definitions and download the shapefiles that correspond with their data needs.

To illustrate this solution, simulated data were created for a hypothetical disease outbreak in Philadelphia, PA for 2010, including case date, case address, and other covariates. This dataset was built from a random sample of public business addresses in Philadelphia (n = 166), each corresponding to a fictitious disease case. The addresses were then geocoded to latitude and longitude, and serve as the starting point for the strategy described herein. Next, TIGER/Line shapefiles were retrieved for 2010 (the hypothetical outbreak year) for Philadelphia County, Pennsylvania and used for mapping the latitudes and longitudes to their respective census tracts via the *tractLookup* algorithm defined in the eAppendix (http://links.lww.com/EDE/A876). The shapefiles were further used to plot a map (Figure), color-coded according to the number of incident cases per census tract (code also available in eAppendix, http://links.lww.com/EDE/A876).

This strategy has important limitations. First, when analysis datasets span multiple years, census tract geography may change if a decennial census was conducted in the middle of the dataset year. Although the census tract mapping algorithm uses only a given year's shapefiles (eg, first, last, or midpoint in the study period), information bias may be induced as individual cases which may be incorrectly aggregated. Second, the simulated dataset used herein contained public business addresses; for actual case addresses, privacy concerns exist. For example, geocoding through Google Maps Application Programming Interface sends the address to Google's servers. While the address can be sent securely, Google's privacy policy indicates that the address may be stored on their servers. Local institutional review boards need to be consulted before undertaking such an analysis.

In short, this solution allows for the researcher to resolve census-defined regions without need for an external paid utility, and create powerful maps for presentation or publication.

ACKNOWLEDGEMENTS

Thank you to Seth L. Welles for helpful comments in reviewing this manuscript.

Neal D. Goldstein
Department of Epidemiology and Biostatistics
Drexel University School of Public Health
Philadelphia, PA
ng338@drexel.edu

REFERENCES

1. Goldstein ND, Auchincloss AH, Lee BK. A no-cost geocoding strategy using R. *Epidemiology*. 2014;25:311–313.
2. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at: http://www.R-project.org/. Accessed October 13, 2014.
3. Bivand R, Lewin-Koh N. Maptools: tools for reading and handling spatial objects. R package version 0.8–30. Available at: http://CRAN.R-project.org/package=maptools. Accessed October 13, 2014.
4. Stack Overflow. How can I batch geocode street addresses from a csv file in to R? Available at: http://stackoverflow.com/questions/18456075/how-can-i-batch-geocode-street-addresses-from-a-csv-file-in-to-r. Accessed October 13, 2014.
5. U.S. Census Bureau. Geographic Terms and Concepts – Census Tract. Available at: http://www.census.gov/geo/reference/gtc/gtc_ct.html. Accessed October 13, 2014.
6. Rizzardi M, Mohr MS, Merrill DW, Selvin S. Interfacing U.S. census map files with statistical graphics software: application and use in epidemiology. *Stat Med*. 1993;12:1953–1964.
7. U.S. Census Bureau. TIGER Products. Available at: http://www.census.gov/geo/maps-data/data/tiger.html. Accessed October, 11 2014.

SAS Macro for Causal Mediation Analysis with Survival Data

To the Editor:

Mediation analysis investigates the mechanisms that underlie an observed relation between an exposure variable and an outcome variable and examines the role of an intermediate factor, the mediator. Such an analysis can help explain biological and social mechanisms and inform policy making. In 2013, we released a SAS (SAS Institute Inc, Cary, NC) macro for causal mediation analysis for binary, continuous and count outcomes, and binary and continuous mediators,[1] implementing the regression-based results of VanderWeele and Vansteeland,[2,3] and Valeri and VanderWeele[1] for natural direct and indirect effects.[4–6] Here, we have extended the SAS macro for mediation analysis to survival outcomes.

The methods for causal mediation analysis yield valid inferences for natural direct and natural indirect effects under the assumptions that the measured covariates control for confounding of the (1) exposure–outcome, (2) mediator–outcome, and (3) exposure–mediator relations, and (4) that none of the mediator–outcome confounders are affected by the exposure. The methods also require correct specification of the model for the *outcome* given exposure, mediator and confounders, as well as correct specification of the model for the *mediator* given the exposure and confounders. Unlike traditional approaches to mediation, the causal inference methods allow for effect decomposition even in the presence of exposure–mediator interaction. Lange and Hansen[7] and VanderWeele[8] extended these approaches to survival outcomes and continuous mediator. We show that estimators of direct and indirect causal effects derived in Valeri and VanderWeele[1] for the case of binary outcome and binary or continuous mediator are valid with a failure time outcome (see eAppendix, http://links.lww.com/EDE/A877, sections 1 and 4 and VanderWeele[6]). We extend the SAS statistical software in Valeri and VanderWeele[1] to allow for survival outcomes modeled under the Cox proportional hazard or accelerated failure time models (AFT) assuming exponential or Weibull distributions. The causal effects are estimated on the hazard ratio scale if the Cox proportional hazard is employed and on the mean survival ratio scale if the AFT model is chosen. The Cox proportional hazards model mediation results require a

The research was funded by the National Institute of Health grants 5P01CA134294 and R01 ES017876.
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). This content is not peer-reviewed or copy-edited; it is the sole responsibility of the author.

ISSN: 1044-3983/15/2602-0e23
DOI: 10.1097/EDE.0000000000000253

rare outcome at the end of follow-up to be valid; the AFT model does not require this assumption. See eAppendix (http://links.lww.com/EDE/A877) for more details.

In eAppendix (http://links.lww.com/EDE/A877) section 2, we provide the macro user manual, and in section 3, we provide an example of mediation analysis with survival data of colorectal cancer patients from Surveillance, Epidemiology, and End Results Program carried out using the macro.[9] We highlight that the present example is for illustration purposes only, as several of the identification conditions are not met. We might, for example, want to investigate whether socio-economic position, measured by percentage of people living below the poverty line in the county of residence, affects survival of colorectal cancer patients and whether stage at diagnosis may mediate some of this effect. In this example, stage at diagnosis may be a potential mediator of the relation between residing in poor counties and the survival outcome. Here, we briefly present the results of the analyses. An AFT regression assuming exponential distribution is run for survival among colorectal cancer patients on the exposure (county percent below poverty line), adjusting for the mediator (stage at diagnosis: advanced versus non-advanced) and potential confounders (age at diagnosis, year at diagnosis, race-ethnicity, cancer registry). A logistic regression model for stage at diagnosis on the exposure adjusting for potential confounders is fitted. The Table displays the output of the estimated direct and indirect effects at the mean level of the covariates. The analysis, presented in full in the eAppendix (http://links.lww.com/EDE/A877) indicates a negative significant effect of poverty on survival. A positive, significant interaction between stage at diagnosis and poverty is detected. The socio-economic position measure displays a positive and significant association with stage at diagnosis. We find that the mean survival time of individuals living in counties with 30% of the population living below the poverty level is 11% lower than that of individuals living in counties that have no people living below the poverty line. On the mean survival-time ratio scale, the direct effect is 0.94 (95% confidence interval = 0.87–0.99) and the indirect effect is 0.95 (95% confidence interval = 0.93–0.97). Stage at diagnosis is estimated to mediate 42% of the effect of poverty on survival.

To the best of our knowledge, this is the first automated macro software for mediation for survival data allowing for exposure–mediator interactions. We anticipate that this additional feature of our SAS macro will foster the application of causal mediation analysis in life course studies. The macro was developed under SAS 9.3 and is available for download at the authors' websites. Further details are available in the eAppendix (http://links.lww.com/EDE/A877).

ACKNOWLEDGEMENTS

The authors acknowledge Jarvis Chen and Brent Coull for assistance with the Surveillance Epidemiology End Results (SEER) cancer registry data and helpful discussions.

Linda Valeri
Department of Biostatistics
Harvard School of Public Health
Boston, MA
liv839@mail.harvard.edu

Tyler J. VanderWeele
Departments of Biostatistics and Epidemiology
Harvard School of Public Health
Boston, MA

TABLE. Output of Estimated Indirect, Indirect, Total Effects and Proportion Mediated from the SAS Macro

Effect	Estimate	95% Confidence interval	*P* value
cde	0.94344	(0.75896–1.17276)	0.59996
nde	0.93693	(0.87855–0.99919)	0.04717
nie	0.95126	(0.93017–0.97284)	0.00001
Total effect	0.89127	(0.83278–0.95386)	0.00089
Proportion mediated	0.41995	-	-

REFERENCES

1. Valeri L, VanderWeele TJ. Mediation analysis allowing for exposure-mediator interaction and causal interpretation: SAS and SPSS macros. *Psychol Methods*. 2013;18:137–150.
2. VanderWeele TJ, Vansteelandt S. Conceptual issues concerning mediation, interventions and composition. *Stat Interface*. 2009; 2:457–468.
3. Vanderweele TJ, Vansteelandt S. Odds ratios for mediation analysis for a dichotomous outcome. *Am J Epidemiol*. 2010;172: 1339–1348.
4. Pearl J. Direct and indirect effects. In Proceedings of the 17th Conference in Uncertainty in Artificial Intelligence. San Francisco, CA: Morgan Kaufmann Publishers Inc.; 2001, 411–420.
5. Robins JM, Greenland S. Identifiability and exchangeability for direct and indirect effects. *Epidemiology*. 1992;3:143–155.
6. VanderWeele TJ. *Explanation in Causal Inference: Methods for Mediation and Interaction*. Oxford University Press: New York; 2014.
7. Lange T, Hansen JV. Direct and indirect effects in a survival context. *Epidemiology*. 2011;22:575–581.
8. VanderWeele TJ. Causal mediation analysis with survival data. *Epidemiology*. 2011;22:582–585.
9. Surveillance, Epidemiology, and End Results (SEER) Program Research Data (1973–2010), National Cancer Institute, DCCPS, Surveillance Research Program, Surveillance Systems Branch, released April 2013.

Time-Dependent Bias in Hepatitis C Classification

To the Editors:

Implicit assumptions that hepatitis C virus infection antedates that of HIV and that clearance of hepatitis C is rare

Three investigators are the recipients of salary support from the Ontario HIV Treatment Network (JMR, CC) and the Canadian Institutes for Health Research (ANB). The OHTN Cohort Study is supported by the Ontario Ministry of Health and Long-Term Care.
Disclosure: There were no conflicts of interest related to this work.

SDC Supplemental digital content is available through direct URL citations in the HTML and PDF versions of this article (www.epidem.com). This content is not peer-reviewed or copy-edited; it is the sole responsibility of the author.

ISSN: 1044-3983/15/2602-0e24
DOI: 10.1097/EDE.0000000000000255