
CONTROLO DE QUALIDADE EM REGISTOS DE CANCRO: UMA NOVA ABORDAGEM PARA O ROR-SUL

QUALITY CONTROL IN CANCER REGISTRIES: A NEW APPROACH FOR ROR-SUL

Autores: Sónia Dória Nóbrega

- Consultora Estatística do Registo Oncológico Regional Sul Instituto Português de Oncologia de Francisco Gentil

Carlos Daniel Paulino

- Professor Associado do Departamento de Matemática do Instituto Superior Técnico da Universidade Técnica de Lisboa

RESUMO:

- No presente estudo, criou-se um modelo de predição das observações com maior probabilidade de erro de exactidão dos dados e causadores de maior impacto na qualidade dos indicadores epidemiológicos produzidos pelo ROR-Sul. Utilizaram-se modelos de regressão logística e, para avaliar a sua reprodutibilidade, calculou-se a sensibilidade e especificidade. O modelo encontrado dá-nos a indicação de existirem variáveis capazes de prever a existência de erros de registo; no entanto, a avaliação da reprodutibilidade do modelo leva-nos a concluir que caso não se encontrem meios que permitam alcançar resultados com melhores sensibilidades, este método de predição de erros de registo não é eficiente nem rendível.

PALAVRAS-CHAVE:

- *Registos de Cancro; Controlo de Qualidade; Exactidão; Validade; Fiabilidade.*

ABSTRACT:

- In this study we develop a prediction model to detect observations with high probability of data accuracy errors with impact on quality of epidemiological measures produced by ROR-Sul. Logistic regression was used to obtain the model and its reproducibility was evaluated by means of sensibility and specificity. The results identified predicting factors of registry errors; however, the model's reproducibility indicate that, if there is no way to reach better sensitivities, this method of predicting cancer errors is neither efficient nor profitable.

KEY-WORDS:

- *Cancer Registry; Quality Control; Accuracy; Data Validity; Data Reliability.*

1. INTRODUÇÃO

O registo de cancro de base populacional é um instrumento essencial de qualquer programa de controlo de cancro a nível europeu (Muir *et al.*, 1985). À sua actividade está associada a gestão de todos os registos de casos de cancro diagnosticados numa população definida, onde estão documentadas as características individuais dos doentes assim como os dados clínicos e anátomo-patológicos dos tumores, recolhidos de forma contínua e sistemática em diferentes fontes de informação. O registo analisa e interpreta estes dados, proporcionando anualmente informação sobre a incidência de todos os tumores malignos, através da sua distribuição geográfica, por sexo e grupo etário. Esta informação não serve apenas como ponto de partida para a investigação epidemiológica sobre os agentes que determinam o cancro, mas também para o planeamento dos cuidados a prestar, tanto ao nível da prevenção como do diagnóstico e tratamento da doença oncológica.

O registo de cancro é, acima de tudo, uma fonte de informação, e como tal a busca da excelência deve ser uma prioridade. Como tem sido alegado, uma informação que não merece confiança é pior do que a falta de informação (Skeet, 1991).

Os princípios básicos do controlo de qualidade em registos de cancro surgiram e evoluíram independentemente do seu desenvolvimento em outras disciplinas, nomeadamente na indústria. Para tal contribuíram programas de controlo de qualidade de prestação de cuidados de saúde a nível hospitalar e de outros sistemas de informação médica (Zippin *et al.*, 1985).

Existem cinco características reconhecidas como indicadores de qualidade de registos de cancro (Hilsenbeck, 1994, Goldberg *et al.*, 1980):

- a **exaustividade do registo** (*case completeness*) – definida como a proporção de todos os casos de cancro incidentes numa população definida que constam da base de dados do registo, relativamente a todos os casos de cancro incidentes nessa população (Parkin *et al.*, 1992);
- a **exactidão** (*accuracy*) – definida como a proporção de casos registados com uma dada característica (sexo, idade, diagnóstico) que realmente têm esse atributo. Na prática, a exactidão traduz-se na percentagem de concordância entre a informação que consta no registo de cancro e a informação de uma fonte independente (Hilsenbeck *et al.*, 1985);
- a **disponibilização atempada da informação** (*timeliness*) - a sua definição está associada ao cumprimento de um determinado nível de calendarização das fases de registo, que permita a tomada de acções decorrentes da informação produzida (Hilsenbeck, 1994);
- a **estabilidade de codificação** (*coding constancy*) - a sua definição está associada a práticas de codificação uniformizadas entre e intra instituições e, tanto quanto possível, constantes ao longo dos anos;

- e a **estabilidade de publicação** (*publication constancy*) - a sua definição está associada à padronização das publicações periódicas, quer ao nível do formato de apresentação quer ao do uso de metodologias estatísticas.

Os dois primeiros indicadores são os mais conhecidos e usados, concentrando-se neles a maior parte do desenvolvimento de metodologias para a sua avaliação (Goldberg *et al.*, 1980).

De acordo com os resultados demonstrados num estudo desenvolvido pela *North American Association of Central Cancer Registries* (Roffers, 1996), o autor refere que, na utilização dos dados para o controlo do cancro, a exaustividade do registo é menos importante do que a exactidão dos dados registados.

A exactidão dos dados é, assim, uma componente essencial no controlo de qualidade de qualquer registo de cancro. Depende da fiabilidade da informação contida nos documentos que dão origem ao registo (e.g. processo clínico, relatórios anátomo-patológicos), do nível de destreza na recolha dos dados, da codificação e da introdução dessa informação na base de dados.

Com base nesta perspectiva, a evolução dos procedimentos de trabalho do Registo Oncológico Regional Sul (ROR-Sul) nos últimos anos resultou na passagem de um registo feito em suporte de papel para um sistema de registo via *Internet*.

O ROR-Sul é responsável pelo registo de cancro da região Sul de Portugal. A sua área de actividade inclui o distrito de Santarém, Lisboa, Setúbal, Portalegre, Évora, Beja e Faro, cobrindo cerca de metade da área e população de Portugal (aproximadamente 4.200.000 habitantes). Funciona como registo de cancro de base-populacional desde 1989.

É constituído por uma estrutura central, sediada nas instalações do Centro Regional de Lisboa do IPOFG, e por núcleos locais sediados nos hospitais centrais e distritais, nas Sub-Regiões de Saúde e, ainda, em infraestruturas de saúde privadas, nomeadamente nos laboratórios de Anatomia Patológica e instituições hospitalares privadas.

Recebe, por ano, dados de cerca de 12.000 novos casos de cancro e cerca de 7000 óbitos por tumor maligno (Miranda, 1998).

As vantagens do sistema de transmissão de dados que está implementado são:

- permitir que a maioria das validações de edição dos dados sejam feitas no momento da introdução dos dados;
- reduzir substancialmente o tempo que decorre entre a identificação dos casos, em cada instituição, e a aceitação e integração numa base de dados central de toda a informação respeitante aos novos casos de cancro.

Este sistema depara-se agora com uma dificuldade; embora exista um programa de formação contínua que visa a manutenção da padronização dos conceitos e procedimentos de registo, dada a diversidade de pessoas que contribuem para o

sistema, o ROR-Sul necessita de encontrar mecanismos capazes de detectar falhas em todo o sistema de informação, por forma a garantir uma maior qualidade dos dados.

Uma das deficiências apontadas aos estudos de avaliação da qualidade realizados até à data pelo ROR-Sul, é o facto de estes terem um carácter pontual, ou seja, realizam-se anualmente apenas em alguns dos núcleos locais. Por outro lado, com o arranque da rede de transmissão de dados só agora se torna viável ao ROR-Sul centralizar toda a informação dos diferentes núcleos locais com o espaço temporal suficiente para poder trabalhar a base de dados da incidência, previamente à publicação dos seus indicadores.

Os estudos de avaliação da qualidade dos dados apontam para a existência de uma considerável proporção de erros na exactidão dos dados, produzidos de uma forma sistemática, ou seja, associados a determinadas características dos registos (por exemplo, lateralidade do tumor associada à topografia do tumor, datas de incidência e instituição que regista o caso, etc.).

Por esta razão, pretende-se com o presente trabalho testar uma metodologia que vise criar um modelo de predição das observações com maior probabilidade de erro de exactidão dos dados e causadoras de maior impacte na qualidade dos indicadores epidemiológicos produzidos pelo ROR-Sul (e.g. taxas de incidência). Pretende-se ainda avaliar a reprodutibilidade dos resultados obtidos pelo modelo.

Por razões de limitação de tempo, decidiu-se restringir a população-alvo a duas fontes de informação do ROR-Sul, designadas, por razões de confidencialidade, por *Instituição A* e *Instituição B*.

2. MATERIAL E MÉTODOS

2.1. TIPO DE ESTUDO

Trata-se de um estudo transversal dirigido a todos os casos de tumores malignos incidentes (novos casos) num período de um ano e registados pela *Instituição A* e *Instituição B*.

2.2. POPULAÇÃO-ALVO

A população estudada compreende todos os novos casos de tumores malignos diagnosticados em 1996 que acorreram à *Instituição A* (634 casos), e diagnosticados em 1997 que acorreram à *Instituição B* (3509 casos).

2.3. AMOSTRA

Foram recolhidas duas amostras aleatórias simples, por instituição, uma para se proceder à aplicação do modelo de predição e outra para avaliação da reprodutibilidade dos resultados obtidos.

Não foi possível determinar a dimensão da amostra de acordo com os objectivos do estudo, uma vez que se desconhece a existência de qualquer tipo de estudo idêntico ao presente que fornecesse alguma informação *a priori*. Como tal, optou-se por recolher uma amostra que conseguisse garantir a ocorrência de uma determinada prevalência de erros, conforme demonstravam estudos anteriores de avaliação da qualidade dos dados do ROR-Sul.

Considerando uma situação de amostragem sem reposição, o cálculo da dimensão das amostras de cada instituição fez-se com base na fórmula derivada do intervalo de confiança para uma proporção numa população finita

$$n \geq \frac{p(1-p) \frac{N}{N-1}}{\frac{\Delta^2}{z_{1-\frac{\alpha}{2}}^2} + \frac{p(1-p)}{N-1}}$$

onde se assumiram os seguintes valores:

- uma estimativa da proporção p de concordância dos dados (i.e. de não haver neles erros de registo considerados graves) de 80%;
- uma margem de precisão de $\Delta=5\%$;
- um nível de confiança $1-\alpha=95\%$, pelo que $z_{1-\frac{\alpha}{2}}=1,96$;
- N é a dimensão da população da instituição em estudo.

Obteve-se, assim, uma amostra de 177 casos para a *Instituição A* e 230 casos para a *Instituição B*.

Para o estudo da reprodutibilidade do modelo, estipulou-se, por questões práticas, que a dimensão da amostra fosse 50% da amostra inicial, de cada instituição.

Assim, o total de casos seleccionados, para a *Instituição A*, foi de 266 (177+89) e de 345 para a *Instituição B* (230+115).

2.4. COLHEITA DA INFORMAÇÃO

Depois de seleccionadas as amostras de cada instituição, procedeu-se à revisão dos dados dos casos identificados, consultando, para tal, os processos clínicos em cada uma das instituições.

A informação recolhida é aquela que consta do instrumento de “Entrada para Registo” do ROR-Sul, contendo variáveis referentes ao doente tais como nome, sexo, data de nascimento, naturalidade, residência, profissão e situação do doente, ou variáveis clínicas referentes ao diagnóstico e tratamento do tumor, nomeadamente, meios de diagnóstico utilizados, localização do tumor, tipo histológico, datas de início dos sintomas/diagnóstico/tratamento, estágio da doença, tipo de tratamento, etc.

2.5. MODELO DE PREDIÇÃO DOS ERROS DE REGISTO

Antes de iniciar uma breve abordagem ao modelo de regressão logística, é importante realçar que o objectivo da análise, por este método, é igual ao objectivo postulado em qualquer técnica de construção de um modelo de regressão: encontrar o modelo que melhor se ajusta e mais parcimonioso, e que seja biologicamente plausível na descrição da variabilidade de uma resposta (variável dependente) em função de um conjunto de variáveis independentes.

O que distingue o modelo de regressão logística de um modelo de regressão linear é o facto de a variável resposta, no primeiro, ser dicotómica (ou, pelo menos, categorizada). Esta diferença reflecte-se nos pressupostos e na escolha do modelo paramétrico subjacente aos dados.

Relembremos que o presente estudo tem como objectivo criar um modelo de predição das observações com maior probabilidade de erro nos dados e causadoras de maior impacto na qualidade dos indicadores epidemiológicos produzidos pelo ROR-Sul (e.g. taxas de incidência);

Para cada observação o nível de desacordo é definido em função da repercussão dos erros nos indicadores epidemiológicos. Assim, um erro que altere os dados referentes ao cálculo das taxas de incidência (alteração do número de novos casos) é considerado um desacordo máximo (e.g. alteração do sexo ou do grupo etário); ao invés, uma alteração da informação registada que não produza mudanças nos dados para o cálculo das taxas de incidência (e.g. alteração da idade mas manutenção do mesmo grupo etário) é considerada uma discordância mínima. Com base nestes níveis de desacordo, definem-se os seguintes níveis de erro para cada observação da amostra:

- Nível 1 – quando ocorreu um erro com repercussões no cálculo da incidência, ou seja, desacordo máximo em pelo menos uma das variáveis sexo, idade, distrito de residência, data do diagnóstico, topografia, comportamento;
- Nível 0 – quando não ocorreu nenhum erro de registo ou o erro é mínimo, não tendo qualquer influência nas variáveis que interferem com o cálculo da incidência.

Uma vez revistos todos os casos, cada variável em estudo foi classificada segundo o nível de discordância encontrado, de acordo com os critérios propostos por Hilsenbeck (Hilsenbeck, 1994 e Hilsenbeck *et al.*, 1985). Posteriormente, cada caso foi classificado de acordo com o nível de erro 0 ou 1, definido anteriormente. Esta será a variável resposta do modelo de predição de erros de registo.

As variáveis independentes, candidatas a este modelo, são as que têm os valores que constam do registo original. Nalguns casos decidiu-se definir alguns agrupamentos por forma a favorecer a discriminação de uma observação com maior probabilidade de erro nos dados.

Assim, foram seleccionadas as seguintes variáveis:

- Instituição
 - A
 - B
- Idade (anos) ou, alternativamente, a variável categorizada
 - 0-14 anos
 - ≥ 15 anos
- Sexo
 - M
 - F
- Região de Naturalidade
 - Portugal Continental
 - Ilhas
 - Estrangeiro
- Região de Residência
 - Lisboa
 - Santarém/Setúbal/Alentejo/Algarve
 - Outros Distritos do Continente
 - Ilhas
 - Estrangeiro
- Situação do Caso
 - Vivo
 - Morto
- Tipo de Diagnóstico
 - Histol. Primário
 - Histol. Metástases
 - Autópsia c/ Histologia
 - Citologia/Hematologia
 - Exames Macroscópicos
- Intervalo entre a 1ª Observação e a 1ª Consulta (dias) ou, alternativamente, a variável categorizada
 - < 0 dias
 - $[0, 180]$ dias
 - > 180 dias
- Intervalo entre o Diagnóstico e a 1ª Consulta (dias) ou, alternativamente, a variável categorizada
 - < -365 dias
 - $[-365, 90]$ dias
 - > 90 dias
- Intervalo entre os 1^{os} Sintomas e a 1ª Consulta (dias) ou, alternativamente, a variável categorizada
 - < -365 dias
 - $[-365, 365]$ dias
 - > 365 dias
- Intervalo entre os 1^{os} Sintomas e a 1ª Observação (dias) ou, alternativamente, a variável categorizada
 - ≤ 365 dias
 - > 365 dias
- Intervalo entre os 1^{os} Sintomas e o Diagnóstico (dias) ou, alternativamente, a variável categorizada
 - ≤ 365 dias
 - > 365 dias
- Intervalo entre a 1ª Observação e o Diagnóstico (dias) ou, alternativamente, a variável categorizada
 - ≤ 60 dias
 - > 60 dias

- Topografia
 - Lábio,Cavidade Oral e Faringe (C00-C14)
 - Órgãos Digestivos (C15-C26)
 - Sist. Respirat. Órg. Intratoráx. (C30-39)
 - Ossos,Articul. Cartil. Articul. (C40-C41)
 - Sist. Hematop. Retículo Endotelial (C42)
 - Pele (C44)
 - Nervos Perif. e Sist. Nerv. Autônomo (C47)
 - Retroperitoneu e Peritoneu (C48)
 - Tec.Conjunt., Subcut. e Out.Tec.Moles (C49)
 - Mama (C50)
 - Órgãos Genitais Femininos (C51-C58)
 - Órgãos Genitais Masculinos (C60-C63)
 - Tracto Urinário (C64-C68)
 - Olho, Céreb. e Out.P.Sist.Nerv.Central (C69-C72)
 - Tiróide e Out. Glândulas Endócrinas (C73-C75)
 - Outras Localiz. Mal Definidas (C76)
 - Gânglios Linfáticos (C77)
 - Localiz. Primária Desconhecida (C80)
- Morfologia
 - Neoplas. Pouco Específicas (9990, 8000-8041)
 - Neoplasias Específicas (8042-9989)
- Comportamento do Tumor
 - Benigno, Incerto se Malig/Benig., Carc. In Situ
 - Maligno
- Lateralidade do Tumor
 - Direita, Esquerda, Bilateral
 - Não Aplicável, Desconhecida
- Estádio na Apresentação
 - Tumor Localizado
 - Com Metástases
 - Não Aplicável
 - Desconhecido
- T
 - Conhecido
 - Omisso
- N
 - Sem invasão ganglionar (N=0)
 - Com invasão ganglionar (N=1..3)
 - Omisso
- M
 - Sem Metástases à Distância (M=0)
 - Com Metástases à Distância (M=1)
 - Omisso
- Grau de Diferenciação
 - Conhecido (Bem/Moderado/Pouco/Indif.)
 - Desconhecido

A escolha dos níveis das variáveis categorizadas fez-se com base em diversos critérios: ou pela representatividade de cada um dos níveis na amostra (e.g. região de naturalidade e residência, tipo de diagnóstico, topografia), ou pelo conhecimento prévio de categorias mais problemáticas (e.g. grupo etário, morfologia, grau de diferenciação, comportamento do tumor) ou ainda pela criação de grupos que eventualmente descrevam situações menos vulgares, que possam corresponder a erros nos dados (e.g. intervalos entre datas). Intervalos negativos entre a data da 1ª observação pela doença e a data da 1ª consulta na instituição são certamente indicadores de erro, uma vez que a 1ª observação da doença não pode ocorrer após a 1ª consulta numa instituição por essa doença. Com a data dos sintomas não existem

fronteiras tão claras, dado que a doença pode ser observada e diagnosticada sem que o doente apresente sintomas; no entanto, sintomas que se manifestem há mais de um ano, sem que ocorra uma consulta motivada pelas queixas, é uma situação menos vulgar.

O modelo de regressão logística será ajustado à 1ª amostra, definida para cada instituição, num total de 407 casos.

2.6. CONSTRUÇÃO DO MODELO MÚLTIPLO

Em função do tipo de estudo, existem diferentes objectivos para o ajustamento do modelo de regressão logística:

- no *modelo para predições*, o objectivo é construir um modelo que, de entre uma colecção de variáveis, selecione apenas aquelas que melhor predigam ou se ajustem à resposta;
- no *modelo para factores de risco*, o objectivo é estimar a associação entre a exposição a determinados factores de risco e a resposta.

No primeiro caso, pretende-se que o modelo seja o mais parcimonioso possível e com o melhor ajustamento aos dados, enquanto que, no segundo caso, poderão obter-se modelos com mais variáveis uma vez que se avaliam efeitos de confundimento e interacções com o factor de risco em estudo. O critério de selecção das variáveis de modelos para predições prende-se mais com critérios estatísticos do que com questões de plausibilidade biológica das variáveis (Hosmer *et al.*, 1989 e Kleinbaum *et al.*, 1982).

Assim, optou-se pelo uso do método *stepwise* com eliminação regressiva, que se baseia no ajustamento do modelo saturado, com as variáveis definidas como candidatas ao modelo, e na fixação de uma regra de decisão com um valor-p de saída, $p_s=0.050$, isto é, com base no Teste de Razão de Verosimilhanças de Wilks saem do modelo saturado todas as variáveis cujo valor-p, dado por este teste, seja superior a p_s . Com base num critério meramente estatístico, proposto por Hosmer e Lemeshow (Hosmer *et al.*, 1989), considerou-se uma variável candidata ao modelo múltiplo se o valor-p do teste de razão de verosimilhanças for inferior a 0.250.

O valor preditivo de cada variável foi caracterizado através da razão das chances (*odds ratio* - OR) e respectivo intervalo de 95% de confiança.

2.7. DIAGNÓSTICO DO MODELO FINAL

Como em qualquer modelo ajustado, antes de se prosseguir com a inferência sobre este, dever-se-á verificar o seu nível de ajustamento. Existe um conjunto de medidas de avaliação da qualidade do ajustamento que descrevem o comportamento

do modelo estimado relativamente aos dados observados, de uma forma global ou particular.

A avaliação da qualidade de ajustamento iniciou-se pela aplicação das metodologias de avaliação global - o Teste do Qui-quadrado de Pearson e o Teste de Hosmer e Lemeshow (Hosmer *et al.*, 1989) - e prosseguiu com uma avaliação cuidadosa das medidas de diagnóstico do modelo que permitem identificar observações com comportamento problemático, sendo estas medidas de três tipos (Hosmer *et al.*, 1989; Hosmer *et al.*, 1980; Lemeshow *et al.*, 1982; Hosmer *et al.*, 1988 e Hosmer *et al.*, 1997):

- medidas de resíduos
- medidas de “leverages”
- medidas de influência

As medidas de resíduos servem para identificar observações em que o valor esperado obtido pelo modelo não está próximo do valor observado da variável resposta.

Quando um indivíduo apresenta um padrão de valores para as covariáveis que é muito pouco usual, quando comparado com o dos restantes elementos da amostra, as medidas de “leverages” são as indicadas para identificar situações deste tipo.

Quando se pretende avaliar o impacte no modelo da presença ou ausência de determinado elemento na amostra devemos observar as alterações produzidas nos coeficientes estimados do modelo. A contribuição de uma observação para a alteração dos coeficientes depende tanto do valor do resíduo como do valor do “leverage”. As medidas de diagnóstico destas situações são chamadas medidas de influência.

Para avaliar simultaneamente a relação da sensibilidade (proporção de casos com erros de nível 1 correctamente identificados pelo modelo) e 1-especificidade (sendo a especificidade a proporção de casos com erros de nível 0 correctamente identificados pelo modelo) dos resultados obtidos pelo modelo, faz-se variar o ponto de corte de 0 a 1 e representa-se graficamente a curva ROC (*Receiver Operating Characteristic* – Metz, 1978). Este método tem tido uma utilização cada vez mais frequente por fornecer uma medida de discriminação do modelo ajustado (Hanley *et al.*, 1982; Hanley *et al.*, 1983; Beck *et al.*, 1986; Schultz *et al.*, 1995; Brickley *et al.*, 1995; Hagen *et al.*, 1995; Fortescue *et al.*, 2000; Dreiseitl *et al.*, 2000 e Beggl *et al.*, 2000).

Hosmer (1995) considera a seguinte regra de avaliação do poder de discriminação do modelo de regressão logística baseada na área sob a curva ROC:

Tabela 1: Classificação do nível de discriminação do modelo de regressão logística

Área sob a Curva ROC	Interpretação
0,50	Não discrimina
≥ 0,70	Discriminação aceitável
≥ 0,80	Discriminação excelente
≥ 0,90	Discriminação marcante (pouco usual)

2.8. AVALIAÇÃO DA REPRODUTIBILIDADE DO MODELO

Uma vez encontrado o modelo final, com uma boa qualidade de ajustamento aos dados, pretendemos avaliar qual o verdadeiro valor discriminatório do modelo encontrado, isto é, até que ponto as predições obtidas pelo modelo correspondem à realidade numa amostra independente daquela que deu origem ao modelo.

Assim, através da segunda amostra retirada em cada instituição (204 casos) pretende-se avaliar a reprodutibilidade dos resultados através do modelo obtido na primeira amostra, avaliando a estimativa da sensibilidade, especificidade e valor preditivo positivo (proporção de casos com erros de nível 1 entre os casos assinalados pelo modelo como errados).

3. RESULTADOS

3.1. REPRESENTATIVIDADE DA AMOSTRA

Dos 611 casos seleccionados para as duas amostras a recolher nas instituições, apenas foram revistos os processos clínicos de 252 casos da *Instituição A* (94,7%) e 338 casos da *Instituição B* (98,0%). Em 11 casos os processos clínicos não foram localizados pelo arquivo clínico, em 9 o processo clínico estava muito incompleto e 1 caso esteve com o processo clínico requisitado pelo serviço responsável pelo doente, durante todo o período de revisão dos dados deste estudo.

A amostra ficou assim reduzida a 590 casos das duas instituições. Dado que o cálculo da dimensão das amostras previa 611 casos, 407 para a amostra que ajustaria o modelo de predição dos erros e 204 para a da avaliação da reprodutibilidade do modelo, decidiu-se que os 21 casos em que não foi possível rever a informação registada seriam retirados da 2ª amostra (dada a menor importância desta amostra nos objectivos do trabalho). Assim, a 2ª amostra ficou reduzida a 183 casos.

Não se rejeitou a hipótese de homogeneidade das amostras com o total de casos registados por cada uma das instituições, no que se refere à distribuição dos sexos e grupos etários. De igual modo não se encontraram diferenças estatisticamente significativas nestas variáveis relativamente aos casos em que foi possível rever os processos clínicos e os 21 casos retirados da amostra por não haver possibilidade de revisão dos dados.

Na globalidade dos dados apenas algumas variáveis interferem directamente na determinação de alguns indicadores epidemiológicos; para o cálculo das taxas de incidência dos tumores malignos estão envolvidas as variáveis sexo, idade, distrito de residência, data do diagnóstico, comportamento e localização primária do tumor (órgão). Basta que ocorra um desacordo máximo numa destas variáveis para que a observação tenha um erro com impacte na incidência .

A distribuição destes erros (nível 0 e 1), na amostra, fez-se da seguinte forma:

Tabela 2: Distribuição dos erros de registo por instituição

Erros de Registo	Total	Instituição	
		A	B
<i>Nível 0</i>	547 casos	240 casos	307 casos
	92,7%	95,2%	90,8%
<i>Nível 1</i>	43 casos	12 casos	31 casos
	7,3%	4,8%	9,2%

Como podemos verificar o nível de exactidão global foi de 92,7%, ocorrendo uma maior frequência de erros de registo com impacte na incidência na *Instituição B* do que na *A*, sendo esta diferença estatisticamente significativa ($p=0,042$).

Tabela 3: Distribuição dos erros de registo por amostra - a de ajustamento do modelo (1ª) e a de avaliação da reprodutibilidade (2ª)

Erros de Registo	Total	Amostra	
		1ª	2ª
<i>Nível 0</i>	547 casos	376 casos	171 casos
	92,7%	92,4%	93,4%
<i>Nível 1</i>	43 casos	31 casos	12 casos
	7,3%	7,6%	6,6%

Verifica-se que não existe uma diferença estatisticamente significativa ($p=0,647$) na distribuição do nível dos erros segundo a amostra.

3.2. MODELO DE PREDIÇÃO DOS ERROS DE REGISTO

3.2.1. MODELOS UNIVARIADOS

Encontraram-se algumas variáveis que individualmente estão significativamente associadas à ocorrência de erros de registo de nível 1. A variável Instituição é um dos factores mais associados onde se verifica que os casos registados pela *Instituição B* têm uma chance de ocorrência de um erro de nível 1 3,5 (valor da OR) vezes superior à dos casos da *Instituição A*. Os casos registados com idades inferiores a 15 anos são também aqueles que apresentam maior probabilidade de erro de registo (OR=6,61) relativamente aos casos com idades superiores ou iguais a 15 anos. Encontrou-se

também uma associação entre o sexo e a ocorrência de erros de nível 1 - os casos do sexo masculino têm uma chance de erro cerca de 50% menor que os casos do sexo feminino; no entanto, esta associação não revela uma significância estatística muito forte ($p=0,113$).

De assinalar, também, uma associação entre a ocorrência dos erros de registo e a região de naturalidade dos casos. São os casos oriundos das ilhas ou do estrangeiro aqueles que maiores probabilidades revelam de dar origem a um erro de registo, quando comparados com os casos naturais de Portugal Continental.

Para o estudo da região de residência, agruparam-se os distritos em: Lisboa, restantes distritos do ROR/Sul (Santarém, Setúbal, Évora, Beja, Portalegre e Faro), restantes distritos de Portugal Continental não pertencentes ao ROR/Sul, Ilhas e Estrangeiro. A razão pela qual se separou Lisboa dos restantes distritos do ROR/Sul deveu-se ao facto de ser conhecido que muitos casos são registados erradamente como residentes no distrito de Lisboa (onde se concentra a maioria dos hospitais com oferta de tratamento oncológico) e por isso darem origem a maiores erros de registo com impacte no cálculo das taxas de incidência por distrito. Quando avaliamos essa associação com a região de residência agrupada deste modo, verifica-se que as diferenças existentes não são estatisticamente significativas. Se englobarmos as 3 primeiras categorias numa só, denominada Portugal Continental, observa-se que a associação da região de residência com a ocorrência de erros ainda não tem significância estatística.

Quanto ao tipo de diagnóstico, são as categorias de autópsia com histologia e histologia de metástases as que apresentam maior probabilidade de ocorrência de erro de registo, quando comparadas com os casos com histologia do tumor primário. Esta variável associa-se significativamente com a ocorrência de erros de nível 1.

Existem localizações onde a incidência de erros de registo é mais elevada (e.g. C00-C14, C73-C75, C76, C77 e C80), no entanto, pela fraca representatividade destas categorias na amostra, não existe uma associação estatisticamente significativa. O mesmo acontece com a morfologia do tumor, agrupada em neoplasias pouco especificadas (e.g. 9990+8000-8041) e neoplasias especificadas (restantes códigos).

Quanto ao comportamento do tumor, verifica-se que é o grupo de tumores de comportamento benigno, incerto se benigno ou maligno e os carcinomas *in situ* que apresenta um risco 4 vezes superior de ocorrência de erros de registo, quando comparados com os tumores de comportamento maligno. Efectivamente esta variável está directamente implicada na selecção dos tumores malignos para o cálculo das incidências e é preenchida de acordo com a opção seleccionada para a histologia do tumor, aquando da introdução dos dados em computador. Muitas vezes, por exemplo, para um tumor basocelular podem surgir diversos códigos, um para maligno e outro para incerto quanto à malignidade.

Outra variável que mostrou uma associação estatisticamente significativa foi o grau de diferenciação do tumor. Os casos cujo grau de diferenciação do tumor é desconhecido têm uma chance de ocorrência de um erro de registo 3 vezes maior do que os casos cujo grau de diferenciação é conhecido.

Nenhum dos intervalos de tempo entre datas de 1^{os} sintomas, de 1^a observação, de 1^a consulta ou de diagnóstico se mostrou directamente associado com a ocorrência de erros de registo, quer como variáveis contínuas (representadas em número de dias) quer categorizadas em intervalos de tempo.

3.2.2. CONSTRUÇÃO DO MODELO MÚLTIPLO

Identificaram-se as seguintes variáveis candidatas ao modelo saturado:

- INSTITUIÇÃO
- GRUPO ETÁRIO
- SEXO
- REGIÃO DE NATURALIDADE
- TIPO DE DIAGNÓSTICO
- COMPORTAMENTO
- GRAU DE DIFERENCIAÇÃO

Aplicando o método *stepwise* de selecção de variáveis atinge-se o seguinte modelo de efeitos principais:

Tabela 4: Modelo de efeitos principais - coeficientes do modelo (β), odds ratios (OR), respectivos intervalos de confiança dos OR e valor-p do Teste de Wald

Variável	Coef. β	SE(β)	OR	I.C. 95%	T. Wald valor-p
CONSTANTE	-4,83	0,64	---	---	---
INSTITUIÇÃO					
Ref.: A					
B	1,99	0,59	7,35	2,31 – 23,37	0,001
GRUPO ETÁRIO					
Ref.: + 15 anos.					
0-14 anos	2,33	0,89	10,24	1,81 – 58,13	0,009
TIPO DIAGNÓST.					
Ref.: Hist. Primário					
Hist. Metástases	2,91	0,83	18,43	3,64 – 93,21	< 0,001
Autópsia c/ Hist.	2,26	1,42	9,62	0,60 – 153,99	0,110
Citologia	-0,63	0,91	0,53	0,09 – 3,14	0,486
GRAU DIFERENC.					
Ref.: Conhecido					
Desconhecido	1,09	0,45	2,97	1,24 – 7,12	0,015

O interesse em estudar o efeito das interações no modelo é o de avaliar o “paralelismo” entre os efeitos principais, ou seja, se a associação entre uma dada covariável e a ocorrência dos erros de registo é constante segundo os níveis de outra covariável. Uma interação é considerada relevante, do ponto de vista quantitativo, quando o seu efeito de entrada no modelo é estatisticamente significativo ($\alpha=0,050$). O critério é idêntico ao adoptado para os efeitos principais.

A inclusão de uma interação não significativa no modelo apenas complicará as inferências a realizar e conduzirá a uma menor precisão das estimativas obtidas dos coeficientes do modelo. O critério de selecção das interações faz-se através de *selecção propositada progressiva*.

O estudo das interações não se deve iniciar pela inclusão de todas as interações de interesse no modelo, dado que muitas vezes conduzem a problemas numéricos no processo inferencial.

Neste caso, o modelo de regressão logística que incluía as interações abaixo assinaladas a negrito pôde ser estimado por haver categorias com 0 elementos, quando cruzadas as variáveis.

INST. x GR.ETARIO	GR.ETARIO x T.DIAGN.
INST. x T.DIAGN.	GR. ETARIO x G.DIFER.
INST. x G.DIFER.	T.DIAGN. x G.DIFER.

Ajustou-se o modelo de efeitos principais e avaliou-se o efeito de inclusão de cada interação no modelo. Nenhuma interação se mostrou de importância estatisticamente significativa no modelo de efeitos principais.

Chegamos, assim, ao modelo final de regressão logística para a probabilidade $\Pi(x)$ de erro de nível 1:

$$\Pi(\mathbf{x}) = \frac{e^{-4.83+1.99 \times \text{INSTIT}_2+2.33 \times \text{GR.ETARIO}_2+2.91 \times \text{T.DIAGN}_2+2.26 \times \text{T.DIAGN}_3-0.63 \times \text{T.DIAGN}_4+1.09 \times \text{G.DIFER}_2}}{1+e^{-4.83+1.99 \times \text{INSTIT}_2+2.33 \times \text{GR.ETARIO}_2+2.91 \times \text{T.DIAGN}_2+2.26 \times \text{T.DIAGN}_3-0.63 \times \text{T.DIAGN}_4+1.09 \times \text{G.DIFER}_2}}$$

onde as covariáveis x são assim definidas:

INSTIT_2=0	→	se INSTITUIÇÃO é A
INSTIT_2=1	→	se INSTITUIÇÃO é B
GR.ETARIO_2=0	→	se GRUPO ETÁRIO é +15 anos
GR.ETARIO_2=1	→	se GRUPO ETÁRIO é 0-14 anos
T.DIAGN_2=0	→	se TIPO DE DIAGNÓSTICO não é Histologia de Metástases
T.DIAGN_2=1	→	se TIPO DE DIAGNÓSTICO é Histologia de Metástases
T.DIAGN_3=0	→	se TIPO DE DIAGNÓSTICO não é Autópsia com Histologia
T.DIAGN_3=1	→	se TIPO DE DIAGNÓSTICO é Autópsia com Histologia
T.DIAGN_4=0	→	se TIPO DE DIAGNÓSTICO não é Citologia
T.DIAGN_4=1	→	se TIPO DE DIAGNÓSTICO é Citologia
G.DIFER_2=0	→	se GRAU DE DIFERENCIAÇÃO é Conhecido
G.DIFER_2=1	→	se GRAU DE DIFERENCIAÇÃO é Desconhecido

Suponhamos um exemplo correspondente a um caso que ocorreu à *Instituição B*, com 36 anos de idade, do sexo feminino, natural de Cabo Verde e residente em Lisboa, com um carcinoma espino-celular do ovário, moderadamente diferenciado, diagnosticado por biópsia com um estágio TNM de 100. Este indivíduo apresenta, segundo o modelo ajustado, uma probabilidade estimada de erro de registo com impacte nos cálculos da incidência, dada por

$$\Pi(\mathbf{x}) = \frac{e^{-4,83+1,99 \times 1+2,33 \times 0+2,91 \times 0+2,26 \times 0-0,63 \times 0+1,09 \times 0}}{1 + e^{-4,83+1,99 \times 1+2,33 \times 0+2,91 \times 0+2,26 \times 0-0,63 \times 0+1,09 \times 0}} = 0,055$$

3.2.3. AVALIAÇÃO DA QUALIDADE DE AJUSTAMENTO DO MODELO FINAL

Para a avaliação da qualidade de ajustamento do modelo obtido utilizou-se o Teste do Qui-quadrado de Pearson e o Teste de Hosmer e Lemeshow.

Tabela 5: Avaliação da qualidade do ajustamento do modelo final

Teste	Estatística	Graus de Liberdade	Valor-p
QUI-QUADRADO DE PEARSON	4,69	9	0,860
HOSMER E LEMESHOW	0,54	5	0,990

Concluiu-se que, globalmente, não existe evidência de que o modelo de regressão logística esteja mal ajustado aos dados (para um nível de significância de 5%).

Individualmente, a avaliação do impacte no modelo, pela presença ou ausência de determinado elemento na amostra, é dada no gráfico que se segue:

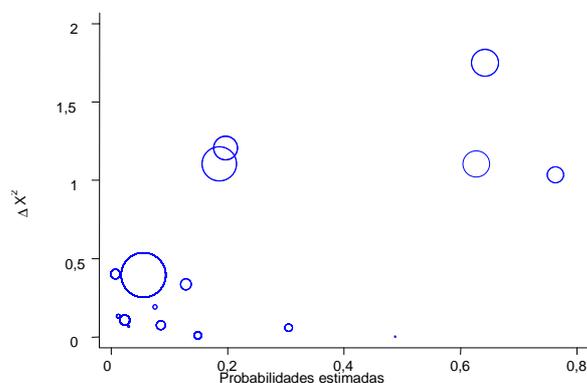


Figura 1: Gráfico de avaliação das medidas de influência de cada observação

As observações assinaladas com um círculo de maior diâmetro correspondem a observações influentes no modelo, mas pelo facto de não serem consideradas mal ajustadas (o decréscimo no valor da estatística de teste do Qui-quadrado de Pearson - ΔX^2 - devida à eliminação dos indivíduos com as convariáveis x_j é, para um nível de significância de 5%, inferior ao valor do quantil de probabilidade de um Qui-quadrado com um grau de liberdade - 3,84), decidiu-se mantê-las no modelo por forma a garantir o maior número de observações na amostra.

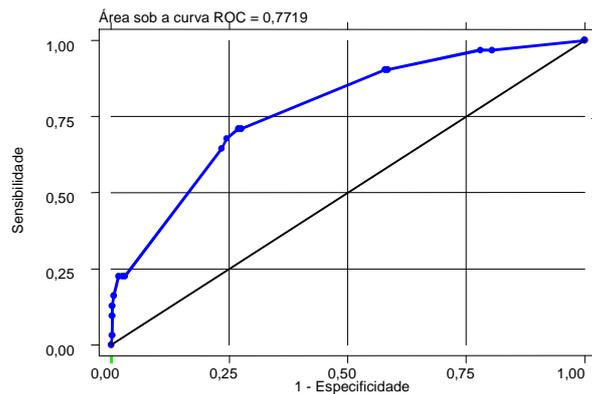


Figura 2: Curva ROC obtida pela aplicação do modelo final de regressão logística

O valor da área sob a curva ROC é de 0,772, indicando, segundo a classificação da tabela 1, que o poder de discriminação do modelo é considerado aceitável.

De facto, pelo gráfico da sensibilidade e especificidade (figura 3) verifica-se que os valores da sensibilidade do modelo (probabilidade do modelo detectar erros de registo, quando eles existam) diminuem drasticamente entre as probabilidades estimadas de 0 e 0,25.

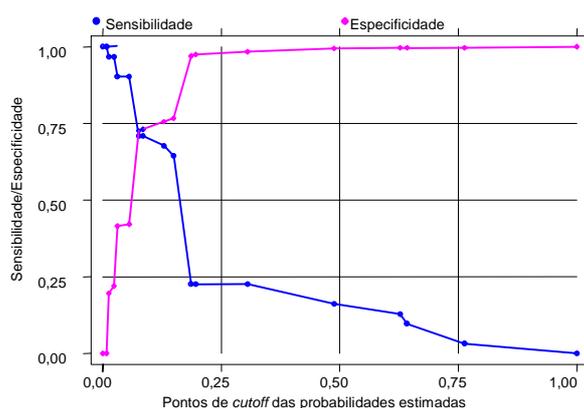


Figura 3: Sensibilidade e especificidade obtida pela aplicação do modelo final de regressão logística, para diversos valores de pontos cutoff das probabilidades estimadas

3.3. AVALIAÇÃO DA REPRODUTIBILIDADE DO MODELO FINAL

Aplicando o modelo final à segunda amostra, a discriminação do modelo, segundo os diferentes pontos de corte, foi a seguinte:

		Erro Observado			
		Cutoff	0	1	
Erro Estimado	0,5 e 0,4				Sensibilidade = $P(EE=1/EO=1) = 0,083$ Especificidade = $P(EE=0/EO=0) = 0,982$ Valor Preditivo Positivo = $P(EO=1/EE=1) = 0,250$ Valor Preditivo Negativo = $P(EO=0/EE=0) = 0,939$
	0	168	11	179	
	1	3	1	4	
	Total	171	12	183	

		Erro Observado			
		Cutoff	0	1	
Erro Estimado	0,3 e 0,2				Sensibilidade = $P(EE=1/EO=1) = 0,167$ Especificidade = $P(EE=0/EO=0) = 0,971$ Valor Preditivo Positivo = $P(EO=1/EE=1) = 0,286$ Valor Preditivo Negativo = $P(EO=0/EE=0) = 0,943$
	0	166	10	176	
	1	5	2	7	
	Total	171	12	183	

		Erro Observado			
		Cutoff	0	1	
Erro Estimado	0,1				Sensibilidade = $P(EE=1/EO=1) = 0,500$ Especificidade = $P(EE=0/EO=0) = 0,649$ Valor Preditivo Positivo = $P(EO=1/EE=1) = 0,091$ Valor Preditivo Negativo = $P(EO=0/EE=0) = 0,949$
	0	111	6	117	
	1	60	6	66	
	Total	171	12	183	

Mesmo assumindo um *cutoff point* de 0,1 a sensibilidade assume valores de 50%, significando que o modelo apenas assinala 50% das observações com erros de registo. O verdadeiro valor preditivo do modelo é de 9,1%, indicando que apenas esta proporção das observações estimadas como tendo erro é que realmente o têm.

4. DISCUSSÃO E CONCLUSÕES

O uso de modelos de regressão logística para a caracterização da exactidão dos dados foi apenas encontrado num estudo americano (Polissar *et al.*, 1984), embora num âmbito diferente daquele aqui estudado. Os autores fazem uso desta ferramenta para determinar diferenças estatisticamente significativas nos níveis de desacordo, ajustando simultaneamente para diferentes variáveis categorizadas.

O nosso objectivo vai um pouco mais além; pretendemos com a regressão logística encontrar um modelo que preveja a probabilidade de um determinado registo conter erros que possam alterar as estimativas das incidências por tumores malignos.

Dada a vasta dimensão de cobertura do ROR-Sul e os escassos recursos humanos de que dispõe para a implementação de procedimentos de monitorização da qualidade dos dados, pareceu-nos que este poderia ser um caminho de optimização do esforço a realizar.

Chegando ao final da revisão das duas amostras seleccionadas, foi com apreensão que nos defrontámos com apenas 43 casos (7,3%) que apresentavam desacordos de nível 1.

Na verdade, o estudo iniciou-se pressupondo uma percentagem de erros da ordem dos 20%, com um erro de previsão de 5%. Este facto levantou-nos uma primeira hipótese de que as amostras seleccionadas poderiam não ser suficientemente representativas para ajustar um modelo de predição dos erros de registo. A amostra, seleccionada aleatoriamente antes da revisão, que iria dar origem ao ajustamento do modelo de erros de registo, continha uma percentagem ligeiramente superior de erros de nível 1 (7,6%) do que a 2ª amostra de avaliação da reprodutibilidade desse modelo (6,6%).

O modelo múltiplo encontrado identificou 4 variáveis como relevantes na predição dos erros de registo: a instituição responsável pelo registo, o escalão etário, o tipo de diagnóstico e o grau de diferenciação do tumor. Não ocorreram interacções estatisticamente significativas entre estas variáveis.

A presença destas variáveis no modelo não é de todo de difícil compreensão:

- ✓ A Instituição faz sentido na medida em que dependendo do nível de formação e envolvimento das pessoas na tarefa do registo se conseguem ou não melhores resultados. Muitas vezes, se apenas uma pessoa fizer o registo dos casos, mais padronizados ficam os procedimentos e menor variabilidade encontramos. Por outro lado, a forma como os processos clínicos estão organizados na instituição pode influenciar claramente a interpretação dos dados e a sua recolha;
- ✓ O escalão etário definiu que os tumores infantis têm maior probabilidade de conter erros de nível 1. Este facto deveu-se em grande parte às crianças vindas dos PALOP cuja residência tinha sido codificada com a morada da embaixada dos seus países em Portugal. Outra explicação poderá ser também o comportamento incerto, se benigno ou maligno, dos tumores de algumas crianças que podem dar origem a erros de interpretação e codificação;
- ✓ O tipo de diagnóstico revelou que os diagnósticos feitos com base em histologias de metástases e autópsias com histologia são aqueles que maior probabilidade têm de conter erros. Estes ocorrem porque muitas vezes se regista como tumores primários o aparecimento de metástases, registando-se a localização das metástases e data do seu aparecimento como topografia e data do diagnóstico, respectivamente;

- ✓ O grau de diferenciação é o factor cuja interpretação é menos objectiva. Verificou-se que são os tumores cujo grau de diferenciação é omissivo que maiores probabilidades têm de conter erro. Poderíamos pensar que estes casos corresponderiam a diagnósticos histológicos menos claros ou mesmo a casos sem diagnóstico histológico; no entanto, o grau de diferenciação não revelou significância estatística na interacção com o tipo de diagnóstico.

O modelo revelou uma boa qualidade de ajustamento, com algumas observações influentes mas sem capacidade para questionar a hipótese de bom ajustamento.

Apesar da discriminação do modelo ser considerada aceitável, verifica-se que a sensibilidade deste diminui drasticamente quando se aumenta a probabilidade de *cutoff* de 0 a 1. Ao avaliarmos a reprodutibilidade do modelo verificamos que, considerando um ponto de corte de 0,1, o modelo apenas é capaz de prever 50% dos erros que existem (sensibilidade) e apenas 9% dos registos estimados como contendo um erro, o têm na realidade (valor preditivo positivo). Um modelo com este desempenho não pode ser considerado uma boa ferramenta de optimização de esforços de monitorização da qualidade.

Outra fragilidade que nos pareceu clara é que dependendo da amostra seleccionada o modelo múltiplo pode não encontrar variáveis que predigam os erros de registo, dada a sua falta de representatividade.

Verificou-se que a variabilidade dos erros é bastante grande, fazendo com que a amostra deva ter uma dimensão muito maior para que seja possível caracterizar todos os erros cometidos de forma sistemática, com a prevalência necessária para serem modelados.

Outro caminho sugerido pela experiência indica que a opção dos modelos de regressão logística pode não ser a mais adequada. Outros modelos menos utilizados mas igualmente apropriados para variáveis respostas dicotómicas poderão ser alternativas a testar.

BIBLIOGRAFIA

BECK, J. R., SHULTZ, E. K., (1986), "The Use of Relative Operating Characteristic (ROC) Curves in Test Performance Evaluation", *Archives of Pathology and Laboratory Medicine*, Vol. 110, 13-20.

BEGGL, C. B., CRAMER, L. D., VENKATRAMAN, E. S., ROSAI, J., (2000), "Comparing Tumour Staging and Grading Systems: a Case Study and a Review of the Issues, Using Thyoma as a Model", *Statistics in Medicine*, Vol. 19, Nº 15, pp. 1997-2014.

- BRICKLEY, M. R., PRYTHERCH, I. M., KAY, E. J., SHEPERD, J. P., (1995), "A New Method of Assessment of Clinical Teaching: ROC Analysis", *Medical Education*, Vol. 29, pp. 150-153.
- DREISEITL, S., OHNO-MACHADO, L., BINDER, M., (2000), "Comparison Three-Class Diagnostic Tests by Three-Way ROC Analysis", *Medical Decision Making*, Vol. 20, Nº 3, pp. 323-331.
- FORTESCUE, E. B., KAHN, K., BATES, D. W., (2000), "Prediction Rules for Complications in Coronary Bypass Surgery: a Comparison and Methodological Critique", *Medical Care*, Vol. 38, Nº 8, pp. 820-835.
- GOLDBERG, J., GELFAND, H. M., LEVY, P. S., (1980), "Registry Evaluation Methods: a Review and a Case Study", *Epidemiologic Reviews*, Vol. 2, pp. 210-220.
- HAGEN, M. D., (1995), "Test Characteristics: How Good Is That Test?", *Medical Decision Making*, Vol. 22, Nº 2, pp. 213-233.
- HANLEY, J. A., MCNEIL, B. J., (1982), "The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve", *Radiology*, Vol. 143, pp. 29-36.
- HANLEY, J. A., MCNEIL, B. J., (1983), "A Method of Comparing the Areas under a Receiver Operating Characteristic Curves Derived from the Same Cases", *Radiology*, Vol. 148, pp. 839-843.
- HILSENBECK, S. G., (1994) "Quality Control", *Central Cancer Registries: Design, Management, and Use*, Harwood Academic Publishers, Illinois, pp. 131-177.
- HILSENBECK, S. G., GLAEFKE, G. S., FEIGL, P., LANE, W. W., GOLENZER, H., AMES, C., DICKSON, C., (1985), *Quality Control for Cancer Registries*, U. S. Department of Health and Human Services, Seattle-WA.
- HOSMER, D. W., LEMESHOW, S., (1989), *Applied Logistic Regression*, John Wiley & Sons, New York.
- HOSMER, D., HOSMER, T., LE CESSIE, S., LEMESHOW, S., (1997), "A Comparison of Goodness-of-Fit Tests for the Logistic Regression Model", *Statistics in Medicine*, Vol. 16, Nº 9, pp. 965-980.
- HOSMER, D., LEMESHOW, S., (1980), "A Goodness-of-Fit Test for the Multiple Logistic Regression Model", *Communications in Statistics*, Series A10, pp. 1043-1069.
- HOSMER, D., LEMESHOW, S., KLAR, J., (1988), "Goodness-of-Fit Testing for Multiple Logistic Regression Analysis When the Estimated Probabilities are Small", *Biometrical Journal*, Vol. 30, Nº 7, pp. 1-14.
- KLEINBAUM, D. G., KUPPER, L. L., MORGENSTERN, H. (1982), *Epidemiologic Research: Principles and Quantitative Methods*, Van Nostrand Reinhold, New York;
- LEMESHOW, S., HOSMER, D., (1982), "A Review of Goodness-of-Fit Statistics for Use in the Development of Logistic Regression models", *American Journal of Epidemiology*, Vol. 115, pp. 92-106.

- METZ, C., (1978), "Basic Principles of ROC Analysis", *Seminars in Nuclear Medicine*, Vol. 8, pp. 283-297.
- MIRANDA, A. C., (1998), "Registo Oncológico Regional Sul", *Automated Data Collection in Cancer Registration*, IARC Technical Reports N° 32, Lyon, pp. 45-47.
- MUIR, C. S., DÉMARET, E., and BOYLE, P., (1985) "The Cancer Registry in Cancer Control: an Overview", *The Role of the Registry in Cancer Control*, IARC Scientific Publication N° 66, Lyon, pp. 13-26.
- PARKIN, D. M., MUIR, C. S., (1992), "Comparability and Quality of Data", *Cancer Incidence in Five Continents, Volume VI*, IARC Scientific Publications N° 120, Lyon, pp. 45-56.
- POLISSAR, L., FEIGL, P., LANE, W. W., GLAEFKE, G., DAHLBERG, S., (1984), "Accuracy of Basic Cancer Patient Data: Results from an Extensive Recoding Survey", *Journal of National Cancer Institute*, Vol. 72, N° 5, pp. 1007-1013.
- ROFFERS, S. D., (1996), "Case Completeness and Data Quality Assessments in Central Cancer registries and Their Relevance to Cancer Control", <http://www.naaccr.org/data/papers/sect94.pdf>.
- SCHULTZ, E. K., (1995), "Multivariate Receiver-Operating Characteristic Curve Analysis: Prostate Cancer Screening as an Example", *Clinical Chemistry*, Vol. 41, Series 8B, pp. 1248-1255.
- SKEET, R. G., (1991) "Quality and Quality Control", *Cancer Registration: Principles and Methods*, IARC Scientific Publication N° 95, Lyon, pp. 101-107.
- ZIPPIN, C., AKERS, C., (1985), "Quality Control in Cancer Registration Programs", *Current Problems in Cancer*, Vol. 9, N° 3, pp. 49-60.