

中間変数が複数存在する繰り返し測定データにおける mediational g-formula を用いた直接・間接効果の推定

エーザイ株式会社クリニカルデータサイエンス部 山室 晋太郎
東京大・医 松山 裕
東京理科大・工 篠崎 智大

1. はじめに

疾患の発症・抑制に対する治療の効果（総合効果）が検証されたのち、治療がその結果をどのようにもたらすかという生物学的なメカニズムに興味がつく。このメカニズムを統計学的に明らかにするためには、治療A・中間変数M・結果変数Yの間にそれぞれ想定される交絡因子を適切に考慮した mediation analysis と呼ばれる手法を用いる必要がある、これによって総合効果は、中間変数を介さない効果（直接効果）と、中間変数を介する効果（間接効果）に分解できる。

mediation analysis では、反事実モデルを用いた natural direct/indirect effect (NDE/NIE) による直接・間接効果の識別がよく行われるが、中間変数が複数考えられる場合、それらが条件付き独立でなければ NDE/NIE は直接・間接効果を識別できない。ここで NDE/NIE と異なる直接・間接効果の定義として interventional direct/indirect effect (IDE/IE) が存在し、Vansteelandt et al.によって IDE/IE を用いた中間変数が複数存在する場合の直接・間接効果が定義されている。しかし、観察研究等では治療の有無や中間変数の値が継時的に変化（時間依存性交絡が存在）することが多いが、この手法はそのような場合には発展していない。

こうした時間依存性交絡存在下で因果効果を推定する手法として g-formula があげられる。VanderWeele et al.は g-formula を mediation analysis に対応させた mediational g-formula を提案し、これによって時間依存性交絡存在下でも IDE/IE を用いて直接・間接効果を適切に推定することができる。しかし、この手法は中間変数が複数存在する場合は考慮されていない。

2. 目的

既存の手法は 1) 継時的に治療・中間変数の値が変化（時間依存性交絡が存在）し、2) 中間変数が複数存在する場合に拡張されていない。そこで本講演では、このような場合における直接・間接効果の推定方法を提案する。簡単のため中間変数は2つの場合を考える。

3. 方法

VanderWeele et al.が提案した mediational g-formula を中間変数が複数存在する場合に拡張し、潜在結果変数 $E(Y_T^{\bar{a}_1, \bar{G}_{I\bar{a}_2|v}, \bar{G}_{II\bar{a}_3|v}} | V)$ (以降は $Q(\bar{a}_1, \bar{a}_2, \bar{a}_3)$ と表す) を推定する。ここで $Q(\bar{a}_1, \bar{a}_2, \bar{a}_3)$ は、時点Tにおいて、治療Aを $\bar{a}_1$ 、中間変数 (M_I, M_{II}) を $(\bar{G}_{I\bar{a}_2|v}, \bar{G}_{II\bar{a}_3|v})$ ((M_I, M_{II}) をランダムサンプリングした値) に固定した場合に観察されるであろう結果変数Yを表す。また変数の上線はその変数の履歴を表す。この潜在結果変数は治療と中間変数の介入ごとに推定され、この比較によって

$$\begin{aligned} IDE &= E\{Q(\bar{a}, \bar{a}^*, \bar{a}^*) - Q(\bar{a}^*, \bar{a}^*, \bar{a}^*)\} \\ IIE \text{ via } M_I &= E\{Q(\bar{a}, \bar{a}, \bar{a}^*) - Q(\bar{a}, \bar{a}^*, \bar{a}^*)\} \\ IIE \text{ via } M_{II} &= E\{Q(\bar{a}, \bar{a}, \bar{a}) - Q(\bar{a}, \bar{a}, \bar{a}^*)\} \end{aligned}$$

と IDE/IE が定義できる（ここで $\bar{a}$ 及び $\bar{a}^*$ は治療Aの相異なる水準を表す）。

4. データ解析

脂質異常症を合併する糖尿病患者にスタチンを処方することで、循環器系疾患（cardiovascular disease; CVD）発症リスクが低下するが、この疾患抑制効果はスタチンの LDL（low-density lipoprotein）コレステロール低下作用によるものか、抗炎症作用によるものかは明らかでない。そこで日常診療で繰り返し測定を行うことで得られた実データに対して提案手法を適用し、スタチンの治療効果を検討した。6年間のスタチン治療による CVD 発症リスクに対する総合効果 (95%CI) は -4.52% (-10.43%, 2.93%)、IDE は -3.72% (-10.05%, 4.05%)、LDL コレステロール低下作用を介す IIE は -0.75% (-1.77%, -0.06%)、抗炎症作用を介す IIE は -0.07% (-0.21%, 0.11%) となった。

参考文献

- Vansteelandt S, Daniel RM. Interventional effects for mediation analysis with multiple mediators. *Epidemiology*. 2017; 28(2): 258-265.
VanderWeele TJ, Tchetgen Tchetgen EJ. Mediation analysis with time varying exposures and mediators. *J R Stat Soc Series B Stat Methodol*. 2017; 79(3): 917-938.