

厚生省特定疾患難病の疫学調査研究班ワークショップ

全国調査による患者推定方法

— 厚生省特定疾患全国有病者数推定 —

編者 稲 葉 裕
青 木 國 雄

厚生省特定疾患難病の疫学調査研究班

(班長 青 木 國 雄)

厚生省特定疾患難病の疫学調査研究班ワークショップ

全国調査による患者推定方法

昭和63年2月5日

東京、順天堂大学

目 次

I. 序	青木 國雄	1
はじめに	稲葉 裕	3
II. 患者数推計の一方法とその検討	水野 正一 他	5
全国疫学調査一次調査による		
患者数推定と推定誤差について	高木 廣文 他	10
肝内結石症患者全国調査	池田 正人	16
受療率推定の一方法	中江 公裕	19
III. コメント		
患者数推定へのコメント	長岡 省悟	22
全国患者数推計上の回収率に関する補正方法について	橋本 修二	23
全国調査による患者数推計方法についてのコメント	柳井 晴夫	28
方法論の確立を願って	山本 俊一	32

全国疫学調査一次調査による 患者数推定と推定誤差について

高木廣文*、佐藤俊哉**、稲葉 裕***

はじめに

厚生省は、昭和47年以来いくつかの難病を「特定疾患」として指定している。各特定疾患に対する調査研究班を中心に、各種の研究が行われているが、その中の一つに、全国の患者数を把握するために全国疫学調査一次調査がある（たんに「一次調査」と呼ばれている）。一次調査は、各調査研究班と難病の疫学調査研究班の事務局により、全国の特定の病院の特定診療科（時に特定の診療所）に対して、特定年での一年間（時には数年間）の各特定疾患の受診患者数を性別に葉書に記載するように依頼し、それを報告してもらうという郵送法により行われている。調査対象施設数もしくは診療科数は数千に達し、その回収率は50%前後が一般的であると思われる。

調査対象のうち未回収分に関しては、患者数を推定する必要が生じる。最も簡単な方法は、患者数の回収率で除すというものであるが、患者のいない施設や診療科では、患者のいる施設などに比べて回収率が低いものと考えられ、ほとんどこの方法は採用されていない。

厚生省特定疾患の全国患者数推定には、「柳井らの方法¹⁾」が発表された昭和50年以来、一次調査に基づく標準的な推定方法としてよく用いられている。同方法は、原論文では重回帰式に基づく推定であったが、現在ではより簡単な計算方法として、回収率を説明変数として受診率を基準変数とする単回帰式に基づく推定方法が取られている。しかし、同方法は回収率と受診率の単相関係数が負の場合、全く推定が行えないという欠点がある。最近ではその欠点を改良するために、原点

を通る回帰直線に基づいて推定を行うという「中江らの方法」²⁾、「水野らの方法」³⁾、さらに「中江らの方法の修正方法」⁴⁾（修正中江らの方法）などが提唱されている。

しかし、多くの手法のうち、どの方法によれば全国患者数をよりよく推定できるのか、また推定値はどの程度信頼できるのかという問題に関しては、各推定方法が単回帰式を基にしているが、後述するように通常とは異なる計算式を用いるために、理論的な面から客観的な議論ができないままに放置されていた。

各推定方法を比較する場合、各手法にどのような特徴や偏りがあるかを知る必要がある。最も重要な点は、推定の信頼度を示す指標となる推定誤差（推定値の標準誤差）を求めて、相互に比較することである。これまで、各推定方法による推定値の標準誤差を提示し、各手法に関して理論的および実際の比較を行った報告は見あたらない。本報告では、主要ないくつかの方法による患者数の推定値の理論的な推定誤差の計算方法を提示し、さらに厚生省特定疾患の全国疫学調査における患者数推定の方法に関するいくつかの問題点に関して、理論的な考察を若干行う。

方 法

ここでは、全国患者数推定のための各方法の簡単な説明と、推定値の誤差の算出方法を提示する。

まず、各方法の説明のために使用する記号の定義をしよう。各地域（都道府県や地域ブロックなど）で調査対象となった施設数や診療科数を M_i とし、そのうち報告があった診療科数を N_i とする ($i = 1, \dots, k$; k は地域数)。また、 N_i 個の診療科から報告された患者数を t_i とする。

調査の回収率 R_i は、

$$R_i = N_i / M_i \quad (1)$$

* 聖路加看護大学統計学教室

** 東京大学医学部疫学教室

*** 順天堂大学医学部衛生学教室

である。地域*i*での人口を R_i とすると、報告受診率 y_i は、

$$y_i = t_i / P_i \quad (2)$$

となる。

全国での調査対象診療科数 M 、報告診療科数 N 、回収率 R は、

$$M = \sum_i M_i, N = \sum_i N_i, R = N / M \quad (3)$$

となり、全国人口を P とすると、報告患者総数 t は、

$$t = \sum_i t_i \quad (4)$$

となるので、報告受診率 y は、

$$y = t / P \quad (5)$$

により求められる。ここで、 $\sum_i$ は地域についての総和をとることを示す。

1. 柳井らの方法¹⁾

「柳井らの方法」では、まず地域別の報告受診率 y_i と回収率 R_i ($i = 1, 2, \dots, k$) の間に、

$$y_i = \beta_1 R_i + \alpha + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2_1) \quad (6)$$

が成り立つものと仮定する。なお、 $N(0, \sigma^2_1)$ は平均 0、分散 σ^2_1 である正規分布を示し、 ε_i は誤差項であり、その分布が $N(0, \sigma^2_1)$ に従うもの仮定する。

母回帰係数 β_1 と定数項 α (切片) の推定値 b_1 、 a はよく知られているように、

$$\begin{cases} b_1 = \sum_i (R_i - \bar{R})(y_i - \bar{y}) / \sum_i (R_i - \bar{R})^2 \\ a = \bar{y} - b_1 \bar{R} \end{cases} \quad (7)$$

である⁵⁾。ここで、 $\bar{y}$ 、 $\bar{R}$ は報告受診率と回収率の平均値である。また、誤差分散 σ^2_1 の推定値 $\hat{\sigma}^2_1$ (残差分散) は、

$$\hat{\sigma}^2_1 = (1 - r_{Ry})^2 \sum_i (y_i - \bar{y})^2 / (k - 2) \quad (8)$$

である⁵⁾。ここで、 r_{Ry} は報告受診率と回収率の単相関係数であり、 k は地域数である。

(7) 式による回帰係数 b_1 の分散は、

$$V(b_1) = \hat{\sigma}^2_1 / \sum_i (R_i - \bar{R})^2 \quad (9)$$

である⁵⁾。

柳井らの方法では、地域 i の未回収分 $(1 - R_i)$ について、回帰係数 b_1 をかけた数値を、報告受診率 y_i に加え、推定受診率 $\hat{\mu}_i$ を求めている。すなわち、

$$\hat{\mu}_i = y_i + (1 - R_i) b_1 \quad (10)$$

により、推定受診率を求めている。なお、実際には回帰式による「予測 prediction」と考えられる

が、本論文では慣用に従って「推定 estimation」という用語を用いることにする。

次に、推定受診率 $\hat{\mu}_i$ の推定の信頼度を検討するために、その期待値と分散を求めよう。

(10) 式による $\hat{\mu}_i$ の期待値 $E(\hat{\mu}_i)$ は、

$$\begin{aligned} E(\hat{\mu}_i) &= E(y_i) + (1 - R_i) E(b_1) \\ &= \beta_1 R_i + \alpha + (1 - R_i) \beta_1 \\ &= \beta_1 + \alpha \end{aligned} \quad (11)$$

と求められる。

従って、 $\hat{\mu}_i$ の分散 $V(\hat{\mu}_i)$ は、(6) 式の回帰式が成り立つと仮定できるならば、

$$\begin{aligned} V(\hat{\mu}_i) &= E\{y_i + (1 - R_i)b_1 - \beta_1 - \alpha\}^2 \\ &= E\{y_i - \beta_1 R_i - \alpha + (1 - R_i)(b_1 - \beta_1)\}^2 \\ &= V(\varepsilon_i) + (1 - R_i)^2 V(b_1) + \\ &\quad 2(1 - R_i) E(\varepsilon_i b_1) \\ &= V(\varepsilon_i) + (1 - R_i)^2 V(b_1) \\ &\quad + 2(1 - R_i)(R_i - \bar{R}) V(\varepsilon_i) / \sum_i (R_i - \bar{R})^2 \end{aligned} \quad (12)$$

と求められる。

誤差項 ε_i が、先に仮定したように正規分布 $N(0, \sigma^2_1)$ に従うものとし、 σ^2_1 の推定値として (8) 式による $\hat{\sigma}^2_1$ を用いると、推定受診率 $\hat{\mu}_i$ の分散は以下のように近似できる。

すなわち、

$$V(\hat{\mu}_i) = \hat{\sigma}^2_1 \left\{ 1 + \frac{(1 - R_i)(1 + R_i - 2\bar{R})}{\sum_i (R_i - \bar{R})^2} \right\} \quad (13)$$

と求められる。

地域 i の推定患者数 $\hat{\tau}_i$ は、推定受診率 $\hat{\mu}_i$ に人口 P_i をかければよい。すなわち、

$$\hat{\tau}_i = P_i \hat{\mu}_i, \quad V(\hat{\tau}_i) = P_i^2 V(\hat{\mu}_i) \quad (14)$$

と求められる。

全国推定患者数 $\hat{\tau}$ とその分散は、各地域を独立とすれば、

$$\hat{\tau} = \sum_i P_i \hat{\mu}_i, \quad V(\hat{\tau}) = \sum_i P_i^2 V(\hat{\mu}_i) \quad (15)$$

により求められる。

2. 水野らの方法³⁾

「柳井らの方法」による回帰直線は原点を通るとは限らないので、(6) 式に基づく回帰式で回収率が 0 の場合に受診率が負となることがある。また、報告受診率と回収率の相関が負の場合、事実上推定は不可能であり、理論的にも実際的にも不都合である。これらの問題を避けるために、中江らは、回収率 0、受診率 0 というデータを恣意

的に多数加え、原点の近くを通る回帰直線を求める方法を提唱している²⁾。しかし、この方法は、データを作為的に追加するという点に大きな問題がある。そこで、水野らは調査データから直接原点を通る回帰式を求めるという、以下の方法を提唱している³⁾。

「柳井らの方法」と異なり、定数項 α を除いて、報告受診率 y_i と回収率 R_i ($i = 1, 2, \dots, k$) の間に、

$$y_i = \beta_2 R_i + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2_2) \quad (16)$$

が成り立つと仮定する。ここで、 ε_i は「柳井らの方法」と同様に誤差項であり、正規分布 $N(0, \sigma^2_2)$ に従うものとする。

母回帰係数 β_2 の推定値 b_2 は、よく知られているように、

$$b_2 = \sum_i R_i y_i / \sum_i R_i^2 \quad (17)$$

である⁶⁾。

この場合、誤差分散 σ^2_2 の推定値 $\hat{\sigma}^2_2$ は、

$$\hat{\sigma}^2_2 = (\sum_i y_i^2 - b_2 \sum_i R_i y_i) / (k-1) \quad (18)$$

となり、推定値 b_2 の分散は、

$$V(b_2) = \hat{\sigma}^2_2 / \sum_i R_i^2 \quad (19)$$

となる⁶⁾。

「水野らの方法」では、回収率を 1 とした場合の推定受診率を求め、それに全国人口 P を乗じて全国患者数の推定を行う。従って、推定受診率 $\hat{\mu}_i$ は回帰係数 b_2 と一致する。

上記のことから、推定受診率 $\hat{\mu}_i$ の期待値と分散は、以下のように求められる。

推定受診率 $\hat{\mu}_i$ の期待値は、

$$E(\hat{\mu}_i) = E(b_2) = \beta_2 \quad (20)$$

と求められる。また、推定受診率 $\hat{\mu}_i$ の分散は、

$$V(\hat{\mu}_i) = E(b_2 - \beta_2)^2 = V(b_2) \quad (21)$$

と求められ、(19) 式による回帰係数の分散に一致する。

全国患者数 τ の推定値 $\hat{\tau}$ とその分散は、各地域を独立とすれば、

$$\hat{\tau} = P \hat{\mu}, \quad V(\hat{\tau}) = P^2 V(\hat{\mu}) \quad (22)$$

により求められる。

3. 修正中江らの方法⁴⁾

回帰直線は「水野らの方法」と同様に、原点を通るように求めるが、地域別の調査対象の未回収分についての補正の方法は、「柳井らの方法」と

同様に行うという、上記の 2 方法の中間的方法を考えてみよう。

報告受診率と回収率の回帰直線が原点を通るように定める点までは、「水野らの方法」と同一式、すなわち、(16) ~ (19) 式を用いるが、地域 i での推定受診率 $\hat{\mu}_i$ は「柳井らの方法」と同一式、すなわち、

$$\hat{\mu}_i = y_i + (1 - R_i) b_2 \quad (23)$$

により求めることにする⁴⁾。ここで回帰係数 b_2 は、「柳井らの方法」と異なり、「水野らの方法」と同一である。

地域 i の推定受診率 $\hat{\mu}_i$ の期待値は、

$$E(\hat{\mu}_i) = E(y_i) + (1 - R_i) E(b_2) = \beta \quad (24)$$

と求められ、「水野らの方法」の推定受診率の期待値と一致する。また、推定受診率 $\hat{\mu}_i$ の分散は、

$$\begin{aligned} V(\hat{\mu}_i) &= E\{y_i + (1 - R_i) b_2 - \beta\}^2 \\ &= E\{y_i - \beta_2 R_i + (1 - R_i)(b_2 - \beta_2)\}^2 \\ &= V(\varepsilon_i) + (1 - R_i)^2 V(b_2) + \\ &\quad 2(1 - R_i) E(\varepsilon_i b_2) \\ &= V(\varepsilon_i) + (1 - R_i)^2 V(b_2) \\ &\quad + 2(1 - R_i) R_i V(\varepsilon_i) / \sum_i R_i^2 \end{aligned} \quad (25)$$

と求められる。上式に (19) を代入することで、

$$V(\hat{\mu}_i) = \hat{\sigma}^2_2 \{1 + (1 - R_i)^2 / \sum_i R_i^2\} \quad (26)$$

と推定受診率 $\hat{\mu}_i$ の分散が求められる。

地域 i の推定患者数 $\hat{\tau}_i$ および全国推定患者数 $\hat{\tau}$ とその分散は、「柳井らの方法」と同様に、それぞれ (14) 式、(15) 式により求められる。

上記の方法が、「修正中江らの方法」である。

表 1 に、上記の 3 方法による各統計量の推定式をまとめて示した。

考 察

厚生省特定疾患の全国疫学調査一次調査に基づく全国患者数推定方法についての簡単な解説と、推定誤差の計算方法を提示した。ここでは、患者数推定に関するいくつかの問題点に関して考察を行う。

なお、念のために付け加えるが、厚生省特定疾患全国疫学調査は、一次調査だけでなく、一次調査の結果を踏まえて「二次調査」を行うのが一般的な手順である。「二次調査」は、一次調査で患者ありと報告のあった診療科を調査対象とし、患者の性、年齢、職業、診療の確実性、日常生活、

臨床症状などを調査するために設定された質問紙を郵送し、主治医の記載後に返送するように依頼するという、郵送法による調査である。

難病の患者数推定には、上記の二次調査に基づく重複例の除去、疑い例の検討、未受診者の考慮などを加える必要があることは言うまでもないだろう。

1. 推定誤差の変動と地域区分について

推定誤差の大小は、表1の③残差分散と⑤推定受診率の分散の各式から判るように、報告受診率の分散もしくは2乗和、報告受診率と回収率の相関係数もしくは積和、標本数（地域区分の個数）、そして各地域の回収率の影響を受けることが判る。簡単にいえば、相関係数は正で大きいほど、報告受診率の変動は小さいほど、各地域の回収率は1に近いほど、そして標本数は大きいほど、推定誤差は小さくなると考えてよいだろう。

上記の点と関連して、推定に用いるデータの地域区分をどの様にすればよいかという問題がある。通常は、地域として都道府県別の区分を用いるか、もしくは適当に区分した地域ブロックを用いるようである。どちらが患者数推定に対して優れているかは、一概には結論できない。地域区分を変えただけで、推定値に大きな相違が生じることは稀ではない。報告受診率と回収率の相関が極めて大きくなり、そのため推定誤差が極めて小さくなることもあるだろうし、逆に正の相関が負となり、推定上好ましくなくなることもある。このような問題は、実データに大きく依存することが普通であると考えられるので、実際の患者数推定を地域区分を変えて数通り行い、推定誤差の小さいものを採用するという便法を取る以外はないように思われる。

何れにしても、地域区分の問題は、ある程度は試行錯誤的に分析を行い、推定結果を比較しながら、継続的に検討を重ねる必要があると考えられる。

2. 各地域での患者数推定について

全国患者数の推定を行う場合「水野らの方法」以外は、地域別の推定患者数を求めて、その合計から全国患者数を算出している。従って、各地域

別の推定患者数を比較することは極めて容易である。しかし、個々の地域での推定患者数もしくは推定受診率を求めて比較することに関しては、以下の理由から意味があるとは考えられない。

「柳井らの方法」は、原論文では推定受診率が回収率と無相関になるように、その推定式が設定されたと述べられている¹⁾。このこと自体は事実であるが、より重要な点は、表1の推定受診率の期待値をみれば判るように、「各地域の推定受診率の期待値は一定であり、各地域の真の受診率は等しい」ことが暗黙に仮定されていると考えねばならないことである。これは、他の2方法についても、表1から明らかなように同様である。

真の受診率が地域によらず一定ならば、推定受診率と回収率が無相関になるということは、当然期待されることであり、真の受診率が一定という仮定の方が理論的には重要な点であろう。上記のように、真の受診率が一定であると仮定される場合に、各地域の推定受診率の比較に意味があるとは考えにくいし、現実問題として各地域の推定誤差は無視できないほど小さくはないだろう。

3. 方法論的問題点と他の推定方法

本報告で取り上げた各手法は、特定疾患の患者数推定方法として比較的よく用いられているものである。しかし、これらの方法に共通する問題がある。すなわち、「柳井らの方法」では、すでに何度も指摘したように、報告受診率と回収率の相関が負の場合、推定は困難であり、患者数の推定結果は実際上意味のあるものとはいえなくなる。

他の2方法では、相関係数が負の場合であっても、形式的に患者数の推定は可能である。だが、仮に推定ができるからといって、それが妥当であるかは別の問題である。確かに、原点を通る回帰直線を用いるので、相関が負の場合の推定上の不都合を避けることはできる。しかし、そのような場合に回帰直線を求めることが妥当とは考えにくい。回帰分析は予測のための方法であり、2変数間の単相関係数が小さければ、予定値の誤差は急激に増加する。さらに、単相関係数が正でなければ原点を通る直線を求めても、それは単に形式的意味しかない。従って、今後は回帰によらない推定方法も考慮すべきであろう。

Takagi⁷⁾は報告のあった各施設の患者数の分布に着目し、多項分布による患者数の最尤推定法を提示しており、別法として多重超幾何分布による最尤推定法についても触れている。また、全国疫学調査によらない他の既存資料に基づく患者数推定の手法との比較も、当然今後の課題となるであろう。

実際に推定値の良さを問題にする場合、推定値の分散とともに、その偏りについても評価する必要があるだろう。しかし、本報告で扱った患者数の推定に関しては、患者の発見率（患者が対象施設を受診する確率に関する分布など）に適当な仮定を置かないと、推定上の偏りを評価できないためこの点に関しては触れていない。

また実際問題として、我々は真の患者数を知ることではできないのであるから、理論的に最も妥当と思われる方法を考える必要があり、問題を抱えた方法を適当に修正するだけでは、結果が類似しているからといって、疫学研究上満足することはできないだろう。従って、今後は報告受診率などにどのような仮定を置けばよいか、また未回収が起こるメカニズムをどう捉えるかを検討し、それらを踏まえて、偏りが小さく推定誤差も小さい患者数推定方法を開発する必要があるだろう。

要 約

厚生省特定疾患全国疫学調査一次調査による地域別の報告受診率と回収率に基づく患者数推定方法のうち、「柳井らの方法」、「修正中江らの方法」、「水野らの方法」について簡単な説明とともに、全国患者数の推定値の推定誤差の理論式を提示した。

今後の課題として、回帰式による患者数推定に付随するいくつかの問題点を指摘し、併せて他の患者数推定方法に関しても若干の考察を加えた。（本報告は、日本公衆衛生雑誌印刷中の論文⁸⁾の一部を改訂したものである。）

参考文献

- 1) 柳井晴夫, 他: 疫学調査における有病率補正の方式、日本公衛誌, 22 (2), 71-81, 1975
- 2) 中江公裕, 他: 膠原病四疾患第一次全国疫学調査成績; 全国患者数の推定、厚生省特定疾患難病の疫学調査研究班、昭和59年度研究業績報告書、135-150, 1985
- 3) 水野正一, 他: 全国調査成績を用いての有病率の一推計方法、日本公衛誌, 33 (10), 255, 1986
- 4) 高木廣文, 他: レックリングハウゼン病と結節性硬化症の全国受診患者数の推定について、厚生省特定疾患神経皮膚症候群調査研究班 昭和61年度研究報告書、15-19, 1987
- 5) 豊川裕之、柳井晴夫編著: 医学・保健学の例題による統計学 (初版)、50-59、現代数学社 (京都)、1982
- 6) Snedecor, G.W. and Cochran, W.G.: STATISTICAL METHODS (6th edition). the Iowa State Univ. Press, Ames., 1967; 畑村又好, 他共訳: スネデカー, コ克蘭 統計的方法 原書第6版, 131-165, 岩波書店 (東京), 1972
- 7) Takagi, H. A mode prediction for the number of patients with an incurable disease in the nationwide epidemiological survey, Bulletin of the Biometric Society of Japan. 8 (2), 51-60, 1987
- 8) 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推定と推定誤差について、日本公衛誌 (印刷中)

表 1 患者数推定方法のまとめ

統計量	柳井らの方法	修正中江らの方法	水野らの方法
①推定受診率	$\hat{\mu}_i = y_i + (1 - R_i) b_1$	$\hat{\mu}_i = y_i + (1 - R_i) b_2$	$\hat{\mu}_i = b_2$
②回帰係数	$b_1 = \sum_i (R_i - \bar{R})(y_i - \bar{y}) / \sum_i (R_i - \bar{R})^2$	$b_2 = \sum_i R_i y_i / \sum_i R_i^2$	$b_2 = \sum_i R_i y_i / \sum_i R_i^2$
③残差分散	$\hat{\sigma}_1^2 = (1 - r_{y_R}^2) \sum_i (y_i - \bar{y})^2 / (k-2)$	$\hat{\sigma}_2^2 = (\sum_i y_i^2 - b_2 \sum_i R_i y_i) / (k-1)$	$\hat{\sigma}_2^2 = (\sum_i y_i^2 - b_2 \sum_i R_i y_i) / (k-1)$
④推定受診率の期待値	$\beta_1 + \alpha$	β_2	β_2
⑤推定受診率の分散	$V(\hat{\mu}_i) = \hat{\sigma}_1^2 \left[1 + \frac{(1-R_i)(1+R_i-2\bar{R})}{\sum_i (R_i - \bar{R})} \right]$	$V(\hat{\mu}_i) = \hat{\sigma}_2^2 \left[1 + (1-R_i^2) / \sum_i R_i^2 \right]$	—————
⑥推定患者数	$\hat{\tau}_i = P_i \hat{\mu}_i$	$\hat{\tau}_i = P_i \hat{\mu}_i$	—————
⑦推定値の分散	$V(\hat{\tau}_i) = P_i^2 V(\hat{\mu}_i)$	$V(\hat{\tau}_i) = P_i^2 V(\hat{\mu}_i)$	—————
⑧全国患者数	$\hat{\tau} = \sum_i \hat{\tau}_i$	$\hat{\tau} = \sum_i \hat{\tau}_i$	$\hat{\tau} = P b_2$
⑨全国値の分散	$V(\hat{\tau}) = \sum_i P_i^2 V(\hat{\mu}_i)$	$V(\hat{\tau}) = \sum_i P_i^2 V(\hat{\mu}_i)$	$V(\hat{\tau}) = P^2 \hat{\sigma}_2^2 / \sum_i R_i^2$

(注) 記号は本文中で使用しているものに従っている。