Published in final edited form as:

Epidemiology. 2018 May ; 29(3): e21–e23. doi:10.1097/EDE.0000000000000803.

Statistical Power for Trend-in-Trend Design

Ashkan Ertefaie,

Department of Biostatistics and Computational Biology, University of Rochester

Dylan Small,

Department of Statistics, University of Pennsylvania, Philadelphia, Pennsylvania

Xinyao Ji,

Department of Statistics, University of Pennsylvania, Philadelphia, Pennsylvania

Charles Leonard, and

Center for Pharmacoepidemiology Research and Training, Center for Clinical Epidemiology and Biostatistics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania

Sean Hennessy

Center for Pharmacoepidemiology Research and Training, Center for Clinical Epidemiology and Biostatistics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania

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^{2,3,4}, 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

Corresponding author: Ashkan Ertefaie, Address: 265 Crittenden Boulevard, Rochester, New York 14642-0630 USA. ashkan_ertefaie@urmc.rochester.edu.

Conflict of interest: None to declare

Code for replication: The computer code is available as an R package “TrendInTrend” on CRAN

study outcome during any study interval; 3) the c-statistic of the cumulative probability of exposure model; 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 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, panel a). We assumed that the study period was divided into 10 time intervals, and generated 500 datasets of size 10,000 (1000 of whom were ever-exposed) for each scenario.

Panels b to d of the Figure display 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. Panel 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. Panel c shows the influence of the outcome probability when the c-statistic is set at 0.75 and the number of cumulative probability of exposure strata is set to 5. As the outcome probability increases, power increases dramatically. Panel d assesses the influence of number of cumulative probability of exposure strata, setting the c statistic to 0.75 and outcome probability to 0.015. It shows that when there is a strong exposure trend (dotted lines), the power function is not sensitive to the number of strata, while for the weak trend (dashed line), increasing the number of strata from 5 (black) to 10 (red) actually reduces the power. Comparing panels b-d suggests that the influence of the outcome probability is stronger than that of the c-statistic or number of cumulative probability of exposure strata. Nonetheless, our results show that when there is a strong exposure prevalence trend, the trend-in-trend design has a reasonable statistical power regardless of outcome rates, c-statistic values and number of strata.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Sources of financial support: National Science Foundation: NSF grant SES-1260782

References

1. Ji X, Small DS, Leonard CE, Hennessy S. The Trend-in-Trend Research Design for Causal Inference. *Epidemiology*. 2017
2. Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables. *Journal of the American statistical Association*. 1996; 91:444–455.
3. Hernan MA, Robins JM. Instruments for causal inference: an epidemiologist's dream. *Epidemiology*. 2006; 17:360–372. [PubMed: 16755261]
4. Ertefaie A, Small DS, Flory JH, Hennessy S. A tutorial on the use of instrumental variables in pharmacoepidemiology. *Pharmacoepidemiology and Drug Safety*. 2017
5. Steyerberg EW, Harrell FE, Borsboom GJ, Eijkemans MJC, Vergouwe Y, Habbema JDF. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *Journal of clinical epidemiology*. 2001; 54(8):774–781. [PubMed: 11470385]

Figure 1.

The trend-in-trend design power analyses. The dashed and dotted lines represent a weak and a strong exposure trend. Panel a: the exposure prevalence over time. Panel b: statistical power for odds ratios of 1-2 for three different values of c-statistic of the cumulative probability of exposure model: 0.75 (green), 0.60 (red), and 0.50 (black). The outcome rate is set to be 0.01. Panel c: statistical power for odds ratios of 1-2 where outcome probability is 0.049 (green), 0.018 (red), and 0.007 (black), where the c-statistic is set to 0.75. Panel d: statistical power for odds ratios of 1-2 for number of cumulative probability of exposure strata = 5 (black) or 10 (red). In all scenarios, the type-I error rate is 0.05, the number of intervals is 10, the number of individuals is 10,000 and the number of replications is 500.