

Reducing the variance of the instrumental variable-based estimates of treatment effect in pharmaco-epidemiology

MICHAL ABRAHAMOWICZ

Marie-Eve Beauchamp, Raluca Ionescu-Ittu, Joseph “Chris” Delaney

Department of Epidemiology & Biostatistics,

McGill University,

Montreal, Quebec, CANADA

OUTLINE

- Background:

Physician Preference-based Instrumental Variable (IV) Brookhart's method

- New IV approach

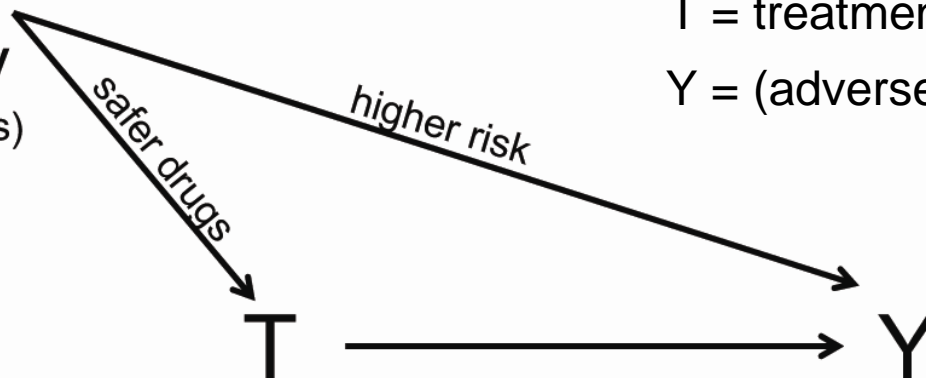
- Simulations

BACKGROUND:

Confounding-by-Indication in Pharmaco-epidemiology

- Observational (database) pharmacoepidemiological studies are essential to detect real-life adverse effects of medications
- Yet, many clinical “risk factors” are often NOT recorded in administrative databases [Wolfe et al., J Rheumatol 2006]
- Such studies suffer from confounding-by-indication bias

Disease
Severity
(sicker patients)

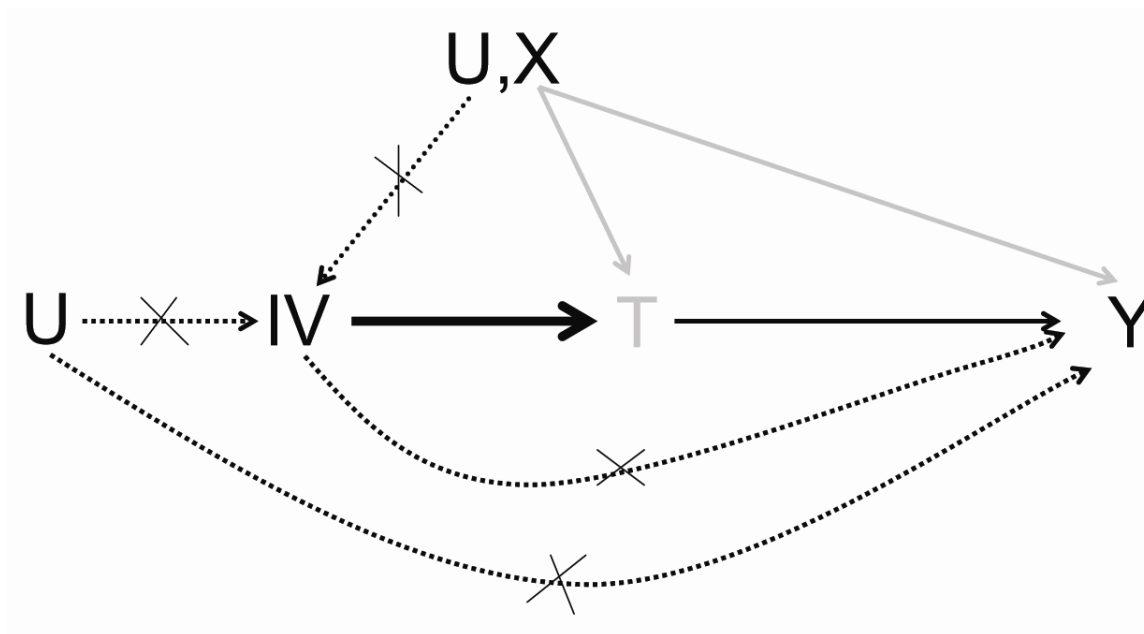


T = treatment (drug prescribed);
Y = (adverse) outcome

BACKGROUND:

Basics of the instrumental variables methodology

- Instrumental Variables (IV) provide a “generic” tool to deal with unobserved confounding
 - IV is associated with T
 - IV is NOT associated (independently of T) with either the unobserved (“U”) or the observed (“X”) covariates



BACKGROUND:

Brookhart's Physician Preference-based IV approach

- IV use in pharmaco-epi was limited by the difficulties in defining IVs that can be measured in database studies
- Brookhart et al's approach [Epidemiology 2006]

IV = subjective Physician Preference (PP) for one of two competing drugs (drug "B" vs drug "A")

(PP measured as ~ the drug prescribed to the previous patient of the same physician)

- PP affects drug choice for most patients of a given physician (PP → T)
- PP is independent of the clinical characteristics of the "index" patient for whom the outcome is observed (PP ⊥ U)

BACKGROUND:

Brookhart's Physician Preference-based IV approach

- **Rationale for using ONLY the “Last Patient” IV**
 - Physician Preference's may CHANGE over time
- **Limitations**
 - Forces dichotomization of the preferences (binary IV)
 - Ignores information on all other prescriptions written by the same physician
 - Ignores observed characteristics which may explain why individual patients receive specific drugs

“Extended” IV Approach

- **We propose to extend Brookhart’s PP-based IV approach by:**
 - Considering ALL Prescriptions written by a given physician
 - Assessing IF/WHEN a physician “i” might have changed his/her preferences (use change-point methods to estimate the possible “**Change-Time**” (k_i^*))
 - Defining “Preference for drug B” as:
 $PP_i = \% \text{ of all prescriptions for drug B written by physician “i” after the change-time } k_i^*$

“Extended” IV approach: BASIC STEPS

1. Order all “relevant” n patients of physician “ i ” in time
 $t_1 < t_2 < \dots < t_n$

2. Fit a Change-Time model for each time
 $\text{logit}\{P[T_j = \text{“B”}]\} = \alpha_0 + \alpha_x X + \theta I\{j > k\}$

where:

T_j = treatment actually received by the j^{th} patient

X = vector of Observed Covariates;

k = potential Change-Time

$I\{j > k\}$ = Indicator of time $>$ change-time

“Extended” IV approach: BASIC STEPS

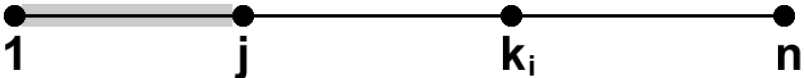
3. **Identify the “optimal” Change-Time model** (with the lowest deviance) for a given physician & consider the corresponding time k_i^* as a potential Change-Time
4. **Fit the “Constant-Preference” (no change) model**
 $\text{logit}\{P[T_j = \text{“B”}]\} = \alpha_0 + \alpha_x X$
5. **Compare AIC of models in steps 3 and 4** (add penalty for additional parameter k_i^* in the Change-Time model)
 - *IF AIC {“Optimal” Change-Time model} < AIC {Constant-Preference model}*
THEN accept k_i^ as the Time-Change for Physician “i”*
 - *ELSE: accept no change in physician “i” preferences*


“Extended” IV approach: BASIC STEPS

6. Define the IV for each patient ($j=1,\dots,n_i$) of each physician “i” ($i=1,\dots,m$)

IV = % of all prescriptions for drug B written by physician “i” after the change-time k_i^* (if any):

(i) IF step 5 identified k_i^* as Change-Time, THEN

(i¹) for $j < k_i^*$ include: 

(i²) for $j > k_i^*$ include: 

(ii) IF step 5 identified NO Change-Point for physician “i”, THEN include:



2SLS Estimation of the IV models

- Use the 2-Stage Least-Squares (2SLS) IV approach to estimate the causal effect of T on Y:

β_1 = adjusted “Risk Difference” associated with T=1 (drug “B”) relative to T=0 (“A”) [Brookhart et al., Epidemiology 2006]

Stage 1: regress P(T=1) on IV

$$P(T_{ji} = 1) = \alpha_0 + \alpha_1 * IV_{ji} + \sum \alpha_p * X_{pji} + \epsilon_{ji}$$

Stage 2: regress Y on P(T=1)

$$P(Y_{ji} = 1) = \beta_0 + \beta_1 * P(T_{ji} = 1) + \sum \beta_p * X_{pji} + \epsilon_{ji}$$

SIMULATIONS: Objectives

- **Compare and assess, in the presence of unobserved confounding, the treatment (B vs A) effect estimates from 3 models:**
 1. Conventional model
(NO adjustment for the unobserved confounder)
 2. Brookhart's model with IV based on Last Prescription
 3. Proposed "Extended" version of the IV model in (2)

SIMULATIONS: Design and Assumptions

- We simulated a hypothetical study comparing adverse events risk of 2 drugs (B vs A), under the following assumptions:
 - 200 prescribing physicians ($i=1, \dots, 200$)
 - # patients/physician: $n_i \sim U[10;50]$
 - Initially, 60% of physicians “prefer” drug A & 40% “prefer” drug B
 - During the study period, some physicians CHANGE their preferences:
 - 33% of the initial 60% who prefer drug A “switch” to drug B
 - 10% of the initial 40% who prefer drug B “switch” to drug A
 - Time of “switch” for physician “i”: $\tau_i \sim U [1; n_i]$
(patient’s index $j=1, \dots, n_i$ is used as time metric)

SIMULATIONS: Data Generation

■ 4 patient-level Covariates:

- $X_1 \sim U [0, 1]$, observed confounder
- $X_2 \sim \text{Bin} [1, p=0.65]$, observed confounder
- $U \sim U [0, 1]$, unobserved confounder
- C = unobserved predictor of T (drug A “counterindication”)
 - **IF** $C=1$ **THEN** patient **MUST** be prescribed drug B
 - **IF** $C=0$ **THEN** patient can receive either drug
[Ionescu-Ittu et al., PDS 2009]
 - $p(C=1)$ was varied from 0 to 0.25

Data Generation: 'TREATMENT'

- **Binary treatment** ($T=1$ for drug B) was generated from the binomial distribution
 - **For patients with counterindication to drug A ($C_{ij} = 1$)**
 $P(T_{ij} = 1) = 1$, regardless of physician preferences (PP) and covariates
 - **For patients without counterindication to drug A ($C_{ij} = 0$)**
Similar to Brookhart, T_{ij} was generated from a linear RD model [Epidemiol. 2006]
$$P(T_{ij}=1) = 0.05 + 0.70*PP_j + 0.03*X_{1i} + 0.03*X_{2i} + 0.15*U_i \quad (1)$$

where PP_i :

- = indicator of physician's "current" preference ($PP_i = 1$ if physician prefers drug B)
- e.g.: physician "i" switched from preferring A to B after seeing patient with index k_i^* :
 - $PP_i = 0$ for his 'earlier' patients: $0 < j \leq k_i^*$
 - $PP_i = 1$ for his 'later' patients $k_i^* < j \leq n$

Data Generation: 'OUTCOME'

- **Binary outcomes** Y_{ij} were generated, conditional on treatment and other patients covariates from a linear RD model [Brookhart et al., Epidemiology 2006]

$$P(Y_{ij}=1) = 0.001 + 0.02*T_i + 0.10*X_{1i} + 0.04*X_{2i} + 0.20*U_i \quad (2)$$

- Note:
 - The unobserved variable “U” is an unobserved confounder (affects both the treatment in (1) and the outcome in (2))
 - In contrast, the risk of outcome in (2) does NOT depend on either “counterindication” (C) or physician preferences (PP)

ANALYSIS of Simulated Data

- Each of the 500 simulated samples were analyzed with 4 Linear RD models:
 - **Model 1A: Artificial (“True”) model**, adjusted for both observed AND unobserved confounders (*OLS estimation*): $P[Y=1] = f(T, X_1, X_2, \underline{U})$
 - **Model 1: Conventional model**, adjusted ONLY for the observed covariates (*OLS estimation*): $P[Y=1] = f(T, X_1, X_2)$
 - **Model 2: Brookhart et al’s IV model** with IV based on the Last Prescription (*2SLS estimation*): $P[Y=1] = f(P(T|PP), X_1, X_2)$
 - **Model 3: New “Extended” IV model** based on estimated Time-Change for Changing Preferences (*2SLS estimation*): $P[Y=1] = f(P(T|PP^*), X_1, X_2)$

SIMULATION RESULTS: Time-Change Estimation for “Extended” IV model

- For the proposed Extended IV model, we first assessed IF and WHEN physician’s ‘i’ preferences change over time

RESULTS: Identification of change based on AIC (steps 1-5)

- True Negative Rate (among physicians with NO change): **0.69**
- True Positive Rate for detecting True Change
 - **0.84** for switch **A** → **B**
 - **0.86** for switch **B** → **A**

RESULTS: Accuracy of Time-Change estimates (for true positives)

- No bias
 - Mean difference (Estimated-True) $\leq 0.3^*$,
 - Median = 0
- Small variance: $SD(\text{Difference}) \leq 5.2^*$

**Maximum Range of Differences [-50,+50]*

MAIN SIMULATION RESULTS (500 samples): Comparison of Treatment Effect Estimates

	Model 1A “TRUE”	Model 1 Conventional	Model 2 Brookhart’s IV	Model 3 “Extended” IV
Relative BIAS (%)	+ 0.7%	+ 51.3%	- 4.8%	+ 1.2%
Variance Ratio (rel. to model 1A)	1.0	1.016	4.431	2.716
MSE Ratio (rel. to model 1A)	1.0	2.074	4.439	2.571
MSE Ratio (rel. to model 2)	0.225	0.467	1.0	0.560

CONCLUSIONS

- **The Proposed “Extended IV model”, that attempts to extend the Brookhart et al’s model through estimating/accounting for changes-over-time in physician’s preferences:**
 - (1) Provides reasonable TP and TN rates for detecting changes and accurate estimates of the Time-Change
 - (2) Yields as unbiased estimates of treatment effect (with unobserved confounding) as the Brookhart’s model
 - (3) Reduces by > 40% both the (inflated) variance and the MSE, relative to the Brookhart et al’s IV model
 - (4) (In our simulations) yields similar (but slightly worse) bias/variance trade-off as the conventional model (MSE ratio < 1.25) BUT avoids the huge bias induced by the conventional model

FUTURE CHALLENGES

- Assess implications for hypothesis testing (type I error vs power)
- Evaluate alternative (to AIC) criteria for detecting “true change”, in order to improve specificity and, thus, avoid relying on “spurious changes” that reduce the amount of observations and, thus, inflate the variance of the estimates from the “Extended IV model”
- Explore how the bias/variance trade-off between the “Extended IV model” vs the conventional model changes with increasing (a) sample size; (b) frequency of changes; (c) strength of physician preferences; (d) strength of unobserved confounding
- Assess robustness of our (preliminary) conclusions with respect to a variety of simulated scenarios and underlying assumptions

REFERENCES

- Wolfe F, Michaud K, Simon T. Can severity be predicted by treatment variables in rheumatoid arthritis administrative data bases? *J Rheumatol* 2006;33(10):1952-1956
- Brookhart MA, Wang PS, Solomon DH, Schneeweiss S. Evaluating short-term drug effects using a physician-specific prescribing preference as an instrumental variable. *Epidemiology* 2006;17(3):268-275
- Ionescu-Ittu R, Delaney JAC, Abrahamowicz M. Bias-Variance Trade-Off in Pharmacoepidemiological Studies Using Instrumental Variables. *Pharmacoepidemiology and Drug Safety* 2009; 18(7):562-71.