

ON BAYESIAN G-COMPUTATION AND INSTRUMENTAL VARIABLES FOR OBSERVATIONAL DATA

P. GUSTAFSON AND H. CAMPBELL

1. INTRODUCTION

The power of instrumental variable is discussed in Burgess (2014) [4].

TABLE 1. default

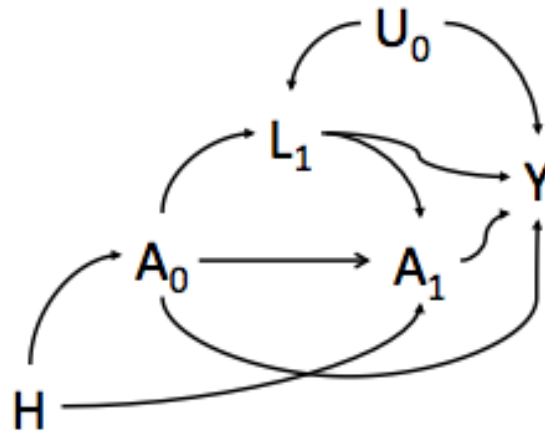
	$H = 1$	$H = 1$
	$A_0 = 0$	$A_0 = 1$
	$P(A_1 = 1)$	$P(A_1 = 1)$
$L_1 = 0$	0.7	1
$L_1 = 1$	0.2	1

TABLE 2. default

	$P(A_0 = 1)$
$H = 0$	0.05
$H = 1$	0.50

During the study observation period, 23.8% of patients (Table 1) had been exposed to a DMD, cytotoxic immunosuppressant or other clinical trial drugs. The percentage of treated patients was 1.0%, 8.3%, 24.2%, and 43.4% for the earliest to the most recent time interval, respectively. However, the actual proportion of the observation time exposed to drug was only 7.0% overall, and 0.2%, 1.4%, 5.0%, and 16.0% for the earliest to most recent time-period group, respectively.

Lee [20] [21] considers Instrumental Variable (IV) analysis as an alternative to g -computation and discusses how effects can be “direct” or “indirect”.



<http://andrewgelman.com/2015/03/20/bayesian-models-causal-inference-time-varying-exposures/>

Given the following change in uptake from period 1 to period 2, what is the required sample size, n , to obtain 80% (90%) power by instrumental variable analysis.

Similar research for power of IV analysis :

Brion (2013) [3]: Calculating statistical power in Mendelian randomization studies.

Burgess(2014) [4]: “Sample size and power calculations in Mendelian randomization with a single instrumental variable and a binary outcome”;

Pierce (2010) [33]:

“Power and instrument strength requirements for Mendelian randomization studies using multiple genetic variants”

and [34].

To verify the IV assumptions : Glymour (2012) [12].

For a given effect size, δ , in order to obtain 80% power, one needs a sample size of approximately XX for successful g-computation.

Examples of calendar period as instrumental variables: Chen (2011) [?].

Treatment effect analysis is possibly the most important area in science, because an important reason to do science is to change a cause to improve the outcome. Much of the literature is concerned with “one-shot” static treatment. But many treatments are repeated over time in reality and, when this is done, they are often modified depending on the interim outcomes. This is natural, as people always try to do better given the information at hand that accumulates over time, and the interim outcomes are part of the information.

Clinical trials are often not practical for long-term outcomes. As such, researchers often find themselves in the difficult position of relying on observational data for determining the effect, or lack thereof, of a given exposure, Shinozaki (2014) [38]. The primary obstacle for such efficacy studies is that of uncontrollable bias due to confounding by indication (exposure is correlated with an unmeasured predisposition to a given outcome), McMahon (2000) [25].

Even in situations when all possible confounders are recorded in the data, adjusting for these may not alleviate one of potential bias, if treatment or exposure is time-varying, [?]. Robbins (1992) identifies two situations in which biased results will occur:

- there exists a time-dependent factor that impacts the outcome while also impacting subsequent treatment;
- past treatment history predicts subsequent risk factor level.

The main obstacles with the analysis of time-varying observational data is the presence of: (1) unknown confounders and (2) known time-dependent confounders affected by prior components of the exposure, “causal intermediates”, Daniel (2013) [8]. While standard regression models can adequately adjust for confounding of fixed-time variables, adjusting for causal intermediates may not be appropriate, and result in misleading inference, Weinberg(1993) [46], Rosenbaum (1984) [36].

One strategy is to apply ***g-computation***, Borsi (2012) [1]. Another, seemingly cruder strategy, would involve treating calendar period of reaching eligibility for treatment as an **instrumental variable**, Greenland (2000) [13].

Note: non-parametric g-computation is identical to IPW:

When all models are nonparametric, the two approaches are identical, as was noted by Hernn et al. [8] when first introducing IPW of MSMs.

Robins in his original paper [1] highlighted a problem with the parametric version of the g-computation formula described above, the g-null paradox.

The theoretical properties of the *g-computation* approach have been previously examined, see Johnston et al. (2008) [15], and applications exist across a wide range of disciplines: assessing the effect of antiretroviral therapy on incident AIDS, Cain et al. (2009) [5]; estimating the risk of coronary heart disease, Taubman (2009), [41]; determining the effect of pillbox organizer use on adherence and viral suppression, Petersen (2007) [32]; evaluating time trends in associations between ozone concentrations and hospitalization for asthma, Moore (2008) [26]. The *Bayesian g-computation*, while not as prevalent in the literature, shows promise, Gustafson (2015) [14], Wang (2011) [45], Keil (2015) [17]. Whether one uses the frequentist or bayesian framework, the validity of the *g-computation* approach relies on the assumption that all confounders are measured in the data.

The alternative **instrumental variable** strategy circumvents the requirement that all confounders are known and measured at the cost of inducing exposure misclassification, Cain et al. (2009) [5]. Three important assumptions are required: 1. calendar periods are associated with therapy because therapies were introduced sequentially over time; 2. calendar period cannot be affected by indications for treatment with therapy; and 3. calendar period is independent of outcome given indications for and actual use of therapy. The strength of the instrument (calendar period) is determined by the correlation between exposure and calendar period, with strong instruments being preferable, Johnston and Gustafson (2008) [15].

Application- Since they were approved 20 years ago, the proportion of MS patients taking beta-interferons has gone up dramatically, but outcomes, as measured by time from disease onset to disease progression, have not improved substantially over calendar time. Tremlett et al. (2008) [42] and Tremlett et al. (2010) [43] review recent advances in the understanding of the natural history of MS. In order to determine whether or not beta-interferons are an effective treatment, one can look over extensive subject-level observational data, as in the analysis of Karim et al. (2014) [16]. If calendar time can be considered an “instrumental variable”, other methods of analysis may prove advantageous.

This paper investigates the merits of each approach in terms of power to detect efficacy (or lack thereof) and robustness to violations of assumptions.

2. CALENDAR PERIOD AS INSTRUMENTAL VARIABLE

Cain et al. (2009) [5]:

Simple calendar period approaches may circumvent confounding by indication at the cost of inducing exposure misclassification. To correct this misclassification, the authors propose an instrumental-variable estimator analogous to ones previously used for noncompliance corrections in randomized clinical trials.

These studies used calendar period as a proxy for actual HAART use to circumvent confounding by indication (3). To the extent that calendar period is a misclassified version of actual HAART exposure, estimates using calendar period may be subject to information bias (M. A. Hernan, Harvard School of Public Health, unpublished manuscript, 2009).

3 assumptions of an instrumental variable:

1. calendar periods are associated with therapy because therapies were introduced sequentially over time.
2. calendar period cannot be affected by indications for treatment with therapy.
3. calendar period is independent of outcome given indications for and actual use of therapy.

calendar period has been used as an instrument in cancer research ([2]), and geographic location has been used as an instrument in cardiovascular research ([40]).

Johnston and Gustafson (2008) [15]:

Given an outcome measure, a treatment variable, a set of unmeasured confounders, and, possibly, a set of measured confounders, a variable satisfies the conditions of an instrumental variable (also referred to as an instrument) if the following conditions hold:

- (1) it is correlated with the treatment variable;
- (2) given the treatment variable and all confounders (measured and unmeasured), it is conditionally independent of outcome; and
- (3) it is independent of the entire set of unmeasured confounders [4].

The strength of the instrument is determined by the correlation between treatment and instrument, with strong instruments being preferable.

Stukel (2007) [40]:

“The instrumental variable behaves like a natural randomization” of patients to regional ‘treatment groups’ that differ in likelihood of receiving cardiac catheterization. Unlike randomization, the difference in likelihood of treatment is not 100%, and one can explore but not prove that the groups are similar in unmeasured patient characteristics.”

3. BAYESIAN G-COMPUTATION

Vansteelandt and Keiding (2010) [44]: Invited Commentary: G-Computation: Lost in Translation?

Daniel (2011) [9]:

This article describes the `gformula` command, an implementation of the g-computation procedure, used to estimate the causal effect of time-varying exposure(s) on an outcome in the presence of time-varying confounders that are themselves also affected by the exposure(s). **The procedure can also be used to address the related problem of estimating controlled direct effects and natural direct/indirect effects when the causal effect of the exposure(s) on an outcome is mediated by intermediate variables, and in particular when confounders of the mediator-outcome relationships are themselves affected by the exposure(s).** A brief overview of the theory and a description of the command and its options are given, and an illustration using two simulated examples is provided.

Snowden (2011) [39]:

The growing body of work in the epidemiology literature focused on G-computation includes theoretical explanations of the method but very few simulations or examples. The authors provide a step-by-step demonstration of G-computation that is intended to familiarize the reader with this procedure.

The distinction between marginal and conditional parameters is important, as it further highlights the limitations of a traditional regression approach when a population-level estimate is of interest.

Neugebauer (2007) [27]:

Beyond allowing a more flexible causal analysis, HRMSMs improve computational tractability and mitigate statistical power concerns when designing longitudinal studies. We also develop three consistent estimators of HRMSM parameters under sufficient model assumptions:

- 1) the Inverse Probability of Treatment Weighted (IPTW),
- 2) G-computation and
- 3) Double Robust (DR) estimators.

Park and Kaplan (2015) [28]:

A fully Bayesian approach to causal mediation analysis for group-randomized designs is presented. A unique contribution of this article is the combination of Bayesian inferential methods with G-computation to address the problem of heterogeneous treatment or mediator effects.

The results of the first simulation study demonstrate that our proposed Bayesian approach to causal mediation analysis satisfies Bayesian as well as frequentist properties. The second simulation study suggests that the proposed approach provides unbiased results even under heterogeneous treatment or mediator effects by using G-computation as long as models are correctly specified.

Keil (2014) [18]:

Unlike standard regression approaches, the parametric g-formula can be used to adjust for time-varying confounders that are affected by prior exposures. To date, there are few published examples in which the method has been applied.

Epidemiologists often use regression models (for example, the Cox proportional hazards model) to adjust for confounding; this is equivalent to estimating stratum-specific hazard ratios and then averaging the information-weighted hazard ratios. When some of those confounders are also causal intermediates, this amounts to adjusting away some of the effect of exposure (Weinberg(1993) [46], Rosenbaum (1984) [36]). The g-formula works differently: first, one finds weighted averages of the stratum-specific hazards, and then those averaged (standardized) hazards are combined in a summary hazard ratio. Thus, bias resulting from time-varying covariates that can be both confounders and causal intermediates is a shortcoming of using regression models to control for confounding, rather than a general principle of observational data analysis. The g-formula is a tool that overcomes this shortcoming, but its use in the literature has been sparse – we could find only 9 examples using observational data.

Keil (2015) [17]:

Bayesian approaches have shown promise for dealing with under-identified models in environmental epidemiology through the use of parameter prior distributions to stabilize estimates that would be highly variable (or possibly inestimable) in an unpenalized maximum likelihood procedure

Because standard g-formula algorithms typically require large Monte Carlo samples and a non-parametric bootstrap, the use of MCMC machinery in the Bayesian g-formula does not result in a substantial computational burden relative to the frequentist approach.

As in our time-fixed simulations, under moderately informative, null priors, the bias was larger and the variance was smaller for the Bayesian g-formula than for the standard g-formula. The Bayesian g-formula under moderately informative priors outperformed the standard g-formula at sample sizes under 100 (with respect to MSE), suggesting that the utility of the Bayesian g-formula increases with reasonable, prior information (Table 2).

With respect to unknown confounders, Recently, Bayesian methods have been proposed to address McCandless (2007)[24].

4. APPLICATIONS

Cole (2013) [7] :

Analysis of Occupational Asbestos Exposure and Lung Cancer Mortality Using the G Formula

The G null paradox states that, under the null hypothesis, it may be impossible to correctly specify the parametric models required for the G formula (11). Therefore, in large samples, the parametric G formula may reject the null hypothesis, even when it is true. In our context we are not overly concerned about the G null paradox, because asbestos is an established cause of lung cancer (31).

Future implementations might explore the introduction of prior knowledge into the parametric G formula through Bayesian methods.

Taubman (2009) [41]:

Unlike standard methods, the parametric g-formula can deliver consistent estimates of risk even when time-dependent confounders are affected by prior components of the intervention; although, in this setting, the parametric g-formula is subject to the ‘g-null paradox’ theorem, which implies it can be essentially impossible to correctly specify the needed parametric models under the causal null hypothesis. As a consequence, the method will reject the causal null, even when true, in sufficiently large samples.

Estimating the population risk of disease under hypothetical interventions – such as the population risk of coronary heart disease (CHD) were everyone to quit smoking and start exercising or to start exercising if diagnosed with diabetes?may not be possible using standard analytic techniques. The parametric g-formula, which appropriately adjusts for time-varying confounders affected by prior exposures, is especially well suited to estimating effects when the intervention involves multiple factors (joint interventions) or when the intervention involves decisions that depend on the value of evolving time-dependent factors

(dynamic interventions). We describe the parametric g-formula, and use it to estimate the effect of various hypothetical lifestyle interventions on the risk of CHD using data from the Nurses' Health Study.

We discuss whether the assumptions required for the validity of the parametric g-formula hold in the Nurses' Health Study data. This work represents the first large-scale application of the parametric g-formula in an epidemiologic cohort study.

Petersen (2007) [32]:

Marginal structural models were used to estimate the effect of pillbox organizer use on adherence and viral suppression, adjusting for confounding by CD4+ T cell count, viral load, prior adherence, recreational drug use, demographic characteristics, and current and past treatment.

Three distinct estimators:

1- inverse probability of treatment weighted (IPTW),

2-G-computation

3- double robust

G-computation (in the current application) relies on standard multivariable regression of adherence on pillbox organizer use and confounders. IPTW estimation, in contrast, relies on multivariable logistic regression of pillbox organizer use on confounders (the treatment mechanism).

The double robust estimator makes use of (1) the multivariable regression of adherence on pillbox organizer use and confounders used in G-computation and (2) the multi- variable logistic regression of pillbox organizer use on confounders (treatment mechanism) used in IPTW estimation.

Time-lagged confounder measurements were used to ensure that confounders occurred before (and, therefore, could not be influenced by) pillbox organizer use.

Moore (2008) [26]:

This traditional approach does not answer directly our original question of interest:

We implemented two estimators of history-restricted marginal structural models (HRMSMs) (Neugebauer et al. 2007) causal parameters: the inverse probability of treatment weighted

(IPTW) and G-computation. Confidence intervals (CIs) and p-values for the two estimates were obtained with 10,000 bootstrap iterations, where resampling was based on the 195 independent grids.

We applied a diagnostic tool to assess the bias in the IPTW estimator due to the experimental treatment assignment (ETA) violation (Wang et al. 2006) and showed a 76% bias in comparison to the G-computation estimate (Table 4). Therefore, we relied on the G-computation estimator.

This would represent the same effect estimated by the conventional regression analysis. In other words, the results from the MSM and conventional analyses, in this particular analysis, give identical parameter estimates because there are no interaction terms in the conventional model.

The results of this analysis indicated that the conventional statistical association model, in this particular analysis, was equivalent to the G-computation estimates of the HRMSM parameters?an observation that is not surprising, given that there were no interactions in the association model.

5. DF

The objective is to determine the significance (superiority or non-inferiority) of the target value $T_{(1,1)-(0,0)}$:

$$(1) \quad T_{(1,1)-(0,0)} = Pr(Y_2 = 1|do(X_1 = 1, X_2 = 1)) - Pr(Y_2 = 1|do(X_1 = 0, X_2 = 0))$$

	IV g-comp with H g-comp without H		
Assumption			
$H \perp\!\!\!\perp Z, \theta_{10} = 0$	✓	✗	✓
$(H X, C, Z) \perp\!\!\!\perp Y, \theta_8 = 0$	✓	✗	✓
H is correlated with X, $\theta_2 > 0$	✓	✗	✗
No unknown confounders, $\theta_9 = 0$	✗	✓	✓
Required Data			
H	✓	✓	✗
X	✗	✓	✓
C	✗	✓	✓

TABLE 3. Assumptions and required data for both approaches.

	Representation		Interpretation
Variable	if = 0	if = 1	
H	early	late	Instrumental Variable Calendar time
C	low	high	Known Confounder Relapse Rate
Y	ill	healthy	Outcome of interest Disease progression
Z	Unknown additional confounder		
X	no treatment	treatment	Intervention of interest Treatment

TABLE 4. Model Components.

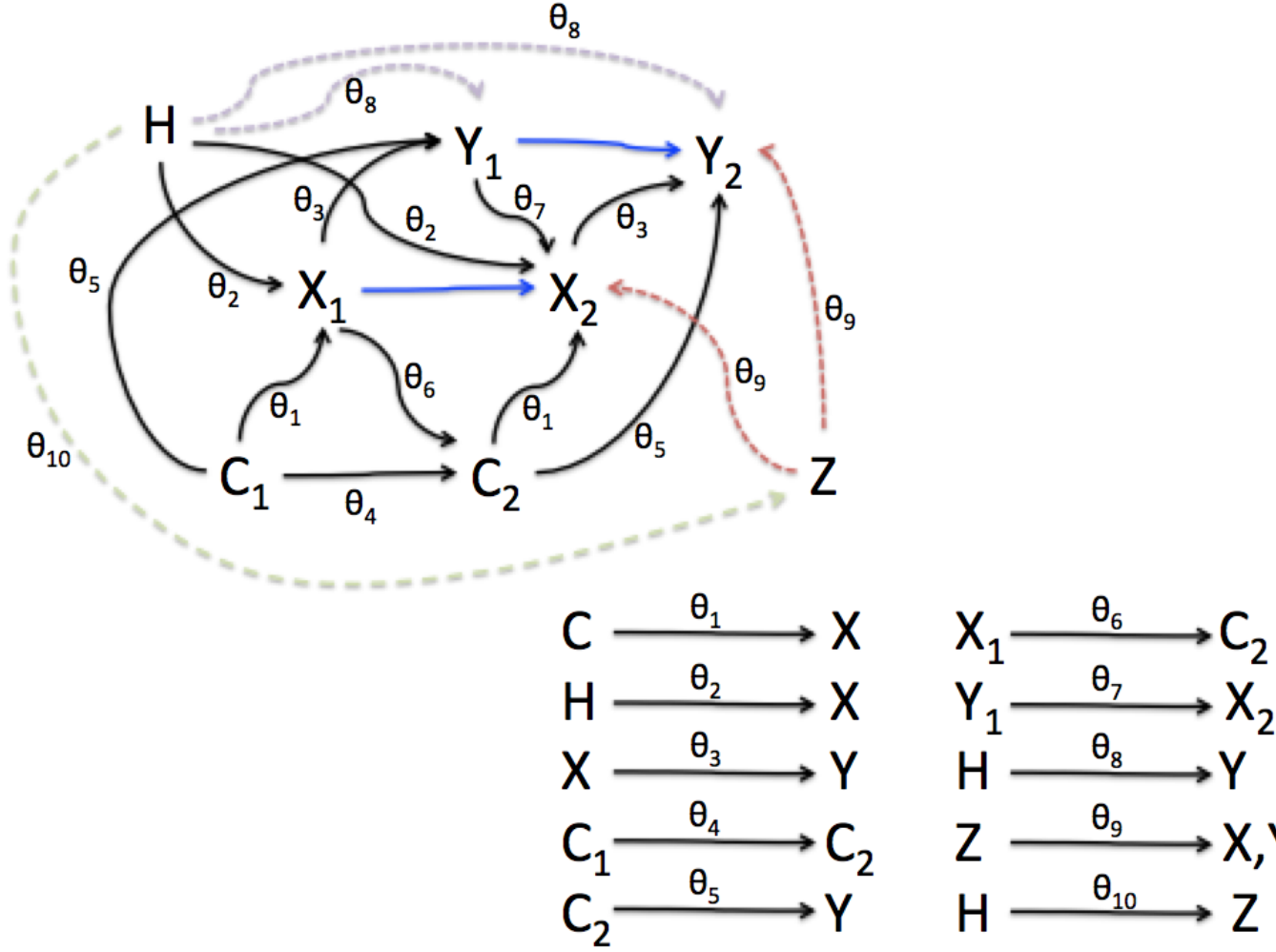


FIGURE 1. Consider the variables as unfolding in the temporal order $(H, C_1, X_1, Y_1, C_2, X_2, Y_2)$.

Given an outcome measure, a treatment variable, a set of unmeasured confounders, and, possibly, a set of measured confounders, a variable satisfies the conditions of an instrumental variable (also referred to as an instrument) if the following conditions hold:

- (1) it is correlated with the treatment variable
- (2) given the treatment variable and all confounders (measured and unmeasured), it is conditionally independent of outcome; and
- (3) it is independent of the entire set of unmeasured confounders [13].

6. METHODS

Frequentist g-comp with logistic model:

- (1) Fit a regression model: $Y_2 \sim H$
- (2) Significant association ($Y_2 \sim H$) \Rightarrow Significant association ($Y_2 \sim X$)

Bayesian g-comp with Uniform priors: $Pr(A = 1|B = b, C = c) \sim \text{Beta}(1 + \sum_{i=1}^n 1_{(A=1, B=b, C=c)}(x_i), 1 + \sum_{i=1}^n 1_{(A=0, B=b, C=c)}(x_i))$

$$f(x) = \frac{\Gamma(a+b)}{(\Gamma(a)\Gamma(b))} x^{(a-1)} (1-x)^{(b-1)}$$

- (1) Obtain a sample from the posterior distribution γ .
- (2) By summation (integration), use samples of γ to obtain samples from the posterior distribution of $Pr(Y_2 = 1|do(X_1 = x_1, X_2 = x_2))$
- (3) Calculate posterior predictive p -value.

7. 4 g -COMPUTATION MODELS

With Y_1 , without H :

$$\begin{aligned}
 Pr(Y_2 = 1|do(X_1 = x_1, X_2 = x_2)) &= \sum_{y_1=0}^1 \sum_{c_1=0}^1 \sum_{c_2=0}^1 Pr(C_1 = c_1) \\
 &\quad \cdot Pr(C_2 = c_2|C_1 = c_1, X_1 = x_1) \\
 &\quad \cdot Pr(Y_1 = y_1|C_1 = c_1, X_1 = x_1) \\
 &\quad \cdot Pr(Y_2 = 1|C_2 = c_2, X_2 = x_2, Y_1 = y_1)
 \end{aligned}$$

With Y_1 , and with H :

$$\begin{aligned}
Pr(Y_2 = 1|do(X_1 = x_1, X_2 = x_2)) &= \sum_{y_1=0}^1 \sum_{c_1=0}^1 \sum_{c_2=0}^1 \sum_{h=0}^1 Pr(H = h) \cdot Pr(C_1 = c_1|H = h) \\
&\cdot Pr(C_2 = c_2|C_1 = c_1, X_1 = x_1, H = h) \\
&\cdot Pr(Y_1 = y_1|C_1 = c_1, X_1 = x_1, H = h) \\
&\cdot Pr(Y_2 = 1|C_2 = c_2, X_2 = x_2, Y_1 = y_1, H = h)
\end{aligned}$$

Without Y_1 , without H :

$$\begin{aligned}
Pr(Y_2 = 1|do(X_1 = x_1, X_2 = x_2)) &= \sum_{c_1=0}^1 \sum_{c_2=0}^1 Pr(C_1 = c_1) \\
&\cdot Pr(C_2 = c_2|C_1 = c_1, X_1 = x_1) \\
&\cdot Pr(Y_2 = 1|C_2 = c_2, C_1 = c_1, X_2 = x_2, X_1 = x_1)
\end{aligned}$$

Without Y_1 , with H :

$$\begin{aligned}
Pr(Y_2 = 1|do(X_1 = x_1, X_2 = x_2)) &= \sum_{c_1=0}^1 \sum_{c_2=0}^1 \sum_{h=0}^1 Pr(H = h) \cdot Pr(C_1 = c_1|H = h) \\
&\cdot Pr(C_2 = c_2|C_1 = c_1, X_1 = x_1, H = h) \\
&\cdot Pr(Y_2 = 1|C_2 = c_2, X_2 = x_2, C_1 = c_1, X_1 = x_1, H = h)
\end{aligned}$$

- run down of the simple sandbox of two time-points, everything binary - g-computation (perhaps motivated several different ways) - IV idea

Comparison in a focussed scenario - confounding by indication present (C encourages treatment start) - different possible forms of treatment effect (through C and/or around C) - both sets of assumptions met

Comparison across a broad range of scenarios

- set up distribution giving rise to a wide array of states of the world - look for patterns, e.g., is the power of the g-computation driven only by θ_{11} - θ_{00} , or are there other important factors

Violations of assumptions

- any defensible sense of one procedure being more robust than the other - how quickly is power lost as we relax assumptions

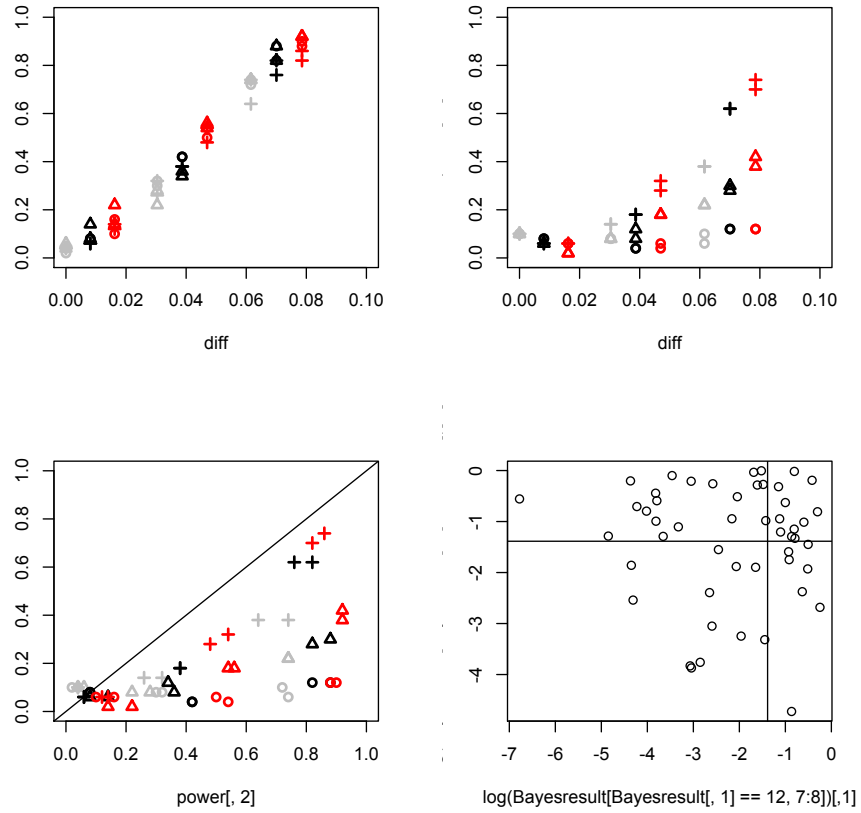


FIGURE 2. Results of Simulation study.

8. APPLICATION

9. DISCUSSION

REFERENCES

- [1] BORSI, L. *Estimating the Causal Effect of Switching to Second-line Antiretroviral HIV Treatment Using G-computation*. PhD thesis, ETH, Department of Mathematics, 2012.
- [2] BRENNER, H. Long-term survival rates of cancer patients achieved by the end of the 20th century: a period analysis. *The Lancet* 360, 9340 (2002), 1131–1135.
- [3] BRION, M.-J. A., SHAKHBAZOV, K., AND VISSCHER, P. M. Calculating statistical power in mendelian randomization studies. *International journal of epidemiology* 42, 5 (2013), 1497–1501.
- [4] BURGESS, S. Sample size and power calculations in mendelian randomization with a single instrumental variable and a binary outcome. *International journal of epidemiology* (2014), dyu005.
- [5] CAIN, L. E., COLE, S. R., GREENLAND, S., BROWN, T. T., CHMIEL, J. S., KINGSLEY, L., AND DETELS, R. Effect of highly active antiretroviral therapy on incident aids using calendar period as an instrumental variable. *American journal of epidemiology* (2009), kwp002.
- [6] CHAKRABORTY, B., AND MOODIE, E. E. G-computation: Parametric estimation of optimal dtrs. In *Statistical Methods for Dynamic Treatment Regimes*. Springer, 2013, pp. 101–112.
- [7] COLE, S. R., RICHARDSON, D. B., CHU, H., AND NAIMI, A. I. Analysis of occupational asbestos exposure and lung cancer mortality using the g formula. *American journal of epidemiology* (2013), kws343.
- [8] DANIEL, R., COUSENS, S., DE STAVOLA, B., KENWARD, M., AND STERNE, J. Methods for dealing with time-dependent confounding. *Statistics in Medicine* 32, 9 (2013), 1584–1618.
- [9] DANIEL, R. M., DE STAVOLA, B. L., COUSENS, S. N., ET AL. gformula: Estimating causal effects in the presence of time-varying confounding or mediation using the g-computation formula. *Stata Journal* 11, 4 (2011), 479.
- [10] DECKER, A. L., HUBBARD, A., CRESPI, C. M., SETO, E. Y., AND WANG, M. C. Semiparametric estimation of the impacts of longitudinal interventions on adolescent obesity using targeted maximum-likelihood: Accessible estimation with the ltmle package. *Journal of Causal Inference J. Causal Infer.* 2, 1 (2014), 95–108.
- [11] DESQUILBET, L., MARGOLICK, J. B., FRIED, L. P., PHAIR, J. P., JAMIESON, B. D., HOLLOWAY, M., AND JACOBSON, L. P. Relationship between a frailty-related phenotype and progressive deterioration of the immune system in hiv-infected men. *Journal of acquired immune deficiency syndromes (1999)* 50, 3 (2009), 299.
- [12] GLYMOUR, M. M., TCHETGEN, E. J. T., AND ROBINS, J. M. Credible mendelian randomization studies: approaches for evaluating the instrumental variable assumptions. *American journal of epidemiology* 175, 4 (2012), 332–339.
- [13] GREENLAND, S. An introduction to instrumental variables for epidemiologists. *International journal of epidemiology* 29, 4 (2000), 722–729.

- [14] GUSTAFSON, P. Discussion of ?on bayesian estimation of marginal structural models? *Biometrics* 71, 2 (2015), 291–293.
- [15] JOHNSTON, K., GUSTAFSON, P., LEVY, A., AND GROOTENDORST, P. Use of instrumental variables in the analysis of generalized linear models in the presence of unmeasured confounding with applications to epidemiological research. *Statistics in medicine* 27, 9 (2008), 1539–1556.
- [16] KARIM, M. E., GUSTAFSON, P., PETKAU, J., ZHAO, Y., SHIRANI, A., KINGWELL, E., EVANS, C., VAN DER KOP, M., OGER, J., AND TREMLETT, H. Marginal structural cox models for estimating the association between β -interferon exposure and disease progression in a multiple sclerosis cohort. *American journal of epidemiology* 180, 2 (2014), 160–171.
- [17] KEIL, A. P., DAZA, E. J., ENGEL, S. M., BUCKLEY, J. P., AND EDWARDS, J. K. A bayesian approach to the g-formula. *arXiv preprint arXiv:1512.04809* (2015).
- [18] KEIL, A. P., EDWARDS, J. K., RICHARDSON, D. R., NAIMI, A. I., AND COLE, S. R. The parametric g-formula for time-to-event data: towards intuition with a worked example. *Epidemiology (Cambridge, Mass.)* 25, 6 (2014), 889.
- [19] KUROKI, M., AND CAI, Z. Instrumental variable tests for directed acyclic graph models.
- [20] LEE, M.-J., AND HUANG, F. Structural instrumental variable estimator for dynamic treatment effects: Spanking effect on behavior.
- [21] LEE, M.-J., AND HUANG, F. Finding dynamic treatment effects under anticipation: the effects of spanking on behaviour. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 175, 2 (2012), 535–567.
- [22] LOK, J., GILL, R., VAN DER VAART, A., AND ROBINS, J. Estimating the causal effect of a time-varying treatment on time-to-event using structural nested failure time models. *Statistica Neerlandica* 58, 3 (2004), 271–295.
- [23] MACK, C. D., BROOKHART, M. A., GLYNN, R. J., MEYER, A. M., CARPENTER, W. R., SANDLER, R. S., AND STÜRMER, T. Comparative effectiveness of oxaliplatin versus 5-fluorouracil in older adults: An instrumental variable analysis. *Epidemiology* 26, 5 (2015), 690–699.
- [24] MCCANDLESS, L. C., GUSTAFSON, P., AND LEVY, A. Bayesian sensitivity analysis for unmeasured confounding in observational studies. *Statistics in medicine* 26, 11 (2007), 2331–2347.
- [25] MCMAHON, A. D., AND MACDONALD, T. M. Design issues for drug epidemiology. *British journal of clinical pharmacology* 50, 5 (2000), 419–425.
- [26] MOORE, K., NEUGEBAUER, R., LURMANN, F., HALL, J., BRAJER, V., ALCORN, S., AND TAGER, I. Ambient ozone concentrations cause increased hospitalizations for asthma in children: an 18-year study in southern california. *Environmental Health Perspectives* 116, 8 (2008), 1063.
- [27] NEUGEBAUER, R., VAN DER LAAN, M. J., JOFFE, M. M., AND TAGER, I. B. Causal inference in longitudinal studies with history-restricted marginal structural models. *Electronic journal of statistics* 1 (2007), 119.

- [28] PARK, S., AND KAPLAN, D. Bayesian causal mediation analysis for group randomized designs with homogeneous and heterogeneous effects: Simulation and case study. *Multivariate behavioral research* 50, 3 (2015), 316–333.
- [29] PETER, K. *Marginal structural models and causal inference*. PhD thesis, Master thesis, ETH, 2011.
- [30] PETERSEN, M., SCHWAB, J., GRUBER, S., BLASER, N., SCHOMAKER, M., AND VAN DER LAAN, M. Targeted maximum likelihood estimation for dynamic and static longitudinal marginal structural working models. *Journal of causal inference* 2, 2 (2014), 147–185.
- [31] PETERSEN, M. L. Commentary: Applying a causal road map in settings with time-dependent confounding. *Epidemiology (Cambridge, Mass.)* 25, 6 (2014), 898–901.
- [32] PETERSEN, M. L., WANG, Y., VAN DER LAAN, M. J., GUZMAN, D., RILEY, E., AND BANGSBERG, D. R. Pillbox organizers are associated with improved adherence to hiv antiretroviral therapy and viral suppression: a marginal structural model analysis. *Clinical Infectious Diseases* 45, 7 (2007), 908–915.
- [33] PIERCE, B. L., AHSAN, H., AND VANDERWEELE, T. J. Power and instrument strength requirements for mendelian randomization studies using multiple genetic variants. *International journal of epidemiology* (2010), dyq151.
- [34] PIERCE, B. L., AND BURGESS, S. Efficient design for mendelian randomization studies: subsample and 2-sample instrumental variable estimators. *American journal of epidemiology* 178, 7 (2013), 1177–1184.
- [35] ROBINS, J. M. General methodological considerations. *Journal of Econometrics* 112, 1 (2003), 89–106.
- [36] ROSENBAUM, P. R. The consequences of adjustment for a concomitant variable that has been affected by the treatment. *Journal of the Royal Statistical Society. Series A (General)* (1984), 656–666.
- [37] SAARELA, O., STEPHENS, D. A., MOODIE, E. E., AND KLEIN, M. B. On bayesian estimation of marginal structural models. *Biometrics* (2015).
- [38] SHINOZAKI, T., MATSUYAMA, Y., AND OHASHI, Y. Estimation of controlled direct effects in time-varying treatments using structural nested mean models: application to a primary prevention trial for coronary events with pravastatin. *Statistics in medicine* 33, 18 (2014), 3214–3228.
- [39] SNOWDEN, J. M., ROSE, S., AND MORTIMER, K. M. Implementation of g-computation on a simulated data set: demonstration of a causal inference technique. *American Journal of Epidemiology* (2011), kwq472.
- [40] STUKEL, T. A., FISHER, E. S., WENNBERG, D. E., ALTER, D. A., GOTTLIEB, D. J., AND VERMEULEN, M. J. Analysis of observational studies in the presence of treatment selection bias: effects of invasive cardiac management on ami survival using propensity score and instrumental variable methods. *Jama* 297, 3 (2007), 278–285.
- [41] TAUBMAN, S. L., ROBINS, J. M., MITTLEMAN, M. A., AND HERNÁN, M. A. Intervening on risk factors for coronary heart disease: an application of the parametric g-formula. *International Journal of Epidemiology* 38, 6 (2009), 1599–1611.
- [42] TREMLETT, H., ZHAO, Y., AND DEVONSHIRE, V. Natural history of secondary-progressive multiple sclerosis. *Multiple sclerosis* 14, 3 (2008), 314–324.

- [43] TREMLETT, H., ZHAO, Y., RIECKMANN, P., AND HUTCHINSON, M. New perspectives in the natural history of multiple sclerosis. *Neurology* 74, 24 (2010), 2004–2015.
- [44] VANSTEELANDT, S., AND KEIDING, N. Invited commentary: G-computation—lost in translation? *American journal of epidemiology* (2011), kwq474.
- [45] WANG, W., SCHARFSTEIN, D., WANG, C., DANIELS, M., NEEDHAM, D., AND BROWER, R. Estimating the causal effect of low tidal volume ventilation on survival in patients with acute lung injury. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 60, 4 (2011), 475–496.
- [46] WEINBERG, C. R. Toward a clearer definition of confounding. *American Journal of Epidemiology* 137, 1 (1993), 1–8.