

Technics Guide

## Capture-mark-recapture法

—インスリン依存型糖尿病発症率への応用—

東京慈恵会医科大学第3内科・講師

田 嶋 尚 子

Professor, Department of Epidemiology, Graduate School of Public  
Health, University of Pittsburg

RE LaPorte

兵庫県立成人病センター・総長

馬 場 茂 明

**Diabetes Journal** (糖尿病と代謝) 1992年6月1日発行 第20巻, 第2号

# Capture-mark-recapture 法

## —インスリン依存型糖尿病発症率への応用—

Capture-mark-recapture method ; application for IDDM incidence rate

田嶋 尚子\*, RE LaPorte\*\*, 馬場 茂明\*\*\*

### はじめに

ある地域で一定の期間に、心筋梗塞の患者や乳癌の患者がどのくらい発生しているのか、私たちは正確に把握しているだろうか。糖尿病についても同じである。インスリン依存型糖尿病(IDDM)の新規に発生した患者の登録制度が確立しているところは別として、その他の地域では、およその数を把握するにすぎない。ところが、生態学者はとうの昔から、蝶の数やバッタの数を正確に把握していた。名前もなく、住所不定で、顔つきからも識別できない昆虫の数を、どのような方法によって数えたのだろうか。捕らえ所のない“昆虫”を数えることができ、住所・氏名の明らかな“糖尿病の人”を数えられない理由は何なのか。それは、生態学者はcapture-mark-recapture method(C-M-R: 標識再捕法)<sup>1-7)</sup>を知っており、私たちは知らなかったからである。

C-M-R法とは昆虫のみならず、クジラ、スイギュウ、ツルなどあらゆる動物を数えるときにひろく用いられてきた方法である。この方法は、いろいろな疾患の頻度を知るための手段としても、すでに一部で用いられていたが<sup>8-10)</sup>、近年、IDDMなど慢性疾患の真の発症率を推定するために応用しようという考えが盛んになってきた<sup>11-13)</sup>。

そこで本稿では、C-M-Rとはどのような方法なのか、疾患の発症率を把握するためには具体的にどうするかについて解説する。

### I. 個体数を推定する方法 (population estimation)

個体群の大きさをはかる最も確実な方法は、ある区域内に存在する個体すべてを数え上げる全数調査(complete census)である。しかしこの方法は、ひろい区域を対象としたときや、穴にもぐって見付けにくい動物などを対象としたときには労力的に実行不可能であり、実際的な方法ではない。したがって一般的には、これに代わるものとして、対象個体群のごく一部を標本として抽出調査し、その情報をもとに全体を推定するという方法がとられる。これをサンプリング調査(sampling census)という。

サンプリング調査の1例が区画法である。区画法では、調べようとする場所の全域を方形の小区画に分割し、そのうちの何区画かについて実測して全体を推定する。これに対して、草むらや林の中、あるいは水中などみえにくい環境にいる動きの早い動物の個体数の推定として、もっぱら用いられてきたのがC-M-R法である。

### II. Capture-mark-recapture 法

#### 1. Petersen-Lincoln 法

ノネズミの個体数の推定を例にとって説明する(図1)。まず調べようとする場所の各所に罠を仕掛けてノネズミを捕まえる。その数(M)を数え、個体識別標識を施し元の場所に戻す。何日かあとに同じ場所に罠を仕掛け、捕らえたノネズミ(n)のうち標識のついたノネズミ(m)の割合から、そ

\*Naoko TAJIMA/東京慈恵会医科大学第3内科・講師●〒105 港区西新橋3-25-8, \*\*Ronald E LaPorte/Professor, Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh ● A529 Craik Hall, Pittsburgh, PA 15261, USA, \*\*\*Shigeaki BABA/兵庫県立成人病センター・総長●〒673 兵庫県明石市北王子町13-70

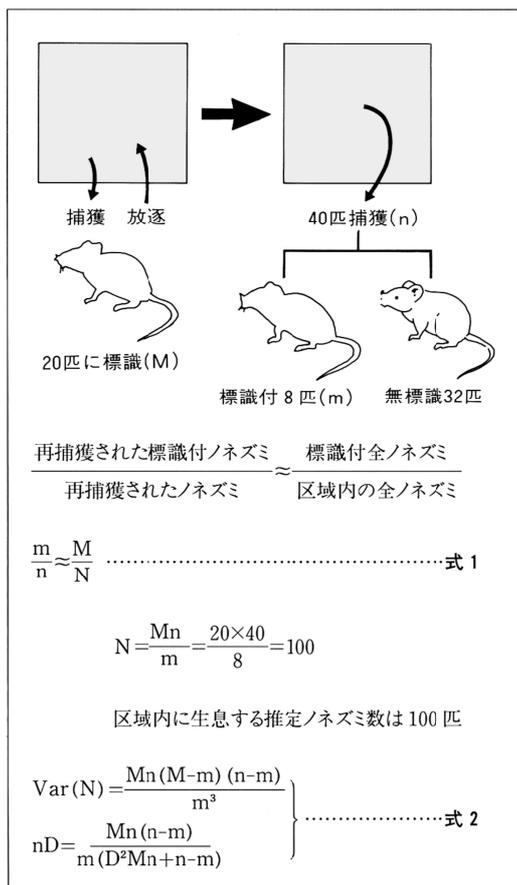


図 1 Petersen-Lincoln法

の場所に生息する総個体数(N)を推定する。放した個体が次の捕獲までの時間内に集団の中に無作為に混ざり、また、印をつけた個体と印をつけない(未採集)個体との間で、捕獲の効果に差がないとすれば、その地域の総個体数は、 $N : M = n : m$ の比例関係によって求められる(図1式1)。さらに、分散の推定値[ $\text{Var}(N)$ ]および任意の精度(D)を確保するために必要なサンプルの大きさ(nD)を計算するためには、図1式2を用いる。

C-M-R法は、1896年デンマークのPetersen<sup>1)</sup>が魚に、1930年アメリカのLincoln<sup>2)</sup>が鳥に、それぞれ独立に用いたもので、水産学者にはPetersen法、陸上動物学者にはLincoln指数法と呼ばれている。

2. Chapman法

さて、Petersenの式のうちで、n個体のサンプル

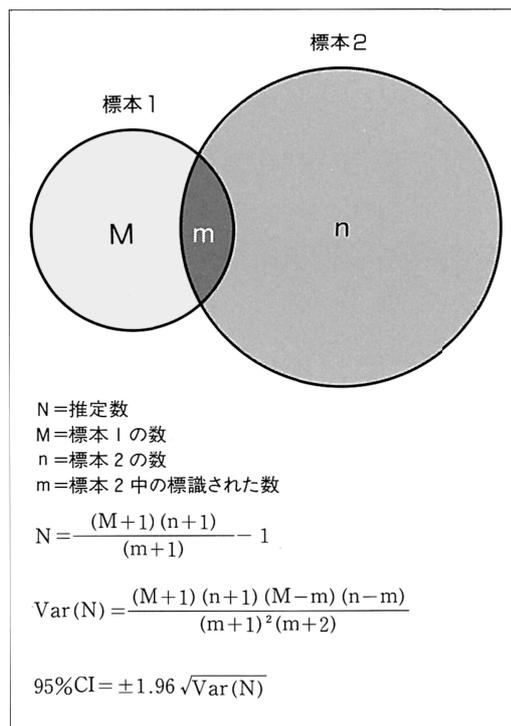


図 2 Chapman法

ル中での標識個体数mのみが確率変数である。このmの分布は、「赤玉M個と白玉N-M個が入っている壺の中から、非復元抽出によって無作為に取り出したn個の玉の中に含まれる赤玉の個数」の確率分布と同じであり、これに誤差法則をあてはめると、Petersenの推定量は一般に正の偏りをもつことになる<sup>7)</sup>。この偏りを補正したものがChapman法<sup>3)</sup>である(図2)。C-M-R法を疾病の頻度を推定する場合には、もっぱらこの改法が用いられている<sup>12,13)</sup>。

3. Jolly-Seber法

C-M-R法は以下の二つの条件を前提としている。①第1回目と第2回目の捕獲時に、各個体は独立に、かつ標識の有無にかかわらず等しい確率で捕獲されること、②標識個体を元の場所に戻してから第2回目の捕獲までの間に、個体群への新規加入(出生、移入)や消失(死亡、移出)がないことである。しかし、野外で動き廻っている動物の動態を解析するとき、②の前提を満たすのはなかなか難しい。そこで、これを解除するために用

表 マドリッドにおける IDDM 発症率—C-M-R 法による推定値—<sup>12)</sup>

	病院から①	糖尿病協会から②	①と②の重複	①+②-③	C-M-R法による推定値	$N = \frac{(M+1)(n+1)}{m+1} - 1$ $= \frac{(432+1)(138+)}{119+1} - 1$ $= \frac{60,187}{120} - 1 \approx 501$
症例数(例)	432(M)	138(n)	119(m)	451	501(N)	
人口10万人あたりの発症率	9.8	3.1	-	10.2	11.3	
捕捉率(%)	86	28	-	90	100	

いられるようになったのが、Jolly-Seber 法<sup>4,5)</sup>である。この方法は、標識・放逐・再捕を間隔をおいて2回以上繰り返すことにより、各時点で存在する個体数の情報のみならず、新規加入や消失に関する情報も盛り込めるようにしたものである。したがって Jolly-Seber 法では、前提①は生きているが、②は無視してよい。しかし、調査期間中、出戻りはないものと仮定されている。現在、動物の個体数を推定する際に、最もよく使われている方法である。

このように動物の生態学の分野では、19世紀の終わりから個体数の推定法が考案され、時代とともに改善されてきた<sup>14)</sup>。

### III. IDDM の真の発症率推定への応用

#### 1. 方法論の解釈

さて、疾病の発症率を推定するときに、生態学で用いられてきた C-M-R 法を、そのままあてはめることができるのだろうか。一定の地域にいる動物を数えることは有病率をみることに等しいが、私たちが知りたいのは発症率である。そこで「1 回目の捕獲個体数=病院調査によって見出された(第1の情報源)、1年間に発生した新規 IDDM 患者数」と置き換えることにする。また、「2 回目の捕獲個体数=病院以外の情報から見出された(第2の情報源)、1年間に発生した新規 IDDM 患者数」と解釈すればよい。ただし、ここで第1の情報源と第2の情報源の間の「独立性」が問題となる。たとえば、生態学では印をつけたことによって、その個体が捕らえやすくなったり、その逆であったりしないということが大前提である。IDDM も同様に、第1の情報源で発見された患者が、発見されなかった患者に比べて、第2の情報源でも発見されやすいということがないという条件が必要

である。このように「独立性」の問題は、生態学と発症率の場合異なるようにみえるが、本質的に何ら違いはない。

生態学の個体数推定の際には、新規移入と消失なしという条件を解除するために、Jolly-Seber 法がよく用いられると前述した。では、IDDM の場合はどうであろうか。この場合には、新規に発症した患者を対象にしているので移入はあり得ない。また、有病率を問題にしているわけではないので、消失(死亡)の心配をしなくてもよい。ただし、診断がつかないまま疾病の発生時に死亡した症例は問題となる。しかし、先進諸国ではこのような症例はまれと思われる。したがって、IDDM やその他慢性疾患の発症率を推定するときには、もっぱら Chapman 法が用いられる。

#### 2. 第1,2の情報源

わが国で IDDM の調査を行うときにしばしば使われる資料は、病院へのアンケート調査、小児特定慢性疾患公費給付申請書、小児糖尿病サマーキャンプ参加者名簿などである。しかし、これらは情報提供に同じ医師が関与している可能性が大きいので、独立した別個の情報とはなり得ない。相互依存性の高い情報は、C-M-R 法の大前提を満足し得ないのである。また、それぞれの調査による症例の捕捉率が低いと予想されるときには、相互に関連している複数の情報源をまとめ、重複している症例を除外し、これを第1の情報源からの症例数(M)とする。

これに対して、学校に対するアンケート調査や健康保険請求明細書からの情報は、医療機関関係からの情報とはまったく独立していると考えてよい。したがって、これらは第2の情報源となる。学校調査では5歳以下の症例を把握することはできないが、これは大きな問題とはならない。なぜ

ならば、第2の情報源から真の症例数(N)の20~30%を把握できれば、C-M-R法を適応することができるので<sup>17)</sup>、発症率の低い0~5歳の症例がたとえここから落ちてしまっても、全体に及ぼす影響は少ないのである。地域によっては、健康保険請求明細書からインスリンを使用している5歳以下の症例を抽出するという方法を、第2の情報源の補足とし得るだろう。しかし、時間や労力、そして経費との兼ね合いを考えて、これを用いるかどうか決めるべきであろう。また、情報源はそれぞれが独立しているのであれば、二つよりは三、四と情報源が多くなるほど発症率を正確に推定することができる<sup>15~17)</sup>。

最近、スペイン・マドリッド行われたIDDMの発症率の調査研究では、病院の診療録からの情報と糖尿病協会会員名簿を用いて情報を得て、C-M-R法を使い真の発症率を推定している。その結果、IDDMの発症率は病院のみの調査では人口10万人あたり9.8であったが、C-M-R法を用いて検討したところ、真の推定値は11.3であることが明らかになった<sup>12,17)</sup>(表)。

### おわりに

先進諸国では、多大な労力と時間をかけて医療機関およびその関連の情報を最大限に駆使すれば、IDDMの新規症例の大多数を把握することは可能であろう。これによって算出した発症率もほぼ正確と考えてよいであろう。しかし、科学的にその数値の確かさを証明するためには、客観的な評価方法が必要である。また、同じ方法論を用いなければ、地域間であるいは国際間で発症率の比較をすることはできない。しかし、C-M-R法を応用すれば、たとえ患者の捕捉率が90%に満たなくても真の発症率を推定できるし、他のデータと比較することも可能である。とくに調査方法として診断の標準化を明確にし、地域人口動態別にその推移年次的に追跡することができる。また、疫学調査として比較的経済的に実施できる。これらの点でC-M-R法は臨床疫学調査方法として非常に優れた方法である。今後、慢性疾患とくに疾病の病態変化のはいIDDMやその他の代謝疾患、あるいは心筋梗塞などの疫学調査として有用であると考えている。

謝辞：稿を終えるにあたり、貴重な助言を賜った東京大学農学部森林動物学教室の宮下直博士に深謝いたします。

### (文献)

- 1) Petersen CGJ : The yearly immigration of young plaice into the Limfjord from the German Sea. **Report of the Danish Biological Station to the Ministry of Fisheries**, 6 : 1-48, 1896
- 2) Lincoln FC : Calculating waterfowl abundance on the basis of banding returns. **US Dept Circular, No 118**, 1930
- 3) Chapman DG : Some properties of the hypergeometric distribution with applications to zoological sample censuses. **Berkeley : University of California Publications in Statistics**, 1 : 131-160, 1951
- 4) Jolly GM : Explicit estimates from capture-recapture data with both death and immigration-stochastic model. **Biometrika**, 52 : 225-247, 1965
- 5) Seber GAF : A note on the multiple-recapture census. **Biometrika**, 52 : 249-259, 1965
- 6) 伊藤嘉昭：第IV章マーキング法による個体数の推定。 **動物生態学入門**。古今書院，東京，pp197-220，1963
- 7) 久野英二：動物の個体群動態研究法I—個体数推定法。 **生態学研究法講座17**。共立出版，東京，pp9-30，1986
- 8) Sekar CC, Deming WE : On a method of estimating birth and death rates and the extent of registration. **Amer Stat Assoc J**, 44 : 100-115, 1949
- 9) Robles SC, Marrett LD, Clarke EA, et al : An application of capture-recapture methods to the estimation of completeness of cancer registration. **J Clin Epidemiol**, 41 : 495-501, 1988
- 10) Cochi SL, Edmonds LE, Dyer K, et al : Congenital rubella syndrome in the United States, 1970-1985 on the verge of elimination. **Am J Epidemiol**, 129 : 349-361, 1989
- 11) Hamman RF, Gay EC, Cruickshanks KJ, et al : Colorado IDDM registry. Incidence and validation of IDDM in children aged 0-17yr. **Diabetes Care**, 13 : 499-506, 1990
- 12) Serrano-Rios M, et al : Incidence of Type 1 (insulin-dependent) diabetes mellitus in subjects 0j-14 years of age in the Comunidad of Madrid, Spain. **Diabetologia**, 33 : 422-424, 1990
- 13) LaPorte RE, McCarty DJ, Tull ES and Tajima N : Counting birds, bees, and NCDs. **Lancet**, 339 : 494-495, 1992
- 14) Bailey NTJ : Improvements in the interpretation of recapture data. **J Anim Ecol**, 21 : 120-127, 1952
- 15) Bishop YM, Fienberg SE, Holland PW, et al : Estimating the size of a closed population. **Discrete Multivariate Analysis. Theory and Practice**. MIT Press, Cambridge, MA, pp229-256, 1975
- 16) Wittes JT, Sidel VW : Capture-recapture methods for assessing the completeness of case ascertainment when using multiple sources. **J Chron Dis**, 27 : 25-36, 1974
- 17) McCarty DJ, Tull ES, Moy CS and LaPorte RE : Ascertainment corrected incidence rates. Submitted to **Am J of Epidemiol**