

ケースコホート研究の理論と統計手法

野間 久史[†]

(受付 2013 年 9 月 7 日；改訂 12 月 9 日；採択 12 月 24 日)

要 旨

臨床研究・疫学研究におけるコホート研究では、しばしば規模の大きな集団を長期間に渡って追跡する必要があり、コホートにおけるすべての対象者において、完全な共変量情報を収集するためには、膨大な費用と労力が必要とされる。ケースコホート研究は、統計的な精度を保持しつつ、このような共変量測定のコスト・労力を節減することを目的とした研究デザインである。本稿では、このケースコホート研究の理論と統計解析の方法論についての解説を行う。特に、曝露効果の指標の推定における方法論とその歴史的な経緯についての解説を行い、古典的な分割表の解析手法から、交絡調整の方法、Cox 回帰の修正部分尤度による推定方法についての一連の方法論を紹介する。また、近年の研究で大きく発展した、全コホートの補助情報を利用した推定量の構成方法についての解説を行う。事例として、Wilms 腫瘍の臨床試験データをもとにした、仮想的なケースコホート研究のシミュレーション実験の結果を示す。

キーワード：疫学，ケースコホート研究，2 段階デザイン，重みつき推定方程式，セミパラメトリック推測，不完全データ。

1. はじめに

臨床研究・疫学研究におけるコホート研究では、しばしば規模の大きな集団を長期間に渡って追跡する必要があり、コホートにおけるすべての対象者において、完全な共変量情報を収集するためには、膨大な費用と労力が必要とされる。米国で 1970～80 年代に行われた Multiple Risk Factor Intervention Trial (MRFIT; MRFIT Research Group, 1982) を例に挙げると、この研究では、12,866 名の男性に冠動脈心疾患の予防プログラムがランダムに割り付けられ、平均 7 年間に渡る追跡が行われている。この研究でかかった費用は、1980 年代当時で 1 億ドル以上といわれており、特に、血清サンプルの生化学的な分析や栄養調査に膨大な費用がかけられたという報告がある (Prentice, 1986)。一方で、対象となる冠動脈心疾患による死亡は、239 名 (1.9%) にしか観測されておらず、MRFIT 試験の費用のほとんどは、追跡期間中にイベントを発生しなかった 12,627 名に費やされたこととなる。統計的な評価に寄与するのは、基本的にはイベントを起こした対象者のみであるため、この膨大な費用のほとんどは、残念ながら、予防プログラムの効果の評価には貢献しなかったといえる。

このような問題を解決するため、疫学理論の研究は、1970 年代以降に大きな発展を遂げ、古典的なコホート研究、ケースコントロール研究に加えて、さまざまな研究デザインが提案されてきた。最も代表的かつ、実践でも広く用いられているものは、ネステッドケースコントロール研究 (nested case-control studies; Oakes, 1981; Thomas, 1977) である。このデザインは、前向

[†] 統計数理研究所：〒190-8562 東京都立川市緑町 10-3

きのコホート研究の中から、イベントを発生したケースと、それに時点マッチングしたコントロールをサンプリングし、高価な共変量の情報を、選択されたケース・コントロールにのみ測定するというデザインである。MRFIT 試験の例で言えば、仮に1:4のコントロールのマッチングを行ったとしても、ケース・コントロールの総数は、たかだか全体の10%程度であり、研究のコストは大幅に節減することができたことになる。理論的にも、「コホート研究の中で行うケースコントロール研究」と考えればよいから、古典的なケースコントロール研究の統計手法をそのまま利用することができ、疫学者にも馴染みやすいデザインであるといえる。

一方、MRFIT 試験のような大規模な前向き研究では、しばしば単一のアウトカムだけではなく、複数のアウトカムについての評価が並行して行われることが多い。このような場合、ネステッドケースコントロール研究では、サンプリングデザインの構造上、アウトカムごとに別々のコントロールグループが必要となる。すなわち、アウトカムの数が増えるほど、それに比例して、研究のコストは大きくなる。ケースコホート研究(case-cohort studies)は、このような難点を解決するために Prentice (1986) によって提案された研究デザインである。このデザインでは、共通の1つのコントロールグループを用いて、複数のアウトカムについてのリスク評価を行うことができ、より効率的な評価が可能になる。

本稿では、このケースコホート研究の理論と統計解析の方法についての解説を行う。近年までの理論疫学・生物統計学における、ケースコホート研究の方法論の発展はめざましく、その対象は広範なものに及ぶが、本稿では、特に、基本的なモデルを利用した曝露効果の推定の方法についての解説を与えることを目的とする。2節では、ケースコホート研究の枠組みと曝露効果の指標の推定方法についてまとめ、2値結果変数を対象とした設定での交絡調整の方法を紹介する。3節において、実践でより一般的なものとなっている、生存時間解析の方法についての研究を概説する。特に、実践における標準的な方法となっている Cox (1972) の比例ハザード回帰モデルでの解析手法を、近年までの新しい研究成果を含めて紹介する。4節では、米国の National Wilms Tumor Studies の事例に基づく、シミュレーション実験を行い、近年までに提案された推定方法の性能についての評価を行う。

2. ケースコホート研究のデザインと曝露効果の推定

2.1 ケースコホート研究のデザイン

前向きのコホート研究の中でケースコントロール研究を行うネステッドケースコントロール研究は、古典的なケースコントロール研究の変法と見なすことができる。ケースコホート研究も、同じく、ケースコントロール研究の変法であり、コホート研究の中で、イベントを発生したケースと、その比較対照となるコントロールのサンプリングを行うデザインである。ただし、従来のケースコントロール研究とは大きく異なり、ケースコホート研究は、コントロールグループを「イベントを発生していない非ケース」ではなく「ケースも含めたコホートすべての構成員」からランダムに選択するという研究デザインである。

例として、MRFIT 試験をもとにした、仮想的なケースコホート研究を行うことを考えよう。まず、「ケースグループ」は、コホート内のすべてのケース(239名)から構成する。これは、従来のケースコントロール研究と同じである。これに対して、コントロールグループは、先述の通り、MRFIT 試験の対象者集団全体からのランダムサンプルとして構成する。すなわち、コントロールグループは「コホート内の非ケースを代表するグループ」ではなく、「コホート全体を代表する、コホート内のサブグループ」となる。このため、このコントロールグループは、一般的に、サブコホート(subcohort)と呼ばれる。ここでは、MRFIT 試験全体の5%にあたる643名の対象者をサブコホートとして選択することを考える。サンプリングデザインの模式図を図1に

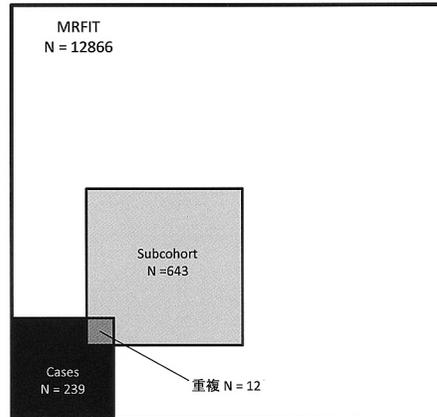


図 1. MRFIT 試験に基づく仮想的なケースコホート研究. 高価な共変量の測定は、12,866 名の対象者のうち、選択されたケース・サブコホートの 870 名 (6.8%) にのみ行えばよい。

示す。高価な共変量の測定は、このサンプリングされた 870 名 (全体の 6.8%) にのみ行えばよいこととなり、大幅なコストの節減が可能であることがわかる。

曝露効果の評価と、具体的な推定効率については、次節以降で示すこととして、この研究デザインでは、

- 複数のアウトカムについての評価に関心がある場合にも、このコントロールグループは、特定のアウトカムに対して選ばれたものではない。したがって、1つのサブコホートを、複数のアウトカムに対する共通のコントロールグループに利用することができる。アウトカムごとにコントロールを別々に選択する必要があるネステッドケースコントロール研究に比べて、研究全体のコストを節減することができる。
- サブコホートは、コホートからのランダムサンプルとなっているので、これから、母集団における共変量の分布 (例えば、遺伝子型の頻度など) を直接推定することができる。ネステッドケースコントロール研究では、これは、必ずしも容易ではない。

現在でも、実践においては、ネステッドケースコントロール研究のほうが圧倒的に適用事例は多いが、近年になって、これらのケースコホート研究の利点が広く知られるようになり、疫学者の間でも徐々に普及しつつあるようである (Breslow et al., 2009b)。代表的な事例として挙げられるのが、米国で行われた Atherosclerosis Risk in Communities (ARIC) 研究 (The ARIC Investigators, 1989) であり、この研究では、保存された生体試料をもとに、遺伝子型や炎症バイオマーカーの測定を行い、冠動脈心疾患や脳卒中など複数のアウトカムについてのリスク評価を行うケースコホート研究が行われている (Afshar-Kharghan et al., 2004; Ballantyne et al., 2005; Ballantyne et al., 2004; Li et al., 2000; Rasmussen et al., 2001)。また、近年では、分子遺伝学などの急速な発展により、臨床研究・疫学研究でも、高価な遺伝子情報の測定が行われる研究が増加している。これらの研究では、ケースコホート研究が重要な役割を果たしている (Kulathinal et al., 2007; Lin, 2014)。

Prentice (1986) による提案以前にも、Kupper et al. (1975)、Miettinen (1982) によって、同様の研究デザインが提案されており「hybrid epidemiologic designs」「case-base studies」と呼ばれているが、現在では、ケースコホート研究という名称が一般的である。ネステッドケースコホート

表 1. MRFIT 試験に基づく仮想的なケースコホート研究.

(a) MRFIT 試験の結果 (7年目追跡時点).

	冠動脈心疾患による死亡		合計	
	あり	なし		
介入群	115 (1.8%)	6,313	6,428	$\widehat{RR} = 0.929$ (95% C.I.: 0.722-1.194)
対照群	124 (1.9%)	6,314	6,438	$\widehat{OR} = 0.927$ (95% C.I.: 0.717-1.198)

(b) 仮想的なケースコホート研究 (サブコホートのサンプリング確率 5%).

	介入群	対照群
ケースグループのみに所属するケース	109	118
ケースグループ・サブコホートに重複するケース	6	6
非ケース (サブコホートにのみ所属)	315	316
サブコホートの合計	321	322

$$\widehat{RR}_{ML} = 0.932 \text{ (95\% C.I.: 0.695-1.249)}, \widehat{OR}_{ML} = 0.931 \text{ (95\% C.I.: 0.692-1.254)}$$

ロール研究との実践的な面でのより詳細な比較については, Wacholder(1991)に詳しい.

2.2 曝露効果の指標の推定

2.2.1 分割表の解析

表 1 に, MRFIT 試験の結果の要約(ただし, 打ち切りは無視している)と, 前節における仮想的なケースコホート研究の内訳を示している. コントロールグループを「コホートからのランダムサンプル」として選んでいるため, 当然の問題として, 同一人物がケースグループとコントロールグループに重複して所属することがあり得る. ケースコントロール研究は, もともと「ケース・コントロールの比較により, 曝露効果の評価を行うもの」というデザインであったため, 一見, 前提からしておかしなデザインであるが, 実際には, これによって, 正しく相対リスクの推定を行うことができる.

表 2 にケースコホート研究のデータを模式的に示す. ケースは, 全コホートにおけるケースから全数サンプリングされるものと仮定している. 仮に, 全コホートにおける曝露群, 非曝露群の総数を ν_1, ν_0 で表すと, 全コホートにおけるリスク比 (risk ratio; RR) の推定量は, $\widehat{RR} = A_1\nu_0/(A_0\nu_1)$ と表される. ケースコホート研究では, サブコホートが全コホートからのランダムサンプルであることを利用して, この ν_0/ν_1 をサブコホートにおける曝露群・非曝露群の人数の比 N_0/N_1 で推定することができる. すなわち, \widehat{RR} の ν_0/ν_1 を N_0/N_1 に置き換えることによって, リスク比の経験的な推定量 (Kupper et al., 1975)

$$\widehat{RR}_E = \frac{A_1 N_0}{A_0 N_1}$$

を得ることができる. 古典的なケースコントロール研究では, ケース・コントロールのサンプリング確率の比などの付加的な情報が得られない限り, オッズ比しか推定することができず, 「対象となる疾病が稀である」という前提のもとで, 近似的にリスク比を推定することしかできないのが難点とされていたため, ケースコホート研究が提案された当初は, こうして直接リスク比を推定することができる点は大きな利点と考えられた. ただし, この \widehat{RR}_E は全コホートのリスク比の一致推定量になるが, 漸近有効性は持たないことが知られている.

表 2. ケースコホート研究のデータのモデル(ただし, ケースが全数サンプリングされた場合).

	曝露群	非曝露群	合計
ケース	A_1	A_0	M_1
ケースグループのみに所属	A_{11}	A_{01}	M_{11}
ケースグループ・サブコホートに重複して所属	A_{10}	A_{00}	M_{10}
非ケース (サブコホートにのみ所属)	B_1	B_0	M_0
サブコホートの合計	N_1	N_0	N

漸近有効なリスク比の最尤推定量は, Sato(1992b)によって与えられており, Kupper et al. (1975)の経験的な推定量におけるサブコホートの合計 N_0, N_1 を

$$N_0^* = \frac{A_0 M_{10}}{M_1} + B_0, N_1^* = \frac{A_1 M_{10}}{M_1} + B_1$$

に置き換えることによって得られる. すなわち,

$$\widehat{RR}_{ML} = \frac{A_1 N_0^*}{A_0 N_1^*}$$

である. 理論的には, この N_0^*/N_1^* のほうが, 直感的な推定量 N_0/N_1 よりも, 安定した推定量となる. Sato(1992b)の推定量は, 表2のデータに適当な多項分布モデルを仮定したもとの最尤推定量として導出される. 対数リスク比の漸近分散の推定量は, デルタ法により,

$$\begin{aligned} \widehat{V}[\log(\widehat{RR}_{ML})] &= \frac{1}{A_1} + \frac{1}{A_0} + \left(\frac{M_{11} - M_{10}}{M_1} \right) \left(\frac{1}{N_1^*} + \frac{1}{N_0^*} \right) \\ &\quad - \left(\frac{1}{A_1} + \frac{1}{A_0} \right) \left(\frac{1}{N_1^*} + \frac{1}{N_0^*} \right)^2 \left(\frac{M_{11} M_{10}}{M_1^2} \right) \end{aligned}$$

となる. したがって, リスク比の近似的な 95%信頼区間は,

$$\exp \left[\log(\widehat{RR}_{ML}) \pm 1.96 \sqrt{\widehat{V}[\log(\widehat{RR}_{ML})]} \right]$$

で得られる. 先ほどの MRFIT 試験の仮想的な事例においては, $\widehat{RR}_{ML} = 0.932$ (95%信頼区間: 0.695–1.249)となる. 全集団におけるリスク比は, $\widehat{RR} = 0.929$ (95% 信頼区間: 0.722–1.194)なので, これをほぼ偏りなく推定することができている. また, 対数リスク比のスケールでの漸近相対効率率は 73.7%であり, コストの節減の割には, 効率の損失はかなり小さいことがわかる. ケースのサンプリング確率が 100%でない場合には, もう少し複雑な推定量になるが, 関心のある読者は, Sato(1992b, 1994)をご参照いただきたい.

ケースコホート研究では, 同様にオッズ比(odds ratio; OR)を推定することも可能である. Prentice(1986)は, ケースグループとサブコホートの重複を無視して, 表2を単純な 2×2 表と見なしたときのオッズ比

$$\widehat{OR}_{ML} = \frac{A_1 B_0}{A_0 B_1}$$

が, 母集団のオッズ比の最尤推定量になることを示している. これは, ケースコホート研究のサンプリングデザインが, 全コホートからのケース・非ケースからのサンプリングと見なした場合, 古典的なケースコントロール研究のサンプリングと等価になることから説明することができる(Noma and Tanaka, 2013). 言い換えると, ケースコントロール研究における解析手法と理論的な結果は, すべてケースコホート研究に適用することができる. このことから, このオッ

ズ比の推定量も、漸近有効性を持つ一致推定量になる (Breslow et al., 2000; Prentice and Pyke, 1979). 対数オッズ比の漸近分散の推定量も、

$$\widehat{V}[\log(\widehat{OR}_{ML})] = \frac{1}{A_1} + \frac{1}{A_0} + \frac{1}{B_1} + \frac{1}{B_0}$$

と得られる. MRFIT 試験の仮想的な事例においては, $\widehat{OR}_{ML} = 0.931$ (95%信頼区間: 0.692–1.254)となる. これも, 全集団でのオッズ比は $\widehat{OR} = 0.927$ (95%信頼区間: 0.717–1.198)であり, 漸近相対効率も 74.6%となっている.

曝露効果の指標の推定効率は, サブコホートのサイズに比例することになるが, これは, 研究の計画段階で, 適切に評価する必要がある. リスク比を指標にする場合には, 複雑な議論が必要になるが, オッズ比を指標にする場合には, 従来のケースコントロール研究の検出力・サンプルサイズの公式 (例えば, Wickramaratne, 1995) をそのまま利用することができる. ただし, ケース・サブコホートの重複は調整する必要がある.

2.2.2 層別解析と回帰モデルによる交絡調整の方法

前節の MRFIT 試験の仮想的な事例のように, ケースコホート研究は, ランダム化比較試験にも適用することができるデザインではあるが, 実践では, 観察研究の適用例のほうが圧倒的に多い. 観察研究においては, 治療効果・曝露効果を偏りなく推定するために, 交絡の調整が不可欠となるため, 層別解析や回帰モデルによる解析を用いるのが一般的である.

層別解析におけるリスク比の推定においては, Kupper et al. (1975) が要約リスク比の推定量を与えているが, Greenland (1986) によって, これには漸近的なバイアスがあることが示されている. 現在のところ, 妥当なリスク比の推定量は, 標準化法については Greenland (1986), 共通リスク比の Mantel-Haenszel 推定量については Sato (1992a) の方法がある. オッズ比の推定では, 前節に記した通り, ケースコントロール研究における方法がそのまま適用できるため, 例えば, 最尤法, 重みつき最小二乗法, Mantel-Haenszel 法などによる共通オッズ比の推定方法 (Breslow and Day, 1980) が利用できる.

回帰モデルによる交絡調整については, リスク比の推定においては, 対数リンクを用いた 2 項回帰モデルにおける擬似尤度法が Schouten et al. (1993) によって提案されている. しかし, この方法は, 標準的なソフトウェアには実装されておらず, 実践でも普及していない.むしろ, 実践で広く用いられているのは, ロジスティック回帰モデルである. 前節までの議論の通り, ケースコホート研究から得られたデータは, ケースコントロール研究と見なして解析することができるため, ロジスティック回帰によって偏りなく母集団のオッズ比を推定することができる. ケースコントロール研究の理論から, 切片項は識別不可能となるが, それ以外の回帰パラメータについては, 漸近有効性を持つ一致推定量が得られる (Breslow et al., 2000; Prentice and Pyke, 1979).

実践においては, ほとんどのケースコホート研究が稀な疾病を対象としていることもあり, 現在では, このロジスティック回帰による方法が, 最も広く用いられている. 一方で, van der Klauw et al. (1999) など, 直接, リスク比の推定量を利用した事例の報告もある.

また, 3 節で述べる通り, 近年では, ケースコホート研究を 2 段階デザインと見なして, 全コホートの情報を利用した, 推定量の改善方法が多く議論されている. ケースコホート研究のサンプリングは, ケースコントロール研究と等価であることから, これを 2 段階デザインと見なした場合, 2 段階ケースコントロール研究 (White, 1982) に一致する. すなわち, 2 段階ケースコントロール研究で開発された解析方法は, すべてそのままケースコホート研究の解析に利用することができる (Noma and Tanaka, 2013). 例えば, 2 段階ケースコントロール研究で発展した, 漸近有効なセミパラメトリック最尤法 (Breslow et al., 2003; Lawless et al., 1999; Scott and

Wild, 1997)などもそのまま適用することができる。また、この枠組みのもとでは、解析対象集団を全コホートと見なして、不完全データの枠組みによって解析を行えばよいことになり、任意のリンク関数を用いた2項回帰モデルでの解析が可能である。すなわち、「対象となる疾病が稀である」という仮定がなくても、直接、リスク差(恒等リンクを使用)やリスク比(対数リンクを使用)を推定することができる。詳細については、3.3~3.5節の解説をご参考いただきたい。

3. Cox 比例ハザード回帰モデルによる解析

ケースコホート研究が有効なのは、高価な測定を必要とする共変量がある場合であり、これまでの事例にも、MRFIT 試験のような大規模かつ長期間の追跡を伴うコホートを元にしたものが多い。したがって、実際には、打ち切りを考慮した生存時間解析の方法を用いる研究が大部分を占めており、これらの研究では、Cox(1972)の比例ハザード回帰モデルを用いた解析が中心となっている。本節では、ケースコホート研究における、Cox 回帰による曝露効果の推定方法について解説する。

3.1 修正部分尤度によるハザード比の推定

本節では、打ち切りを伴うイベントまでの時間をアウトカムとした議論を行うため、2節とは異なる表記法を用いる。全コホートは、母集団からのランダムサンプリングによって得られた N 人の対象者によって構成されるものとして、 n_1 人の対象者にイベントが観測され、サブコホートは、コホートからランダムに選ばれた n_0 人の対象者によって構成されるものとする。関心のあるイベントが起こるまでの時間を T 、全コホートにおける対象者すべてで測定される共変量を X 、ケースコホート研究にサンプリングされた対象者のみに測定される共変量を Z として、Cox(1972)の比例ハザード回帰モデル：

$$\lambda(t|X, Z) = \lambda_0(t) \exp(\beta_1 X + \beta_2 Z)$$

を仮定する。関心のあるパラメータは、回帰係数 $\beta = (\beta_1, \beta_2)$ である。この設定では、イベントが観測された対象者がケースであり、追跡期間中にイベントが観測されず、打ち切りとなった対象者が非ケースとなる。

当然ながら、ケースコホート研究にサンプリングされた対象者は、イベントの有無に依存して選択されているため、母集団からのランダムサンプルと見なすことはできない。したがって、Cox(1972, 1975)の部分尤度による回帰パラメータの推定量は妥当なものとならない。そこで、いくつかの修正部分尤度に基づく推定方法が提案されている。ケースコホート研究が提案された当初は、これらの推定量の妥当性・最適性についての理論も発展途上であったため、試行錯誤的にさまざまな方法が提案されたが、Barlow et al.(1999)は、これらの方法が、個人ごとに適当な重みを付与した擬似尤度関数

$$\mathcal{L}(\beta) = \prod_{t_j} \left[\frac{\exp(\beta_1 X_j + \beta_2 Z_j)}{\sum_{k \in \mathcal{R}(t_j)} \exp(\beta_1 X_k + \beta_2 Z_k) w_k(t_j)} \right]$$

に基づく推定方式と見なすことができるとして、比較・評価を行っている。ここでは、一般性を失うことなく、 $t_1 < t_2 < \dots < t_{n_1}$ はイベント発生時点、 $\mathcal{R}(t_j)$ は時点 t_j におけるリスクセット、 $w_i(t_j)$ は時点 t_j における対象者 i の重みである。本稿でも、この解釈に従って、これらの方法の解説を行うこととする。それぞれの方法で提案されている重みを表3にまとめる。Prentice(1986)は、各イベント発生時点におけるサブコホートの対象者のみを、擬似尤度の分母におけるリスクセット $\mathcal{R}(t_j)$ とする重みを用いている。つまり、サブコホートに含まれなかったケー

表 3. 修正部分尤度による Cox 回帰の重みつけの方法.

	Prentice (1986)	Self and Prentice (1988)	Barlow et al., (1999) [†]	Kalbfleisch and Lawless (1988)
非ケース (サブコホートに所属)	1	1	N/n_0	$1/\pi$
サブコホートに所属するケース				
イベント発生まで	1	1	N/n_0	1
イベントが発生した時点	1	1	1	1
サブコホートに所属しないケース				
イベント発生まで	0	0	0	1
イベントが発生した時点	1	0	1	1

[†]Barlow (1994) は、これに類似した時間依存性の重みを提案している.

π : サブコホートのサンプリング確率.

スは、自身のイベント発生時点でのみ、擬似尤度に貢献する。Self and Prentice (1988) は、これを少し修正して、サブコホートに含まれなかったケースは、自身のイベント発生時点でも $\mathcal{R}(t_j)$ に含めないという方法を与えている。Kalbfleisch and Lawless (1988) は、すべてのケースを重み 1 でリスクセット $\mathcal{R}(t_j)$ に含め、残りのサブコホートの非ケースについては、サブコホートのサンプリング確率の逆数を重みとする方法を提案している。Barlow (1994) は、サブコホートの対象者についての重みを「時点ごとの全コホートの At risk な対象者数」と「サブコホートの At risk な対象者数」の比とする、時間依存性の重みを提案している。Barlow et al. (1999) は、これを単純化して、サブコホートのサンプリング割合の逆数で近似する方法を提案している。

これらの異なる重みつけの方法は、Barlow et al. (1999), Onland-Moret et al. (2007), Petersen et al. (2003) で比較されている。シミュレーション実験による数値的な比較では、Onland-Moret et al. (2007) は、サブコホートのサイズが十分に大きい場合 (この報告では、全コホートの 15% 以上) には、いずれの方法も似たような推定値を与え、またこれは、全コホートにおける最大部分尤度推定値に概ね近い値をとったと報告している。一方、サブコホートのサイズが小さい場合には、Prentice (1986) の方法が良好な性能を示したとしている。Petersen et al. (2003) でも、概ね同様の結論が与えられており、特に、Prentice (1986), Kalbfleisch and Lawless (1988) の推定量が良い挙動を示したと結論している。

3.2 分散の推定量

3.1 節で紹介した方法は、通常の部分尤度に基づく推定ではなく、擬似尤度に基づく推定を行うため、回帰パラメータの分散の推定に、モデル分散を用いることはできない。特に、これらの方法は、擬似尤度の最大化におけるスコアの寄与を独立と見なすことができないため、マルチンゲール理論を直接的には適用することができず、漸近的な評価もより複雑なものとなる (Prentice, 1986; Self and Prentice, 1988)。

Prentice (1986) は、スコアの要素間の相関を修正した分散の推定量を提案している。Self and Prentice (1988) は、前節での擬似尤度推定量が、一致性と漸近正規性を満たすための条件を与えており、その分散の推定量も与えている。Lin and Ying (1993) は、より一般的に、共変量に欠測がある場合の Cox 回帰についての議論を行っており、その特別な例として、ケースコホート研究を挙げている。彼らは、Lin and Wei (1989), Reid and Crépeau (1985) による無限小ジャックナイフ推定量を用いることができることを示している。Barlow (1994) は、Lin and Ying (1993)

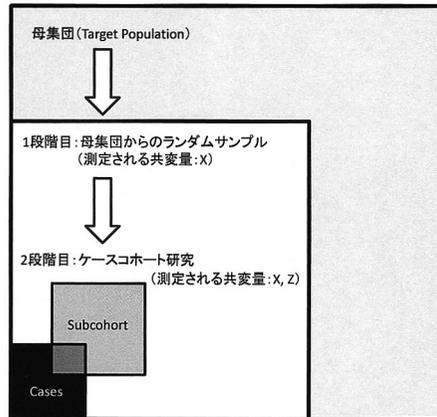


図 2. 母集団からの 2 段階サンプリングのモデル. 2 段階目でサンプリングされた対象者についてのみ, 高価な共変量 Z が測定される. 1 段階目のコホートを解析対象集団と見なすと, Z が部分的に MAR で欠測した不完全データと見なすことができる.

とは独立に, ケースコホート研究の分散の推定量を, このジャックナイフ法を用いて求めることを提案している. これらのロバスト分散の推定については, Therneau and Li(1999)に詳しい.

現在では, ケースコホート研究における Cox 回帰の方法は, 分散の推定も含めて, 標準的な統計ソフトウェアで普及しつつあり, 例えば, R では survival パッケージの cch で, 代表的な解析手法については実行することができる. SAS, S-PLUS などによる解析プログラムも, Barlow et al.(1999), Kulathinal et al.(2007), Langholz and Jiao(2007), Therneau and Li (1999) で与えられている.

3.3 重みつき部分尤度による推定方法

3.3.1 不完全データとしての定式化

ここまで述べてきた通り, ケースコホート研究の方法論の研究は, 当初は, ケースコホート研究で得られたサンプルのデータから, 曝露効果を偏りなく推定するための方法の開発が中心であった. 一方で, 実践においては, ケースコホート研究を行う母体となるコホートにおいて, ケースグループ・サブコホートに選択されなかった対象者にも, 部分的な共変量情報が(その多くは非ケースであるが)得られていることが多い. これらの情報を曝露効果の推定に活用すれば, 推定精度を高めることができるのではないかとするのは自然な発想であり, 2000 年代に入って, この副次的な情報を利用した解析方法の研究が大きく発展した.

全コホートにおける情報を解析に利用するにあたり, ケースコホート研究を, 母集団からの 2 段階のサンプリングに基づくデザイン (Zhao and Lipsitz, 1992) であると定式化するとわかりやすい. 図 2 は, 母集団から, ケースコホート研究のサンプルが得られるまでのサンプリングの過程を模式的に示したものである. すなわち, ケースコホート研究は, 母集団からのランダムサンプリングによって, 母体となるコホートが得られる過程を 1 段階目, そこからケース・サブコホートを選択する過程を 2 段階目と見なすことによって, 2 段階のサンプリングデザインと定式化することができる. 多くの場合, 1 段階目の全コホートで, 結果変数と, 測定に大きなコストを必要としない共変量 X (年齢, 性別, 人種など) は測定されており, 2 段階目で選択された対象者には, 付加的に高価な測定を必要とする共変量 Z の情報が測定される. つまり, ケー

スコホート研究から得られるデータは、1段階目の全コホートを解析対象集団とすると、2段階目で測定される共変量 Z が欠測している不完全データと見なすことができる。さらに、2段階目で測定される共変量が観測されるか否かは「デザインによって生じる欠測」であり、欠測のメカニズムは、1段階目のコホートで観測された情報によって完全に規定される MAR (Missing at random) での欠測であることが保証される。これにより、ケースコホート研究には、MAR のもとの不完全データの解析の方法をそのまま適用することができる。

3.3.2 層別サンプリングと重みつき推定方程式

2段階目のサンプリング確率は完全に既知であるため、MAR のもとの妥当な解析方法として、この既知のサンプリング確率の逆数を重みに使った、逆確率重みつき推定法 (inverse probability weighting [IPW]; Robins et al., 1994) を利用することができる。実際、後付的に考えると、3.1 節に与えた推定方法は、いずれも個人ごとのスコアの寄与を重みづけによって調整することにより、不偏な推定量を得ようという方式であると解釈することができる。

Borgan et al. (2000) は、2段階目で測定される共変量と相関がある変数が、1段階目の全コホートで測定されている場合、これによって層別を行い、2段階目のサンプリングを行うデザインを提案した。2段階ケースコントロール研究の理論 (Breslow, 2005; White, 1982) でもよく知られている通り、このような層別サンプリングを行うことによって、2段階目での共変量 Z のばらつきを小さくすることができ、回帰パラメータの推定精度を改善することができる。Borgan et al. (2000) は、いくつかの重みつき推定量を提案しており、層ごとに Self and Prentice (1988) の重みを付与する推定量 (Type-I 推定量) と、層ごとのケース・非ケースのサンプリング割合の逆数を重みとして付与する IPW 推定量 (Type-II 推定量) がよく知られている。現在では、実践においても、多くのケースコホート研究で、この層別サンプリングデザインは用いられており、特に、Borgan et al. (2000) の Type-II 推定量がよく用いられているようである。

3.3.3 標本調査法におけるキャリブレーション推定法の応用

Breslow et al. (2009a, 2009b) は、標本調査法で発展した Horvitz-Thompson 推定量の改良版であるキャリブレーション法をケースコホート研究の IPW 解析に応用することを提案している。

標本調査法で発展した Horvitz-Thompson 推定量 (Horvitz and Thompson, 1952) については、上記の2段階サンプリングデザインの議論と同様、源泉集団で測定された補助変数の情報を利用して、推定精度を改善する方法の研究が早い段階から進められており、一般化回帰推定量、事後層別推定量、レイキング推定量など、多くの推定量が提案されている。キャリブレーション法は、Deville and Särndal (1992) によって提案された、これらを一般化した推定方法の枠組みである。

ここでは、キャリブレーション法を説明するために、いくつかの表記法を新しく導入する。いま、 N 人の対象者から構成される母集団における変数 Y_1, \dots, Y_N の総和 $Y_{tot} = \sum_{i=1}^N Y_i$ を推定する問題を考える。ここからサンプリングされた対象者の添え字集合を Λ とする。 $w_i = \Pr(i \in \Lambda)^{-1}$ とすると、 Y_{tot} の Horvitz-Thompson 推定量は、 $\hat{Y}_{HT} = \sum_{i \in \Lambda} w_i Y_i$ と書ける。ここで、 Y_i に強い相関のある補助変数 A_i が母集団における N 人の対象者全員に観測されていたら、 A_i の母集団での既知の総和 $A_{tot} = \sum_{i=1}^N A_i$ に対して、

$$A_{tot} = \sum_{i=1}^N A_i = \sum_{i \in \Lambda} \omega_i A_i \quad (*)$$

となる適当な重み ω_i を与えることができる。直感的には、仮に Y_i と A_i の間に強い相関がある場合、真のサンプリング確率の逆数 w_i の代わりに ω_i を重みに用いることで、 $\hat{Y}_{tot} = \sum_{i \in \Lambda} \omega_i Y_i$ は、 Y_{tot} により近い推定値を与えることが期待できる。この制約を与える方程式 (*) を、キャ

リブレーション方程式(calibration equation)という。実際には、キャリブレーション方程式を満たす重みは無数にあるため、この方程式の解の中から、真の重み w_i に最も近い重み ω_i を用いるのが自然である。Deville and Särndal(1992)のキャリブレーション法では、この ω_i を一意に定めるために、 w_i と ω_i の「近さ」を測る距離関数 $G(\omega, w)$ を具体的に与える。例えば、 $G_{GREG}(\omega, d) = (\omega - d)^2/2d$ という距離関数を用いると、これによる制約つき最適化問題からは、一般化回帰推定量が得られる。キャリブレーション法は、他にも、適当な距離関数を用いることで、事後層別推定量、レイキング推定量をはじめ、さまざまな推定量を構成することができる推定方式の枠組みである (Deville and Särndal, 1992; Deville et al., 1993)。

ケースコホート研究の設定では、IPW 推定量 $\hat{\beta}_{IPW}$ を真値 β_0 のまわりでテイラー展開し、1 次の近似式を求めると、

$$\hat{\beta}_{IPW} \approx \beta_0 + \sum_{\Lambda} w_i I^{-1}(\beta_0) U_i(\beta_0)$$

となる。ここでは、 Λ は 2 段階目のサンプルの添え字集合であり、 w_i は 2 段階目でのサンプリング確率の逆数、 U_i は i 番目の対象者のスコア関数、 I は情報行列を表す。上記の近似式において、 $I^{-1}(\beta_0)U_i(\beta_0)$ に相関が高い変数を補助変数 A_i として、キャリブレーション法を行うことで、推定量の改善が期待できる。補助変数 A_i の自然な候補として、Breslow et al.(2009a, 2009b)は、dfbeta

$$A_i = I^{-1}(\tilde{\beta})U_i(\tilde{\beta})$$

を用いることを提案している。 $\tilde{\beta}$ は、全コホートにおける Cox 回帰の推定値である。ただし、当然ながら、実際には、この全コホートの dfbeta は観測されないため(2 段階目に選択されなかった対象者には Z が欠測しているため)、Breslow et al.(2009a, 2009b)は、以下の Kulich and Lin(2004)の Plug-in 式の方法を用いて、近似的な dfbeta を求めることを提案している。

- (1) 2 段階目にサンプリングされなかった対象者についての Z の予測値を求める。(X, Z)の完全データが観測されている対象者(2 段階目の対象者)に対して、 Z を結果変数、全コホートで観測されている共変量 X を説明変数とした、適当な回帰モデルを仮定して、重みつき推定法(例えば、サンプリング確率の逆数を重みとした IPW 法)によって回帰パラメータの推定値を求める。これを予測モデルとして、2 段階目にサンプリングされなかった対象者についての Z の予測値を求める。
- (2) 1 段階目の全コホートのデータセットの Z の欠測値に、(1)で求めた予測値を補完し、擬似的な完全データを作る。これに対して、Cox 回帰を行い、ここから得られる dfbeta を、前述の $I^{-1}(\tilde{\beta})U_i(\tilde{\beta})$ の近似値として利用する。

Breslow et al.(2009a, 2009b)は、こうして求められた dfbeta によって、キャリブレーション法を行い、求めた重みを利用して IPW 法を実行することを提案している。分散の推定は、重みのキャリブレーションによる不確実性も加わるため、単純なサンドイッチ推定量では推定できないが、この解析は、R の survey パッケージ(Lumley, 2004; Lumley et al., 2011)で実行することができる。

3.3.4 推定された重みに基づく方法

不完全データの解析における IPW 法では、真の重み(既知のサンプリング確率の逆数)を重みにとるよりも、適当な回帰モデルで推定した重み(真のサンプリング確率からは、必ず乖離が生じる)を用いたほうが、回帰パラメータの推定精度が改善されることが知られている (Robins et al., 1994)。すなわち、2 段階目の共変量 Z が観測されるか否かを表す指示変数を結果変数としたロジスティック回帰モデルを用いて、1 段階目の全コホートで観測されている共変量を説

明変数として、「 Z が観測される確率」を推定し、この逆数を重みに用いるという IPW 法である。これは、一見、矛盾しているかのように見える結果であるが、この推定された重みを用いることによる精度の改善は、理論的にも証明されている (Henmi and Eguchi, 2004; Pierce, 1982; Lumley et al., 2011)。

重みを予測するための回帰モデルには、まず、層別サンプリングを行う場合には、層を表すダミー変数を含めることは必須である。層ごとのサンプリング割合が不均一である場合には、これを含まなければ、モデルが誤特定されたものになってしまうので、推定量の一致性が保証されない。また、精度の改善のために、前節の方法で求められる df_{beta} もモデルに加えることが勧められる。

分散の推定においては、やはり、重みの推定における不確実性が加わるため、単純なサンドイッチ推定量ではバイアスが入る(ただし、一般的に、保守的なバイアスとなる)。分散の一致推定量は、この重みの不確実性を加味することで得ることができ、少し複雑なものとなるが、これについても、R の `survey` パッケージ (Lumley, 2004; Lumley et al., 2011) で計算することができる。

3.4 セミパラメトリック有効な推定方法

MAR の欠測のもとでは、単純な IPW 法は、一致推定量は与えるが、漸近有効な方法ではないため、前節の方法のように、精度を改善するための方法論についての議論は広く行われている。理論的には、Robins et al. (1994) によって、一般的には局外要因となる、2 段階目で測定される Z の分布について、特定のパラメトリックな仮定を置かないセミパラメトリックモデルのもとでの漸近有効な方法が与えられている。いわゆる AIPW (augmented inverse probability weighting) 法であるが、実際には未知量である IPW 推定関数への Augmented 項の推定量を構成する必要があり、それは必ずしも容易ではない。ひとつの例としては、Qi et al. (2005) が、カーネル法を用いて、Cox 回帰の AIPW 推定関数の構成を行う方法を与えており、ケースコホート研究への応用を議論している。また、Kulich and Lin (2004) は、AIPW 法と同等の漸近分散を持つ、2 重の重みづけ推定量を与えている。Nan (2004) は有効スコア方程式の解を近似する推定量を提案しており、これが漸近的にセミパラメトリック有効になることを示している。

一方、これらの方法は、比較的新しい方法論であるとともに、強い制約的な仮定や複雑な計算を必要とするため、実践においては広く用いられてはいない (Breslow and Wellner, 2007)。これは、ケースコホート研究への応用に限った話ではなく、MAR のもとでの欠測に対するセミパラメトリック有効な推定の方法論全般の問題であるといえる。今後、理論的な研究が進むにつれて、解決されていくことが期待されるが、現状としては、これらの問題から、より現実的な方法として、3.3 節に紹介した方法が、実践的には有用であるだろう。ただし、2.2.2 節で述べた、2 項回帰モデルの解析では、Breslow et al. (2003), Lawless et al. (1999), Scott and Wild (1997) などの 2 段階ケースコントロール研究で確立された方法がある。

3.5 Multiple Imputation 法

MAR のもとでの漸近有効性を持つ、もうひとつの代表的な方法として、Multiple Imputation 法 (Rubin, 1987) が挙げられる。Multiple Imputation 法は、当初はベイズ流の枠組みにおいて提案された方法であるが、近年の研究により、一定の正則条件のもとで、頻度論的な漸近有効性を有する推定量を与えることが示されている (Robins and Wang, 2000; Wang and Robins, 1998). Keogh and White (2013), Marti and Chavance (2011), 野間 他 (2012) は、Multiple Imputation 法をケースコホート研究の解析に応用することを提案しており、より精度の高い推定値が得られることを、数値実験等で示している。

Multiple Imputation 法では、観測されなかった Z に対して、複数の適当な補完値の組 $Z^{(1)}, Z^{(2)}, \dots, Z^{(m)}$ を生成し、全コホートの擬似的な完全データを作成する。これらをそれぞれ Cox 回帰により解析し、推定値 $\hat{\beta}^{(1)}, \hat{\beta}^{(2)}, \dots, \hat{\beta}^{(m)}$ と分散の推定値 $\hat{V}^{(1)}, \hat{V}^{(2)}, \dots, \hat{V}^{(m)}$ を得る。これにより、 β の推定値は $\hat{\beta}_{IM} = \sum_{i=1}^m \hat{\beta}^{(i)} / m$ で得ることができる。また、補完内分散 $W_{IM} = \sum_{i=1}^m \hat{V}^{(i)} / m$ 、補完間分散 $B_{IM} = \sum_{i=1}^m (\hat{\beta}^{(i)} - \hat{\beta}_{IM})(\hat{\beta}^{(i)} - \hat{\beta}_{IM})^T / (m - 1)$ から、分散の推定値は、 $\hat{V}(\hat{\beta}_{IM}) = W_{IM} + (1 + m^{-1})B_{IM}$ と得ることができる。補完値の数 m は、以前は 5~10 回でも十分と言われてきたが、近年では、十分大きな値をとるべきであるとする文献が多い(例えば、Royston and White, 2011)。

Multiple Imputation 法では、一般的に補完値の生成において、付加的なパラメトリックな仮定を置く必要があり、重みつき解析の方法に比べて、より強い制約を必要とすることが欠点として挙げられることが多い。一方で、Multiple Imputation 法は、医学研究では、欠測データの解析方法として、既に広く普及しており、多くの標準的なソフトウェアに解析用のモジュールが付属しているという利点もある。

4. National Wilms Tumor Studies によるシミュレーション実験

本節では、米国の National Wilms Tumor Study (NWTs) のデータをもとにした仮想的なケースコホート研究を行い、種々の解析手法の性能の評価を行う。Wilms 腫瘍患者では、「unfavourable histology (UH)」とされる特殊な細胞型の腫瘍を持つ少数の患者が、「favourable histology (FH)」と診断される患者に比べ、治療後の再発や死亡が起りやすいことが報告されている (Beckwith and Palmer, 1978)。NWTs では、100 以上の参加施設で組織診断が行われた後、中央の研究施設に検体を送られ、病理専門家によって中央診断が行われている。両者の診断結果にはある程度、不一致があることが確認されており、施設での診断結果の精度は、全体で感度 74%、特異度 98%であったとされている。施設診断に比べ、中央診断のほうが診断の誤差は小さいと考えられるが、一方で追加コストが必要となる。ここでは、中央での診断結果を 2 段階目で測定される共変量、施設ごとの診断結果をその代替変数とした評価を行う。1 段階目のコホートを NWTs による 3915 名のコホートとし、これに基づく仮想的なケースコホート研究を 1000 回シミュレーションし、解析手法の評価を行う。同様のシミュレーションは、Breslow et al. (2009b), Kulich and Lin (2004) でも行われている。

2 段階目のサブコホートのサンプリングは、Breslow et al. (2009b) に従い、施設病理診断の結果 (UH, FH)、進行度 (Stage I, II or III, IV)、年齢 (1 歳未満, 以上) に基づく 8 つの層による、層別サンプリングによって行った。その内訳を表 4 に示す。このコホートをもとに、仮想的なケースコホート研究を 1000 回行い、それぞれの試行で、全コホートの完全データを用いた解析 [WC; whole cohort], Borgan et al. (2000) の Type-II 推定量 [standard weights], Breslow et al. (2009a, 2009b) によるキャリブレーション法 [calibrated weights], 推定された重みによる IPW 法 [estimated weights], Multiple Imputation 法による解析を行った。解析には、共変量に中央病理診断の結果 (Histology; UH, FH)、年齢 (区分線形関数による適当な変換を行っている; Age0 が 1 歳未満, Age1 が 1 歳以上に対応)、進行度 (Stage I, II or III, IV)、腫瘍径 (Diameter) と、3 つの交互作用項 (進行度 \times 腫瘍径, 中央病理診断 \times Age0, 中央病理診断 \times Age1) をモデル化した Cox 回帰モデルを用いた。Multiple Imputation 法では、マルコフ連鎖モンテカルロ法を用いた補完値の生成を行った。

表 5 に、それぞれの解析方法による 1000 回の回帰係数の推定値の平均と標準偏差 (SD)、漸近的な標準誤差 (ASE; asymptotic standard error) の推定値の平均を示している。ただし、WC の列には、1 段階目のコホートの対象者全員の完全データに対しての Cox 回帰の最大部分尤度

表 4. NWTS コホートの内訳と層別ケースコホート研究のデザイン. 施設病理診断の結果(UH or FH), Stage, Age によって 8 つの層に層別を行っている.

	合計	Favorable Histology (FH)				Unfavorable Histology (UH)			
		Stage I-II		Stage III-IV		Stage I-II		Stage III-IV	
		Age < 1	Age ≥ 1	Age < 1	Age ≥ 1	Age < 1	Age ≥ 1	Age < 1	Age ≥ 1
1 段階目の全コホート (N = 3,915)									
Cases	669	57	232	10	208	15	41	29	77
Non-cases	3246	452	1620	40	914	12	107	2	99
% Relapsed	17.1	11.2	12.5	20	18.5	55.5	27.7	93.5	43.8
ケースコホート研究における層別サンプリング (n = 1,329)									
Cases	669	57	232	10	208	15	41	29	77
Non-cases	660	120	160	40	120	12	107	2	99

による推定値と、その標準誤差の推定値を示している。

全般的な結果としては、いずれの方法による推定値も、WC の推定値と概ね近い値をとっており、全コホートにおけるハザード比の推定値を、ケースコホート研究によって、偏りなく再現できたことがわかる。ただし、交互作用項に関しては、IPW 推定量ではある程度のバイアスが見受けられる。ケースコホート研究では、通常の臨床研究の欠測とは異なり、欠測データの割合が、全データの大半を占めることになるため、特に、不安定な交互作用項については、この違いが大きく影響するものと思われる。一方で、Multiple Imputation 法では、ほとんどバイアスはない。ただし、先述の通り、Multiple Imputation 法は、補完値の生成において、特定のパラメトリックなモデルの仮定を置く必要があるため、これを誤ると、逆に、深刻なバイアスを生じさせてしまう可能性も考えられる。

回帰パラメータの推定精度は、方法間でかなりの差が見受けられる。Borgan et al. (2000) の Type-II 推定量は、原則として、全コホートの情報を層別にしか使っていないため、相対的な SD は大きくなっており、一方で、全コホートの情報を取り入れているキャリブレーション法と推定された重みによる IPW 法は、かなりの改善が認められている。特に、全コホートで観測されている Histology 以外の共変量に対応する回帰パラメータでは、大きな精度の改善が認められている。もちろん、Histology についての改善も認められる。他方、Multiple Imputation 法は、これらよりもさらに大きな精度の改善が認められている。これは、理論的な結果に合致し、Multiple Imputation 法は、補完値の生成モデルにもパラメトリックなモデルを仮定する分、漸近分散がそもそも小さい。

5. まとめ

ケースコホート研究が、ネステッドケースコントロール研究に比べて、実践での普及が遅れたのは、統計手法の敷居の高さと計算の困難さも原因のひとつであったといわれているが (Barlow et al., 1999; Wacholder, 1991)、現在では、標準的な統計ソフトウェアでも、これらの解析は簡単に行えるようになっており、解析手法のオプションも大幅に増えている。シミュレーションによる効率の比較においても、ネステッドケースコントロール研究とは、ほとんど違いはないことが示されており (Langholz and Thomas, 1990, 1991)、実践面でもむしろ利点は多いように思われる。

本邦においても、近年では、環境省エコチル調査 (<http://www.env.go.jp/chemi/ceh/>) で採用

表 5. NWTS コホートに基づくシミュレーション実験の結果. 仮想的な 1,000 回のケースコホート研究における推定結果.

		IPW 推定量				
		WC	Standard weights	Calibrated weights	Estimated weights	Multiple Imputation
Histology	推定値	4.042	4.049	4.056	4.059	4.040
	SD	—	0.187	0.132	0.135	0.036
	ASE	0.413	0.536	0.523	0.518	0.512
Age ₀	推定値	-0.661	-0.677	-0.662	-0.673	-0.662
	SD	—	0.162	0.037	0.059	0.007
	ASE	0.326	0.359	0.325	0.325	0.322
Age ₁	推定値	0.104	0.107	0.104	0.107	0.101
	SD	—	0.022	0.006	0.011	0.001
	ASE	0.017	0.026	0.017	0.017	0.016
Stage	推定値	1.346	1.341	1.339	1.334	1.344
	SD	—	0.232	0.071	0.112	0.018
	ASE	0.244	0.346	0.272	0.271	0.261
Diameter	推定値	0.069	0.070	0.070	0.070	0.072
	SD	—	0.015	0.005	0.008	0.001
	ASE	0.014	0.021	0.015	0.015	0.015
Stage × Diameter	推定値	-0.076	-2.652	-2.662	-2.661	-0.076
	SD	—	0.280	0.235	0.242	0.001
	ASE	0.019	0.612	0.595	0.590	0.020
Histology × Age ₀	推定値	-2.635	-0.052	-0.050	-0.051	-2.651
	SD	—	0.047	0.047	0.049	0.061
	ASE	0.464	0.051	0.049	0.048	0.566
Histology × Age ₁	推定値	-0.058	-0.076	-0.075	-0.075	-0.031
	SD	—	0.021	0.006	0.010	0.012
	ASE	0.034	0.029	0.021	0.021	0.037

推定値, SD: 1000 回の試行での推定値の平均と標準偏差. SD は, 2 段階目のサンプリングによる推定量のばらつきを反映している.

ASE: 1000 回の試行における漸近標準誤差の推定値の平均. 1 段階目, 2 段階目のサンプリングにおける推定値の総合的なばらつきを反映している.

されており, 実践での応用の場は少しずつ広がっていくものと思われる. また, 海外の動向を鑑みると, 遺伝子情報・ゲノム情報の測定など, 高価な共変量情報の測定を伴う研究は, 近年になって, ますます増加しており, ケースコホート研究の需要は, 確実に高まっているといえるだろう (Kulathinal et al., 2007; Lin, 2014). こうした実務におけるニーズに応えるためにも, さらなる理論の確立と普及は, 今後も重要な課題になると思われる.

謝 辞

本研究は, 科学研究費補助金(課題番号: 24800081)の助成を受けた.

参 考 文 献

- Afshar-Kharghan, V., Matijevic-Aleksic, N., Ahn, C., Boerwinkle, E., Wu, K. K. and Lopez, J. A. (2004). The variable number of tandem repeat polymorphism of platelet glycoprotein Iba1 and risk of coronary heart disease, *Blood*, **103**, 963–965.
- The ARIC Investigators (1989). The Atherosclerosis Risk in Communities (ARIC) study: Design and objectives, *American Journal of Epidemiology*, **129**, 687–702.
- Ballantyne, C. M., Hoogeveen, R. C., Bang, H., et al. (2004). Lipoprotein-associated phospholipase A2, high-sensitivity C-reactive protein, and risk for incident coronary heart disease in middle-aged men and women in the Atherosclerosis Risk in Communities (ARIC) study, *Circulation*, **109**, 837–842.
- Ballantyne, C. M., Hoogeveen, R. C., Bang, H., et al. (2005). Lipoprotein-associated phospholipase A2, high-sensitivity C-reactive protein, and risk for incident ischemic stroke in middle-aged men and women in the Atherosclerosis Risk in Communities (ARIC) study, *Archives of Internal Medicine*, **165**, 2479–2484.
- Barlow, W. E. (1994). Robust variance estimation for the case-cohort design, *Biometrics*, **50**, 1064–1072.
- Barlow, W. E., Ichikawa, L., Rosner, D. and Izumi, S. (1999). Analysis of case-cohort designs, *Journal of Clinical Epidemiology*, **52**, 1165–1172.
- Beckwith, J. B. and Palmer, N. F. (1978). Histopathology and prognosis of Wilms tumor, *Cancer*, **41**, 1937–1948.
- Borgan, Ø., Langholz, B., Samuelsen, S. O., Goldstein, D. R. and Pogoda, J. (2000). Exposure stratified case-cohort designs, *Lifetime Data Analysis*, **6**, 39–58.
- Breslow, N. E. (2005). Case-control study, Two-phase, *Encyclopedia of Biostatistics*, 2nd ed. (eds. P. Armitage and T. Colton), 670–678, Wiley, Chichester.
- Breslow, N. E. and Day, N. E. (1980). *Statistical Methods in Cancer Research, Vol. 1: The Analysis of Case-control Studies*, International Agency for Research on Cancer, Lyon.
- Breslow, N. E. and Wellner, J. A. (2007). Weighted likelihood for semiparametric models and two-phase stratified samples, with application to Cox regression, *Scandinavian Journal of Statistics*, **34**, 86–102.
- Breslow, N. E., Robins, J. M. and Wellner, J. A. (2000). On the semi-parametric efficiency of logistic regression under case-control sampling, *Bernoulli*, **6**, 447–455.
- Breslow, N. E., McNeney, B. and Wellner, J. A. (2003). Large sample theory for semiparametric regression models with two-phase, outcome dependent sampling, *Annals of Statistics*, **31**, 1110–1139.
- Breslow, N. E., Lumley, T., Ballantyne, C. M., Chambless, L. E. and Kulich, M. (2009a). Improved Horvitz-Thompson estimation of model parameters from two-phases stratified samples: Applications in epidemiology, *Statistics in Bioscience*, **1**, 32–49.
- Breslow, N. E., Lumley, T., Ballantyne, C. M., Chambless, L. E. and Kulich, M. (2009b). Using the whole cohort in the analysis of case-cohort data, *American Journal of Epidemiology*, **169**, 1398–1405.
- Cox, D. R. (1972). Regression models and life tables (with discussion), *Journal of the Royal Statistical Society, Series B*, **34**, 187–220.
- Cox, D. R. (1975). Partial likelihood, *Biometrika*, **62**, 269–276.
- Deville, J. C. and Särndal, C.-E. (1992). Calibration estimators in survey sampling, *Journal of the American Statistical Association*, **87**, 376–382.
- Deville, J. C., Särndal, C.-E. and Sautory, O. (1993). Generalized raking procedures in survey sampling, *Journal of the American Statistical Association*, **88**, 1013–1020.
- Greenland, S. (1986). Adjustment of risk ratios in case-base studies (hybrid epidemiologic designs),

- Statistics in Medicine*, **5**, 579–584.
- Henmi, M. and Eguchi, S. (2004). A paradox concerning nuisance parameters and projected estimating functions, *Biometrika*, **91**, 929–941.
- Horvitz, D. and Thompson, D. (1952). A generalization of sampling without replacement from a finite population, *Journal of the American Statistical Association*, **47**, 663–685.
- Kalbfleisch, J. D. and Lawless, J. F. (1988). Likelihood analysis of multi-state models for disease incidence and mortality, *Statistics in Medicine*, **7**, 149–160.
- Keogh, R. H. and White, I. R. (2013). Using full-cohort data in nested case-control and case-cohort studies by multiple imputation, *Statistics in Medicine*, **32**, 4021–4043.
- Kulathinal, S., Karvanen, J., Saarela, O. and Kuulasmaa, K. (2007). Case-cohort design in practice: Experiences from the MORGAM Project, *Epidemiologic Perspectives & Innovations*, **4**, 15.
- Kulich, M. and Lin, D. Y. (2004). Improving the efficiency of relative-risk estimation in case-control studies, *Journal of the American Statistical Association*, **99**, 832–844.
- Kupper, L. L., McMichael, A. J. and Spirtas, R. (1975). A hybrid epidemiologic study design useful in estimating relative risk, *Journal of the American Statistical Association*, **70**, 524–528.
- Langholz, B. and Jiao, J. (2007). Computational methods for case-cohort studies, *Computational Statistics and Data Analysis*, **51**, 3737–3748.
- Langholz, B. and Thomas, D. C. (1990). Nested case-control and case-cohort methods of sampling from a cohort: A critical comparison, *American Journal of Epidemiology*, **131**, 169–176.
- Langholz, B. and Thomas, D. C. (1991). Efficiency of cohort sampling designs: Some surprising results, *Biometrics*, **47**, 1563–1571.
- Lawless, J. F., Kalbfleisch, J. D. and Wild, C. J. (1999). Semiparametric methods for response-selective and missing data problems, *Journal of the Royal Statistical Society, Series B*, **61**, 413–438.
- Li, R., Boerwinkle, E., Olshan, A. F., et al. (2000). Glutathione S-transferase genotype as a susceptibility factor in smoking-related coronary heart disease, *Atherosclerosis*, **149**, 451–462.
- Lin, D. Y. (2014). Survival analysis with incomplete genetic data, *Lifetime Data Analysis*, **20**, 16–22.
- Lin, D. Y. and Wei, L. J. (1989). The robust inference for the Cox proportional hazards model, *Journal of the American Statistical Association*, **84**, 1074–1078.
- Lin, D. Y. and Ying, Z. (1993). Cox regression with incomplete covariate measurements, *Journal of the American Statistical Association*, **88**, 1341–1349.
- Lumley, T. (2004). Analysis of complex survey samples, *Journal of Statistical Software*, **9**, Issue 8.
- Lumley, T., Shaw, P. A. and Dai, J. Y. (2011). Connections between survey calibration estimators and semiparametric models for incomplete data, *International Statistical Review*, **79**, 200–220.
- Marti, H. and Chavance, M. (2011). Multiple imputation analysis of case-cohort studies, *Statistics in Medicine*, **30**, 1595–1607.
- Miettinen, O. S. (1982). Design options in epidemiologic research: An update, *Scandinavian Journal of Work and Environmental Health*, **8**, 7–14.
- MRFIT Research Group (1982). Multiple risk factor intervention trial: Risk factor changes and mortality results, *Journal of the American Medical Association*, **248**, 1465–1477.
- Nan, B. (2004). Efficient estimation for case-cohort studies, *Canadian Journal of Statistics*, **32**, 403–419.
- Noma, H. and Tanaka, S. (2013). Analysis of case-cohort designs with binary outcomes: Improving the efficiency using whole cohort auxiliary information (unpublished manuscript).
- 野間久史, 田中司朗, 田中佐智子, 和泉志津恵 (2012). Multiple Imputation 法によるネステッドケースコントロール研究, ケースコホート研究の解析, 計量生物学, **33**, 101–124.
- Oakes, D. (1981). Survival times: Aspects of partial likelihood, *International Statistical Review*, **49**, 235–252.

- Onland-Moret, N. C., van der A., D. L., van der Schouw, Y. T., et al. (2007). Analysis of case-cohort data: A comparison of different methods, *Journal of Clinical Epidemiology*, **60**, 350–355.
- Petersen, L., Sørensen, T. I. A. and Andersen, P. K. (2003). Comparison of case-cohort estimators based on data on premature death of adult adoptees, *Statistics in Medicine*, **22**, 3795–3803.
- Pierce, D. A. (1982). The asymptotic effect of substituting estimators for parameters in certain types of statistics, *Annals of Statistics*, **10**, 475–478.
- Prentice, R. L. (1986). A case-cohort design for epidemiologic cohort studies and disease prevention trials, *Biometrika*, **73**, 1–11.
- Prentice, R. L. and Pyke, R. (1979). Logistic disease incidence models and case-control studies, *Biometrika*, **66**, 403–411.
- Qi, L., Wang, C. Y. and Prentice, R. L. (2005). Weighted estimators for proportional hazards regression with missing covariates, *Journal of the American Statistical Association*, **100**, 1250–1263.
- Rasmussen, M. L., Folsom, A. R., Catellier, D. J., Tsai, M. Y., Garg, U. and Eckfeldt, J. H. (2001). A prospective study of coronary heart disease and the hemochromatosis gene (HFE) C282Y mutation: The Atherosclerosis Risk in Communities (ARIC) study, *Atherosclerosis*, **154**, 739–746.
- Reid, N. and Crépeau, H. (1985). Influence functions for proportional hazards regression, *Biometrika*, **72**, 1–9.
- Robins, J. M. and Wang, N. (2000). Inference for imputation estimators, *Biometrika*, **87**, 113–124.
- Robins, J. M., Rotnitzky, A. and Zhao, L. P. (1994). Estimation of regression-coefficients when some regressors are not always observed, *Journal of the American Statistical Association*, **89**, 846–866.
- Royston, P. and White, I. R. (2011). Multiple Imputation by Chained Equations (MICE): Implementation in Stata, *Journal of Statistical Software*, **45**, Issue 4.
- Rubin, D. B. (1987), *Multiple Imputation for Nonresponse in Surveys*, John Wiley, New York.
- Sato, T. (1992a). Estimation of a common risk ratio in stratified case-cohort studies, *Statistics in Medicine*, **11**, 1599–1605.
- Sato, T. (1992b). Maximum likelihood estimation of the risk ratio in case-cohort studies, *Biometrics*, **48**, 1215–1221.
- Sato, T. (1994). Risk ratio estimation in case-cohort studies, *Environmental Health Perspectives*, **102**, 53–56.
- Schouten, E. G., Dekker, J. M., Kok, D. F., et al. (1993). Risk ratio and rate ratio estimation in case-cohort designs: Hypertension and cardiovascular mortality, *Statistics in Medicine*, **12**, 1733–1745.
- Scott, A. J. and Wild, C. J. (1997). Fitting regression models to case-control data by maximum likelihood, *Biometrika*, **84**, 57–71.
- Self, S. G. and Prentice, R. L. (1988). Asymptotic distribution theory and efficiency results for case-cohort studies, *Annals of Statistics*, **16**, 64–81.
- Therneau, T. M. and Li, H. (1999). Computing the Cox model for case cohort designs, *Lifetime Data Analysis*, **5**, 99–112.
- Thomas, D. C. (1977). Addendum to a paper by F. D. K. Liddell, J. C. McDonald and D. C. Thomas, *Journal of the Royal Statistical Society, Series A*, **140**, 483–485.
- van der Klauw, M. M., Goudsmit, R., Hlie, M. R., et al. (1999). A population-based case-cohort study of drug-associated agranulocytosis, *Archives of Internal Medicine*, **159**, 369–374.
- Wacholder, S. (1991). Practical considerations in choosing between the case-cohort and nested case-control designs, *Epidemiology*, **2**, 155–158.
- Wang, N. and Robins, J. M. (1998). Large sample theory for parametric multiple imputation procedures,

Biometrika, **85**, 935–948.

White, J. E. (1982). A two stage design for the study of the relationship between a rare exposure and a rare disease, *American Journal of Epidemiology*, **115**, 119–128.

Wickramaratne, P. J. (1995). Sample size determination in epidemiologic studies, *Statistical Methods in Medical Research*, **4**, 311–337.

Zhao, L. P. and Lipsitz, S. (1992). Designs and analysis of two-stage studies, *Statistics in Medicine*, **11**, 769–782.

Theory and Methods for the Case-cohort Studies

Hisashi Noma

The Institute of Statistical Mathematics

Clinical and epidemiological cohort studies often require large numbers of subjects and long-time follow ups, and assembly of covariate histories on all cohort members can be prohibitively expensive. Case-cohort designs are common means of reducing the cost of covariate measurements for large cohort studies. This article provides a comprehensive review of theory and statistical methods for case-cohort studies. It reviews conventional methods for estimating the measures of exposure effects, risk-ratio and odds-ratio, and modified partial likelihood methods for the Cox regression model. In addition, it reviews recently developed estimating methodologies using auxiliary information of whole cohort subjects. Simulation studies using data taken from Wilms' tumor studies are provided.