American Journal of Epidemiology

Vol. 191, No. 11
https://doi.org/10.1093/aje/kwac136
Advance Access publication:
July 27, 2022

Practice of Epidemiology

Estimation of the Average Causal Effect in Longitudinal Data With Time-Varying Exposures: The Challenge of Nonpositivity and the Impact of Model Flexibility

Jacqueline E. Rudolph*, David Benkeser, Edward H. Kennedy, 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: jacqueline.rudolph@jhu.edu).

Initially submitted October 4, 2021; accepted for publication July 25, 2022.

There are important challenges to the estimation and identification of average causal effects in longitudinal data with time-varying exposures. Here, we discuss the difficulty in meeting the positivity condition. Our motivating example is the per-protocol analysis of the Effects of Aspirin in Gestation and Reproduction (EAGeR) Trial. We estimated the average causal effect comparing the incidence of pregnancy by 26 weeks that would have occurred if all women had been assigned to aspirin and complied versus the incidence if all women had been assigned to placebo and complied. Using flexible targeted minimum loss-based estimation, we estimated a risk difference of 1.27% (95% CI: −9.83, 12.38). Using a less flexible inverse probability weighting approach, the risk difference was 5.77% (95% CI: −1.13, 13.05). However, the cumulative probability of compliance conditional on covariates approached 0 as follow-up accrued, indicating a practical violation of the positivity assumption, which limited our ability to make causal interpretations. The effects of nonpositivity were more apparent when using a more flexible estimator, as indicated by the greater imprecision. When faced with nonpositivity, one can use a flexible approach and be transparent about the uncertainty, use a parametric approach and smooth over gaps in the data, or target a different estimand that will be less vulnerable to positivity violations.

average causal effect; causal inference; longitudinal data; parametric model; positivity

Abbreviations: ACE, average causal effect; CI, confidence interval; EAGeR, Effects of Aspirin in Gestation and Reproduction; IPW, inverse probability weighting; IQR, interquartile range; LTMLE, longitudinal targeted minimum loss-based estimation.

Analyses seeking to estimate the causal effect of a time-varying exposure on an outcome in the presence of time-varying confounding are becomingly increasingly common in epidemiology. In such settings, epidemiologists often seek to estimate average causal effects (ACEs)—for example, the effect contrasting counterfactual outcomes that would have arisen had the entire population always been exposed versus never exposed. However, the complex nature of longitudinal data with time-varying variables can potentially threaten the validity of ACE estimation in several ways.

First, these data typically include time-varying confounders of the exposure-outcome relationship that can be simultaneously mediators or colliders. Controlling for these confounders using traditional regression can result in blocking of a mediating pathway or collider stratification bias (1). In this context, standard regression approaches rarely target quantities that are appropriate for evaluating causal effects. Instead, we must rely on specialized techniques such as inverse probability weighting (IPW), g-computation, or targeted minimum loss-based estimation (2–9).

Second, meeting the identification conditions sufficient to interpret the estimated associations as causal effects can be more challenging than in time-fixed analyses. For example, the positivity condition for the average treatment effect requires sufficient numbers of individuals who are continuously exposed and unexposed in all confounding strata *across all time points* (10, 11). Even if this condition holds at the population level (i.e., there are no structural positivity violations), the likelihood of having enough data for positivity to hold in moderate-sized samples decreases as the number of time points in the analysis increases.

Third, an analysis incorporating time-varying variables may be particularly vulnerable to bias due to violations of the correct model specification assumption because a researcher will need to correctly specify a potentially large number of models capturing the relationships between variables at the current time point as well as past time points.

Here, we delve into the feasibility of estimating ACEs in longitudinal settings with time-varying exposures and draw attention to the repercussions of violations of certain causal and statistical assumptions. Our motivating example is the per-protocol analysis of the Effects of Aspirin in Gestation and Reproduction (EAGeR) Trial (12, 13).

MOTIVATING EXAMPLE

EAGeR was a double-blind, placebo-controlled trial designed to evaluate the effect of taking preconception low-dose aspirin on pregnancy outcomes among 1,228 women at high risk for pregnancy loss (13). Women were followed for up to 6 menstrual cycles and, if they became pregnant, through 36 weeks of gestation. The primary outcome of interest was live birth, with secondary outcomes including occurrence of human chorionic gonadotropin–confirmed pregnancy, pregnancy loss, pregnancy complications, and preterm birth. In the intention-to-treat analysis, the investigators reported similar rates of live birth and pregnancy loss across the trial arms but a small increase in pregnancy incidence among those assigned to aspirin relative to placebo (12).

One drawback to the trial's intention-to-treat results was significant noncompliance with assigned treatment across follow-up (illustrated in Figure 1). The proportion of women who complied dropped from 96.3% in week 1 to 45.2% in week 26. To better describe the biological efficacy of aspirin in the presence of noncompliance, we can consider a per-protocol analysis of the EAGeR Trial, wherein we target the ACE on each pregnancy outcome of being assigned to aspirin and complying across follow-up versus being assigned to placebo and complying (14):

$$E\left[Y(t)^{(r=1,\bar{c}=1)}\right] - E\left[Y(t)^{(r=0,\bar{c}=1)}\right],$$

where $Y(t)$ indicates the outcome at time t, $r = 1$ indicates assignment to aspirin, $r = 0$ indicates assignment to placebo, and $\bar{c} = 1$ indicates compliance from baseline to time t.

Here we focused on the effect of compliance with assigned treatment on incidence of pregnancy through 26 weeks of follow-up (roughly 6 menstrual cycles). Compliance was defined as taking the assigned pill at least 5 out of 7 days per week, as determined by bottle-weight compliance measurements (13, 14). We considered as baseline confounders age, body mass index (weight (kg)/height (m)^2), and smoking; time-varying confounders included weekly reported bleeding or nausea. Women were considered right-censored upon dropout or reaching 26 weeks of follow-up without becoming pregnant. The data structure for our analysis is represented by the nonparametric structural equations given in Table 1.

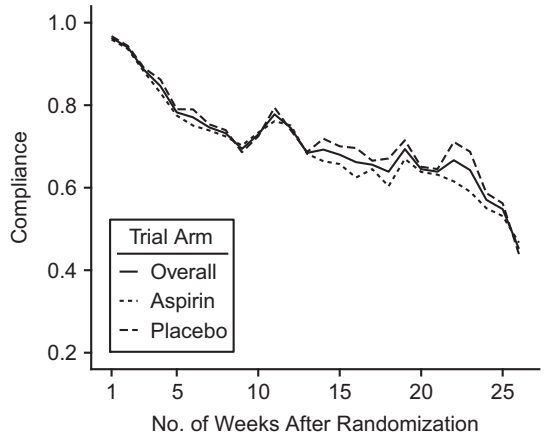

Figure 1. Proportion of women who complied with assigned treatment in a given week of follow-up, overall and by treatment arm, in the Effects of Aspirin in Gestation and Reproduction (EAGeR) Trial, 2007–2011.

ESTIMATION

There are a number of estimators we can use to estimate the per-protocol effect (14–18). For instance, the published per-protocol analysis of the EAGeR Trial used g-computation with parametric models (14). However, we chose an estimator that allowed us to make as few parametric assumptions as possible, namely longitudinal targeted minimum loss-based estimation (LTMLE) (8, 19–21). This estimator accounts for the complexities of our longitudinal data with survival outcomes, right-censoring, time-varying exposures, and time-varying covariates (22). Importantly, LTMLE is double-robust and can be used in conjunction with machine learning methods (23). Incorporating machine learning methods allows us to make fewer assumptions about the functional forms of the relationships between confounders, exposures, and outcomes than traditional parametric models. Single-robust estimators like IPW and g-computation typically require special implementations if one wants to use flexible learning algorithms (24, 25). For this reason, IPW and g-computation are commonly coupled with parametric models, which may render their estimates more susceptible to statistical model misspecification bias (23).

The implementation of LTMLE that is available in the R package "ltmle" is highly flexible for 2 additional reasons. First, it does not smooth the data over time; exposure, censoring, and outcome models are fitted separately at each time point. Second, this implementation does not smooth over treatment; individuals are only retained in the estimation process for as long as they follow the specified treatment trajectory (8). When the intervention is "always compliant with aspirin use," the exposure and outcome models in a given week of follow-up only use the information from individuals who had been assigned to aspirin use and complied in all previous weeks of follow-up. While this approach limits the number of assumptions about the observed data, it reduces the sample size available to the model and can lead

Table 1. Nonparametric Structural Equations Used for a Per-Protocol Analysis of the Effects of Low-Dose Aspirin on Pregnancy Incidence, Effects of Aspirin in Gestation and Reproduction (EAGeR) Trial, 2007–2011

Variable	Notation	Structural Equation
Age	V_1	$V_1 \sim f(U_{V_1})$
Body mass index[a]	V_2	$V_2 \sim f(U_{V_2})$
Smoking	V_3	$V_3 \sim f(U_{V_3})$
Randomization	R	$R \sim f(U_R)$
Bleeding	Z_{1t}	$Z_{1t} \sim f(\overline{C}_{t-1}, \overline{Z}_{1(t-1)}, \mathbf{V}, U_{Z_1})$
Nausea	Z_{2t}	$Z_{2t} \sim f(\overline{C}_{t-1}, \overline{Z}_{2(t-1)}, \mathbf{V}, U_{Z_2})$
Compliance	C_t	$C_t \sim f(\overline{C}_{t-1}, \mathbf{Z}_t, R, \mathbf{V}, U_C)$
Outcome	Y_t	$Y_t \sim f(\overline{C}_t, \mathbf{Z}_t, \mathbf{V}, U_Y)$
No dropout	D_t	$D_t \sim f(\overline{C}_t, \mathbf{Z}_t, \mathbf{V}, U_D)$

[a] Weight (kg)/height (m)^2.

to sparse data cells. An alternative would be to include all person-time, even among those who were not continuously compliant (i.e., an individual could be compliant in one week and noncompliant in the next), in the model. This approach uses more of the data but requires strong modeling assumptions to inform the scenario in which everyone was continuously compliant.

In our implementation of LTMLE, the exposure models regressed compliance at time t against the baseline confounders, current (at time t) and historical (all times $j < t$) values for time-varying confounders, and historical values for exposure. The censoring and outcome models included current exposure, in addition to the above covariates. We fitted these models using SuperLearner, with a library consisting of generalized linear models, generalized additive models, multivariate adaptive regression splines, neural networks, and Bayesian generalized linear models (all using default hyperparameters) (26). We also implemented LMTLE using only main-terms logistic regression. We compared the cumulative incidence of pregnancy under 2 exposure interventions: 1) Set baseline treatment to aspirin and set compliance to 1 (i.e., compliant) across follow-up and 2) set baseline treatment to placebo and set compliance to 1 across follow-up. We obtained 95% confidence intervals (CIs) using the variance from the estimated influence curve (8).

We estimated that there would have been 1.27 (95% CI: −9.83, 12.38) more pregnancies per 100 women by 26 weeks postrandomization if all women had complied with aspirin use versus complied with placebo (and not been censored). This risk difference is closer to the null than the original EAGeR per-protocol analysis using g-computation, which found that there would be 8 more pregnancies per 100 women (95% CI: 4.64, 10.96) under full compliance with aspirin relative to placebo (12, 14). Perhaps the more noticeable finding, though, was the uncertainty implied by our results. In our analysis, the 95% CI width was 22.21, as compared with 6.32 in the original per-protocol analysis (14). Moreover, the imprecision remained in our supplementary analysis using logistic regression instead of SuperLearner, where we estimated a risk difference of 7.51 (95% CI: −8.49, 23.52), indicating that it was not our use of machine learning algorithms that led to the wide CIs.

Our question was then: Why were our results considerably more imprecise than the original published analysis? As we discuss in depth below, the answer lies in a lack of positivity present in our data, which was laid bare by the flexibility of our estimation approach.

THE ROLE OF POSITIVITY

Let us more formally define the positivity condition for the per-protocol analysis of the EAGeR Trial. To identify the additive ACE at 26 weeks of follow-up, there must be a nonzero probability that a woman assigned to aspirin or placebo complied in all 26 weeks of follow-up, conditional on baseline ($\mathbf{V}$) and history of time-varying confounders ($\overline{\mathbf{Z}}_t$) (27, p. 241):

$$P\left(R = r, C_t = c_t | \overline{C}_{t-1} = \overline{c}_{t-1}, \overline{\mathbf{Z}}_t, \mathbf{V}, \overline{D}_{t-1} = \overline{1}\right) > 0, \forall\, (r, \overline{c}_t, \overline{z}_t)\,,$$

where $R = r$ indicates randomization to treatment r, C_t indicates compliance with the protocol at time t, $\overline{C}_t$ indicates history of compliance through $t-1$, and $\overline{D}_{t-1}$ indicates that a woman has not dropped out through $t-1$ (27, p. 159).

Positivity violations can occur for 2 main reasons: 1) Structural violations occur when there exist subgroups in the population who will never be exposed (or unexposed); and 2) practical violations occur when there are few or no individuals exposed within a particular stratum of confounders by chance due to a finite sample size (10). Determining the source of the positivity violation generally requires substantive knowledge and is important for informing decisions on the best solution to overcome the violation. In our example, we were largely concerned with practical, not structural, positivity violations.

If the positivity condition is violated, we cannot interpret the risk difference estimated in the observed data as a causal effect. Moreover, the estimand can become difficult to define because we condition on empty sets, divide by 0-value probabilities, or extrapolate beyond the observed data. A lack of positivity can also lead to wide CIs, just as we observed in our LTMLE per-protocol analysis (10). Fortunately, unlike the exchangeability and consistency conditions, we can empirically evaluate whether there is nonpositivity present in our data by estimating the conditional probability in the above definitions (27, p. 30).

We estimated the probability of compliance given covariates for each woman in EAGeR (i.e., the cumulative propensity scores) using logistic regression models stratified by week of follow-up, regressing compliance in week t against assigned treatment, baseline covariates, compliance in weeks $t-1$ and $t-2$, and time-varying confounders in weeks t, $t-1$, and $t-2$. Similar to our LTMLE analysis, we specified a compliance model that did not smooth across time (because we stratified by week of follow-up). However, this model did smooth across treatment assignment. For the models to converge, we had to use the data from all individuals who were still in the study in a given week of follow-up, regardless of their exposure trajectory to that point. We then estimated the cumulative propensity score at each t by taking the cumulative product of the propensity scores through t. For this diagnostic, we did not consider the probabilities of dropout.

As is visualized in Figure 2, the distribution of cumulative propensity scores steadily shrank toward 0 as weeks of follow-up accrued. For example, in week 1, the average cumulative propensity score was 0.963 (interquartile range (IQR), 0.959–0.987). The average cumulative propensity score then dropped to 0.214 (IQR, 0.045–0.361) by week 10, to 0.025 (IQR, <0.001–0.032) by week 20, and finally to 0.004 (IQR, <0.001–0.002) by week 26. We were likely to have had a practical violation of the positivity condition in our analysis. Accounting for the probability of dropout, as is necessary for this analysis, would only exacerbate the problem.

These findings tell us that the wide 95% CIs observed in our LTMLE analysis were probably a result of the length of follow-up and our choice of the ACE as the target estimand. The more weeks of follow-up we considered, the fewer the observed women who remained continuously compliant with assigned treatment, within strata of the covariates. It is easy to see how this problem—a practical violation of the positivity condition—is not unique to our data set but rather is a general challenge to estimation of an ACE in longitudinal studies with long follow-up periods and time-varying exposures (21).

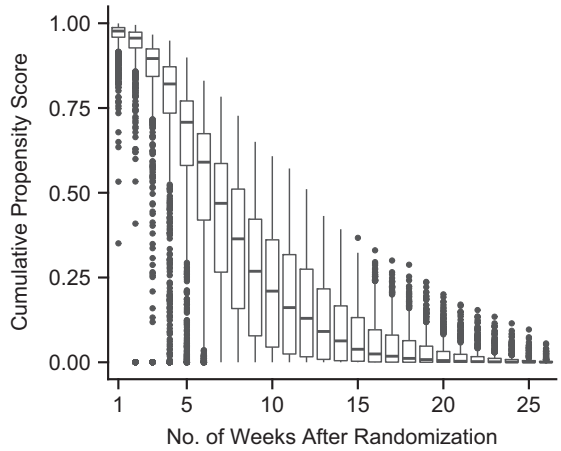

Figure 2. Distribution of the cumulative propensity score (probability of remaining compliant, conditional on covariates), by week of follow-up, in the Effects of Aspirin in Gestation and Reproduction (EAGeR) Trial, 2007–2011. The horizontal line inside the box represents the median cumulative propensity score, the top and bottom of the box represent the 25th and 75th percentiles, and the whiskers extend to the largest (or smallest) value no further than 1.5 times the interquartile range from the top and bottom of the box. The dots show any data points that fell beyond the whiskers.

THE ROLE OF SMOOTHING

There is one more important factor playing a role in the results observed in our LTMLE per-protocol analysis. Specifically, our results were especially vulnerable to practical positivity violations because our implementation stratified by both time and treatment. This stratification reduced the amount of data available for model-fitting, thereby leading to higher variability in estimates of model parameters. This in turn led to more extreme estimated propensity scores. If we had not stratified by time point or treatment, the model parameters would have been shared across time points and treatment arms (a process we refer to as "smoothing"), leading to greater stability in their estimation. This stability, though, comes at the cost of a modeling assumption. If the true underlying propensities varied significantly over time or across treatment arms, the smoothing strategy would have caused bias due to model misspecification. Furthermore, because of the increased stability, the results would not just be incorrect; they would be precisely incorrect (28).

To illustrate the difference between approaches that do and do not smooth, we reestimated the per-protocol effect using 2 different IPW implementations. The first limited the amount of smoothing (mirroring the LTMLE analysis above), while the second introduced smoothing into each step of the analysis (29). We obtained 95% CIs using the 2.5th and 97.5th percentiles of point estimates from 1,000 bootstrap resamples.

Our first IPW estimator, with limited smoothing, used weights at time t defined as

$$w_{rt} = \frac{I(R=r)}{P(R=r)}\prod_{j=1}^{t} \times \frac{I(C_j=1,D_j=1)}{\begin{array}{l}\hat{P}(C_j=1|R=r,\overline{C}_{j-1}=\overline{1},\overline{\mathbf{Z}}_j=\overline{\mathbf{z}}_j,\mathbf{V},\overline{D}_{j-1}=\overline{1}\,)\\ \times\hat{P}(D_j=1|R=r,\overline{C}_j=\overline{1},\overline{\mathbf{Z}}'_j=\overline{\mathbf{z}}'_j,\mathbf{V},\overline{D}_{j-1}=\overline{1})\end{array}}.$$

We estimated the probability of compliance using the logistic regression models described above. The probability of not dropping out by week t was estimated using pooled logistic

regression, conditional on assigned treatment, compliance with the protocol, and baseline and time-varying covariates. The model included indicator terms for week of follow-up and first-order interaction terms between week of follow-up and all other variables in the model. We then stabilized these weights by dividing the above weight by the mean value of the weights:

$$sw_{rt} = \frac{w_{rt}}{\frac{1}{n}\sum_{i=1}^{n} w_{irt}},$$

where i indexes individuals and n is the sample size. While the stabilization method differs, these weights are similar to the inverse-probability-of-censoring weights often used in per-protocol analyses, which censor individuals when they are no longer adherent to the protocol (18).

Finally, we estimated the risk function for each of the 2 treatment assignments using the complements of the respective weighted Kaplan-Meier survival curves, and we took the risk difference at $t = 26$ weeks. This risk estimator inherently includes no smoothing across time, and, because the weights defined above become 0 when an individual stops complying, the outcome model also does not smooth across patterns of compliance.

Using this approach, we estimated that there would have been 3.02 (95% CI: −9.82, 17.83) more pregnancies per 100 women if all women had complied with aspirin use versus all complied with placebo. Comparing this against the results obtained using LTMLE, we see that the risk difference was larger and that the 95% CI was in fact even wider (with a width of 27.63).

Our second IPW estimator, with more smoothing, redefined the stabilized weights (3) as

$$sw_{rt} = \prod_{j=1}^{t} \frac{I\left(C_j = 1\right)\pi_{nj} + \left[1 - I\left(C_j = 1\right)\right]\left(1 - \pi_{nj}\right)}{I\left(C_j = 1\right)\pi_{dj} + \left[1 - I\left(C_j = 1\right)\right]\left(1 - \pi_{dj}\right)} \times \frac{I\left(D_j = 1\right)\hat{P}\left(D_j = 1 | R = r, \overline{C}_j = \overline{c}_j\right)}{\hat{P}\left(D_j = 1 | R = r, \overline{C}_j = \overline{c}_j, \overline{\mathbf{Z}}_j = \overline{\mathbf{z}}_j, \mathbf{V}\right)},$$

where

$$\pi_{nj} = \hat{P}\left(C_j = 1 | R = r, \overline{C}_{j-1} = \overline{c}_{j-1}, \overline{D}_{j-1} = \overline{1}\right)$$

and

$$\pi_{dj} = \hat{P}\left(C_j = 1 | R = r, \overline{C}_{j-1} = \overline{c}_{j-1}, \overline{\mathbf{Z}}_j = \overline{\mathbf{z}}_j, \mathbf{V}, \overline{D}_{j-1} = \overline{1}\right).$$

We estimated π_{nj} and π_{dj} using pooled logistic regression, with the models stratified by treatment arm. Unlike the previous analysis, we did not stratify the compliance models by week of follow-up and instead directly modeled the relationship between time and compliance by including time in our model, thereby introducing smoothing across time. We estimated the probability of not dropping out using main-terms pooled logistic regression. Another important difference to notice is that these weights do not become 0 when an individual stops complying with the protocol; the information from such individuals can then be used when modeling the risk of the outcome, implying more smoothing across patterns of compliance than we allowed in our previous IPW analysis.

In addition to the smoothing in the compliance model, we specified a working marginal structural model for the outcome that made further parametric assumptions (27; 30, p. 264). Here, we parametrically modeled the hazard of the outcome and its dose-response relationship with cumulative exposure using weighted pooled logistic regression:

$$\text{logit}\left(P\left(Y_t = 1 | R, \overline{C}_t\right)\right) = \beta_0 + \beta_1 \overline{A}_t + \beta_2 \overline{N}_t + f(t),$$

where $\overline{A}_t$ was the cumulative average of an indicator term specifying "took aspirin versus did not take aspirin" (which could mean taking placebo or taking nothing), $\overline{N}_t$ was the cumulative average of an indicator term specifying "did not comply versus complied," and $f(t)$ was a function for time (specifically, indicator terms). We defined the cumulative averages as

$$\overline{A}_t = \frac{\sum_{j=1}^{t} I\left(R = 1\right) \times I\left(C_j = 1\right)}{t};$$

$$\overline{N}_t = \frac{\sum_{j=1}^{t} I\left(C_j = 0\right)}{t}.$$

To estimate the per-protocol effect, we used the coefficients from the weighted model to predict the hazard of the outcome at $t = 1, \ldots, 26$ when A_t was always 1 and N_t was always 0 (the "took aspirin at all time points" scenario) and the hazard when A_t was always 0 and N_t was always 0 (the "took placebo at all time points" scenario). We then estimated the risk functions by plugging these hazards into the Kaplan-Meier survival estimator. Note that this second IPW approach, while still being used to estimate the ACE, targets a different statistical estimand than the previous IPW estimator. The statistical estimands are now the model parameters of a marginal structural model with terms for cumulative exposure. While subtle, this in fact adds an additional layer of smoothing.

Using this approach, we estimated that there would have been 5.77 (95% CI: −1.13, 13.05) more pregnancies per 100 women had all women complied with aspirin versus had they complied with placebo through 26 weeks of follow-up. As illustrated in Figure 3, this risk difference was similar to those estimated using the earlier approaches, but the 95% CI width was 14.18—almost half the width seen for the previous IPW estimator.

All analyses described above were carried out using R, version 4.0.4 (R Foundation for Statistical Computing, Vienna, Austria).

CONCLUSIONS

In analyses of longitudinal data, the data structure, targeted estimand, and amount of parametric smoothing can critically affect our ability to meet the positivity condition. Positivity in the time-varying setting is defined relative to the probability of following a specified treatment regimen

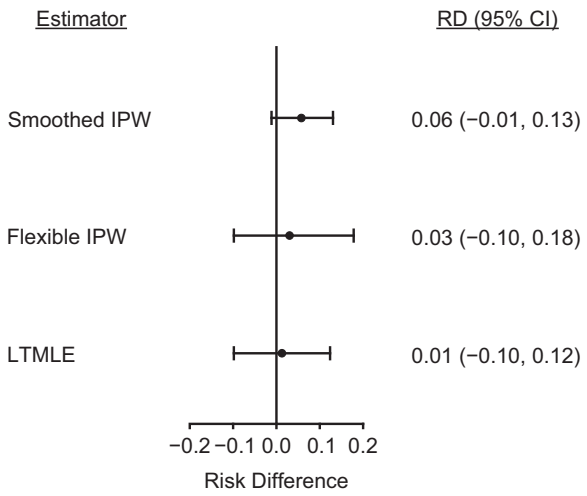

Figure 3. Risk differences (RDs) from 3 estimators (inverse probability weighting (IPW) with high amounts of smoothing, IPW with limited smoothing, and longitudinal targeted minimum loss-based estimation (LTMLE)) used to estimate the per-protocol effect in the Effects of Aspirin in Gestation and Reproduction (EAGeR) Trial, 2007–2011. Bars represent 95% confidence intervals (CIs).

across follow-up, and this probability must be bounded away from 0 for all treatment regimens examined, which we can evaluate by looking at distributions of the cumulative propensity score. As shown here, we may see the effects of violations of this condition manifested in wide 95% CIs, unless our approach is highly parametric. Perhaps more importantly, with a lack of positivity, it becomes risky to interpret an estimated association as a causal effect—regardless of the estimator used.

This discussion was motivated by an attempt to estimate the per-protocol effect of the EAGeR Trial in a robust and flexible manner using LTMLE (8, 21). The risk difference we obtained, though, had wide 95% CIs, and further investigation revealed a clear practical violation of the positivity condition. The imprecision persisted when we changed our estimator to IPW and estimated the weights using parametric models with limited smoothing across treatment and time. We were only able to shrink the size of our CIs by introducing further, strong parametric assumptions. This was the approach taken in the published EAGeR per-protocol analysis; the authors used parametric models in their g-computation algorithm to extrapolate across gaps in the data (14).

We might then ask ourselves whether the wide 95% CIs we obtained when using an estimator with limited smoothing were a flaw or a feature. On the one hand, we might say this uncertainty makes it difficult to interpret or learn from the analysis. From this vantage point, we would view the imprecision as a flaw to be overcome, perhaps by introducing models with stronger parametric assumptions. Smoothing is thus a practical solution for obtaining answers in a complex, longitudinal data structure. On the other hand, we might argue that the wide CIs were simply an accurate reflection of the high uncertainty we should have in our estimates. They remind us of the limited amount of information we can draw from our data set alone. One might also say that introducing strong parametric assumptions to overcome the "flaw" would obscure the underlying lack of positivity.

However, another solution to the issue of nonpositivity in longitudinal studies exists beyond introducing parametric assumptions: targeting a different estimand (10). This could mean estimating an effect through, say, 3 months rather than 6 months of follow-up or estimating an effect within the subset of the sample that was capable of being exposed (a solution that is particularly attractive when there are structural positivity violations). This could also mean not estimating the ACE. One example of a causal estimand that is targetable in the context of time-varying data with long follow-up is the incremental effect (31). This estimand quantifies the effect on the outcome of shifting each individual's propensity score by a specified odds ratio. In our EAGeR example, we could have estimated the incidence of pregnancy if we had doubled each woman's probability of compliance with her assigned treatment at each time point, relative to her observed probability of compliance. The advantage of estimating the incremental propensity score effect over the ACE is that it does not require that we meet the positivity condition for identification. Moreover, like LTMLE, estimation of incremental effects is a doubly robust approach that can be implemented using machine learning algorithms. Other estimands that relax the positivity condition are also available (10, 32).

In the case of practical positivity violations, another solution would be to design a study that has sufficient sample size to estimate the ACE across the entirety of follow-up. Unlike structural positivity violations, practical positivity violations will disappear as the sample size increases. However, if the required sample size is impractical or the analysis is being conducted in existing data, one of the other solutions discussed above may be more appropriate.

While our discussions have focused on the practical positivity violations in our data, the impact of those violations on the estimated results, and solutions for overcoming the violations, we must also remember that positivity is just one of the identifiability conditions necessary for causal inference. For the estimators discussed here, we must also meet the exchangeability and consistency conditions. Conditional exchangeability in the time-varying setting requires that the observed exposure (and censoring) be independent of the counterfactual outcome at each time point, conditional on history of exposure and history of baseline and time-varying confounders. Selection of the confounders to include should be based on one's causal model, but in practice decisions such as the final confounder set and the number of historical values to include in the model are closely linked to some of the trade-offs we discussed here. The positivity assumption is directly linked to the confounder set that is chosen, and one solution for nonpositivity is to trade confounding bias (by not including a confounder in one's model that leads to a positivity violation) for the bias due to practical violations of positivity.

Data with long follow-up periods and time-varying exposures are common in epidemiology, and there are a growing number of methods we can use to estimate the ACE in

these settings. However, it is a mistake to jump straight from choosing the target estimand to estimation to reporting of results, without assessing how well one can meet the causal identifiability conditions. It should be standard practice to examine the potential for positivity violations in applied analyses, by using cumulative propensity score plots like Figure 2 or by examining the distribution of inverse probability weights. Researchers can even go one step further by using methods such as the parametric bootstrap to quantitatively assess the potential for bias due to positivity violations (10). Furthermore, it is crucial that we recognize how inherently difficult it can be to meet the positivity condition in longitudinal analyses estimating the ACE and what solutions are available. In such settings, we must decide whether making strong parametric assumptions is justified to obtain a precise answer, whether exposing one's true level of uncertainty by staying more flexible is instead desired, or whether an alternative estimand would answer one's research question just as well.

ACKNOWLEDGMENTS

Author affiliations: Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland, United States (Jacqueline E. Rudolph); Department of Biostatistics and Bioinformatics, Rollins School of Public Health, Emory University, Atlanta, Georgia, United States (David Benkeser); Department of Statistics and Data Science, Dietrich College of Humanities and Social Sciences, Carnegie Mellon University, Pittsburgh, Pennsylvania, United States (Edward H. Kennedy); 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 funded by National Institutes of Health grants R01-HD093602, R01-CA250851, and U01-DA036297.

Researchers can apply to the National Institute of Child Health and Human Development (https://dash.nichd.nih.gov/) to access data from the EAGeR Trial. Software code for the analysis can be found on GitHub (33).

Conflict of interest: none declared.

REFERENCES

1. Robins JM, Hernán MA. Estimation of the causal effects of time-varying exposures. In: Fitzmaurice G, Davidian M, Verbeke G, et al., eds. *Longitudinal Data Analysis*. New York, NY: Chapman & Hall/CRC Press; 2009:553–597.
2. Naimi AI, Cole SR, Kennedy EH. An introduction to g methods. *Int J Epidemiol.* 2017;46(2):756–762.
3. Hernán MA, Brumback B, Robins JM. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology.* 2000;11(5):561–570.
4. 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 Model.* 1986;7(9):1393–1512.
5. van der Laan M, Coyle J, Hejazi N, et al. *Targeted Learning in R: Causal Data Science With the tlverse Software Ecosystem.* https://tlverse.org/tlverse-handbook/. Published March 25, 2019. Updated August 10, 2022. Accessed August 3, 2020.
6. Gruber S, van der Laan M. tmle: an R package for targeted maximum likelihood estimation. *J Stat Softw.* 2012;51(13):1–35.
7. Glynn AN, Quinn KM. An introduction to the augmented inverse propensity weighted estimator. *Polit Anal.* 2010;18(1):36–56.
8. Lendle SD, Schwab J, Petersen ML, et al. ltmle: an R package implementing targeted minimum loss-based estimation for longitudinal data. *J Stat Softw.* 2017;81(1):1–21.
9. Tran L, Yiannoutsos C, Wools-Kaloustian K, et al. Double robust efficient estimators of longitudinal treatment effects: comparative performance in simulations and a case study. *Int J Biostat.* 2019;15(2):20170054.
10. Petersen ML, Porter KE, Gruber S, et al. Diagnosing and responding to violations in the positivity assumption. *Stat Methods Med Res.* 2012;21(1):31–54.
11. Westreich D, Cole SR. Invited commentary: positivity in practice. *Am J Epidemiol.* 2010;171(6):674–677.
12. 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.
13. 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.
14. 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.
15. Murray EJ, Hernán MA. Adherence adjustment in the Coronary Drug Project: a call for better per-protocol effect estimates in randomized trials. *Clin Trials.* 2016;13(4):372–378.
16. Murray EJ, Hernán MA. Improved adherence adjustment in the Coronary Drug Project. *Trials.* 2018;19(1):158.
17. Hernán MA, Alonso A, Logan R, et al. Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease. *Epidemiology.* 2008;19(6):766–779.
18. Cain LE, Cole SR. Inverse probability-of-censoring weights for the correction of time-varying noncompliance in the effect of randomized highly active antiretroviral therapy on incident AIDS or death. *Stat Med.* 2009;28(12):1725–1738.
19. Petersen ML, Schwab J, Gruber S, et al. Targeted maximum likelihood estimation for dynamic and static longitudinal marginal structural working models. *J Causal Inference.* 2014;2(2):147–185.
20. Schnitzer ME, Moodie EE, van der Laan M, et al. Modeling the impact of hepatitis C viral clearance on end-stage liver disease in an HIV co-infected cohort with targeted maximum likelihood estimation. *Biometrics.* 2013;70(1):144–152.

21. Schomaker M, Luque-Fernandez MA, Leroy V, et al. Using longitudinal targeted maximum likelihood estimation in complex settings with dynamic interventions. *Stat Med.* 2019; 38(24):4888–4911.
22. 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):Article 9.
23. Naimi AI, Mishler AE, Kennedy EH. Challenges in obtaining valid causal effect estimates with machine learning algorithms [published online ahead of print July 15, 2021]. *Am J Epidemiol.* (https://doi.org/10.1093/aje/kwab201).
24. van der Laan M, Benkeser D, Cai W. Causal inference based on undersmoothing the highly adaptive lasso. Presented at the Association for the Advancement of Artificial Intelligence Spring Symposia, Palo Alto, California, March 25–27, 2019.
25. Newey WK, Hsieh F, Robins JM. Twicing kernels and a small bias property of semiparametric estimators. *Econometrica.* 2004;72(3):947–962.
26. van der Laan MJ, Polley EC, Hubbard AE. Super learner. *Stat Appl Genet Mol Biol.* 2007;6:Article 25.
27. Hernán MA, Robins JM. *Causal Inference: What If.* Boca Raton, FL: Chapman & Hall/CRC; 2020.
28. Rudolph JE, Cole SR, Edwards JK. Parametric assumptions equate to hidden observations: comparing the efficiency of nonparametric and parametric models for estimating time to AIDS or death in a cohort of HIV-positive women. *BMC Med Res Methodol.* 2018;18(1):142.
29. Toh S, Hernández-Diaz S, Logan R, et al. Estimating absolute risks in the presence of nonadherence: an application to a follow-up study with baseline randomization. *Epidemiology.* 2010;21(4):528–539.
30. Robins JM, Hernán MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology.* 2000;11(5):550–560.
31. Kennedy EH. Nonparametric causal effects based on incremental propensity score interventions. *J Am Stat Assoc.* 2019;114(526):645–656.
32. Diaz I, Williams N, Hoffman KL, et al. Non-parametric causal effects based on longitudinal modified treatment policies [preprint]. *arXiv*. 2021. (https://doi.org/10.48550/arXiv.2006.01366). Accessed March 26, 2021.
33. Rudolph JE. Time-varying-positivity. https://github.com/jerudolph13/time-varying-positivity. Published March 22, 2021. Accessed June 22, 2022.