American Journal of Epidemiology

Vol. 191, No. 2
https://doi.org/10.1093/aje/kwab248
Advance Access publication:
October 13, 2021

Practice of Epidemiology

The Role of the Natural Course in Causal Analysis

Jacqueline E. Rudolph*, Abigail Cartus, Lisa M. Bodnar, Enrique F. Schisterman, and Ashley I. Naimi

* Correspondence to Dr. Jacqueline E. Rudolph, Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, 615 N. Wolfe Street, Baltimore, MD 21205 (e-mail: jrudulp1@jhmi.edu).

Initially submitted April 6, 2021; accepted for publication October 5, 2021.

The average causal effect compares counterfactual outcomes if everyone had been exposed versus if everyone had been unexposed, which can be an unrealistic contrast. Alternatively, we can target effects that compare counterfactual outcomes against the factual outcomes observed in the sample (i.e., we can compare against the natural course). Here, we demonstrate how the natural course can be estimated and used in causal analyses for model validation and effect estimation. Our example is an analysis assessing the impact of taking aspirin on pregnancy, 26 weeks after randomization, in the Effects of Aspirin in Gestation and Reproduction trial (United States, 2006–2012). To validate our models, we estimated the natural course using g-computation and then compared that against the observed incidence of pregnancy. We observed good agreement between the observed and model-based natural courses. We then estimated an effect that compared the natural course against the scenario in which participants assigned to aspirin always complied. If participants had always complied, there would have been 5.0 (95% confidence interval: 2.2, 7.8) more pregnancies per 100 women than was observed. It is good practice to estimate the natural course for model validation when using parametric models, but whether one should estimate a natural course contrast depends on the underlying research questions.

causal inference; g-computation; model validation; natural course; parametric model

Abbreviations: CI, confidence interval; EAGeR, Effects of Aspirin in Gestation and Reproduction.

In recent years, causal inference has become an explicit goal of many epidemiologic analyses. An important first step in such analyses is to clearly define the exposure effect of interest. Investigators typically estimate the average causal effect, which compares counterfactual outcome summaries (e.g., risk, rate, or odds) that would be observed if everyone in the sample were exposed versus if everyone were unexposed. However, there is a growing recognition that the average causal effect does not always correspond to a realistic contrast because, for many clinical or public health interventions, it would be impossible or infeasible to ensure all individuals in a population are fully exposed or fully unexposed (1, 2).

When thinking about policy-relevant effects, an important alternative to estimating the average causal effect is an effect that compares an intervention with the "natural course," which we here call natural-course effects. The natural course represents a factual summary of the outcome under the conditions that naturally occurred in the sample, with no interventions to change how individuals were exposed. In general, such contrasts provide information on the effects of interventions in the population represented by the sample, rather than effects that would be observed if everyone were exposed or unexposed. To give an example, the Effects of Aspirin in Gestation and Reproduction (EAGeR) trial was designed to evaluate whether assignment to low dose aspirin before conception might have an effect on pregnancy-related outcomes (3, 4). Instead of comparing outcomes among women who always took aspirin against women who never took aspirin, we might ask the question, "How would the incidence of pregnancy have differed if all women in the trial randomized to aspirin always took aspirin relative to the compliance that was actually observed in the study?"

When thinking about the role of the natural course in causal analyses, there are 2 related considerations. The first is as we just described, where the natural course refers to

the descriptive outcome summaries typically obtained in the sample. The second is the natural course predicted from the parametric models being used for estimation (5). These are both estimates of the natural course, and each has its own use. In an ideal scenario, in which our models closely reflect the causal mechanisms at work, the model-based natural course would match the observed natural course. On the other hand, large deviations between the observed natural course and the model-based natural course suggest problems with the fitted models. Thus, it is common to see the model-based natural course used for model validation in causal inference analyses. If we sought to do this in the EAGeR example above, we might ask ourselves, "How well can our models replicate the observed incidence of pregnancy among women assigned to aspirin?"

The natural course is not a new concept. However, the causal inference literature that has made use of the natural course has generally not devoted attention to explaining or demonstrating the natural course as a topic of interest unto itself (6–9). Thus, we here explore more fully the role of the natural course in causal analyses. Using the EAGeR trial and the effect of taking aspirin on pregnancy as an example, we walk through how to estimate the model-based natural course, how to use the model-based natural course for model validation, and how to use the natural course as a comparator in effect estimation (while contrasting this effect against the average causal effect). In this illustration, we estimate the natural course effect using g-computation, which is one example of an estimator that can be used to quantify these effects. Last, we comment on how natural-course effects fit into the broader class of effects that can be targeted by epidemiologists and public health practitioners.

METHODS

Data

The EAGeR trial (United States, 2006–2012) was a double-blind, placebo-controlled randomized trial designed to evaluate whether taking daily, preconception low-dose aspirin had an effect on birth outcomes among 1,228 women at high risk for pregnancy loss (3, 4). After assignment to aspirin or placebo, women were followed for up to 6 menstrual cycles if they did not become pregnant, through live birth or pregnancy loss if they did become pregnant, or until they withdrew from the study.

Our outcome of interest was human chorionic gonadotropin–confirmed pregnancy. Pregnancy was determined by either a positive result on a "real-time" urine test of human chorionic gonadotropin done at home or at a study visit or from testing conducted on stored samples after study completion (4). Over the course of follow-up, women either became pregnant, were right-censored when they withdrew from the study, or were administratively censored if they did not become pregnant by 6 months (26 weeks). Compliance with randomized treatment (either aspirin or placebo) was the exposure of interest, as derived from bottle weight measurements (3). We defined our compliance protocol such that a woman was deemed compliant in a given week of follow-up if she took her assigned pill at least 5 out of 7 days (10).

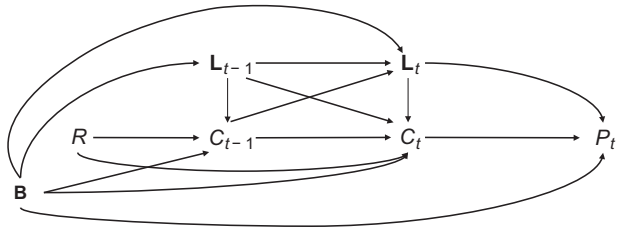

Figure 1. Directed acyclic graph for our trial example (Effects of Aspirin in Gestation and Reproduction, United States, 2006–2012), where P_t is pregnancy in week t of follow-up, C_t is compliance in week t, $\mathbf{L_t}$ is the set of time-varying confounders in week t, R is assigned treatment at baseline, and **B** is the set of baseline confounders.

We considered 4 baseline covariates as potential confounders: age, body mass index, any smoking, and which trial eligibility criteria a woman met. We additionally controlled for 2 important time-varying confounders: reporting any bleeding or any nausea or vomiting in a given week.

Estimating the natural course

In causal analyses, the g-computation algorithm can be used to generate predicted outcomes under a given exposure intervention (6). For example, we can use g-computation to estimate what the incidence of pregnancy would be under an intervention that made all EAGeR women adhere to aspirin. We then contrast these predicted outcomes against a second scenario (say, under an intervention that made all women not adhere to aspirin) to estimate causal effects. We can also use the algorithm to estimate the natural course, which in our example is the incidence of pregnancy observed among women assigned to aspirin. To do this, we might follow these steps: prepare, model key variables, and predict the outcome.

Preparation. Before beginning the analysis, we must specify the causal model. Causal analyses (regardless of analytical methods) require an understanding, informed by background knowledge, of the causal structure of the data relevant to one's research question. This structure is often visualized graphically, for example, using a directed acyclic or single world intervention graph (11, 12). Figure 1 shows the directed acyclic graph we constructed for the EAGeR example. We must also choose the estimand (i.e., the quantity we seek to estimate) that answers our question of interest and the statistical estimator that will allow us to target that estimand. Here, we focused on estimating the natural course and on estimating a causal effect that contrasts the natural course against a counterfactual scenario in which all women assigned to aspirin adhered. In all scenarios, we aimed to estimate the incidence of pregnancy through 26 weeks after randomization. Our measure of incidence was the risk function, obtained via the complement of the Kaplan-Meier estimator (13, 14):

$$F(t) = P(T \leq t) = 1 - \prod_{k<t}\left(1 - \frac{d_k}{n_k}\right),$$

where d_k is the number of events at time k and n_k is the number at risk.

Modeling key variables. After deciding on the causal model and target estimand, we can start building our g-computation algorithm. We begin by modeling the necessary variables as determined by the causal model, regressing each variable on those that affect it. For those who have used g-computation to predict outcomes under a (nonstochastic) exposure intervention, one important difference when estimating the natural course is that we need to include a model for the exposure. When we "set" each participant's exposure to a single value, the probability of the exposure being at that set level is exactly one. However, when estimating the natural course, our goal is to model and replicate the observed probability of being exposed, which is unlikely to be equal to one. Similarly, if one's intervention scenario involves eliminating right-censoring, then estimating the outcome under that intervention does not require a model for censoring. However, there was censoring under the natural course, so a model was required for censoring when estimating the natural course.

In our EAGeR example, we used pooled logistic regression to model time-varying compliance to aspirin, time-varying confounders, censoring, and pregnancy (15). The dependent variables included in each model were those that affect the specific independent variable, as determined by Figure 1. For instance, we included in the model for compliance at time t: compliance at $t-1$, baseline confounders, and time-varying confounders at $t-1$ and t. Given that we did not have data on women after they dropped out of EAGeR, all models were conditional on having not dropped out through time t. In specifying all the models, we sought to be as flexible as the data would allow, to help relax parametric assumptions. This meant including spline terms for continuous variables and fitting our models stratified by randomized treatment arm (aspirin or placebo). The latter step implies a term for interaction between treatment and all variables included in the model.

Predicting the outcome. In simple implementations of g-computation (time-fixed and few confounders), we can use the estimated coefficients from the outcome model and the observed data to predict when or simply whether an event occurred. In more complex settings (many, rare confounders or time-varying data), we first need to take a Monte Carlo resample of the observed data (8). This resample should be as large as is computationally feasible (6).

Here, we took a Monte Carlo resample of 5,000 women and recreated each resampled woman's follow-up visit by visit. For a given week of follow-up, we predicted the value of her time-varying confounders and her exposure (with the exception of the first week, where we used the observed values). We then used those predicted values to determine whether she was censored or had the outcome during that week. If she had the outcome or was censored, the g-computation algorithm terminated; otherwise, we repeated this process in the following week. Alternatively, we can save each individual's predicted outcome probability at each time point, without simulating the outcome directly, which is generally more efficient.

To estimate the natural course stratified by treatment arm, we stratified the data prior to predicting follow-up and took Monte Carlo resamples separately within each treatment arm.

Model validation

To check how well we specified the models, we graphically compared the risk functions for the model-based natural course (based on the outcomes predicted from the g-computation algorithm) and the observed natural course (based on the outcomes actually observed in the data). We compared the cumulative incidence (i.e., risk) of pregnancy by 26 weeks, as well as the median time to pregnancy. In addition to validating the outcome, we compared the distributions of all other after-baseline variables predicted through the g-computation algorithm (compliance, censoring, bleeding, and nausea) with their observed distributions.

In the event that the model-based natural course does not match the observed data well, the models should be altered to increase correspondence between them. This could be accomplished by increasing model flexibility or changing the parametric form. Reasonable agreement is typically determined informally. Regardless, it is important to recognize that our goal in analyses where we estimate the natural course is generally causal and not predictive. Thus, we do not necessarily wish to choose the model(s) that will best predict the observed outcomes. Instead, we wish to model the observed natural course as best we can while still accounting for the proposed causal structure (i.e., controlling for a sufficient set of variables to achieve conditional exchangeability). In other words, we choose the model with the best fit, as long as the model includes the necessary variable set. As in other causal analyses, our causal models should be informed by substantive background knowledge—not by the data alone (16).

It is worth noting that, while this model-validation step is good practice, having good agreement between the observed and model-based natural course is not sufficient to guarantee no model-misspecification bias. Furthermore, for many analyses, estimating the natural course parametrically requires fitting more models than are necessary for effect estimation (e.g., models for exposure and censoring), which can complicate assessments of model misspecification.

Effect estimation

Finally, we can use the natural course as a comparator to assess the impact of a clinical or public health intervention in our sample. Assuming the model-based natural course matches the observed natural course well, one can choose to use either as the comparison value; both have been done in practice (7, 9).

Here, we were interested in how the cumulative incidence of pregnancy might have differed from the observed cumulative incidence if we set all women who had been assigned to aspirin to be compliant across every week of follow-up. We followed the same g-computation algorithm used to estimate the natural course. The major difference was that, instead of modeling exposure and predicting its value at every week of follow-up, we always set exposure to be

Table 1. Baseline Characteristics of Trial Participants, Overall and According to Compliance[a], Effects of Aspirin in Gestation and Reproduction, United States, 2006–2012

Variable	All Participants ($n = 1{,}226$)		Complied ($n = 983$)		Did Not Comply ($n = 243$)	
	No.	%	No.	%	No.	%
Randomized to aspirin	615	50.2	479	48.7	136	56.0
Met new eligibility criteria	678	55.3	521	53.0	157	64.6
Body mass index[b]	26.3 (6.6)		26.1 (6.5)		27.1 (6.9)	
Age, years[b]	28.7 (4.8)		28.9 (4.7)		28.1 (5.1)	
Any smoking in past year	152	12.4	92	9.4	60	24.7

[a] Compliance defined by whether a woman's average compliance across all of follow-up met the protocol of taking the assigned pill 5 out of 7 days of the week.
[b] Values are expressed as mean (standard deviation). Body mass index is calculated as weight (kg)/height (m)^2.

"compliant." We then compared the risk function obtained under this intervention with the observed natural course among women assigned to aspirin. Using counterfactual notation, we sought to estimate:

$$P\left(T^{(\bar{c}=1)} \leq t | R=1\right) - P\left(T \leq t | R=1\right),$$

where T denotes time to event, $R = 1$ denotes randomized to aspirin, and superscript $\bar{c} = 1$ denotes that all women were set to comply across all of follow-up. We additionally estimated the risk of pregnancy if we added a second intervention to eliminate censoring in the sample and compared this risk against the natural course:

$$P\left(T^{(\bar{c}=1,\bar{d}=0)} \leq t | R=1\right) - P\left(T \leq t | R=1\right),$$

where D denotes drop-out and $\bar{d} = 0$ denotes an intervention to allow no censoring across all of follow-up.

To show the differences between the natural-course contrast and a more traditional average causal effect, we also sought to compare the counterfactual outcomes had all women who were assigned to aspirin been compliant versus if they had all been noncompliant:

$$P\left(T^{\bar{c}=1} \leq t | R=1\right) - P\left(T^{\bar{c}=0} \leq t | R=1\right).$$

As with the natural-course effect, we estimated the average causal effect with no intervention on drop-out and with an intervention to eliminate drop-out. We estimated 95% confidence intervals (CIs) using the standard error of the point estimates from 200 bootstrap resamples. All analyses were carried out using R, version 4.0.2 (R Foundation for Statistical Computing, Vienna, Austria), and the code can be found on GitHub (San Francisco, California) (17).

RESULTS

Our analyses included 1,226 EAGeR participants, who contributed over 17,000 person-weeks of follow-up, with a median 12 weeks of follow-up. At baseline, half the sample (50.2%) was randomized to aspirin, and just over half met the new eligibility criteria (55.3%). The mean age was 28.7 years, and the mean body mass index was 26.3 (weight (kg)/meters (m)^2). Few women reported any smoking (12.4%) in the prior year. See Table 1 for these distributions overall and by compliance status (compliance averaged across all weeks of follow-up).

EAGeR participants complied with their assigned treatment in 75.5% of person-weeks (76.3% among those assigned to placebo and 74.6% among those assigned to aspirin). The proportion of women who complied varied by week of follow-up, from a maximum of 96.3% in week 1 (placebo arm: 96.7%; aspirin arm: 95.9%) to a minimum of 45.2% in week 26 (placebo arm: 43.9%; aspirin arm: 46.7%). Bleeding was reported in 64.8% of person-weeks; the proportion of women reporting bleeding varied by week at the start but stabilized at around 75% starting at week 10). Nausea was reported in 34.3% of person-weeks; the proportion of women reporting nausea increased from 20.3% in week 1 to around 40%–45% from week 10 onward. Web Figures 1–4 (available at https://doi.org/10.1093/aje/kwab248) show good agreement between the distributions of all time-varying variables (compliance, bleeding, nausea, and drop-out) as observed and as predicted by the parametric models.

By 26 weeks of follow-up, 773 women became pregnant, with a median event time of 13 weeks, and 116 women withdrew from the study. The cumulative incidence of becoming pregnant by 26 weeks was 67.1%. Stratifying by treatment arm, the observed cumulative incidence by 26 weeks was 70.4% (median time to event = 13 weeks) among those assigned to aspirin and 63.8% (median time to event = 14 weeks) among those assigned placebo. Using g-computation, we estimated that the cumulative incidence of pregnancy by 26 weeks was 66.4%, with a median time to event of 14 weeks. Figure 2 shows close agreement between the g-computation-estimated risk function and the observed natural course. Stratifying by treatment arm, we estimated using g-computation that the cumulative incidence by 26 weeks was 69.8% (median time to event = 13 weeks) among those assigned to aspirin and 63.1% (median time to event = 14 weeks) among those assigned to placebo.

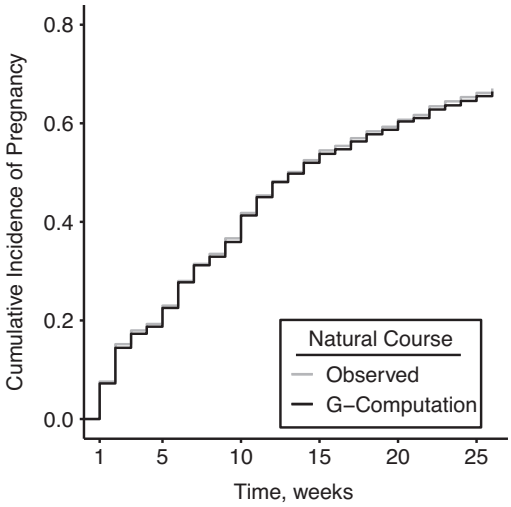

Figure 2. Comparison of the risk functions for the g-computation-estimated natural course and the observed natural course (data from Effects of Aspirin in Gestation and Reproduction, United States, 2006–2012).

Under the intervention where all women assigned to aspirin were compliant across all of follow-up, the estimated cumulative incidence of pregnancy by 26 weeks was 75.4%, with a median time to event of 11 weeks. This implies that, if these participants had always complied, there would have been 5.0 (95% CI: 2.2, 7.8) more pregnancies per 100 women, compared with what was actually observed in the sample. When we additionally eliminated drop-out, the risk difference was 4.9 (95% CI: 2.1, 7.7) per 100 women. Figure 3A illustrates the risk function for this intervention scenario (intervening on compliance and drop-out) relative to the observed natural course, and Figure 3B shows the risk difference (with 95% CI) comparing those risks.

If we had instead been interested in quantifying the average causal effect, comparing the above intervention scenario with the scenario in which all women assigned to aspirin never complied, we would have observed a much larger effect. We estimated that there would have been 25.9 (95% CI: 14.1, 37.7) more pregnancies per 100 women if everyone had been compliant relative to if everyone had been noncompliant across all 26 weeks of follow-up. When we eliminated drop-out, the risk difference was 22.9 (95% CI: 10.9, 34.8) per 100 women. See Figure 4 for a comparison of the risk functions for these 2 intervention scenarios, under elimination of drop-out, and the associated risk difference function.

DISCUSSION

Here, we outlined the concept of the natural course and demonstrated how it can be used for model validation and effect estimation. In the EAGeR example, our parametric models were able to replicate the observed natural course well. We then estimated a small increase in cumulative incidence of pregnancy (5 more pregnancies by 26 weeks, per 100 women) if all participants assigned to aspirin had been compliant relative to the natural course with the observed levels of compliance. In contrast, when we quantified the average causal effect, we estimated that there would be 24 more pregnancies per 100 women if everyone had been compliant compared with if everyone had been noncompliant.

In these data, the natural-course effect and the average causal effect tell very different stories, but which is more relevant will depend on the research question. Natural-course effects are likely to be of interest when estimating the impact of a specific policy or intervention in a population: for example, when examining how many traffic accident fatalities could be prevented by implementing specific road-engineering interventions, relative to doing nothing. Average causal effects are likely to be of interest when estimating the impact of a treatment on patients in a clinical setting, such as in a clinical trial to compare the effects of taking a particular dose of statin medications relative to placebo on average blood cholesterol. We should note, however, that we expect average causal effects to be larger in magnitude than natural-course effects because they compare more extreme scenarios. Specifically, average causal effects compare the always-exposed versus never-exposed scenarios rather than comparing against a scenario in which there was a mixture of exposures. Additionally, in our example, most women in EAGeR were compliant through 26 weeks of follow-up (recall that, in the aspirin arm, women complied in 74.6% of their contributed person-weeks). Thus, the counterfactual scenario in which "everyone complied" was closer to what was observed, while the counterfactual scenario in which "everyone was noncompliant" was further from what was observed.

While not pursued here, in other analyses we might be interested in an intervention scenario that depends directly on the natural course. For example, imagine we wished to examine the effect of setting a policy that would cap exposure levels in our target population. The intervention of interest might then take the form "if an individual has a natural course exposure value above the cap, we set their exposure to be at the cap." Effects of such interventions may be estimated using the extended g-formula (7, 18). The steps for the extended g-formula are similar to those described above for the natural course. At each time point, one predicts the exposure. If the predicted exposure does not meet the intervention criteria (e.g., is above exposure cap), one intervenes on that value. Otherwise, the exposure value remains the predicted value. One then proceeds with predicting the outcome and censoring using the predicted or set exposure. Risk under this intervention can be contrasted against any desired comparator, including the natural course.

It is important to note that we could have used estimators other than g-computation to estimate the natural-course effect in this demonstration. Natural-course effects may be estimated with any estimator that can target the incidence of the outcome under a given intervention, including (but not limited to) inverse probability weighted estimators, targeted minimum loss-based estimators, or incremental effect estimators (19, 20). As we did here with g-computation, these estimators can be used to compare the outcomes under the intervention with the observed outcomes and thus estimate

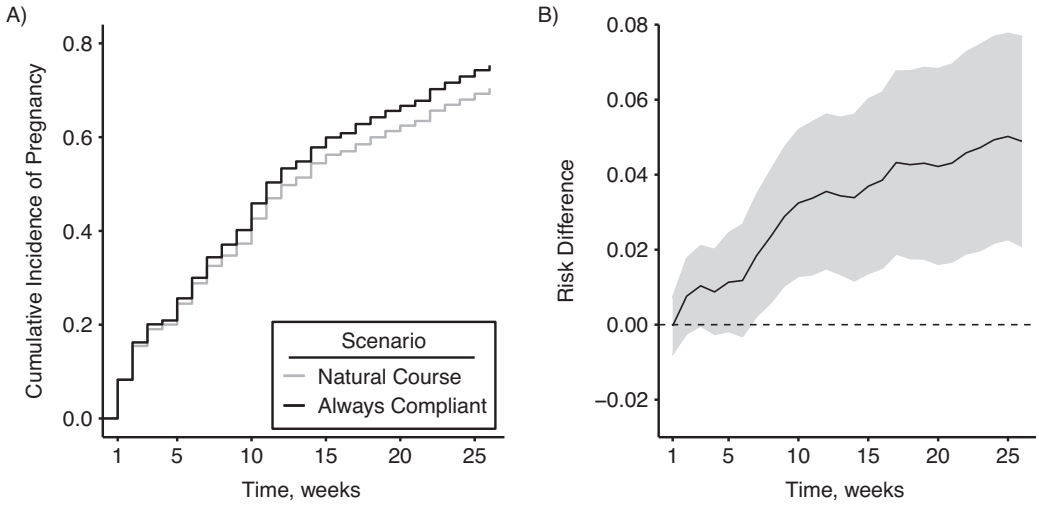

Figure 3. Comparison of the cumulative incidence of pregnancy by a given week of follow-up under the observed natural course among women assigned to aspirin with a scenario in which those women always complied with aspirin treatment (data from Effects of Aspirin in Gestation and Reproduction, United States, 2006–2012). A) The 2 risk curves; B) the risk difference comparing those curves, with the gray band representing the 95% confidence interval.

a natural-course effect. Similarly, to use the natural course for model validation, one could use the models required by the estimator of choice to predict the observed data structure (e.g., using the propensity score model to predict observed exposure) and then compare the prediction with the actual data structure.

The average treatment effect and natural course contrasts discussed here are just 2 of numerous possible estimands one can define for a given analysis. Different estimands include average treatment effects in the treated and untreated, "blip" effects obtained from structural nested models, and various stochastic or population intervention effects (1, 21–23). Many of these estimands can be related to one another. For example, for a time-fixed exposure with a homogeneous effect and prevalence of 50%, one can equate the average treatment effect in the treated and certain natural-course effects. In a time-varying setting, the average treatment effect in the treated, blip effects, and natural-course effects can also be related under a certain set of assumptions (24). In general, it is important for researchers to understand which estimand best answers their substantive research question of interest. Natural-course effects can play an important role,

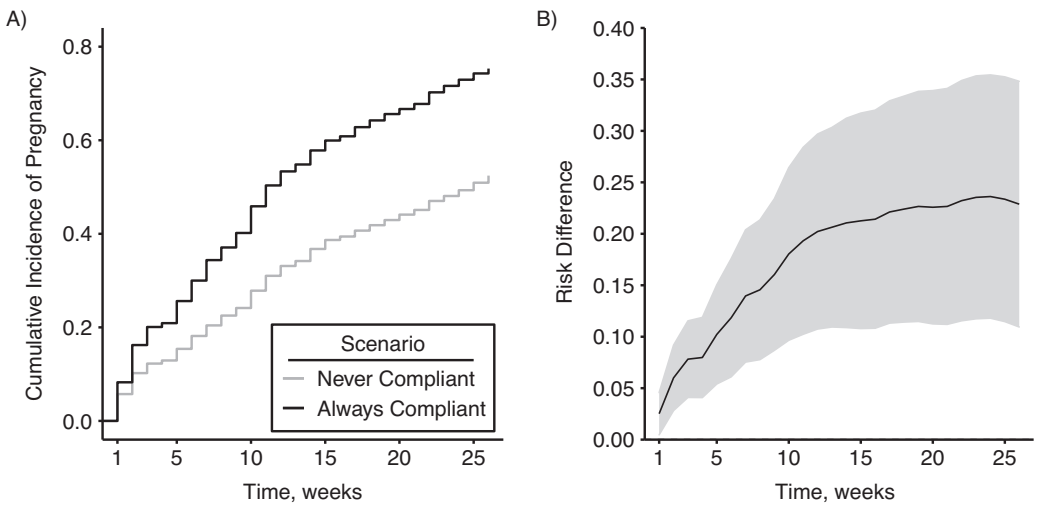

Figure 4. Comparison of the cumulative incidence of pregnancy by a given week of follow-up, under the scenario in which all participants assigned to aspirin always complied, with the scenario in which they never complied (data from Effects of Aspirin in Gestation and Reproduction, United States, 2006–2012). A) The 2 risk curves; B) the risk difference comparing those curves, with the gray band representing the 95% confidence interval.

but other estimands should be targeted when the contrast captured by the natural course effect does not align with research interests.

There are a few limitations of our demonstrations worth mentioning. First, our estimated risk differences should be interpreted cautiously. To interpret our estimates as causal, we would need to meet the standard identification criteria of conditional exchangeability, positivity, and consistency (with no interference) (25). We also need to assume that all of our data were accurately measured and that both our causal and statistical models were correct (26). While our covariate set was identical to that used in the published per-protocol analysis, we cannot confirm that we have controlled for all confounders of the relationship between compliance and pregnancy (10). Additionally, while the model-validation step showed good agreement between the model-based natural course and the observed natural course, that similarity is not sufficient to confirm that there was no statistical model misspecification.

Second, our examples will not be representative of all implementations of g-computation or of all methods that could be used to estimate natural-course effects. How one carries out g-computation and the models that need to be included will always be question- and data-specific. In addition, there are other ways to implement g-computation, such as iterated conditional expectations (27). Third, a limitation of g-computation generally is that we are required to use parametric modeling. While it might seem attractive to use nonparametric machine-learning methods in place of pooled logistic or accelerated failure time models to limit the potential for bias due to model misspecification, this approach is not recommended (28). A doubly-robust alternative to g-computation should be used instead, if incorporating machine-learning approaches is desired (19, 20, 29).

We have explored here how the natural course can be used in causal analyses. To minimize the chance for bias due to model misspecification, we recommend that all analyses using parametric models, such as our g-computation analysis here, compare the estimated natural course with the observed course as a way to check model validity and to adjust the model form (e.g., by increasing its flexibility) as needed based on the results. We further believe that causal estimands that compare an intervention scenario with the natural course could be of great interest to epidemiologists, particularly when assessing the effects of a public health policy. In many scenarios, this estimand could be more relevant than the commonly estimated average causal effect.

ACKNOWLEDGMENTS

Author Affiliations: Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland, United States (Jacqueline E. Rudolph); Department of Epidemiology, School of Public Health, Brown University, Providence, Rhode Island, United States (Abigail Cartus); Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania, United States (Lisa M. Bodnar); Department of Biostatistics, Epidemiology, and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, United States (Enrique F. Schisterman); and Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, Georgia, United States (Ashley I. Naimi).

This work was supported in part by the National Institutes of Health (grant R01 HD093602) and the Intramural Research Program of the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development (contract numbers HHSN267200603423, HHSN267200603424, and HHSN267200603426).

Researchers can apply to access data from the Effects of Aspirin in Gestation and Reproduction trial from the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development (https://dash.nichd.nih.gov/).

Conflict of interest: none declared.

REFERENCES

1. Westreich D. From patients to policy: population intervention effects in epidemiology. *Epidemiology.* 2017; 28(4):525–528.
2. Heckman JJ, Vytlacil E. Policy-relevant treatment effects. *Am Econ Rev.* 2001;91:107–111.
3. Schisterman EF, Silver RM, Lesher LL, et al. Preconception low-dose aspirin and pregnancy outcomes: results from the EAGeR randomised trial. *Lancet.* 2014;384(9937):29–36.
4. Schisterman EF, Silver RM, Perkins NJ, et al. A randomised trial to evaluate the effects of low-dose aspirin in gestation and reproduction: design and baseline characteristics. *Paediatr Perinat Epidemiol.* 2013;27(6):598–609.
5. 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. *Mathematical Modelling.* 1986;7(9):1393–1512.
6. Keil AP, Edwards JK, Richardson DB, et al. The parametric g-formula for time-to-event data: intuition and a worked example. *Epidemiology.* 2014;25(6):889–897.
7. Cole SR, Richardson DB, Chu H, et al. Analysis of occupational asbestos exposure and lung cancer mortality using the g formula. *Am J Epidemiol.* 2013;177(9): 989–996.
8. Westreich D, Cole SR, Young JG, et al. The parametric g-formula to estimate the effect of highly active antiretroviral therapy on incident AIDS or death. *Stat Med.* 2012;31(18): 2000–2009.
9. 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.
10. Naimi AI, Perkins NJ, Sjaarda LA, et al. The effect of preconception-initiated low-dose aspirin on human chorionic gonadotropin-detected pregnancy, pregnancy loss, and live birth: per protocol analysis of a randomized trial. *Ann Intern Med.* 2021;174(5):595–601.
11. Richardson TS, Robins JM. Single world intervention graphs (SWIGs): a unification of the counterfactual and graphical approaches to causality. *Cent Stat Soc Sci Univ Washingt Ser Work Pap.* 2013;(128).

12. Pearl J. Causal diagrams for empirical research. *Biometrika.* 1995;82(4):669–688.
13. Cole SR, Edwards JK, Naimi AI, et al. Hidden imputations and the Kaplan-Meier estimator. *Am J Epidemiol.* 2020; 189(11):1408–1411.
14. Kaplan EL, Meier P. Nonparametric-estimation from incomplete observations. *J Am Stat Assoc.* 1958;53(282): 457–481.
15. D'Agostino RB, Lee ML, Belanger AJ, et al. Relation of pooled logistic regression to time dependent Cox regression analysis: the Framingham Heart Study. *Stat Med.* 1990;9(12): 1501–1515.
16. Robins JM, Wasserman L. On the impossibility of inferring causation from association without background knowledge. In: Glymour C, Cooper G, eds. *Computation, Causation, and Discovery*. Cambridge, MA: AAAI Press/The MIT Press; 1999:305–321.
17. Rudolph JE. Natcourse. https://github.com/jerudolph13/Natcourse. Updated June 1, 2021. Accessed September 7, 2021.
18. Young JG, Herńan MA, Robins JM. Identification, estimation and approximation of risk under interventions that depend on the natural value of treatment using observational data. *Epidemiol Methods.* 2014;3(1):1–19.
19. van der Laan MJ, Gruber S. Targeted minimum loss based estimation of causal effects of multiple time point interventions. *Int J Biostat.* 2012;8(1):21.
20. Kennedy EH. Nonparametric causal effects based on incremental propensity score interventions. *J Am Stat Assoc.* 2019;114(526):645–656.
21. Shpitser I, Pearl J. Effects of treatment on the treated: identification and generalization. [Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence (UAI2009)]. *arXiv.* 2012. https://arxiv.org/abs/1205.2615. Accessed May 28, 2020.
22. Naimi AI, Moodie EE, Auger N, et al. Stochastic mediation contrasts in epidemiologic research: interpregnancy interval and the educational disparity in preterm delivery. *Am J Epidemiol.* 2014;180(4):436–445.
23. Munoz ID, van der Laan M. Population intervention causal effects based on stochastic interventions. *Biometrics.* 2012; 68(2):541–549.
24. Vansteelandt S, Joffe M. Structural nested models and g-estimation: the partially realized promise. *Stat Sci.* 2014; 20(4):707–731.
25. Hernan MA, Robins JM. *Causal Inference: What If*. Boca Raton, FL: Chapman & Hall/CRC; 2020.
26. Keil AP, Mooney SJ, Jonsson Funk M, et al. Resolving an apparent paradox in doubly robust estimators. *Am J Epidemiol.* 2018;187(4):891–892.
27. Petersen ML. Commentary: applying a causal road map in settings with time-dependent confounding. *Epidemiology.* 2014;25(6):898–901.
28. Naimi AI, Mishler A, Kennedy EH. Challenges in obtaining valid causal effect estimates with machine learning algorithms [Published online ahead of print July 15, 2021]. *Am J Epidemiol.* 2021. https://doi.org/10.1093/aje/kwab201.
29. Gruber S, van der Laan M. tmle: an R package for targeted maximum likelihood estimation. *J Stat Soft.* 2012;51(13): 1–35.