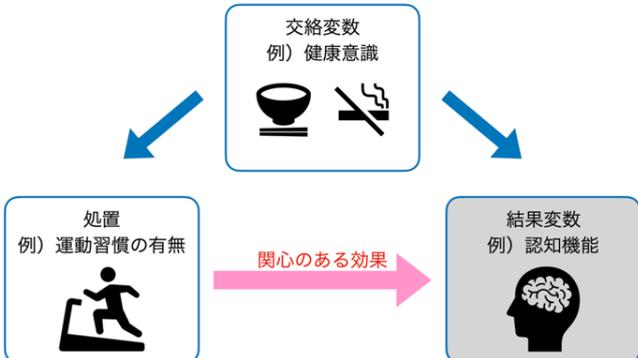


<p>データサイエンス</p>	<p>【代表的な研究テーマ】</p> <p>□ 欠測を伴う調査データにおける因果効果推定と精度向上のための共変量選択</p>
<p>key word</p>	<p>課題解決に役立つシーズの説明</p>
<ul style="list-style-type: none"> ■ 統計科学 ■ 欠測データ解析 ■ 統計的因果推論 ■ 変数選択 	<p>【概要】</p> <p>多くの実証研究や政策評価では、データの欠測が避けられず、これを無視すると推定結果にバイアスが生じる可能性があります。私は、欠測するかどうかが欠測値そのものと関連して生じる状況において、因果効果を適切に推定する方法を研究しています。さらに、推定の精度を高めるために、モデルに含めるべき・除くべき共変量の選択方針についても検討しています。これにより、途中離脱や未回答が避けられない調査データにおいても、信頼性の高い効果検証を可能にします。</p> <p style="text-align: center;">◇ ◇ ◇</p> <p>【研究内容】</p> <p>ランダム化比較試験のように研究者が処置を割り当てられない調査データを用いた効果検証では、欠測や交絡の問題が避けられません。</p> <p>例えば、訪問調査により、運動習慣が高齢者の認知機能に与える影響を推測する場合を考えます。認知機能は調査員がテストを通じて計測し、運動習慣など他の項目については、家族や介護者同席のもとアンケート調査を実施するとします。</p> <p>高齢者を対象とした認知機能のテストでは、認知機能が低下している人ほどテストの負担に耐えきれず脱落してしまう可能性が高いと考えられます。この場合、欠測データを無視した認知機能に関する推論は、認知機能の過大評価につながるため適切ではありません。したがって、適切な推論を行うために欠測の調整が必要です。このように、欠測するかどうかが欠測値そのものに依存するメカニズムは、Missing Not at Random (MNAR) と呼ばれています。MNAR の場合は、他の欠測メカニズムよりも適切な調整が難しいとされています。</p> <p>また、運動習慣のある人はそもそも健康意識が高いと考えられます。そのため、運動習慣のある群・ない群の間での認知機能の単純な比較は、運動習慣の影響だけでなく、健康意識に基づく食事習慣や喫煙などの違いも含めた結果となってしまいます。このように処置と結果の両方に関連する変数は交絡変数と呼ばれ、適切に調整しなければ因果効果を正しく推定できません。</p>
	
<p>米山 慎太郎 Shintaro Yoneyama</p>	
<p>データサイエンス・AI イノベーション 研究推進センター助教</p>	
<p>【プロフィール】</p> <ul style="list-style-type: none"> ・2019年3月 慶應義塾大学大学院 理工学 研究科 基礎理工学専攻 前期博士課程 修了 ・2021年4月-2023年4月 シミック株式会社 データサイ エンス事業本部 統計解析部 (臨床試験における統計解析計 画書作成や、統計解析の実施) ・2023年3月-2025年4月 日本たばこ産業株式会社 D- LAB マネージャー (ユニークなシーズを持つ大学 やスタートアップなどと連携し、 新規事業の企画・開発を実施) ・2025年3月 慶應義塾大学大学院 理工学 研究科 基礎理工学専攻 後 期博士課程 修了 ・2025年5月-11月 株式会社フックフィット (スマートインソールの開発や、 それをを用いた歩行分析アルゴ リズム開発の実施、及び、大 学との共同研究で、認知症や フレイルと歩容の関連を分析) ・2025年12月- 現職 <p>【主な社会的活動】</p> <ul style="list-style-type: none"> ・日本行動計量学会 所属 <p>【researchmap】</p> <p>https://researchmap.jp/shintaro.yoneyama</p>	<div style="text-align: center;">  <p>図：想定している状況</p> </div> <p>近年、結果変数に MNAR の欠測がある場合に、変数間の独立性を利用して母平均を推定する方法が複数提案されています(例:操作変数法:Sun et al. (2018), shadow variable 法:Miao and Tchetgen Tchetgen (2016))。これらの方法は、比較的弱い仮定のもとで推定対象を理論的に定めることができる点や、欠測が生じる結果変数そのものをモデリングしなくてもよいという利点があります。しかし、推定できる対象が母平均に限定されることや、欠測メカニズムのモデリングには依然として定性的考察が求められるという課題がありました。</p> <p>これらの課題に対し、私は以下の2つを提案しました。</p> <ul style="list-style-type: none"> ➢ ランダム化ができない調査観察研究において、結果変数に MNAR の欠測がある場合に、平均処置効果、処置群・対照群での平均処置効果を推定する方法 ➢ 操作変数法を用いる際に、欠測メカニズムのモデリングにおいてバイアスと分散の観点から望ましい共変量選択の指針 <p>これらの提案を通じ、途中離脱や未回答が避けられない調査データにおいても、因果効果や母平均を理論的に妥当かつ精度高く推定することが可能になります。</p> <p>今後は理論的考察を深めるとともに、機械学習手法との統合や実データへの応用を通じて、実務に資する効果検証手法の発展を目指します。</p> <p>企業・自治体へのメッセージ</p> <p>未回答や途中離脱が避けられない調査・施策評価データにおいても、欠測や交絡によるバイアスを抑えた信頼性の高い効果検証を可能にします。臨床試験における統計解析業務や IoT デバイスデータの解析といった実務経験も踏まえ、理論と実務の両面から企業・自治体の皆様と連携できれば幸いです。</p>