ESTMATING THE TRUE DISEASE RATES WITH
CONSIDERATION OF FALLIBILITY OF DIAGNOSTIC TESTS

Takasi Yanagawa,* Fumiyoshi KASAGI** and Yasuko UKAJI*

*Department of Mathematics, Kyushu University 33, Fukuoka 812

**Department of Epidemiology, National Cardiovascular Center,
Suita 560

SUMMARY

The disease rates estimated through the use of a diagnostic test can be substantially different from the true rates (Yanagawa T. and Gladen, B. (1984). A method is developed in Yanagawa T., Kasagi F. and Yoshimura T. (1984) for estimating the true disease rates of a progressive disease based on an incomplete data. It adjusts for the fallibility of a diagnostic test. The mathematical characteristics of the method are studied in Yanagawa T. and Kasagi F. (1985). Although the method is applied well to Guatemalan onchocerciasis data, it is only the first step toward the solution of this complex problem. The method is extended by Yanagawa T. and Kasagi F. (1986) in their paper presented at the VIIIth Biometric Conference. They formulate the problem as estimation of a survival function with covariates under uncertainty of the response function. The method is briefly reviewed in this paper with emphasis placed on the importance of adjusting for the impact of diagnostic tests in estimating the disease rates. The conventional method which are frequently employed in epidemiologic studies are also critically reviewed.

1. 目的

ある特定の疾病について、ある地域の住民を対象に、あらかじめ定めた期間に実施した追跡調査から、しかも次のような枠組みの中で、真の発病率および有病率を推定するのが目的である。すなわち

(イ) 対象とする疾病(疾患)は、追跡期間内での真の"自然治癒"がありえないものに限る。これを progressive disease とよぶことにする。

(ロ) Diagnostic tests による診断は "positive", 又は "negative" の二項選択で表す。

(ハ) 検診は、追跡期間内に少なくとも3回行う。ある時点で "positive" と診

断された個体であっても、続けて、つぎの検診時点で検診する。

表1は、以上のような追跡調査からの典型的データである。通常の発病率調査データと様相が異なることに注意したい。すなわち、通常は固定2時点での検診で、しかも2回目の検診受診者は初期検診で negative であった個体のみに限られる。しかしながら、Diagnostic tests が示す "positive" が真の発病と常に一致するとは限らない。いいかえれば、真に発病していても Diagnostic tests の結果は negative (偽陰性)でありえるし、また、真に非発病であっても positive (偽陽性)でありえる。通常の調査法では、この相違を正しく把握して、真の発病率を推定することが難しい。これを可能にするために(ハ)のような調査が工夫されたわけである。検診を3回以上行うため、いくつかの時点で受診しない個体が出現し、いわゆる "欠損データ" が多数生じるのが特徴である。

表1

Data, 1977年9月—1979年7月

No.	examination month					
	September	April	October	January	March	July
	1977	1978	1978	1979	1979	1979
1	—		—			
2		—		+		
3			—		—	—
4			—			
5		+				
6		+	+		+	
7	+		—	+		

+ : tested positive, — : tested negative

2. Diagnostics tests の影響を調整しない発病率の推定

Diagnostics tests の影響を調整しないということは、完全な診断を行っていると考えていることに相当する。いいかえれば、

発病率 = Diagnostics tests の "positive" 率

を意味する。この場合の発病率の推定について、2つの方法を考察する。

2.1 Person-months 法

追跡調査から発病率を推定する方法として、つぎのような公式が多くのテキストで与えられ、person-months 法をよばれている。

$$\text{月間発病率} = \frac{\text{\# of new incident cases}}{\text{person-months followed-up}}$$

$$\text{年間発病率} = \text{月間発病率} \times 12$$

表1のような"不完全"データにこの公式の適用は妥当ではない。このことを次の2つの例で明らかにする。

[例2.1]

初回検査結果が negative の個体だけが follow-up されたとする図1の場合について考える。図より

図1

個体	月				person-months
	0	1	2	3	
(a)	—			—	3
(b)	—	—	—		2
(c)	—	+			1
(d)	—			+	3
合計					9

$$\text{月間発病率} = \frac{2}{9} = 0.22$$

を得る。しかるに、progressive disease が対象とされているから、(d)はつぎの3つの 場合のいずれでもありえる。

(d 1)	-	-	-	+
(d 2)	-	-	+	+
(d 3)	-	+	+	+

欠落

postive になれば follow-up が打ち切られると、例に想定すれば、person-months は(d 1), (d 2), (d 3)の各場合に異なる値をとる。すなわち、(d)の様なパタンの欠落データに person-months 法を適用するには問題があることが分かる。なお、このとき、図 1 (d)の person-monthsを $3 \rightarrow 3/2$ に修正しておき person-months 法を適用すると、次節の条件つき最尤法から求める推定値にかなり近い値が得られる。

〔例 2. 2〕

つぎに図 2 に示されたような "late arrival" がある場合について考えてみよう。

図 2

	月			
個体	0	1	2	3
(a)	-			-
(b)		-	-	
(c)		-	-	+
(d)			-	
(e)			+	

通常の person-months は各個体の第 1 回目の検査時点を基点として計算される。

たとえば、図2. 個体(b)の person-months は1とされる。しかしながら、対象とされているのが progressive disease であることに思いを馳せるなら、たとえば(b)は調査開始時点で negative であることは明らかである。よって基点を調査開始時点、すなわち図2の $t = 0$, において person-months 法を適用してもよいはずである。なお、このように基点を定めるときは、唯1回しか受診しなかった個体(e)の様なデータも考慮にいれなければ片手落ちになる。

Person-months 法は、Yanagawa, T. and Kasagi, F. (1984)で指摘されているように、いくつかの数学的な仮定と、追跡調査の1つの典型的データをモデルとして導かれたものであって、これを公式として一人歩きさせると、上の例の様なパタンのデータに対して、その適用に苦慮することになる。出発点にもどって、発病率の推定法を第1歩から考え直す方がよい。

2.2 最尤法

Progressive disease に関する発病率の推定は、生存データ解析におけるハザード関数、あるいは同じことであるが生存関数の推定問題としてとらえることができる。いま、調査開始時点 t_0 を $t_0 = 0$ と定め、 T を任意の個体が発病するまでの期間(月)とする。

$$F(t) = \Pr(T \leq t \mid T > 0)$$

$$P_0 = \Pr(T \leq 0)$$

とおく。 P_0 は調査開始時点での有病率、 $F(t)$ は、区間 $(0, t]$ での発病率を表す。

また、時点 t での有病率は

$$\Pr(T_t) = 1 - (1 - P_0) [1 - F(t)]$$

で与えられている。時刻 t でのハザード関数を $\lambda(t)$ とおく、

$$F(t) = 1 - \exp \left\{ - \int_0^t \lambda(x) dx \right\}$$

特に、person-months 法が仮定しているように、ハザードが時刻に依存せず一定、すなわち $\lambda(t) = \lambda$ と仮定すれば、

$$F(t) = 1 - \exp(-\lambda t) \quad : \quad \text{指数分布} \\ = \lambda t \quad (\lambda \text{ が十分小のとき})$$

このことより、 $F(t)$ 、あるいは λ を推定すれば、発病率が推定でき、さらに P_0 を推定すれば、有病率も推定できる。 λ の推定には、条件つき最尤法と無条件最尤法が適用できる。その相違は、例えば図2の個体(a)，(e)のデータの尤度を前者は

$$(a) \Pr(T > 3 \mid T > 0) = 1 - F(3)$$

$$(e) \Pr(T \leq 2 \mid T > 0) = F(2)$$

とするのに対して、後者は

$$(a) \Pr(T > 3) = (1 - P_0)(1 - F(3))$$

$$(e) \Pr(T \leq 2) = 1 - (1 - P_0)(1 - F(2))$$

とする。条件つき最尤法からは λ だけが推定できる。無条件最尤法からは λ と P_0 が推定できる。

3. Diagnostic tests の影響

D_t , E_t をつぎのように定義する

$$D_t = \begin{cases} 0 : \text{時刻 } t \text{ で真に非発病} \\ 1 : \text{時刻 } t \text{ で真に発病} \end{cases}$$

$$E_t = \begin{cases} 0 : \text{時刻 } t \text{ で negative と診断} \\ 1 : \text{時刻 } t \text{ で positive と診断} \end{cases}$$

このとき、Diagnostic tests の挙動は

$$\alpha = \Pr(E_t = 1 \mid D_t = 0) : \quad \text{偽陽性率}$$

$$\beta = \Pr(E_t = 0 \mid D_t = 0) : \quad \text{偽陰性率}$$

で把握できる。これらは、数理統計学における第I種、第II種の誤りの確率に対応する。

つぎの記号を導入して固定2時点 $t_0 = 0$, $t_1 = 1$ での検診の場合のDiagnostic tests の影響に評価を行う。すなわち

$P_0 = \Pr(D_0 = 1) :$ $t = 0$ での真の有病率

$P = \Pr(E_0 = 1) :$ $t = 0$ でのみせかけの有病率

$I_0 = \Pr(D_1 = 1 | D_0 = 0) :$ 真の発病率

$I = \Pr(E_1 = 1 | E_0 = 0) :$ みせかけの発病率

$R_0 = \Pr(D_1 = 0 | D_0 = 1) :$ 真の逆転率

$R = \Pr(E_1 = 0 | E_0 = 1) :$ みせかけの逆転率

Progressive disease の仮定より $R_0 = 0$ であることに注意する。

Yanagawa-Gladen(1984)は、 P, I, R を α, β, P_0, I_0 で表現する方法を開発し、 $\alpha = \beta = 0.05$ ぐらいでも、 P と P_0, I と I_0 の間に本質的な差が生じる場合があることを指摘している。とくに、 P_0 と I_0 の値がなんであれ、

不等式

$$\alpha \leq P \leq 1 - \beta$$

$$\alpha \leq I \leq 1 - \beta$$

$$\beta \leq R \leq 1 - \alpha$$

が成り立つ。例えば、真の発病率が $5/10000$ で $\alpha = 0.05$ のとき、この不等式より $I \geq 0.05 = 5/100$ 。したがって $\alpha = 0.05$ という極めて精度の良い diagnostic tests を用いても、真の発病率を 100 倍以上過大推定することになる。

4. 偽陰性を推定する「真の罹患率」の推定

2.2 節の生存関数の推定問題に、前節の偽陰性、偽陽性の考えを導入して、尤度関数を構成する。属性変数を考慮しない場合とする場合の2つの方法を考慮する。

4.1 属性変数を考慮しない場合 YKY法

Yangawa, T, Kasagi, F and Yoshimura, T. (1984) で開発された方法である。偽陽性率 $= 0$ とされている。この方法の特徴は、(1) 真の有病率、真の発病率、及び偽陰性率を同時に推定する (2) 無条件尤度法による。すなわち唯1回受診者のデータをも用いて、情報を最大に利用しつつ発病率推定を行う。

4.2 属性変数を考慮する場合 拡張YKY法

上述のYKY法は、層別を行って各層内で、発病率及び Diagnostic tests の影響を均一にすることができるとき、有効である。しかしながら、例えば、年齢の変化が発病率を急激に変えるとき、層の巾を小さくしなければ層内での均一性は得られないが、他方このとき層内のデータ数が少なくなって、発病率の推定ができない場合が生じる（少なくとも発病者が1人含まれていなければ、意味ある推定は出来ない）。このため、性や年齢など、いわゆる共変量を回帰式を利用して、発病率、有病率及び偽陰性率にとり込むことにする。

いま、 z を年齢や性を表す共変量ベクトルとして

$$P_0(z) = P_z(T \leq 0)$$

$$F_z(t) = P_z(T \leq t \mid T > 0)$$

β_z : 共変量の値が z の個体に対する偽陰性率

(ただし、偽陽性率 = 0 としておく)

とおき、YKY法の理論を適用する。例えば、共変量の値が z で、 $t = 0$, t_1 , t_2 で受診し、結果が $-$, $+$, $-$ であった個体の尤度は

$$P_z(E_0=0, E_t=1, E_2=0)$$

$$= P_z(0 < T \leq t_1) \beta_z(1 - \beta_z) + P_z(T \leq 0) \beta_z^2(1 - \beta_z)$$

$$= [1 - P_0(z)] F_z(t) \beta_z(1 - \beta_z) + P_0(z) \beta_z^2(1 - \beta_z)$$

のように構成できる。Yanagawa, T. and Kasagi, F. (1986) は

z : 年齢

$$P_0(z) = 1 - \exp(-\theta_3 z^{\theta_4})$$

$$F_z(t) = 1 - \exp(-\theta_1 z^{\theta_2} t)$$

$$\beta_z(t) = (1 + \exp(\theta_5 z^{\theta_6}))^{-1}$$

とおき、拡張YKY法を Gutemalan onchocerciasis の部分的データに適用した。

表2は、パラメータ $\theta_1, \theta_2, \dots, \theta_6$ の推定値。表3は、これらの推定値を代入して求めた偽陰性率、有病率および年間発病率である。

表 2 パラメータの推定値

パラメータ	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6
推定値 (S D)	0.05116 (0.0326)	0.06213 (0.2666)	0.01892 (0.0101)	1.22262 (0.5231)	0.11804 (0.0589)	0.80704 (0.1455)

表 3 年齢別偽陰性率、有病率、年間発病率

年齢	β_z	Po (z)	年間発病率
5	0.39	0.13	0.055
10	0.32	0.27	0.057
15	0.26	0.40	0.059
20	0.21	0.52	0.060
30	0.14	0.70	0.061
40	0.09	0.82	0.062

このように拡張 Y K Y 法では、Diagnostic tests の影響を取り除いた真の発病率、有病率の推定ができる。同時に偽陰性率も推定できる。拡張 Y K Y 法の難点は、尤度関数が未知パラメータの高度な混合で与えられるため、極大点が数個存在する可能性があること、また、尤度関数の値の変化は微少であってもパラメータの値は大きく変わりえることである。このため、計算にかなりの熟練を要し、また良い初期値を必要とする。この難点を補うため、Yanagawa Kasagi (1986) は部分尤度法を開発した。この方法を用いるとパラメータの推定値を求める計算は格段に省略される。他方、この方法は部分データから、1 ～ 2 個のパラメータを推定し、その推定値を既知として残りのパラメータを推定するという手続きをとるため、まだ解明されていない何らかの事情のため、残りのパラメータの推定が不可能になる場合がかなり生じる。今後さらに研究を続け、より安定した簡易な推定法を確立したい。

参考文献

- Yanagawa T. and Gladen B. (1984). Estimating incidence and remission from a diagnostic test. *Am. J. Epidemiology* 119, 1015-23.
- Yanagawa T., Kasagi F. and Yoshimura T. (1984). A method for estimating incidence rates of onchocerciasis from skin-snip biopsies with consideration of false negative. *Biometrics* 40, 301-11.
- Yanagawa T. and Kasagi F. (1985). Estimating prevalence and incidence of disease from a diagnostic test. In *Recent Development in Statistical Inference and Data Analysis* (K. Matushita, Editor) 783-801, North Holland, Amsterdam.
- Yanagawa T. and Kasagi F. (1986). Estimating the true disease rate with consideration of fallibility of diagnostic tests. Abstract, The VIIIth International Biometric Conference.

[Received January 1987]