645	99,5	164	25,4
	não	3	0,5	-	-
sintoma: dor	sim	135	20,8	29	21,5
	não	513	79,2	135	26,3
sintoma: inflamação	sim	94	14,5	31	33,0
	não	554	85,5	133	24,0
sintoma: disúria	sim	80	12,3	18	22,5
	não	568	87,7	146	25,7
sintoma: prurido	sim	145	22,4	30	20,7
	não	503	77,6	134	26,6
tempo de evolução da doença (mês)	1 - 6	239	38,6	58	24,3
	7 - 12	178	28,8	46	25,8
	13 - 23	31	5,0	12	38,7
	24 - 348	171	27,6	37	21,6
	ignorado	29	4,5	11	37,9
Total		648	100,0	164	25,3

*: porcentagem calculada excluindo-se pacientes sem informação, com referência à cada variável. A porcentagem da categoria "ignorada" foi calculada em relação aos 648 pacientes

** : porcentagem calculada em relação ao total de pacientes da categoria

Não se conseguiu a informação a respeito da presença de doença venérea em 55% (355/648) dos pacientes. Isso prejudicou as conclusões a respeito desse fator de risco que vem sendo pesquisado em outros estudos. Naqueles em que se obteve a

informação, havia presença de pelo menos uma doença venérea em pouco mais da metade ($56\% = 165/293$). Porém a porcentagem de óbito foi 21%, independente da presença de doença venérea.

Não se obteve informação se o paciente havia realizado tratamento não oncológico prévio em 54% (351/648) deles. Naqueles com informação, cerca de 82% (245/297) haviam realizado algum tipo de tratamento não oncológico anteriormente, sendo que a porcentagem de óbito nesse grupo foi pouco superior ao do grupo sem tratamento (23,7%/19,2%). A maioria dos pacientes (86%) não havia realizado nenhum tratamento relativo ao tumor antes de procurar o Hospital A.C. Camargo. A porcentagem de óbito daqueles pacientes que já haviam realizado algum tratamento oncológico prévio foi quase o dobro dos restantes (40,9%/22,9%).

O sintoma mais comum foi ferida no pênis (99,5%). Poucos referiam prurido (22,4%), dor (20,8%), inflamação (14,5%) ou disúria (12,3%). No entanto, a porcentagem de óbito esteve entre 20 e 26%, independente da presença ou não de qualquer sintoma, exceto entre aqueles que referiram inflamação peniana, onde a porcentagem de óbito foi de 33%.

O tempo de evolução da doença, antes de procurar tratamento, foi de 1 a 348 meses. O tempo médio de evolução foi de 20,5 meses (desvio padrão de 32,4 meses), porém 50% dos pacientes procuraram o médico com até 10 meses de evolução da doença. A porcentagem de óbitos foi pouco mais de 20% para todos, com exceção do grupo que procurou tratamento entre 13 e 23 meses, cuja porcentagem de óbito foi 38,7%.

A tabela 6 e os gráficos de 18 a 31 mostram os resultados referentes às variáveis relacionadas ao tumor, propriamente dito. Apenas 13 pacientes (2%) dos pacientes tinham lesões associadas ao câncer de pênis, porém nenhum deles morreu. Também poucos indivíduos (2,5%) apresentavam outro tumor concomitante ao câncer de pênis; e, desses, apenas um morreu (6,3%).

O comprometimento mais freqüente foi na glândula (79%) e no prepúcio (74%), seguidos de corpo (44%), uretra (18%), raiz (12%), escroto (5%) e outras partes do pênis (4%). Observa-se que quanto mais proximal a região de comprometimento, maior a porcentagem de óbito.

Tabela 6. Variáveis relacionadas ao tumor dos pacientes atendidos no Hospital A.C. Camargo, 1953 - 1985.

variável	categoria	total		óbito	
		Nº	(%)*	Nº	(%)**
lesão associada	não	635	98,0	164	25,8
	sim	13	2,0	-	-
tu concomitante	não	632	97,5	163	25,8
	sim	16	2,5	1	6,3
local: glândula	não	17	3,2	2	11,8
	sim	514	96,8	122	24,7
	ignorado	117	18,1	40	34,2
local: prepúcio	não	44	8,4	6	13,6
	sim	482	91,6	116	24,1
	ignorado	122	18,8	42	34,4
local: corpo	não	211	42,7	32	15,2
	sim	283	57,3	86	30,4
	ignorado	154	23,8	46	29,9
local: escroto	não	506	94,2	113	22,3
	sim	31	5,8	14	45,2
	ignorado	111	17,1	37	33,3
local: uretra	não	341	74,9	65	19,1
	sim	114	25,1	39	34,2
	ignorado	193	29,8	60	31,1
local: raiz	não	457	85,6	94	20,6
	sim	77	14,4	33	42,9
	ignorado	114	17,6	37	32,5
local: outro	não	509	95,1	109	21,4
	sim	26	4,9	19	73,1
	ignorado	113	17,4	36	31,9
total de gânglios positivos	0	525	82,0	120	22,9
	1	35	5,4	11	31,4
	2 - 3	40	6,3	13	32,5
	4 e +	40	6,3	16	40,0
	ignorado	8	1,2	4	50,0
estadiamento T	0+I	47	9,0	3	6,4
	II	190	36,3	33	17,4
	III	237	45,2	64	27,0
	IV	50	9,5	23	46,0
	ignorado	124	19,1	41	33,1
estadiamento N	0	336	55,7	44	13,1
	I	61	10,1	15	24,6
	II	92	15,3	28	30,4
	III	114	18,9	69	60,5
	ignorado	45	6,9	8	17,8
estadiamento M	0	627	99,1	160	25,5
	I	6	0,9	1	16,7
	ignorado	15	2,3	3	20,0
grau de diferenciação do tumor	bem	371	57,5	76	20,5
	mod+pouco	35	5,4	16	45,7
	SOE	239	37,1	72	30,1
	ignorado	3	0,5	-	-
Total		648	100,0	164	25,3

*: porcentagem calculada excluindo-se pacientes sem informação, com referência à cada variável. A porcentagem da categoria "ignorada" foi calculada em relação aos 648 pacientes

** : porcentagem calculada em relação ao total de pacientes da categoria

Gráfico 22: comprometimento do corpo

Gráfico 23: comprometimento do escroto

Gráfico 24: comprometimento da uretra

Gráfico 25: comprometimento da raiz

Gráfico 26: comprometimento de outros locais

Em 82% dos pacientes avaliados, não havia gânglios positivos. A porcentagem de óbitos foi crescente de acordo com o número de gânglios positivos avaliados.

Gráfico 27: total de gânglios positivos

O estadiamento T não pode ser avaliado em 19% (124/648) dos pacientes. Dos que foram classificados, a maioria se encontrava em T₂ (37%) e T₃ (46%). A

porcentagem de óbitos foi crescente quanto maior a extensão do tumor, passando de 6,4% no grupo T_{0+1} , para 46,0% no estadiamento T_4 .

Houve pouca perda na avaliação da classificação N (7%). A maioria dos pacientes (56%) estava sem evidências de envolvimento de linfonodos regionais (N_0). Porém quanto maior o comprometimento dos linfonodos, maior a porcentagem de óbitos, chegando a 60,5% de óbitos no estadiamento N_3 .

A maioria dos pacientes (99%) não apresentava metástases à distância. Dos 6 indivíduos com estadiamento M_1 , apenas um morreu (16,7%).

Em mais da metade dos pacientes (58%), o tumor era bem diferenciado, porém em 37% deles (239/645) não havia referência ao grau de diferenciação do tumor (SOE). No grupo de tumor moderado ou pouco diferenciado foi encontrada a maior porcentagem de óbitos (45,7%).

A tabela 7 e os gráficos 32 a 40 apresentam as variáveis relacionadas aos procedimentos e terapia a que se submeteram os pacientes com câncer de pênis atendidos no Hospital A.C. Camargo. Em 202 pacientes (31%) foi realizada amputação total do pênis e em 196 (30%) não foi feita nenhuma cirurgia. A maior porcentagem de óbito se observou dentre os pacientes não operados (45,4%).

Tabela 7. Variáveis relacionadas aos procedimentos e terapia dos pacientes com câncer de pênis atendidos no Hospital A.C. Camargo. 1953 - 1985.

variável	categoria	total		óbito	
		Nº	(%)*	Nº	(%)**
cirurgia	amputação total	202	31,2	49	24,3
	outras cirurgias	250	38,6	26	10,4
	não realizada	196	30,2	89	45,4
linfadenectomia	profilática	68	10,5	17	25,0
	necessária	139	21,5	42	30,2
	não realizada	441	68,0	105	23,8
esvaziamento ganglionar direito	não realizado	462	71,3	108	23,4
	inguinal	61	9,4	18	29,5
	inguinal+ilíaco	125	19,3	38	30,4
esvaziamento ganglionar esquerdo	não realizado	461	71,1	112	24,3
	inguinal	61	9,4	18	29,5
	inguinal+ilíaco	126	19,5	34	27,0
motivo do esvazia- mento ganglionar direito	não realizado	462	71,3	108	23,4
	profilático	55	8,5	10	18,2
	necessário	131	20,2	46	35,1
motivo do esvazia- mento ganglionar esquerdo	não realizado	461	71,1	112	24,3
	profilático	53	8,2	16	30,2
	necessário	134	20,7	36	26,9
tempo esvaziamento ganglionar	não realizado	482	74,4	115	23,9
	tempo dir. ou esq<7 m	115	17,7	35	30,4
	demais	51	7,9	14	27,5
quimioterapia	não	584	90,1	127	21,7
	sim	64	9,9	37	57,8
radioterapia	não	548	84,6	122	22,3
	sim	100	15,4	42	42,0
Total		648	100,0	164	25,3

*: porcentagem calculada excluindo-se pacientes sem informação, com referência à cada variável. A porcentagem da categoria "ignorada" foi calculada em relação aos 648 pacientes

**: porcentagem calculada em relação ao total de pacientes da categoria

A linfadenectomia profilática foi realizada em 10% dos pacientes, sendo que em esse procedimento foi necessário em 21% deles. Verifica-se que a porcentagem de óbito foi maior no segundo grupo (30,2%).

Em 71% dos pacientes não foi realizado esvaziamento ganglionar, seja direito ou esquerdo, com porcentagem de óbito semelhantes (próximas a 24%). Ao serem observados os grupos que se submeteram a intervenções profiláticas, houve maior porcentagem de óbito dentre os que a realizaram no lado esquerdo (30,2%). Cerca de 18% dos pacientes realizaram esvaziamento com menos de 7 meses após o início do tratamento, porém a porcentagem de óbitos foi semelhante aos que se submeteram à essa cirurgia após esse tempo. O motivo do esvaziamento não apresentou diferenças quanto aos lados, nem em relação à porcentagem de óbitos.

Grande parte dos pacientes não se submeteu à quimioterapia (90%) ou à radioterapia (85%). No entanto, as maiores porcentagens de óbito foram entre os pacientes que receberam esses tratamentos, quais sejam, 57,8% dentre os que fizeram quimioterapia e 42,0% dentre os que fizeram radioterapia.

VII.2. Comparação entre os bancos de dados dos pacientes com câncer de pênis

Para facilidade da apresentação, os quatro bancos de dados serão identificados de acordo com as siglas apresentadas no capítulo VI desta tese (Material e Métodos), quais sejam:

- **B1:** o banco de dados original (648 casos),
- **A1:** amostra aleatória de 25% dos pacientes do banco de dados original (162 casos),
- **A2:** selecionados apenas os pacientes que tinham informações completas (391 casos), e
- **A3:** amostra aleatória do conjunto de dados A2 (98 casos).

Como as tabelas com os resultados são extensas, optou-se por deixar no texto apenas aquelas que resumem o assunto discutido. As tabelas com os resultados em detalhes encontram-se no anexo 3.

Ao se compararem os diversos bancos de dados, verifica-se que as características sócio-demográficas dos pacientes são muito semelhantes (tabela 8 - anexo 3). Tanto as porcentagens das diversas categorias, como as porcentagens de óbitos são muito parecidas para todas as variáveis. O mesmo se verifica entre as variáveis clínicas e sintomas pré-tratamento (tabela 9 - anexo 3), as variáveis relacionadas ao tumor (tabela 10 - anexo 3) e as variáveis relacionadas aos procedimentos e terapia (tabela 11 - anexo 3). Porém é importante notar que algumas categorias em que havia poucos casos e/ou poucos óbitos já no banco de dados original, muitas vezes não puderam ser avaliadas nas duas amostras sorteadas (A1-162 casos e A3-98 casos), pois apresentaram caselas zero em diversas categorias.

Também vale destacar o comportamento dos quatro grupos de pacientes em relação à idade dos pacientes (gráfico 41), ano de início no estudo (gráfico 42) e porcentagem de óbitos (gráfico 43). Em todos, tanto a idade média quanto a mediana foi de 56 anos. As idades mínimas e máximas também foram semelhantes indicando que os quatro bancos de dados são comparáveis em relação à idade.

A maioria dos pacientes (cerca de 65%) entrou no estudo entre as décadas de 60 e de 70, sendo que por volta de 16% iniciou no seguimento entre 1980 e 1985, em todos os bancos de dados.

Em todas as amostras, a porcentagem de óbitos esteve próxima a 21%, sendo que no grupo todo essa porcentagem foi de 25%. Isso indica que os bancos de dados são semelhantes em relação à ocorrência dos óbitos.

Ao se compararem as curvas de porcentagem de sobrevida acumulada $[S(t)]$, para os quatro bancos de dados, observa-se que o comportamento das curvas é semelhante (gráficos 44 a 47).

VII.3. Análise dos fatores prognósticos dos pacientes com câncer de pênis.

Resultados com o banco de dados original (B1)

Inicialmente foi realizada a análise de sobrevida univariada pelo método de Kaplan-Meier (tabelas 12 a 15 - anexo 3). As variáveis que apresentaram diferença estatisticamente significativa entre as suas categorias foram:

- ano de início no estudo (gráfico 48). Os pacientes que iniciaram no estudo após 1980 tiveram uma sobrevida melhor e após 60 meses de seguimento, 83% deles estavam vivos. Essa porcentagem foi superior ao que se encontrou para as décadas anteriores, ou seja, 61%, 67% e 73%, respectivamente para 1953-1959, 1960-1969 e 1970-1979. Observa-se, também, que a sobrevida, de uma maneira geral, foi melhor, quanto mais recente a década.

- tratamento oncológico prévio (gráfico 49). Verificou-se que 72% dos pacientes que não faziam tratamento oncológico antes de iniciar o tratamento no Hospital A.C. Camargo, permaneciam vivos após 60 meses, contra 49% do outro grupo.

Gráfico 49 - S(t) para tratamento oncológico prévio

- sintoma inflamação (gráfico 50). Os pacientes que apresentaram inflamação no pênis tiveram uma sobrevida acumulada pior do que os que não tiveram esse sintoma.

Gráfico 50 - S(t) para sintoma=inflamação

- local de acometimento (gráficos 51 a 55). Os pacientes que tiveram acometimento de corpo, escroto, uretra, raiz ou outro local do pênis tiveram sobrevida após 60 meses pior do que os que não tinham lesão nesses locais (respectivamente, 63%, 41%, 56%, 38% e 18%). Observa-se que a probabilidade de sobrevida acumulada foi sempre pior para os pacientes que tinham acometimento nesses locais..

Gráfico 51 - S(t) para local = corpo

Gráfico 52 - S(t) para local = escroto

Gráfico 53 - S(t) para local = uretra

Gráfico 54 - S(t) para local = raiz

Gráfico 55 - S(t) para local = outro

total de gânglios positivos (gráfico 56). Quanto maior o número de gânglios positivos detectados, menor a probabilidade de sobrevida após 60 meses de acompanhamento. Essa porcentagem foi de 53% nos pacientes que tinham mais de 4 nódulos e de 72% entre os que não tinham nenhum gânglio positivo. O grupo com 4 ou mais gânglios positivos teve a probabilidade de sobrevida acumulada sempre pior que a dos outros pacientes.

Gráfico 56 - S(t) para o total de gânglios positivos

- estádios T e N (gráficos 57 e 58). Quanto maior o grau do estadiamento, menor a porcentagem de sobrevida após 5 anos de acompanhamento. No estadiamento T_{0+1} essa probabilidade foi 95% passando para 77%, 69% e 33%, respectivamente, para T_2 , T_3 e T_4 . Para o estadiamento N essas porcentagens foram 83%, 72%, 63% e 22%, respectivamente para N_1 , N_2 , N_3 e N_4 .

Gráfico 57 - S(t) para o estadiamento T

Gráfico 58 - S(t) para estadiamento N

- grau de diferenciação do tumor (gráfico 59). Os pacientes que tinham tumor bem diferenciado tiveram as melhores probabilidades de sobrevida, quando comparado aos outros grupos. Especificamente para o grupo com pouco ou moderado grau de diferenciação, que teve a pior porcentagem de sobrevida, após 60 meses de seguimento, apenas 39% dos pacientes permaneciam vivos.

A seguir iniciou-se o processo de modelagem para se definirem os fatores prognósticos dos pacientes com câncer de pênis. Primeiramente foi feita a análise univariada da regressão de Cox (tabelas 16 a 19 - anexo 3). Nessa análise, os fatores prognósticos significantes foram: ano de início, tratamento oncológico prévio, todas as localizações, a presença de tumor concomitante, o estadio T, o estadio N e o grau de diferenciação do tumor. Como no grau de diferenciação do tumor só havia três casos na categoria “ignorado”, optou-se por retirar estes pacientes e refazer toda a análise univariada, sendo selecionandas, nesta segunda etapa, as mesmas variáveis. O modelo multivariado foi elaborado, pois com 645 pacientes. As variáveis citadas são, praticamente, as mesmas que tinham sido significantes pelo Método de Kaplan-Meier. Apesar de lesão associada também ter sido significativa, não foi selecionada para a análise multivariada porque dentre os 13 pacientes com lesão nenhum morreu. As variáveis glândula e prepúcio também não foram selecionadas para a análise multivariada pelo fato de que a categoria significativa, estatisticamente, ter sido a dos “ignorados”.

O estadio N foi a variável que iniciou o processo de modelagem. A seguir acrescentou-se o estadio T, que além de significativa, mostrou ser de confusão para o estadio N. Após foram analisadas, separadamente, as variáveis raiz, outro, uretra, corpo e escroto. Nenhuma delas melhorou a precisão do modelo e nem foram

estatisticamente significantes de *per si*. Por último o modelo foi acrescido da variável grau de diferenciação do tumor e das variáveis de controle ano e idade. A tabela 20 apresenta os resultados do modelo final.

Os gráficos 60 a 63 mostram as razões de risco ajustadas e respectivos intervalos de confiança para os fatores prognósticos dos pacientes dessa coorte (B1), a saber: estadiamento T, estadiamento N, grau de diferenciação do tumor e ano de início no estudo.

Tabela 20. Resultados da análise multivariada pelo modelo de Cox dos fatores prognósticos dos pacientes* com câncer de pênis atendidos no Hospital A.C. Camargo. 1953 - 1985. Modelo B1

variável	categoria	coef.(β i)	HR _{bruta}	HR _{ajustada}	IC(HRa) _{95%}	p(categ.)	p(variável)*
estad. T	0+I	referência	1,00	1,00	-	-	0,016
	II	0,9414	2,82	2,56	[0,78;8,39]	0,1198	
	III	1,1373	5,00	3,12	[0,97;10,07]	0,0571	
	IV	1,8106	12,85	6,11	[1,78;20,95]	0,0040	
	ignorado	1,0532	6,21	2,87	[0,86;9,60]	0,0875	
estad. N	0	referência	1,00	1,00	-	-	<0,000
	I	0,6863	1,94	1,99	[1,09;3,61]	0,0246	
	II	0,6966	2,64	2,01	[1,22;3,31]	0,0065	
	III	2,0252	9,54	7,58	[4,87;11,79]	<0,0001	
	ignorado	0,2756	1,32	1,32	[0,56;3,08]	0,5246	
grau de diferenc. do tumor	bem	referência	1,00	1,00	-	-	0,007
	mod+pouco	1,0676	3,61	2,91	[1,66;5,11]	0,0002	
	SOE	0,1976	1,58	1,22	[0,85;1,74]	0,2798	
ano	1953 - 1959	referência	1,00	1,00	-	-	0,003
	1960 - 1969	-0,0396	0,90	0,96	[0,64;1,45]	0,8497	
	1970 - 1979	-0,2743	0,67	0,76	[0,50;1,16]	0,2030	
	1980 - 1985	-1,0980	0,38	0,33	[0,17;0,65]	0,0014	

* excluídos três pacientes sem informação do grau de diferenciação do tumor

** valor de p da variável ao ser acrescentada no modelo multivariado

modelo final

log-likelihood=-914,2351

likelihood ratio=166,6858 $\Rightarrow \chi^2=166,6858$ com 13 gl $\Rightarrow p<0,0001$

A razão de riscos para cada paciente (i) desta coorte fica, então definida pela fórmula:

$$HR(i) = \exp \left(\begin{array}{l} 0,9414.T1 + 1,1373.T2 + 1,8106.T3 + 0,0532.Ti + 0,6863.N1 + 0,6966.N2 + 2,0252.N3 + 0,2756.Ni \\ + 1,0676.G1 + 0,1976.G2 - 0,0396.A1 - 0,2743.A2 - 1,0980.A3 \end{array} \right)$$

(38)

onde:

$$\begin{aligned} T1 &= \begin{cases} 1 \text{ se estadio T} = \text{II} \\ 0, \text{ estadio T} \neq \text{II} \end{cases} \\ T2 &= \begin{cases} 1 \text{ se estadio T} = \text{III} \\ 0, \text{ estadio T} \neq \text{III} \end{cases} \\ T3 &= \begin{cases} 1 \text{ se estadio T} = \text{IV} \\ 0, \text{ estadio T} \neq \text{IV} \end{cases} \\ Ti &= \begin{cases} 1 \text{ se estadio T} = \text{ignorado} \\ 0, \text{ estadio T} \neq \text{ignorado} \end{cases} \end{aligned}$$

$$\begin{aligned} N1 &= \begin{cases} 1 \text{ se estadio N} = \text{I} \\ 0, \text{ estadio N} \neq \text{I} \end{cases} \\ N2 &= \begin{cases} 1 \text{ se estadio N} = \text{II} \\ 0, \text{ estadio N} \neq \text{II} \end{cases} \\ N3 &= \begin{cases} 1 \text{ se estadio N} = \text{III} \\ 0, \text{ estadio N} \neq \text{III} \end{cases} \\ Ni &= \begin{cases} 1 \text{ se estadio N} = \text{ignorado} \\ 0, \text{ estadio N} \neq \text{ignorado} \end{cases} \end{aligned}$$

$$\begin{aligned} A1 &= \begin{cases} 1 \text{ se ano} < 1960 \\ 0, \text{ se ano} > 1959 \end{cases} \\ A2 &= \begin{cases} 1 \text{ se } 1959 < \text{ano} < 1970 \\ 0, \text{ caso contrario} \end{cases} \\ A3 &= \begin{cases} 1 \text{ se } 1969 < \text{ano} < 1980 \\ 0, \text{ caso contrario} \end{cases} \\ A1 &= \begin{cases} 1 \text{ se ano} > 1979 \\ 0, \text{ caso contrario} \end{cases} \end{aligned}$$

$$\begin{aligned} G1 &= \begin{cases} 1 \text{ se grau dif.} = \text{mod} + \text{pouco} \\ 0, \text{ se grau dif.} \neq \text{mod} + \text{pouco} \end{cases} \\ G2 &= \begin{cases} 1 \text{ se grau dif.} = \text{SOE} \\ 0, \text{ se grau dif.} \neq \text{SOE} \end{cases} \end{aligned}$$

A tabela 21 e o gráfico 64 mostram a frequência dos pacientes, segundo grupo de razões de riscos.

Tabela 21. Número de pacientes com câncer de pênis segundo grupos de risco, calculados pelo modelo de Cox multivariado Hospital A.C. Camargo. 1953 - 1985. Modelo B1.

Razão de Riscos	Nº	%
0-1	122	18,8
2 - 5	330	50,9
6 e +	196	30,3
Total	648	100,0

VII.4. Análise dos fatores prognósticos dos pacientes com câncer de pênis. Resultados com amostra de 162 pacientes do banco de dados original (A1)

Como o objetivo desta análise foi validar os resultados obtidos pelo banco de dados total, irão ser omitidos alguns detalhes apresentados na análise multivariada de B1. Isso para que a descrição dos resultados não se torne repetitiva.

Os resultados da análise univariada pelo Método de Kaplan-Meier encontram-se nas tabelas 12 a 15 (anexo 3). Nesta etapa, as variáveis que foram estatisticamente significantes foram: tratamento oncológico prévio, comprometimento da uretra e estadiamento N. A análise de algumas variáveis nesta amostra de 162 pacientes ficou prejudicada, devido ao pequeno tamanho de algumas categorias (por exemplo, algumas localizações do tumor) ou ausência de óbito (como, por exemplo, em T_{0+1}).

Através da análise univariada pelo modelo de Cox (tabelas 31 a 34 - anexo 3) selecionaram-se as variáveis estadiamento N, uretra, raiz e tratamento oncológico prévio, para a análise multivariada, sendo essa a ordem de entrada no modelo. Vale salientar que no estadio T a única categoria significativa foi a “ignorada”. O ano de

início no estudo e a idade também foram testados no modelo multivariado. Na presença do estadiamento N, nenhuma das variáveis apresentou significância estatística, o que fez com que no modelo final ficasse apenas uma variável. Mesmo para essa variável apenas a categoria N₃ foi significativa, com um intervalo de confiança para a HR muito grande. A tabela 35 e o gráfico 65 apresentam esse resultado.

Tabela 35. Resultados da análise multivariada pelo modelo de Cox dos fatores prognósticos dos pacientes com câncer de pênis atendidos no Hospital A.C. Camargo. 1953 - 1985. Modelo A1

variável	categoria	coef.(βi)	HR _{bruta}	IC(HRa) _{95%}	p(categ.)	p(variável)*
estad. N	0	referência	1,00	-	-	<0,0001
	I	0,3599	1,43	[0,40;5,09]	0,5775	
	II	-0,2235	0,80	[0,18;3,58]	0,7699	
	III	3,1006	22,21	[9,54;51,72]	<0,0001	
	ignorado	0,5386	1,71	[0,48;6,08]	0,4343	

* valor de p da variável ao ser acrescentada no modelo multivariado

modelo final

$\log\text{-likelihood} = -142,9105$

$\text{likelihood ratio} = 112,458 \Rightarrow \chi^2 = 112,458$ com 4 gl $\Rightarrow p < 0,0001$

A razão de riscos para cada paciente (i) desta coorte fica, então, definida pela fórmula:

$$HR(i) = \exp(0,3599.N1 - 0,2235.N2 + 3,1006.N3 + 0,5386.Ni) \quad (39)$$

onde :

$$N1 = \begin{cases} 1 & \text{se estadio N = I} \\ 0 & \text{estadio N} \neq \text{I} \end{cases}$$

$$N2 = \begin{cases} 1 & \text{se estadio N = II} \\ 0 & \text{estadio N} \neq \text{II} \end{cases}$$

$$N3 = \begin{cases} 1 & \text{se estadio N = III} \\ 0 & \text{estadio N} \neq \text{III} \end{cases}$$

$$Ni = \begin{cases} 1 & \text{se estadio N = ignorado} \\ 0 & \text{estadio N} \neq \text{ignorado} \end{cases}$$

A tabela 36 e o gráfico 66 mostram a frequência dos pacientes, segundo grupo de risco. A maioria dos pacientes (85%) teve risco entre 0 e 1,99 e nenhum teve risco entre 2 e 5,99, mostrando o baixo poder discriminatório desta função de riscos..

Tabela 36. Número de pacientes com câncer de pênis segundo grupos razão de riscos, calculados pelo modelo de Cox multivariado Hospital A.C.Camargo.1953 - 1985.Modelo A1.

Razão de Riscos	Nº	%
0 - 1	138	85,2
2 - 5	-	-
6 e +	24	14,8
Total	162	100,0

VII.5. Análise dos fatores prognósticos dos pacientes com câncer de pênis. Resultados com amostra de 391 pacientes do banco de dados original (A2)

A análise univariada pelo Método de Kaplan-Meier (tabelas 12 a 15 - anexo 3) mostrou que as variáveis cujas categorias diferiam quanto às curvas de sobrevida acumulada foram: estadio N, estadio T, local de acometimento do tumor (corpo, escroto, uretra, raiz e outros), grau de diferenciação do tumor e sintoma de inflamação. Essas variáveis também foram estatisticamente significantes pela análise univariada do modelo de Cox (tabelas 37 a 40 - anexo 3), sendo pois selecionadas para a análise multivariada. A ordem de entrada no modelo foi a sequência apresentada para o método de Kaplan-Meier. Além disso também foram testadas o ano de início no estudo e idade do paciente. A tabela 41 apresenta os resultados do modelo final.

Tabela 41. Resultados da análise multivariada pelo modelo de Cox dos fatores prognósticos dos pacientes com câncer de pênis atendidos no Hospital A.C. Camargo. 1953 - 1985. Modelo A2

variável	categoria	coef.(β i)	HR _{bruta}	HR _{ajustada}	IC(HRa) _{95%}	p(categ.)	p(variável)
estad. T	0+I	referência	1,00	1,00	-	-	0,013
	II	0,8255	2,43	2,28	[0,68;7,65]	0,1807	
	III	1,0356	4,11	2,82	[0,85;9,39]	0,0918	
	IV	1,8719	13,30	6,50	[1,86;22,78]	0,0034	
estad. N	0	referência	1,00	1,00	-	-	<0,000
	I	0,6686	1,74	1,95	[0,87;4,36]	0,1033	
	II	0,8991	3,64	2,46	[1,30;4,63]	0,0054	
	III	2,0441	10,74	7,72	[4,34;13,74]	<0,0001	
grau de diferenc. do tumor	bem	referência	1,00	1,00	-	-	0,0279
	mod+pouco	1,3276	2,98	3,77	[1,70;8,39]	0,0011	
	SOE	0,2910	1,63	1,34	[0,82;2,19]	0,2485	
ano	1953 - 1959	referência	1,00	1,00	-	-	0,0175
	1960 - 1969	0,1308	1,02	1,14	[0,66;1,98]	0,6418	
	1970 - 1979	-0,1656	0,73	0,85	[0,47;1,55]	0,5890	
	1980 - 1985	-1,3158	0,32	0,27	[0,10;0,73]	0,0102	

* valor de p da variável ao ser acrescentada no modelo multivariado

modelo final

$\log\text{-likelihood} = -420,9689$

$\text{likelihood ratio} = 99,7786 \Leftrightarrow \chi^2 = 99,7786 \text{ com } 11 \text{ gl} \Leftrightarrow p < 0,0001$

Os gráficos 67 a 70 mostram os razões de risco ajustados e respectivos intervalos de confiança para os fatores prognósticos dos pacientes dessa coorte (A2), a saber: estadio T, estadio N, grau de diferenciação do tumor e ano de início no estudo.

O modelo multivariado final, para o banco de dados A2, contém as mesmas variáveis do modelo com o banco total, embora os coeficientes sejam um pouco diferentes. A única variável que tem um comportamento um pouco desigual é o ano de entrada no estudo. A fórmula que descreve a razão de riscos de um paciente de A2 é dado pela fórmula:

$$HR(i) = \exp \left(0,8255.T1 + 1,0356.T2 + 1,8719.T3 + 0,6686.N1 + 0,8991.N2 + 2,0441.N3 + \right. \\ \left. + 1,3276.G1 + 0,2910.G2 + 0,1308.A1 - 0,1656.A2 - 1,3158.A3 \right) \quad (40)$$

onde:

$$T1 = \begin{cases} 1 & \text{se estadio T} = \text{II} \\ 0 & \text{estadio T} \neq \text{II} \end{cases}$$

$$T2 = \begin{cases} 1 & \text{se estadio T} = \text{III} \\ 0 & \text{estadio T} \neq \text{III} \end{cases}$$

$$T3 = \begin{cases} 1 & \text{se estadio T} = \text{IV} \\ 0 & \text{estadio T} \neq \text{IV} \end{cases}$$

$$N1 = \begin{cases} 1 & \text{se estadio N} = \text{I} \\ 0 & \text{estadio N} \neq \text{I} \end{cases}$$

$$N2 = \begin{cases} 1 & \text{se estadio N} = \text{II} \\ 0 & \text{estadio N} \neq \text{II} \end{cases}$$

$$N3 = \begin{cases} 1 & \text{se estadio N} = \text{III} \\ 0 & \text{estadio N} \neq \text{III} \end{cases}$$

$$G1 = \begin{cases} 1 & \text{se grau dif.} = \text{mod} + \text{pouco} \\ 0 & \text{se grau dif.} \neq \text{mod} + \text{pouco} \end{cases}$$

$$G2 = \begin{cases} 1 & \text{se grau dif.} = \text{SOE} \\ 0 & \text{se grau dif.} \neq \text{SOE} \end{cases}$$

$$\begin{aligned}
 A1 &= \begin{cases} 1 & \text{se ano} < 1960 \\ 0 & \text{se ano} > 1959 \end{cases} \\
 A2 &= \begin{cases} 1 & \text{se } 1959 < \text{ano} < 1970 \\ 0 & \text{caso contrario} \end{cases} \\
 A3 &= \begin{cases} 1 & \text{se } 1969 < \text{ano} < 1980 \\ 0 & \text{caso contrario} \end{cases} \\
 A1 &= \begin{cases} 1 & \text{se ano} > 1979 \\ 0 & \text{caso contrario} \end{cases}
 \end{aligned}$$

O gráfico 71 mostra a frequência dos pacientes, segundo grupo de risco. Dos 391 pacientes deste grupo, 51% (199) tiveram razão de riscos abaixo de 1 e para os restantes a razão de riscos chegou a 2, demonstrando que essa fórmula tem pouco poder discriminatório para esse grupo de pacientes

VII.6. Análise dos fatores prognósticos dos pacientes com câncer de pênis.

Resultados da amostra de 98 pacientes do banco de dados original (A3)

A análise de sobrevida neste banco de dados também ficou prejudicada pelo fato de haver caselas com frequência zero para algumas variáveis, como, por exemplo, idade, tempo de evolução da doença, comprometimento da glândula e outras

(tabelas 12 a 15 - anexo 3). A análise de sobrevida pelo método de Kaplan-Meier mostrou que as variáveis com diferentes curvas de probabilidade de sobrevida acumulada, ao se compararem as suas categorias, foram: estadio N, estadio T e comprometimento do escroto, raiz e outro local do pênis.

A análise univariada pela regressão de Cox (tabelas 42 a 45 - anexo 3) mostrou que apenas o estadio N e o comprometimento da raiz entrariam no processo de modelagem multivariada, bem como ano de início no estudo e idade do paciente. No modelo multivariado (tabela 46, gráfico 72), apenas o estadio N foi estatisticamente significativo, pois a presença das outras variáveis não melhorou a precisão do modelo.

Tabela 46. Resultados da análise multivariada pelo modelo de Cox dos fatores prognósticos dos pacientes com câncer de pênis atendidos no Hospital A.C. Camargo. 1953 - 1985. Modelo A3

variável	categoria	coef.(β_i)	HR _{bruta}	IC(HRb) _{95%}	p(categ.)	p(variável)*
estad. N	0	referência	1,00	-	-	0,0211
	I	-0,4085	0,67	[0,09;5,16]	0,6959	
	II	1,0124	2,75	[0,96;7,93]	0,0607	
	III	1,8551	6,39	[1,98;20,61]	0,0019	

* valor de p da variável ao ser acrescentada no modelo multivariado

modelo final

$\log\text{-likelihood} = -87,7582$

$\text{likelihood ratio} = 9,7154 \Rightarrow \chi^2 = 9,7154$ com 3 gl $\Rightarrow p = 0,0211$

O risco individual para cada paciente desta coorte (A3) fica, então, definido pela fórmula:

$$HR(i) = \exp(-0,4085.N1 + 1,0124.N2 + 1,8551.N3) \quad (41)$$

onde :

$$N1 = \begin{cases} 1 & \text{se estadio N} = \text{I} \\ 0 & \text{estadio N} \neq \text{I} \end{cases}$$

$$N2 = \begin{cases} 1 & \text{se estadio N} = \text{II} \\ 0 & \text{estadio N} \neq \text{II} \end{cases}$$

$$N3 = \begin{cases} 1 & \text{se estadio N} = \text{III} \\ 0 & \text{estadio N} \neq \text{III} \end{cases}$$

A tabela 48 e o gráfico 73 mostram a frequência dos pacientes, segundo grupo de risco. A maioria dos pacientes (71%) teve risco igual a 1.

Tabela 47. Número de pacientes com câncer de pênis segundo grupos de razões de riscos calculados pelo modelo de Cox multivariado Hospital A.C. Camargo. 1953 - 1985. Modelo A3.

Razões de Riscos	Nº	%
0 - 1	79	80,7
2 - 5	12	12,2
6 - 9	7	7,1
Total	98	100,0

VII.7. Comparação entre os resultados referentes ao estudo dos fatores prognósticos dos pacientes com câncer de pênis com as diversas amostras.

Validação interna do Modelo de Cox convencional

A análise multivariada para definirem-se os fatores prognósticos dos pacientes com câncer de pênis iniciou-se, sempre, pela análise univariada. A tabela 48 mostra as variáveis estatisticamente significantes, ou seja, selecionadas para a análise multivariada, segundo os diversos bancos de dados.

Tabela 48. Ordem de entrada no modelo multivariado(O*) e coeficientes de regressão (β^{**}) das variáveis selecionadas na análise univariada pelo modelo de Cox, segundo banco de dados. Hospital A.C. Camargo. 1953 - 1985.

variável	B1		A1		A2		A3	
	O*	β^{**}	O*	β^{**}	O*	β^{**}	O*	β^{**}
sintoma: inflamação	não	não	não	não	9	0,6181	não	não
sintoma: prurido	não	não	não	não	11	-0,5470	não	não
trat.oncológico prévio	não	não	3	1,0549	não	não	não	não
local: corpo	6	0,9170 0,8403	não	não	6	0,8937	3	0,8497
local: escroto	8	1,1045 0,5142	não	não	7	1,2171	5	1,6175
local: uretra	5	0,8499 0,6369	2	1,0327 1,0578	5	0,9243	não	não
local: raiz	4	1,1761 0,6054	4	0,9990 0,9884	2	1,5849	2	1,2184
local: outro	3	1,6315 0,5174	5	1,2318 0,8426	3	1,9560	não	não
total de gangl. posit.	9	0,1476 0,6011 -0,0634 0,4107	não	não	não	não	não	não
estadio T	2	1,3078 1,6094 2,5534 1,8257	não	não	4	0,8887 1,4138 2,5875	4	0,7572 1,5422
estadio N	1	0,6623 0,9696 2,2550 0,2788	1	0,3599 -0,2235 3,1006 0,5386	1	0,5520 1,2925 2,3741	1	-0,4085 1,0124 1,8551
grau de difer. do tu	7	1,2830 0,4573 -14,887	não	não	8	1,0913 0,4859	não	não
ano	10	-0,1009 -0,3997 -0,9660	6	0,4258 0,1523 0,0094	10	0,0156 -0,3122 -1,1268	6	0,0954 -0,8239 -1,5082
idade	11	-0,3333 -0,3889	7	-0,3527 -0,6570	12	-0,2590 -0,0668	7	0,5165
no. óbitos/total pacientes	164/ 648		37/ 162		84/ 391		21/ 98	

O*: ordem de entrada no modelo multivariado

β^{**} : valor do coeficiente de regressão de cada categoria da variável

Verifica-se que o estadiamento N esteve presente em todas as análises e sempre com a maior significância estatística. A raiz foi outra variável que também foi selecionada em todas as análises univariadas, porém a ordem de importância não

foi selecionada em todas as análises univariadas, porém a ordem de importância não foi a mesma. Os sintomas inflamação e prurido apareceram apenas em um dos bancos de dados (A2), o mesmo acontecendo com o total de gânglios positivos(B1). Observa-se que quanto maior o número de pacientes que existem no banco de dados, maior o número de variáveis selecionadas. Em B1 e A2 foram selecionadas, respectivamente, 11 e 12 variáveis para a modelagem multivariada. Nos outros bancos de dados foram escolhidas, em ambos, apenas 7. Os valores dos coeficientes de regressão são muito diferentes, inclusive em relação ao sinal, e se a comparação entre os modelos fosse feita diretamente por eles não se poderia chegar a nenhuma conclusão.

A tabela 49 mostra os resultados do modelo multivariado para cada um dos bancos de dados.

A única variável que, consistentemente, permanece como fator prognóstico para o óbito de pacientes com câncer de pênis é o estadio N. No entanto nem sempre todas as suas categorias foram significantes do ponto de vista estatístico. O estadiamento N₃ (evidências de envolvimento de linfonodos fixos) foi sempre muito significativo ($p < 0,001$). Os resultados apresentados pelos bancos de dados B1 e A2 são muito parecidos, não só em relação às variáveis selecionadas, mas, também, nas significâncias estatísticas das diversas categorias e nos valores dos coeficientes de regressão. O mesmo ocorre ao se compararem os resultados dos bancos de dados A1 e A3, que foram amostras sorteadas.

Tabela 49. Modelo multivariado de Cox para determinação de fatores prognósticos de pacientes com câncer de pênis, segundo banco de dados. Hospital A.C. Camargo. 1953 - 1985.

variável	categoria	B1			A1			A2			A3		
		HR _{a1}	IC(HR _{a1}) _{95%}	p	HR _{a1}	IC(HR _{a1}) _{95%}	p	HR _{a1}	IC(HR _{a1}) _{95%}	p	HR _{a1}	IC(HR _{a1}) _{95%}	p
estad. T	0+I	1,00	-	-	-	-	-	1,00	-	-	-	-	-
	II	2,56	[0,78;8,39]	0,1198	-	-	-	2,28	[0,68;7,65]	0,1887	-	-	-
	III	3,12	[0,97;10,07]	0,0571	-	-	-	2,82	[0,83;9,39]	0,0918	-	-	-
	IV	6,11	[1,78;20,95]	0,0040	-	-	-	6,50	[1,86;22,78]	0,0034	-	-	-
	ignorado	2,87	[0,86;9,60]	0,0875	-	-	-	-	-	-	-	-	-
estad. N	0	1,00	-	-	1,00	-	-	1,00	-	-	1,00	-	-
	I	1,99	[1,09;3,61]	0,0246	1,43	[0,40;5,09]	0,5775	1,95	[0,87;4,36]	0,1033	0,67	[0,09;5,16]	0,6
	II	2,01	[1,22;3,31]	0,0065	0,80	[0,18;3,58]	0,7699	2,46	[1,30;4,63]	0,0054	2,75	[0,96;7,93]	0,0
	III	7,58	[4,87;11,79]	<0,0001	22,2	[9,54;51,72]	<0,0001	7,72	[4,34;13,74]	<0,0001	6,39	[1,98;20,61]	0,0
	ignorado	1,32	[0,56;3,08]	0,5246	1,71	[0,48;6,08]	0,4343	-	-	-	-	-	-
grau de diferenc. do tumor	bem	1,00	-	-	-	-	-	1,00	-	-	-	-	-
	mod+posso	2,91	[1,66;5,11]	0,0002	-	-	-	3,77	[1,70;8,39]	0,0011	-	-	-
	SOE	1,22	[0,83;1,74]	0,2798	-	-	-	1,34	[0,82;2,19]	0,2485	-	-	-
ano	1953 - 1959	1,00	-	-	-	-	-	1,00	-	-	-	-	-
	1960 - 1969	0,96	[0,64;1,45]	0,8497	-	-	-	1,14	[0,66;1,98]	0,6418	-	-	-
	1970 - 1979	0,76	[0,50;1,16]	0,2030	-	-	-	0,85	[0,47;1,55]	0,5890	-	-	-
	1980 - 1985	0,33	[0,17;0,65]	0,0014	-	-	-	0,27	[0,10;0,73]	0,0102	-	-	-
signific. modelo	do final	p <0,0001			p <0,0001			p <0,0001			p= 0,0211		

A análise multivariada para cada um dos grupos definiu uma fórmula para se estimar a razão de riscos para cada paciente em estudo. As fórmulas, já apresentadas nas seções anteriores, são:

$$HR_{B1}(i) = \exp \left(\begin{array}{l} 0,9414.T1 + 1,1373.T2 + 1,8106.T3 + 0,0532.Ti + 0,6863.N1 + 0,6966.N2 + 2,0252.N3 + 0,2756.Ni \\ + 1,0676.G1 + 0,1976.G2 - 0,0396.A1 - 0,2743.A2 - 1,0980.A3 \end{array} \right) \quad (38)$$

$$HR_{A1}(i) = \exp(0,3599.N1 - 0,2235.N2 + 3,1006.N3 + 0,5386.Ni) \quad (39)$$

$$HR_{A2}(i) = \exp \left(0,8255.T1 + 1,0356.T2 + 1,8719.T3 + 0,6686.N1 + 0,8991.N2 + 2,0441.N3 + \right. \\ \left. + 1,3276.G1 + 0,2910.G2 + 0,1308.A1 - 0,1656.A2 - 1,3158.A3 \right) \quad (40)$$

e

$$HR_{A3}(i) = \exp(-0,4085.N1 + 1,0124.N2 + 1,8551.N3) \quad (41).$$

Para serem cotejadas as razões de riscos estimados pelos quatro modelos através da análise multivariada, essas quatro fórmulas foram aplicadas nos 391 pacientes do banco de dados com todas as informações completas (A2). A seguir foram determinados os grupos de risco, segundo cada fórmula. A tabela 50 e o gráfico 74 mostram esses resultados.

Tabela 50- Número e porcentagem de pacientes com câncer de pênis, segundo grupo de razões de risco e fórmula utilizada. Hospital A.C. Camargo. 1953 - 1985.

Grupo de risco	B1* (38)		A1* (39)		A2* (40)		A3* (41)	
	Nº	%	Nº	%	Nº	%	Nº	%
0 - 1	37	9,5	293	75,0	37	9,5	344	88,0
2 - 5	214	54,7	51	13,0	234	59,9	-	-
6 e mais	140	35,8	47	12,0	120	30,6	47	12,0
Total	391	100,0	391	100,0	391	100,0	391	100,0

* fórmula para a função de riscos definida pelo modelo multivariado de Cox.

Verifica-se que as fórmulas para as razões de riscos definidas pelos modelos de B1 e A2 determinam grupos de riscos semelhantes, o mesmo acontecendo com as fórmulas de A1 e A3. Os dois primeiros tendem a quantificar riscos mais altos que os dois últimos, provavelmente porque estão incorporando mais variáveis em suas fórmulas. Vale lembrar que na fórmula de A2 não entram os coeficientes de

regressão para as categorias de ignorado e na fórmula de B1, sim. No entanto os resultados são muito semelhantes.

A seguir, quando possível, foi calculada a estatística Kappa, comparando os grupos de riscos das diferentes fórmulas, duas a duas. As tabelas 51 a 56 apresentam esses resultados.

Tabela 51- Número e porcentagem de pacientes com câncer de pênis, segundo grupo de razões de risco e fórmula utilizada. Hospital A.C. Camargo. 1953 - 1985.

Grupo de Risco (fórmula 38)	Grupo de risco (fórmula 39)			Total
	0 - 1	2 - 5	6 e mais	
0 - 1	37	-	-	37
2 - 5	208	6	-	214
6 e mais	48	45	47	140
Total	293	51	47	391

Kappa=0,05504

O valor da estatística Kappa entre B1 e A2 mostra que, descontando o efeito do acaso, existe concordância em apenas 5,5% dos casos, indicando que as duas fórmulas não determinam os mesmos grupos de risco. A fórmula (38) tende a classificar os pacientes em grupos de risco mais altos.

Tabela 52- Número e porcentagem de pacientes com câncer de pênis, segundo grupo de razões de risco e fórmula utilizada. Hospital A.C. Camargo. 1953 - 1985.

Grupo de Risco (fórmula 38)	Grupo de risco (fórmula 40)			Total
	0 - 1	2 - 5	6 e mais	
0 - 1	37	-	-	37
2 - 5	-	214	-	214
6 e mais	-	20	120	140
Total	37	234	120	391

Kappa=0,90760

Houve uma concordância quase perfeita (próxima a 91%) entre esses dois critérios de classificação, indicando que se o objetivo do trabalho fosse determinar grupos de risco essas duas fórmulas são equivalentes.

Tabela 53 . Número e porcentagem de pacientes com câncer de pênis, segundo grupo de razões de risco e fórmula utilizada. Hospital A.C. Camargo. 1953 - 1985.

Grupo de Risco (fórmula 38)	Grupo de risco (fórmula 41)			Total
	0 - 1	2 - 5	6 e mais	
0 - 1	37	-	-	37
2 - 5	214	-	-	214
6 e mais	93	-	47	140
Total	344	-	47	391

Foi impossível se calcular a estatística Kappa, pois o número de categorias entre os grupos de risco de cada fórmula é diferente. Verifica-se que em apenas 21% dos casos $[(37+46)/391]$ houve concordância, porém esse é um resultado bruto, sem se levar em conta o efeito do acaso.

Tabela 54- Número e porcentagem de pacientes com câncer de pênis, segundo grupo de razões de risco e fórmula utilizada. Hospital A.C. Camargo. 1953 - 1985.

Grupo de Risco (fórmula 39)	Grupo de risco (fórmula 40)			Total
	0 - 1	2 - 5	6 e mais	
0 - 1	37	228	28	293
2 - 5	-	6	45	51
6 e mais	-	-	47	47
Total	37	234	120	391

Kappa = 0,05443

As fórmulas 39 e 40 não definem os mesmos grupos de riscos, com a segunda fórmula tendendo a classificar os pacientes em grupos de risco mais alto.

Tabela 55- Número e porcentagem de pacientes com câncer de pênis, segundo grupo de razões de risco e fórmula utilizada. Hospital A.C. Camargo. 1953 - 1985.

Grupo de Risco (fórmula 39)	Grupo de risco (fórmula 41)			Total
	0 - 1	2 - 5	6 e mais	
0 - 1	293	-	-	293
2 - 5	51	-	-	51
6 e mais	-	-	47	47
Total	344	-	47	391

Não se calculou a estatística Kappa, pois o número de categorias entre os grupos de risco é diferente. Embora não leve em conta o efeito do acaso, a concordância bruta é de 87% $[(293+47)/391]$. Todos os pacientes classificados no grupo de risco mais alto, pela fórmula 39, também o foram pela fórmula 41. O mesmo acontecendo com a maioria dos pacientes do grupo de risco mais baixo.

Tabela 56- Número e porcentagem de pacientes com câncer de pênis, segundo grupo de razões de risco e fórmula utilizada. Hospital A.C. Camargo. 1953 - 1985.

Grupo de Risco (fórmula 40)	Grupo de risco (fórmula 41)			Total
	0 - 1	2 - 5	6 e mais	
0 - 1	37	-	-	37
2 - 5	234	-	-	234
6 e mais	73	-	47	120
Total	344	-	47	391

Mesmo não se calculando a estatística Kappa, verifica-se que as duas fórmulas não definem grupos de risco semelhantes, com exceção dos pacientes classificados como de maior risco.

Também foi avaliada a concordância dos resultados através dos valores das funções de risco sem serem categorizadas. Primeiramente calculou-se o coeficiente de correlação de Pearson e, a seguir, o coeficiente de confiabilidade alfa de Chronbach. A tabela 57 apresenta esses resultados.

Tabela 57 - Coeficiente de correlação de Pearson (ρ) e coeficiente de confiabilidade alfa de Chronbach (α), segundo fórmula utilizada.

fórmula	B1* (38)		A1* (39)		A2* (40)		A3* (41)	
	ρ	α	ρ	α	ρ	α	ρ	α
B1* (38)	1,00	1,00	0,78	0,31	0,99	0,98	0,79	0,75
A1* (39)			1,00	1,00	0,72	0,31	0,93	0,63
A2* (40)					1,00	1,00	0,71	0,08
A3* (41)							1,00	1,00

* fórmula para a função de riscos definida pelo modelo multivariado de Cox.

α entre as quatro fórmulas = 0,8238

Pelos resultados acima, observa-se que as razões de riscos estimadas pelas fórmulas definidas pelos modelos multivariados de B1 e de A2 são muito correlacionadas, o mesmo acontecendo quando se avalia a correlação intra-classes (α). O mesmo se verifica, porém com menos intensidade, ao se compararem as razões de riscos fornecidas pelas fórmulas de A1 e A3.

Por tudo que se apresentou, verificou-se que as estimativas feitas pelo modelo de Cox convencional no banco de dados originais se equivalem às feitas no banco de dados onde todos os pacientes têm informações completas.

VII.8. Análise da proporcionalidade das razões de riscos e elaboração do modelo de Cox modificado.

Nessa optou-se por trabalhar com o banco de dados com pacientes com informação completa(A2).

Para cada uma das covariáveis elaborou-se o gráfico de $\ln(-\ln(S(t)))$ em relação ao tempo, utilizando o estimador produto-limite de Kaplan-Meier (24) para estimar $S(t)$. As variáveis cujas curvas de $S(t)$, para as diversas categorias, não diferiram estatisticamente (tabelas 12 a 15 - anexo 3) apresentaram gráficos em que todas as curvas coincidiam e, por isso, não serão apresentadas aqui. A seguir estão as curvas referentes à idade (gráfico 75), ano de início no estudo (gráfico 76), corpo (gráfico 77), escroto (gráfico 78), uretra (gráfico 79), raiz (gráfico 80), outro (gráfico 81), tratamento não oncológico prévio (gráfico 82), grau de diferenciação do tumor (gráfico 83) e estadiamentos T (gráfico 84) e N (gráfico 85).

Gráfico 79 . $\ln(-\ln(S(t)))$ para uretra Gráfico 80 . $\ln(-\ln(S(t)))$ para raiz Gráfico 81 . $\ln(-\ln(S(t)))$ para outro Gráfico 82 . $\ln(-\ln(S(t)))$ para trat.não onc. prévio Gráfico 83 . $\ln(-\ln(S(t)))$ para grau de difer. tumor Gráfico 84 . $\ln(-\ln(S(t)))$ para estadiamento T Gráfico 85 . $\ln(-\ln(S(t)))$ para estadiamento N

Pela visualização dos gráficos, aparentemente, as variáveis de local de acometimento (corpo, escroto, uretra, e raiz), os estadiamentos N e T, não têm os riscos proporcionais durante o período de seguimento.

Porém ao construírem-se os modelos univariados, incorporando o tempo como função linear da covariável, nenhum dos termos que incorporou o tempo foi estatisticamente significativo, como se observa na tabela 58.

Tabela 58 . Resultados da análise univariada do modelo de Cox, considerando a variável tempo-dependente. Hospital A.C. Camargo. 1953 - 1985.

variável	categoria [#]	coef.(β i)	HR	p(categ.)	p(variável)
local: corpo	não	referência	1,00	-	0,0003
	sim	1,2271	3,41	0,0005	
	t*corpo	-0,0296	0,97	0,1537	
local: escroto	não	referência	1,00	-	0,0028
	sim	1,6453	5,18	0,0003	
	t*escroto	-0,0428	0,96	0,2944	
local: uretra	não	referência	1,00	-	0,0014
	sim	1,0095	2,74	0,0019	
	t*uretra	-0,0103	0,99	0,6289	
local: raiz	não	referência	1,00	-	<0,0001
	sim	1,4007	4,06	<0,0001	
	t*raiz	0,0226	1,02	0,2926	
local: outro	não	referência	1,00	-	<0,0001
	sim	1,7650	5,84	<0,0001	
	t*outro	0,0234	1,02	0,3613	
estad. T	0+I	referência	1,00	-	<0,0001
	II	1,3741	3,95	0,0919	
	III	1,8800	6,55	0,0324	
	IV	3,1685	23,77	0,0008	
	t*T	-0,0041	1,00	0,7148	
estad. N	0	referência	1,00	-	<0,0001
	I	0,6326	1,88	0,1560	
	II	1,5094	4,52	0,0001	
	III	2,7235	15,23	<0,0001	
	t*N	-0,0096	0,99	0,2703	
grau de diferenc. do tumor	bem	referência	1,00	-	0,0387
	mod+pouco	1,0900	2,97	0,0113	
	SOE	0,4967	1,64	0,1372	
	t*(graudif)	<0,0001	1,00	0,9407	

#: t*X_i = termo tempo-dependente

Os coeficientes de regressão (β_{1i}) referentes ao termo tempo-dependente são muito próximos de zero, ou seja, não são estatisticamente significantes. Isso faz com que a equação

$$h(t/X_i) = h_o(t) \cdot \exp[(\alpha_{1i} + \alpha_{2i} \cdot t) \cdot X_i], \text{ para cada um dos } X_i$$

se reduza a $\boxed{h(t/X_i) = h_o(t) \cdot \exp[\alpha_{1i} \cdot X_i]}$, ou seja, não é necessário se considerar o tempo como função linear da variável independente.

Mesmo com esses resultados negativos, algumas simulações foram realizadas com o modelo multivariado, porém em nenhuma delas houve significância estatística no termo tempo-dependente. Observa-se que o resultado desta análise univariada, para os termos dependentes do tempo, é muito semelhante ao que foi obtido no modelo de Cox convencional. Por isso decidiu-se não comparar o modelo de Cox incorporando variável tempo-dependente com o modelo de processos de contagem.

VII.9. Modelo incorporando processos de contagem

Na realização desta etapa utilizou-se o banco de dados A2, ou seja, trabalhou-se com os pacientes em que haviam informações completas. Relembrando, esse banco de dados constituiu-se de 391 pacientes. Primeiramente foi realizada a análise univariada para algumas variáveis relacionadas ao tumor, idade e ano de início no estudo, que haviam sido estatisticamente significantes na análise univariada de B1. A variável “tratamento oncológico prévio” não entrou nesta análise porque foi considerada sem importância como fator prognóstico.

Tabela 59. Resultados da análise univariada pelo modelo usando processos de contagem para algumas variáveis. Hospital A.C. Camargo. 1953 - 1985.

variável	categoria	coef.(β i)	HR	IC(HR) _{95%}	p(categ.)	p(variável)
local: glândula	não	referência	1,00	-	-	0,992
	sim	0,0019	1,00	[0,68;1,48]	0,992	
local: prepúcio	não	referência	1,00	-	-	0,166
	sim	0,536	1,71	[0,75;3,91]	0,203	
local: corpo	não	referência	1,00	-	-	<0,0001
	sim	0,886	2,43	[1,57;3,76]	<0,0001	
local: escroto	não	referência	1,00	-	-	0,0007
	sim	0,348	1,42	[1,23;1,63]	<0,0001	
local: uretra	não	referência	1,00	-	-	0,0001
	sim	0,87	2,39	[1,56;3,66]	<0,0001	
local: raiz	não	referência	1,00	-	-	<0,0001
	sim	1,51	4,52	[2,89;7,07]	<0,0001	
local: outro	não	referência	1,00	-	-	<0,0001
	sim	1,69	5,41	[3,18;9,19]	<0,0001	
total de gânglios positivos	0	referência	1,00	-	-	0,420
	1	-0,126	0,88	[0,32;2,41]	0,807	
	2 - 3	0,488	1,63	[0,79;3,38]	0,190	
	4 e +	0,528	1,70	[0,74;3,90]	0,214	
estad. T	0+I	referência	1,00	-	-	<0,0001
	II	0,959	2,61	[0,79;8,62]	0,0116	
	III	1,503	4,49	[1,39;14,55]	0,0012	
	IV	2,601	13,48	[4,04;44,99]	<0,0001	
estad. N	0	referência	1,00	-	-	<0,0001
	I	0,452	1,57	[0,72;3,41]	0,2533	
	II	1,109	3,03	[1,70;5,41]	0,0002	
	III	2,385	10,86	[6,61;17,86]	<0,0001	
estad. M	0	referência	1,00	-	-	0,474
	I	-3,01	0,05	[0,00;∞]	0,728	
grau de diferenc. do tumor	bem	referência	1,00	-	-	0,0244
	mod+pouco	1,014	2,76	[1,30;5,85]	0,0084	
	SOE	0,402	1,49	[0,97;2,30]	0,0683	
idade (anos)	18 - 39	referência	1,00	-	-	0,609
	40 - 59	-0,291	0,75	[0,41;1,36]	0,342	
	60 - 87	-0,148	0,86	[0,46;1,62]	0,644	
ano de início	1953 - 1959	referência	1,00	-	-	0,0176
	1960 - 1969	-0,0406	0,96	[0,59;1,56]	0,8705	
	1970 - 1979	-0,3122	0,73	[0,42;1,28]	0,2727	
	1980 - 1985	-1,2409	0,29	[0,11;0,75]	0,0102	

A análise univariada pelo modelo de processos de contagem (tabela 59) mostrou que as variáveis que foram estatisticamente significantes, também o foram na análise univariada do modelo de Cox convencional, com exceção da glândula e prepúcio. A idade e o total de gânglios positivos também não se mostraram estatisticamente significantes nos dois modelos. A maioria das estimativas das razões de risco e respectivos intervalos de confiança são muito semelhantes.

A análise multivariada está apresentada na tabela 60, a seguir. O critério de entrada no modelo multivariado foi a significância estatística da análise univariada. O modelo multivariado foi iniciado pelo estadiamento N, seguido do estadiamento T, raiz, outro, corpo, uretra, escroto, grau de diferenciação do tumor, ano de início e idade.

Os fatores prognósticos significantes, estimados pelo modelo de sobrevida utilizando processos de contagem, foram os estadiamentos T e o N, o grau de diferenciação do tumor e o ano de início no estudo, como pode ser visualizado nos gráficos de 86 a 89. As variáveis de local de acometimento do tumor perderam sua significância na presença das variáveis de estadiamento, à semelhança do que aconteceu na modelagem utilizando o modelo de Cox convencional. A idade continuou não sendo significante, do ponto de vista estatístico. A década de 80 permaneceu como um fator protetor para o óbito do paciente com câncer de pênis

Tabela 60. Resultados da análise multivariada pelo modelo utilizando processos de contagem dos fatores prognósticos dos pacientes com câncer de pênis atendidos no Hospital A.C. Camargo. 1953 - 1985. Modelo B1

variável	categoria	coef.(β_i)	HR _{bruta}	HR _{ajustada}	IC(HRa) _{95%}	p(categ.)
estad. T	0+I	referência	1,00	1,00	-	-
	II	0,87	2,61	2,40	[0,72;7,98]	0,1550
	III	1,10	4,49	3,00	[0,91;9,90]	0,0719
	IV	1,86	13,48	6,41	[1,84;22,34]	0,0036
estad. N	0	referência	1,00	1,00	-	-
	I	0,5647	1,57	1,76	[0,80;3,89]	0,1630
	II	0,7081	3,03	2,03	[1,09;3,78]	0,0254
	III	2,0835	10,86	8,03	[4,65;13,89]	<0,0001
grau de diferenc. do tumor	bem	referência	1,00	1,00	-	-
	mod+pouco	1,2805	2,76	3,60	[1,64;7,94]	0,0015
	SOE	0,2063	1,49	1,23	[0,76;1,97]	0,3900
ano	1953 - 1959	referência	1,00	1,00	-	-
	1960 - 1969	0,0973	0,96	1,10	[0,65;1,87]	0,7170
	1970 - 1979	-0,1906	0,73	0,83	[0,47;1,47]	0,5140
	1980 - 1985	-1,4148	0,29	0,24	[0,09;0,66]	0,0053

modelo final

likelihood ratio=107,0000 $\Rightarrow \chi^2=107,0000$ com 11 gl $\Rightarrow p<0,0001$

VII.10. Comparação entre os resultados referentes ao estudo dos fatores prognósticos dos pacientes com câncer de pênis definidos pelo modelo de Cox convencional e o modelo utilizando processos de contagem

Para serem comparados os dois modelos foram utilizados os resultados do modelo multivariado definido por A2, por duas razões. Em primeiro lugar, foi esse o banco de dados analisado na modelagem de processos de contagem. Depois, os resultados dos fatores prognósticos de A2 foram muito semelhantes aos encontrados utilizando o banco de dados original. Em que pese o número maior de pacientes, como no banco de dados original havia as categorias “ignorado”, perdem-se graus de liberdade na análise desnecessariamente, pois essa informação não é utilizada na

discussão dos fatores prognósticos. A tabela 61 apresenta a ordem de entrada das variáveis no modelo multivariado dos dois modelos.

Com exceção dos sintomas (que não foram testados no modelo utilizando processos de contagem) e do total de gânglios positivos, os dois modelos selecionaram as mesmas variáveis, com, praticamente, a mesma ordem de entrada. A maior diferença foi na segunda variável a ser acrescentada no modelo multivariado. No primeiro modelo (A2) foi a raiz e, no segundo, o estadiamento T. Porém em ambos os modelos, essa variável, raiz, perde a significância estatística diante da variável estadiamento T. Vale ressaltar que o estadio T teve como categoria significativa apenas o T₄, nos dois modelos. Porém isso não interferiu na escolha do modelo final.

O modelo multivariado definido pelos dois métodos foi muito parecido, tanto no que diz respeito à variáveis selecionadas (estadiamentos T e N, grau de diferenciação do tumor e ano de início), quanto à significância estatística de cada uma das categorias. Para a maioria das categorias o intervalo de confiança da razão de riscos foi muito parecido, o mesmo acontecendo com a sua estimativa por ponto.

Isso leva a concluir que os fatores prognósticos definidos pelos dois métodos são os mesmos, com a mesma significância estatística. Comparando-se as duas razões de verossimilhança (*likelihood ratio*) verifica-se que a diferença entre elas é imperceptível. A razão de verossimilhança estimada para o modelo de Cox convencional foi de 99,7786 com 11 graus de liberdade. A mesma razão para o modelo de processos de contagem, foi de 107,0000, também com 11 graus de liberdade. Para ambos os modelos, a significância estatística da razão de verossimilhança foi alta ($p < 0,0001$).

Tabela 61. Ordem de entrada no modelo multivariado das variáveis selecionadas na análise univariada, segundo modelo. Hospital A.C. Camargo. 1953 - 1985.^A

variável	modelo definido por A2	Processos de contagem
sintoma: inflamação	9	não testada
sintoma: prurido	11	não testada
local: corpo	6	5
local: escroto	7	7
local: uretra	5	6
local: raiz	2	3
local: outro	3	4
total de gangl. posit.	não	não
estadio T	4	2
estadio N	1	1
grau de difer. do tu	8	não
ano	10	8
idade	12	9

A: o número nas caselas corresponde à ordem de entrada no modelo multivariado

VIII. DISCUSSÃO

VIII.1. Do ponto de vista estatístico

Ao se desejar analisar fatores prognósticos de determinada doença, o modelo de Cox tem sido o mais utilizado, seja por fornecer as estimativas das razões de riscos desses fatores, ajustadas pela presença de outras covariáveis, seja pelo fato de estar presente em grande número de pacotes estatísticos, tornando fácil o acesso a essa técnica.

No entanto, há várias questões que se apresentam na elaboração de modelos multivariados. A primeira delas refere-se à validação dos resultados, identificando possíveis associações espúrias.

Há dois aspectos na validação de resultados de um específico estudo : a validação interna e a validação externa. Não pretendeu-se esgotar esse assunto nessa discussão, porém foram abordadas questões referentes a validação por meio amostras do banco de dados original. O ideal, para a generalização dos resultados, é fracionar o banco de dados original e fazer a modelagem em cada um dos sub-conjuntos, como se fossem bancos de dados provenientes de diferentes pesquisas, ou seja, independentes. A consistência dos achados nesses diferentes bancos de dados daria maior força à associação entre a variável dependente e as covariáveis. Há várias técnicas para se sortear amostras com propósito de validação, como, por exemplo, o *bootstrap* ou o *split-sample*. FEINSTEIN e colaboradores⁽²⁵⁾ e KLEINBAUM e colaboradores⁽⁴⁴⁾ argumentam que, quando os achados de um estudo não são validados por meio de amostras independentes do banco de dados original, os resultados devem ser vistos com cautela e devem ser realizados, posteriormente, novos estudos, utilizando a mesma metodologia para dar credibilidade a esses resultados.

Neste trabalho, optou-se por trabalhar com o banco de dados original total, pois o objetivo primeiro era se compararem alguns modelos de análise de sobrevida e necessitava-se de uma coorte com tamanho de amostra grande e com muitos eventos. Com um banco de dados menor poderiam ficar comprometidas a significância estatística de algumas variáveis e a comparação entre os modelos seria prejudicada. Não se poderia avaliar se a não significância era pelo tamanho da amostra ou porque, realmente, as diferentes técnicas resultavam em modelos diferentes. A partir do banco de dados original foram selecionadas três amostras, para se validar, internamente, os resultados do modelo de Cox convencional. Através da comparação das frequências relativas das categorias das variáveis de estudo, nas várias amostras, verificou-se que todas as três eram dependentes da coorte original de 648 pacientes, mesmo tendo sido utilizadas diferentes estratégias de amostragem. Existe ciência, por parte da autora deste trabalho, que, com isso a generalização dos resultados aqui encontrados, ficou comprometida. Uma limitação deste trabalho é que as diferentes técnicas não foram avaliadas em bancos de dados que possuam diferentes porcentagens de ocorrência do evento de interesse.

Todas as estratégias de amostragem com propósito de validação dependem do tamanho do banco de dados original e de um número considerável de eventos positivos. Sem isso não se consegue fracionar adequadamente esse banco de dados. No presente estudo a coorte era composta por 648 pacientes e nos três sub-conjuntos selecionados havia 162, 391 e 98 pacientes. Obteve-se reprodutibilidade dos resultados dos modelos multivariados utilizando o modelo de riscos proporcionais de Cox convencional, quando as comparações foram feitas com os dois conjuntos de dados com maior número de pacientes. A presença da variável estadiamento N, nos modelos multivariados dos dois outros sub-conjuntos, fez com nenhuma outra covariável melhorasse a significância do modelo, do ponto de vista estatístico.

FEINSTEIN e colaboradores⁽²⁵⁾ defendem a idéia de que devem ser consideradas como relevantes apenas aquelas variáveis que foram significantes em todas as análises, pois só elas estariam livres de serem achados espúrios. Neste trabalho, apenas uma variável, o estadiamento N, esteve presente em todos os modelos. Porém vale ressaltar que os quatro fatores prognósticos definidos pelo banco de dados original também foram encontrados na análise multivariada do banco de dados cujos pacientes tinham todas as informações (391 pacientes). Esses fatores estão de acordo com o descrito em literatura, o que leva a pensar que talvez o tamanho das amostras menores tenha sido insuficiente para mostrar as associações mais fracas.

No banco de dados original havia pacientes com algumas informações ignoradas. A retirada desses pacientes resultou no banco de dados com 391 indivíduos. Os modelos multivariados desses dois bancos de dados foram muito semelhantes, seja pelos fatores prognósticos definidos, como pela significância estatística. Além disso, as funções das razões de riscos desses dois modelos, categorizam, semelhantemente, os pacientes.

Os pacientes com algumas informações faltantes (ou seja, os *missing values*) precisam ser avaliados com maior detalhe. Em primeiro lugar é importante se considerar o número das perdas. Neste estudo ocorreram 40% (257/648) de perdas. Esse é um valor significativo, porém como o banco de dados original era grande (648 pacientes), com os 391 pacientes restantes ainda conseguiu-se fazer a análise multivariada, sem perda da precisão estatística. Para algumas variáveis essa perda foi maior, como, por exemplo, o estado civil e a presença de doenças venéreas. Em segundo lugar, deve ser avaliado se essas perdas formam um grupo particular dentro do conjunto de dados ou se ocorreram aleatoriamente. Dependendo da causa das perdas nas observações há várias soluções, como por exemplo, retirarem-se esses pacientes, ou só se trabalhar com as variáveis que

possuam informações completas, ou estimarem-se valores para as perdas. Há algumas técnicas para essas estimativas e os pesquisadores devem se aconselhar com especialistas dessa área para se decidir a melhor maneira de se tratar os *missing values*. Neste trabalho a perda parece ser aleatória, pois os resultados, incluindo ou não as perdas, foram semelhantes.

Outro ponto importante, em modelagem estatística, é verificar se as suposições do modelo estão sendo obedecidas. A principal delas, no modelo de Cox, é a de que a razão de riscos entre as categorias de uma variável independente, é proporcional ao longo do seguimento. A avaliação da proporcionalidade pode ser visual, através do gráfico do $\ln(-\ln(S(t)))$ pelo tempo. Se as curvas das razões de riscos não forem paralelas, deve-se incorporar o tempo no modelo de Cox convencional.

Neste estudo, verificou-se, graficamente, que as curvas das razões de risco de algumas das variáveis independentes mais significantes no prognóstico de um paciente com câncer de pênis, pareciam não ser paralelas. Por isso, elaboraram-se modelos univariados incorporando esse termo. No entanto, em nenhuma das análises, o termo tempo-dependente se mostrou significativo, ou seja, todos os modelos voltaram a ser o modelo de Cox convencional.

Há várias razões para isso. A primeira é a de que os riscos eram efetivamente proporcionais e não se precisaria, efetivamente, fazer a análise através do modelo de Cox modificado. Outro motivo poderia ser que o tempo de seguimento não foi suficientemente longo. Séries consideradas longas são as com dez anos ou mais de seguimento e, nessas, recomenda-se que seja testado o termo tempo-dependente⁽²²⁾.

Por último, talvez a função de tempo que foi incorporada ao modelo não seja a mais adequada para descrever o processo. Neste trabalho incluiu-se o tempo como

uma função linear da variável. Essa é considerada uma “solução ingênua” que Cox propôs para solucionar a não proporcionalidade dos riscos. Pelo fato de não se mostrar como a melhor solução, em termos matemáticos, ainda há discussão sobre qual a melhor técnica a ser adotada quando os riscos não são proporcionais ou a variável é tempo dependente.

Neste trabalho utilizou-se como técnica alternativa, o modelo de análise de sobrevida utilizando processos de contagem. Os resultados encontrados para os dois modelos, de Cox convencional e de processos de contagem, foram muito semelhantes. Os dois modelos definiram os mesmos fatores prognósticos, com, praticamente, a mesma significância estatística. Os valores estimados para as razões de riscos também foram muito parecidos.

É precoce fazer-se alguma afirmação sobre qual a melhor técnica para ser utilizada em análise de sobrevida. Há várias questões em aberto. Os dois modelos poderiam ser comparados em seguimentos mais longos. Ou, então, poderiam ser feitas simulações de várias situações onde a proporcionalidade não se verifica ao longo do tempo. Estas sugestões ficam para futuros estudos metodológicos em análise de sobrevida.

O modelo utilizando processos de contagem é mais sofisticado, do ponto de vista matemático. Porém o modelo de Cox está disponível em grande número de pacotes estatísticos e a interpretação de coeficientes se faz com maior facilidade. Por isso, talvez, continue a ser a técnica estatística mais utilizada quando o objetivo do estudo é definir fatores prognósticos e grupos de risco.

VIII.2. Do ponto de vista epidemiológico

Estudar epidemiologia de câncer, qualquer que seja a localização, sempre é um desafio. Os estudos etiológicos trabalham com o modelo probabilístico de causalidade, ou seja, são avaliados os fatores cuja presença (ou ausência) leva a um aumento do risco de um indivíduo ficar doente ou ter recorrência ou outro evento de interesse. Em termos de fatores prognósticos, verificam-se quais os fatores que proporcionariam um acréscimo na chance de um indivíduo vir a óbito pela doença.

Especificamente em relação ao câncer de pênis, essa é uma doença rara, cujas maiores incidências se encontram em países sub-desenvolvidos. Por isso, para se analisarem seus fatores prognósticos são necessários vários anos de acompanhamento e muito cuidado na hora de coletar as informações, para que haja padronização de todas as definições e critérios, ao longo do tempo.

Essa questão é particularmente importante em estudos retrospectivos e que coletam informações em prontuários médicos, como foi o caso deste estudo. Geralmente, as anotações de um prontuário médico mostram as ocorrências diárias de evolução do paciente, o tratamento, os resultados de exames de laboratório, etc. Não têm por objetivo a execução de futuras pesquisas. Sendo assim, muitas informações necessárias nos estudos retrospectivos não constam do prontuário médico. Nessa coorte as informações referentes à presença de doença venérea e estado civil, por exemplo, só constavam em cerca de 50% dos prontuários. Além disso, os registros de um prontuário são feitos por diferentes pessoas, sem muita padronização e não se dispõe do paciente no momento da coleta do dado da pesquisa para se esclarecerem as possíveis dúvidas. Por isso, quando se desejar fazer uma coleta de informação em prontuário médico deve-se avaliar, primeiramente, a validade e a confiabilidade do dado.

Em que pese o fato de haver pequena diferença na porcentagem de óbitos entre as diversas categorias das variáveis sócio-demográficas, nenhuma delas foi importante na sobrevida dos indivíduos.

A média de idade dos pacientes (54 anos) está de acordo com o que foi descrito em literatura por BURGERS e colaboradores⁽¹²⁾ e LUCIA & MILLER⁽⁵⁰⁾. Comparando-se com séries de países desenvolvidos, como Estados Unidos⁽²⁷⁾ ou Dinamarca⁽³⁹⁾, a idade média dessa coorte foi menor. A variação das idades (entre 18 e 87 anos) foi semelhante ao estudo indiano⁽⁴⁰⁾. A maioria dos pacientes era da raça branca, porém isso parece não ser de importância para a definição de fatores prognósticos. A raça parece descrever mais as condições sociais do que, propriamente, uma predisposição genética. Infelizmente não se dispõem de dados a respeito da classe social desses pacientes. Uma alternativa seria comparar a procedência dos indivíduos (área urbana/rural e região). Porém não foram detectadas diferenças na sobrevida entre os pacientes, nas categorias dessas variáveis. Importante ressaltar que a maioria dos pacientes era proveniente da região Sudeste, com pouco predomínio da zona urbana.

O valor prognóstico da presença do HPV nos tumores não pode avaliada, pois não se dispunha desse dado. Em poucos pacientes havia a informação de presença de doença venérea. Naqueles em que havia essa informação, verificou-se que a sobrevida do paciente não se alterava, estatisticamente, na presença ou ausência da mesma.

O câncer de pênis é pobre em sintomatologia dolorosa. Alguns autores comentam que esse é um dos fatores que mais influenciam na demora do início do tratamento. Nessa casuística, mais da metade dos pacientes não referiu nenhuma sintomatologia. O principal sintoma foi a ferida no pênis, porém nenhum dos sintomas foi importante para o prognóstico dessa localização de câncer. O tempo

médio entre o início da sintomatologia e a procura de atendimento foi de 20,5 meses mostrando a demora em procurar o médico após o início do aparecimento dos sintomas. Esse resultado está acima do que foi descrito por outros autores, que referem que os pacientes procuram o tratamento, em geral, após um ano do aparecimento dos sintomas^(2,23,40).

As variáveis relacionadas ao tumor foram as mais importantes como fator prognóstico dos pacientes dessa coorte. Na análise univariada, verificou-se que quanto mais proximal o local de acometimento, menor a sobrevida do paciente. Neste estudo, a análise univariada mostrou, pelo método de Kaplan-Meier, que pacientes com acometimento de corpo, escroto, raiz e outros locais do pênis tiveram porcentagem de sobrevida aos 5 anos de seguimento menor, quando comparados com indivíduos sem acometimento. Utilizando os resultados do banco de dados original, as piores porcentagens de sobrevida foram quando o acometimento era de escroto (41%), raiz (38%) e outro (18%).

Os diversos trabalhos de literatura não referem o local de acometimento como fator prognóstico e, sim, a classificação TNM. FRALEY e colaboradores⁽²⁷⁾ citam que 100% dos pacientes de T₁, sem a presença de linfonodos positivos estavam vivos após 5 anos de seguimento, contra 0% dos pacientes de T₄. Fatores que também influenciaram na melhora da porcentagem de sobrevida do estudo desses autores foi a realização de linfadenectomia em T₂ e T₃. A demora na linfadenectomia aumentou o risco de morte dos pacientes em T₂.

GERBAULET e colaboradores⁽³⁰⁾ referem que a porcentagem de sobrevida, após 5 anos de acompanhamento, foi de 82% para pacientes sem linfonodos positivos e 36% para pacientes com linfonodos positivos.

HORENBLAS & van TITEREN⁽³⁸⁾ encontraram uma porcentagem de sobrevida de 93% para pacientes em N_1 e 55% e 30% para pacientes de estadios N_{2+3} e N_4 , respectivamente.

KAMAT e colaboradores⁽⁴⁰⁾ também mostraram que quanto maior o grau do estadiamento, menor a porcentagem de sobrevida. Por exemplo, após 5 anos de seguimento, 83% dos pacientes em N_1 estavam vivos, contra 60% e 40% para estadios mais avançados (N_2 e N_3 , respectivamente).

Outros autores analisaram apenas a presença de linfonodos positivos. RAVI⁽⁶⁷⁾ verificou que 95% dos pacientes sem linfonodos positivos e 75% dos com linfonodos positivos estavam vivos após 5 anos de seguimento. SRINIVAS e colaboradores⁽⁷¹⁾ também encontraram um decréscimo na sobrevida quanto maior o número de gânglios positivos, chegando a 12% a sobrevida após 5 anos, dos pacientes com acometimento de linfonodos bilaterais. SUFRIN e HUBEN⁽⁷²⁾ descreveram uma porcentagem de sobrevida, após 5 anos, de 66% em pacientes sem linfonodos metastáticos e 27% em pacientes com linfonodos metastáticos.

Em que pese o fato de todos os trabalhos citados analisarem porcentagem de sobrevida após 5 anos de seguimento, em nenhum desses trabalhos há referência de qual o método utilizado para esse cálculo, com exceção do trabalho de HORENBLAS & van TITEREN⁽³⁸⁾ que mencionaram que as porcentagens de sobrevida foram calculadas pelo método de Kaplan e Meier. Por isso não se comparou os resultados deste trabalho com os de literatura, pois as porcentagens seriam comparáveis apenas se as técnicas fossem iguais. Não há, também, menção da presença de censura nas observações e como as mesmas foram tratadas no cálculo das porcentagens de sobrevida, se fosse o caso.

Ao se analisar o número de linfonodos metastáticos, os resultados aqui apresentados estão de acordo com a literatura, pois quanto maior o número, menor a porcentagem de sobrevida. Uma outra variável importante foi o grau de diferenciação do tumor. Cerca de 76% dos pacientes com tumor bem diferenciado estavam vivos após 5 anos de seguimento, contra apenas 54% dos que tinham tumor pouco ou moderadamente diferenciados.

Neste trabalho, a análise multivariada identificou como fatores prognósticos para a sobrevida do câncer de pênis os estadiamentos T e o N, grau de diferenciação do tumor e ano de diagnóstico. Os pacientes com tumor pouco ou moderadamente diferenciado têm maior risco de morte quando comparados com os que têm tumor bem diferenciado. Quanto maior a classificação T e N, maior o risco de óbito do paciente. O ano de início no estudo, também foi um fator prognóstico importante, pois os pacientes que entraram no seguimento na década de 80 tiveram um menor risco de óbito. Porém esse resultado tem que ser visto com cautela, pois não se conhece os reais motivos.

Na revisão de literatura só foram encontrados dois trabalhos que estudaram os fatores prognósticos de câncer de pênis por análise multivariada. HORENBLAS & van TITEREN⁽³⁸⁾ relataram resultados parecidos aos encontrados por este estudo, pois os fatores prognósticos significantes de sua coorte foram o estadiamento N e o grau de diferenciação do tumor. Esses autores não apresentaram as estimativas das razões de risco. LOPES⁽⁴⁹⁾, estudando apenas pacientes que foram submetidos à amputação peniana e linfadenectomia, encontrou, na análise multivariada, que os fatores importantes para o risco de óbito foram: presença de linfonodos metastáticos, estadio clínico T e a infiltração eosinofílica. Pacientes com linfonodos metastáticos presentes tiveram razão de riscos igual a 2,7 (IC(HR)=[1,5;4,7]), quando comparados com os outros. Pacientes em T₃, tiveram razão de riscos 2 (IC(HR)=[1,1;3,6]), quando comparados com pacientes em T₁₊₂. Pacientes com

infiltrado eosinofílico moderado tiveram razão de riscos 2,1 ($IC(HR)=[1,0;4,3]$), quando comparados com aqueles que não tinham infiltração.

Isso reforça a importância do diagnóstico precoce da doença, para que se melhore não só a sobrevida, mas a qualidade de vida do paciente com câncer de pênis.

IX. CONCLUSÕES

1. O modelo de riscos proporcionais de Cox e o modelo de análise de sobrevida baseado na teoria de processos de contagem apresentaram resultados semelhantes, na definição de fatores prognósticos de uma coorte de pacientes com câncer de pênis.
2. Os resultados referentes ao banco de dados original foram confirmados por uma das amostras desse conjunto de dados. Apenas o estadiamento N foi confirmado como fator prognóstico também nas outras duas amostras.
3. Os fatores prognósticos para a sobrevida de pacientes com câncer de pênis foram os estadiamentos T e N e o grau de diferenciação do tumor. Esses resultados estão ajustados pelo ano de diagnóstico.
4. Os pacientes com maior risco de morte foram aqueles que apresentaram tumor pouco diferenciado, em T₃ ou em T₄, e N₁ ou N₂ ou N₃.

X. REFERÊNCIAS BIBLIOGRÁFICAS

1. AALEN, O.O. A linear regression model for the analysis of life time. **Stat. Med.**, **8**:907-25, 1989.
2. AALEN, O.O. Further results on the non-parametric linear regression model in survival analysis. **Stat. Med.**, **12**:1569-88, 1993.
3. ABI-AAD, A.S. & deKERNION, J.B. Controversies in ilioinguinal lymphadenectomy for cancer of the penis. **Urol. Clin. North Am.**, **19**:319-24, 1992.
4. ALTMAN, D.G. & ANDERSEN, P.K. Bootstrap investigation of the stability of a Cox regression model. **Stat. Med.**, **8**:771-83, 1989.
5. ALTMAN, D.G. Statistics in medical journals: developments in the 1980s. **Stat. Med.**, **10**:1897-913, 1991.
6. ANDERSEN, P.K. & GILL, R.D. Cox's regression model for counting processes: a large sample study. **The Ann. Stat.**, **4**:1100-20, 1982.
7. ANDERSEN, P.K. Survival analysis 1982-1991: the second decade of the proportional hazards regression model. **Stat. Med.**, **10**:1931-44, 1991.
8. ANDERSEN, P.K.; BORGAN, Ø.; GILL, R.D. & KEIDING, N. **Statistical models based on counting process**. New York, Springer-Verloag, 1993.
9. AYYAPAN, K.; ANANTHAKRISHMAN, N. & SANKARAN, V. Can regional lymph node involvement be predicted in patients with carcinoma of penis? **Br. J. Urol.**, **73**:549-53, 1994.

10. BECKER, N.G. **Analysis of infectious disease data**. London, Chapman and Hall, 1989 (Monographs on Statistics and Applied Probability).
11. BRINTON, L.A.; JUN-YAO, L.; SHOU-DE, R.; HUANG, S.; SHENG, X.B.; BAI-GAO, S.; ZHE-JUN, Z.; SCHIFFMAN, M.H. & DAWSEY, S. Risk factors for penile cancer: results from a case-control study in China. **Int. J. Cancer**, 47:504-9, 1991.
12. BURGERS, J.K., BADALAMENT, R.A. & DRAGO, J.R. Penile cancer: clinical presentation, diagnosis, and staging. **Urol. Clin. North Am.**, 19:247-56, 1992.
13. CANCER STATISTICS 1995. **Am. Cancer Soc.**, 45, 1995.
14. CHAMBERS, J.M. & HASTIE, T.S. **Statistical models in S**. California, Wadsworth & Brooks/Cole Computer Science Series, 1992.
15. CHEUVART, B. & RYAN, L. Adjusting for age-related competing mortality in long-term cancer clinical trials. **Stat. Med.**, 10:65-77, 1991.
16. CIAMPI, A.; HOGG, S.A. & KATES, L. Regression analysis of censored survival data with generalized F family - an alternative to the proportional hazards model. **Stat. Med.**, 5:85-96, 1986.
17. CONCATO, J.; FEINSTEIN, A.R. & HOLFORD, T.R. The risk of determining risk with multivariable models. **Ann Intern. Med.**, 118: 201-10, 1993.
18. COX, D.R. Regression models and life tables (with discussion). **J. R. Stat. Soc.**, Series B, 34: 187-220, 1972.

- 19.COX, D.R. & OAKES, D. **Analysis of survival data**. London, Chapman & Hall, 1984.
- 20.CROLEY, J. & BRESLOW, N. Statistical analysis of survival data. **Assoc. Rev. Public Health**, 5: 385-411, 1994.
- 21.DALING, J.R.; SHERMAN, K.J.; HISLOP, T.G.; MADEN, C.; MANDELSON, M.T.; BECKMANN, A.M. & WEISS, N.S. Cigarette smoking and the risk of anogenital cancer. **Am. J. Epidemiol.**, 135:180-9, 1992.
- 22.ESDAILE, J.M.; ABRAHAMOWICZ, M.; MacKENZIE, T.; HAYSLETT, J.P. & KASHGARIAN, M. The time-dependence of long-term prediction in Lupus Nephritis. **Arthritis Rheum.**, 37:359-68, 1994.
- 23.FAIR, W.R., FUKS, Z.Y. & SCHER, H.I. Cancer of the urethra and penis. In DeVita Jr, Hellman, S. & Rosenberg, S.A. (eds). **Cancer principles & practice of oncology**. 4th edition, Philadelphia , Lippincott Co, 1993.v.1.
- 24.FARANGI, D. & SIMON, R. A neural network model for survival data. **Stat. Med.**, 14:73-82, 1995.
- 25.FEINSTEIN, A.R.; WELLS, C.K. & WALTER, S.D. A comparison of multivariable mathematical methods for predicting survival - I. Introduction, rationale and general strategy. **J. Clin. Epidemiol.**, 43:339-47, 1990.
- 26.FOWLER JR, J.E. An approach to the management of penile cancer. **Surg. Ann.**, part 2, 139-56, 1992.

- 27.FRALEY, E.E.; ZHANG, G.; MANIVEL, C. & NIEHANS, G.A. The role of ilioinguinal lymphadenectomy and the significance of histological differentiation in treatment of carcinoma of penis. **J. Urol.**, **142**:1478-82, 1989.

- 28.FRANCO, E.L.; CAMPOS FILHO, N.; VILLA, L.L. & TORLONI, H.
Correlation patterns of cancer relative frequencies with some socioeconomic and demographic indicators in Brazil: an ecologic study. **Int. J. Cancer**, **41**:24-9, 1988.

- 29.GAIL, M.H. A bibliography and comments on the use of statistical models in epidemiology in the 1980s. **Stat. Med.**, **10**: 1819-85, 1991.

- 30.GERBAULET, A. & LAMBIN, P. Radiation therapy of cancer of the penis. **Urol. Clin. North Am**, **19**:325-32, 1992.

- 31.GREENLAND, S. Modelling and variable selection in epidemiologic analysis. **Am. J. Public Health**, **79**: 340-9, 1989.

- 32.GORE, S.M.; POCOCK, S.J. & KERR, G.R. Regression models and non-proportional hazards in the analysis of breast cancer survival. **Appl. Stat.**, **33**:176-95, 1984.

- 33.HANHAM, I.W.F. & BOSCH, F. Câncer do Pênis. In: União Internacional Contra o Câncer (UICC). **Manual de Oncologia Clínica**. 5ª ed., trad. Fundação Oncocentro de São Paulo. São Paulo, 1991.

- 34.HARREL JR, F.E.; LEE, K.L.; CALIFF, R.M.; PRYOR, D.B. & ROSATI, R.A. Regression modelling strategies for improved prognostic prediction. **Stat. Med.**, **10**:143-52, 1984.

- 35.HARREL JR, F.E.; LEE, K.L.; MATCHAR, B.D. & REICHERT, T.A.
Regression models for prognostic prediction: advantages, problems, and suggested solutions. **Cancer Treatment Report**, 69: 1071-7, 1985.
- 36.HARRIS, E.K. & ALBERT, A. **Survivorship analysis for clinical studies**. New York, Marcel Dekker Inc, 1991.
- 37.HENDERSON, R. & MILNER, A. Aalen plots under proportional hazards.**Appl. Stat.**,40:401-9, 1991.
- 38.HORENBLAS, S. & van TINTEREN, H. Squamous cell carcinoma of the penis. IV. Prognostic factors of survival: analysis of tumor, nodes and metastasis classification system. **J. Urol.**, 151: 1239-43, 1994.
- 39.JENSEN, M.S. Cancer of the penis in Denmark- 1942 to 1962 (511 cases). **Dan. Med. Bull.**, 24: 66-72, 1977.
- 40.KAMAT, M.R.; KULKARNI, J.N. & TONGAONKAR, H.B. Carcinoma of the penis: the Indian experience. **J. Surg. Oncol.**, 52:50-5, 1993.
- 41.KAPLAN, E.L. & MEIER, P. Nonparametric estimation from incomplete observations. **J. Am. Stat. Assoc.**, 53: 457-81, 1958.
- 42.KATTAN, J.; COURT, B.; CULINE, S.; PERRIN, J-L.; DROZ, J-P; WIBAULT, P.; FADEL, E. & HAIE-MEDER, C. Penile cancer chemotherapy: twelve years' experience at Institut Gustave-Roussy. **Urology**,42: 559-62, 1993.

- 43.KLEINBAUN, D.G., KUPPER, L.L. & MORGENSTERN, H. **Epidemiologic Research**. California, Lifetime Learning Publications, 1982.
- 44.KLEINBAUN, D.G., KUPPER, L.L. & MULLER, K.E. **Applied regression analysis and other multivariable methods**. California, PWS-KENT Publishing Company, 1988.
- 45.KOCH, M.O. & SMITH JR., J.A. Local recurrence of squamous cell carcinoma of penis. **Urol. Clin. North Am**, **21**:739-43, 1994.
- 46.LATORRE, M.R.D.O. Mortalidade como preditor da morbidade. São Paulo, 1992. [Dissertação de Mestrado - Faculdade de Saúde Pública da USP]
- 47.LEE, E.T. **Statistical methods for survival data analysis**. 2nd ed. New York, John Wiley & Sons INC, 1992.
- 48.LERNER, S.E; JONES, J.G. & FLEISCHMANN, J. Management of recurrent penile cancer following partial or total penectomy. **Urol. Clin. North Am**, **21**:729-37, 1994.
- 49.LOPES, A. Fatores clínicos e anatomopatológicos envolvidos no risco de metástases linfonodais e no prognóstico dos pacientes portadores de carcinoma de pênis tratados por cirurgia. São Paulo, 1995. [Tese de Doutorado - Faculdade de Medicina da Universidade de São Paulo]
- 50.LUCIA, M.S. & MILLER, G.J. Histopatology of malignant lesions of the penis. **Urol. Clin. North Am**, **19**(2):227-46, 1992.

- 51.MAICHE, A.G. & PYRHÖNEN, S. Clinical staging of cancer of penis: by size? By localization? Or by depth of infiltration? **Eur.Urol.**, 16:18-22, 1990.
- 52.MASSAD, E. Sobre a biomatemática das moléstias infecciosas: um estudo teórico exemplificado por dois modelos originais. São Paulo, 1987.[Tese de Livre-Docência - Faculdade de Medicina da USP]
- 53.MASSAD, E.; STRUCHINER, C.J.; AZEVEDO NETO, R.S.; HALLORAN, M.E. & BRUNET, R.C. The impact of heterogeneity in individual frailty on the dynamics of infections. **Mathematical modelling in Scientific Computer.** *in press.*
- 54.MAU, J. On a graphical method for the detection of time-dependent effects of covariates in survival data. **Appl. Statist.**, 35:245-55, 1986.
- 55.McDOUGAL, W.S.; KIRCHNER JR., F.K.; EDWARDS, R.H. & KILLION, L.T. Treatment of carcinoma of the penis: the case for primary lymphadenectomy. **J. Urol.**, 136:38-41, 1986.
- 56.McKNIGHT, B. Survival Analysis. Lecture notes from the New England Epidemiology Institute, Boston, 1994.
- 57.MILLER JR., R.G. **Survival Analysis**. New York, John Wiley & Sons, 1981.
- 58.MINISTÉRIO DA SAÚDE. Instituto Nacional de Câncer. Câncer no Brasil - dados dos registros de base populacional. Brasília, 1991.

- 59.MIRRA, A.P. & FRANCO, E.L. **Cancer incidence in São Paulo County - Brazil (1969-1973-1978)**. São Paulo, Registro de Câncer de São Paulo/Instituto Ludwig de Pesquisa sobre o Câncer, 1985.
- 60.MIRRA, A.P. & FRANCO, E.L. **Cancer mortality in São Paulo - Brazil**. São Paulo, Registro de Câncer de São Paulo/Instituto Ludwig de Pesquisa sobre o Câncer, 1987 (LICR Cancer Epidemiology Series - v.3)
- 61.NATH, S.; DESAI, G. & MUNKONGE, L. Carcinoma of penis in Zambia: associated problems in management. **Cent. Afr. J. Med.**, **38**:108-11, 1992.
- 62.OLMO, J.M.C.; NAVARRO, J.; LA PEÑA, J.; CISNEROS, J.; VALOR, P.C.; GARCIA-MATREZ, M.J., MARTINEZ-PIÑEIRO, L. & MARTINEZ-PIÑEIRO, J.A. Controversias sobre la linfadenectomía regional en el carcinoma de pene. **Arch. Esp. Urol.**, **44**: 951-5, 1991.
- 63.ORNELLAS, A.A.; SEIXAS, A.L.C. & MORAES, J.R. Analysis of 200 lymphadenectomies in patients with penile carcinoma. **J. Urol.**, **146**:330-2, 1991.
- 64.PARKIN, D.M.; MUIR, C.S.; WHELAN, S.L.; GAO, Y-T.; FERLAY, J. & POWELL, J. **Cancer incidence in five continents - Volume VI**. Lyon, 1992. (IARC Scientific Publications nº 120).
- 65.PRENTICE, R.L. & GLOECKLER, L.A. Regression analysis of grouped survival data with application to breast cancer data. **Biometrics**, **34**:57-67, 1978.
- 66.RASHID, A-M.H.; WILLIAMS, R.M. & HORTON, L.W.L. Malignant melanoma of penis and male urethra. **Uropathology**, **41**: 470-1, 1993.

67. RAVI, R. Correlation between the extent of nodal involvement and survival following groin dissection for carcinoma of the penis. **Br. J. Urol.**, 72:817-9, 1993.
68. ROTHMAN, K.J. **Modern Epidemiology**. Boston, Little, Brown and Company, 1986.
69. SEGAL, M.R. & BLOCH, D.A. A comparison of estimated proportional hazards models and regression trees. **Stat. Med.**, 8:539-50, 1989.
70. SOUZA, J.M.P. Tábua de Vida. São Paulo, Faculdade de Saúde Pública da Universidade de São Paulo, 1992.
71. SRINIVAS, V.; MORSE, M.J.; HERR, H.W.; SOGANI, P.C. & WHITMORE JR., W.R. Penile cancer: relation of extent of nodal metastasis to survival. **J. Urol.**, 137:880-2, 1987.
72. SUFRIN, G. & HUBEN, R. Benign and malignant lesion of the penis. In: GILLET, J.Y. (ed): **Adult and pediatric urology**. 2 ed. Chicago, Year Book Medical Publisher, 1991. p. 1643.
73. TORNESELLO, M.L.; BUONAGURO, F.M.; BETH-GIRALDO, E.; KYALWAZI, S.K. & GIRALDO, G. Human papillomavirus (HPV) DNA in penile carcinomas and in two cell lines from high-incidence areas for genital cancers in Africa. **Int. J. Cancer**, 51:587-92, 1992.
74. UNIÃO INTERNACIONAL CONTRA O CÂNCER (UICC). **Manual de Oncologia Clínica**; trad. Fundação Oncocentro de São Paulo. 5ª ed. São Paulo, 1991.

- 75.VILLA, L.L. & LOPES, A. Human papillomavirus DNA sequences in penile carcinomas in Brazil. **In. J. Cancer**, **37**:853-5, 1986.
- 76.VOELKEL, J.G. & CROWLEY, J. Nonparametric inference for a class of semi-markov processes with censored observations. **Ann. Stati.**, **12**:142-60, 1984.
- 77.WALTER, S.D.; FEINSTEIN, A.R.& WELLS, C.K.. A comparison of multivariable mathematical methods for predicting survival - II. Statistical selection of prognostic variables. **J. Clin. Epidemiol**, **43**:349-59, 1990.
- 78.WELLS, C.K.; FEINSTEIN, A.R. & WALTER, S.D. A comparison of multivariable mathematical methods for predicting survival - III. Accuracy of predictions in generating and challenge sets. **J. Clin. Epidemiol**, **43**:361-72, 1990.
- 79.WISWELL, T.E. Circumcision- an update. **Curr. probl. Pediatr.**, **22**:424-31, 1992.
- 80.WOODBURY, M.A.; MANTON, K.G. & STALLARD, R. Longitudinal models for chronic disease risk: an evaluation of logistic multiple regression and alternatives. **Int. J. Epidemiol**, **10**:187-97, 1981.
- 81.YOUNG, M.J.; REDA, D.J. & WATERS, W.B. Penile carcinoma: a twenty-five-year experience. **Urology**, **38**:529-32, 1991.

ANEXO 1

ESTUDO RETROSPECTIVO DE CA DE PENIS (CEC) - FAP/ILPC

1. Nome (iniciais): _____ 1_1_1_1_1
2. RG: _____ 1_1_1_1_1_1_1_1_1_1
3. Data de admissão: ____/____/____ ____/____/____
4. Idade à admissão para tratamento: _____ anos 1_1_1
5. Estado civil: 11|Solt. 12|Cas. 13|Separ. 14|Outro 19|Ign. 1_1
6. Raça: 11|Branca 12|Amarela 13|Não branca 1_1
7. Procedência (estado): _____ 1_1_1_1
8. Região (vínculo residencial): 11|Rural 12|Urbana 1_1
9. Tipo de sintomas/sinais: 1_1_1_1_1
11|Ferida 12|Dor 13|Inflamação 14|Disúria 15|Prurido 16|Outros
10. Tempo de evolução: _____ meses (999 se ignor.) 1_1_1_1
11. Antecedentes venéreos: 11|Sim: _____ 12|Não 19|Ignor. ... 1_1
12. Postectomia: 11|Não 12|Não relacionada 13|Relacionada 1_1
13. Tratamento prévio(não oncológico): 11|Sim 12|Não 19|Ignor. .. 1_1
14. Situação à 1ª consulta: 1_1_1_1_1
11|Biopsiado 12|Amputado 13|Esvaziado
14|Rxt 15|Qt 16|Intacto
15. Lesões associadas: 11|Não 12|Leucop. 13|Condiloma 14|Outras ... 1_1
16. Tumores concomitantes: _____ 1_1_1_1_1_1_1
17. Estadiamento clínico inicial: a.T: _____ 1_1
b.N: _____ 1_1
c.M: _____ 1_1
18. Localização (cl clinicamente):
a.Glande: 11|Sim 12|Não 19|Ign. 1_1
b.Prepuccio: 11|Sim 12|Não 19|Ign. 1_1
c.Corpo: 11|Sim 12|Não 19|Ign. 1_1
d.Escroto: 11|Sim 12|Não 19|Ign. 1_1
e.Uretra: 11|Sim 12|Não 19|Ign. 1_1
f.Raiz de pênis: 11|Sim 12|Não 19|Ign. 1_1
g.Outros: 11|Sim 12|Não 19|Ign. 1_1
19. Tratamento inicial: 1_1
11|Curativo 12|Paliativo 13|RHD 14|Recusa 15|Proservação
20. Data da amputação: ____/____/____ ____/____/____
21. Cirurgia: 1_1_1_1_1
11|Não 12|Amp. parcial 13|Amp. total 14|Exerese 15|Postec.

ESTUDO RETROSPECTIVO DE CA DE PENIS (CEC) - FAP/ILPC

22. Data do esvaziamento:
 a. Direito: ____/____/____
 b. Esquerdo: ____/____/____
23. Esvaziamento ganglionar:
 a. Direito: 11|Não 12|Inguinal 13|Ing. Ilíaco
 b. Esquerdo: 11|Não 12|Inguinal 13|Ing. Ilíaco
24. Motivo do esvaziamento:
 a. Direito: 11|NR 12|Princípio 13|Necessidade
 b. Esquerdo: 11|NR 12|Princípio 13|Necessidade
25. Complicações relacionadas à:
 a. Amputação:
 11|Infecção 12|Deiscência 13|Estenose 14|Hemorragia 15|
 b. Esvaziamento:
 11|Infecção 12|Deiscência 13|Hemat. 14|Necrose
 15|Linfedema 16|Rupt. vasos 17|Não
26. Radioterapia: 11|Pré-op. 12|Pós-op. 13|NR 19|Ignor.
27. Embolização linfática 11|Sim 12|Não 19|Ign
28. Embolização venosa 11|Sim 12|Não 19|Ign
- 29 Inf. espaço perineural 11|Sim 12|Não 19|Ign
- 30 Inf. linfo plasmocitário 11|Sim 12|Não 19|Ign
31. Quimioterapia: 11|Pré-op. 12|Pós-op. 13|NR 19|Ignor.
32. Tipo de quimioterapia: 11|Bleo 12|Cispl. 13|Outro 14|NR ... 1_
33. Data de recidiva: ____/____/____ ____/____
34. Local: 11|Côto 12|Reg. Homo 13|Reg. Contra 14|Distância .. 1
35. Grau de diferenciação: 11|Bem 12|Moderado 13|Pouco 14|SOE
36. Invasão:
 a. Glande: 11|Sim 12|Não 19|Ign.
 b. Prepúcio: 11|Sim 12|Não 19|Ign.
 c. Corpo: 11|Sim 12|Não 19|Ign.
 d. Escroto: 11|Sim 12|Não 19|Ign.
 e. Uretra: 11|Sim 12|Não 19|Ign.
 f. Raiz de pênis: 11|Sim 12|Não 19|Ign.
 g. Outros: 11|Sim 12|Não 19|Ign.
37. Total de gânglios examinados:
 a. Direito:
 b. Esquerdo: 1_
38. Total de gânglios positivos:
 a. Direito:
 b. Esquerdo: 1_

PENIS

ICD-O T-187

Centre Case No. Ident. No.
 Patient Reg. Year
 Date of Birth Age Sex Ann. Month

PRE-TREATMENT CLINICAL CLASSIFICATION (TNM)

— Please check one or more items in each section as applicable. See minimum requirements for assessment on reverse side.

Site and Extent of Tumour	PRIMARY TUMOUR		Category
Site: Preputium ☐ Glans ☐ Shaft ☐ Size in greatest dimension cm	Size	Local Extension
Pre-invasive carcinoma (carcinoma in situ) ☐		Tis
	No evidence of primary tumour ☐		T0
	2 cm or less ☐	Strictly superficial or exophytic ☐	T1
	More than 2 to 5 cm ☐	Minimal extension ☐	T2
	More than 5 cm ☐	Deep extension including urethra ☐	T3
	Tumour infiltrating neighbouring structures ☐		T4
	Primary tumour not assessed ☐		TX
	REGIONAL LYMPH NODES		Category
	No evidence of involvement ☐		N0
	Evidence of involvement	Movable	Unilateral ☐
Bilateral ☐			N2
Fixed nodes ☐		N3	
Regional lymph nodes not assessed ☐		NX	
DISTANT METASTASES		Category	
No evidence of distant metastases ☐		M0	
Evidence of distant metastases (specify) ☐		M1	
Distant metastases not assessed ☐		MX	
Regional lymph nodes: Histologically: negative ☐ positive ☐ Histological diagnosis: Grade: Category: T N M Stage: G			
Information obtained: Before admission ☐ At Centre ☐ Examiner Date			

POST-SURGICAL HISTOPATHOLOGICAL CLASSIFICATION (p.TNM)

— Please check one or more items in each section as applicable. See minimum requirements for assessment on reverse side.

Site and Extent of Tumour	PRIMARY TUMOUR		Category
Site: Preputium ☐ Glans ☐ Shaft ☐ Size in greatest dimension cm	Size	Local Extension
Pre-invasive carcinoma (carcinoma in situ) ☐		pTis
	No evidence of primary tumour ☐		pT0
	2 cm or less ☐	Strictly superficial or exophytic ☐	pT1
	More than 2 to 5 cm ☐	Minimal extension ☐	pT2
	More than 5 cm ☐	Deep extension including urethra ☐	pT3
	Tumour invading neighbouring structures ☐		pT4
	Primary tumour not assessed ☐		pTX
	REGIONAL LYMPH NODES		Category
	No evidence of invasion ☐		pN0
	Evidence of invasion	Movable	Unilateral ☐
Bilateral ☐			pN2
Fixed nodes ☐		pN3	
Regional lymph nodes not assessed ☐		pNX	
DISTANT METASTASES		Category	
No evidence of distant metastases ☐		pM0	
Evidence of distant metastases (specify) ☐		pM1	
Distant metastases not assessed ☐		pMX	
Regional lymph nodes: Histologically: negative ☐ positive ☐ Histological diagnosis: Grade: Category: pT pN pM Stage: G			
Information obtained: Before admission ☐ At Centre ☐ Clinician Date			

ANEXO 3

TABELAS

Tabela 8. Características sócio-demográficas dos pacientes com câncer de pênis atendidos no Hospital A.C. Camargo, segundo os diferentes bancos de dados. 1953 - 1985.

variável	categoria	B1			A1			A2			A3		
		Nº	%*	% de óbito**	Nº	%*	% de óbito**	Nº	%*	% de óbito**	Nº	%*	% de óbito**
idade (anos)	18 - 39	93	14,3	33,3	17	10,5	29,4	46	11,8	26,1	10	10,2	-
	40 - 59	331	51,1	24,8	84	51,9	25,0	206	52,7	19,9	50	51,0	22,0
	60 - 87	224	34,6	22,8	61	37,7	18,0	139	35,5	22,3	38	38,8	26,3
raça	branco	499	77,0	26,3	123	75,9	22,8	294	75,2	22,4	71	72,4	18,3
	não branco	149	33,0	22,1	39	24,1	23,1	97	24,8	18,6	27	27,6	29,6
estado civil	casado	379	58,6	24,0	99	61,1	22,2	235	60,1	21,3	50	51,0	24,0
	não casado	20	5,0	30,0	4	2,5	25,0	9	2,3	33,3	5	5,1	40,0
	ignorado	249	38,4	26,9	59	36,4	23,7	147	37,6	21,1	43	43,9	16,3
zona	rural	282	45,0	25,5	59	36,4	23,7	174	44,5	23,0	41	41,8	26,8
	urbana	345	55,0	24,9	99	61,1	23,2	200	51,2	20,5	54	55,1	18,5
	ignorado	21	3,2	28,6	4	2,5	-	17	4,3	17,6	3	3,1	-
procedên- cia	SE	484	77,4	26,7	130	80,2	23,8	190	74,2	34,2	72	73,5	20,8
	S	40	6,4	15,0	7	4,3	14,3	27	6,9	7,4	10	10,2	10,0
	NE	62	9,9	19,4	14	8,6	7,1	36	9,2	22,2	7	7,1	57,1
	N+CO	39	6,3	28,2	7	4,3	57,1	19	4,9	31,6	6	6,1	16,7
	ignorado	23	3,5	26,1	4	2,5	-	19	4,9	15,8	3	3,1	-
ano de início	1953 - 1959	137	21,1	34,3	35	21,6	20,0	94	24,0	28,7	17	17,3	29,4
	1960 - 1969	214	33,1	29,0	54	33,3	29,6	130	33,2	25,4	36	36,7	30,6
	1970 - 1979	192	29,6	22,9	51	31,5	21,6	96	24,6	19,8	29	29,6	13,8
	1980 - 1985	105	16,2	10,5	22	13,6	13,6	71	18,2	7,0	16	16,3	6,3
Total		648	100,0	25,3	162	100,0	22,8	391	100,0	21,5	98	100,0	21,4

*: porcentagem calculada em relação ao total da amostra. Havendo categoria ignorada, a porcentagem foi calculada excluindo-se os ignorados

**: porcentagem calculada em relação ao total de pacientes da categoria

Tabela 9. Variáveis clínicas e sintomas pré-tratamento dos pacientes com câncer de pênis atendidos no Hospital A.C. Camargo, segundo os diferentes bancos de dados. 1953 - 1985.

variável	categoria	B1			A1			A2			A3		
		Nº	%*	% de óbito**	Nº	%*	% de óbito**	Nº	%*	% de óbito**	Nº	%*	% de óbito**
doença venérea	sim	165	56,3	21,2	49	30,2	20,4	102	26,1	18,6	27	27,6	14,8
	não	128	43,7	21,9	29	17,9	31,0	81	20,7	18,5	18	18,4	11,1
	ignorado	355	54,8	28,5	84	51,9	21,4	208	53,2	24,0	53	54,1	28,3
tratamento não oncol. prévio	sim	245	82,5	23,7	66	40,7	16,7	160	40,9	21,9	38	38,8	23,7
	não	52	17,5	19,2	12	7,4	16,7	36	9,2	13,9	7	7,1	14,3
	ignorado	351	54,2	27,4	84	51,9	28,6	195	49,9	22,6	53	54,1	20,8
tratamento onc. prévio	sim	88	13,6	40,9	22	13,6	40,9	9	2,3	22,2	2	2,0	50,0
	não	560	86,4	22,9	140	86,4	20,0	382	97,7	21,5	96	98,0	20,8
sintoma: ferida	sim	645	99,5	25,4	161	99,4	23,0	390	99,7	21,5	98	100,0	21,4
	não	3	0,5	-	1	0,6	-	1	0,3	-	-	-	-
sintoma: dor	sim	135	20,8	21,5	33	20,4	15,2	90	23,0	15,6	25	25,5	16,0
	não	513	79,2	26,3	129	79,6	24,8	301	77,0	23,3	73	74,5	23,3
sintoma: inflamação	sim	94	14,5	33,0	24	14,8	29,2	55	14,1	32,7	17	17,3	23,5
	não	554	85,5	24,0	138	85,2	21,7	336	85,9	19,6	81	82,7	21,0
sintoma: disúria	sim	80	12,3	22,5	22	13,6	13,6	52	13,3	17,3	16	16,3	31,3
	não	568	87,7	25,7	140	86,4	24,3	339	86,7	22,1	82	83,7	19,5
sintoma: prurido	sim	145	22,4	20,7	43	26,5	20,9	101	25,8	14,9	26	26,5	15,4
	não	503	77,6	26,6	119	73,5	23,5	290	74,2	23,8	72	73,5	23,6
tempo de evolução da doença (meses)	1 - 6	239	38,6	24,3	59	36,4	22,0	160	40,9	22,5	39	39,8	17,9
	7 - 12	178	28,8	25,8	43	26,5	18,6	106	27,1	23,6	31	31,6	29,0
	13 - 23	31	5,0	38,7	7	4,3	42,9	18	4,6	27,8	4	4,1	-
	24 - 348	171	27,6	21,6	43	26,5	18,6	107	27,4	16,8	22	22,4	22,7
	ignorado	29	4,5	37,9	10	6,2	50,0	-	-	-	2	2,0	-
Total		648	100,0	25,3	162	100,0	22,8	391	100,0	21,5	98	100,0	21,4

*: porcentagem calculada em relação ao total da amostra. Havendo categoria ignorada, a porcentagem foi calculada excluindo-se os ignorados

**: porcentagem calculada em relação ao total de pacientes da categoria

Tabela 10. Variáveis relacionadas ao tumor dos pacientes atendidos no Hospital A.C. Camargo, segundo os diferentes bancos de dados. 1953 - 1985.

variável	categoria	B1			A1			A2			A3		
		Nº	%*	% de óbito**	Nº	%*	% de óbito**	Nº	%*	% de óbito**	Nº	%*	% de óbito**
lesão	não	635	98,0	25,8	155	95,7	23,9	380	97,2	22,1	97	99,0	21,6
associada	sim	13	2,0	-	7	4,3	-	11	2,8	-	1	1,0	-
tu conco-	não	632	97,5	25,8	160	98,8	23,1	381	97,4	22,0	95	96,9	22,1
mitante	sim	16	2,5	6,3	2	1,2	-	10	2,6	-	3	3,1	-
local:	não	17	3,2	11,8	5	3,1	-	12	3,1	8,3	-	-	-
glande	sim	514	96,8	24,7	128	79,0	21,1	379	96,9	21,9	98	100,0	21,4
	ignorado	117	18,1	34,2	29	17,9	34,5	-	-	-	-	-	-
local:	não	44	8,4	13,6	9	5,6	22,2	40	10,2	12,5	7	7,1	-
prepúcio	sim	482	91,6	24,1	124	76,5	20,2	351	89,8	22,5	91	92,9	23,1
	ignorado	122	18,8	34,4	29	17,9	34,5	-	-	-	-	-	-
local:	não	211	42,7	15,2	53	32,7	17,0	195	49,9	14,4	52	53,1	15,4
corpo	sim	283	57,3	30,4	73	45,1	23,3	196	50,1	28,6	46	46,9	28,3
	ignorado	154	23,8	29,9	36	22,2	30,6	-	-	-	-	-	-
local:	não	506	94,2	22,3	129	79,6	20,2	366	93,6	19,9	93	94,9	4,9
escroto	sim	31	5,8	45,2	4	2,5	17,9	25	6,4	44,0	5	5,1	40,0
	ignorado	111	17,1	33,3	29	17,9	34,5	-	-	-	-	-	-
local:	não	341	74,9	19,1	86	53,1	15,1	296	75,7	17,6	75	76,5	18,7
uretra	sim	114	25,1	34,2	30	18,5	30,0	95	24,3	33,7	23	23,5	30,4
	ignorado	193	29,8	31,1	46	28,4	32,6	-	-	-	-	-	-
local: raiz	não	457	85,6	20,6	119	73,5	18,5	334	85,4	16,8	86	87,8	18,6
	sim	77	14,4	42,9	13	8,0	30,8	54	14,6	49,1	12	12,2	41,7
	ignorado	114	17,6	32,5	30	18,5	36,7	-	-	-	-	-	-
local:	não	509	95,1	21,4	129	79,6	19,4	369	94,4	18,2	94	95,9	20,2
outro	sim	26	4,9	73,1	4	2,5	50,0	22	5,6	77,3	4	4,1	50,0
	ignorado	113	17,4	31,9	29	17,9	34,5	-	-	-	-	-	-

continua

continuação da tabela 10...

variável	categoria	B1			A1			A2			A3		
		Nº	%*	% de óbito**	Nº	%*	% de óbito**	Nº	%*	% de óbito**	Nº	%*	% de óbito**
total de	0	525	82,0	22,9	135	83,3	20,0	338	86,4	20,1	82	83,7	19,5
gânglios	1	35	5,4	31,4	13	8,0	30,8	17	4,3	23,5	4	4,1	-
positivos	2-3	40	6,3	32,5	4	2,5	50,0	19	4,9	36,8	8	8,2	50,0
	4 e +	40	6,3	40,0	8	4,9	37,5	17	4,3	29,4	2	2,0	50,0
	ignorado	8	1,2	50,0	2	1,2	50,0	-	-	-	2	2,0	-
estad. T	0+I	47	9,0	6,4	16	9,9	-	44	11,3	6,8	6	6,1	-
	II	190	36,3	17,4	42	25,9	23,8	155	39,6	15,5	47	48,0	17,0
	III	237	45,2	27,0	66	40,7	22,7	147	37,6	23,8	37	37,8	27,0
	IV	50	9,5	46,0	8	4,9	25,0	45	11,5	48,9	8	8,2	37,5
	ignorado	124	19,1	33,0	30	18,5	33,3	-	-	-	-	-	-
estad. N	0	336	55,7	13,1	86	53,1	14,0	252	64,5	11,5	70	71,4	15,7
	I	61	10,1	24,6	17	10,5	17,6	41	10,5	19,5	9	9,2	11,1
	II	92	15,3	30,4	21	13,0	9,5	51	13,0	35,3	12	12,2	41,7
	III	114	18,9	60,5	24	14,8	70,8	47	12,0	61,7	7	7,1	57,1
	ignorado	45	6,9	17,8	14	8,6	21,4	-	-	-	-	-	-
estad. M	0	627	99,1	25,5	154	95,1	23,4	388	99,2	21,6	97	99,0	21,6
	I	6	0,9	16,7	3	1,9	-	3	0,8	-	1	1,0	-
	ignorado	15	2,3	20,0	5	3,1	20,0	-	-	-	-	-	-
grau de	bem	371	57,5	20,5	89	54,9	22,5	230	58,8	17,4	52	53,1	19,2
diferenc.	mod+pouco	35	5,4	45,7	9	5,6	22,2	21	5,4	38,9	3	3,1	33,3
do tumor	SOE	239	37,1	30,1	64	39,5	23,4	140	35,8	25,7	43	43,9	23,3
	ignorado	3	0,5	-	-	-	-	-	-	-	-	-	-
Total		648	100,0	25,3	162	100,0	22,8	391	100,0	21,5	98	100,0	21,4

*: porcentagem calculada em relação ao total da amostra. Havendo categoria ignorada, a porcentagem foi calculada excluindo-se os ignorados

**: porcentagem calculada em relação ao total de pacientes da categoria

Tabela 11. Variáveis relacionadas aos procedimentos e terapia dos pacientes com câncer de pênis atendidos no Hospital A.C. Camargo, segundo os diferentes bancos de dados.. 1953 - 1985.

variável	categoria	B1			A1			A2			A3		
		Nº	%*	% de óbito**	Nº	%*	% de óbito**	Nº	%*	% de óbito**	Nº	%*	% de óbito**
cirurgia	amp total	202	31,2	24,3	52	32,1	19,2	123	31,5	20,3	30	30,6	30,0
	outras cir.	250	38,6	10,4	66	40,7	10,6	166	42,5	8,4	43	43,9	14,0
	NR	196	30,2	45,4	44	27,2	45,4	102	26,1	44,1	25	25,5	24,0
linfadenectomia	profilática	68	10,5	25,0	19	11,7	10,5	41	10,5	17,1	10	10,2	40,0
	necessária	441	68,1	23,8	109	67,3	21,1	283	72,4	21,2	70	71,4	15,7
	NR	139	21,4	30,2	34	21,0	35,3	67	17,1	25,4	18	18,4	33,3
esvazi. gang. direito	NR	462	71,3	23,4	116	71,6	21,6	291	74,4	21,0	72	73,5	15,3
	inguinal	61	9,4	29,5	16	9,9	31,3	34	8,7	23,5	10	10,2	30,0
	ing+ilíaco	125	19,3	30,4	30	18,5	23,3	66	16,9	22,7	16	16,3	43,8
esvazi. gang. esquerdo	NR	461	71,1	24,3	115	71,0	21,7	290	74,2	21,0	71	72,4	15,5
	inguinal	61	9,4	29,5	16	9,9	25,0	30	7,7	23,3	9	9,2	33,3
	ing+ilíaco	126	19,4	27,0	31	19,1	25,8	71	18,2	22,5	18	18,4	38,9
motivo do esv. gang. direito	NR	462	71,3	23,4	116	71,6	21,6	291	74,4	21,0	72	73,5	15,3
	profilático	55	8,5	18,2	18	11,1	5,6	38	9,7	18,4	10	10,2	40,0
	necessário	131	20,2	35,1	28	17,3	29,3	62	15,9	25,8	16	16,3	37,5
motivo do esv. gang. esquerdo	NR	461	71,1	24,3	115	71,0	21,7	290	74,2	21,0	71	72,4	15,5
	profilático	53	8,2	30,2	15	9,3	13,3	33	8,4	21,2	8	8,2	50,0
	necessário	134	20,7	26,9	32	19,8	31,3	68	17,4	23,5	19	19,4	31,6
tempo esv. gang.	NR	482	74,4	24,9	122	75,3	22,1	298	76,2	20,8	73	74,5	15,1
	D/E <7 m.	115	17,7	30,4	26	16,0	15,4	63	16,1	27,0	18	18,4	38,9
	outros	51	7,9	27,5	14	8,6	42,9	30	7,7	16,7	7	7,1	42,9
quimio.	não	584	90,1	21,7	150	92,6	20,0	362	92,6	18,5	92	93,9	20,7
	sim	64	9,9	57,8	12	7,4	58,3	29	7,4	58,6	6	6,1	33,3
radiot.	não	548	84,6	22,3	144	88,9	20,1	335	85,7	18,5	88	89,8	19,3
	sim	100	15,4	42,0	18	11,1	44,4	56	14,3	39,3	10	10,2	40,0
Total		648	100,0	25,3	162	100,0	22,8	391	100,0	21,5	98	100,0	21,4

*: porcentagem calculada em relação ao total da amostra. Havendo categoria ignorada, a porcentagem foi calculada excluindo-se os ignorados

** : porcentagem calculada em relação ao total de pacientes da categoria

Tabela 12. Resultados da análise univariada pelo método de Kaplan-Meier. Características sócio-demográficas dos pacientes com câncer de pênis atendidos no Hospital A.C. Camargo, segundo os diferentes bancos de dados. 1953 - 1985.

variável	categoria	B1			A1			A2			A3		
		% de óbito *	S(60) **	p (log- rank)	% de óbito *	S(60) **	p (log- rank)	% de óbito *	S(60) **	p (log- rank)	% de óbito *	S(60) **	p (log- rank)
idade (anos)	18 - 39	34	59	0,1802	36	59	0,4552	27	70	0,4281	-	100	0,1775
	40 - 59	25	71		26	70		20	77		21	76	
	60 - 87	24	71		20	76		22	71		26	67	
raça	branco	27	68	0,3704	25	72	0,9190	23	72	0,3236	18	78	0,1238
	não branco	22	72		23	67		16	80		29	62	
estado civil	casado	25	72	0,2114	23	73	0,9416	22	75	0,1417	21	75	0,1184
	não casado	38	53		25	75		43	31		67	33	
zona	rural	26	69	0,9259	25	70	0,9685	21	73	0,8785	22	73	0,8936
	urbana	26	70		25	70		22	74		20	76	
procedên- cia	SE	28	67	0,2890	26	69	0,3350	23	72	0,2998	22	73	0,4600
	S	17	79		14	75		8	89		10	86	
	NE	16	80		8	90		17	80		40	53	
	N+CO	27	69		50	50		29	63		-	100	
ano de início	1953 - 1959	36	61	0,0191	22	75	0,8484	29	68	0,1129	29	71	0,1667
	1960 - 1969	30	67		31	66		25	72		32	61	
	1970 - 1979	22	73		23	73		19	77		11	85	
	1980 - 1985	12	83		16	80		8	88		7	91	
no. óbitos /total de pac		164 /648			37 /162			84 /391			21 / 98		

*: porcentagem calculada em relação ao total de pacientes da categoria

** S(60): porcentagem de sobrevida aos 60 meses de seguimento

Tabela 13. Resultados da análise univariada pelo método de Kaplan-Meier. Variáveis clínicas e sintomas pré-tratamento dos pacientes com câncer de pênis atendidos no Hospital A.C. Camargo, segundo os diferentes bancos de dados. 1953 - 1985.

variável	categoria	B1			A1			A2			A3		
		% de óbito *	S(60) **	p (log- rank)	% de óbito *	S(60) **	p (log- rank)	% de óbito *	S(60) **	p (log- rank)	% de óbito *	S(60) **	p (log- rank)
doença venérea	sim	20	75	0,4178	20	77	0,1313	16	80	0,2723	8	92	0,6289
	não	23	71		33	55		20	72		11	87	
trat. não onc. prév.	sim	26	69	0,6024	19	77	0,6063	23	72	0,4420	27	66	0,5598
	não	21	72		18	57		16	78		14	86	
tratamento onc. prévio	não	23	72	<0,000 1	21	75	0,0026	22	74	0,6521	20	75	0,3648
	sim	44	49		47	40		25	69		50	50	
sintoma: ferida	não	-	100	0,4258	_A	_A	_A	_A	_A	_A	_A	_A	_A
	sim	26	69		_A	_A		_A	_A		_A	_A	
sintoma: dor	não	27	68	0,3464	26	69	0,6161	23	73	0,2478	22	73	0,6534
	sim	23	73		17	80		16	79		16	77	
sintoma: inflamação	não	24	71	0,0410	23	72	0,4000	19	77	0,0037	19	75	0,5410
	sim	36	59		32	65		37	58		29	70	
sintoma: disúria	não	26	69	0,7166	25	69	0,3006	22	73	0,5267	18	77	0,1648
	sim	24	71		17	82		17	79		33	61	
sintoma: prurido	não	27	68	0,1881	26	70	0,6958	24	72	0,1289	23	73	0,5360
	sim	22	74		21	74		16	81		16	78	
tempo de evolução da doença (meses)	1 - 6	24	70	0,4002	22	73	0,3723	22	72	0,3302	14	84	0,1108
	7 - 12	28	67		21	76		26	70		35	55	
	13 - 23	35	57		50	40		24	74		-	100	
	24 e +	23	74		20	75		16	81		22	73	
no. óbitos /total		164 /648			37 /162			84 /391			21 / 98		

*: porcentagem calculada em relação ao total de pacientes da categoria

** S(60): porcentagem de sobrevivência aos 60 meses de seguimento

A: todos os pacientes referiram ferida

Tabela 14. Resultados da análise univariada pelo método de Kaplan-Meier. Variáveis relacionadas ao tumor dos pacientes atendidos no Hospital A.C. Camargo, segundo os diferentes bancos de dados. 1953 - 1985.

variável	categoria	B1			A1			A2			A3		
		% de óbito *	S(60) **	p (log- rank)	% de óbito *	S(60) **	p (log- rank)	% de óbito *	S(60) **	p (log- rank)	% de óbito *	S(60) **	p (log- rank)
lesão associada	não	27	68	0,0415	26	70	0,1599	22	73	0,0912	21	74	0,5858
	sim	-	100		-	100		-	100		-	100	
tu conco- mitante	não	27	68	0,0625	25	70	0,4081	22	73	0,0950	21	73	0,3448
	sim	6	93		-	100		-	100		-	100	
local: glande	não	12	85	0,2397	-	100	0,2448	8	92	0,2516	A	A	A
	sim	24	71		22	74		22	73		A	A	
local: prepúcio	não	13	85	0,0903	25	73	0,9569	11	87	0,0994	-	100	0,1770
	sim	25	71		21	75		23	72		23	72	
local: corpo	não	15	81	<0,000 1	18	79	0,1488	14	82	0,0002	16	80	0,1792
	sim	32	63		25	70		29	65		27	67	
local: escroto	não	23	73	<0,000 1	21	75	0,5739	20	76	0,0001	19	77	0,0017
	sim	52	41		25	67		50	41		50	0	
local: uretra	não	19	77	<0,000 1	16	81	0,0294	18	79	0,0001	18	79	0,1900
	sim	36	56		30	61		35	56		32	60	
local: raiz	não	21	75	<0,000 1	20	77	0,2227	17	80	<0,000 1	18	79	0,0257
	sim	48	38		27	51		54	26		44	27	
local: outro	não	22	74	<0,000 1	20	76	0,0611	18	78	<0,000 1	19	76	0,0325
	sim	81	18		67	33		88	9		67	33	
total de gânglios positivos	0	24	72	0,0319	22	75	0,4499	20	75	0,6124	18	77	0,1471
	1	31	66		31	67		24	77		-	100	
	2-3	29	63		50	33		33	64		50	44	
	4 e +	40	53		38	54		29	68		50	50	
estad. T	0+I	5	95	<0,000 1	-	100	0,1717	5	95	<0,000 1	-	100	0,0456
	II	18	77		26	70		16	80		18	78	
	III	27	69		23	73		23	73		23	73	
	IV	55	33		29	42		57	28		50	25	
estad. N	0	14	83	<0,000 1	15	82	<0,000 1	12	85	<0,000 1	16	80	0,0155
	I	21	72		19	79		18	75		11	86	
	II	30	63		10	84		33	59		36	61	
	III	71	22		84	0		72	20		60	25	
estad. M	0	26	69	0,1679	25	71	0,7943	22	74	0,6534	21	75	0,7475
	I	33	67		-	100		-	100		-	100	
grau de	bem	20	76	<0,000 1	23	72	0,7259	17	79	0,0075	16	82	0,3875
diferenc. do tumor	mod+pouco	54	39		25	66		39	53		33	67	
	SOE	32	62		26	70		27	68		26	68	
no. óbitos /total pac.		164 /648			37 /162			84 /391			21 /98		

*: porcentagem calculada em relação ao total de pacientes da categoria

** S(60): porcentagem de sobrevivência aos 60 meses de seguimento

A: todos os pacientes tiveram acometimento de glândula

Tabela 15. Resultados da análise univariada pelo método de Kaplan-Meier. Variáveis relacionadas aos procedimentos e terapia dos pacientes com câncer de pênis atendidos no Hospital A.C. Camargo, segundo os diferentes bancos de dados. 1953 - 1985.

variável	categoria	B1			A1			A2			A3		
		% de óbito *	S(60) **	p (log- rank)	% de óbito *	S(60) **	p (log- rank)	% de óbito *	S(60) **	p (log- rank)	% de óbito *	S(60) **	p (log- rank)
cirurgia	amp total	23	73	<0,000 1	20	75	<0,000 1	19	77	<0,000 1	28	70	0,0959
	outras cir.	10	87		11	88		8	90		12	84	
	NR	54	35		58	20		53	35		31	56	
linfadenectomia	profilática	22	76	0,5435	11	89	0,2240	15	84	0,4319	40	58	0,1863
	necessária	29	66		23	71		23	75		29	69	
	NR	26	69		35	61		22	72		15	80	
esvazi. gang. direito	NR	25	70	0,9241	24	70	0,8953	22	72	0,7201	15	81	0,1624
	inguinal	30	67		31	66		24	75		30	67	
	ing.+ilíaco	28	69		23	75		19	79		40	58	
esvazi. gang. esquerdo	NR	26	68	0,7264	24	71	0,9757	22	72	0,6565	15	81	0,2613
	inguinal	30	67		25	72		23	75		33	65	
	ing.+ilíaco	24	73		26	71		19	79		35	62	
motivo do esv.gang. direito	NR	25	70	0,1081	24	70	0,0703	22	72	0,6076	15	81	0,1359
	profilático	17	82		6	94		16	83		40	58	
	necessário	33	62		39	59		23	75		33	64	
motivo do esv.gang. esquerdo	NR	26	68	0,8265	24	71	0,4148	22	72	0,6502	15	81	0,0552
	profilático	27	72		13	87		19	81		50	50	
	necessário	26	70		31	63		21	76		28	69	
tempo esv. gang.	NR	26	69	0,9387	24	70	0,1252	22	73	0,5709	14	82	0,1346
	D/E <7 m.	28	70		15	84		23	74		35	63	
	outros	28	69		43	50		17	83		43	46	
químio.	não	22	74	<0,000 1	21	75	0,0004	19	77	<0,000 1	20	76	0,2946
	sim	58	24		58	17		57	33		33	60	
radiot.	não	23	73	0,0002	21	75	0,0068	18	78	0,0016	18	78	0,0827
	sim	43	51		50	42		39	54		40	46	
no. óbitos /total pac.		164	/648		37	/162		84	/391		21	/98	

*: porcentagem calculada em relação ao total de pacientes da categoria

** S(60): porcentagem de sobrevivência aos 60 meses de seguimento

Tabela 16. Resultados da análise univariada pelo modelo de Cox das características sócio-demográficas dos pacientes com câncer de pênis atendidos no Hospital A.C. Camargo. 1953 - 1985. Modelo com B1

variável	categoria	coef.(β_i)	HR	IC(HR) _{95%}	p(categ.)	p(variável)
idade (anos)	18 - 39	referência	1,00	-	-	0,2196
	40 - 59	-0,3333	0,72	[0,47;1,08]	0,1143	
	60 - 87	-0,3889	0,68	[0,43;1,06]	0,0879	
raça	branco	referência	1,00	-	-	0,5034
	não branco	-0,1287	0,88	[0,60;1,29]	0,5088	
estado civil	casado	referência	1,00	-	-	0,3017
	não casado	0,4279	1,53	[0,67;3,51]	0,3104	
	ignorado	0,2171	1,24	[0,91;1,70]	0,1779	
zona	rural	referência	1,00	-	-	0,8232
	urbana	-0,0954	0,91	[0,67;1,24]	0,5504	
	ignorado	0,0242	1,02	[0,45;2,36]	0,9546	
procedên- cia	SE	referência	1,00	-	-	0,5132
	S	-0,6166	0,54	[0,24;1,22]	0,1399	
	NE	-0,1700	0,84	[0,47;1,53]	0,5734	
	N+CO	0,1574	1,17	[0,63;2,17]	0,6163	
	ignorado	-0,0936	0,91	[0,40;2,07]	0,8227	
ano de início	1953 - 1959	referência	1,00	-	-	0,0065
	1960 - 1969	-0,1009	0,90	[0,62;1,32]	0,6019	
	1970 - 1979	-0,3997	0,67	[0,44;1,01]	0,0568	
	1980 - 1985	-0,9660	0,38	[0,20;0,73]	0,0040	

Tabela 17. Resultados da análise univariada pelo modelo de Cox das variáveis clínicas e sintomas pré-tratamento dos pacientes com câncer de pênis atendidos no Hospital A.C. Camargo. 1953 - 1985. Modelo com B1

variável	categoria	coef.(β_i)	HR	IC(HR) _{95%}	p(categ.)	p(variável)
doença venérea	sim	referência	1,00	-	-	0,1013
	não	0,0319	1,03	[0,63;1,70]	0,8998	
	ignorado	0,3533	1,42	[0,97;2,09]	0,0717	
tratamento não oncol. prévio	sim	referência	1,00	-	-	0,4903
	não	-0,2002	0,82	[0,42;1,60]	0,5588	
	ignorado	0,1310	1,14	[0,82;1,58]	0,4309	
tratamento onc. prévio	não	referência	1,00	-	-	0,0001
	sim	0,7849	2,19	[1,51;3,18]	<0,0001	
sintoma: ferida	não	referência	1,00	-	-	0,2438
	sim	15,0080	∞	nd*	...	
sintoma: dor	não	referência	1,00	-	-	0,2924
	sim	-0,2108	0,81	[0,54;1,21]	0,3032	
sintoma: inflamação	não	referência	1,00	-	-	0,1403
	sim	0,3037	1,36	[0,92;2,00]	0,1279	
sintoma: disúria	não	referência	1,00	-	-	0,6194
	sim	-0,1222	0,89	[0,54;1,44]	0,6249	
sintoma: prurido	não	referência	1,00	-	-	0,0985
	sim	-0,3228	0,72	[0,49;1,08]	0,1100	
tempo de evolução da doença (mês)	1 - 6	referência	1,00	-	-	0,2653
	7 - 12	0,1476	1,16	[0,79;1,71]	0,4553	
	13 - 23	0,6011	1,82	[0,98;3,40]	0,0581	
	24 - 348	-0,0634	0,94	[0,62;1,42]	0,7631	
	ignorado	0,4107	1,51	[0,79;2,87]	0,2118	

*nd: não definido

Tabela 18. Resultados da análise univariada pelo modelo de Cox das variáveis relacionadas ao tumor dos pacientes atendidos no Hospital A.C. Camargo. 1953 - 1985. Modelo com B1

variável	categoria	coef.(βi)	HR	IC(HR) _{95%}	p(categ.)	p(variável)
lesão associada	não	referência	1,00	-	-	0,0034
	sim	-15,0424	0,00	nd*	...	
tu conco- mitante	não	referência	1,00	-	-	0,0181
	sim	-1,7184	0,18	[0,03;1,28]	0,0867	
local: glande	não	referência	1,00	-	-	0,0219
	sim	0,8899	2,43	[0,60;9,85]	0,2119	
	ignorado	1,3123	3,72	[0,90;15,37]	0,0701	
local: prepúcio	não	referência	1,00	-	-	0,0097
	sim	0,6541	1,92	[0,85;4,37]	0,1182	
	ignorado	1,0773	2,94	[1,25;6,91]	0,0136	
local: corpo	não	referência	1,00	-	-	<0,0001
	sim	0,9170	2,50	[1,67;3,76]	<0,0001	
	ignorado	0,8403	2,32	[1,48;3,64]	0,0003	
local: escroto	não	referência	1,00	-	-	0,0003
	sim	1,1045	3,02	[1,73;5,27]	0,0001	
	ignorado	0,5142	1,67	[1,15;2,43]	0,0067	
local: uretra	não	referência	1,00	-	-	<0,0001
	sim	0,8499	2,34	[1,57;3,48]	<0,0001	
	ignorado	0,6369	1,89	[1,33;2,69]	0,0004	
local: raiz	não	referência	1,00	-	-	<0,0001
	sim	1,1761	3,24	[2,17;4,83]	<0,0001	
	ignorado	0,6054	1,83	[1,25;2,68]	0,0018	
local: outro	não	referência	1,00	-	-	<0,0001
	sim	1,6315	5,11	[3,13;8,34]	<0,0001	
	ignorado	0,5174	1,68	[1,15;2,45]	0,0072	
total de gânglios positivos	0	referência	1,00	-	-	
	1	0,1610	1,18	[0,63;2,18]	0,6095	0,1581
	2 -3	0,2050	1,22	[0,69;2,18]	0,4827	
	4 e +	0,6542	1,92	[1,14;3,25]	0,0142	
	ignorado	0,6790	1,97	[0,73;5,34]	0,1816	
estad. T	0+I	referência	1,00	-	-	<0,0001
	II	1,0378	2,82	[0,87;9,21]	0,0853	
	III	1,6094	5,00	[1,57;15,92]	0,0064	
	IV	2,5534	12,85	[3,85;42,85]	<0,0001	
	ignorado	1,8257	6,21	[1,92;20,05]	0,0023	
estad. N	0	referência	1,00	-	-	<0,0001
	I	0,6623	1,94	[1,08;3,49]	0,0268	
	II	0,9696	2,64	[1,64;4,24]	0,0001	
	III	2,2550	9,54	[6,48;14,03]	<0,0001	
	ignorado	0,2788	1,32	[0,62;2,81]	0,4683	
estad. M	0	referência	1,00	-	-	0,6200
	I	0,9215	2,51	[0,35;18,10]	0,3603	
	ignorado	-0,3105	0,73	[0,23;2,30]	0,5941	
grau de diferenc. do tumor	bem	referência	1,00	-	-	0,0001
	mod+pouco	1,2830	3,61	[2,10;6,20]	<0,0001	
	SOE	0,4573	1,58	[1,14;2,18]	0,0054	
	ignorado	-14,8867	0,00	nd*	0,9911	

*nd: não definido

Tabela 19. Resultados da análise univariada pelo modelo de Cox das variáveis relacionadas aos procedimentos e terapia dos pacientes com câncer de pênis atendidos no Hospital A.C. Camargo, 1953 - 1985. Modelo com B1

variável	categoria	coef.(βi)	HR	IC(HR) _{95%}	p(categ.)	p(variável)
linfadenec- tomia	profilática	referência	1,00	-	-	0,7965
	necessária	0,1405	1,15	[0,69;1,92]	0,5911	
	NR	0,1909	1,21	[0,69;2,13]	0,5067	
esvazi. gang. direito	NR	referência	1,00	-	-	0,7765
	inguinal	0,0193	1,02	[0,62;1,68]	0,9396	
	ing +iliaco	0,1352	1,15	[0,79;1,66]	0,4735	
esvazi. gang. esquerdo	NR	referência	1,00	-	-	0,9310
	inguinal	-0,0595	0,94	[0,57;1,55]	0,8149	
	ing+iliaco	-0,0651	0,94	[0,64;1,38]	0,7396	
motivo do esv.gang direito	NR	referência	1,00	-	-	0,0807
	profilático	-0,4402	0,64	[0,34;1,23]	0,1830	
	necessário	0,2642	1,30	[0,92;1,84]	0,1336	
motivo do esv.gang. esquerdo	NR	referência	1,00	-	-	0,8586
	profilático	0,0230	1,02	[0,61;1,73]	0,9316	
	necessário	-0,0992	0,91	[0,62;1,32]	0,6049	
tempo esv. gang.	NR	referência	1,00	-	-	0,6857
	D/E <7 m.	0,1260	1,13	[0,78;1,66]	0,5140	
	outros	-0,1310	0,88	[0,50;1,53]	0,6437	

Tabela 31. Resultados da análise univariada pelo modelo de Cox das características sócio-demográficas dos pacientes com câncer de pênis atendidos no Hospital A.C. Camargo. 1953 - 1985. Modelo com A1

variável	categoria	coef.(β i)	HR	IC(HR) _{95%}	p(categ.)	p(variável)
idade (anos)	18 - 39	referência	1,00	-	-	0,4651
	40 - 59	-0,3527	0,70	[0,27;1,87]	0,4792	
	60 - 87	-0,6570	0,52	[0,18;1,49]	0,2236	
raça	branco	referência	1,00	-	-	0,6908
	não branco	0,1548	1,17	[0,55;2,48]	0,6868	
estado civil	casado	referência	1,00	-	-	0,9293
	não casado	-0,1074	0,90	[0,12;6,67]	0,9163	
	ignorado	0,1205	1,13	[0,58;2,21]	0,7247	
zona	rural	referência	1,00	-	-	0,2551
	urbana	-0,0574	0,94	[0,49;1,83]	0,8608	
	ignorado	-15,0977	0,00	nd*	0,9875	
procedên- cia	SE	referência	1,00	-	-	0,0990
	S	-0,5800	0,56	[0,08;4,10]	0,5682	
	NE	-1,0293	0,36	[0,05;2,62]	0,3111	
	N+CO	1,0417	2,83	[1,00;8,04]	0,0502	
	ignorado	-15,2420	0,00	nd*	0,9885	
ano de início	1953 - 1959	referência	1,00	-	-	0,7554
	1960 - 1969	0,4258	1,53	[0,63;3,72]	0,3474	
	1970 - 1979	0,1523	1,16	[0,45;3,01]	0,7529	
	1980 - 1985	0,0094	1,01	[0,26;3,92]	0,9892	

nd*: não definido

Tabela 32. Resultados da análise univariada pelo modelo de Cox das variáveis clínicas e sintomas pré-tratamento dos pacientes com câncer de pênis atendidos no Hospital A.C. Camargo. 1953 - 1985. Modelo com A1

variável	categoria	coef.(βi)	HR	IC(HR) _{95%}	p(categ.)	p(variável)
doença venérea	sim	referência	1,00	-	-	0,4282
	não	0,5851	1,80	[0,73;4,43]	0,2038	
	ignorado	0,1207	1,13	[0,52;2,45]	0,7597	
tratamento não oncol. prévio	sim	referência	1,00	-	-	0,2467
	não	0,2768	1,32	[0,29;5,96]	0,7190	
	ignorado	0,5897	1,80	[0,88;3,68]	0,1055	
tratamento onc. prévio	não	referência	1,00	-	-	0,0129
	sim	1,0549	2,87	[1,35;6,10]	0,0061	
sintoma: dor	não	referência	1,00	-	-	0,5497
	sim	-0,2785	0,76	[0,30;1,96]	0,5629	
sintoma: inflamação	não	referência	1,00	-	-	0,4660
	sim	0,3166	1,37	[0,60;3,13]	0,4510	
sintoma: disúria	não	referência	1,00	-	-	0,2451
	sim	-0,6401	0,53	[0,16;1,72]	0,2881	
sintoma: prurido	não	referência	1,00	-	-	0,6255
	sim	-0,1842	0,83	[0,39;1,76]	0,6308	
tempo de evolução da doença (mês)	1 - 6	referência	1,00	-	-	0,2164
	7 - 12	-0,1266	0,88	[0,37;2,13]	0,7782	
	13 - 23	0,8231	2,28	[0,65;8,02]	0,1997	
	24 - 348	-0,2898	0,75	[0,31;1,81]	0,5192	
	ignorado	0,9525	2,59	[0,92;7,28]	0,0706	

Tabela 33. Resultados da análise univariada pelo modelo de Cox das variáveis relacionadas ao tumor dos pacientes atendidos no Hospital A.C. Camargo. 1953 - 1985. Modelo com A1

variável	categoria	coef.(βi)	HR	IC(HR) _{95%}	p(categ.)	p(variável)
local:	não	referência	1,00	-	-	0,1389
prepúcio	sim	-0,0073	0,99	[0,24;4,19]	0,9920	
	ignorado	0,7817	2,19	[0,48;9,99]	0,3134	
local:	não	referência	1,00	-	-	0,0851
corpo	sim	0,6260	1,87	[0,83;4,20]	0,1293	
	ignorado	0,9608	2,61	[1,08;6,32]	0,0329	
local:	não	referência	1,00	-	-	0,1256
escroto	sim	0,4925	1,64	[0,22;12,07]	0,6290	
	ignorado	0,8032	2,23	[1,07;4,64]	0,0315	
local:	não	referência	1,00	-	-	0,0074
uretra	sim	1,0327	2,81	[1,20;6,58]	0,0175	
	ignorado	1,0578	2,88	[1,37;6,06]	0,0053	
local: raiz	não	referência	1,00	-	-	0,0195
	sim	0,9990	2,72	[0,93;7,92]	0,0674	
	ignorado	0,9884	2,69	[1,30;5,56]	0,0077	
local:	não	referência	1,00	-	-	0,0507
outro	sim	1,2318	3,43	[0,81;14,48]	0,0939	
	ignorado	0,8426	2,32	[1,11;4,85]	0,0248	
total de	0	referência	1,00	-	-	0,6564
gânglios	1	0,3515	1,42	[0,50;4,07]	0,5124	
	2 -3	0,6011	1,82	[0,98;3,40]	0,0581	
positivos	4 e +	-0,0634	0,94	[0,62;1,42]	0,7631	
	ignorado	0,4107	1,51	[0,79;2,87]	0,2118	
estad. T	0+I+II	referência	1,00	-	-	0,1555
	III	0,4767	1,61	[0,72;3,59]	0,2433	
	IV	0,8586	2,36	[0,52;10,80]	0,2685	
	ignorado	1,0068	2,74	[1,14;6,59]	0,0247	
estad. N	0	referência	1,00	-	-	<0,0001
	I	0,3599	1,43	[0,40;5,09]	0,5775	
	II	-0,2235	0,80	[0,18;3,58]	0,7699	
	III	3,1006	22,21	[9,54;51,73]	<0,0001	
	ignorado	0,5386	1,71	[0,48;6,08]	0,4043	

Tabela 34. Resultados da análise univariada pelo modelo de Cox das variáveis relacionadas aos procedimentos e terapia dos pacientes com câncer de pênis atendidos no Hospital A.C. Camargo, 1953 - 1985. Modelo com A1

variável	categoria	coef.(β i)	HR	IC(HR) _{95%}	p(categ.)	p(variável)
linfadenectomia	profilática	referência	1,00	-	-	0,1798
	necessária	0,9718	2,64	[0,62;11,22]	0,1876	
	NR	1,1244	3,44	[0,77;15,36]	0,1062	
esvazi. gang. direito	NR	referência	1,00	-	-	0,8911
	inguinal	0,1254	1,13	[0,43;2,96]	0,7980	
	ing.+ilíaco	-0,1487	0,86	[0,37;1,99]	0,7282	
esvazi. gang. esquerdo	NR	referência	1,00	-	-	0,9541
	inguinal	-0,1506	0,86	[0,30;2,47]	0,7799	
	ing+ilíaco	-0,0675	0,94	[0,42;2,07]	0,8680	
motivo do esv.gang direito	NR	referência	1,00	-	-	0,0387
	profilático	-1,6368	0,20	[0,03;1,44]	0,1086	
	necessário	0,3964	1,49	[0,73;3,02]	0,2736	
motivo do esv.gang. esquerdo	NR	referência	1,00	-	-	0,3543
	profilático	-0,8391	0,43	[0,10;1,83]	0,2540	
	necessário	0,1523	1,16	[0,56;2,43]	0,6842	
tempo esv. gang.	NR	referência	1,00	-	-	0,1408
	D/E <7 m.	-0,6721	0,51	[0,18;1,46]	0,2101	
	outros	0,5817	1,79	[0,74;4,34]	0,1984	

Tabela 37 . Resultados da análise univariada pelo modelo de Cox das características sócio-demográficas dos pacientes com câncer de pênis atendidos no Hospital A.C. Camargo. 1953 - 1985. Modelo com A2

variável	categoria	coef.(β i)	HR	IC(HR) _{95%}	p(categ.)	p(variável)
idade (anos)	18 - 39	referência	1,00	-	-	0,6169
	40 - 59	-0,2590	0,77	[0,41;1,47]	0,4301	
	60 - 87	-0,0668	0,94	[0,48;1,82]	0,8442	
raça	branco	referência	1,00	-	-	0,6785
	não branco	-0,1091	0,90	[0,53;1,51]	0,6816	
estado civil	casado	referência	1,00	-	-	0,4968
	não casado	0,7541	2,13	[0,66;6,83]	0,2053	
	ignorado	0,1036	1,11	[0,71;1,74]	0,6505	
zona	rural	referência	1,00	-	-	0,5979
	urbana	-0,1808	0,84	[0,54;1,29]	0,4160	
	ignorado	-0,4339	0,65	[0,20;2,10]	0,4686	
procedên- cia	SE	referência	1,00	-	-	0,1256
	S	-1,1843	0,31	[0,08;1,25]	0,0990	
	NE	0,1250	1,13	[0,54;2,36]	0,7387	
	N+CO	0,6291	1,88	[0,81;4,33]	0,1406	
	ignorado	-0,5005	0,61	[0,19;1,93]	0,3968	
ano de início	1953 - 1959	referência	1,00	-	-	0,0347
	1960 - 1969	0,0156	1,02	[0,61;1,69]	0,9522	
	1970 - 1979	-0,3122	0,73	[0,41;1,32]	0,2975	
	1980 - 1985	-1,1268	0,32	[0,13;0,84]	0,0208	

Tabela 38. Resultados da análise univariada pelo modelo de Cox das variáveis clínicas e sintomas pré-tratamento dos pacientes com câncer de pênis atendidos no Hospital A.C. Camargo. 1953 - 1985. Modelo com A2

variável	categoria	coef.(βi)	HR	IC(HR) _{95%}	p(categ.)	p(variável)
doença venérea	sim	referência	1,00	-	-	0,2951
	não	0,1470	1,16	[0,59;2,28]	0,6705	
	ignorado	0,3938	1,48	[0,87;2,52]	0,1441	
tratamento não oncol. prévio	sim	referência	1,00	-	-	0,6538
	não	-0,3914	0,68	[0,27;1,72]	0,4130	
	ignorado	0,0147	1,02	[0,65;1,58]	0,9482	
tratamento onc. prévio	não	referência	1,00	-	-	0,7887
	sim	0,1980	1,22	[0,30;4,96]	0,7822	
sintoma: dor	não	referência	1,00	-	-	0,1745
	sim	-0,3815	0,68	[0,39;1,21]	0,1926	
sintoma: inflamação	não	referência	1,00	-	-	0,0289
	sim	0,6181	1,86	[1,10;3,13]	0,0202	
sintoma: disúria	não	referência	1,00	-	-	0,5255
	sim	-0,2177	0,80	[0,40;1,61]	0,5372	
sintoma: prurido	não	referência	1,00	-	-	0,0423
	sim	-0,5470	0,58	[0,33;1,01]	0,0549	
tempo de evolução da doença (mês)	1 - 6	referência	1,00	-	-	0,5091
	7 - 12	0,1007	1,11	[0,66;1,84]	0,6989	
	13 - 23	0,3273	1,38	[0,54;3,54]	0,4930	
	24 - 348	-0,2883	0,75	[0,43;1,32]	0,3181	

Tabela 39. Resultados da análise univariada pelo modelo de Cox das variáveis relacionadas ao tumor dos pacientes atendidos no Hospital A.C. Camargo. 1953 - 1985. Modelo com A2

variável	categoria	coef.(β i)	HR	IC(HR) _{95%}	p(categ.)	p(variável)
local: glânde	não	referência	1,00	-	-	0,1511
	sim	1,1738	3,23	[0,45;23,24]	0,2434	
local: prepúcio	não	referência	1,00	-	-	0,1032
	sim	0,6788	1,97	[0,80;4,87]	0,1411	
local: corpo	não	referência	1,00	-	-	0,0001
	sim	0,8937	2,44	[1,55;3,85]	0,0001	
local: escroto	não	referência	1,00	-	-	0,0012
	sim	1,2171	3,38	[1,79;6,38]	0,0002	
local: uretra	não	referência	1,00	-	-	0,0001
	sim	0,9243	2,52	[1,62;3,92]	<0,0001	
local: raiz	não	referência	1,00	-	-	<0,0001
	sim	1,5849	4,88	[3,07;7,75]	<0,0001	
local: outro	não	referência	1,00	-	-	<0,0001
	sim	1,9560	7,07	[4,13;12,12]	<0,0001	
total de gânglios positivos	0	referência	1,00	-	-	0,6470
	1	-0,1064	0,90	[0,33;2,47]	0,8363	
	2 -3	0,4226	1,53	[0,70;3,32]	0,2873	
	4 e +	0,3917	1,48	[0,60;3,67]	0,3984	
estad. T	0+I	referência	1,00	-	-	<0,0001
	II	0,8887	2,43	[0,73;8,08]	0,1467	
	III	1,4138	4,11	[1,26;13,37]	0,0188	
	IV	2,5875	13,30	[3,97;44,57]	<0,0001	
estad. N	0	referência	1,00	-	-	<0,0001
	I	0,5520	1,74	[0,79;3,80]	0,1672	
	II	1,2925	3,64	[2,02;6,57]	<0,0001	
	III	2,3741	10,74	[6,36;18,15]	<0,0001	
grau de diferenc. do tumor	bem	referência	1,00	-	-	0,0117
	mod+pouco	1,0913	2,98	[1,39;6,38]	0,0050	
	SOE	0,4859	1,63	[1,04;2,55]	0,0345	

Tabela 40. Resultados da análise univariada pelo modelo de Cox das variáveis relacionadas aos procedimentos e terapia dos pacientes com câncer de pênis atendidos no Hospital A.C. Camargo, 1953 - 1985. Modelo com A2

variável	categoria	coef.(β_i)	HR	IC(HR) _{95%}	p(categ.)	p(variável)
linfadenectomia	profilática	referência	1,00	-	-	0,4872
	necessária	0,4515	1,57	[0,72;3,44]	0,2587	
	NR	0,4047	1,50	[0,62;3,62]	0,3677	
esvazi. gang. direito	NR	referência	1,00	-	-	0,8855
	inguinal	-0,1508	0,86	[0,41;1,80]	0,6887	
	ing.+ilíaco	-0,0969	0,91	[0,52;1,60]	0,7369	
esvazi. gang. esquerdo	NR	referência	1,00	-	-	0,8256
	inguinal	-0,1827	0,83	[0,38;1,82]	0,6473	
	ing.+ilíaco	-0,1318	0,88	[0,51;1,52]	0,6391	
motivo do esv. gang. direito	NR	referência	1,00	-	-	0,6806
	profilático	-0,3317	0,72	[0,33;1,57]	0,4061	
	necessário	-0,0046	1,00	[0,57;1,73]	0,9868	
motivo do esv. gang. esquerdo	NR	referência	1,00	-	-	0,7873
	profilático	-0,2478	0,78	[0,36;1,71]	0,5351	
	necessário	-0,1005	0,90	[0,52;1,57]	0,7208	
tempo esv. gang.	NR	referência	1,00	-	-	0,3834
	D/E <7 m.	0,1113	1,12	[0,65;1,91]	0,6844	
	outros	-0,5390	0,58	[0,23;1,45]	0,2465	

Tabela 42. Resultados da análise univariada pelo modelo de Cox das características sócio-demográficas dos pacientes com câncer de pênis atendidos no Hospital A.C. Camargo. 1953 - 1985. Modelo com A3

variável	categoria	coef.(β i)	HR	IC(HR) _{95%}	p(categ.)	p(variável)
idade (anos)	18 - 59	referência	1,00	-	-	0,2430
	60 - 87	0,5165	1,68	[0,71;3,96]	0,2386	
raça	branco	referência	1,00	-	-	0,6785
	não branco	-0,1091	0,90	[0,53;1,51]	0,6816	
estado civil	casado	referência	1,00	-	-	0,4334
	não casado	0,9009	2,46	[0,55;11,04]	0,2393	
	ignorado	-0,2497	0,78	[0,31;1,98]	0,5999	
zona	rural	referência	1,00	-	-	0,4654
	urbana	-0,3806	0,68	[0,29;1,61]	0,3840	
	ignorado	-15,2494	0,00	nd*	0,9932	
procedên- cia	SE	referência	1,00	-	-	0,2478
	S	-1,0773	0,34	[0,05;2,59]	0,2975	
	NE	1,0020	2,72	[0,90;8,25]	0,0763	
	N+CO	-0,0129	0,99	[0,13;7,48]	0,9900	
	ignorado	-15,2346	0,00	nd*	0,9939	
ano de início	1953 - 1959	referência	1,00	-	-	0,1383
	1960 - 1969	0,0954	1,10	[0,38;3,17]	0,8600	
	1970 - 1979	-0,8239	0,44	[0,12;1,64]	0,2197	
	1980 - 1985	-1,5082	0,22	[0,03;1,90]	0,1690	

*nd:não definido

Tabela 43. Resultados da análise univariada pelo modelo de Cox das variáveis clínicas e sintomas pré-tratamento dos pacientes com câncer de pênis atendidos no Hospital A.C. Camargo. 1953 - 1985. Modelo com A3

variável	categoria	coef.(β i)	HR	IC(HR) _{95%}	p(categ.)	p(variável)
doença venérea	sim	referência	1,00	-	-	0,2951
	não	0,1470	1,16	[0,59;2,28]	0,6705	
	ignorado	0,3938	1,48	[0,87;2,52]	0,1441	
tratamento não oncol. prévio	sim	referência	1,00	-	-	0,7238
	não	-0,6097	0,54	[0,07;4,30]	0,5634	
	ignorado	-0,3033	0,74	[0,31;1,79]	0,5019	
tratamento onc. prévio	não	referência	1,00	-	-	0,5089
	sim	0,7553	2,13	[0,29;15,88]	0,7365	
sintoma: dor	não	referência	1,00	-	-	0,4115
	sim	-0,4371	0,65	[0,22;1,92]	0,4316	
sintoma: inflamação	não	referência	1,00	-	-	0,7983
	sim	0,1443	1,16	[0,39;3,44]	0,7953	
sintoma: disúria	não	referência	1,00	-	-	0,3512
	sim	0,5007	1,65	[0,60;4,51]	0,3291	
sintoma: prurido	não	referência	1,00	-	-	0,3188
	sim	-0,5265	0,59	[0,20;1,76]	0,3437	
temp.evol. doença(m)	1 - 12	referência	1,00	-	-	0,6563
	13 - 348	-0,2240	0,80	[0,29;2,18]	0,6623	

Tabela 44. Resultados da análise univariada pelo modelo de Cox das variáveis relacionadas ao tumor dos pacientes atendidos no Hospital A.C. Camargo. 1953 - 1985. Modelo com A3

variável	categoria	coef.(β i)	HR	IC(HR) _{95%}	p(categ.)	p(variável)
local: prepúcio	não	referência	1,00	-	-	0,0485
	sim	15,1521	0,00	nd*	0,9854	
local: corpo	não	referência	1,00	-	-	0,0545
	sim	0,8497	2,34	[0,97;5,65]	0,0592	
local: escroto	não	referência	1,00	-	-	0,0826
	sim	1,6175	5,04	[1,10;23,17]	0,0377	
local: uretra	não	referência	1,00	-	-	0,1988
	sim	0,6197	1,86	[0,75;4,61]	0,1811	
local: raiz	não	referência	1,00	-	-	0,0361
	sim	1,2184	3,38	[1,22;9,39]	0,0194	
local: outro	não	referência	1,00	-	-	0,1590
	sim	1,2312	3,43	[0,79;14,80]	0,0991	
total de gânglios positivos	0	referência	1,00	-	-	0,2909
	1	0,8025	2,23	[0,74;6,69]	0,1519	
	2 -3	1,1104	3,04	[0,40;23,23]	0,2849	
	4 e +	-15,0254	0,00	nd*	0,9918	
estad. T	0+I+II	referência	1,00	-	-	0,0727
	III	0,7572	2,13	[0,84;5,41]	0,1108	
	IV	1,5422	4,68	[1,22;17,94]	0,0246	
estad. N	0	referência	1,00	-	-	0,0211
	I	-0,4085	0,67	[0,09;5,16]	0,6959	
	II	1,0124	2,75	[0,96;7,93]	0,0607	
	III	1,8551	6,39	[1,98;20,61]	0,0019	
grau de diferenc. do tumor	bem	referência	1,00	-	-	0,7702
	mod+pouco	0,8311	2,30	[0,29;18,03]	0,4293	
	SOE	0,1149	1,12	[0,47;2,70]	0,7975	

nd*: não definido

Tabela 45. Resultados da análise univariada pelo modelo de Cox das variáveis relacionadas aos procedimentos e terapia dos pacientes com câncer de pênis atendidos no Hospital A.C. Camargo, 1953 - 1985. Modelo com A3

variável	categoria	coef.(β_i)	HR	IC(HR) _{95%}	p(categ.)	p(variável)
linfadenectomia	profilática	referência	1,00	-	-	0,3234
	necessária	-0,8611	0,42	[0,13;1,33]	0,1405	
	NR	-0,3864	0,68	[0,19;2,42]	0,5504	
esvazi. gang. direito	NR	referência	1,00	-	-	0,1829
	inguinal	0,4391	1,55	[0,43;5,57]	0,5009	
	ing.+ilíaco	0,9264	2,53	[0,98;6,52]	0,0556	
esvazi. gang. esquerdo	NR	referência	1,00	-	-	0,2860
	inguinal	0,5107	1,67	[0,46;5,98]	0,4336	
	ing.+ilíaco	0,7660	2,15	[0,83;5,55]	0,1135	
motivo do esv.gang. direito	NR	referência	1,00	-	-	0,2222
	profilático	0,9057	2,47	[0,79;7,77]	0,1211	
	necessário	0,6640	1,94	[0,72;5,27]	0,1918	
motivo do esv.gang. esquerdo	NR	referência	1,00	-	-	0,1773
	profilático	1,1562	3,18	[1,01;9,99]	0,0479	
	necessário	0,4564	1,58	[0,58;4,28]	0,3695	
tempo esv. gang.	NR	referência	1,00	-	-	0,1787
	D/E <7 m.	0,8198	2,27	[0,88;5,86]	0,0902	
	outros	0,8480	2,34	[0,65;8,42]	0,1951	