臨床研究・疫学研究の多変量解析における変数選択

野間 久史
統計数理研究所
2022年10月24日
国立がん研究センター生物統計セミナー
e-mail: noma at ism.ac.jp
URL: http://www.ism.ac.jp/~noma/

臨床研究における多変量解析

- 複数の説明変数をモデル化した多変量モデルによる分析方法
- アウトカム変数の型ごとに異なるモデルが用いられる
  - 連続アウトカム：線形回帰モデル
  - 2値アウトカム：ロジスティック回帰モデル
  - 生存時間アウトカム：Cox回帰モデル
- STROBE声明，TRIPOD声明などのガイドラインでも、交絡調整のための多変量モデルによる調整解析は必須とされており、国際誌の査読をクリアするためには、現在、ほぼ必ず求められる統計解析手法となっている

Original Investigation

Clinical Risk Score for Persistent Postconcussion Symptoms Among Children With Acute Concussion in the ED

Roger Zemek, MD; Nick Barrowman, PhD; Stephen B. Freedman, MDCM, MSc; Jocelyn Gravel, MD; Isabelle Gagnon, PhD; Candice McGahern, BA; Mary Aglipay, MSc; Gurinder Sangha, MD; Kathy Boutis, MD; Darcy Beer, MD; William Craig, MDCM; Emma Burns, MD; Ken J. Farion, MD; Angelo Mikrogianakis, MD; Karen Barlow, MD; Alexander S. Dubrovsky, MDCM, MSc; Willem Meeuwisse, MD, PhD; Gerard Gioia, PhD; William P. Meehan III, MD; Miriam H. Beauchamp, PhD; Yael Kamil, BSc; Anne M. Grool, MD, PhD, MSc; Blaine Hoshizaki, PhD; Peter Anderson, PhD; Brian L. Brooks, PhD; Keith Owen Yeates, PhD; Michael Vassilyadi, MDCM, MSc; Terry Klassen, MD; Michelle Keightley, PhD; Lawrence Richer, MD; Carol DeMatteo, MSc; Martin H. Osmond, MDCM; for the Pediatric Emergency Research Canada (PERC) Concussion Team

IMPORTANCE Approximately one-third of children experiencing acute concussion experience ongoing somatic, cognitive, and psychological or behavioral symptoms, referred to as persistent postconcussion symptoms (PPCS). However, validated and pragmatic tools enabling clinicians to identify patients at risk for PPCS do not exist.

Editorial page 987

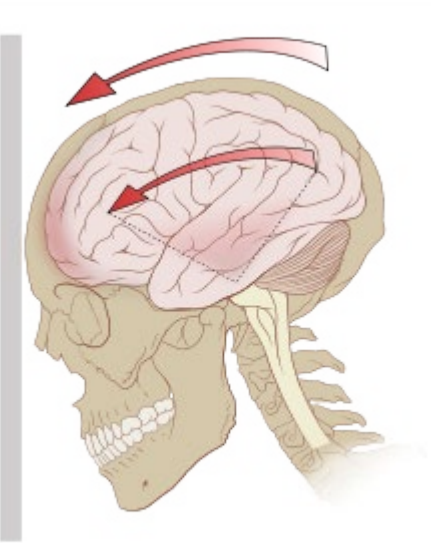
JAMA Report Video at jama.com

Supplemental content at jama.com

JAMA 2016; 315(10): 1014-25.

持続的な脳震盪後症候群

- 脳震盪の多くは、一過性のものであるが、小児の患者の３分の１ほどは、持続的な脳震盪後の症状に悩まされる
  - Persistent Postconcussion Symptoms (PPCS)
- しかし、このPPCSが起こるか否かを予測するための、Validateされた実用的な予測ツールは開発されていなかった
- Zemek et al. (2016) では、カナダのオタワ大学のグループを中心に、脳震盪で救急を受診した子どもを対象として、PPCSの発症を予測するためのリスクスコアを開発した



https://ja.wikipedia.org/wiki/%E8%84%B3%E9%9C%87%E7%9B%AA

PPCSのリスクスコアの作製

- 研究デザイン：前向きの多施設共同コホート研究
- 対象者：5歳から18歳の小児で、受傷後48時間以内に救急を受診した患者
- プライマリアウトカム：受傷後、28日以内のPPCSの発症

- あらかじめ先行研究によるエビデンスと専門家による意見をもとにして選定した、**46個の予後因子の候補を収集していた**
- **一定数の変数に絞り込みを行った上で**、ロジスティック回帰モデルによる、PPCSの発症を予測するためのリスクスコアを開発した

Zemek et al. (2016)

本日のお話

- 多変量モデルによる解析を行う際には、一般的に、多くの変数が候補として上がってくる（先ほどのJAMAの事例でも、46の候補があった）
- どの変数をモデル化すればよいのか？というのは、臨床研究のプラクティスの上でも、しばしば問題になる
- モデル化される変数の組によって、最終的な結論が変わることも
- どのような方法で、モデル化する変数の組を選べばよいのか？

多変量モデルによる解析の目的

- 臨床研究における多変量モデルによる解析には、大きく分けて２つの目的がある
- 臨床予測モデルの構築のため
  - 複数の予測に寄与する変数をモデル化して、将来の患者集団において、正確なアウトカムの予測を行うことが目的
- 交絡によるバイアスを調整するため
  - 関心のある治療法の効果を評価する際に、比較群間で分布の偏りがあるリスク要因がバイアス（交絡）を生じさせるため、それを調整することが目的

臨床予測モデルの構築

- ロジスティック回帰モデル

$$\Pr(Y = 1|\boldsymbol{x}) = \frac{\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p)}{1 + \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p)}$$

- 説明変数 $x_1, x_2, \ldots, x_p$ によって、結果変数 Y $(= 0,1)$ を予測するモデル
- 複数の予測に寄与する変数を組み合わせることによって、より高い精度での予測を行うことが可能に

Steyerberg (2019)

交絡によるバイアス

1980~90年代（日本）：乳がんの外科手術後のタモキシフェンの使用と再発の関係について調べた臨床研究（後ろ向きコホート研究；Sato and Matsuyama, 2003）

タモキシフェン	再発あり	再発なし	合計
使用	464	2085	2549
非使用	424	1928	2352

タモキシフェン使用群の再発割合：$464/2549 = 18.2\%$
タモキシフェン非使用群の再発割合：$424/2352 = 18.0\%$

リスク差（Risk Difference）：
$18.2\% - 18.0\% = 0.2\%\ (95\%\text{CI}: -2.0\%, 2.3\%), \text{P} = 0.873$

交絡によるバイアス

	リンパ節転移あり			リンパ節転移なし		
タモキシフェン	再発あり	再発なし	合計	再発あり	再発なし	合計
使用	368	847	1215	96	1238	1334
未使用	253	507	760	171	1421	1592

リスク差：−3.0%
(95%CI: − 7.2%, 1.2%), P=0.162

リスク差： − 3.5%
(95%CI: −5.6%, − 1.5%), P < 0.001

Matsuyama et al. (2000), Sato and Matsuyama (2003)

交絡によるバイアス

- 重症（高リスク）である患者ほど、有望であると思われる治療を割り付けられやすい傾向がある
- 「治療ありグループ」の患者に高リスクの患者が多く、「治療なしグループ」の患者に低リスクの患者が多いならば？？
- 単純な比較（全体の結果）は、純粋な「治療のありなし」の比較ではなく、「背後にある別なリスク要因（交絡要因）による高リスク群と低リスク群の比較」になってしまう
  - 真の治療効果 ≠ 見せかけの関連
  - 「交絡（Confounding）」というバイアスとして知られる

多変量モデルによる交絡調整

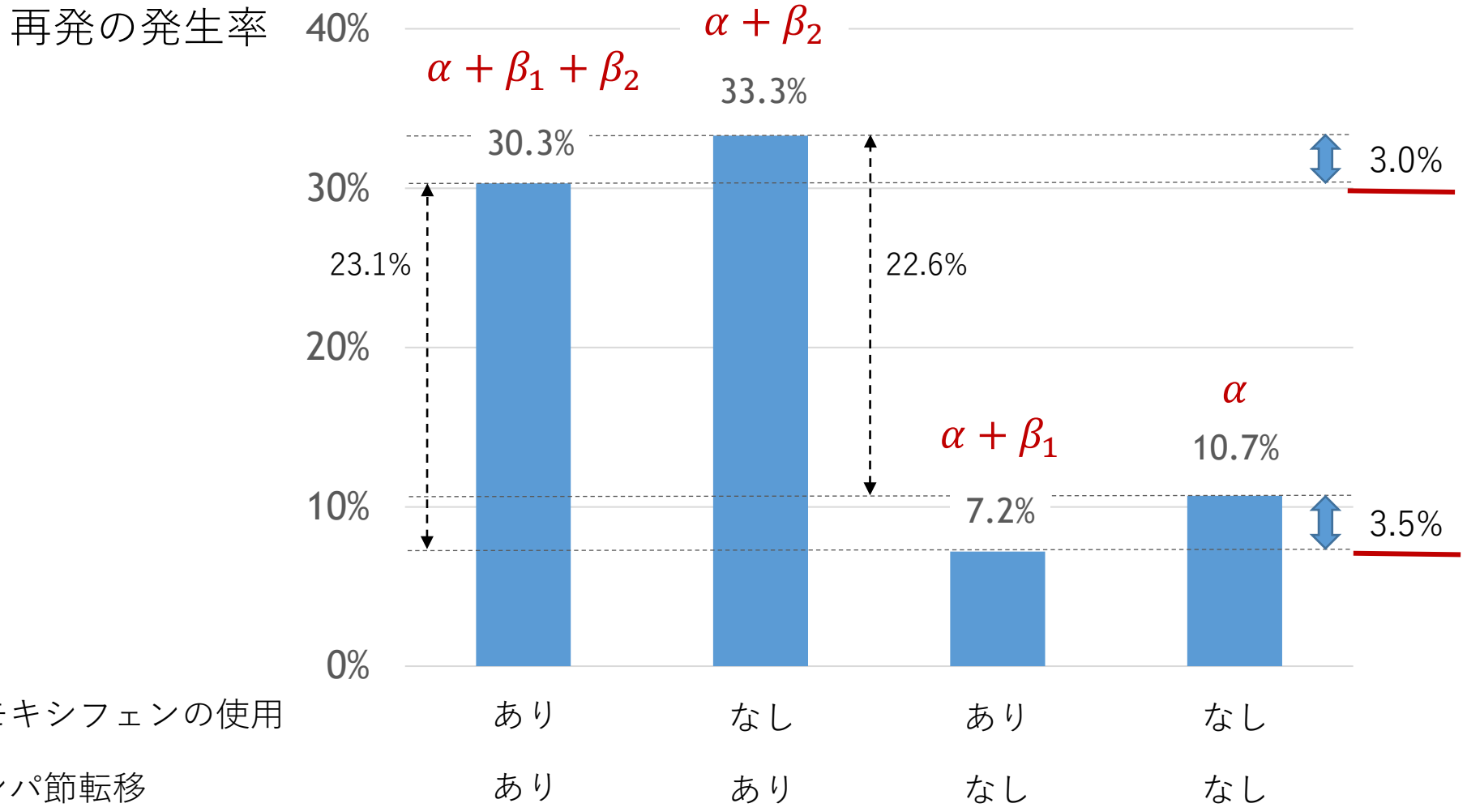
- イベント発生の確率のモデル

$$\Pr(Y = 1|X_1, X_2) = \alpha + \beta_1 X_1 + \beta_2 X_2$$

  - $Y = 1$（乳がんの再発あり），$= 0$（再発なし）
  - $X_1 = 1$（タモキシフェンの治療あり），$= 0$（治療なし）
  - $X_2 = 1$（リンパ節転移あり），$= 0$（転移なし）
- 乳がんの再発を起こす確率が、「タモキシフェンの治療の有無」と「リンパ節転移の有無」で説明することができ、上記のような数式でモデル化できるという仮定を置く
- 交絡要因 X_2 を調整した X_1 の効果は、β_1によって推定することができる

回帰モデルによる交絡調整

$$\Pr(Y = 1|X_1, X_2) = \alpha + \beta_1 X_1 + \beta_2 X_2$$



モデル構築の原理

- 臨床予測モデルの構築
  - 予測モデルの予測精度の最大化（将来の患者のアウトカムをより精度良く予測できるモデルが良いモデル）
- 交絡調整のための回帰モデル
  - 関心のある治療効果のパラメータの推定のバイアスの最小化（真の治療効果をよりバイアスなく推定できるモデルが良いモデル）
- それぞれの目的に応じて、いくつかの変数選択のための数学的な方法が開発されている

変数選択のための代表的な方法

- Stepwise法
  - 有意性による基準，情報量規準
  - Forward, Backward法
- Univariate Screening
- Regularized regression（LASSOなど）
- Change-in-estimate criterion
- Background Knowledge
- Bayesian Model Averaging
- Causal graph

Heinze et al. (2018), Steyerberg (2019)

Stepwise法：有意性による基準

- 数学的な基準で、最も有意性の大きな変数の組を、機械的に選択するためのアルゴリズム
- Forward法：なにも説明変数を加えていないモデルからスタートして、有意性の大きな候補の変数を１つ１つ加えていく
- Backward法：すべての候補変数を含むFull modelからスタートして、有意性の小さな変数を１つ１つ除外していく

Akaike (1973), Mallows (1973), Schwarz (1978)

Stepwise法：Forward法

- 1. すべての候補となる共変量を１つずつ含めた単変量モデルと、なにも説明変数を加えていないモデルとの間に差があるかどうかを検定し、最もP値が小さくなった変数を１つ選ぶ。そのP値が、0.05未満であれば、その変数を加える。
- 2. １つ前のモデルと、そのモデルに対して残りの候補となる共変量を１つずつ加えたモデルの間に、差があるかどうかを検定し、最もP値が小さくなった変数を１つ選ぶ。そのP値が、0.05未満であれば、その変数を加える。
- 3. 2. のプロセスを繰り返し、P値が0.05を切る変数がなくなったら終了。

Stepwise法：Backward法

- 1. すべての候補となる変数を加えたフルモデルと、１つずつの変数を除いたモデルに、差があるかどうかを検定し、最もP値が小さくなった変数を選ぶ。そのP値が、0.05未満であれば、その変数を除く。
- 2. １つ前のモデルと、そのモデルから、残りの候補となる共変量を１つずつ除いたモデルの間に、差があるかどうかを検定し、最もP値が小さくなった変数を選ぶ。そのP値が、0.05未満であれば、その変数を除く。
- 3. 2. のプロセスを繰り返し、P値が0.05を切る変数がなくなったら終了。

Stepwise法のアルゴリズム

- 有意水準は緩めに設定してもよい（e.g., 20%，50%など）
  - サンプルサイズが小さい条件下では、有意水準を厳しくしすぎると（5%など）、予測モデルの性能は悪くなる
  - ただし、弱い関連しか持たない変数が残ることもある
- 一般的には、Backward法のほうが好まれる
  - Full Modelからスタートするので、すべての変数の関連を同時に評価することができる
  - 相関のある予測性能の高い変数の組が同時にモデルに残る可能性がある（Forward法ではすべてドロップする可能性もある）

Steyerberg (2019)

Stepwise法：情報量規準による方法

- Akaike’s Information Criterion (AIC)
  - AIC = −2 loglikelihood + 2p
  - p: モデル中のパラメータ数
- Bayesian Information Criterion (BIC)
  - BIC = −2 loglikelihood + p log(n)
  - n: サンプルサイズ
- 「推定されたモデル」と「真の分布」の近さ・遠さを測るための基準スコア
- 小さいほど、より良いモデルであると判断される

https://www.ism.ac.jp/akaikememorial/index.html

Stepwise法：Forward法

- 1. すべての候補となる共変量を１つずつ含めた単変量モデルにおけるAIC/BICを計算し、最もAIC/BICが小さくなった変数によるモデルを初期モデルとする。
- 2. １つ前のモデルに対して、残りの候補となるすべての説明変数を１つずつ加えたモデルのAIC/BICを計算し、最もAIC/BICが小さくなった変数によるモデルを候補モデルとする。そのAIC/BICが１つ前のモデルよりも小さければ、その変数を加える。
- 3. 2. のプロセスを繰り返し、AIC/BICが小さくなる変数がなくなった時点で終了とする。

Stepwise法：Backward法

- 1. すべての候補となる変数を加えたフルモデルを初期モデルとする。
- 2. １つ前のモデルから、候補となる説明変数を１つずつ除いたモデルのAIC/BICを計算する。その中で、最もAIC/BICが小さくなったモデルを候補モデルとする。そのAIC/BICが１つ前のモデルよりも小さければ、その変数を除く。
- 3. 2. のプロセスを繰り返し、AIC/BICが小さくなる変数がなくなった時点で終了とする。

Stepwise法の問題点

- 変数選択の不安定性
  - 選ばれる「変数の組」が安定しない
  - 同じコホートから、一部の少し違う対象者のサブセットを除外してStepwise法にかけると、違う変数の組が残るなど
- 回帰係数の推定量は不偏性を失う（バイアスが生じる）
- 標準誤差，信頼区間，P値も不正確なものとなる
  - 最終モデルは、何度も何度も回帰モデルの分析を行い、それぞれの数学的な基準で選定された後のモデルであるため
  - 交絡調整のためのモデル選択には不適切な方法である

Greenland (1989), Steyerberg et al. (1999, 2000), Steyerberg (2009)

“Noise Variables” の影響

- 実際にはアウトカムと関連のない “Noise Variables” は、Stepwise法でも、かなりのものが選択されることも
  - Derksen and Keselman (1992) のシミュレーションでは、30-70%の選ばれた変数がノイズであったとも
- Noise Variables が誤って含まれてしまうと、当然ながら、予測性能も下がる
- また、少数の Noise Variables の混入はそれほど大幅に予測性能を下げることはなく、重要な予後因子が含まれないことのほうが予測性能への影響は大きい

Steyerberg (2009)

Univariate Screening

- 単変量解析を行って、P < 0.05, 0.1, 0.2, 0.5 などの基準に合致した変数を残して、多変量モデルを構築する
- または、そうして選ばれた変数の組を、Backward Stepwise法にかけるというアプローチ
- あらかじめ、ほとんど関連の認められなかった変数を除いて検討を行うことができる
- 大規模なデータベースを利用した、共変量の多い研究などでは、計算コストを節減することもできる
- 実践的には、よく採用される方法であるが、Stepwise法全般と同じ欠点を有する方法である

Steyerberg (2009)

Regularized Regression

- ロジスティック回帰モデルなどでは、サンプルサイズが小さい条件下では、回帰パラメータの大きさを過大に推定するバイアスがあることが理論的に示されている
- 過大推定バイアスを補正するために、回帰パラメータに制約をつけた推定方法
- 制約の数理的な性質により、全体的に回帰係数は小さめの値をとる（帰無仮説の方向に縮小される）
- LASSO法：一部の変数が、強く縮小され、点推定値が「0」になるというアルゴリズム（自動的に変数選択のアルゴリズムにもなる）

Vach et al. (2001), Steyerberg (2000, 2001)

LASSOの原理

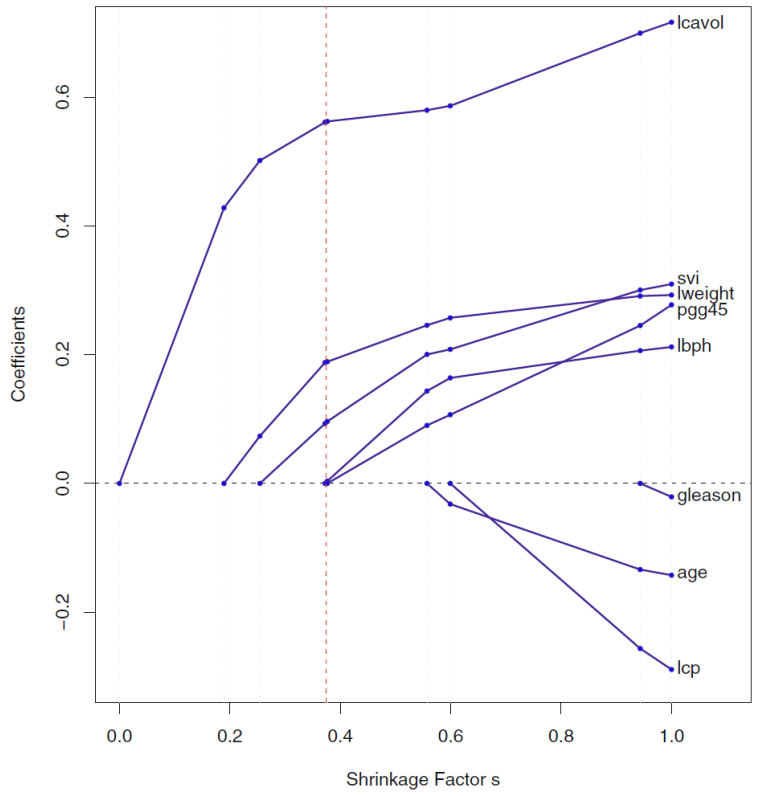


FIGURE 3.10. *Profiles of lasso coefficients, as the tuning parameter t is varied. Coefficients are plotted versus $s = t/\sum_1^p |\hat{\beta}_j|$. A vertical line is drawn at $s = 0.36$, the value chosen by cross-validation. Compare Figure 3.8 on page 65; the lasso profiles hit zero, while those for ridge do not. The profiles are piece-wise linear, and so are computed only at the points displayed; see Section 3.4.4 for details.*

- 通常の最尤法とは異なる回帰パラメータの推定を行う
- ロジスティック回帰にモデル化した説明変数の回帰係数が、共通の分布（ラプラス分布）に従うという仮定を置き、帰無仮説方向への縮小を行う
- バイアスはかかるものの、全体としての回帰パラメータの真値との誤差（平均二乗誤差）は小さくなる（＝予測精度が改善される可能性）
- Shrinkage Factor *s* が小さくなるほど、関連が相対的に弱い変数の回帰パラメータは 0 に近づいていく（＝変数選択が自動的に行われる！）
- *s* は、Cross-validationなどで最適値を求めることが一般的

Hastie, Tibshirani and Friedman (2009)

Change-in-estimate criterion

- X_1という治療・曝露の変数の効果に関心があるケース
- $M_1: \beta_0 + \beta_1 X_1 + \beta_2 X_2$
- $M_2: \gamma_0 + \gamma_1 X_1$
  - X_2という調整変数を加えたモデルと加えなかったモデル
- $\text{Relative Change}(\%) = 100 \times (\gamma_1 - \beta_1)/\beta_1$
- X_2を除くことによって、X_1の回帰係数がどの程度変わるか？
- 一定の水準（%）以上の変化（e.g., 10%）があるかどうかで、調整変数 X_2 を加えるかどうかを決める方法

Background Knowledge

- Stepwise法やChange-in-estimate criterionなどの完全に数理的なアルゴリズムのみで変数の選択を行うと、数学的な基準に合致した変数のみが機械的に選ばれてしまう
- 最終的に得られる変数の組は、しばしば臨床的・生物学的な知見から既知の重要な変数が抜け落ちてしまう
- １度１度の研究の統計的なばらつきによって、誤った変数の組が選ばれることもしばしばある
- 臨床的・生物学的な知見から、モデルに含める変数の候補を選ぶ（e.g., 肺がんの予後研究では、年齢は必ず調整すべき）

Greenland (1989), Rothman et al. (2008)

Bayesian Model Averaging

- p 個の説明変数をもつ観測データ D に対して、すべての考え得るモデル $M = \{M_1, \ldots, M_m\}$ を考える（$m = 2^p$）。このとき、関心のあるパラメータ δ の「平均化」された事後分布

$$\Pr(\delta|D) = \sum_{i=1}^{m} \Pr(\delta|M_i, D) \Pr(M_i|D)$$

モデル M_i のもとでの事後分布

モデル M_i の事後確率（寄与率を調整する重みに！）

を用いた推測を行う方法

- モデル選択の不確実性を推測・予測に取り込んだ方法

Bayesian Model Averaging

- 「どの変数が含まれれば、より確からしいモデルになるか」を白黒（含める・含めない）で決めるのではなく、濃淡をつけ、その重みづけ平均によって、最終的なモデルを作成する
- シミュレーション実験によって、Stepwise法よりも正確に、真の関連がある変数を採用し、ノイズ変数を除外する傾向があることが示されている
- Framingham研究への応用などでは、Stepwise法よりも高い予測性能を示した

Raftery et al. (1997), Wang et al. (2003)

Causal graphs

- 治療・曝露，交絡，アウトカム間の原因・結果の関連性を、矢印を用いたグラフで表したもの
  - Directed Acyclic Graph (DAG) と言われる
- 統計的因果推論の領域で発展した方法
- 交絡調整のための解析の変数選択の手法がいくつか開発されている

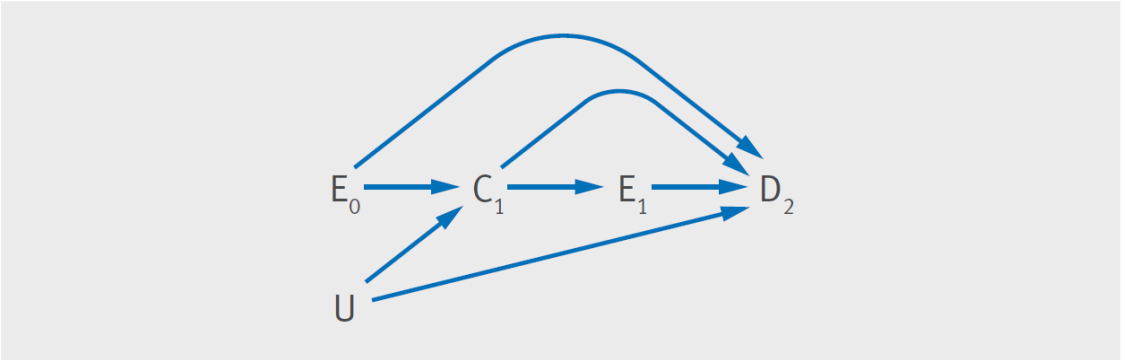
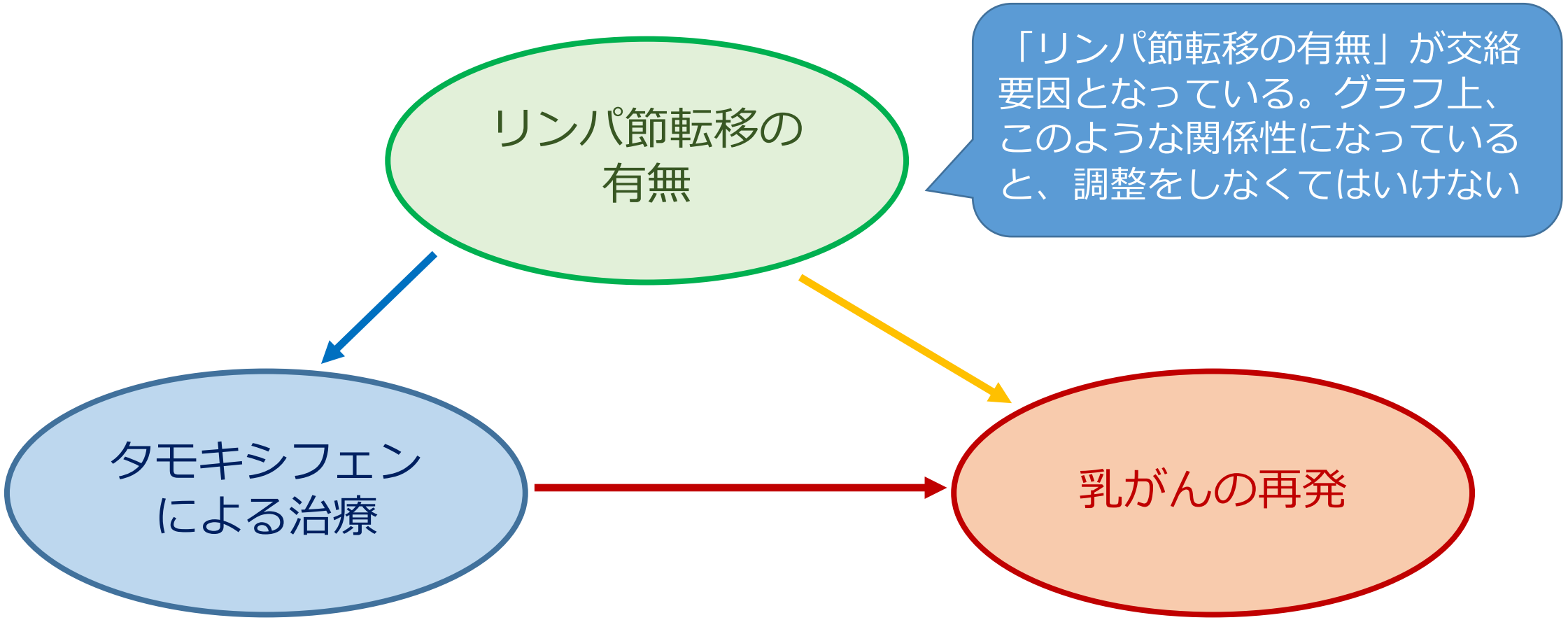


Fig 1 | Causal diagram showing time varying confounding affected by past exposure

Mansournia et al. (2017)

DAGと交絡の例



Causal graphによる変数選択アルゴリズム

- バックドアアルゴリズムなどの方法によって、設定された因果グラフが正しいという仮定のもとで、バイアスのない治療効果の推定値を得ることができる
- より発展的な Disjunctive Causal Criterion などによる変数選択の手法も提案されている
- ただし、①測定されていない交絡要因がひとつもない、②設定されている因果グラフが完全に正しい、という検証不可能な強い仮定があり、現実的にバイアスのない推定値が得られる保証は必ずしもない

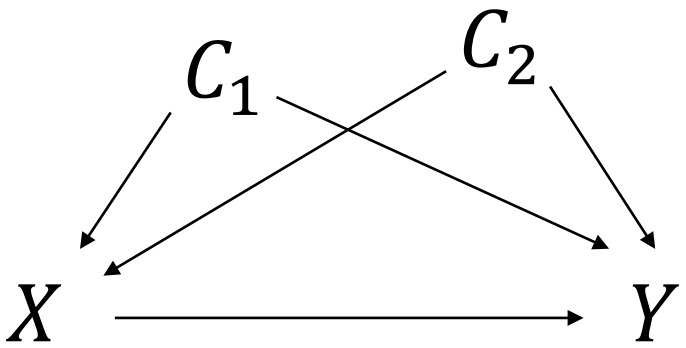
Greenland and Robins (1999), VanderWeele and Shpitser (2011)

Causal graphsの利用

- 実践において、バックドア法などの変数選択アルゴリズムの直接的な適用は難しいが、変数間の臨床的・生物学的な関連性を要約し、背景知識に基づく変数選択を行う際に、Causal graphそのものは有用なツールになる
- 実践においても、交絡調整のための多変量解析の変数選択の参考ツールとして、広く用いられている

交絡要因の必要条件とDAG

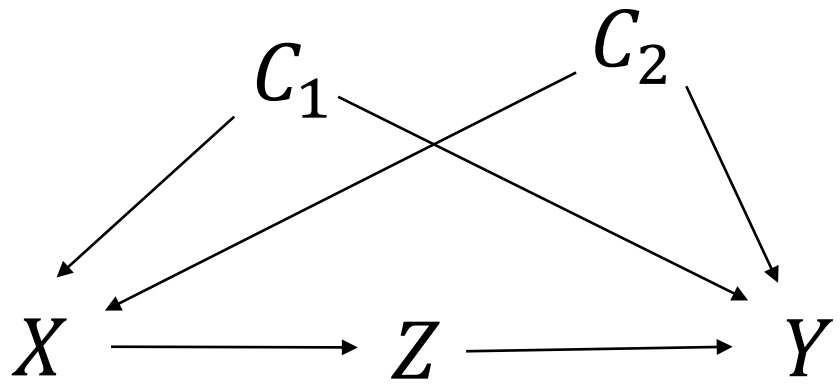
- 交絡要因の定義（必要十分条件）は論理的に与えられないが、必要条件は与えられる
- ① アウトカム変数のリスク要因（予防的要因）である
- ② 治療・曝露変数と関連を持つ



C_1, C_2は調整すべき交絡要因

交絡要因の必要条件とDAG

- ③ 治療・曝露とアウトカムの間の中間的な変数にはなっていない
  - 中間変数になっている変数は交絡要因にはならない上に、調整するとバイアスを生じさせる



Zは、中間変数なので、
調整してはいけない

“Parsimony” の原理

- 重要性の低い、無駄な情報は、加えても予測の性能を向上させないだけではなく、悪化させる可能性もある
- 少数の高い予測能力のある変数を用いることが、オーバーフィッティングを防ぎ、予測性能も高くなる傾向がある
- 予測モデルの実用化の際も、測定するべき必要な情報が絞られる

Ockham chooses a razor

https://www.google.co.jp/url?sa=i&source=images&cd=&cad=rja&uact=8&ved=2ahUKEwj8muvk6aXeAhXGzLwKHd1zBxYQjRx6BAgBEAU&url=https%3A%2F%2Fwww.safalniveshak.com%2Flatticework-mental-models-occams-razor%2F&psig=AOvVaw0fzAf-z8CDpgRJZs-hAc6S&ust=1540702375243042

Event-per-variable (EPV) の基準

- 対象者集団に対して、モデル化する変数の数が多すぎると、少ないデータの情報量で複雑なモデルを推定しなくてはならなくなるため、推定が不安定／不可能に
- ロジスティック回帰，Cox回帰モデルによる予測モデルの構築においては、EPVは10 or 15以上はあることが望ましい（Harrell et al, 1984; Harrell, 2015）
- 条件によっては、それ以下のEPVでもモデルの構築は可能であるとの報告もあるが、一般的には、変数を多くし過ぎることは推奨されない（van Smeden, 2016, 2018; Riley et al., 2018）

モデルに対する格言

- そもそも現象を単純化したロジスティック回帰，Cox回帰モデルなどの数学的なモデルが完全に正しいケースはある？
- “We do not accept the notion that there is a simple “true model” in the biological science.” (Burnham and Anderson, 2002)
- “We recognize that true models do not exist. … A model will only reflect underlying patterns, and hence should not be confused with reality.” (Steyerberg, 2009)

Sauerbrei *et al. Diagnostic and Prognostic Research* (2020) 4:3
https://doi.org/10.1186/s41512-020-00074-3

Diagnostic and
Prognostic Research

COMMENTARY Open Access

State of the art in selection of variables and functional forms in multivariable analysis—outstanding issues

Willi Sauerbrei[1*], Aris Perperoglou[2], Matthias Schmid[3], Michal Abrahamowicz[4], Heiko Becher[5], Harald Binder[1], Daniela Dunkler[6], Frank E. Harrell Jr[7], Patrick Royston[8], Georg Heinze[6] and for TG2 of the STRATOS initiative

Abstract

Background: How to select variables and identify functional forms for continuous variables is a key concern when creating a multivariable model. Ad hoc 'traditional' approaches to variable selection have been in use for at least 50 years. Similarly, methods for determining functional forms for continuous variables were first suggested many years ago. More recently, many alternative approaches to address these two challenges have been proposed, but knowledge of their properties and meaningful comparisons between them are scarce. To define a state of the art and to provide evidence-supported guidance to researchers who have only a basic level of statistical knowledge, many outstanding issues in multivariable modelling remain. Our main aims are to identify and illustrate such gaps in the literature and present them at a moderate technical level to the wide community of practitioners, researchers and students of statistics.

Methods: We briefly discuss general issues in building descriptive regression models, strategies for variable selection, different ways of choosing functional forms for continuous variables and methods for combining the selection of variables and functions. We discuss two examples, taken from the medical literature, to illustrate problems in the practice of modelling.

Results: Our overview revealed that there is not yet enough evidence on which to base recommendations for the selection of variables and functional forms in multivariable analysis. Such evidence may come from comparisons between alternative methods. In particular, we highlight seven important topics that require further investigation and make suggestions for the direction of further research.

Conclusions: Selection of variables and of functional forms are important topics in multivariable analysis. To define a state of the art and to provide evidence-supported guidance to researchers who have only a basic level of statistical knowledge, further comparative research is required.

Keywords: Descriptive modelling, Methods for variable selection, Spline procedures, Fractional polynomials, Categorisation, Bias, Shrinkage, Empirical evidence, STRATOS initiative

STRATOS Initiative から公表された変数選択に関する “state of the art” についてまとめられたレビュー論文。

結論としては「現状、多変量解析における変数選択や関数形についての推奨を与えるための基礎となるエビデンスはまだ十分にないことが明らかになった」とのこと。

STRATOSでのガイドライン整備のための研究も、まだこれから進められるというのが現状のようである。

予測モデルによる変数選択

- いかなる数学的基準による変数選択アルゴリズムでも、理論上の「真のモデル」を正確に同定することはできない
- そもそも、現実的には「真のモデル」自体が存在せず、予測モデルの目的は「高い予測精度を達成すること」である
- 予測モデルは、予測のためのツールとしての便宜上の ”Working model”
- それぞれの手法の長短を理解した上で、慎重に活用する必要がある

Original Investigation

Clinical Risk Score for Persistent Postconcussion Symptoms Among Children With Acute Concussion in the ED

Roger Zemek, MD; Nick Barrowman, PhD; Stephen B. Freedman, MDCM, MSc; Jocelyn Gravel, MD; Isabelle Gagnon, PhD; Candice McGahern, BA; Mary Aglipay, MSc; Gurinder Sangha, MD; Kathy Boutis, MD; Darcy Beer, MD; William Craig, MDCM; Emma Burns, MD; Ken J. Farion, MD; Angelo Mikrogianakis, MD; Karen Barlow, MD; Alexander S. Dubrovsky, MDCM, MSc; Willem Meeuwisse, MD, PhD; Gerard Gioia, PhD; William P. Meehan III, MD; Miriam H. Beauchamp, PhD; Yael Kamil, BSc; Anne M. Grool, MD, PhD, MSc; Blaine Hoshizaki, PhD; Peter Anderson, PhD; Brian L. Brooks, PhD; Keith Owen Yeates, PhD; Michael Vassilyadi, MDCM, MSc; Terry Klassen, MD; Michelle Keightley, PhD; Lawrence Richer, MD; Carol DeMatteo, MSc; Martin H. Osmond, MDCM; for the Pediatric Emergency Research Canada (PERC) Concussion Team

IMPORTANCE Approximately one-third of children experiencing acute concussion experience ongoing somatic, cognitive, and psychological or behavioral symptoms, referred to as persistent postconcussion symptoms (PPCS). However, validated and pragmatic tools enabling clinicians to identify patients at risk for PPCS do not exist.

Editorial page 987

JAMA Report Video at jama.com

Supplemental content at jama.com

JAMA 2016; 315(10): 1014-25.

Statistical Analysisより

- Forty-six variables were selected a priori for assessment based on a national planning meeting, recent systematic reviews, previous studies, and clinical experience.[16]
- All reliable variables associated with PPCS (*P* < .20) were entered into a multivariable model using forward stepwise binary logistic regression analysis (*P* = .05 included but *P* = .10 removed).

- 46の変数が、研究計画会議に基づく事前の評価、最近のシステマティックレビュー、先行研究、臨床的な経験によって選択された（＝Background Knowledge）。
- PPCSと相応の関連があったもの（単変量解析で P < .20）から、Forward Stepwise法によって、多変量モデルに含める変数の選択を行った (P = .05 included but P = .10 removed)。

Resultsより

- The final multivariable model included (1) age, (2) sex, (3) prior concussion with symptom duration of longer than 1 week, (4) physician-diagnosed migraine history, (5) headache, (6) sensitivity to noise, (7) fatigue, (8) answering questions slowly, and (9) abnormal tandem stance (Table 5).

- 最終的な多変量モデルには、上記の９つの変数が選ばれた。

Zemek et al. (2016)

Table 5. Selected Predictor Variables for Multivariable Model of Persistent Postconcussive Symptoms (PPCS) at 28 Days in the Derivation Cohort[a]

	No. of Risk Points for PPCS	Odds Ratio (95%CI)	*P* Value
Age group, y			
5-7	0	1 [Reference]	<.001
8-12	1	1.54 (1.09-2.19)	
13-<18	2	2.31 (1.62-3.32)	
Sex			
Male	0	1 [Reference]	<.001
Female	2	2.24 (1.78-2.82)	
Prior concussion and symptom duration			
No prior concussion; symptom duration <1 wk	0	1 [Reference]	.01
Prior concussion; symptom duration ≥1 wk	1	1.53 (1.10-2.13)	
Physician-diagnosed migraine history			
No	0	1 [Reference]	.001
Yes	1	1.73 (1.24-2.43)	
Answering questions slowly			
No	0	1 [Reference]	.008
Yes	1	1.37 (1.08-1.74)	
Balance Error Scoring System tandem stance No. of errors			
0-3	0	1 [Reference]	.02
≥4 or Physically unable to undergo testing	1	1.31 (1.04-1.66)	
Headache			
No	0	1 [Reference]	.01
Yes	1	1.66 (1.11-2.48)	
Sensitivity to noise			
No	0	1 [Reference]	.002
Yes	1	1.47 (1.15-1.87)	
Fatigue			
No	0	1 [Reference]	<.001
Yes	2	1.84 (1.37-2.46)	

[a] There were 1701 patients in the derivation cohort included in the primary analysis.

リスクスコアの計算・解釈のしやすさのために、連続変数をカテゴリ化して、予測モデルを構築する方針もよく採られる

この研究では、左記のカテゴリに割り振られたスコアごとに、整数値のスコア（0-12点）を計算し、リスクスコアを計算することに（Sullivan et al. (2004) の方法による）

Zemek et al. (2016)

交絡調整モデルの変数選択

- 現状では、いかなる変数選択アルゴリズムでも、バイアスのない推定値を確実に得るためには、現実的にはまず成立することのない強い仮定が必要／バイアスや推測の妥当性の深刻な問題がある（理想的な答えを与えてくれる数学的な手法は存在しない）
- 安易に使用すると、臨床上、重要な変数が、数学的な基準に合致せずに、モデルから除外されるリスクも
- 既存のエビデンスから、交絡要因の候補となる変数の選定（Background Knowledgeの活用）を行うことは必須

疫学の一流誌のレビュー①

Table 1 Variable selection methods used in major epidemiologic journals in 2008

Selection technique	American Journal of Epidemiology		Epidemiology		European Journal of Epidemiology		International Journal of Epidemiology	
	n	%	*n*	%	*n*	%	*n*	%
Prior knowledge	50	29	11	28	13	30	9	20
Effect estimate change	31	18	6	15	3	7	4	9
Stepwise selection	27	16	9	23	10	23	13	29
Modern methods (shrinkage, penalized regression)	0	0	0	0	0	0	0	0
Other (e.g., principal components, propensity scores)	2	1	4	10	1	2	2	4
Not described	61	36	10	25	17	39	17	38
Total	171		40		44		45	

交絡調整のための多変量解析が多い疫学の領域では、Prior Knowledgeを利用した変数選択が行われることが多い（2008年の疫学上位4誌）

Walter and Tiemeier (2009)

疫学の一流誌のレビュー②

Table 1: Variable selection methods used in explicative studies published in four major epidemiological journals in 2015

	American Journal of Epidemiology	Epidemiology	European Journal of Epidemiology	International Journal of Epidemiology	Total
Prior knowledge or causal graphs	55 (47%)	33 (59%)	27 (46%)	31 (52%)	146 (50%)
Prior knowledge or causal graphs only	*40 (34%)*	*29 (52%)*	*19 (32%)*	*28 (47%)*	*116 (40%)*
Change in estimate	20 (17%)	5 (9%)	5 (8%)	4 (7%)	34 (12%)
Stepwise	5 (4%)	3 (5%)	7 (12%)	1 (2%)	16 (5%)
Univariate analyses	16 (14%)	4 (7%)	5 (8%)	1 (2%)	26 (9%)
Other	3 (3%)	1 (2%)	1 (2%)	0 (0%)	5 (2%)
Insufficiently detailed	42 (36%)	16 (29%)	24 (41%)	25 (42%)	107 (37%)
Total	118	56	59	59	292

Results are reported as frequency (%). More than one method could be used in each study; as such, percentages do not add up to 100%.

交絡調整のための多変量解析が多い疫学の領域では、Prior Knowledgeを利用した変数選択が行われることが多い（2015年の疫学上位4誌）

Talbot and Massamba (2019)

Research

Open access

Work related sexual harassment and risk of suicide and suicide attempts: prospective cohort study

BMJ 2020 ; 370 doi: https://doi.org/10.1136/bmj.m2984 (Published 02 September 2020)
Cite this as: *BMJ* 2020;370:m2984

See other articles in issue 8258

Linked Editorial
Sexual harassment and suicide

Article Related content Metrics Responses Peer review

Linda L Magnusson Hanson, associate professor [1], Anna Nyberg, associate professor [1], Ellenor Mittendorfer-Rutz, professor [2], Fredrik Bondestam, director [3], Ida E H Madsen, senior researcher [4]

Author affiliations

Correspondence to: L Magnusson Hanson linda.hanson@su.se

Accepted 13 July 2020

Article tools
PDF 4 responses
Respond to this article
Data supplement
Print
Alerts & updates
Citation tools
Request permissions

https://www.bmj.com/content/370/bmj.m2984

Statistical Analysisより

- We estimated the risk of suicide or suicide attempt by using proportional hazard regression analyses with age as the underlying time scale.
- The analyses were adjusted for sex, family type, country of birth, educational level, and income, as these types of factors have been found to be associated with workplace sexual harassment and risk factors for suicidal behaviour.[7 17 18]

- 自殺や自殺企図のリスクを評価するために、比例ハザード回帰分析（Cox回帰分析）を行った。
- 職場でのセクシャルハラスメントと関連し、自殺行動のリスク要因として、既存の研究から知られている、性別、家族の種類、出生国、教育水準、収入などの要因を調整した（＝Background Knowledgeによる変数選択）。

Table 2 | Results from Cox regression analyses on workplace sexual harassment stratified by sex, presented as hazard ratios (HR) and 95% confidence intervals with and without adjustment for covariates

	All			Men			Women		
	No with valid data	No of cases	HR (95% CI)	No with valid data	No of cases	HR (95% CI)	No with valid data	No of cases	HR (95% CI)
Suicide									
Model 0*	85 205	125	2.23 (1.19 to 4.16)	40853	87	2.99 (1.09 to 8.18)	44 353	38	3.20 (1.39 to 7.33)
Model 1†	84 238	124	2.82 (1.49 to 5.34)	40421	86	2.62 (0.95 to 7.19)	43 817	38	2.94 (1.28 to 6.76)
Model 2‡	83 048	121	2.51 (1.29 to 4.90)	39877	84	2.62 (0.95 to 7.22)	43 171	37	2.39 (0.98 to 5.80)
Model 3§	82 860	121	2.47 (1.25 to 4.87)	39794	84	2.60 (0.92 to 7.34)	43 066	37	2.25 (0.91 to 5.56)
Suicide attempts									
Model 0*	84 556	816	1.54 (1.19 to 2.01)	40540	397	1.79 (1.03 to 3.11)	44 016	419	1.49 (1.10 to 2.02)
Model 1†	83 600	799	1.59 (1.21 to 2.08)	40111	391	1.80 (1.03 to 3.13)	43 489	408	1.49 (1.09 to 2.02)
Model 2‡	82 419	786	1.55 (1.18 to 2.04)	39570	385	1.78 (1.02 to 3.11)	42 849	401	1.44 (1.06 to 1.98)
Model 3§	82 233	785	1.56 (1.18 to 2.05)	39488	384	1.77 (1.01 to 3.12)	42 745	401	1.47 (1.07 to 2.02)

*Unadjusted analyses.
†Adjusted for sex, birth country, family situation, education, and income at baseline.
‡Adjusted for sex, birth country, family situation, education, income, and poor mental health at baseline.
§Adjusted for sex, birth country, family situation, education, income, demands, control, social support at work, workplace bullying, and poor mental health at baseline.

Model 0が単変量解析、Model 1が前頁の変数による多変量解析、その他、いくつかの追加のリスク要因を加えた調整解析の結果が、Model 2, 3。このように、複数の変数の組での解析結果を併記して、考察を行う論文は増えている。

Hanson et al. (2020)

Resultsより

- In the Cox regression analyses (table 2), the hazard ratio for completed suicide was 2.23 (95% confidence interval 1.19 to 4.16) for any workplace sexual harassment. The hazard ratio was considerably higher when we adjusted for sex. After adjustment for sex, birth country, family type, educational level, and income, the hazard ratio was 2.82 (1.49 to 5.34). This corresponded to a population attributable fraction of 0.06.

- 単変量解析のHRは、2.23 (95%CI: 1.19, 4.16)。性別を調整すると、HRはかなり大きくなる。さらに、Model 1の多変量モデルでの調整解析を行うと、HRは2.82 (95%CI: 1.49, 5.34) となる。集団寄与割合としては、6%ほどとなる。

Resultsより

- Further adjustment for baseline mental health and working conditions resulted in a more than twofold higher risk of suicide among people exposed to workplace sexual harassment (hazard ratios 2.51 (1.29 to 4.90) and 2.47 (1.25 to 4.87), respectively).

- ベースラインでのメンタルヘルスや就労条件による、さらなる調整解析（Model 3, 4）を行っても、2倍以上の自殺のリスクが認められた。
- このような複数の変数の組み合わせでの解析が行われることも一般的である。変数間の真の関係性は未知であるため、調整する変数の組を変えて、結果がどの程度変わるか（変わらないか）の感度解析を行い、補足的な情報とする。

まとめ

- 多変量モデルによる解析には、予測モデルの作製と交絡調整という２つの目的がある
- 変数選択のストラテジーは、目的に応じて策定されるべき
- いずれの目的の解析においても、理想的な答えを与えてくれる数学的な手法は存在しないため（STRATOS Initiativeでのガイドラインも、現在進行形で整備中）、手法ごとの長短を理解した上で、解析に用いる手法は選択するべき
- 現状では、投稿先のジャーナルの最新の研究で用いられている方法などを参考にして慎重に選択するのがよいと思われる

文献

- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *2nd International Symposium on Information Theory*, B. N. Petrov, and F. Csaki (eds), 267-281. Budapest: Akademiai Kiado.
- Burnham, K. P., and Anderson, D. R. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. New York: Springer.
- Derksen, S., and Keselman, H. (1992). Backward, forward and stepwise automated subset selection algorithms: frequency of obtaining authentic and noise variables. *British Journal of Mathematical and Statistical Psychology* **45**, 265-282.
- Greenland, S. (1989). Modeling and variable selection in epidemiologic analysis. *American Journal of Public Health* **79**, 340-349.
- Greenland, S., Pearl, J., and Robins, J. M. (1999). Causal diagrams for epidemiologic research. *Epidemiology* **10**, 37-48.
- Harrell, F. E., Jr. (2015). *Regression Modeling Strategies*. New York: Springer.

- Harrell, F. E., Jr., Lee, K. L., Califf, R. M., Pryor, D. B., and Rosati, R. A. (1984). Regression modelling strategies for improved prognostic prediction. *Statistics in Medicine* **3**, 143-152.
- Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd edition. New York: Springer.
- Heinze, G., Wallisch, C., and Dunkler, D. (2018). Variable selection - A review and recommendations for the practicing statistician. *Biometrical Journal* **60**, 431-449.
- Magnusson Hanson, L. L., Nyberg, A., Mittendorfer-Rutz, E., Bondestam, F., and Madsen, I. E. H. (2020). Work related sexual harassment and risk of suicide and suicide attempts: prospective cohort study. *BMJ* **370**, m2984.
- Mallows, C. L. (1973). Some comments on C_p. *Technometrics* **15**, 661-675.
- Mansournia, M. A., Etminan, M., Danaei, G., Kaufman, J. S., and Collins, G. (2017). Handling time varying confounding in observational research. *BMJ* **359**, j4587.
- Raftery, A. E., Madigan, D., and Hoeting, J. (1997). Bayesian model averaging for linear regression models. *Journal of the American Statistical Association* **92**, 179-191.

- Riley, R. D., Snell, K. I., Ensor, J., *et al*. (2019a). Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes. *Statistics in Medicine* **38**, 1276-1296.
- Sauerbrei, W., Perperoglou, A., Schmid, M., et al. (2020). State of the art in selection of variables and functional forms in multivariable analysis-outstanding issues. *Diagnostic and Prognostic Research*. **4**, 3.
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics* **6**, 461-464.
- Steyerberg, E. W. (2009). *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*. New York: Springer.
- Steyerberg, E. W., Eijkemans, M. J., and Habbema, J. D. (1999). Stepwise selection in small data sets: a simulation study of bias in logistic regression analysis. *Journal of Clinical Epidemiology* **52**, 935-942.
- Steyerberg, E. W., Eijkemans, M. J., and Habbema, J. D. (2001a). Application of shrinkage techniques in logistic regression analysis: a case study. *Statistica Neerlandica* **55**, 76-88.

- Steyerberg, E. W., Eijkemans, M. J., Harrell, F. E., Jr., and Habbema, J. D. (2000). Prognostic modelling with logistic regression analysis: a comparison of selection and estimation methods in small data sets. *Statistics in Medicine* **19**, 1059-1079.
- Talbot, D., and Massamba, V. K. (2019). A descriptive review of variable selection methods in four epidemiologic journals: there is still room for improvement. *European Journal of Epidemiology* **34**, 725-730.
- Vach, W., Sauerbrei, W., and Schumacher, M. (2001). Variable selection and shrinkage: comparison of some approaches. *Statistica Neerlandica* **55**, 53-75.
- van Smeden, M., de Groot, J. A., Moons, K. G., *et al*. (2016). No rationale for 1 variable per 10 events criterion for binary logistic regression analysis. *BMC Medical Research Methodology* **16**, 163.
- van Smeden, M., Moons, K. G., de Groot, J. A., *et al*. (2019). Sample size for binary logistic prediction models: Beyond events per variable criteria. *Statistical Methods in Medical Research* **28**, 2455-2474.

- VanderWeele, T. J., and Shpitser, I. (2011). A new criterion for confounder selection. *Biometrics* **67**, 1406-1413.
- Walter, S., and Tiemeier, H. (2009). Variable selection: current practice in epidemiological studies. *European Journal of Epidemiology* **24**, 733-736.
- Wang, D., Lertsithichai, P., Nanchahal, K., and Yousufuddin, M. (2003). Risk factors of coronary heart disease: a Bayesian model averaging approach. *Journal of Applied Statistics* **30**, 813-826.
- Zemek, R., Barrowman, N., Freedman, S. B., *et al*. (2016). Clinical risk score for persistent postconcussion symptoms among children with acute concussion in the ED. *JAMA* **315**, 1014-1025.