

厚生労働省科学研究費補助金循環器疾患等生活習慣病対策政策研究事業  
循環器疾患における集団間の健康格差の実態把握とその対策を目的とした大規模コホート共同研究  
(H26 - 循環器等 (政策) - 一般 - 001)  
分担研究報告書

4. 複数のコホート研究データを用いたリスク評価モデルの構築・妥当性評価に関する文献的検討

研究分担者 村上 義孝 東邦大学医学部社会医学講座医療統計学分野 教授  
研究協力者 大庭 真梨 横浜市立大学大学院医学群臨床統計学講座 助教

研究要旨

統合コホートデータベースに対し、複数の研究からなることを考慮した、より精緻な解析を実施するための統計学的手法、とくに生存時間に対するリスク予測モデル構築とその妥当性の吟味について文献収集し、概要をまとめた。研究間差を考慮する方法としては各研究で解析し加重平均を算出する方法と、全データをまとめ層別解析や変量効果モデルで解析する手法があること、曝露効果の研究間差を調べるための検討(効果修飾の検討)では、統合コホートデータベースから詳細な検討が可能であり、図示は有効な手段であること、などが要約された。また疾患予測モデルを構築する際は、関心のある曝露の全体効果の推定のみに着目するのではなく、コホートごとの効果や共変量の影響に異質性はないか、異質性の原因は何であるかなど検討し、作成したモデルの妥当性ととも比較検討しながら解析を実施する必要があることが確認された。今回の検討からEPOCH-JAPANにおいて本報告で整理した一連の手法は有用であり、現段階で有効な候補であると思われた。

A. 研究目的

循環器疾患を対象とした大規模統合データ解析プロジェクトである Emerging risk factors collaboration (ERFC)では、30以上の観察研究(以下コホート)から個人データ(Individual participants data, IPD)を収集し、統合・解析されている。今回ERFCにおけるIPDメタアナリシス(IPDMA)で用いられた生存時間の予測モデルの構築手法およびモデル評価手法 [Thompson 2010, Pennells 2013]を調査し、整理したので報告する。

B. 研究方法

IPDMAの解析手順

IPDMAの解析手順には、コホートごとに要約統計量を算出してから統合する二段階法と、IPDを直接解析する一段階法がある。二段階法は各コホートで興味のあるパラメータを推定してそれを統合するため、データが膨大な場合であっても計算負

荷が比較的低く、またIPDがなく文献ベースの要約統計量しかないコホートの結果も解析に含めることができるという利点をもつ。一方でコホート数が少ない場合に検出力が低下する、個人レベルでの交絡調整が行えないという欠点も存在する。一段階法は統計モデルを作成する際にコホートを層別因子とするなどして考慮し、すべての個人データを同時に解析する方法である。この方法はデータ数が多いため、検出力の低下がなく、交互作用を個人レベルで検討することができ、交互作用、時間依存性共変量などを用いた複雑な解析も可能である。交絡調整やモデルの柔軟性の観点からは一段階法が望ましいが、二段階法でも曝露効果の推定値自体には大きなバイアスとならないとされている[Riley2008]。

基本モデル

複数のコホートを統合した解析では、コホートの効

果を考慮したモデル化が必須である。二段階法の場合、コホート  $s$  ごとにハザード比を推定し、次に統合するので区別は容易である。

$$h_{si}(t) = h_{0s}(t) \exp(\beta_s E_{si} + \gamma_s X_{si}) \quad (1)$$

ここで  $h_{0s}(t)$  は研究  $s$  のベースラインハザード、 $E_{si}$  と  $X_{si}$  はコホート  $s$  の対象者  $i$  の曝露因子と共変量、 $k$  は任意の層別因子（人種や年齢など）、 $\beta_s$  と  $\gamma_s$  は研究  $s$  の曝露効果と共変量効果である。ベースラインハザード  $\lambda_{0sk}(t)$  の添え字  $s$  と  $k$  は層によってベースラインリスクが異なることを表す。一段階法では、同じ式が同時にデータに当てはめられる。

変量効果モデルを用いる場合、コホートごとの効果  $\beta_s$  に分布を仮定する。推定された対数ハザード比  $\hat{\beta}_s$  は、推定誤差  $\varepsilon_s$  をもち、さらにコホートごとの効果の期待値  $\beta_s$  は全体平均を中心に分布していると仮定する。

$$\begin{aligned} \hat{\beta}_s &= \beta_s + \varepsilon_s, \quad \varepsilon_s \sim N(0, \sigma_s^2) \\ \beta_s &= \beta + \eta_s, \quad \eta_s \sim N(0, \tau^2) \end{aligned} \quad (2)$$

統合効果  $\beta$  の推定は重み  $w_s = (\sigma_s^2 + \tau^2)^{-1}$  を用いた重み付平均によって行う。 $\tau^2$  の推定はモーメント法を用いる DerSimonian-Laird 法が一般的である。

$$\hat{\beta} = \frac{\sum_s w_s \hat{\beta}_s}{\sum_s w_s}, \quad \hat{\sigma}^2 = \frac{1}{\sum_s w_s} \quad (3)$$

固定効果モデルによってコホートの結果を統合する場合は、真のコホート間差はないと仮定するため、 $\tau^2$  は存在せず上式において重みが  $w_s = (\sigma_s^2)^{-1}$  となる。

#### 効果の異質性(研究間ばらつき)の評価

統合研究では効果の研究間差の評価が不可欠である。推定値  $\hat{\beta}_s$  の分散に占める研究間差の割合は  $Q$  統計量で定量化される。

$$Q = \sum_s w_s (\hat{\beta}_s - \beta)^2$$

真の研究間差がないときの  $Q$  の期待値は自由度に

一致するため、過大な研究間差を  $I^2(\%)$  で表し

$$I^2 = (Q^2 - df) / Q \times 100$$

推定値に併記する ( $df = \text{研究数} - 1$ )。  $I^2 = 0$  は全く研究間の効果の違いがないことを意味し、100%に近いほど異質性が高いことを意味する。さらに  $Q$  が研究間差がない下ではカイ二乗分布に従うことを利用して検定も可能である。

ERFC ではフィブリノゲンの CHD に対するリスクを変量効果モデル、固定効果モデル両方で推定した。変量効果モデルの方が研究間差  $\tau^2$  の分、標準偏差はやや大きくなった。各パラメータに対して  $I^2$  の評価が行われ、強い研究間差が示唆された(表1上段)。

#### 曝露因子の変数型

連続尺度の曝露因子を統計モデルに投入する場合、連続値のまま、変数変換する、カテゴリ化する、など変数型の選択肢が多い。連続値として扱う場合は線形性の評価が必要である。線形性の評価方法としてはカテゴリ化して図示し、線形性を目視で確認することも重要である。多項式変換法や平滑化も可能である。連続尺度のまま用いる場合でも、変数変換によって推定値や研究間差が異なる可能性がある。表1の下段は連続尺度として扱う場合の変数変換による推定値の違いを示している。なお ERFC の例では推定値の違いは見られなかった。

#### 共変量調整

共変量として調整が必須な因子として年齢がある。曝露因子と同様に連続値のまま、変数変換、カテゴリ化が可能である。年齢の一次項による調整は不十分な可能性があり、多項式、他の因子との交互作用項の検討が必要な場合がある。年齢以外の因子は多くの場合、一次項を入れた調整で十分だが、推定値の変化や研究間差の確認は有用である。

表2に年齢を調整することによるフィブリノゲン効果の推定値の違いを示した。年齢は連続(線形)を共変量、5歳刻みのカテゴリを共変量、5歳刻みのカテゴリを層別因子、性別との交互作用、二次項、交互作用の二次項など様々な方法で行い、

フィブリノゲンの推定値が変わらないことを確認している。

### 曝露効果の修飾

効果の修飾の検討は、交互作用項を用いてサブグループによる曝露効果の違いを推定することである。個人データを直接解析する場合は、共変量  $X$  (たとえば性別;  $X=0$  男性、 $X=1$  女性) によって曝露効果が違う状態を(4)式でモデル化する。

$$h_{si}(t) = h_{0s}(t) \exp(\beta_s E_{si} + \gamma_s X_{si} + \delta_s E_{si} X_{si}) \quad (4)$$

男性の曝露効果は  $\beta_s$ 、女性の曝露効果は  $\beta_s + \delta_s$ 、性別による曝露効果の修飾(交互作用)は  $\delta_s$  に推定される。ERFC の検討(表 3 上段)では年齢、収縮期血圧、BMI による効果の修飾が示された。

人種や性別、曝露の測定系などコホート単位で測定された共変量は変量効果メタ回帰で検討する。(2) 式に共変量を加えて  $\delta_B$  を推定すると、たとえば、男性の割合が 10%多いコホートではフィブリノゲンの効果が  $\delta_B$  高い、といった推測ができる。

$$\beta_s = \beta + \delta_B X_s + \eta_s, \quad \eta_s \sim N(0, \tau^2) \quad (5)$$

ERFC の解析例では性別について、(4)で研究内交互作用の平均  $\delta_W$  を、(5)のメタ回帰で研究間交互作用  $\delta_B$  を推定している(表 3 下段)。標準誤差を比較すると前者が小さいこと、後者には研究間交絡が存在しうるかことから、研究内交互作用を重視すべきだとの例である。

### モデルの妥当性評価

予後を予測するためのモデル(以下、予測モデル)構築では、上記のような曝露効果の推定だけでなく、予測の妥当性の評価、複数モデルの比較が行われる。モデル評価や比較においても、複数のコホートを統合したことを考慮する必要がある。

複数コホートの予測モデルの妥当性評価の手段として Concordance index( C index )[Harrell 1992]、

Discrimination measure ( D measure ) [Royston 2004] が有用であると言われてきた。C index は、予測モデルで推定されたイベント発生順と実際の生存時間の長さが一致する程度の指標である。D measure は予測リスク分布の上半分と下半分のハザード比の平均値である。両者とも、コホートごとに指標を計算し(3)式と同様の重み付き平均を使って統合する。

これらの用途の一つとして、予測力を低くする要因の探索がある。ERFC では、37 コホートからなるデータベースにおける CHD イベントの予測モデルの判別力を評価するため、メタ回帰が行われた(図 1.A,B)。図 1.の A,B は C index、D measure それぞれを年齢に回帰した図である。丸の大きさはメタ回帰の際の重み(分散の逆数)を表す。年齢が高いコホートほど予測が当てはまりやすく、若い集団ほど予測が難しいことが推測される。

モデル間で妥当性を比較し、より良いモデルを選択する際には、C index や D measure の変化量が有用である。事例では既存のモデルに対数 CRP を追加したときの 2 指標の変化およびそのコホート間差について検討されている(図 4.C,D)。この事例では、対数 CRP によって年齢問わず当てはまりが改善することが示されている。

いずれの指標もコホートごとの推定値を統合する際、重みに何をを用いるかで結果は異なる。多くの場合、推定値の分散の逆数またはイベント数が用いられる。イベント数を用いた場合、イベント数の少ない大規模コホートの重みが軽くなることに注意が必要である。

### 考察

観察研究ではデザインに由来するバイアスのためにメタアナリシスによって精密な信頼区間のついた推定値を得ることに細心の注意が求められる(Greenland 1994, Thompson 1994)。ここでレビューした内容は通常の観察研究の解析でも求められるものであるが、統合研究ではさらに、データが複数のコホートからなることを常に意識し、研究間差を確認しながら評価を進めることが重要であ

る。研究数が十分にあれば、メタ回帰によって研究間差の原因を探索することも可能である。ここで紹介した方法は、ERFCからStataコードが提供されている。これらを使用することでより妥当な解析モデルの構築が期待できる。

統合研究は、その解析手法と一般化する集団の定義を入念に行うことで、精度が高い情報が提供できる手段であり、今後の観察研究に欠かせない研究方法である。統合研究に要求される解析手法のレビューや、実際のデータに適用する際の問題点の抽出は統計学的にも重要な課題であるといえる。本報告で言及していない統計学的課題として、測定誤差によるバイアス、効果の希薄化、研究単位での欠測等への対処がある。また、リスク予測モデルの評価に欠けているものとして、デザインの異なる研究の統合や欠測補完の影響、外的妥当性の評価がある。

EPOCH-JAPAN の各コホートにも対象者の特性、地域特性、コホートの時代区分など多くの多様性が存在し、要因・結果の関連がコホート間で一様であるかは検討すべき課題の一つである。今後、EPOCH-JAPAN で疾患予測モデルを構築する際は、関心のある曝露の全体効果の推定のみに着目するのではなく、コホートごとの効果や共変量の影響に異質性はないか、異質性の原因は何であるかなど検討し、作成したモデルの妥当性も比較検討しながら解析を実施する必要がある。本報告で整理した一連の手法は、現段階で最も有用な候補であると思われた。

## 文献

Borenstein M, Hedges LV, Higgins JPT, et al. Introduction to Meta-analysis. 2009. Wiley, West Sussex UK.

Greenland S. Can meta-analysis be salvaged? Am J Epidemiol. 1994;140:783-7.

Greenland S. A critical look at some popular meta-analytic method. Am J Epidemiol. 1994;140:290-6.

Harrell F, Califf RM, Pryor DB, et al. Evaluating the

Yield of Medical Tests. JAMA,1982;247:2543-6.

Pennells L, Kaptoge S, White IR, et al. Assessing risk prediction models using individual participant data from multiple studies. Am J Epidemiol. 2014;179:621-632

Riley RD, Lamvert PC, Staessen JA, et al. Meta-analysis of continuous outcomes combining individual patient data and aggregate data. Statist. Med. 2008; 27:1870-93.

Royston P and Sauerbrei W. A new measure of prognostic separation in survival data. Statist Med 2004;23:723-48.

Thompson SG. Why sources of heterogeneity in meta-analysis should be investigated BMJ. 1994;309:1351-5.

Thompson SG, Kaptoge S, White IR, et al. Statistical methods for the time-to-event analysis of individual participant data from multiple epidemiological studies. Int J Epidemiol. 2010;39:1345-59.

## E . 結論

Emerging risk factors collaboration におけるIPD メタアナリシスで用いた生存時間の予測モデルの構築手法およびモデル評価手法を整理、報告した。今後 EPOCH-JAPAN で疾患予測モデルを構築する際はこれら手法による解析、検討が有用であると思われた。

## F . 研究発表

### 1 . 論文発表

なし。

### 2 . 学会発表

なし。

## G . 知的財産権の出願・登録状況（予定を含む）

### 1 . 特許取得

なし。

### 2 . 実用新案登録

なし。

### 3 . その他

なし

表 1. 年齢で調整したフィブリノゲンの CHD イベントに対するハザード比  
(Emerging risk factors collaboration (ERFC)より)

Method	HR (95% CI)	Log HR, $\hat{\beta}$ (SE)	Heterogeneity		
			Between-study variance $\hat{\tau}^2$	P-value	$I^2$ (95% CI)
<b>Untransformed fibrinogen: log HRs per 1g/l increase</b>					
Random-effects meta-analysis	1.57 (1.47–1.67)	0.450 (0.033)	0.018	<0.0001	64% (48, 76)
Fixed-effects meta-analysis	1.52 (1.47–1.57)	0.419 (0.018)	NA	<0.0001	NA
<b>Transformed fibrinogen: log HRs per SD increase—random-effects meta-analysis</b>					
Untransformed fibrinogen	1.34 (1.29–1.40)	0.294 (0.022)	0.008	<0.0001	64% (48, 76)
Log fibrinogen	1.38 (1.32–1.45)	0.325 (0.025)	0.010	<0.0001	65% (48, 76)
Study-specific SD score fibrinogen	1.34 (1.29–1.40)	0.292 (0.021)	0.007	<0.0001	63% (45, 75)
Study-specific SD score log fibrinogen	1.37 (1.31–1.44)	0.316 (0.024)	0.009	<0.0001	64% (47, 76)
<b>Untransformed fibrinogen</b>					
Quadratic term for fibrinogen	0.96 (0.91–1.01)	−0.045 (0.027)	0.007	0.013	40% (7, 61)

NA: not applicable; SE: standard error.

表 2. 年齢調整によるフィブリノゲン効果の推定値への影響  
(Emerging risk factors collaboration (ERFC)より)

With adjustment for	HR (95% CI)	Log HR $\hat{\beta}$ (SE)	$\chi^2$	Heterogeneity		
				Between-study variance $\hat{\tau}^2$	P-value	$I^2$ (95% CI)
Age	1.57 (1.47–1.67)	0.450 (0.033)	181	0.018	<0.0001	64% (48, 76)
Age as 5-year age bands	1.57 (1.47–1.68)	0.451 (0.033)	183	0.018	<0.0001	64% (48, 76)
Stratification by 5-year age bands	1.57 (1.47–1.68)	0.451 (0.033)	182	0.018	<0.0001	64% (48, 76)
Age sex × age	1.57 (1.47–1.68)	0.450 (0.034)	180	0.018	<0.0001	65% (48, 76)
Age age <sup>2</sup>	1.56 (1.46–1.67)	0.447 (0.033)	179	0.018	<0.0001	64% (48, 76)
Age age <sup>2</sup> sex × age sex × age <sup>2</sup>	1.57 (1.47–1.67)	0.448 (0.034)	177	0.019	<0.0001	65% (49, 76)
Age smoking tchol sbp bmi <sup>a</sup>	1.38 (1.31–1.45)	0.320 (0.026)	156	0.006	0.028	35% (0, 58)

表 3. 共変量の効果修飾の程度とその研究間差の評価  
(Emerging risk factors collaboration (ERFC)より)

Potential effect modifier	Estimated interaction between the potential effect modifier and fibrinogen				Heterogeneity $I^2$ (95% CI)
	Number of cohorts	Number of subjects	Estimate $\delta$ (SE)	P-value	
Age (10 years)	31	154 211	−0.095 (0.029)	0.001	0% (0, 40)
Systolic blood pressure (10 mmHg)	31	154 211	−0.021 (0.010)	0.032	21% (0, 50)
Body mass index (5 kg/m <sup>2</sup> )	31	154 211	−0.079 (0.023)	<0.0001	3% (0, 31)
Total cholesterol (1 mmol/l)	31	154 211	−0.025 (0.014)	0.081	1% (0, 41)
<b>Sex: women vs men</b>					
Between-study interaction	31	154 211	0.120 (0.092)	0.21	NA
Within-study interaction	16	90 529	0.089 (0.061)	0.15	0% (0, 52)
Overall pooled interaction <sup>a</sup>	31	154 211	0.098 (0.051)	0.054	NA

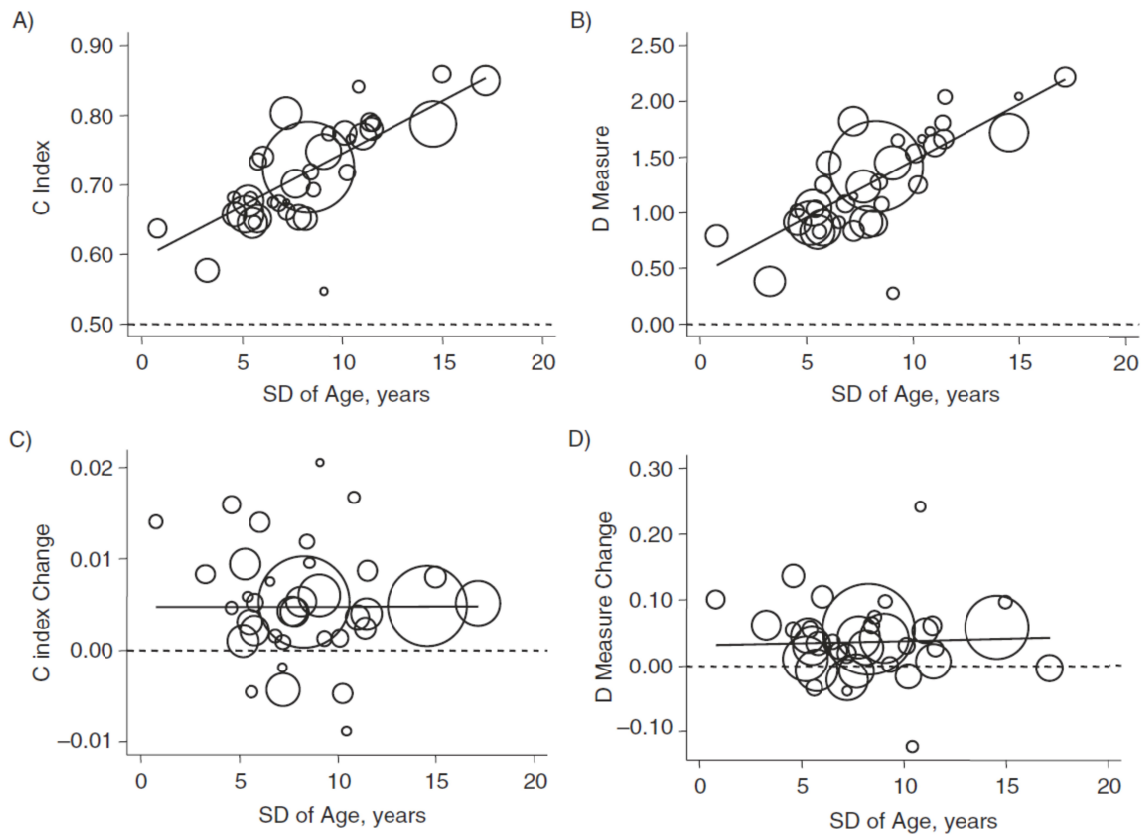


図 1. 判別指標の推定値の共変量へのメタ回帰  
(Emerging risk factors collaboration (ERFC)より)