

Método de captura e recaptura: nova metodologia para pesquisas epidemiológicas

Capture-recapture: a new methodology for epidemiological research

John Dunn*, Sérgio Baxter Andreoli**

DUNN, J. & ANDREOLI, S. B. Método de captura e recaptura: nova metodologia para pesquisas epidemiológicas. *Rev. Saúde Pública*, 28:449-53, 1994. O método de captura e recaptura, embora não seja novo, tem sido pouco usado em estudos epidemiológicos. Trata-se de método bem adaptável e adequado para estudar populações incomuns ou esquivas, como usuários de drogas endovenosas. Tem sido usado para estudar populações diversas como prostitutas que trabalham na rua ou volume de células vermelhas no homem. Permite ainda estimar a incidência e a prevalência de doenças de forma mais precisa do que os métodos tradicionais e com uma melhor relação custo-benefício. Devido à sua relevância dentro do campo da epidemiologia, decidiu-se realizar uma revisão sobre esse método, enfocando a história, as principais aplicações e apontando as suposições teóricas que o fundamentam. Seu potencial para futuras pesquisas epidemiológicas é promissor.

Descritores: Métodos epidemiológicos. Incidência. Prevalência.

Introdução

Em recente editorial, de LaPorte¹⁹ (1994) considerou que conhecemos os mais sobre o número de águias, baleias e bisões do que sobre o número de pessoas desempregadas, doentes ou com fome em nossa sociedade. Isto porque, ecologistas que trabalham com animais têm desenvolvido e refinado técnicas para estimar o número dessas populações sem se utilizar das técnicas de contagem direta. Diferentemente, os epidemiologistas e cientistas que estudam populações, utilizam cada vez mais registros de caso, banco de dados e censos populacionais, como se qualquer método que não envolva a enumeração da população fosse imperfeito. O método de captura e recaptura é uma técnica usada em situações onde não existe qualquer possibilidade de contagem direta, como por exemplo em estudos no campo da biologia, e tem mostrado que pode resultar em estimativas acuradas do número de populações, sendo bem adaptável e adequado para estudar populações incomuns, esquivas ou de difícil acesso (Sudman e col²⁹.,

1988), tais como pacientes com doença de Huntington's (Hook¹², 1992), doenças do tecido conjuntivo (McCarty²⁰, 1992) ou usuário de drogas endovenosas (Frischer e col⁸., 1991).

Devido a relevância do citado método, dentro do campo da epidemiologia, o presente trabalho apresenta uma revisão sobre o método de captura e recaptura, enfocando a história, as principais aplicações e apontando as suposições teóricas que o fundamentam.

O histórico e o desenvolvimento do método

O método de captura e recaptura foi inventado por La Place¹⁷ (1783) para estimar a população da França, e foi recentemente usado para ajustar o censo dos Estados Unidos, de 1990 (Wolter³⁴, 1991). O método tem sido desenvolvido e refinado por ecologistas trabalhando com animais para estimar e monitorizar o tamanho das populações de várias espécies (Seber²⁶, 1982), especialmente de populações fechadas, ou seja, populações que não têm migração (Seber²⁶, 1982; Pollock²⁴, 1991). Por exemplo, para se estimar o número total de cobras na ilha das Cobras poder-se-ia capturar uma amostra de cobras (*A*), marcá-las e soltá-las para que se misturem com o resto da população. Depois de alguns dias poder-se-ia capturar uma outra amostra (*B*) e anotar o número de cobras marcadas (*m*). Se todas as cobras tiverem a mesma probabilidade de serem capturadas, então o número de cobras marcadas na amostra *B* estará relacionado ao número total de

* Pesquisador visitante do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) no Departamento de Psiquiatria e Psicologia Médica da Escola Paulista de Medicina - São Paulo, SP - Brasil.

** Departamento de Psiquiatria e Psicologia Médica da Escola Paulista de Medicina - São Paulo, SP - Brasil, Separatas/Reprints: S. B. Andreoli - Rua Botucatu, 740 - 3º andar - 04023-900 - São Paulo, SP - Brasil
Edição subvencionada pela FAPESP. Processo 94/0500-0.

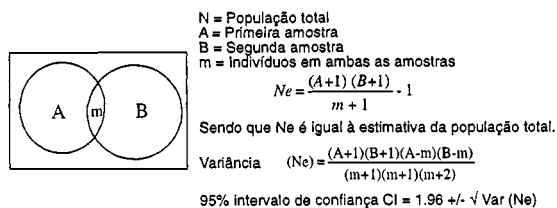


Figura 1. Superposição de casos em duas amostras.

cobras na ilha (N) e, usando uma fórmula simples poder-se-ia obter uma estimativa deste número (Ne) e calcular o intervalo de confiança desta estimativa (La Porte e col.¹⁸, 1992).

Um outro exemplo para uma situação clínica (Tabela), seria a estimativa da população de pacientes esquizofrênicos em uma determinada comunidade através da coleta de duas amostras, uma conseguida através do registro de internação em hospitais psiquiátricos (A) e uma segunda amostra (B) obtida através de informação direta dos pacientes ou familiares depois de veicular propaganda, para este fim, nos meios de comunicação escrita e falada. O fato das amostras não serem compostas por todos os indivíduos atingidos não preocupa, uma vez que as estimativas são feitas com base no número de pacientes que aparecem nas duas amostras, ou seja, a estimativa é feita sobre o grau de concomitância dos casos (McCarty e col.²¹, 1993).

Para comparar a prevalência e a incidência de uma doença de um local com outro, é essencial que se saiba com exatidão o número de casos ou casos novos respectivamente. Tradicionalmente, tem sido usado para este fim os registros de casos, as listas de altas hospitalares e outras formas de registros onde são contados o número de indivíduos registrados em um período determinado. Nos países mais organizados onde existem serviços de coletas de casos de dados avançados, tais como na Suíça, estudos mostram que esses métodos podem subestimar a verdadeira incidência em 20% (La

Tabela. Pacientes com determinada condição distribuídos pelas fontes de identificação.

		casos identificados na fonte A	
		sim	não
Casos identificados na fonte B	sim	a	b
	não	c	d=?

OBS: A estimativa de $d=bc/a$, sendo d a estimativa do número de casos que não foram achados em nenhuma das fontes. A população total N (com uma correção para amostras pequenas) é:

$$N = a + b + c + \frac{bc}{a + 1}$$

Porte e col.¹⁸, 1992). Por este motivo, as comparações feitas entre Estados ou Países, ao longo do tempo tornam-se imprecisas devido a este alto nível de subestimação. O método de captura e recaptura pode elucidar o grau de subestimação.

Um exemplo desta situação foi a reanálise da prevalência da doença de Huntington's realizado no Estado de Maryland, USA (Hook¹², 1992). Este estudo procedeu uma pesquisa ampla utilizando 14 fontes diferentes de informação (registro de casos, altas hospitalares, casas de repouso, investigações genealógicas, noticiários em jornais, e outras) numa tentativa de identificar todos os pacientes com sintomas de doença de Huntington's, em abril de 1980. Uma soma simples de casos não duplicados mostrou uma prevalência de 217 casos no Estado. Na reanálise dos dados, utilizando o método de captura e recaptura, foi encontrada uma nova estimativa, mostrando que com um intervalo de confiança de 90%, cerca de 24% a 41% de casos não tinham sido identificados na metodologia anterior e que a prevalência estava entre 284 a 366 casos.

O método de captura e recaptura, além de prover estimativas mais corretas de incidência e prevalência das doenças, tem outras vantagens: consome menos tempo e é mais barato. Isto porque não existe necessidade de contar todos os casos de uma determinada condição dentro de uma população. A experiência do Departamento do Censo dos EUA mostrou que o aumento dos gastos não eliminou a subestimação dos resultados e, em 1990, o referido Departamento admitiu que esta contagem da população não era completa, e introduziu a metodologia de captura e recaptura para calcular o nível de subestimação (Wolter³⁴, 1991).

Aplicação do método de captura e recaptura

Estudos populacionais

Em alguns países o censo populacional, registros de nascimentos e de óbitos conseguem contar a maioria da população, mas não são completos. O método de captura e recaptura, por suas características, permite fazer estimativas mais precisas da população (Wolter³⁴, 1991; Sekar e Deming²⁷, 1949).

Incidência e prevalência de doenças comuns ou raras

As estimativas de prevalência e incidência de doenças têm derivado da combinação de in-

formações de várias fontes onde se somam os casos não duplicados. Quando a condição estudada tem alta prevalência na comunidade, a estimativa de prevalência é feita através de uma amostra aleatória (Almeida-Filho e col.¹, 1992). A menos que a condição tenha realmente uma alta prevalência, o tamanho da amostra aleatória precisa ser proibitivamente grande o que torna o estudo caro, demorado e ineficaz por causa da proporção pequena de casos em relação aos não-casos.

O método de captura e recaptura permite fazer estimativas corretas de incidência e prevalência, mesmo que sejam usados dados provenientes de fontes incompletas. Pode ser usado, portanto, em situações onde não é possível conhecer o número exato de casos. Uma grande variedade de doença tem sido estudada utilizando este método, por exemplo: estimativa da fidedignidade dos registros de casos de câncer (Robles e col.²⁵, 1988), a incidência de infarto do miocárdio (LaPorte e col.¹⁸, 1992), a prevalência de espinha bífida (Hook e col.¹¹, 1980) e "lábio leporino" (Chapman³, 1983), incidência de doenças do tecido conjuntivo (McCarty²⁰, 1992), a incidência de diabetes em crianças (WHO³⁰, 1990) e a incidência de rubéola congênita (Cochi e col.⁴, 1989).

Populações esquivas e difíceis de encontrar

Algumas populações são difíceis de estudar pois não são encontradas nos serviços hospitalares, ou têm altas percentagens de migração, ou o comportamento a ser estudado é estigmatizado pela sociedade ou ilegal. Nestas, o método captura e recaptura tem sido usado com sucesso, por exemplo: o tamanho da população dos sem tetos em Londres (Fisher e col.⁷, 1994); a prevalência de usuário de drogas endovenosas em Glasgow, Escócia (Frischer e col.⁸, 1991); o número de prostitutas que trabalham na rua, também em Glasgow (McKeganey e col.²², 1992); e a prevalência de infecção HIV no Bronx, Nova Iorque (Drucker e Vermund⁶, 1989).

Os usuários de drogas injetáveis constituem um exemplo de uma população difícil de encontrar devido ao caráter ilegal de suas atividades. Um dos problemas encontrados nos levantamentos populacionais é a omissão de informação para a elaboração do diagnóstico, isto porque, o indivíduo nesta situação teme ser identificado e, com isso, sofrer algum tipo de punição devido a característica ilegal desta atividade. Nesta situação, o método de captura e recaptura pode ser útil, uma vez que ele permite a utilização de fontes de casos onde o problema da identificação não exista. Frischer

e col.^{8,9} (1991, 1992) estudaram a prevalência de usuários de drogas injetáveis, utilizando três fontes: serviços de tratamento de drogados, registro de teste de HIV na Escócia e registro de pessoas apreendidas pela polícia por uso de drogas. Para manter o sigilo no estudo, os detalhes de identificação dos pacientes não foram registrados, mas foi construído um algoritmo com as iniciais do nome do paciente, data do nascimento, sexo e, também, a primeira parte do número do código de endereçamento postal.

Técnicas de investigação clínica

O método de captura e recaptura pode ser utilizado para estimar o volume e a sobrevivência de células vermelhas usando o rádio-isotopo para marcá-las (Internacional Committee for Standardization in Hematology¹⁴, 1980) e, também, pode ser usado para estimar o número de falsos negativos em procedimentos de rastreamento médico (Goldberg e Wittes¹⁰, 1978). Mais recente foi usado para analisar proteínas e DNA (Morrissey e col.²³, 1989; Sidow e col.²⁸, 1992).

Suposições Teóricas

A mesma probabilidade de serem capturados

A suposição mais básica do método é que todos os membros da população tenham a mesma chance de serem capturados ("equal catchability"). Mas, freqüentemente, isto não acontece, por exemplo, pessoas de renda baixa têm menor chance de serem capturadas numa amostra de pacientes de um serviço de saúde particular. No caso de um grupo de pacientes com menor chance de ser capturado, em todas as fontes, é possível calcular N_e (veja Figura 1) para cada grupo e, depois somar estas estimativas (Hook e Regal¹³, 1993). Quando um grupo tem menor chance de ser capturado numa fonte, mas maior chance de ser capturado por outra, pode-se combinar as duas fontes para criar uma. Por exemplo, se uma fonte fosse de pacientes de um serviço de saúde particular e a outra de pacientes de um serviço público, a combinação das duas criaria apenas uma fonte de 'saúde'.

Dependência das fontes

No modelo simples, as fontes são geralmente independentes, mas as fontes utilizadas nas pesquisas clínicas mostram freqüentemente algum grau de dependência positiva. Isto ocorre

quando um caso encontrado na fonte A tem chance maior de ser capturado pela fonte B do que um caso que não tenha sido capturado pela fonte A. Por exemplo, um paciente que aparece num registro de notificação compulsória de casos deveria também aparecer numa lista de altas hospitalares. A dependência positiva resulta num grau maior de intersecção (m é maior, veja Figura 1), e pode levar a uma subestimativa da população. Existem casos, entretanto, de dependência negativa das fontes. Isto ocorre quando um caso capturado pela fonte A tem uma chance grande de não ser capturado pela fonte B. Por exemplo, um paciente que frequenta um ambulatório particular tem chance muito menor de frequentar um outro ambulatório particular por causa da mesma queixa. Dependência negativa resultam num grau menor de intersecção (m é menor, veja Figura 1) e, pode levar a uma superestimativa (Hook¹², 1992; Wittes e col.³¹, 1974).

O grau de dependência das fontes pode ser medido através da análise log-linear (Cormack⁵, 1989; Frischer e Leyland⁹, 1992), mas somente nos casos em que existem mais do que duas fontes. Nos casos de duas fontes apenas, não é possível determinar o grau de dependência entre elas.

Fontes múltiplas

Na sua forma simples, o método de captura e recaptura é utilizado com duas fontes, representadas por duas amostras consecutivas. Na pesquisa clínica esta situação é rara (Fig. 2). Por exemplo, no estudo de prevalência da doença de Huntington's foram utilizadas 14 fontes diferentes. Wittes e Sidel³³ (1968) demonstraram matematicamente que é possível utilizar este método para fontes múltiplas e testar a eficiência de cada fonte, calculando as estimativas de máxima

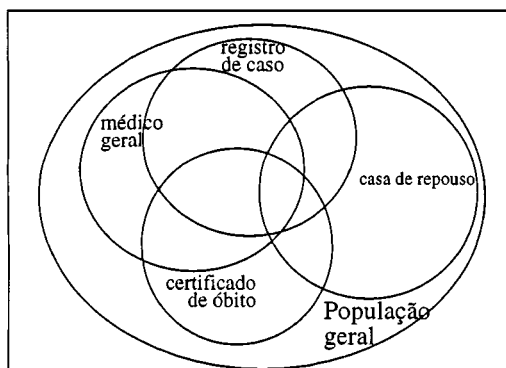


Figura 2. Superposição de casos em fontes múltiplas.

*maximum likelihood

verossimilhança*. Estas estimativas são usadas posteriormente no cálculo da população total e das suas variâncias.

Erros devidos a amostras pequenas

Retornando para o modelo de duas amostras, se a amostra A tem um tamanho n_1 , a amostra B um tamanho n_2 e o tamanho da população total é N , então a precisão da estimativa da população total (Ne) depende da relação entre estes tamanhos de amostras. Wittes³² (1972) demonstrou que se n_1 e n_2 são muito pequenos em relação ao N , então a estimativa Ne tem um nível de erro grande e inaceitável. Quando os tamanhos das amostras são aumentados, Ne começa a ser mais próximo do verdadeiro valor de N . Wittes³² (1972) mostrou o seguinte exemplo, se $N = 1.000$, então $n_1 + n_2$ precisa ser pelo menos igual a 50 para se obter uma estimativa Ne próxima de N . Na clínica, utilizando múltiplas fontes, sem dependência positiva entre si, o somatório das amostras $n_1 + n_2 + n_3 \dots$ etc pode ser igual ou maior que N e nesta situação, a estimativa Ne será próxima de N .

Precisão diagnóstica

Quando se utiliza várias fontes de casos, a questão da precisão do diagnóstico começa a ganhar muita importância. No estudo de prevalência da doença de Huntington's (Hook¹², 1992), todos os casos de todas as fontes foram entrevistados e examinados por um médico. Uma alternativa seria rever os diagnósticos através de uma amostra randômica de casos, estratificando também as fontes. Deste modo, a precisão diagnóstica poderia ser obtida sem o exame de todos os casos.

Sistemas abertos e programas de computador

Algumas populações não são fechadas, e são sujeitas a um alto grau de migração, por exemplo, prostitutas que trabalham na rua (Bloor e col.², 1991). Então, adicionais adaptações ao método têm sido feitas para atender a estas condições (Jolly^{15,16}, 1965, 1982). Estas análises podem ser realizadas através de programas de computador especialmente desenhados, tais como o GLIM e o POPAN (Cormack⁵, 1989). Os pacotes estatísticos que incluem a análise log-linear podem ser adaptados para analisar dados provenientes de estudos que utilizam o método de captura e recaptura.

DUNN, J. & ANDREOLLI, S. B. [Capture-recapture method: a new methodology for epidemiological research]. *Rev. Saúde Pública*, 28:449-53, 1994. Although capture-recapture method is not new, their use in epidemiological studies has so far been limited. The method was developed by animal ecologists to estimate the size of various species populations, but in recent years it has been used to measure the incidence and prevalence of a variety of non-communicable diseases. Capture-recapture allows more accurate estimates to be made than is the case with traditional methods and is more cost-effective. The method is ideally suited to studying rare or elusive populations, such as intravenous drug misusers; but they are highly adaptable and have been used to study populations as diverse as street working prostitutes and the human red-cell volume. The objectives of this review are to explain the mathematical theory behind capture-recapture, give examples of its use and application in clinical epidemiology and outline some of the modifications that have been made to the theory, which take into account the complex nature of samples used in clinical research. An increasing number of medical epidemiologists have started to use the method, and its potential for future epidemiological research is enormous.

Keywords: Epidemiologic methods. Incidence. Prevalence.

Referências Bibliográficas

- ALMEIDA-FILHO, N. et al. Estudo multicêntrico de morbidade psiquiátrica em áreas urbanas brasileiras (Brasília, São Paulo, Porto Alegre). *Rev. ABP-APAL*, 14:93-104, 1992.
- BLOOR, M. et al. Estimating hidden populations: a new method of calculating the prevalence of drug-injecting and non-injecting female street-working prostitutes. *Br. J. Addiction*, 86:1477-83, 1991.
- CHAPMAN, C. J. Ethnic differences in the incidence of cleft lip and/or cleft palate in Auckland, 1960-1976. *N. Z. Med. J.*, 96:327-9, 1983.
- COCHI, S. L. et al. Congenital rubella syndrome in the United States, 1970-85. *Am. J. Epidemiol.*, 129: 349-61, 1989.
- CORMACK, R. M. Log-linear models for capture-recapture. *Biometrics*, 45:395-413, 1989.
- DRUCKER, E. & VERMUND, S. H. Estimating population prevalence of human immunodeficiency virus infection in urban areas with high rates of intravenous drug use: a model of the Bronx in 1988. *Am. J. Epidemiol.*, 130:133-42, 1989.
- FISHER, N. et al. Estimating numbers of homeless and homeless mentally ill people in north east Westminster by using capture-recapture analysis. *BMJ*, 308:27-30, 1994.
- FRISCHER, M. et al. A new method of estimating prevalence of injecting drug use in an urban population: results from a Scottish city. *Int. J. Epidemiol.*, 20:997-1000, 1991.
- FRISCHER, M. & LEYLAND, A. Reliability of population and prevalence estimates. *Lancet*, 339:995, 1992.
- GOLDBERG, J. D. & WITTES, J. T. The estimation of false negatives in medical screening. *Biometrics*, 34:77-86, 1978.
- HOOK, E. B. et al. 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. *Am. J. Epidemiol.*, 112:750-8, 1980.
- HOOK, E. B. The value of capture-recapture methods for even apparent exhaustive surveys. *Am. J. Epidemiol.*, 135:1060-7, 1992.
- INTERNATIONAL COMMITTEE FOR STANDARDIZATION IN HEMATOLOGY. Recommended method for radioisotope red-cell survival studies. *Br. J. Hematol.*, 45:659-66, 1980.
- JOLLY, G. M. Explicit estimates from capture-recapture data with both death and immigration - stochastic model. *Biometrika*, 52:225-47, 1965.
- JOLLY, G. M. Mark-recapture models with parameters constant in time. *Biometrics*, 38:301-21, 1982.
- LAPLACE, P. S. Sur les naissances, les mariages, et les morts. In: *Histoire de l'Académie Royale des Sciences*. Paris, Impr. Royale, 1783. p. 693.
- LAPORTE, R. E. et al. Monitoring the incidence of myocardial infarction: applications of capture-recapture technology. *Int. J. Epidemiol.*, 21:258-62, 1992.
- LAPORTE, R. E. Assessing the human condition: capture-recapture techniques. *BMJ*, 308:5-6, 1994.
- McCARTY, D. J. THE importance of incidence registries for connective tissue diseases. *J. Rheumatol.*, 19:1-6, 1992.
- McCARTY, D. J. ET al. Ascertainment corrected rates: applications of capture-recapture methods. *Int. J. Epidemiol.*, 22:559-65, 1993.
- McKEGANEY, N. et al. Female street-working prostitutes and HIV infection in Glasgow. *BMJ*, 305:801-4, 1992.
- POLLOCK, K. H. Modeling capture, recapture, and removal statistics for fish and wild life populations: past, present and future. *J. Am. Stat. Assoc.*, 86:225-38, 1991.
- ROBLES, S. C. et al. An application of capture-recapture methods to the estimation of completeness of cancer registration. *J. Clin. Epidemiol.*, 41:495-501, 1988.
- SEBER, G. A. F. *The estimation of animal abundance and related parameters*. 2nd ed. London, Charles Griffin, 1982.
- SEKAR, C. C. & DEMING, W. E. On a method of estimating birth and death rates and the extent of registration. *Am. Stat. Assoc. J.*, 44:101-15, 1949.
- SUDMAN, S. et al. Sampling rare and elusive populations. *Science*, 240: 991-6, 1988.
- WHO DIAMOND PROJECT GROUP. Who multinational project for childhood diabetes. *Diabetes Care*, 13:1062-8, 1990.
- WITTES, J. et al. Capture-recapture methods for assessing the completeness of case ascertainment when using multiple information sources. *J. Chron. Dis.*, 27:25-36, 1974.
- WITTES, J. On the bias and estimated variance of Chapman's two-sample capture-recapture population estimate. *Biometrics*, 28:592-7, 1972.
- WITTES, J. & SIDEL, V. W. A generalization of the simple capture-recapture model with applications to epidemiological research. *J. Chron. Dis.*, 21:287-301, 1968.
- WOLTER, K. M. Accounting for America's uncounted and miscounted. *Science*, 253:12-5, 1991.

Recebido para publicação em 17.3.1994
Aprovado para publicação em 18.10.1994