A Practical Example Demonstrating the Utility of Single-world Intervention Graphs

To the Editor:

Causal diagrams[1,2] have become widespread in epidemiologic research. Recently developed single-world intervention graphs explicitly connect the potential outcomes framework of causal inference with causal diagrams.[3] Here, we provide a practical example demonstrating how single-world intervention graphs can supplement traditional causal diagrams.

A randomized controlled trial is conducted to evaluate whether a vaccine (A = 1 if vaccine, 0 if placebo) decreases the risk of disease ($Y = 1$ if disease, 0 otherwise). Individuals are enrolled at baseline, randomized to vaccine or placebo, followed 6 months, and monitored for disease. The vaccine is more likely to result in injection site pain ($W = 1$ if pain, 0 otherwise), and those with pain are more likely to drop out and have unobserved outcomes ($S = 1$ if dropped out, 0 otherwise). Participants with poor (unmeasured) health ($U = 1$ if poor health, 0 otherwise) are more likely to experience pain and get the disease. The scenario is summarized in Figure A.

There is selection bias if we condition on not dropping out ($S = 0$) because the path $A \rightarrow W \leftarrow U \rightarrow Y$ is opened. Stratifying on W does not block this path and may in fact induce more bias. Based on this causal diagram, it is not immediately clear how to identify the causal effect of the vaccine using the observed data (although see references 4, 5, or 6).

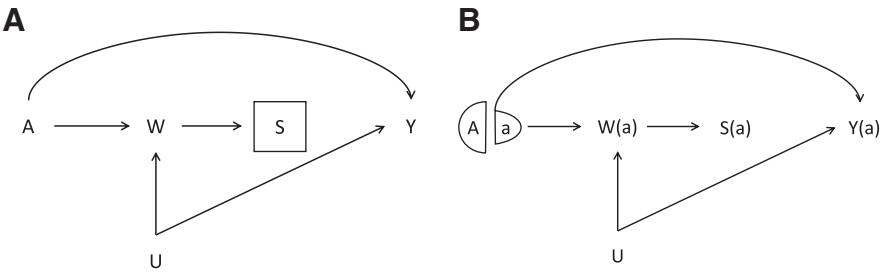

FIGURE. The causal diagram (A) corresponding to the vaccine trial. The single-world intervention template (B), the template used to construct single-world intervention graphs, corresponding to the vaccine trial is constructed by splitting the treatment node of the causal diagram, and replacing all descendants of the assigned treatment with their potential outcomes.

The single-world intervention graph in Figure B, however, clearly displays the independencies necessary to identify the effect of the vaccine from the observed data as follows (here, a variable $X(a)$ represents the value of X had the individual received vaccine level a):

$$
\begin{aligned}
E[Y(a)] &= \sum_w E[Y(a) \mid W(a) = w] P(W(a) = w) \\
&= \sum_w E[Y(a) \mid W(a) = w, S(a) = 0] \\
&\qquad P(W(a) = w) \\
&= \sum_w E[Y(a) \mid W(a) = w, S(a) = 0] \\
&\qquad P(W(a) = w \mid A = a) \\
&= \sum_w E[Y(a) \mid W(a) = w, S(a) = 0, A = a] \\
&\qquad P(W(a) = w \mid A = a) \\
&= \sum_w E[Y \mid W = w, S = 0, A = a] \\
&\qquad P(W = w \mid A = a)
\end{aligned}
$$

The first equality holds by the law of total probability, the second by d-separation of $S(a)$ and $Y(a)$ given $W(a)$, the third by d-separation of $W(a)$ and A, the fourth by d-separation of $Y(a)$ and A given $W(a)$ and $S(a)$, and the last by causal consistency. All components of the final line of the equation, which is Robins' g-formula,[7] can be estimated from observed data. The key insight provided by the single-world intervention graph is that $S(a)$ is independent of $Y(a)$ given $W(a)$, but conditioning on $W(a)$ does not open any paths between A and $Y(a)$.

We conducted a simulation of 1,000,000 individuals for illustration (SAS code is available in the eAppendix; http://links.lww.com/EDE/B306). Individuals were randomly assigned vaccine with probability 0.5 and had probability 0.3 of being in poor health. The probability of injection site pain for healthy individuals was 0.2 if assigned placebo and 0.6 if assigned vaccine. Poor health increased the probability of pain by 0.3. The probability of dropping out was 0.1 for those without pain and 0.9 for those with pain. Finally, the probability of disease was 0.3 for healthy individuals assigned placebo, and it was increased by 0.5 by poor health and decreased by 0.2 by the vaccine.

The true effect of the vaccine on the disease was a 0.20 decrease in risk. The complete case analysis gave a 0.24 decrease in risk. Stratifying on injection site pain worsened the bias, giving a 0.26 decrease in risk. Finally, the g-formula with empirically estimated

Availability of code: The computing code is available in the eAppendix (http://links.lww.com/EDE/B306) for this letter.

Supported by the National Institutes of Health grants R01AI100654 (SRC) and R01AI085073 (MGH).

The authors report no conflicts of interest.

SDC Supplemental digital content is available through direct URL citations in the HTML and PDF versions of this article (www.epidem.com).

expectations and probabilities yielded the true decrease of 0.20.

An anonymous reviewer noted that the derivation above also holds with certain additional edges in the causal diagram, such as $W \rightarrow Y$ or $A \rightarrow S$. These would lead to, respectively, edges $W(A) \rightarrow Y(a)$ or $a \rightarrow S(a)$ in the single-world intervention graph. In the latter case, $S(a)$ is d-separated from $Y(a)$ given $W(a)$ and a, thus $S(a)$ would remain independent of $Y(a)$ conditional on $W(a)$ (Theorem 12 in Richardson and Robins[3]). The reviewer also noted that the derivation fails with unmeasured confounding between A and W or between S and Y.

ACKNOWLEDGMENTS

The authors thank the anonymous reviewer of this letter for their helpful comments.

Alexander Breskin
Stephen R. Cole
Department of Epidemiology
University of North Carolina at Chapel Hill
Chapel Hill, NC
abreskin@unc.edu

Michael G. Hudgens
Department of Biostatistics
University of North Carolina at Chapel Hill
Chapel Hill, NC

REFERENCES

1. Pearl J. Causal diagrams for empirical research. *Biometrika*. 1995;82:669–688.
2. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology*. 1999;10:37–48.
3. 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;128:2013.
4. Bareinboim E, Pearl J. Controlling selection bias in causal inference. *Proc Learn Res*. 2012;22:100–108.
5. Bareinboim E, Tian J, Pearl J. Recovering from selection bias in causal and statistical inference. *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*. 2014;2410–2416.
6. Bareinboim E, Tian J. Recovering causal effects from selection bias. *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*. 2015:3475–3481.
7. Robins J. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Math Model*. 1986;7:1393–1512.

Statistical Power for Trend-in-trend Design

To the Editor:

Unmeasured confounding is often a major concern in observational studies. The recent article by Ji et al.[1] proposed a novel design known as trend-in-trend that, under certain assumptions including a strong time trend in exposure prevalence, provides unbiased estimates of the effects of exposures in the presence of unmeasured confounding. It accomplishes this by examining trends in outcome occurrence as a function of trends in exposure prevalence across strata defined by the cumulative probability of exposure, which models exposure as function of measured baseline variables and effectively stratifies on rate of adoption of the exposure. It therefore extends and improves on studies using calendar time as an instrumental variable,[2–4] eliminating its reliance on the assumption of an absence of a secular trend in the outcome.

Several factors affect the statistical power/detectable alternative of trend-in-trend studies, although we know of no closed-form solution to their estimation. We therefore developed a Monte Carlo simulation approach for estimating statistical power or detectable alternative when planning a trend-in-trend study. This approach requires the investigator to specify six parameters: (1) the type-1 error rate; (2) the probability of a study subject experiencing the study outcome during any study interval; (3) the *c* statistic of the cumulative probability of exposure model[5]; (4) the number of cumulative probability of exposure strata into which the population is divided; (5) the shape of the exposure trend, expressed as a linear or quadratic function of time on log scale; and (6) the desired statistical power or minimum detectable causal odds ratio. The simulation procedure (which has been incorporated into the TrendInTrend package for the R: https://cran.r-project.org/web/packages/TrendInTrend/index.html) provides an estimate for either the statistical power or the minimum detectable odds ratio, whichever was specified in (6) above. The eAppendix (http://links.lww.com/EDE/B308) provides technical information about the simulation procedure.

To illustrate the simulations and assess the influence of the required parameters, we estimated the statistical power of a hypothetical trend-in-trend study under different scenarios. We assumed (1) a type-1 error rate of 5%; (2) a proportion of the population experiencing the outcome during any study interval of 0.007, 0.018, and 0.049; (3) a *c* statistic for the cumulative probability of exposure model of 0.50, 0.60, and 0.75; (4) a number of strata of 5 and 10; (5) a linear time trend in exposure prevalence with slopes (α_t) 0.07 and 0.20 relative percent over the study period (Figure A). We assumed that the study period was divided into 10 time intervals, and generated 500 datasets of size 10,000 (1,000 of whom were ever-exposed) for each scenario.

The Figure B–D displays simulated statistical power over odds ratios ranging from 1 to 2, with the dashed and the dotted lines representing weak ($\alpha_t = 0.07$) and strong ($\alpha_t = 0.20$) time trends in exposure, respectively. The Figure B shows the effect of the *c* statistic of the cumulative probability of exposure model and strength of exposure trend on power when the outcome rate over the entire study period is set to 0.015 and the number of cumulative probability strata is set to 5. The strength of exposure trend has a bigger influence on power than does the *c* statistic, although *c* statistic does have an observable effect. The Figure C shows the influence of the outcome probability when the *c* statistic is

The computer code is available as an R package "TrendInTrend" on CRAN.

Supported by National Science Foundation: NSF grant SES-1260782.

The authors report no conflicts of interest.

SDC Supplemental digital content is available through direct URL citations in the HTML and PDF versions of this article (www.epidem.com).