American Journal of Epidemiology Advance Access published March 16, 2011

American Journal of Epidemiology

DOI: 10.1093/aje/kwq472

Practice of Epidemiology

Implementation of G-Computation on a Simulated Data Set: Demonstration of a Causal Inference Technique

Jonathan M. Snowden*, Sherri Rose, and Kathleen M. Mortimer

* Correspondence to Jonathan M. Snowden, Division of Epidemiology, School of Public Health, University of California, Berkeley, 1918 University Avenue, Suite 3C, Berkeley, CA 94704 (e-mail: jsnowden@berkeley.edu).

Initially submitted April 8, 2010; accepted for publication October 8, 2010.

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 of application. The small number of G-computation analyses in the epidemiology literature relative to other causal inference approaches may be partially due to a lack of didactic explanations of the method targeted toward an epidemiology audience. The authors provide a step-by-step demonstration of G-computation that is intended to familiarize the reader with this procedure. The authors simulate a data set and then demonstrate both G-computation and traditional regression to draw connections and illustrate contrasts between their implementation and interpretation relative to the truth of the simulation protocol. A marginal structural model is used for effect estimation in the G-computation example. The authors conclude by answering a series of questions to emphasize the key characteristics of causal inference techniques and the G-computation procedure in particular.

air pollution; asthma; causality; methods; regression analysis

Abbreviations: FEV_1, forced expiratory volume in 1 second; IPTW, inverse probability-of-treatment weighting; MSM, marginal structural model.

Editor's note: *An invited commentary on this article appears on page **000**, and the authors' response appears on page **000**.*

Statistical methods from the causal inference literature are used with increasing frequency in epidemiology to estimate causal effects from observational data (1–6). G-computation, a maximum likelihood substitution estimator of the G-formula, is one such approach to causal-effect estimation (7). Application of this method allows investigators to use observational data to estimate parameters that would be obtained in a perfectly randomized controlled trial. Under certain assumptions, these estimates can be interpreted causally.

Like other causal inference approaches, G-computation can be understood and implemented through the use of the counterfactual framework, which posits the existence of unobserved outcomes corresponding to theoretical unobserved exposures in addition to the observed data that are collected. Because it is impossible to observe each study participant under all possible treatment or exposure regimens (words that we use interchangeably), outcomes that would have occurred under this alternate exposure scenario can be considered missing data, the absence of which prevents the straightforward estimation of unbiased causal effects. G-computation and other methods for causal inference can use the existence of counterfactuals (8) (i.e., the entire set of possible outcomes) to enable unbiased estimation of marginal causal effects.

There is a growing body of work in the epidemiology literature focused on G-computation, including theoretical explanations of the method (9) and a didactic demonstration of the application of the approach to an intervention setting that uses real data (10). With some notable exceptions (2, 11, 12), however, relatively few examples of the implementation of G-computation exist. The implementation of the G-computation approach is not more difficult than inverse

probability-of-treatment weighting (IPTW), another method for causal inference that has been used more frequently thus far in epidemiology (1, 4, 6, 13–15). Both of these estimators can be implemented in standard statistical software and can estimate parameters in a marginal structural model (MSM). The relative predominance of the IPTW approach in published analyses may be partially due to a greater number of didactic explanations of the method targeted toward an epidemiologic audience (16–18). Implementation of the G-computation estimator is equivalent to using the marginal distribution of the covariates as the standard in standardization, a familiar class of procedures in epidemiology (19, 20).

As epidemiologists become more familiar with causal inference methods, the decision to use a statistical approach should be dictated by the technique that is best suited to answer the research question of interest in a given data set, rather than which option is most accessible or most commonly used. Especially in light of concerns about the statistical inefficiency of the IPTW estimator (21–23), researchers will benefit from having more than one analytic technique at their disposal. Also, other more advanced causal inference techniques, such as targeted maximum likelihood estimation (24), build on the G-computation approach, which may further motivate researchers to invest in learning this technique.

The present article details a step-by-step demonstration of G-computation that is intended to familiarize the reader with the procedure. We aimed to clarify both the conceptual basis and the actual implementation of G-computation, not to provide detailed information about the estimator's statistical properties (7) or to replace the statistical support often needed to implement causal inference techniques. The intended audience for this didactic explanation includes researchers who are well-versed in standard statistical techniques, such as regression, but who are less familiar with causal inference methodology. We demonstrated both G-computation and traditional regression to draw connections and illustrate contrasts between their implementation and interpretation. For the purpose of this article, we used "traditional regression" to denote conventional regression approaches such as maximum likelihood estimation for parametric models. As the purpose of this article was to clearly demonstrate a method rather than a subject-matter analysis, we analyzed simulated data whose simulation protocol is known. Through simulation, we know the "truth" about the data-generating distribution (if there is a true underlying effect), a luxury we never have in observational data analysis. Causal inference methods have been applied most frequently in the epidemiology literature in the context of MSMs; thus, we implemented G-computation to estimate a parameter in an MSM, although this approach can also be used to estimate parameters without an MSM. We conclude by posing and answering a series of questions related to G-computation, to further emphasize the key characteristics of this approach.

SIMULATED DATA

Suppose we are interested in the causal effect of ozone on pulmonary function in asthmatic individuals. Although we need to account for possible confounding and interaction in our sample, we are interested in a marginal estimate of the exposure effect in this vulnerable population that can be used to inform regulatory standards. Traditional approaches for controlling confounding and dealing with effect modification produce conditional estimates, which are not of interest in some observational settings. (G-computation can also be used to estimate adjusted effects that are conditional only upon stratification variables of interest—e.g., a priori effect modifiers—rather than nuisance variables that are not relevant to the question of interest (22).)

We simulated a single data set with a sample size of 300, a realistic size for a study, and generated 1 outcome variable, a single exposure variable, and 2 additional covariates. We used R (version 2.10.0; R Foundation for Statistical Computing, Vienna, Austria) for all simulations and analyses. The variable for gender (W_1) was generated from a Bernoulli distribution, with a probability of being male of 40% ($P(W_1 = 1) = 0.4$). A binary baseline variable for controller medication use (W_2) was generated from a Bernoulli distribution, with the probability of controller medication use of 50% ($P(W_2 = 1) = 0.5$). The covariates W_1 and W_2 may collectively be referred to as the vector $\boldsymbol{W} = \{W_1, W_2\}$ and are independent of each other. The binary exposure variable (A) was ozone exposure. The following equation describes the probability of being exposed to levels of ozone above the US Environmental Protection Agency regulatory standard of 75 parts per billion, where, for this teaching example, the simulation protocol assigned higher exposures to males and nonmedication users:

$$\begin{aligned}\text{Generation of ozone}(A) : \text{Logit}(P(A = 1|W_1, W_2)) \\ = 0.5 + 0.2 \times W_1 - 0.3 \times W_2.\end{aligned}$$

The continuous outcome (Y) represents forced expiratory volume in 1 second (FEV_1) measured in liters and was generated from a normal distribution with the following mean, and a standard deviation of 0.4 L:

$$\begin{aligned}\text{Generation of FEV}_1(Y) : \text{E}(Y|A, W_1, W_2) \\ = 3 - 0.5 \times A + 1 \times W_1 + 0.3 \times A \times W_2.\end{aligned}$$

This simple simulated data set, the properties of which are described in Table 1, has 1 exposure of interest (ozone; A) and 2 covariates (W_1 and W_2) that must be considered when analyzing the effects of exposure on the outcome (FEV_1; Y). Gender (W_1) affects both A and Y and therefore confounds the exposure-outcome association, requiring some method of adjustment for an unbiased effect estimate of A. In contrast, controller medication use (W_2) affects the exposure and the outcome, yet W_2 has no independent effect on Y, as evidenced by the lack of a main term for W_2. Instead, W_2 is an effect modifier for the effect of A on Y, with no independent contribution beyond its joint effect with A.

This data generation protocol implies that ozone has a negative effect among non-controller medication users and an attenuated negative effect among controller medication users. The effect of ozone is described by both the main term A and the $A \times W_2$ interaction term. The main term for A taken independently estimates the ozone effect in the non-controller medication group (−0.5 L), in contrast to the

Table 1. Summary Statistics and Protocol for Sex, Controller Medication Use, High Ozone Exposure, and Forced Expiratory Volume in 1 Second[a] ($n = 300$)

Variable	Notation	Binary		Continuous		Simulation Protocol
		No.	%	Median	Interquartile Range	
Male sex: 1 = yes, 0 = no	W_1	122	40.7			$P(W_1 = 1) = 0.4$
Controller medication user: 1 = yes, 0 = no	W_2	148	49.3			$P(W_2 = 1) = 0.5$
High ozone exposure: 1 = high, 0 = low	A	119	39.7			$\text{Logit}(P(A = 1 \mid W_1, W_2)) = 0.5 + 0.2 \times W_1 - 0.3 \times W_2$
Forced expiratory volume in 1 second, L	Y			3.12	2.75–3.74	$E(Y \mid A, W_1, W_2) = 3 - 0.5 \times A + 1 \times W_1 + 0.3 \times A \times W_2$ (standard deviation, 0.4)

[a] All variables are binary except for forced expiratory volume in 1 second, which is continuous.

estimate for controller medication users, which requires summation of the A main-term coefficient and the coefficient for the $A \times W_2$ interaction term ($-0.5 + 0.3 = -0.2$ L). A weighted average based on the population distribution of W_2 results in a population-level effect of -0.35 L, which is the "truth" of the marginal effect of exposure A on outcome Y.

Our simulated data set takes the place of the reader's own real data, with the benefit of knowing that there is a true effect because we generated the relation between variables. We note that evaluation of a statistical estimator requires repeated simulation from the data-generating distribution, and we use a single simulated data set in this article to demonstrate the implementation of regression and G-computation only. We also discuss similarities and differences in assumptions and parameter interpretation for both regression and G-computation.

REGRESSION IMPLEMENTATION

Because epidemiology focuses on data collected in observational settings regarding human beings living freely, it is unlikely that investigators can, a priori, specify the correct model for the data-generating distribution of the outcome given covariates and exposure. We fitted 4 different linear regression models, each one representing a different model that might have been reasonably selected by the investigator before analyzing the data. Other, more complex models could also have been selected.

One analysis used the correct model specification, 2 intentionally misspecified the model, and another contained the correct specification nested within it. We subsequently carried forward these regression models to the model-fitting step of the G-computation demonstration.

The first regression model we fitted was a crude regression of the outcome on the exposure.

$$\text{Regression model 1}: \mathrm{E}(Y|A, W_1, W_2) = \alpha_0 + \alpha_1 \times A.$$

This model, denoted model 1, is incorrect in that it excludes the confounder W_1 and does not account for the interaction by W_2. Model 2 is another misspecified model, with main terms for A, W_1, and W_2:

$$\text{Regression model 2}: \mathrm{E}(Y|A, W_1, W_2) = \alpha_0 + \alpha_1 \times A + \alpha_2 \times W_1 + \alpha_3 \times W_2.$$

A third model has the correct specification nested within it. Model 3 contains the main terms for A, W_1 and the $A \times W_2$ interaction term, but also includes a main term for W_2, despite the absence of an independent effect for controller medication use:

$$\text{Regression model 3}: \mathrm{E}(Y|A, W_1, W_2) = \alpha_0 + \alpha_1 \times A + \alpha_2 \times W_1 + \alpha_3 \times W_2 + \alpha_4 \times A \times W_2.$$

Model 4 is the correct model fit, with main terms for A and W_1, and a single interaction term between A and W_2.

$$\text{Regression model 4}: \mathrm{E}(Y|A, W_1, W_2) = \alpha_0 + \alpha_1 \times A + \alpha_2 \times W_1 + \alpha_3 \times A \times W_2.$$

Because we are interested only in the effects of exposure A on outcome Y, our discussion of the model results focuses on the effects of ozone rather than the confounders (W_1, W_2), which in this case may be considered nuisance parameters. Traditional regression model 1 estimates a crude ozone effect of approximately -0.23 L, as indicated in Table 2. This effect estimate is not adjusted for the confounding by gender, nor does it reflect the interaction present between controller medication use and ozone. Under traditional model 2, the A coefficient by itself summarizes the population-wide effect of ozone, which is a decrement of -0.36 L. The effect of ozone in model 3 is described by 2 coefficients: the effect of ozone among non-medication users is estimated by α_1 to be -0.48 L, and the effect of ozone among medication users is estimated by the sum of $\alpha_1 + \alpha_4$ to be -0.19 L. Similarly, the effect of ozone is not summarized in 1 single effect estimate in the correctly specified traditional model 4, which estimates an ozone effect of -0.49 L among the non-medication users and -0.18 L among controller medication users.

The presence of interaction between a covariate and the exposure means that the coefficients in a traditional regression do not estimate a population-level marginal effect. The effect estimate for the exposure in the correctly specified model for our data is conditional and, thus, is a fundamentally different parameter than the one we will estimate with G-computation. This situation might be desirable if the interaction is of a priori interest, such as in clinical settings where treatment would be effective or safe only in certain

Table 2. Results in Liters of Forced Expiratory Volume in 1 Second From Traditional Regression and G-Computation Marginal Structural Model Analyses of the Simulated Data ($n = 300$) for Each of the Regression Models[a]

Variable[b]	Variable Notation	Model 1		Model 2		Model 3		Model 4 (Correct Model)	
		Traditional Coefficient (SE)[c]	MSM Coefficient (SE)[c]	Traditional Coefficient (SE)[c]	MSM Coefficient (SE)[c]	Traditional Coefficient (SE)[c]	MSM Coefficient (SE)[c]	Traditional Coefficient (SE)[c]	MSM Coefficient (SE)[c]
Intercept		3.31 (0.05)	3.31 (0.05)	2.87 (0.05)	3.36 (0.04)	2.93 (0.05)	3.37 (0.04)	2.95 (0.04)	3.38 (0.04)
Ozone: 1 = high, 0 = low[d]	A/a	−0.23 (0.08)	−0.23 (0.08)	−0.36 (0.05)	−0.36 (0.06)	−0.48 (0.07)	−0.33 (0.06)	−0.49 (0.06)	−0.34 (0.06)
Male gender: 1 = yes, 0 = no	W_1			1.06 (0.05)		1.05 (0.05)		1.05 (0.05)	
Controller medication use: 1 = yes, 0 = no	W_2			0.13 (0.05)		0.02 (0.06)			
Ozone-medication interaction[d]	$A/a \times W_2$					0.29 (0.10)		0.31 (0.08)	

Abbreviations: MSM, marginal structural model; SE, standard error.

[a] For each of the 4 analyses, the same model was used for the traditional regression and the Q-model in the G-computation analysis.

[b] The true parameter values for the data-generating distribution are as follows: −0.5 L for ozone (A), 1 L for male (W_1), 0 for controller medication (W_2), and 0.3 L for the ozone-medication interaction term ($A \times W_2$).

[c] Standard errors were validly derived from a bootstrap procedure with 10,000 repetitions.

[d] The traditional regression coefficients and the MSM coefficients are numerically identical due to the lack of exposure/covariate interactions. The true marginal effect based on the simulation protocol is −0.35 L.

subgroups of patients. However, with policy-related research questions, marginal effects are often of greater interest. 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.

G-COMPUTATION IMPLEMENTATION

The first step of G-computation is to fit a regression of the outcome on the exposure and relevant covariates, using the observed data set. This regression model is frequently called the "Q-model" in the context of G-computation. We emphasize that despite the unfamiliar terminology and applications of the model, our Q-model is no different than a traditional regression of Y on A and W, such as the ones fitted above (regression models 1–4), and can also be implemented in standard statistical software. The equivalence of the regression model used in the traditional approach and the Q-model used in G-computation is demonstrated in Figure 1, which schematically outlines the steps of both approaches. For G-computation to estimate an unbiased exposure effect, the Q-model must be correctly specified, just as correct model specification is required for unbiased estimation in traditional regression. However, investigators may select from additional tools to nonparametrically or semiparametrically estimate the Q-model, including machine-learning algorithms that cannot be incorporated into a traditional regression approach (for example, super learner (25)).

Once estimated, the Q-model is used to predict counterfactual outcomes for each observation under each exposure regimen. This is accomplished by plugging $a = 1$ and then subsequently $a = 0$ into the regression fit to obtain a predicted outcome under these 2 settings. The counterfactual outcome associated with exposure intervention a can be represented as Y_a. Because we have a dichotomous treatment setting, the counterfactual outcomes correspond to Y_1 for exposed and Y_0 for unexposed. Generating counterfactual outcomes for each observation in under exposed and unexposed settings furnishes the investigator with a full data set that is free of confounding under causal assumptions and resolves the missing data problem described above. The decision to implement G-computation using continuous exposures raises many important considerations concerning the research question and resulting inference (26), including frequent violations of the experimental treatment assignment assumption (discussed in the next section). We have assumed time-ordering, where W was generated before A, and A before Y.

We fitted each of the 4 models described earlier as possible Q-models to predict Y for each of the 300 observations. We used the coefficients from each model to predict the values of Y_1 and Y_0 for each observation, leaving their covariates at the observed values but intervening on the value of a as described above ($a = 1$ when exposed, $a = 0$ when unexposed), thus generating, for each individual, Y_1 and Y_0, respectively. A step-by-step outline, including an illustration of the full data, is presented in the Web Appendix (available at http://aje.oxfordjournals.org/).

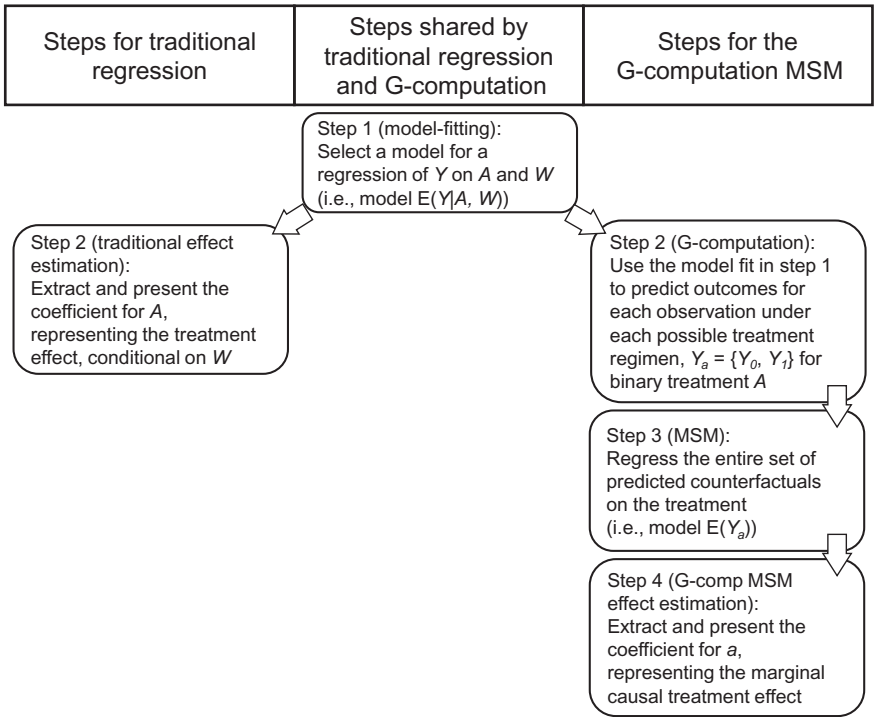

Figure 1. Contrasts between the implementation of traditional regression and the G-computation marginal structural model (G-comp MSM). *A* is a binary treatment, $W = \{W_1, W_2\}$ a vector of confounders, and *Y* a continuous outcome.

Having generated the full data with G-computation, we subsequently fitted an MSM of the outcome Y_a on the treatment a, to estimate the marginal effect of ozone on FEV_1:

$$\text{MSM}: E(Y_a) = \beta_0 + \beta_1 \times a.$$

Because we were interested in a marginal effect, our MSM included only the exposure of interest, ozone. Confounders and nuisance effect modifiers were adjusted for in the Q-model. We used the same MSM (presented above) for each of the 4 Q-models and corresponding full data sets, targeting the same parameter in each of the 4 MSM regressions. In this example, the causal risk difference estimated by our MSM is identical to the estimate that we could have calculated without an MSM, still defining our parameter of interest as $E(Y_1 - Y_0)$, as described in the Web Appendix. Our described G-computation implementation does not immediately extend to more complicated data structures (e.g., continuous *A*) and MSM specifications (e.g., unsaturated (16)).

Table 2 shows the results of the MSMs fitted using the G-computation estimator and the 4 separate Q-models described in the previous section. We used the notation MSM_{Q1} to refer to the MSM run on the counterfactual outcomes obtained by using model 1 for the Q-model. Recall from the discussion of model 1 that the corresponding MSM_{Q1} parameter estimate of −0.23 L for *A* is not adjusted for confounding by W_1 and does not account for the effect modification by W_2. Based upon the main-term-only Q-model that failed to include the $A \times W_2$ interaction term from the simulation protocol, MSM_{Q2} produced an effect estimate of −0.36 L for *A*. MSM_{Q3} relied on a nested model, thus estimating an ozone effect of −0.33 L, which was similar to the ozone effect of −0.34 L that was estimated by the MSM_{Q4}, which relied on the correctly specified Q-model.

Because the standard errors generated by software programs are not correct for G-computation parameter estimators in an MSM, the user must implement resampling-based methodology, such as bootstrapping, to determine accurate standard errors. For an accessible introduction to bootstrapping, we refer the reader to the appendix of another didactic article (10). We implemented a nonparametric bootstrap with 10,000 repetitions.

TRADITIONAL REGRESSION ANALYSIS AND G-COMPUTATION

The 2 approaches discussed here rely on similar assumptions. The G-computation estimator relies on the experimental treatment assignment or positivity assumption, which requires that there be a nonzero probability of each treatment regimen within all subgroups for which the effect of treatment is being assessed. This is analogous to the caution against extrapolating beyond the observed data in a traditional regression framework. Both estimators described here rely on the assumption of no unmeasured confounding for unbiased estimation and also time-ordering of certain

variables (e.g., *A* occurs before *Y*). Correct model specification is assumed in the traditional regression and the G-computation MSM approach (in which both the Q-model and the MSM are assumed to be correctly specified). Assumptions relating to the existence of counterfactuals and the consistency of counterfactual outcomes apply to the G-computation estimator and not traditional regression. Detailed explanations of these assumptions appear elsewhere (7, 16, 27, 28).

As we discuss results from these 2 analyses, it is important to remember that a direct comparison between the traditional regression and G-computation estimators is problematic because they estimate different parameters. The results from the traditional regression analysis and the subsequent G-computation analysis revealed that both approaches estimated the same ozone effect value for models 1 and 2. This was not a coincidence or a rounding artifact, but rather a necessary occurrence with a continuous outcome. Because models 1 and 2 both modeled the effect of ozone as a main term without any ozone-covariate interactions, the effect value of ozone on FEV_1 was estimated to be identical across all strata of the covariates. Therefore, the effect estimate produced by the initial regression is exactly reproduced for each individual in the G-computation step, with only the offset varying according to each subject's controller medication use and sex. Although in these cases the traditional regression coefficient and the MSM coefficient are numerically identical, they have different interpretations based on their assumptions (causal vs. noncausal) that preclude direct comparison.

In addition to the theoretical problems associated with comparing marginal estimates with conditional estimates, models 3 and 4 present additional problems that hinder comparison of the traditional regression findings with the G-computation MSM findings. In contrast to the MSM approach we demonstrated, each of these traditional models targeted different parameters. The treatment-covariate interaction present in traditional models 3 and 4 means that there is not a single effect estimate for ozone as estimated by these models. Rather, there is a heterogeneity of ozone effect, with one effect estimate corresponding to the controller medication users and a separate effect estimate corresponding to the nonusers. In contrast, the G-computation effect estimated by MSM_{Q3} and MSM_{Q4} describes the effect estimate with a single value, weighted by the observed frequency of the effect modifier in the data set. In our simple example, standardization using the marginal distribution of the covariates as the standard also yielded the same marginal effect. If the effect modification is of a priori interest, it may actually be desirable to present multiple effect estimates—something that both the traditional approach and G-computation allow. However, if not, this inability to easily report a single effect estimate in the presence of effect modification represents a shortcoming of traditional regression techniques relative to G-computation.

This highlights a key advantage of G-computation: its implementation and interpretation remove the researcher's focus from estimation of nuisance parameters, drawing attention to the parameter(s) of interest. Because G-computation decouples adjustment for cofounding and nuisance effect modification from estimation of parameter(s) of interest, the final MSM estimates only the intercept and the ozone effect. This encourages investigators to define a research question in advance of the analysis stage, clearly distinguishing between nuisance variables and the exposure(s) of interest.

DISCUSSION

Below we present and answer a series of questions that are commonly asked by colleagues as they learn and implement this methodology.

Question 1. G-computation seems very similar to traditional regression techniques; in fact, the first step of G-computation is a traditional regression. Given this similarity, what are the advantages of G-computation?

Answer. As with other causal inference techniques, the G-computation approach decouples the estimation of effects of interest from the estimation of parameters that are not directly related to the research question (e.g., effects of confounders). Additionally, when the effect of exposure on the outcome varies by strata of a third covariate—in other words, interaction exists for the treatment variable—G-computation permits the estimation of a single, marginal effect estimate averaged across the observed distribution of that third variable. The estimation of a single effect may simplify interpretation of exposure effects as compared with multiple effect estimates, depending on the research question.

Question 2. Is G-computation's only application as an estimator of a parameter in an MSM?

Answer. No, G-computation is a general technique that can estimate many parameters, including parameters estimated without the use of an a priori specified parametric model. One point that is seldom explicitly stated in the epidemiologic literature on causal inference methods is that the reliance on counterfactuals for inference does not necessitate reliance on MSMs. For example, one may go through the Q-model step of G-computation and create Y_1 and Y_0. At that point, the investigator has a choice to merely take the difference between Y_1 and Y_0 and then average across the observed distribution of the confounders and report that.

Question 3. What if I misspecify the Q-model?

Answer. This is common in practice; investigators are frequently (and justifiably) worried about model misspecification. As is the case with model specification in traditional regression, a misspecified Q-model will lead to a biased effect estimate. One of the benefits of the G-computation approach is that it allows, but does not require, the researcher to use data-adaptive methodology to obtain the best estimator for the data. Machine-learning techniques are not new but have only recently been gaining traction in epidemiology. One approach, called super learning (25), selects from a set of so-called candidate learners (e.g., random forests, splines, etc.) to compute an estimator for the predicted values that outperforms each of the candidate estimators. If an investigator believes he or she has a regression that fits the data, it can be included as a candidate.

Question 4. Can G-computation be implemented for longitudinal exposure regimens?

Answer. Yes, but implementation of G-computation for longitudinal data is more complex than in the point

treatment setting. An illustration of a Monte Carlo simulation approach (12) and another, more flexible and less computationally intensive algorithm (29) have been published elsewhere. IPTW is used more frequently in the literature for estimation of causal parameters in longitudinal data (4, 5, 13, 30, 31), and the emerging targeted maximum likelihood estimation approach has also been proposed as a suitable alternative for this purpose (32, 33).

We have provided a justification for the implementation of causal inference methods, specifically G-computation, as well as identified some limitations of traditional regression. Comparison of the implementation and interpretation of both methods has been provided. The actual mechanics used in each method are similar; i.e., standard regression software may be utilized. The G-computation approach requires several steps beyond the initial fitting step, but the process is straightforward. The G-computation procedure has some advantages relative to traditional regression, including the decoupling of confounding adjustment and effect estimation, and the causal parameter interpretation. Although G-computation is not necessarily well-suited to every data structure, it is nonetheless an important building block for more sophisticated estimators in the causal inference literature, and researchers working with these techniques benefit from a comprehensive understanding of it.

ACKNOWLEDGMENTS

Author affiliations: Division of Epidemiology, School of Public Health, University of California, Berkeley, Berkeley, California (Jonathan M. Snowden, Kathleen M. Mortimer); and Division of Biostatistics, School of Public Health, University of California, Berkeley, Berkeley, California (Sherri Rose).

This work was supported by the California Air Resources Board (contract 99-322) and the National Heart, Lung, and Blood Institute's Division of Lung Diseases (grant number R01 HL081521).

Conflict of interest: none declared.

REFERENCES

1. Cole SR, Hernán MA, Robins JM, et al. Effect of highly active antiretroviral therapy on time to acquired immunodeficiency syndrome or death using marginal structural models. *Am J Epidemiol*. 2003;158(7):687–694.
2. Petersen ML, Wang Y, van der Laan MJ, et al. Pillbox organizers are associated with improved adherence to HIV antiretroviral therapy and viral suppression: a marginal structural model analysis. *Clin Infect Dis*. 2007;45(7):908–915.
3. Westreich D, MacPhail P, Van Rie A, et al. Effect of pulmonary tuberculosis on mortality in patients receiving HAART. *AIDS*. 2009;23(6):707–715.
4. Bodnar LM, Davidian M, Siega-Riz AM, et al. Marginal structural models for analyzing causal effects of time-dependent treatments: an application in perinatal epidemiology. *Am J Epidemiol*. 2004;159(10):926–934.
5. Brotman RM, Klebanoff MA, Nansel TR, et al. A longitudinal study of vaginal douching and bacterial vaginosis—a marginal structural modeling analysis. *Am J Epidemiol*. 2008;168(2):188–196.
6. Haight T, Tager I, Sternfeld B, et al. Effects of body composition and leisure-time physical activity on transitions in physical functioning in the elderly. *Am J Epidemiol*. 2005;162(7):607–617.
7. Robins J. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Math Modelling*. 1986;7(9–12):1393–1512.
8. Hernán MA. A definition of causal effect for epidemiological research. *J Epidemiol Community Health*. 2004;58(4):265–271.
9. Petersen ML, Wang Y, van der Laan MJ, et al. Assessing the effectiveness of antiretroviral adherence interventions. Using marginal structural models to replicate the findings of randomized controlled trials. *J Acquir Immune Defic Syndr*. 2006;43(suppl 1):S96–S103.
10. Ahern J, Hubbard A, Galea S. Estimating the effects of potential public health interventions on population disease burden: a step-by-step illustration of causal inference methods. *Am J Epidemiol*. 2009;169(9):1140–1147.
11. Moore K, Neugebauer R, Lurmann F, et al. Ambient ozone concentrations cause increased hospitalizations for asthma in children: an 18-year study in Southern California. *Environ Health Perspect*. 2008;116(8):1063–1070.
12. Taubman SL, Robins JM, Mittleman MA, et al. Intervening on risk factors for coronary heart disease: an application of the parametric g-formula. *Int J Epidemiol*. 2009;38(6):1599–1611.
13. Tager IB, Haight T, Sternfeld B, et al. Effects of physical activity and body composition on functional limitation in the elderly: application of the marginal structural model. *Epidemiology*. 2004;15(4):479–493.
14. Cole SR, Hernán MA, Anastos K, et al. Determining the effect of highly active antiretroviral therapy on changes in human immunodeficiency virus type 1 RNA viral load using a marginal structural left-censored mean model. *Am J Epidemiol*. 2007;166(2):219–227.
15. Westreich DJ, Sanne I, Maskew M, et al. Tuberculosis treatment and risk of stavudine substitution in first-line antiretroviral therapy. *Clin Infect Dis*. 2009;48(11):1617–1623.
16. Robins JM, Hernán MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology*. 2000;11(5):550–560.
17. Mortimer KM, Neugebauer R, van der Laan M, et al. An application of model-fitting procedures for marginal structural models. *Am J Epidemiol*. 2005;162(4):382–388.
18. Cole SR, Hernán MA. Constructing inverse probability weights for marginal structural models. *Am J Epidemiol*. 2008;168(6):656–664.
19. Hernán MA, Robins JM. Estimating causal effects from epidemiological data. *J Epidemiol Community Health*. 2006;60(7):578–586.
20. Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology*. 3rd ed. Philadelphia, PA: Lippincott Williams & Wilkins; 2008.
21. van der Laan MJ, Robins JM. *Unified Methods for Censored Longitudinal Data and Causality*. New York, NY: Springer Publishing Company; 2003.
22. Neugebauer R, van der Laan M. Why prefer double robust estimators in causal inference? *J Stat Plan Infer*. 2005;129(1-2):405–426.
23. Ertefaie A, Stephens DA. Comparing approaches to causal inference for longitudinal data: inverse probability weighting

versus propensity scores. *Int J Biostat*. 2010;6(2):Article 14. (doi: 10.2202/1557-4679.1198).
24. van der Laan MJ, Rubin DB. Targeted maximum likelihood learning. *Int J Biostat*. 2006;2(1):Article 11. (doi: 10.2202/1557-4679.1043).
25. van der Laan MJ, Polley EC, Hubbard AE. Super learner. *Stat Appl Genet Mol Biol.* 2007;6(1):Article 25. (doi: 10.2202/1544-6115.1309).
26. Moore KL, Neugebauer RS, van der Laan MJ, et al. Causal inference in epidemiological studies with strong confounding (UC Berkeley Division of Biostatistics Working Paper 255). Berkley, CA: University of California, Berkeley; 2009. (Available at: http://www.bepress.com/ucbbiostat/paper255). (Accessed February 26, 2010).
27. Robins JM. Association, causation, and marginal structural models. *Synthese*. 1999;121:151–179.
28. Pearl J. *Causality: Models, Reasoning, and Inference*. 2nd ed. New York, NY: Cambridge University Press; 2009.
29. Neugebauer R, van der Laan MJ. G-computation estimation for causal inference with complex longitudinal data. *Comput Stat Data An*. 2006;51(3):1676–1697.
30. Garcia-Aymerich J, Lange P, Serra I, et al. Time-dependent confounding in the study of the effects of regular physical activity in chronic obstructive pulmonary disease: an application of the marginal structural model. *Ann Epidemiol*. 2008;18(10):775–783.
31. Lima VD, Harrigan R, Bangsberg DR, et al. The combined effect of modern highly active antiretroviral therapy regimens and adherence on mortality over time. *J Acquir Immune Defic Syndr.* 2009;50(5):529–536.
32. van der Laan M. Targeted maximum likelihood based causal inference: part I. *Int J Biostat*. 2010;6(2):Article 2. (doi: 0.2202/1557-4679.1211).
33. van der Laan M. Targeted maximum likelihood based causal inference: part II. *Int J Biostat*. 2010;6(2):Article 3. (doi: 10.2202/1557-4679.1241).