A Bayesian latent class approach to causal inference with longitudinal data

Kuan Liu[1,2], PhD(c)

[1]University of Toronto
[2]The Hospital for Sick Children
[3]University Health Network

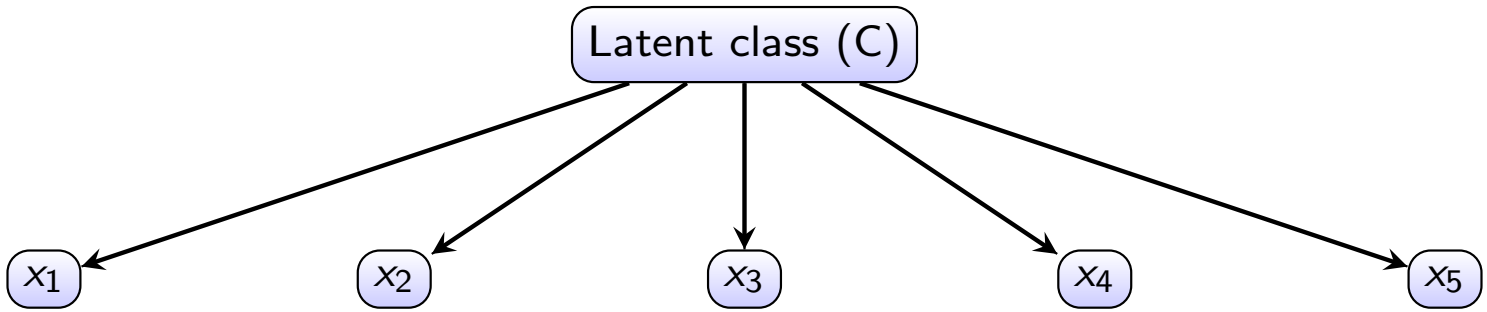
Feb 13th, 2020

Joint work with Dr. Eleanor Pullenayegum[1,2], Dr. Olli Saarela[1] and Dr. George Tomlinson[1,3]

Latent variable

UNIVERSITY OF TORONTO

- **Latent variable** is defined as a random variable whose realizations are hidden from us.
- This is in contrast to **manifest variables** where the realizations are observed.
- can be continuous or categorical (latent class)
- latent class analysis (LCA) is a latent variable modeling technique that identifies latent (unobserved) subgroups of individuals within a population based on nominal or ordinal indicators.



Causal inference with latent variable

UNIVERSITY OF TORONTO

Existing causal inference literature on latent variables (as confounder).

- **"Proxy" confounder**[1]
  - The true confounder values are unknown but we have observed measures which are subject to measurement errors. Causal inference on exposure effect is subject to bias.
  - Several papers have studied the conditions for which the causal effect can be restore using proxy variables.
- **Unmeasured confounder**[2]
  - One key assumption in deriving causal estimators is the "no unmeasured confounder" assumption, an untestable assumption.
  - standard approach to quantify bias due to unmeasured confounder is through sensitivity assessments.

[1] Cai & Kuroki, 2008; Greenland & Lash, 2009; Pearl, 2009

[2] Lin et al, 1998; Robins & Rotnitzky & Scharfstein, 1999; McCandless LC & Gustafson & Levy, 2007; VanderWeele & Arah, 2011; Greoenwold et al, 2016

The CARRA JDM CTP study

UNIVERSITY OF TORONTO

- Newly diagnosed moderate severe JDM patients at CARRA clinics will be prescribed to one of the three therapies: MP, MMP and MMPI by the treating clinician.
- No clinical trial evidence available due to disease rarity. The three therapies are a collective consensus treatment plans developed by CARRA members at an annual meeting.
- At onset and each follow-up clinical visits, patient's demographic information, disease relative clinical measurements as well as other key health indicators (i.e. weight, height etc) and past medical histories are collected.

Motivation causal framework

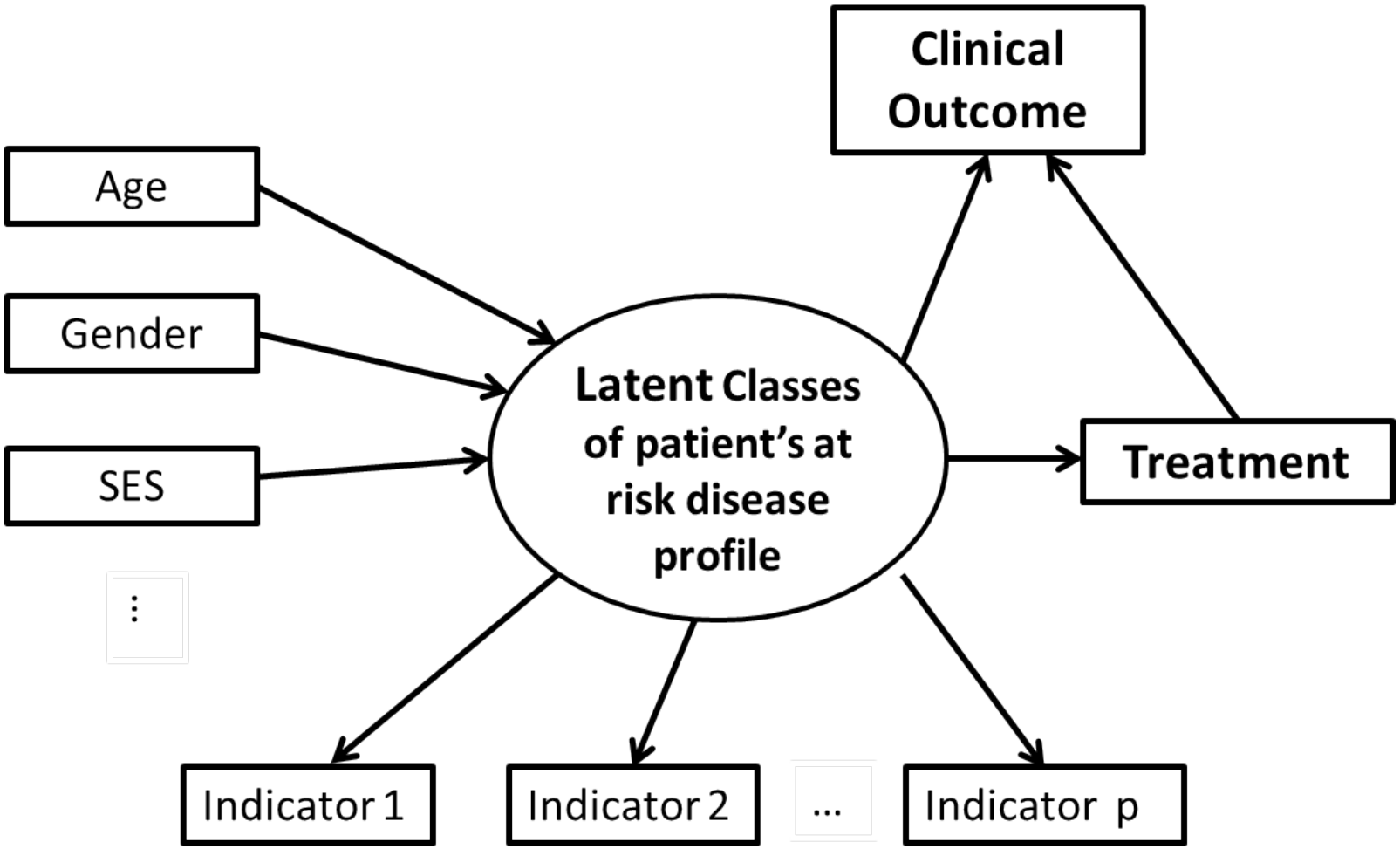


Figure: Hypothesized causal diagram (DAG)

Additional discussion on the framework

UNIVERSITY OF TORONTO

- **Features**
  - **Dimension reduction** of the measured covariates.
  - Two categories of measured covariates: covaraites that predict class membership (**class predictors**) and covaraites that manifested from the latent class (**class indicators**).
- **Flexibility and extensions**
  - **Indicator quality**, in case of a binary class indicator X, X is high quality if $P(X = 1|U = 1) \sim 1$ or 0. We can assess estimation sensitivity with different quality levels.
  - **High dimension indicators**, applying variable selection or model averaging methods.

Traditional causal inference follows a two-stage process: the design stage (treatment assignment) and the analysis stage (outcome model). In principle the two stages should not be jointly estimated. **The proposed framework would permit a full Bayesian estimation.**

Notations

UNIVERSITY OF TORONTO

- Under a simple three-visit setting with an end-of-study outcome Y_i for patient i.
- Z_{ij} represents the treatment patient i received at visit j,
- U_{ij} represent the unobserved latent class for patient i at visit j,
- X_{ij}, $p \times 1$ represents class indicators.
- We consider C_i as time independent baseline variables like gender and age.
- Let $Y_i^{(a_1,a_2)}$ be the potential outcome under treatment combination $z_1 = a_1$ and $z_2 = a_2$.

Simple longitudinal DAG

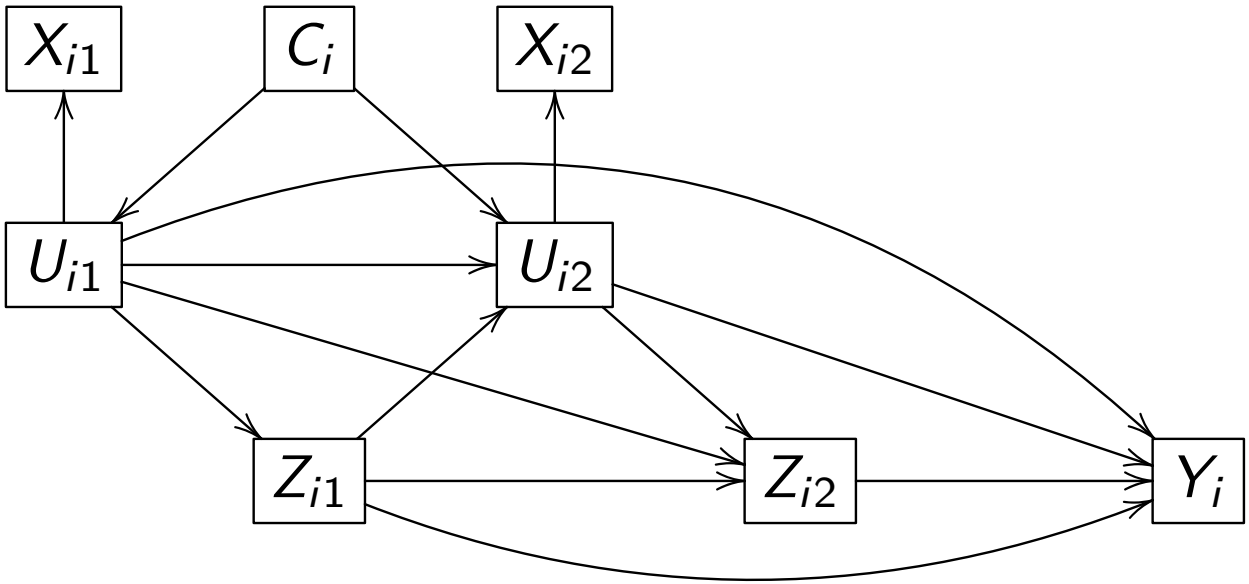


Figure: Longitudinal causal diagram between latent classes, treatment assignments, covariates and outcomes

Potential outcome framework

UNIVERSITY OF TORONTO

- In case of a binary treatment assignment, the collection of all potential variables are,

$$\mathcal{W}_i = \{U_{i2}^0, X_{i2}^0, U_{i2}^1, X_{i2}^1, , Y_i^{00}, Y_i^{01}, Y_i^{10}, Y_i^{11}\}.$$

- Assumptions
  1. Consistency, $Y_i^{\tilde{z}_{i2}=\tilde{a}} \mid (z_{i1} = a_1, z_{i2} = a_2) = Y_i \mid (z_{i1} = a_1, z_{i2} = a_2)$
  2. Sequential randomization (latent unconfoundedness)

$$\begin{aligned} &\text{(i)} \quad Z_{i1} \perp \mathcal{W}_i \mid U_{i1} \quad \text{and} \quad Z_{i2} \perp \mathcal{W}_i \mid (U_{i1}, U_{i2}, Z_{i1}) \\ &\text{(ii)} \quad Z_{i1} \perp C_i \mid U_{i1} \quad \text{and} \quad Z_{i2} \perp C_i \mid (U_{i1}, U_{i2}, Z_{i1}) \end{aligned}$$

  3. Positivity (Identifiability)
  4. Independency between class indicators given class membership, $P(X_{i11}, \ldots, X_{i1p}|U_{i1}) = \prod_{h=1}^{p} P(X_{i1h} \mid U_{i1})$ and $P(X_{i21}, \ldots, X_{i2p}|U_{i2}) = \prod_{h=1}^{p} P(X_{i2h} \mid U_{i1})$.

Causal parameter of interest

Joint likelihood involving counterfactuals,

$$\prod_{i=1}^{n}\prod_{a_1=0}^{1}\prod_{a_2=0}^{1}\Big[\sum_{u_{i2}}\sum_{u_{i1}} P(y^{(a_1,a_2)} \mid u_{i2}^{a_1}, u_{i1})P(z_{i2} \mid z_{i1}, u_{i2}^{a_1}, u_{i1})P(x_{i2}^{a_1} \mid u_{i2}^{a_1})$$

$$\times P(u_{i2}^{a_1} \mid u_{i1}, z_{i1}, c_i)P(z_{i1} \mid u_{i1})P(x_{i1} \mid u_{i1})P(u_{i1} \mid c_i)\Big]^{I_{Z_{i1}=a_1}, I_{Z_{i2}=a_2}} \quad (1)$$

Average potential outcome (APO),

$$E[Y_i^{(a_1,a_2)} \mid C_i] = \sum_{u_{i2}}\sum_{u_{i1}} E(y_i \mid u_{i1}, u_{i2}, z_{i1} = a_1, z_{i2} = a_2)$$

$$\times P(u_{i2} \mid z_{i1} = a_1, u_{i1}, c_i)P(u_{i1} \mid c_i)\ dx_{i1}dx_{i2} \quad (2)$$

Bayesian estimation

UNIVERSITY OF TORONTO

Estimation proceeds with the following steps

1. Specify a model for each component of the joint likelihood in (1);
2. Select prior distributions for each model parameters;
3. Obtain posterior samples of the model parameters through MCMC;
4. Obtain posterior predictive values of APO.

Straightforward Bayesian estimation, can be achieved using standard MCMC software (NO MATH).

Simulation study

UNIVERSITY OF TORONTO

We hope to conduct a thorough simulation study to assess the Bayes estimator under the following settings

- varying numbers of class indicators
- varying qualities of the class indicators
- varying strength of confounding effects of the latent class on the outcome
- sample size
- number of latent classes

Simulation setup

UNIVERSITY OF TORONTO

We simulated 100 iterations of a longitudinal data with $n = 300,500$.

1. Two time independent confounders $ca_i \sim N(10,3)$, $cs_i \sim Bin(n, 0.6)$
2. Latent class at visit 1 $U_{i1} \sim Bin(n, expit(-0.05ca_i + 0.2cs_i))$
3. Class indicators for U_1 $n_x = 5, 10$
   - $X_{i1}^1 \sim Bin(n, expit(-2 + 4u_{i1}))$, $P(x_{i1}^1 = 1 | u_{i1} = 1) \approx 88\%$, high
   - $X_{i1}^2 \sim Bin(n, expit(-1 + 2u_{i1}))$, $P(x_{i1}^2 = 1 | u_{i1} = 1) \approx 73\%$, medium
   - $X_{i1}^3 \sim Bin(n, expit(-0.5 + u_{i1}))$, $P(x_{i1}^3 = 1 | u_{i1} = 1) \approx 62\%$, low
   - $X_{i1}^4 \sim Bin(n, 0.5)$, predicting by chance
4. Treatment at visit 1 $Z_{i1} \sim Bin(n, expit(-1 + u_{i1}))$
5. Latent class at visit 2
   $U_{i2} \sim Bin(n, expit(-0.05ca_i + 0.2cs_i + u_{i1} - z_{i1}))$
6. Class indicators for U_2 repeat step 3
7. Treatment at visit 2 $Z_{i2} \sim Bin(n, expit(-1 + u_{i2} - z_{i1}))$
8. Outcome
   Medium confounding $Y \sim N(0.5z_1 + z_2 + 0.1z_1z_2 - 0.5u_1 - u_2, 1)$;
   High confounding $\mu_y = 0.5z_1 + z_2 + 0.1z_1z_2 - u_1 - 2u_2$.

Comparative models

UNIVERSITY OF TORONTO

We compared three estimation models here,

1. Naive regression model, $y_i = \alpha_1 + \alpha_2 z_{i1} + \alpha_3 z_{i2} + \alpha_4 z_{i1} z_{i2}$. we use bootstrap to return variance estimators for $\mu_y^{00}, \mu_y^{10}, \mu_y^{01}, \mu_y^{11}$.

2. Covariates adjusted regression model (**"Wrong model"**), $y_i = \beta_1 + \beta_2 z_{i1} + \beta_3 z_{i2} + \beta_4 z_{i1} z_{i2} + \beta_5 ca_i + \beta_6 cs_i + \beta_7 X_{i1}^l + \beta_8 X_{i1}^m + \beta_9 X_{i1}^h + \beta_{10} X_{i2}^l + \beta_{11} X_{i2}^m + \beta_{12} X_{i2}^h$. Again, we use bootstrap to return variance estimators for $\mu_y^{00}, \mu_y^{10}, \mu_y^{01}, \mu_y^{11}$.

3. Full Bayesian specification following the joint likelihood (1) with non-informative uniform priors [-5,5] on all coefficient parameters and the normal distribution for y has a precision piror $\tau_0 \sim Gamma(.001, .001)$. 4000 MCMC posterior samples were drew using JAGS.

Simulation Results (n=300, medium confounding)

UNIVERSITY OF TORONTO

Setting	Estimator	Mean	RB	ESE	ASE	CP	U1	U2
Med conf	Naive	1.33	-26.89	0.30	0.28	62		
5 indicators	Adjust	1.56	-14.15	0.27	0.26	87		
high quality	Bayes	1.82	-0.09	0.28	0.27	95	0.98	0.98
Med conf	Naive	1.33	-26.89	0.30	0.28	62		
5 indicators	Adjust	1.49	-18.34	0.29	0.28	80		
medium quality	Bayes	1.83	0.57	0.30	0.28	93	0.83	0.85
Med conf	Naive	1.30	-28.42	0.25	0.30	59		
10 indicators	Adjust	1.56	-14.06	0.24	0.28	82		
high quality	Bayes	1.80	-1.30	0.24	0.27	99	1	1
Med conf	Naive	1.30	-28.42	0.25	0.30	59		
10 indicators	Adjust	1.51	-16.89	0.26	0.29	83		
medium quality	Bayes	1.80	-1.15	0.25	0.27	99	0.92	0.93

- Parameter of interest $\mu_{11} - \mu_{00}$
- Relative Bias (RB); Empirical Standard Error (ESE); Average Standard Error (ASE); Coverage Probability (CP)
- U1 and U2 record the average proportion of times U1 and U2 is correctly predicted under Bayesian estimation.

Simulation Results (n=300, high confounding)

UNIVERSITY OF TORONTO

Setting	Estimator	Mean	RB	ESE	ASE	CP	U1	U2
High conf	Naive	1.07	-47.72	0.39	0.39	31		
5 indicators	Adjust	1.53	-24.80	0.28	0.31	64		
high quality	Bayes	2.05	0.36	0.29	0.29	94	0.98	0.99
High conf	Naive	1.07	-47.72	0.39	0.39	31		
5 indicators	Adjust	1.38	-32.09	0.32	0.36	54		
medium quality	Bayes	2.06	1.00	0.34	0.33	92	0.85	0.9
High conf	Naive	1.03	-49.66	0.34	0.40	28		
10 indicators	Adjust	1.54	-24.22	0.25	0.32	72		
high quality	Bayes	2.01	-1.18	0.25	0.28	99	1	1
High conf	Naive	1.03	-49.66	0.34	0.40	28		
10 indicators	Adjust	1.45	-28.77	0.29	0.37	68		
medium quality	Bayes	2.02	-0.85	0.26	0.30	99	0.93	0.95

- Parameter of interest $\mu_{11} - \mu_{00}$
- Relative Bias (RB); Empirical Standard Error (ESE); Average Standard Error (ASE); Coverage Probability (CP)
- U1 and U2 record the average proportion of times U1 and U2 is correctly predicted under Bayesian estimation.

Simulation Results (n=500, medium confounding)

UNIVERSITY OF TORONTO

Setting	Estimator	Mean	RB	ESE	ASE	CP	U1	U2
Med conf	Naive	1.32	-27.51	0.22	0.23	38		
5 indicators	Adjust	1.57	-13.87	0.22	0.21	78		
high quality	Bayes	1.82	0.02	0.22	0.21	96	0.98	0.98
Med conf	Naive	1.32	-27.51	0.22	0.23	38		
5 indicators	Adjust	1.50	-17.80	0.22	0.22	64		
medium quality	Bayes	1.82	0.24	0.22	0.22	97	0.83	0.85
Med conf	Naive	1.30	-28.71	0.26	0.23	39		
10 indicators	Adjust	1.56	-14.03	0.23	0.21	73		
high quality	Bayes	1.80	-0.94	0.23	0.21	92	1	1
Med conf	Naive	1.30	-28.71	0.26	0.23	39		
10 indicators	Adjust	1.51	-16.78	0.24	0.22	69		
medium quality	Bayes	1.80	-0.82	0.24	0.21	91	0.92	0.93

- Parameter of interest $\mu_{11} - \mu_{00}$
- Relative Bias (RB); Empirical Standard Error (ESE); Average Standard Error (ASE); Coverage Probability (CP)
- U1 and U2 record the average proportion of times U1 and U2 is correctly predicted under Bayesian estimation.

Simulation Results (n=500, high confounding)

UNIVERSITY OF TORONTO

Setting	Estimator	Mean	RB	ESE	ASE	CP	U1	U2
High conf	Naive	1.04	-48.98	0.28	0.31	7		
5 indicators	Adjust	1.54	-24.47	0.23	0.24	45		
high quality	Bayes	2.04	-0.00	0.23	0.22	97	0.98	0.99
High conf	Naive	1.04	-48.98	0.28	0.31	7		
5 indicators	Adjust	1.39	-31.71	0.24	0.28	30		
medium quality	Bayes	2.04	0.16	0.25	0.25	98	0.85	0.91
High conf	Naive	1.01	-50.39	0.34	0.31	7		
10 indicators	Adjust	1.55	-24.16	0.24	0.25	51		
high quality	Bayes	2.02	-0.95	0.24	0.22	93	1	1
High conf	Naive	1.01	-50.39	0.34	0.31	7		
10 indicators	Adjust	1.45	-28.90	0.26	0.29	49		
medium quality	Bayes	2.02	-0.97	0.26	0.23	93	0.93	0.95

- Parameter of interest $\mu_{11} - \mu_{00}$
- Relative Bias (RB); Empirical Standard Error (ESE); Average Standard Error (ASE); Coverage Probability (CP)
- U1 and U2 record the average proportion of times U1 and U2 is correctly predicted under Bayesian estimation.

Discussion and future works

UNIVERSITY OF TORONTO

- In this simple simulation study, the Bayesian estimation outperforms the two regression estimators.
- Not surprising given the setups. If we lower the confounding effect, the regression methods might not be that "bad".
- Here we fixed the number of latent class memberships and analysis in the simulation were conducting using the corrected simulated class size. Also, the models are all correctly specified for the Bayes estimator - expect it to perform well.
- Simulation on three latent classes
- Simulation with Bayesian variable selection

Any other suggestions?

Special Acknowledgments

- This thesis research is supported by the Canadian Institutes of Health Research, Doctoral Research Award GSD-152386.

Reference

1. Rosenbaum, Paul R and Rubin, Donald B. Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome, Journal of the Royal Statistical Society: Series B (Methodological),45(2), 212-218, 1983.
2. Z.Cai and M. Kuroki. On identifying total effects in the presence of latent variables and selection bias. In Proceedings of the Twenty-Fourth Conference on Uncertainty in Artificial Intelligence, AUAI Press, 2008.
3. S. Greenland and T.Lash. Bias analysis. In Modern epidemiology, 2008.
4. J.Pearl. Causality. Cambridge university press, 2009
5. Lin DY, Psaty BM and Kronmal RA. Assessing the sensitivity of regression results to unmeasured confounders in observational studies. Biometrics, 948-63, 1998.
6. Scharfstein DO, Rotnitzky A, Robins JM. Adjusting for nonignorable drop-out using semiparametric nonresponse models. Journal of the American Statistical Association, 94(448), 1096-120, 1999.
7. McCandless LC, Gustafson, P and Levy AR. Bayesian sensitivity analysis for unmeasured confounding in observational studies, Statistics in medicine, 26, 2331-2347, 2007
8. VanderWeele and Tyler J. Sensitivity analysis: distributional assumptions and confounding assumptions, Biometrics, 64(2), 645-649, 2008.
9. Groenwold RH, Sterne JA, Lawlor DA, Moons KG, Hoes AW, Tilling K. Sensitivity analysis for the effects of multiple unmeasured confounders. Annals of epidemiology, 26(9), 605-11, 2016
10. Louizos C., Shalit U., Mooij J, Sontag D., Zemel R. and Willing M., Causal effect inference with deep latent-variable models, In Proceeding of the thirty-first Conference on Neural Information Processing Systems, 2018
11. Wang M., Zhi G. and Tchetgen T. E., Identifying causal effects with proxy variables of an unmeasured confounder, Biometrika, 2018. In press.