


Estimation of the Time-Varying Incremental Effect of Low-dose Aspirin on Incidence of Pregnancy

 Jacqueline E. Rudolph^a, Kwangho Kim^b, Edward H. Kennedy^c, and Ashley I. Naimi^d

Background: In many research settings, the intervention implied by the average causal effect of a time-varying exposure is impractical or unrealistic, and we might instead prefer a more realistic target estimand. Instead of requiring all individuals to be always exposed versus unexposed, incremental effects quantify the impact of merely shifting each individual's probability of being exposed.

Methods: We demonstrate the estimation of incremental effects in the time-varying setting, using data from the Effects of Aspirin in Gestation and Reproduction trial, which assessed the effect of pre-conception low-dose aspirin on pregnancy outcomes. Compliance to aspirin or placebo was summarized weekly and was affected by time-varying confounders such as bleeding or nausea. We sought to estimate what the incidence of pregnancy by 26 weeks postrandomization would have been if we shifted each participant's probability of taking aspirin or placebo each week by odds ratios (OR) between 0.30 and 3.00.

Results: Under no intervention (OR = 1), the incidence of pregnancy was 77% (95% CI: 74%, 80%). Decreasing women's probability of complying with aspirin had little estimated effect on pregnancy incidence. When we increased women's probability of taking aspirin, estimated incidence of pregnancy increased, from 83% (95% confidence interval [CI] = 79%, 87%) for OR = 2 to 89% (95% CI = 84%, 93%) for OR=3. We observed similar results when we shifted women's probability of complying with a placebo.

Conclusions: These results estimated that realistic interventions to increase women's probability of taking aspirin would have yielded little to no impact on the incidence of pregnancy, relative to similar interventions on placebo.

Keywords: Aspirin; Causal inference; Incremental effect; Positivity; Pregnancy

(*Epidemiology* 2023;34: 38–44)

When estimating the effect of an exposure on an outcome in the time-varying setting, epidemiologists routinely target the average causal effect, which compares counterfactual outcomes had one intervened to expose all versus none of a sample across all time points. However, there is a growing recognition that the average causal effect is an unrealistic contrast because, even in time-fixed settings, it is difficult to imagine an intervention that would result in all individuals in a population being exposed or unexposed.¹ The interventions implied by the average causal effect are even more challenging to imagine in the time-varying setting, where we have to assume we can set all individuals to being exposed at all time points. Furthermore, the longer we follow participants, the more unrealistic such intervention becomes.

Fortunately, there exist alternative causal estimands to the average causal effect. One such estimand—the incremental effect—allows us to estimate the effect of shifting each individual's probability of being exposed, instead of intervening on the exact, fixed value of the exposure as in the average causal effect.^{2–4} Estimation of incremental effects has several advantages. First, depending on the exposure of interest and target population, this approach may better reflect the impact of a realistic public health intervention. For example, realistic interventions to increase patient adherence to medication (e.g., daily cellphone notifications) might on average increase patients' likelihood of taking their medication but are unlikely to result in all patients always adhering. Interventions that would lead to perfect adherence (e.g., daily nurse visits) are likely to be costly or unethical. Exposure interventions are also not typically applied uniformly in populations, so a fixed intervention like that imagined by the average causal effect may not be of practical policy interest.³ Second, incremental effects can be estimated using double-robust methods, which enables balancing the tradeoff between bias from the curse of dimensionality and bias from potential statistical model misspecification. Specifically, the approach we use here can achieve optimal statistical properties (root-n convergence)

Submitted March 24, 2022; accepted September 13, 2022

^aDepartment of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD; ^bDepartment of Health Care Policy, Harvard Medical School, Boston, MA; ^cDepartment of Statistics, Carnegie Mellon University, Pittsburgh, PA; and ^dDepartment of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA

This work was supported in part by grants R01-HD093602, R01-CA250851, and U01-DA036297 from the National Institutes of Health, Bethesda, Maryland. Researchers can apply to access data from the EAGeR trial from the National Institute of Child Health and Human Development (<https://dash.nichd.nih.gov/>). The analysis code is available on GitHub (https://github.com/jerudolph13/inc_effect_eager).

The authors report no conflicts of interest.

Correspondence: Jacqueline E. Rudolph, PhD, Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, 615 N. Wolfe St., Baltimore, MD 21205. E-mail: jacqueline.rudolph@jhu.edu

Copyright © 2022 Wolters Kluwer Health, Inc. All rights reserved.

ISSN: 1044-3983/23/341-38-44

DOI: 10.1097/EDE.0000000000001545

regardless of the number of timepoints—even when implemented with flexible machine learning tools.^{4,5} Third, identification of incremental effects does not require meeting the positivity assumption, which makes this an attractive estimator in settings where either structural or random violations of positivity may be likely.⁶

Prior work has described the theory and motivation behind the estimation of incremental effects.^{2,4} Naimi et al.³ demonstrated how to estimate incremental effects in time-fixed settings, using as an example the effect of increasing vegetable intake on the risk of preeclampsia. Here, we demonstrate how to estimate these effects in longitudinal data with a time-varying exposure, time-varying confounding, and drop-out. We build on the applied example in Kim et al.,⁴ by describing in depth how to estimate the effect of taking preconception, low-dose aspirin on the incidence of pregnancy in the Effects of Aspirin in Gestation and Reproduction (EAGeR) trial. This work was motivated by the challenge of analyzing the effects of a time-varying exposure (adherence to aspirin) that suffered from nonpositivity as follow-up accrued.

METHODS

Study Sample

The EAGeR trial was a double-blind trial, designed to investigate whether taking preconception low-dose aspirin had an impact on pregnancy outcomes.^{7,8} The study enrolled 1228 women at high risk for pregnancy loss and randomized women 1:1 to receive 81 mg of aspirin or placebo; all women additionally received 400 mcg of folic acid. Participants were followed up for 6 menstrual cycles if they did not become pregnant and, if they did become pregnant, throughout pregnancy. Women were allowed to leave the study at any point during follow-up. The trial's primary outcome was live birth, with additional outcomes of interest including pregnancy and preterm birth. EAGeR participants provided written informed consent to participate in the trial. Our secondary analysis fell under the approval of the Institutional Review Board of the University of Pittsburgh, who deemed the work not human subjects' research.

Here, we focus on the incidence of pregnancy by 26 weeks of follow-up (approximately six menstrual cycles). Pregnancy was determined by either a positive result on a “real-time” urine pregnancy test carried out at home or at a study visit or from urine testing conducted on stored samples after study completion.⁷ During this time period, 116 (9.5%) of the 1226 women included in our analysis dropped out of the study (two women enrolled in the study were excluded due to all missing data).

The intention-to-treat analysis of the EAGeR trial reported a small increase in rates of pregnancy among those assigned to aspirin relative to placebo.⁸ However, there was noted noncompliance to assigned treatment that increased as follow-up accrued (Figure 1A). Thus, a per-protocol analysis

was conducted that assessed the effect on pregnancy outcomes of being assigned to aspirin and complying in each week of follow-up versus being assigned to placebo and always complying.⁹ Compliance to randomized treatment was determined based on bottle weight measurements and was defined as taking an assigned pill 5 out of 7 days in a given week. This per-protocol analysis reported a small increase in the incidence of pregnancy among those compliant with aspirin, relative to those compliant with placebo; specifically, the estimated risk difference was 7.8% (95% CI = 4.6%, 11%).

Defining the Causal Effect of Interest

We often define the per-protocol effect as the effect of assigning everyone to a given treatment and intervening to ensure they always comply with a specified protocol versus assigning everyone to a comparator treatment and intervening to ensure they always comply.¹⁰

$$E[Y^{r=1, c=1}] - E[Y^{r=0, c=1}]$$

where Y indicates the outcome (here, pregnancy), $R = r$ indicates randomization to treatment r (here, aspirin or placebo), and $\bar{c} = \bar{1}$ indicates compliance to protocol across follow-up (throughout, the overbar denotes past history for a variable). The per-protocol effect is thus a special type of average causal effect that contrasts treatment regimens requiring all participants always be compliant. The exposure in a per-protocol analysis is analogous to the exposure in an observational study, with the difference being that we only need to control for the confounders of the effect of postbaseline C on Y (as there are no confounders of the effect of R on Y). Given that compliance to the protocol is generally a time-varying variable, g-methods such as inverse probability weighting, g-computation, or a double-robust alternative are needed to control for the presence of time-varying confounders that might be affected by past treatment.^{11,12}

However, suppose we thought it unrealistic to model an intervention that would force all women to comply in all weeks of follow-up, but we thought we could instead model an intervention that would increase women's probability of complying. If this were the case, we might be interested in targeting an incremental effect, rather than the average causal effect. Incremental effects in the longitudinal setting have been described in detail elsewhere; here, we provide an overview of the method in the application.⁴

In the EAGeR per-protocol analysis, we could estimate the risk of the outcome through a certain time t (ψ_t) under an intervention that shifts each woman's probability of being compliant with treatment by a specified odds ratio (δ). The usual time-varying propensity score (probability of being exposed) takes the form:

$$\pi_t(h_t) = P(A_t = 1 \mid H_t = h_t, D_{t-1} = 1)$$

where A_t is the exposure at time t ; H_t is the set of all relevant historical variables through t including past exposure,

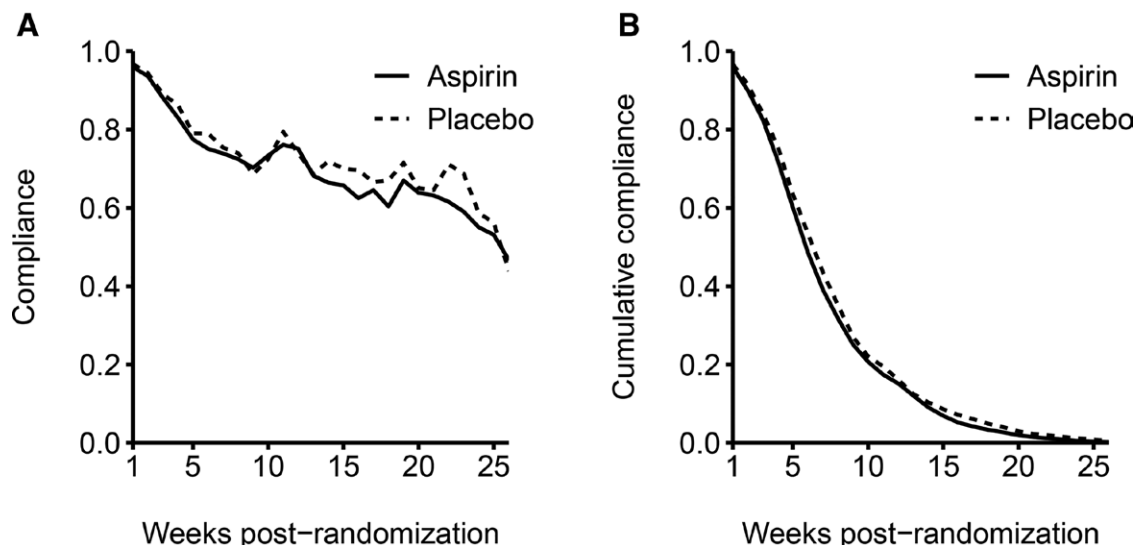


FIGURE 1. Compliance by treatment arm in the EAGeR trial. A, illustrates the observed proportion of EAGeR participants who complied in a given week of follow-up, while (B) illustrates the mean cumulative probability of staying compliant (conditional on covariates), estimated using logistic regression.

time-varying confounders, and baseline confounders; and D_{t-1} is an indicator of not dropping out of the study through $t-1$. We will hereafter define $A_t = 1$ as “complied with aspirin” ($R = 1$ and $C_t = 1$).

To shift π_t by δ , we use:

$$q_t(h_t; \delta, \pi_t) = \frac{\delta \pi_t(h_t)}{\delta \pi_t(h_t) + 1 - \pi_t(h_t)}$$

Rewriting this equation, we can see that δ is interpreted as the odds ratio comparing the odds of q_t to the odds of π_t :

$$\delta = \frac{q_t/(1 - q_t)}{\pi_t/(1 - \pi_t)}$$

We are then interested in estimating what the average counterfactual outcomes (i.e., counterfactual risk of the outcome) would be under exposure to the shifted propensity score q_t across follow-up:

$$\psi_t(\delta) = E[Y^{\mathcal{Q}_t(\delta)}]$$

where $\mathcal{Q}_t(\delta)$ represents random draws from the conditional distributions (q_1, \dots, q_t) that have been shifted by δ . Unlike the average causal effect above, which estimates the effect of a deterministic intervention on the exposure, this estimand represents a dropout-adjusted stochastic intervention on the shifted propensity score q_t . We can then compare the counterfactual risk of the outcome under different values of δ . For example, we can compare the risks under interventions to increase the probability of being exposed ($\delta > 1$) against the observed risk ($\delta = 1$).

The double-robust estimator for longitudinal incremental effects combines information on the shifted

propensity score q_t with the output of an outcome regression ($m_t(h_t, a_t) = m_t(H_t = h_t, A_t = a_t, D_{t+1} = 1)$), as well as the estimated probability for not dropping out ($\omega_t = P(D_{t+1} = 1 | H_t = h_t, A_t = a_t)$). We can estimate $\psi_t(\delta)$, the risk of the outcome at t , by taking $\hat{\psi}_t(\delta) = \sum_{i=1}^n \hat{\psi}_{ti}(\delta)$, where for individual i :

$$\hat{\psi}_{ti}(\delta) = \sum_{s=1}^t \left\{ \frac{\left(\frac{1}{\delta A_s + 1 - A_s} \right) \times \left[\frac{\{\hat{m}_s(H_s, 1) - \hat{m}_s(H_s, 0)\} \delta (A_s - \hat{\pi}_s) \hat{\omega}_s + \delta \hat{\pi}_s + 1 - \hat{\pi}_s}{\delta \hat{\pi}_s + 1 - \hat{\pi}_s} \right]}{\hat{m}_s(H_s, 1) \{\hat{\pi}_s \hat{\omega}_s - A_s D_{s+1}\} + \hat{m}_s(H_s, 0) \{(1 - \hat{\pi}_s) \hat{\omega}_s - (1 - A_s) D_{s+1}\}} \times \prod_{k=1}^s \left(\frac{\delta A_k + 1 - A_k}{\delta \hat{\pi}_k + 1 - \hat{\pi}_k} \times \frac{1}{\hat{\omega}_k} \right) D_k \right\} + \prod_{s=1}^t \left[\frac{\delta A_s + 1 - A_s}{\delta \pi_s + 1 - \pi_s} \times \frac{1}{\omega_s} \right] Y_t D_{t+1}$$

For simplicity, we have removed the subscript i from the righthand side of the equation. The first portion of the equation (labeled M_s) is the pseudo-outcome at time s , whereas the portion labeled U_k is the inverse probability weight for treatment and drop out at k .

Although different in form, $\hat{\psi}_t$ shares a similar structure to the standard double-robust estimator, in that it has an augmentation term for bias correction followed by an inverse-probability-weighted term. The above estimator only requires the second-order product term $(m_t - \hat{m}_t + \pi_t - \hat{\pi}_t)(\pi_s - \hat{\pi}_s + \omega_s - \hat{\omega}_s)$ to be “small enough” for all $s \leq t$. This double robustness property enables us to employ more flexible nonparametric methods for estimating each regression function. Moreover, the estimator $\hat{\psi}_t(\delta)$ can yield important efficiency gains over the classical inverse probability weighted estimator.⁴

Identifying incremental effects

To interpret an estimate obtained in the observed data as the targeted causal effect, meeting certain identification conditions is required. For the standard g-methods, one sufficient set of conditions includes exchangeability, causal consistency, and positivity.¹³ One complication of estimating average causal effects in data with time-varying exposures and long follow-up periods, though, is that violations of the positivity condition are common, particularly random violations due to data sparsity.^{4,6}

In the time-varying setting, positivity requires a nonzero probability of following a given treatment regime (conditional on those variables necessary to achieve exchangeability) across all follow-ups. For example, when estimating the per-protocol effect in EAGeR, we have to assume that the cumulative probability of remaining compliant with assigned treatment (conditional on the baseline and time-varying confounders) in each successive week of follow-up is bounded away from zero (and one). Specifically, we define the positivity condition for the per-protocol effect as:

$$P(A_t = a \mid H_t = h, D_{t-1} = 1) > 0, \forall (a, h)$$

When we estimated these probabilities using logistic regression, we saw that these probabilities approached zero as follow-up accrued, indicating a random positivity violation (Figure 1B).⁶

When positivity is violated, several solutions can be pursued.^{6,14} If the violations are random, one can use parametric models to smooth over the sparse data, at the cost of strong model specification assumptions. This was the solution pursued by the original EAGeR per-protocol analysis.⁹ Another solution, which works regardless of whether the violations are random or structural, is changing the target estimand to one that will not be affected by the positivity violation (e.g., by estimating the average causal effect in the subset of participants with exposure opportunity) or one that does not require the positivity condition.⁶ Incremental effects are one example of the latter type of causal estimand.²⁻⁴

Previous articles on the estimation of incremental effects have demonstrated why this approach does not require positivity for identification.^{3,4} The core idea is that, for individuals and times with $\pi_t = \{0, 1\}$, regardless of whether it was owing to data sparsity or owing to structural violations, the intervention naturally leaves the propensities as is. We do not intervene on them at all. In such cases, an individual's estimated outcome becomes a function of their observed exposure, the specified δ , the estimated probability of not dropping out, and the pseudo-regression functions. Thus, the interpretation of the risk of the outcome under the incremental effect is the average of the outcomes in the entire population when the probability of the exposure is shifted among those with exposure opportunity. The performance of the incremental effect estimator has not been assessed under near violations of positivity when π_t is close to but not exactly equal to zero or one. In such cases, we would likely expect

our 95% confidence intervals to be wide, just as they would be for methods to estimate the average causal effect. It is also worth mentioning that the estimation of incremental effects can be used as a sensitivity analysis for the positivity assumption. When positivity is violated, q_t approaches $I(\pi_t > 0)$ at very large δ (approaching ∞) and approaches $I(\pi_t = 1)$ at very small δ (near 0).²

When controlling for right censoring as we do here, we must also meet the positivity condition for drop out, which is often more likely to hold relative to positivity assumptions for time-dependent treatments.⁴ Finally, it is critically important to note that we must still meet the exchangeability (by controlling for confounding and selection bias due to informative drop out) and consistency conditions to identify the incremental effect.

Estimating Incremental Effects

To estimate incremental effects, we can use the following steps:⁴

- (1) *Sample Splitting.* Split the full sample into k sample splits (here, $k = 2$).^{5,15-17} For a given k , define testing (including all individuals selected into split k) and training (including all individuals not selected into split k) data sets. Sample splitting not only allows us to avoid any restrictions on the complexity of the nuisance estimators so that we can use arbitrarily complex modern machine learning methods but also makes our algorithm easily parallelizable.²
- (2) *Estimate nuisance parameters.* Regress the exposure on historical variables ($A_t \sim H_t$) and the indicator for not dropping out on the exposure and historical variables ($D_t \sim A_t + H_t$) within the training data and use the output from these models to predict $\hat{\pi}_t$ and $\hat{\omega}_t$ in the full sample. Then, use these predicted values and the observed exposure to build cumulative weights:

$$W_t = D_t \prod_{s=1}^t U_s$$

- (3) *Estimate the pseudo regression functions.* Starting at the last time point, let $M_{t+1} = Y_t$ and regress M_{t+1} on exposure and historical variables ($M_{t+1} \sim A_t + H_t$) within the training data and use the model output to predict the outcomes $m_t(H_t, a_t)$ when $a_t = \{0, 1\}$ within the individuals who have not dropped out. Then use these predicted outcomes, $A_t, \hat{\pi}_t, D_t, \hat{\omega}_t, \delta$, to compute the pseudo-outcome at the previous time point M_t . Repeat the process of regression in the training set, prediction in the full sample, and computing M_s , for $s = t - 1, \dots, 1$.
- (4) *Estimate risk.* Within the testing data set, combine the results from the steps above to estimate risk of the outcome for this sample split:

$$\hat{\psi}_t^k(\delta) = \frac{1}{n_k} \sum_{i=1}^{n_k} \hat{\psi}_{it}(\delta)$$

$$\hat{\psi}_{it}^k(\delta) = \sum_{s=1}^t W_s M_s + W_s Y_s D_{s+1}$$

We then repeat these steps for the other sample splits, and the overall estimated risk of the outcome is the average of the estimates obtained within each sample split:

$$\hat{\psi}_t(\delta) = \sum_{j=1}^k \hat{\psi}_t^j(\delta)$$

We should note that all steps are only carried out among those who were “observable” at a given time t , by which we mean they had not dropped out or had the outcome. Because we use sample-splitting estimators that allow arbitrarily complex nuisance estimation, all regressions above can be carried out using not only traditional parametric models (e.g., logistic regression) but also flexible machine learning algorithms or ensemble approaches such as SuperLearner,¹⁸ even when our regression problems are high-dimensional. To obtain 95% confidence intervals (CI), we can estimate the variance of $\psi_t(\delta)$ from the efficient influence function.⁴ When comparing risks under different δ values, 95% CI can be obtained using the delta method.¹⁹

Application to EAGeR

As mentioned above, our outcome of interest in the EAGeR application was the incidence (risk) of pregnancy by 26 weeks of follow-up. Our exposure of interest was compliance with aspirin. We examined δ values that ranged from 0.3 to 3.0. We considered four baseline covariates as potential confounders: age, body mass index, any smoking, and which trial eligibility criteria a woman met. We additionally controlled for two time-varying confounders: reporting of any bleeding or any nausea or vomiting in a given week. This adjustment set matches that used in the published per-protocol analysis.⁹ We ran all regressions using the SuperLearner R package to combine generalized linear models, random forest, and k-nearest neighbors (using default hyperparameters).¹⁸

Because our motivating example was the per-protocol analysis of the EAGeR trial, we repeated the above steps to estimate the incremental effect for an exposure defined as “complied with placebo” ($R = 0$ and $C_t = 1$). We estimated the risk difference comparing the results for the two exposures at specific values of δ and estimated 95% CI using the delta method.

We carried out all analyses using R version 4.1.0 (The R Foundation, Vienna, Austria); code can be found on GitHub (https://github.com/jerudolph13/inc_effect_eager).

RESULTS

As illustrated in Figure 1a, the overall proportion of EAGeR participants who complied with assigned treatment in a given week dropped consistently across follow-up, from a high of 96% in week 1 to a low of 45% in week 26. Compliance was relatively similar for each treatment arm, although there

were weeks in which the proportion of women who complied with aspirin was notably lower than the proportion of women who complied with placebo. For example, in week 23, 69% of women assigned to placebo complied, compared with 59% of women assigned to aspirin (a difference in proportions of 10%). By 26 weeks of follow-up, 773 women became pregnant, with 403 and 370 of those pregnancies occurring among women assigned to aspirin and placebo, respectively. The observed incidence of pregnancy by 26 weeks (not controlling for informative censoring) was 67% among all participants, 70% among those assigned to aspirin, and 64% among those assigned to placebo.

We summarized how the incidence of pregnancy by 26 weeks changed as we shifted women’s probability of complying with aspirin in Figure 2A. When $\delta = 1$, we estimated that the incidence of pregnancy under no intervention on the exposure (but under an intervention to remove censoring) was 77% (95% CI = 74%, 80%). As we decreased women’s probability of complying ($\delta < 1$), the incidence of pregnancy did not meaningfully change, given the width of the confidence intervals. For example, when $\delta = 0.5$, the incidence of pregnancy was 76% (95% CI = 72%, 81%). This implies that the incidence of pregnancy was 0.29% (95% CI = −3.2%, 2.6%) lower when we multiplied women’s odds of complying with aspirin across follow-up by $\delta = 0.5$ compared to when we did not intervene on their exposure ($\delta = 1$). As we increased women’s probability of complying with aspirin ($\delta > 1$), the incidence of pregnancy steadily increased as δ increased. When $\delta = 2$, the incidence of pregnancy was 83% (95% CI = 79%, 87%), and the risk difference relative to $\delta = 1$ was 6.4% (95% CI = 3.8%, 9.0%). When $\delta = 3$, the incidence was 89% (95% CI = 84%, 93%), for a risk difference relative to $\delta = 1$ of 12% (95% CI = 7.8%, 16%). Risks and risk differences for additional δ values are provided in the Table.

We saw a similar pattern in our results when we shifted women’s probability of complying with a placebo, although the increase in pregnancy incidence at $\delta \approx 0.3$ was more pronounced (Figure 2B). When $\delta = 1$ we estimated that the incidence of pregnancy by 26 weeks was 77% (95% CI = 74%, 80%). The risk difference comparing intervening on aspirin to intervening on placebo at $\delta = 1$ was 0.01% (95% CI = −0.62%, 0.64%). When $\delta = 0.5$ the incidence of pregnancy was 78% (95% CI = 74%, 83%). The risk difference relative to $\delta = 1$ was 1.7% (95% CI = −0.75%, 4.2%), and the risk difference comparing aspirin to placebo at $\delta = 0.5$ was −2.0% (95% CI = −6.2%, 2.1%). When $\delta = 2$ the incidence was 85% (95% CI = 81%, 89%), for a risk difference relative to $\delta = 1$ of 8.0% (95% CI = 5.4%, 11%) and a risk difference comparing aspirin to placebo at $\delta = 2$ of −1.5% (95% CI = −5.2%, 2.1%).

DISCUSSION

In this study, we estimated longitudinal incremental effects in the EAGeR trial to assess how an intervention to shift probabilities of complying with aspirin and complying

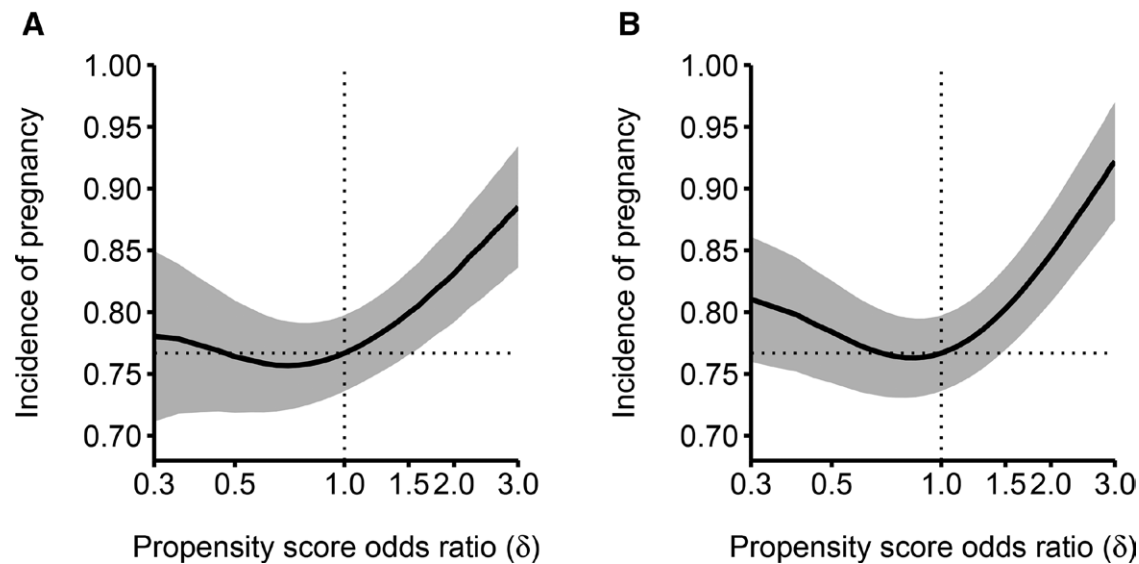


FIGURE 2. Incidence of pregnancy (with 95% confidence intervals) by 26 weeks of follow-up as we shifted women’s probability of complying with (A) aspirin and (B) placebo by δ . The dotted lines highlight the incidence when there was no intervention on the exposure ($\delta = 1$).

TABLE. Incidence of Pregnancy (per 100 women) Under Interventions to Shift EAGeR Participants’ Odds of Complying with Aspirin and Placebo by δ , and the Difference in Incidence Relative to no Intervention ($\delta = 1$)

δ	Compliance to aspirin				Compliance to placebo				Aspirin vs. Placebo	
	Risk	95% CI	RD	95% CI	Risk	95% CI	RD	95% CI	RD	95% CI
0.3	78	71, 85	1.4	−4.4, 7.1	81	76, 86	4.4	0.67, 8.1	−3.0	−10, 4.1
0.5	76	72, 81	−0.29	−3.2, 2.6	78	74, 83	1.7	−0.75, 4.2	−2.0	−6.2, 2.1
0.7	76	72, 79	−1.0	−2.5, 0.42	77	73, 80	−0.07	−1.4, 1.2	−0.93	−3.3, 1.4
1.0	77	74, 80	Ref.	Ref.	77	74, 80	Ref.	Ref.	0.01	−0.62, 0.64
1.5	80	77, 83	3.2	1.7, 4.8	80	77, 84	3.6	2.1, 5.1	−0.34	−2.4, 1.7
2.0	83	79, 87	6.4	3.8, 9.0	85	81, 89	8.0	5.4, 11	−1.5	−5.2, 2.1
2.5	86	82, 91	9.4	6.0, 13	89	84, 93	12	8.7, 15	−2.6	−7.4, 2.2
3.0	89	84, 93	12	7.8, 16	92	87, 97	16	12, 20	−3.7	−9.4, 2.0

δ indicates propensity score odds ratio; CI, confidence interval; RD: risk difference.

with placebo impacted the incidence of pregnancy by 26 weeks of follow-up. In doing so, we sought to provide information that would answer a similar question as a standard per-protocol analysis but in a manner that would not be vulnerable to the random nonpositivity we observed in our data. We estimated that the incidence of pregnancy steadily increased as we increased women’s probability of complying and changed little if we decreased the probability of complying; however, the results were nearly identical regardless of whether women were complying with aspirin or placebo.

The similarity in results seen for the aspirin and placebo exposures seems to indicate that aspirin (even when taken regularly) has little estimated impact on the incidence of pregnancy by 26 weeks in the EAGeR sample. Both the original intention-to-treat and per-protocol analyses reported small increases in incidence of pregnancy for the aspirin arm

relative to placebo arm.^{8,9} Potential reasons for the differing results include that here we simply targeted a different estimand than the original per-protocol analysis and that we used flexible machine learning methods, rather than parametric models. Nonetheless, we saw that incidence of pregnancy steadily increased as we increased women’s probability of complying with either treatment. This finding suggests that the act of complying with and staying involved in the EAGeR trial mattered more for pregnancy incidence than the treatment being taken. The results seen here for both treatment arms demonstrate why the per-protocol effect usually has as its comparator “always complied with placebo.” The goal is to isolate the effect of the drug’s active ingredient. Having a comparator group with the same, perfect level of compliance as the active treatment arm controls for the impact of behaviors related to complying with the trial protocol.¹⁰

Thinking beyond this particular application in EAGeR, there are several important advantages to estimating incremental effects in epidemiologic analyses—many of which we have already mentioned. First, the intervention proposed by incremental effects is interpretable and realistic. The average causal effect imagines one could intervene to make everyone in a given sample exposed and unexposed across all time points, which in some contexts would be infeasible or impossible. In contrast, the intervention implied by incremental effects instead simply increases or decreases everyone's probability of being exposed, which mirrors the sort of impact one could achieve via many public health interventions. Second, this approach does not require the positivity assumption to interpret the risk obtained in observed data as the counterfactual risk of the outcome that would be seen if we intervened to shift everyone's probability of being exposed. This can make an estimation of incremental effects an attractive method to use in analyses where nonpositivity due to either structural or random violations is likely.⁶ Third, incremental effects can be estimated using a double-robust approach implemented with machine learning algorithms, which means it can be less vulnerable to statistical model misspecification bias than other approaches that rely on parametric models (e.g., g-computation, inverse probability weighting, or even double robust estimation of marginal structural models).⁵

There are, however, some caveats in the use of the incremental effect as a causal estimand. In particular, the incremental effect will not answer every research question. Causal estimands and the estimators used to target them should only be chosen if they appropriately answer the scientific question of interest. In particular, incremental effects capture natural increases and decreases in treatment propensity relative to the observational setting; if personalized optimal treatment regimens are of interest, or if treatment assignments really can be finely controlled (e.g., if all in a population can feasibly be treated), then incremental effects may not be the most useful estimand. Furthermore, while we do not need to meet the positivity condition to identify incremental effects, exchangeability and consistency conditions must still be met to interpret one's estimate as causal. As in any other analysis, these assumptions cannot be tested in the data and must be made solely on background knowledge. Finally, if everyone in the population has zero or one probability of treatment, then no method—even estimation of incremental effects—will be able to estimate a meaningful contrast without extrapolation. Our results could also be sensitive to the number of sample splits used. However, this limitation is not unique to our analysis but may be mitigated in larger sample sizes or by making

assumptions about the empirical process conditions, which would impose strong restrictions on the complexity of our nuisance estimators.¹⁷

Despite these limitations, the estimation of incremental effects is a novel approach that provides results that are both highly interpretable and robust. The method described here is thus likely to be attractive for many analyses of epidemiologic studies.

REFERENCES

- Westreich D. From patients to policy: population intervention effects in epidemiology. *Epidemiology*. 2017;28:525–528.
- Kennedy EH. Nonparametric causal effects based on incremental propensity score interventions. *J Am Stat Assoc*. 2019;114:645–656.
- Naimi AI, Rudolph JE, Kennedy EH, et al. Incremental propensity score effects for time-fixed exposures. *Epidemiology*. 2021;32:202–208.
- Kim K, Kennedy EH, Naimi AI. Incremental intervention effects in studies with dropout and many timepoints. *J Causal Inference*. 2021;9:302–344.
- Naimi AI, Mishler A, Kennedy EH. Challenges in obtaining valid causal effect estimates with machine learning algorithms. *Am J Epidemiol*. 2021.
- Petersen ML, Porter KE, Gruber S, et al. Diagnosing and responding to violations in the positivity assumption. *Stat Methods Med Res*. 2012;21:31–54.
- 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;27:598–609.
- Schisterman EF, Silver RM, Leshner LL, et al. Preconception low-dose aspirin and pregnancy outcomes: results from the EAGeR randomised trial. *Lancet*. 2014;384:29–36.
- 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;174:595–601.
- Rudolph JE, Naimi AI, Westreich DJ, et al. Defining and identifying per-protocol effects in randomized trials. *Epidemiology*. 2020;31:692–694.
- 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;7:1393–1512.
- Naimi AI, Cole SR, Kennedy EH. An introduction to g methods. *Int J Epidemiol*. 2017;46:756–762.
- Hernan MA, Robins JM. *Causal Inference: What If*. Boca Raton, FL: Chapman & Hall/CRC; 2020.
- Rudolph JE, Benkeser D, Kennedy EH, et al. Estimation of the average causal effect in longitudinal data with time-varying exposures: the challenge of non-positivity and the impact of model flexibility. [published online ahead of print July 27, 2022]. *Am J Epidemiol*. doi:10.1093/aje/kwac136.
- Zivich PN, Breskin A. Machine learning for causal inference: on the use of cross-fit estimators. *Epidemiology*. 2021;32:393–401.
- Zhong Y, Kennedy EH, Bodnar LM, et al. AIPW: an R package for augmented inverse probability weighted estimation of average causal effects. *Am J Epidemiol*. 2021;190:2690–2699.
- Kennedy EH. Semiparametric theory and empirical processes in causal inference. <https://arxiv.org/abs/1510.04740>. Published 2016. Updated July 22, 2016. Accessed 2020.
- van der Laan MJ, Polley EC, Hubbard AE. Super learner. *Stat Appl Genet Mol Biol*. 2007;6:Article25.
- Cox C. Delta Method. In: *Encyclopedia of Biostatistics*. New York: John Wiley; 1998:1125–1127.