

Metodologia de captura-recaptura: uma opção para a vigilância das doenças não transmissíveis na população idosa

Capture-recapture methodology:
an option for surveillance of non-communicable
diseases in the elderly

Cláudia Medina Coeli ¹

Renato Peixoto Veras ²

Evandro da Silva Freire Coutinho ³

¹ Departamento de Medicina Preventiva, Faculdade de Medicina, Núcleo de Estudos de Saúde Coletiva, Universidade Federal do Rio de Janeiro.

Av. Brigadeiro Trompowski s/nº, Edifício do Hospital Universitário Clementino Fraga Filho, 5º andar, Ala Sul, Rio de Janeiro, RJ 21931-059, Brasil.
coeli@acd.ufrj.br

² Departamento de Epidemiologia, Instituto de Medicina Social, Universidade do Estado do Rio de Janeiro.

Rua São Francisco Xavier 524, 7º andar, Bloco D, Rio de Janeiro, RJ 20559-900, Brasil.
veras@uerj.br

³ Departamento de Epidemiologia e Métodos Quantitativos em Saúde, Escola Nacional de Saúde Pública, Fundação

Oswaldo Cruz, Rua Leopoldo Bulhões 1480, Rio de Janeiro, RJ 21041-210, Brasil.
evandro@manguinhos.ensp.fiocruz.br

Abstract Capture-recapture methodology is used in Ecology to estimate the total size of wild animal populations. This method can be used in Epidemiology to estimate the incidence and prevalence of communicable and non-communicable diseases in a relatively inexpensive and quick way. Surveillance systems based on the use of this methodology are a potential alternative for monitoring non-communicable diseases in Brazil. This paper presents a brief review of fundamental capture-recapture methodology and its applications to Epidemiology. We also present a theoretical model for implementation of a diabetes mellitus surveillance system in the elderly using publicly available morbidity and mortality data sources and the capture-recapture methodology.

Key words Diabetes Mellitus; Epidemiologic Surveillance; Aged

Resumo A metodologia de captura-recaptura é utilizada na área da ecologia para a estimativa do tamanho de populações de animais selvagens. Esta técnica pode ser empregada no âmbito da epidemiologia para a obtenção de estimativas de doenças transmissíveis e não transmissíveis em uma forma relativamente barata e rápida. Sistemas de vigilância baseados na utilização desta metodologia representam uma alternativa potencial para o monitoramento de doenças não transmissíveis em nosso meio. Este artigo apresenta uma revisão sucinta dos fundamentos da metodologia de captura-recaptura e de suas aplicações no âmbito da epidemiologia. Adicionalmente, é apresentado um modelo teórico para a implantação de um sistema de vigilância do diabetes mellitus na população idosa baseado na utilização de fontes de morbidade e mortalidade usualmente disponíveis em nosso meio e na aplicação da metodologia de captura-recaptura.

Palavras-chave Diabetes Mellitus; Vigilância Epidemiológica; Idosos

Introdução

As estatísticas sobre morbidade na população idosa brasileira são escassas, o que em parte é explicado pelo número reduzido de trabalhos científicos na área da Terceira Idade (Veras & Alves, 1995). Por outro lado, esta faixa etária é acometida preponderantemente pelas doenças crônico-degenerativas, para as quais os sistemas de monitoramento sistemático são raros em nosso meio. Recentemente a Organização Mundial da Saúde (OMS) (WHO, 1993a, 1993b) alertou para a importância do desenvolvimento de estratégias para a prevenção e controle das doenças não transmissíveis, incluindo a implantação de sistemas de monitoramento integrados e padronizados, tanto em nível nacional como internacional.

O modelo adotado para a vigilância das doenças transmissíveis, baseado na identificação de casos feita principalmente pela notificação aos órgãos de Saúde Pública (notificação passiva), não é adequado para o monitoramento das doenças não transmissíveis em função da cobertura incompleta de casos geralmente observada (Goldberg et al., 1980; WHO, 1993b). O registro de base populacional foi o modelo adotado para o monitoramento das doenças não transmissíveis. A característica mais marcante deste tipo de sistema é a tentativa de identificação de todos os casos de uma determinada doença em uma população com tamanho e composição conhecidos (Goldberg et al., 1980). O maior empecilho à implantação generalizada dos registros de base populacional é o alto custo, já que é necessária a manutenção de uma equipe de profissionais envolvidos com a busca ativa de casos, a codificação, o processamento e a análise de dados (Goldberg et al., 1980; WHO, 1993b).

Os sistemas de vigilância deveriam buscar ser tão baratos quanto os da vigilância das doenças transmissíveis e, para serem efetivos, deveriam combinar os atributos rapidez, escopo de doenças ampliado e ampla cobertura geográfica (vigilância das doenças transmissíveis), com uma identificação ampla dos casos (registro de base populacional das doenças não transmissíveis). LaPorte et al. (1992) propõem como solução para esta questão a utilização pelos epidemiologistas de uma metodologia de "contar" que permitisse estimar o tamanho de uma população sem a necessidade de enumerar todos os elementos que a compõem. A metodologia de captura-recaptura, utilizada na área da ecologia para a estimativa do tamanho de populações de animais, representa uma alternativa eficiente para ser aplicada em situa-

ções onde se deseja estimar o tamanho de populações humanas.

Neste artigo buscamos apresentar uma revisão sucinta dos fundamentos da metodologia da captura-recaptura e de suas aplicações no âmbito da epidemiologia. Adicionalmente, é apresentado um modelo teórico para a implantação de um sistema de vigilância do diabetes mellitus na população idosa baseado na utilização de fontes de morbidade e mortalidade usualmente disponíveis em nosso meio, e na aplicação da metodologia de captura-recaptura.

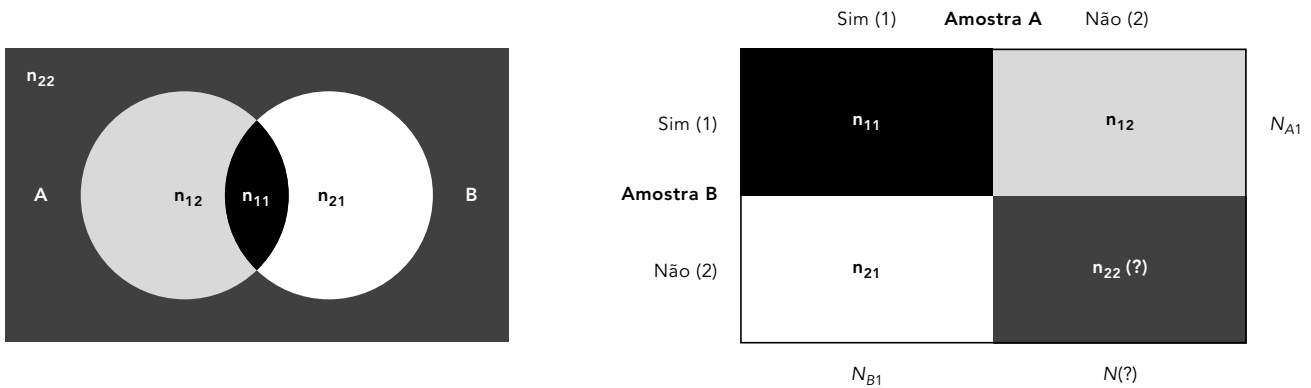
Fundamentos da metodologia

Foi no campo da ecologia, principalmente a partir das décadas de 30 e 40, onde se observou um consistente desenvolvimento teórico e aplicado do método de captura-recaptura para a estimação de populações de animais selvagens (White et al., 1982:1). Na área da ecologia, a forma mais simples de aplicação da metodologia de captura-recaptura envolve a coleta de duas amostras independentes em dois momentos distintos (amostras seqüenciais) de uma população de animais fechada, ou seja, na qual não são observados nascimentos, mortes ou migrações durante o período de estudo. Além disso, todos os elementos desta população devem ser homogêneos em relação aos atributos que determinam a probabilidade de seleção. Em outras palavras, todos os elementos devem apresentar a mesma probabilidade de captura.

Suponha, por exemplo, que se deseja estimar a população de peixes de uma determinada espécie em um lago empregando a metodologia de captura-recaptura. Inicialmente seria capturada uma amostra (amostra A) de peixes, que seriam contados, marcados e libertados. Após um período de tempo suficientemente longo para que os peixes marcados se misturassem aos não marcados, uma nova amostra (amostra B) seria capturada. Os peixes capturados na segunda amostra seriam contados e o número de peixes recapturados seria identificado. O resultado deste processo poderia ser representado por meio do Diagrama de Venn e da tabela de contingência apresentados na Figura 1. Assumindo-se a hipótese de independência entre as amostras, então poderíamos concluir que a proporção de peixes recapturados em relação ao total de peixes capturados pela segunda amostra seria igual à proporção de peixes capturados pela primeira amostra em relação ao total da população:

Figura 1

Diagrama de Venn e tabela de contingência relativos a um processo de captura-recaptura envolvendo duas fontes.



Onde,

- $N = n_{11} + n_{12} + n_{21} + n_{22}$ = tamanho da população.
- $N_{A1} = n_{11} + n_{12}$ = total de peixes capturados na primeira amostra (amostra A).
- $N_{B1} = n_{11} + n_{21}$ = total de peixes capturados na segunda amostra (amostra B).
- n_{11} = número de peixes capturados na primeira e segunda amostras.
- n_{12} = número de peixes capturados somente na primeira amostra.
- n_{21} = número de peixes capturados somente na segunda amostra.
- n_{22} = número de peixes não capturados em nenhuma amostra.

Além disso,

- $p_{A1} = \frac{N_{A1}}{N}$ = probabilidade de ser capturado pela primeira amostra.
- $p_{B1} = \frac{N_{B1}}{N}$ = probabilidade de ser capturado pela segunda amostra.
- $p_{A1B1} = \frac{n_{11}}{N}$ = probabilidade de ser capturado por ambas as amostras.

$$\frac{n_{11}}{N_{B1}} = \frac{N_{A1}}{N} \tag{1}$$

Resolvendo para N , chegamos ao seguinte estimador:

$$\hat{N} = \frac{N_{A1} N_{B1}}{n_{11}} \tag{2}$$

que é conhecido como estimador de Petersen (Hook & Regal, 1995; IWGDME, 1995a). Este estimador é enviesado para amostras pequenas, levando a uma superestimativa do real tamanho populacional. Chapman (Hook & Regal, 1995) propôs um ajustamento que apresenta melhores propriedades e que é conhecido como estimador “quase não enviesado”, sendo expresso por:

$$\hat{N} = \left[\frac{(N_{A1} + 1)(N_{B1} + 1)}{(n_{11} + 1)} \right] - 1 \tag{3}$$

Para que os estimadores acima descritos não sejam enviesados, é fundamental que a premissa de independência entre as amostras não seja violada. Considerando que a captura

de peixes é um processo aleatório, então N_{A1} , N_{B1} e n_{11} são variáveis aleatórias, com os respectivos valores esperados definidos por: $E[N_{A1}] = Np_{A1}$, $E[N_{B1}] = Np_{B1}$ e $E[n_{11}] = Np_{A1B1}$. Sendo assim, para N suficientemente grande (i. e., $N \rightarrow \infty$) teremos aproximadamente (IWGDME, 1995a):

$$E[\hat{N}] \cong \frac{E[N_{A1}] E[N_{B1}]}{E[n_{11}]} \tag{4}$$

Portanto:

$$E[\hat{N}] \cong N \frac{P_{A1} P_{B1}}{P_{A1B1}} \tag{5}$$

Se as amostras são independentes, a probabilidade de ser capturado por ambas as amostras será igual ao produto entre a probabilidade de ser capturado pela primeira amostra e a probabilidade de ser capturado pela segunda amostra ($P_{A1B1} = P_{A1}P_{B1}$). A razão $P_{A1}P_{B1} / P_{A1B1}$ será igual a 1 e \hat{N} será aproximadamente igual a N . Entretanto, se as amostras não são independentes, a razão $P_{A1}P_{B1} / P_{A1B1}$ será diferente de 1 e \hat{N} será enviesado.

Considere-se, por exemplo, que o processo de captura $e/$ ou marca dos peixes na primeira amostra provocasse algum dano, fazendo com que os peixes capturados na primeira amostra fossem mais facilmente recapturados. Neste caso seria observada uma “dependência positiva”, visto que o fato de ter sido capturado pela primeira amostra aumentaria a probabilidade de ser capturado pela segunda amostra, isto é, a probabilidade de ser capturado por ambas as amostras seria maior que o produto entre a probabilidade de ser capturado pela primeira amostra e a probabilidade de ser capturado pela segunda amostra ($P_{A_1B_1} > P_{A_1}P_{B_1}$). Como consequência, a razão $P_{A_1}P_{B_1} / P_{A_1B_1}$ será menor que 1 e \hat{N} irá subestimar N .

Em outra situação, poderíamos imaginar que o processo de captura não provocasse nenhuma alteração nos peixes mas que a espécie em questão fosse capaz de “aprender” a evitar a armadilha após uma primeira captura. Observaríamos, portanto, um padrão oposto ao primeiro (dependência negativa), ficando a probabilidade de ser capturado por ambas as amostras menor que produto entre a probabilidade de ser capturado pela primeira amostra e a probabilidade de ser capturado pela segunda amostra ($P_{A_1B_1} < P_{A_1}P_{B_1}$). Conseqüentemente, a razão $P_{A_1}P_{B_1} / P_{A_1B_1}$ será maior que 1 e \hat{N} irá superestimar N .

Em adição à premissa de independência entre as amostras, a obtenção de estimadores válidos só será alcançada se todos os elementos da população apresentarem-se homogêneos em relação aos atributos que determinam a probabilidade de captura. Neste caso, para cada amostra, cada elemento da população deve apresentar a mesma probabilidade de ser incluído (capturabilidade homogênea). No exemplo dos peixes, esta premissa poderia ser violada caso a nossa população não fosse homogênea, mas sim constituída por dois estratos: o primeiro formado por peixes ágeis e, portanto, mais dificilmente capturáveis; enquanto o segundo, formado por peixes menos ágeis e, conseqüentemente, mais facilmente capturáveis. Nesta situação, mesmo garantindo a independência entre as amostras dentro de cada estrato, o resultado final para o conjunto da população seria de uma “aparente dependência” entre as amostras (IWGDME, 1995a) e os estimadores seriam enviesados.

O efeito da variabilidade de capturabilidade sobre a dependência entre as amostras A e B, pode ser comparado ao efeito de uma variável de confusão sobre a associação entre uma variável de exposição e um determinado desfecho, embora no primeiro caso não se busque

uma interpretação causal da relação entre as duas amostras. No exemplo acima, poderíamos considerar a captura pela primeira amostra como a “exposição” e a captura pela segunda amostra como o “desfecho”. Neste caso o atributo agilidade assumiria um papel semelhante ao de uma variável de confusão, ou seja, estaria associado à captura pela primeira amostra (maior agilidade, menor probabilidade de captura) e determinaria uma menor probabilidade de captura pela segunda amostra. Desta maneira, mesmo que as amostras fossem independentes em cada estrato, se considerássemos o conjunto da população sem levar em conta a presença de heterogeneidades, encontraríamos uma “aparente dependência entre as amostras”.

Hook & Regal (1993, 1995) mostram que a direção final do enviesamento do estimador do tamanho populacional (sub ou superestimativa), na presença de variabilidade de capturabilidade, dependerá da direção da variabilidade em cada amostra nos diferentes estratos. No exemplo dos peixes acima apresentado, a direção é semelhante para as duas amostras (armadilhas), ou seja, a maior agilidade leva a uma menor capturabilidade em ambas as amostras. Sendo assim, o efeito final será de dependência positiva, e o estimador \hat{N} irá subestimar N . Se, todavia, o atributo determinante da variabilidade de capturabilidade fosse o tamanho do peixe e fossem utilizadas armadilhas diferentes, de forma que a primeira capturasse mais facilmente os peixes maiores, enquanto a segunda os peixes menores, a direção seria oposta, isto é, o estrato dos peixes maiores teria maior probabilidade de captura pela primeira armadilha do que pela segunda, enquanto o estrato dos peixes menores apresentaria comportamento oposto. Neste caso, o efeito final seria de dependência negativa, e o estimador \hat{N} iria superestimar N .

Outrossim, pode vir a ser observado um comportamento heterogêneo da dependência entre as amostras nos diferentes estratos; conceito análogo ao de interação causal (efeito modificador). Hook & Regal (1993, 1995) mostram que os efeitos de dependência e variabilidade de capturabilidade podem ocorrer conjuntamente e que a direção final do enviesamento do estimador dependeria da combinação destes fatores.

Aplicações da metodologia de captura-recaptura na epidemiologia

A idéia de estimar o tamanho de uma população com uma determinada característica (presença de doença, por exemplo) no lugar de enu-

merar todos os indivíduos que a compõem, não é estranha aos epidemiologistas. Na realidade, a extrapolação de resultados para a população fundamentada na observação de uma amostra de indivíduos é amplamente empregada na pesquisa epidemiológica, sendo a base dos inquéritos epidemiológicos desenvolvidos para a estimativa da prevalência de doenças na população. Contudo, como LaPorte (1993, 1994) salienta, estes estudos são caros, demorados e cobrem áreas geográficas reduzidas, fazendo com que a sua aplicação para o planejamento e avaliação dos serviços e ações de Saúde Pública seja limitada.

Listas de doenças de notificação compulsória, estatísticas hospitalares e de outros serviços de saúde, registros de óbitos, entre outros, são fontes de dados sobre morbidade usualmente disponíveis. A utilização conjunta de duas ou mais destas fontes e a aplicação da metodologia de captura-recaptura poderiam ser uma alternativa barata e efetiva para o monitoramento de doenças na população.

Nas aplicações em epidemiologia, cada lista (fonte de morbidade) é considerada uma amostra aleatória simples da população alvo (Wittes & Sidel, 1968). “Ser capturado por uma amostra” seria substituído por “estar registrado em uma lista” (IWGDME, 1995a). Cada elemento da lista deve ser identificado univocamente (“marca”) de maneira a permitir o conhecimento do número de indivíduos que aparece simultaneamente em mais de uma lista (relacionamento de registros). Este identificador geralmente é formado com base na combinação de dois ou mais dos seguintes atributos: nome, sobrenome, data de nascimento e endereço (representado geralmente pelo bairro e/ou código de endereçamento postal). Ao contrário do observado nas aplicações na ecologia, não existe uma ordenação temporal das amostras (Hook & Regal, 1995).

As premissas para a obtenção de estimativas válidas são semelhantes àquelas das aplicações na área da ecologia, a saber: a) população fechada; b) sistema de marcação eficiente que permita o relacionamento dos casos nas diferentes amostras (listas); c) as listas devem ser independentes; d) para cada lista cada elemento da população deve apresentar a mesma probabilidade de ser incluído (capturabilidade homogênea). Adicionalmente Hook & Regal (1995a) consideram necessário que os casos tenham sido diagnosticados corretamente nas diferentes listas (validade de diagnóstico) e que todos os casos pertençam a mesma unidade espaço-temporal de interesse (base populacional do estudo).

As aplicações da metodologia da captura-recaptura na área da epidemiologia envolvem a vigilância em Saúde Pública, a avaliação da cobertura de registros (registros de nascidos vivos, por exemplo), o refinamento de estimativas de incidência e prevalência e a derivação de valores plausíveis para os limites superior e inferior de estimativas, nas situações em que estimativas pontuais precisas não possam ser obtidas (Hook & Regal, 1995).

Aplicações na epidemiologia: limitações e soluções propostas

As principais limitações da utilização da metodologia de captura-recaptura nas aplicações epidemiológicas envolvem a violação dos pressupostos básicos do método. Muitos destes problemas também são observados nas aplicações no campo da ecologia; porém, diferenças tais como a existência de ordenação temporal entre as amostras, específica das aplicações ecológicas, e a natureza da variabilidade de capturabilidade, fazem com que as soluções utilizadas nas duas áreas sejam diferentes. Passaremos a seguir a detalhar estas limitações no âmbito da epidemiologia e as soluções empregadas para contorná-las.

População fechada

Considera-se como fechada, toda a população na qual não são observados nascimentos, mortes ou migrações durante o período de estudo. O tamanho de uma população fechada se mantém constante ao longo do período de observação. Acrescente-se que, esta suposição contribui para a eliminação da variabilidade de capturabilidade determinada por diferenças individuais no período de exposição às fontes de captura. Ou seja, indivíduos que entram na população após o início do estudo, ou que abandonam antes do seu término, apresentam uma menor chance de serem capturados e/ou recapturados do que aqueles que permanecem na população ao longo de todo o período de observação. A premissa de população fechada raramente é respeitada nas aplicações no campo da epidemiologia, mas na maioria dos casos ela pode ser considerada aproximadamente atendida (IWGDME, 1995a).

Na área de ecologia considera-se necessário respeitar a premissa de “fechamento geográfico” em adição à premissa de “fechamento demográfico” acima descrito (White et al., 1982: 3). O critério geográfico implica não somente a definição precisa da área habitada pela popu-

lação de animais que se deseja estimar, como também a garantia de que as armadilhas utilizadas cubram toda a área definida. Por exemplo, esta suposição tem maior chance de ser satisfeita quando se deseja estimar a população de peixes de um pequeno lago do que quando se deseja estimar a população de peixes de um reservatório grande. No segundo caso é possível que as armadilhas deixem de cobrir determinadas áreas do reservatório, fazendo com que os peixes que as habitam tenham probabilidade nula de serem capturados.

O “fechamento geográfico” seria alcançado nas aplicações epidemiológicas, se definíssemos a base populacional do estudo e implementássemos uma estratégia amostral que permitisse captar toda a experiência da base selecionada. Com relação a este último aspecto, Hook & Regal (1995) recomendam a diversificação quantitativa (maior número possível) e qualitativa (naturezas diversas) das fontes de dados. Esta estratégia minimizaria a chance de que algum segmento da base populacional apresentasse probabilidade nula de ser captado.

Como as fontes empregadas nas aplicações epidemiológicas não são, na maioria dos casos, geradas pelo pesquisador, mas produzidas com outras finalidades, a pergunta a ser respondida não é mais “Sobre quem queremos falar?”, mas sim “Sobre quem podemos falar?”. Por exemplo, casos subclínicos dificilmente são captados pela maioria das fontes disponíveis.

Validade do diagnóstico e de outras informações

A obtenção de estimativas válidas também dependerá da qualidade das informações disponíveis em cada fonte. Em outras palavras, deve-se evitar a ocorrência de vies de informação. Particularmente nos estudos de captura-recaptura, o interesse é centrado nas informações relativas ao diagnóstico, às variáveis envolvidas no relacionamento de registros (nome, sexo, data de nascimento) e às variáveis que permitem a identificação do caso dentro da base populacional definida (endereço de residência, data em que foi captado em determinada fonte e qualquer atributo adicional utilizado como critério de inclusão na base populacional).

Hook & Regal (1995) ressaltam que existem poucos estudos que tenham procurado modelar quantitativamente a consequência deste tipo de erro. Estes autores lembram que os vieses de diagnóstico nas diferentes fontes podem não ser independentes e que este fenômeno poderia provocar uma dependência entre fontes ou alterar um padrão de dependência espe-

rado pela ação de outros fatores que não o diagnóstico. Quando a metodologia de captura-recaptura é empregada com o objetivo de refinar estimativas de prevalência e incidência durante a realização de estudos especiais sobre uma determinada doença ou dentro de sistemas de vigilância baseados na busca ativa de casos (registro de base populacional), o problema do vies de informação diminui, já que todos os casos são revisados para a confirmação do diagnóstico. No entanto, a implantação da metodologia para a vigilância rotineira de uma vasta gama de doenças simultaneamente, tal qual proposto por LaPorte (LaPorte, 1994), inviabiliza a revisão de todos os casos. Uma possível solução para o problema seria a revisão de uma amostra de registros em cada fonte e o desenvolvimento de fatores de correção baseados na informação sobre a acurácia de cada fonte.

Com relação às variáveis utilizadas para o relacionamento de registros o problema é mais complexo, especialmente em relação ao nome. O registro do nome é sujeito a erros de ortografia e de transposição de caracteres no momento da digitação, o que pode dificultar a identificação de um par verdadeiro. Por outro lado, a existência de homônimos pode levar a identificação de um par falso. Estes problemas podem ser contornados com a utilização de programas especiais que implementam a metodologia de relacionamento probabilístico de registros (Carmargo Jr. & Coeli, 1998).

Dependência entre fontes e variabilidade de capturabilidade

A dependência surgiria de características “intrínsecas” das fontes, fazendo com que indivíduos captados por uma fonte tivessem maior (ou menor) probabilidade de serem capturados por uma segunda fonte em relação aos não capturados pela primeira fonte. Médicos que encaminham seus pacientes para serem internados em um determinado hospital, casos de doenças genéticas sendo encaminhados para confirmação em laboratórios especializados, e pacientes recebendo acompanhamento ambulatorial adequado e, conseqüentemente, apresentando menor probabilidade de hospitalização e óbito, seriam exemplos de dependências entre fontes.

Já a variabilidade de capturabilidade seria determinada por fatores geográficos, sócio-econômicos e biológicos (gravidade da doença e idade, por exemplo) que se distribuem de forma heterogênea na população e influenciam a chance de captura por uma determinada lista. Estes dois fenômenos são, entretanto, intima-

mente relacionados e a distinção entre ambos nem sempre é fácil. Na área da ecologia a questão da “verdadeira” dependência entre as armadilhas não é considerada, isto é, qualquer dependência entre armadilhas é atribuída à variabilidade de capturabilidade que pode ser de três tipos: temporal, devida a heterogeneidades individuais e devida a respostas comportamentais (White et al., 1982).

A variabilidade temporal surge em função da ordenação temporal nos processos de captura e recaptura dos animais fazendo com que mudanças climáticas, por exemplo, possam modificar as probabilidades de captura de uma amostra para outra (White et al., 1982). Nas aplicações em epidemiologia, este fenômeno é considerado de menor importância, o que pelo menos em parte se deve ao fato de o processo de captura pelas diferentes fontes não seguir uma ordenação no tempo.

A variabilidade de capturabilidade devido a heterogeneidades individuais assume que a probabilidade de captura possa variar de animal para animal em função de características tais como o acesso a armadilhas, a dominância social, o sexo e a idade (White et al., 1982), e desta maneira é análoga ao conceito de variabilidade de capturabilidade empregada nas aplicações epidemiológicas. Entretanto, nestas últimas o problema é mais complexo, pois na área da ecologia as armadilhas empregadas nos processos de captura e recaptura são iguais e os fatores que determinam a variabilidade de capturabilidade agem igualmente em todas as armadilhas. Ao contrário, nas aplicações epidemiológicas, a natureza das fontes é bastante distinta, fazendo com que os fatores determinantes da variabilidade de capturabilidade ajam de forma diferente em cada fonte (IWGDME, 1995a). Por exemplo, idosos com 80 anos ou mais apresentam um maior risco de hospitalização e óbito que os idosos mais jovens, mas pela dificuldade de locomoção costumam usar menos os serviços ambulatoriais (especialmente na população de baixa renda).

Por fim, o último tipo de variabilidade é decorrente de respostas comportamentais ao processo de captura anterior (“variabilidade de contágio”), ou seja, o comportamento do animal tende a variar após a primeira captura (White et al., 1982). Os exemplos anteriormente por nós utilizados para a apresentação dos conceitos de dependência positiva e negativa, em que alguns peixes aumentavam a probabilidade de serem capturados em uma segunda armadilha em função de terem se machucado durante a captura pela primeira armadilha, enquanto outros tinham esta probabilidade dimi-

nuída em função de “aprenderem” a evitar a recaptura, são tratados pelos ecologistas não como verdadeiras dependências entre as amostras, mas sim “aparentes dependências” introduzidas pela variabilidade de capturabilidade originada nas respostas comportamentais ao processo de captura.

É interessante observar que o fato de um paciente ser encaminhado para um determinado hospital ou laboratório após ser atendido por um médico e a diminuição do risco de hospitalização e óbito em função de um tratamento efetivo, também poderiam ser interpretados como “variabilidades de contágio” originadas pelo processo de captura anterior. Ou seja, os limites entre os conceitos de “verdadeira dependência” entre fontes e de “aparente dependência” determinada pela variabilidade de capturabilidade, são muito tênues, se é que existem.

Hook & Regal (1995) consideram que na maioria das vezes a dependência entre as fontes em epidemiologia se originaria na realidade na variabilidade de capturabilidade encontrada em numerosos pequenos subgrupos da população. Na situação limite cada estrato seria constituído por um único indivíduo da população, que em função de um conjunto de atributos, apresentaria uma determinada probabilidade de captura. Estes autores recomendam, todavia, que em função da dificuldade de tratar pequenos subgrupos isoladamente, na prática seria mais interessante considerar que qualquer viés do estimador \hat{N} (sendo os demais pressupostos atendidos) seria originado da dependência entre as fontes em estratos populacionais maiores.

No âmbito da epidemiologia, dois procedimentos são mais freqüentemente empregados para o tratamento da dependência entre as listas: o censo de Bernoulli e os modelos loglineares (Wittes et al., 1974; Hook et al., 1980; Hook & Regal, 1995; IWGDME, 1995a). Em ambos os casos se faz necessária a existência de pelo menos três fontes de dados. Nas situações em que só estejam disponíveis duas fontes de dados, o pesquisador poderá suspeitar da dependência entre as listas caso o valor do tamanho populacional estimado seja bastante discrepante daquele esperado. Uma evidência clara da presença de dependência positiva seria a obtenção de um número de casos estimados menor do que o número de casos diferentes observados nas duas fontes em conjunto.

A técnica do censo de Bernoulli envolve o cálculo de várias estimativas obtidas com todas as combinações possíveis entre as listas (Wittes et al., 1974; Hook et al., 1980). Os resul-

tados encontrados são então comparados. Se as diferentes estimativas não apresentarem grandes variações é provável que a premissa de independência não tenha sido violada. A dependência pode ser suspeitada entre as listas cujos valores estimados, com base nelas, se apresentam muito discrepantes dos demais. Neste caso a estimativa mais correta seria obtida pela combinação das listas supostamente dependentes, que passam a ser tratadas como uma única lista.

Apesar de simples e intuitiva, esta técnica tem como limitação o fato de não permitir explorar estruturas mais complexas de dependência (dependência simultânea entre conjuntos de listas diferentes, por exemplo). A estratégia de agrupar duas listas supostamente dependentes tende a aumentar a variância da estimativa populacional, sendo mais útil nos casos de forte dependência entre as listas (Wittes et al., 1974).

A outra abordagem implica a utilização de todas as 2^d listas em conjunto, sendo os dados classificados em categorias que constituirão as caselas de uma tabela de contingência que reproduz a história de captura/recaptura de todos os casos. Esta tabela apresentará um valor desconhecido (indivíduos não capturados por nenhuma das fontes de dados). Empregam-se, então, modelos loglineares, que são ajustados para a tabela incompleta observada. A seguir, seleciona-se um modelo segundo critérios de plausibilidade e simplicidade. Por fim, o modelo selecionado é empregado para a obtenção da estimativa do número de indivíduos não capturados por nenhuma das fontes (Fienberg, 1972). Na Figura 2 é apresentada uma tabela de contingência que reproduz um processo de captura-recaptura envolvendo três fontes ($d = 3$), onde os índices i, j, k tomam o valor 1 para os casos capturados, respectivamente pelas fontes A, B, C , e o valor 2 no caso contrário. A tabela apresenta 8 caselas (2^3), sendo que o efetivo n_{222} representa o número de indivíduos não capturados por nenhuma fonte, e que se deseja estimar.

A introdução de termos de interação permite que a dependência entre as fontes possa ser modelada. Ou seja, com a introdução de termos de interação podemos *ajustar* a estimativa do tamanho do número de casos não captados para a presença de dependência entre as fontes. No caso de três fontes poderiam ser ajustados oito modelos diferentes, indo da situação mais simples (considerando as três fontes independentes) até a situação mais complexa (modelo saturado), em que todas as fontes são dependentes duas a duas (Figura 2).

Quando os atributos que determinam a variabilidade de capturabilidade são conhecidos (sexo e idade, por exemplo), é possível incluir no modelo termos que representam o efeito destas variáveis assim como a interação das mesmas com as diferentes fontes (Fienberg, 1972). Esta estratégia foi utilizada por Hilsenbeck et al. (1992), McGilchrist et al. (1996) e Robles et al. (1987). Hook & Regal (1995) chamam a atenção para a ocorrência de alguns resultados anômalos quando da aplicação desta estratégia. Estes autores recomendam que o processo de modelagem neste caso deva ser precedido por uma análise exploratória inicial detalhada dos dados. A estratégia mais empregada quando o atributo que determina a variabilidade de capturabilidade é conhecido e a informação sobre o mesmo se encontra disponível, é a análise estratificada com o ajuste de modelos para cada estrato e a obtenção de uma estimativa global resultante da soma dos valores estimados em cada estrato.

A metodologia de captura-recaptura na vigilância de doenças na população idosa em nosso meio: o caso do diabetes mellitus

A utilização da metodologia de captura-recaptura para a implantação de sistemas rotineiros para a vigilância de doenças é certamente uma das aplicações mais interessantes da técnica, embora também seja a de mais difícil implementação. Para que um sistema possa ser implantado a um custo que viabilize a sua manutenção, é necessário que estejam disponíveis fontes de dados em saúde diversificadas que permitam cobrir diferentes segmentos da população. Estas fontes devem apresentar os seguintes atributos: possibilitar a captação e processamento ágil dos dados; contemplar informações sobre o nome e endereço de residência dos pacientes; e apresentar informações válidas.

A utilização da metodologia de captura-recaptura no âmbito da epidemiologia ainda é bastante recente, e a maioria dos estudos anteriormente citados aplicaram a metodologia para a avaliação da cobertura de sistemas de registros e refinamento das estimativas de incidência e prevalência de doenças em estudos pontuais. Uma das poucas experiências de aplicação da técnica para a vigilância rotineira de uma doença envolve o projeto mundial para avaliação do diabetes mellitus insulino-dependente (IWGDME, 1995b). Papoz et al. (1996) alertam para a dificuldade da utilização desta

Figura 2

Tabela de contingência que reproduz um processo de captura-recaptura envolvendo três listas e modelos loglineares que poderiam representar a tabela.

Capturado pela Lista A	Capturado pela Lista C		Capturado pela Lista B	
	Sim $k = 1$		Não $k = 2$	
	Sim $j = 1$	Não $j = 2$	Sim $j = 1$	Não $j = 2$
Sim $i = 1$	n_{111}	n_{121}	n_{112}	n_{122}
Não $i = 2$	n_{211}	n_{221}	n_{212}	$n_{222} (?)$

Modelo	Descrição
1) $\log n_{ijk} = u + u_i^A + u_j^B + u_k^C$	As três listas são independentes.
2) $\log n_{ijk} = u + u_i^A + u_j^B + u_k^C + u_{ij}^{AB}$	A lista C é independente das listas A e B.
3) $\log n_{ijk} = u + u_i^A + u_j^B + u_k^C + u_{jk}^{AC}$	A fonte B é independente das listas A e C.
4) $\log n_{ijk} = u + u_i^A + u_j^B + u_k^C + u_{jk}^{BC}$	A lista A é independente das listas B e C.
5) $\log n_{ijk} = u + u_i^A + u_j^B + u_k^C + u_{ij}^{AB} + u_{ik}^{AC}$	A lista A é associada as listas B e C.
6) $\log n_{ijk} = u + u_i^A + u_j^B + u_k^C + u_{ik}^{AC} + u_{jk}^{BC}$	A lista C é associada as listas A e B.
7) $\log n_{ijk} = u + u_i^A + u_j^B + u_k^C + u_{ij}^{AB} + u_{jk}^{BC}$	A lista B é associada as listas A e C.
8) $\log n_{ijk} = u + u_i^A + u_j^B + u_k^C + u_{ij}^{AB} + u_{ik}^{AC} + u_{jk}^{BC}$	As três listas são dependentes duas a duas. (Modelo saturado)

Onde,

$\log n_{ijk}$ = logaritmo da frequência da casela formada pelo cruzamento da categoria i da variável lista A com a categoria j da lista B e com a categoria k da variável lista C.

u é o termo geral (grande média); u_i^A , u_j^B , u_k^C são, respectivamente, os efeitos principais das variáveis lista A, B e C.

u_{ij}^{AB} , u_{ik}^{AC} , u_{jk}^{BC} são, respectivamente, os termos de interação de primeira ordem entre as variáveis duas a duas (listas).

metodologia em países em desenvolvimento em função da baixa disponibilidade de sistemas de informação em saúde.

Nos últimos anos pôde ser observado um crescimento da difusão da tecnologia de informática em nosso país, o que tornou possível o acesso ágil a bases de dados com informações variadas e desagregadas sobre mortalidade, internações hospitalares, registros de nascidos vivos, entre outras. Estes bancos de dados representam fontes importantes que podem ser empregadas rotineiramente na vigilância das doenças. Por exemplo, um sistema relativamente simples para a vigilância do diabetes mellitus em idosos do nosso meio poderia ser desenvolvido utilizando-se a combinação de três fontes de dados, a saber: o Sistema de Informação sobre Mortalidade (SIM/MS), o Sistema de Informações Hospitalares do Sistema Único de Saúde (SIH-SUS/DATASUS/FNS/MS) e as estatísticas ambulatoriais de atendimento ao diabetes mellitus em unidades de saúde credenciadas pelo SUS. Ao contrário do observado para as fontes de dados de óbito e de inter-

nações hospitalares, não existe um Sistema de Informações Ambulatoriais que inclua dados sobre morbidade, padronizado para todas as unidades credenciadas pelo SUS. Esta característica traz uma limitação operacional que torna impossível a avaliação de todas as unidades de saúde credenciadas pelo SUS que prestam atendimento ambulatorial. Uma forma de viabilizar a implementação do sistema seria a captação de dados junto às unidades que apresentam a maior probabilidade de captar um caso de diabetes mellitus em idosos. Unidades de saúde que são pólos de referência para o Programa Nacional de Diabetes Mellitus, serviços de endocrinologia e núcleos de atendimento ao idoso poderiam funcionar como unidades sentinelas para o sistema.

Uma das limitações do sistema acima proposto diz respeito a sua cobertura. Dentre as fontes de dados selecionadas apenas o SIM apresenta cobertura universal. As demais fontes só abrangem a população de idosos usuários do SUS, que é em sua quase totalidade representada pelas classes sociais de baixa ren-

da. As informações relativas à população usuária de planos de saúde e de serviços privados não se encontram disponíveis para fins de avaliação, uma vez que não existe até a presente data uma regulamentação pelos órgãos oficiais sobre esta questão. Uma forma de contornar este problema seria restringir a análise aos registros de óbitos que ocorreram em unidades de saúde credenciadas pelo SUS. Ainda que o conhecimento do perfil de morbidade da população de menor renda já seja de grande valia para o planejamento a curto prazo dos serviços públicos de saúde, a inacessibilidade às informações sobre as classes de maior renda, implica uma limitação do sistema de vigilância, que passa a ter seu escopo de abrangência reduzido.

Na Figura 3 apresentamos um esquema que de forma genérica busca representar as possíveis relações de dependência entre as fontes ambulatorial, hospitalar e de mortalidade, além das principais variáveis que poderiam determinar a variabilidade de capturabilidade nas diferentes fontes, no contexto de um sistema voltado para a vigilância do diabetes mellitus na população idosa. A gravidade do caso é uma variável central na relação entre as fontes, porque boa parte da dependência entre estas fontes é em função do grau de gravidade do diabetes mellitus. Outras variáveis importantes são o

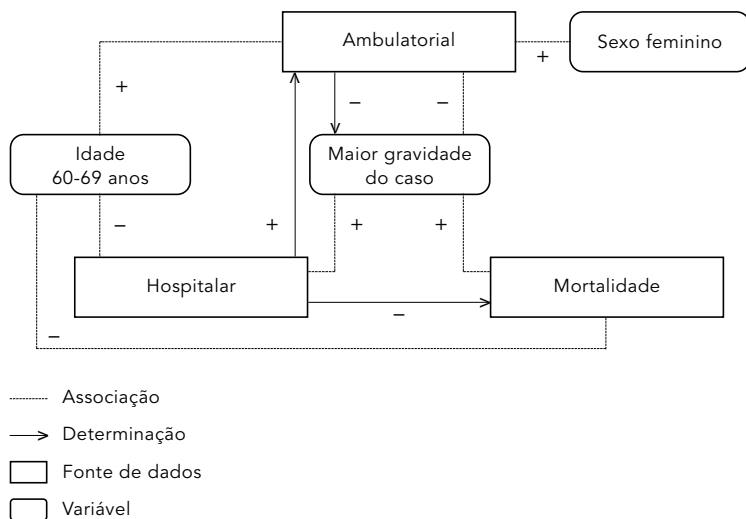
sexo e a idade, que podem influenciar o uso de serviços e/ou a mortalidade. Em uma primeira aproximação vamos simplificar o modelo, considerando que todo paciente portador de diabetes mellitus teria a doença diagnosticada e registrada em cada uma das fontes quando fosse captado.

A ocorrência de dependência negativa entre a fonte ambulatorial e as demais fontes poderia se dar em função de quatro fatores: a) o tratamento efetivo evitando a ocorrência de complicações agudas e crônicas (relação de determinação); b) dependendo do tipo de unidade ambulatorial incluída (ambulatórios especializados/unidades básicas), a maior proporção dos casos atendidos no ambulatório poderia representar casos recém-diagnosticados que apresentam uma menor ocorrência de complicações agudas e crônicas (relação de associação); c) idosos jovens (60 a 69 anos) usam mais o ambulatório em função da maior facilidade para locomoção e ao mesmo tempo têm uma menor probabilidade de serem hospitalizados ou de virem a falecer, já que costumam apresentar um menor número de co-morbidades, incluindo as complicações crônicas do diabetes (relação de associação com sinal positivo para a fonte ambulatorial e negativo para as demais fontes, levando a dependência negativa entre a fonte ambulatorial e as demais fontes); d) a maior utilização de serviços ambulatoriais por pacientes do sexo feminino. Desta maneira, poderia ser esperada uma menor probabilidade de internação e mortalidade neste grupo em função do diagnóstico precoce e do tratamento efetivo do diabetes mellitus. Da mesma forma que, para a idade, observaríamos uma associação positiva entre o sexo feminino e a fonte ambulatorial e negativa com as fontes hospitalar e de mortalidade, levando a ocorrência de uma dependência negativa entre a fonte ambulatorial e as demais fontes.

Apesar de alguns estudos terem evidenciado uma taxa de hospitalização maior nos idosos em relação às idosas (Travassos-Veras, 1992), o mesmo ocorrendo com a taxa de mortalidade (Yazaki & Saad, 1990), outros estudos realizados em populações idosas diabéticas não demonstraram os diferenciais acima descritos para mortalidade (Panzram, 1987). Sendo assim, optamos por não incluir a relação direta entre sexo e as fontes hospitalar e de mortalidade, que, uma vez presentes, contribuiriam para acentuar a dependência negativa entre estas fontes e a fonte ambulatorial. É interessante observar que os pacientes que sofreram internação hospitalar são usualmente encaminhados para o acompanhamento ambulatorial

Figura 3

Modelo representando as relações de dependência entre fontes e as variáveis associadas à variabilidade de capturabilidade no contexto de um sistema de vigilância do diabetes mellitus em idosos.



após a alta, o que poderia minimizar a dependência negativa entre as fontes ambulatorial e hospitalar determinada pelos fatores acima descritos.

Já as fontes hospitalar e de mortalidade tenderiam a apresentar uma dependência positiva. Apesar de a internação hospitalar provavelmente contribuir para uma menor mortalidade dentro do conjunto dos casos com igual gravidade (relação de determinação), é possível que, globalmente, a relação que talvez se manifeste é a de dependência positiva, pois os casos que necessitam de internação e que evoluem para óbito são os de maior gravidade. Conforme discutido acima, as fontes hospitalar e de mortalidade apresentariam associações com igual direção tanto para idade quanto para o sexo, o que contribuiria para ocorrência de dependência positiva entre elas.

O modelo acima apresentado é bastante simplificado e não considera outras variáveis tais como o tipo de tratamento (dieta, hipoglicemiantes orais ou insulina). Dentre as variáveis não consideradas, a variabilidade no regis-

tro do diagnóstico é um fator que pode provocar alterações importantes nas relações anteriormente apresentadas. Esta variabilidade poderia impedir a identificação de casos recapturados, levando a uma diminuição do número de casos nas interseções entre as fontes, e, conseqüentemente, provocando uma atenuação da dependência positiva entre as fontes hospitalar e de mortalidade, além de uma acentuação da dependência negativa entre a fonte ambulatorial e demais fontes.

Somadas aos problemas mencionados, as dificuldades no relacionamento dos registros nas diferentes bases, a dificuldade da identificação do local de residência dos idosos e a validade do diagnóstico são fatores potencialmente limitantes para a implantação de sistemas de vigilância baseados na metodologia de captura-recaptura. A realização de estudos que busquem avaliar estas questões contribuirão para avaliar o real benefício do uso da metodologia de captura-recaptura para a implantação de sistemas de vigilância de doenças na população idosa de nosso meio.

Referências

- CAMARGO Jr., K. R. & COELI, C. M., 1998. RecLink: Aplicativo para o relacionamento de banco de dados implementando o método *probabilistic record linkage*. In: IV Congresso Brasileiro de Epidemiologia, *Resumos*, p. 162. Rio de Janeiro: ABRASCO.
- FIENBERG, S. E., 1972. The multiple recapture census for closed populations and incomplete $2k$ contingency tables. *Biometrika*, 59:591-603.
- GOLDBERG, J.; GELFAND, H. M. & LEVY, P. S., 1980. Registry evaluation methods: A review and case study. *Epidemiologic Reviews*, 2:210-220.
- GREEN, A.; GALE, E. A. M. & PATTERSON, C. C., 1992. Incidence of childhood-onset insulin-dependent diabetes mellitus: The Eurodiab Ace Study Group. *Lancet*, 339:905-909.
- HILSENBECK, S. G.; KURUCZ, C. & DUNCAN, R. C., 1992. Estimation of completeness and adjustment of age-specific and age-standardized incidence rates. *Biometrics*, 48:1249-1262.
- HOOK, E. B.; ALBRIGHT, S. G. & CROSS, P. K., 1980. Use of Bernoulli census and log-linear methods for estimating the prevalence of spina bifida in livebirths and the completeness of vital record reports in New York State. *American Journal of Epidemiology*, 112:750-758.

- HOOK, E. B. & REAGAL, R. R., 1993. Effect of variation in probability of ascertainment by sources ("variable catchability") upon "capture-recapture" estimates of prevalence. *American Journal of Epidemiology*, 137:1148-1166.
- HOOK, E. B. & REAGAL, R. R., 1995. Capture-recapture methods in epidemiology: Methods and limitations. *Epidemiologic Reviews*, 17:243-264.
- IWGDMF (International Working Group for Disease Monitoring and Forecasting), 1995a. Capture-recapture and multiple-record systems estimation. I: History and theoretical development. *American Journal of Epidemiology*, 142:1047-1058.
- IWGDMF (International Working Group for Disease Monitoring and Forecasting), 1995b. Capture-recapture and multiple-record systems estimation. II: Applications in human diseases. *American Journal of Epidemiology*, 142:1059-1068.
- LaPORTE, R. E., 1993. How to improve monitoring and forecasting of disease patterns. *BMJ*, 307: 1573-1574.
- LaPORTE, R. E., 1994. Assessing the human condition: Capture-recapture techniques. *BMJ*, 308:5-6.
- LaPORTE, R. E.; DEARWATER, S. R.; CHANG, Y.; AARON, D. J. & ANDERSON, R. L., 1995. Efficiency and accuracy of disease monitoring systems: Application of capture-recapture methods to injury monitoring. *American Journal of Epidemiology*, 142:1069-1077.
- LaPORTE, R. E.; McCARTY, D. J. & TULL, E. S., 1992. Counting birds, bees, and NCDs. *Lancet*, 339:494-495.
- McGILCHRIST, C. A.; McDONNELL, L. F.; JORM, L. R. & PATEL, M. S., 1996. Loglinear models using capture-recapture methods to estimate the size of a measles epidemic. *Journal of Clinical Epidemiology*, 49:293-296.
- PANZRAM, G., 1987. Mortality and survival in type 2 (non-insulin-dependent) diabetes mellitus. *Diabetologia*, 30:123-131.
- PAPOZ, L.; BALKAU, B. & LELLOUCH, L., 1996. Case counting in epidemiology: Limitations of methods based on multiple data sources. *International Journal of Epidemiology*, 25:474-478.
- ROBLES, S.; MARRETT, L. D.; CLARKE, E. A. & RISCH, H. A., 1987. An application of capture-recapture methods to the estimating of completeness of cancer registration. *Journal of Clinical Epidemiology*, 4:495-501.
- TRAVASSOS-VERAS, C. M., 1992. *Equity in the Use of Private Hospitals Contracted by a Compulsory Insurance Scheme in the City of Rio de Janeiro, Brazil, in 1986*. Ph. D. Thesis, London: Department of Public Administration, School of Economics and Political Science.
- VERAS, R. P. & ALVES, M. I. C., 1995. A população idosa no Brasil: Considerações acerca do uso de indicadores de saúde. In: *Os Muitos Brasis: Saúde e População na Década de 80* (M. C. S. Minayo, org.), pp. 320-337, São Paulo: Editora Hucitec/Rio de Janeiro: ABRASCO.
- WHITE, G. C.; ANDERSON, D. R.; BURTHAM, K. P. & OTIS, D. L., 1982. *Capture-Recapture and Removal Methods for Sampling Closed Populations*. Los Alamos: Los Alamos National Laboratory.
- WHO (World Health Organization), 1993a. Directors of non-communicable disease collaborating Centers and Key Officials. Shanghai declaration on non-communicable diseases. *BMJ*, 306:588.
- WHO (World Health Organization), 1993b. Directors of World Health Organization non-communicable disease collaborating Centers and Key Officials, 1993 needed: Universal monitoring of all serious diseases of global importance. *American Journal of Public Health*, 83:941-943.
- WITTES, J.; COLTON, T. & SIDEL, V. W., 1974. Capture-recapture methods for assessing the completeness of case ascertainment when using multiple information sources. *Journal of Chronic Diseases*, 27:25-36.
- WITTES, J. & SIDEL, V. W., 1968. A generalization of the simple capture-recapture model with applications to epidemiological research. *Journal of Chronic Diseases*, 21:287-301.
- YAZAKI, L. M. & SAAD, P. M., 1990. Mortalidade na população idosa. In: *O Idoso na Grande São Paulo* (Fundação Sistema Estadual de Análise de Dados, org.), pp. 125-159, São Paulo: Fundação Sistema Estadual de Análise de Dados.