

Capture-Mark-Recapture 法による 患者数推定の試みについて

研究協力者 高木 廣文

文部省統計数理研究所

中内 洋一

国立病院医療センター皮膚科

稲葉 裕

順天堂大学衛生学教室

はじめに

従来、難病の患者数の推定は、全国疫学調査を実施し、県別地域別などの回収率と報告患者数を用いて、回帰分析の手法を応用して推定していた¹⁾。最近、心筋梗塞の患者数を推定するための方法として、Capture-Mark-Recapture法を用いる試みが報告された²⁾。

Capture-Mark-Recapture法は、湖にいる魚の数を推定する場合など、野生動物の個体数の推定に用いられていたが、慢性疾患の患者数を推定するのにとも有用ではないかと考えられる。ここでは、同方法の簡単な説明を行い、神経線維腫症(レックリングハウゼン病)と結節硬化症についての実例を用いて、患者数の推定方法を行う。

方 法

まず、使用する変数を定義する。特定の患者集団に対して、2つの標本調査を独立に行ったとしよう。それぞれの標本数を N_1 , N_2 とする。この場合、2つの調査に重複して対象となるケースが存在するだろう。その人数を m とする。Capture-Mark-Recapture法を用いれば、これだけの情報から、もとの母集団の人数 N が推定できる。

$$N = \frac{(N_1 + 1)(N_2 + 1)}{(m + 1)} - 1 \quad (1)$$

となる。この分散は、

$$\text{Var}(N) = \frac{(N_1 + 1)(N_2 + 1)(N_1 - m)(N_2 - m)}{(m + 1)^2(m + 2)} \quad (2)$$

となる。(1), (2)式から全体の人数の95%信頼区間は、

$$N - 1.96 \times \sqrt{[\text{Var}(N)]} < \text{全体の人数} < N + 1.96 \times \sqrt{[\text{Var}(N)]} \quad (3)$$

により求められる。99%信頼区間を求めるには、(3)式の1.96を2.58にかえればよい。

応用例と考察

Croweらは神経線維腫症(レックリングハウゼン病)の症例についての州立病院調査から、同症の患者の有病率を推定している³⁾。同調査は、精神的欠損のためのミシガン州立病院とてんかんのための1施設の入院患者、6,873人について検査を実施した。その結果、神経線維腫症の患者17例と3例の疑い例が認められた。さらに、この調査に含まれた107家族のうち、100家族は州の同一地域に在住しており、100家族中に、当該疾患を有する213例が知られていた。この213人のうち1人が精神的欠損のための上記の病院に入院していた。

Croweらは上記の方法とは若干異なる方法で、1950年での人口は6,069,568人から、同症の有病率は、0.00060と0.00070の間と計算した。これを上記の(1), (2), (3)の方法により計算すると、確実例のみの場合は表1の網かけの部分のように推定値は1,925人、95%信頼区間の下限が229人、上限は3,970人となる。また、疑い例も含めた場合は、表2のように推定値は2,246人、下限は232人、上限は4,653人となる。これを有病率にすると、推定値は0.00032~0.00037となり、上限は0.00065~0.00077程になる。Croweらの報告値はこの上限に近い値になっている。ちなみに、有病率の推定値に日本人口(1億2千万)を乗じると、38,400~44,400人となり、以前報告した4万人程度と一致する⁴⁾。ただし、表1、表2から分かるように、一致する人数が1例増減する影響が極めて大きい。これは、2つの調査対象者数が小さすぎることに由来するものであり、安定した結果を得るには、より大きな標本数が必要である。

表1 神経線維腫症の推定値と95%信頼区間
(確実例のみの場合)

共通数	推定値	下限値	上限値
0	3851	229*	9027
1	1925	229*	3970
2	1283	229*	2424
3	962	229*	1699
4	769	252	1287

$N_1 = 213$, $N_2 = 17$, $m = 1$.

*推定下限値が標本数より小さいので、標本数を代入。

表2 神経線維腫症の推定値と95%信頼区間
(疑い例も含む場合)

共通数	推定値	下限値	上限値
0	4493	232*	10557
1	2246	232*	4653
2	1497	232*	2847
3	1123	245	2000
4	898	277	1518

$N_1 = 213$, $N_2 = 20$, $m = 1$.

*推定下限値が標本数より小さいので、標本数を代入。

次に、結節性硬化症の全国患者数の推定についての例を示そう。2つの調査は、(1)昭和61年度に実施した全国疫学調査二次調査(200床以上の病院での1年間に外来した患者が対象)により把握した患者479例⁴⁾、および(2)昭和34年10月~昭和46年12月までの12年間に、東京大学皮膚科外来を受診した52例、である。これらの調査は、調査時期が大幅に異なるため、対象者の死亡についての対応が必要であるが、ここでは数値上の応用例として考えて貰いたい。

上記の(1)(2)の患者について、①姓のイニシャル、②名のイニシャル、③性別、④誕生日の年号、⑤誕生日の年、⑥誕生日の月、⑦誕生日の日、の7項目の一致を調べ、5項目以上の一致者については原票をあたって、同一人かを確認した。その結果、2例が共通していることが分かった。表3に推定結果を示した。網かけの部分を見ると、推定値は8479人、下限は529人、上限は16526人となる。ちなみに、昭和61年度の調査では、全国の外来受診患者数を1333~1883例とし、全国患者数としては、その10倍程度の約1万5千人と推定している。この例は2つの調査時期が異なりすぎるた

め、対象者の死亡などの問題が生じる。実際には死亡の影響で、観察された共通する患者数は小さくなっているものと考えられる。表3の共通数がより大きな3, 4, 5人の箇所をみると、患者数の推定値は急激に小さくなることが分かる。

すでに述べたように、Capture-Mark-Recapture法を用いる場合、2つの独立な調査の標本数は大きい方が、推定された患者数は安定するし、共通する患者数も大きい方が安定した結果が得られる。さらに、信頼区間の幅も小さくなる。今回示した例は、そのようにはなっておらず、幾つかの問題もあるが、Capture-Mark-Recapture法により、比較的簡単に患者数の推定が行えることが理解されたと考えられる。今後は、患者数推定のための疫学調査法の開発と本方法などの応用面について検討する必要があるだろう。

表3 結節性硬化症の推定値と95%信頼区間

共通数	推定値	下限値	上限値
0	25439	529*	60326
1	12719	529*	26809
2	8479	529*	16526
3	6359	1021	11697
4	5087	1233	8941
5	4239	1300	7178

$N_1 = 479$, $N_2 = 52$, $m = 2$.

*推定下限値が標本数より小さいので、標本数を代入。

文 献

- 1) 高木廣文, 佐藤俊哉, 稲葉 裕: 厚生省特定疾患の全国疫学調査による患者数推定と推定誤差について, 日本公衆衛生雑誌, 35(6), 275-285, 1989.
- 2) R. E. Laporte, E. S. Tull, and D. McCarty: Monitoring the Incidence of Myocardial Infarctions: Applications of Capture-Mark-Recapture Technology, International Journal of Epidemiology, 21(2), 258-262, 1992.
- 3) Crowe, F. W., Schull, W. J., and Neel, J. V.: VI. The Frequency of Neurofibromatosis, A Clinical, Pathological, and Genetic Study of Multiple Neurofibromatosis, 142-146, 1956.
- 4) 高木廣文, 稲葉 裕, 高橋月容, 新村眞人, 中内洋一: レックリングハウゼン病と結節性硬化症の重複率と全国患者数, 厚生省特定疾患神経皮膚症候群調査研究班. 昭和62年度研究報告書, 11-15, 1988.