

Evaluating Public Health Interventions: 7. Let the Subject Matter Choose the Effect Measure: Ratio, Difference, or Something Else Entirely

We define measures of effect used in public health evaluations, which include the risk difference and the risk ratio, the population-attributable risk, years of life lost or gained, disability-adjusted life years, quality-adjusted life years, and the incremental cost-effectiveness ratio. Except for the risk ratio, all of these are absolute effect measures.

For constructing externally generalizable absolute measures of effect when there is superior fit of the multiplicative model, we suggest using the multiplicative model to estimate relative risks, which will often be obtained in simple linear form with no interactions, and then converting these to the desired absolute measure. The externally generalizable absolute measure of effect can be obtained by suitably standardizing to the risk factor distribution of the population to which the results are to be generalized.

External generalizability will often be compromised when absolute measures are computed from study populations with risk factor distributions different from those of the population to whom the results are to be generalized, even when these risk factors are not confounders of the intervention effect. (*Am J Public Health*. 2018;108:73–76. doi:10.2105/AJPH.2017.304105)

Donna Spiegelman, ScD, Polyna Khudyakov, PhD, Molin Wang, PhD, and Tyler J. Vanderweele, PhD

In part 2 of this two-part commentary, the seventh in the series, “Evaluating Public Health Interventions,” we further consider factors informing the choice of the measure for quantifying the effects of public health interventions and the consequences of these choices for study design, analysis, and interpretation. In the first part,¹ we argued that the internal validity and efficiency of results can often be maximized in studies of high-risk populations, populations that can be expected to provide optimal data, and populations relatively homogeneous with respect to risk from other causes. Furthermore, we reviewed evidence for the widespread applicability of the multiplicative model, leading to relative risk (RR) estimates that, if internally valid, are also likely to be generalizable externally when the multiplicative model fits the data and there are no unmeasured effect modifiers. We also noted that the effect measures obtained from the models that best fit the data often will not correspond to the effect measures that are of greatest public health interest, and thus that some conversion may be needed. In this second part of the commentary, we focus more on this issue.

We discuss methods for obtaining externally generalizable absolute measures of effect and estimate years of life gained in the Nurses’ Health Study (NHS) and in the US general population

illustrating this approach. The approach we propose is (1) use the statistical model that best fits the data, often the multiplicative model, to obtain preliminary effect estimates and baseline measures of outcome frequency; (2) transform these estimates to obtain the effect measure that is of public health interest; and (3) standardize this estimate to the risk factor distribution of the target population. The risk factor distribution will, in general, not be estimable in the study population from which the RR estimates were calculated but will need to be obtained from other sources; in some cases, it may not be available at all.

WHICH ADDITIVE EFFECT MEASURE?

In 1959, in the context of the debate about the causal role of cigarette smoking in the etiology of lung cancer, it was argued that the RR provides the best evidence for a causal effect. This was

said to be because, in many settings, the metrics used to assess sensitivity of estimates to unmeasured confounding are more robust for the RR, whereas the risk difference is best for assessing public health impact.^{2,3} Since then, difference measures, that is, the risk or rate difference, have often been described as more relevant for assessing public health impact.^{4–6}

When deciding which of two subpopulations to intervene in to maximize impact, choosing the subpopulation with the largest RR will not necessarily maximize the cases prevented; however, choosing the subpopulation with the largest risk difference will.^{4,6} Many regard this phenomenon, the prevention paradox, as a central public health strategy.⁷ For example, high exposure to vinyl chloride leads to an RR for angiosarcoma of the liver of 36.⁸ Yet, because high exposure to this substance is rare, many more lives would likely be saved by lowering exposure to air pollution even by 10 micrograms

ABOUT THE AUTHORS

Donna Spiegelman is with the departments of Epidemiology, Biostatistics, Nutrition, and Global Health, Harvard T. H. Chan School of Public Health, Boston, MA. Donna Spiegelman and Molin Wang are also with the Channing Division of Network Medicine, Brigham and Women’s Hospital, Boston. Polyna Khudyakov is with the Department of Epidemiology, Harvard T. H. Chan School of Public Health. Molin Wang is with the departments of Epidemiology, Biostatistics, and Medicine, Harvard T. H. Chan School of Public Health. Tyler J. Vanderweele is with the departments of Epidemiology and Biostatistics, Harvard T. H. Chan School of Public Health.

Correspondence should be sent to Donna Spiegelman, 677 Huntington Ave., Kresge Building, Room 802, Boston, MA 02115 (e-mail: stlds@hsph.harvard.edu). Reprints can be ordered at <http://www.ajph.org> by clicking the “Reprints” link.

This commentary was accepted August 25, 2017.

doi: 10.2105/AJPH.2017.304105

per cubic meter because of the widespread nature of this exposure, despite an RR of 1.06 for this increment.

Absolute effect measures include risk and rate differences, population-attributable risks (PARs), years of life lost, incremental cost-effective ratios (ICERs), disability-adjusted life years (DALYs), and quality-adjusted life years (QALYs). Appendix A (available as a supplement to the online version of this article at <http://www.ajph.org>) defines these measures. Public health researchers have offered a wide range of views on which of these parameters is best suited for describing the effect of public health interventions. A former editor-in-chief of *AJPH*, Mary Northridge, asserted, “If the goal is to estimate the amount or proportion of cases of a disease attributable to a given risk factor, or to predict the impact of medical and public health interventions on the health status of a population, then PARs are particularly relevant.”^{9(p1203)}

By contrast, health and development economists favor absolute effect measures related to years of life lost and functions thereof, such as DALYs, QALYs, and ICERs. In 2007, Peter Orszag, former director of the Congressional Budget Office, emphasized the need for increased research and evaluation quantified by the ICER as a means of containing rising health care costs.¹⁰ The World Health Organization, responsible for setting standards and guidelines for health promotion and disease prevention throughout the world, focuses on DALYs and QALYs for intervention effect estimation,¹¹ as does the Global Burden of Disease Project.¹² The Environmental Protection Agency’s publicly available

BENMAP tool for evaluating the impact of various air pollution rollback policies similarly focuses on differences in life years lost and saved (YLL).¹³

As discussed in part 1 of this commentary, when the multiplicative model fits the data, the additive model without interactions will not. Many interactions and other nonlinear representations of model covariates will, thus, often be needed to fit the data adequately on the additive scale. We saw a striking example of this in part 1 of this commentary: the NHS dealing with air pollution exposure’s effect on all-cause mortality risk. Because of this, we advocate using the statistical model that fits the data best (which is often the multiplicative model) and subsequently transforming the estimated coefficients of that model to obtain whatever effect measure (e.g., risk difference, PAR, life years lost) is of public health relevance.

EXTERNALLY GENERALIZABLE RESULTS

PARs, ICERs, YLLs, QALYs, and DALYs are related measures and have common features in terms of their estimation. When the multiplicative model holds, it is often best to estimate the RR using the multiplicative model, because if there are no unmeasured effect modifiers, the effect estimates may be generalizable to other populations. Causal inference ideas and methods, as we began to discuss in a previous commentary in this series,^{14,15} can and should be used for internally valid estimation of the intervention effect, particularly when time-varying confounding is a concern. These

estimates of RR may sometimes be obtained in a “high-risk” epidemiological study. They may also be obtained for an intervention population in which baseline characteristics are not representative of a more general population target to which results are to be applied for policy recommendations and for data-informed decisions about allocation of health care resources. Therefore, externally generalizable functions of this effect estimate need to be put forward.

MODELING FOR EXTERNAL GENERALIZABILITY

When the multiplicative model holds and there are no unmeasured effect modifiers, valid RR estimation can ignore risk factors that are not confounders. In this setting, effect modification will invariably be induced on the absolute scale; this holds even in fully randomized studies, unless the study population has the same joint distribution of these risk factors, which would rarely be the case. Moreover, when converting to these other effect measures (i.e., PARs, ICERs, QALYs, or DALYs), risk factors that are not confounders must be included in the multiplicative model; this is true unless the risk factors’ joint distribution can be assumed to be the same in the study population used for RR estimation and the general population to which the absolute measure is to be applied.

Otherwise, to provide externally generalizable absolute estimates of effect for the identification of optimal public health promotion and disease prevention strategies, data need to be available that accurately characterize the age and risk

factor distributions in the target, or general, population, along with, for some measures, baseline population-based outcome rates. These data may be available, at least in part, from national mortality, morbidity, and disease incidence registers—such as SEER (Surveillance, Epidemiology, and End Results)¹⁶ and the National Death Index¹⁷—and national health surveys of risk factor prevalence—such as US National Health and Nutrition Examination Survey (NHANES)¹⁸ and the National Health Interview Survey (NHIS).¹⁹

In some instances, data from these sources can be pieced together to construct the externally generalizable absolute measures of interest, which must be accompanied by a comprehensive uncertainty measure that takes into account all sources of uncertainty of the estimate. Methods that take this approach have been given for PARs,²⁰ QALYs and DALYs,¹¹ and ICERs.²¹ These methods are closely related to the concepts of standardization, particularly direct standardization.^{22,23} In other instances, the necessary population-based data may be unavailable, particularly for outcomes with a complex risk factor profile and for some countries. In these cases, new designs are needed to allow the production of absolute estimators that are generalizable to these subpopulations and countries.

AIR POLLUTION AND LIVES LOST

Continuing the example introduced in part 1 of this two-part commentary, we consider NHS data used in examining the prospective relationship between fine particulate matter of 2.5

micrograms or less ($PM_{2.5}$) and all-cause mortality; this involved 8617 deaths that occurred among 108 767 nurses between 2000 and 2006. For a moving average exposure of 10 micrograms per cubic meter or greater in $PM_{2.5}$ compared with less, the fully multivariable-adjusted RR was 1.08 (95% confidence interval [CI] = 1.01, 1.15); this is adjusted for age, current smoking status, pack years of cigarettes, US census region, race, family history of myocardial infarction and hypercholesterolemia, body mass index, physical activity, alternative healthy eating index, neighborhood income, mother's occupation, marital status, and husband's education if married.

Under the observed distribution of these risk factors during the follow-up period, we calculated that there is an estimated gain of 0.16 life years per person associated with a reduction in air pollution exposure to below 10 micrograms per cubic meter (95% CI = -0.17, 0.48). This is generalizable to populations with a similar age and risk factor distribution as that of the NHS: White college-educated women born between 1918 and 1945.^{24,25} Even in this quite large and well-annotated study population, although the *P* value for the RR (*P* = .02) indicates evidence for an association, the 95% CIs for the years of life gained because of this reduction in air pollution includes the null; this illustrates a loss of statistical power associated with this more complex effect estimate.

Whether 0.16 life years gained per person is substantial is difficult to assess. One approach for doing so is as follows. If we multiply this value by the number of nurses affected and then divide by the approximate average life expectancy of White females (taken here to be 85 years), we find that 174 lives would have been saved.

Alternatively, 14 790 nurses would have lived an additional year because of a reduction of air pollution exposure to below 10 micrograms per cubic meter among the 85% of nurses so exposed. Let us further consider that among the current US population of approximately 320 000 000, 15% are older than 65 years, 75% are female, 84% have a high school education or more, and 8.6% are exposed to $PM_{2.5}$ levels of more than 10 micrograms per cubic meter. It might reasonably be assumed that these results apply to approximately $0.15 \times 0.75 \times 0.84 \times 0.086 = 0.81\%$ of the US population, or 2 600 640 people.

A 0.16 YLL per person among this segment of the population leads to more than 416 102 years of life saved by reducing $PM_{2.5}$ to below 10 micrograms per cubic meter; with an assumed 85-year life expectancy, this is the equivalent of more than 4895 lives saved because of this exposure reduction in this demographic sector. Whether 416 102 years of life saved is substantial may require comparing with the estimated impacts of reductions of other adverse but prevalent exposures and taking the relative cost of alternative interventions into account.

To construct a YLL that is perhaps more externally generalizable to all US women, rather than those similar in their all-cause mortality risk factor distribution to NHS participants, we reestimated the RR for all-cause mortality in relation to $PM_{2.5}$, adjusting only for smoking status, race, region, and marital status, to obtain an RR of 1.07 (95% CI = 1.01, 1.14). This is nearly identical to the full multivariable-adjusted RR reported above, suggesting no confounding by the other risk

factors included in the full multivariable model. However, the distribution of these risk factors in the NHS is quite different from that in the US general population, which is 67% White rather than 94%, 18% from the Northeast rather than 40%, 8% never married versus 2%, and 9% current smokers versus 12%.

When we standardized the YLL to this distribution of these risk factors, using the NHS distribution for the others, we obtained a YLL of 0.28 (95% CI = 0.06, 0.51), with evidence again for an association (*P* = .01); this is nearly twice as high as the YLL standardized to the NHS risk factor distribution. Again, we emphasize that, despite their striking difference, we obtained these two YLLs under the assumption of no multiplicative modification of the $PM_{2.5}$ effect by any other measured risk factor, as there was no evidence for this in the data for the measured risk factors (although there may still be some unmeasured modifiers).

This analysis could be further refined by standardizing to the full joint distribution of all risk factors in the multivariable RR model. However, NHIS appears to be missing four risk factors (family history of heart disease, hypercholesterolemia, mother's occupation, and healthy eating index), whereas the NHANES is missing two (family history of heart disease and mother's occupation). In addition, because these population-representative risk factor surveys do not jointly estimate the prevalence of risk factors in the same group at once, only a partial joint distribution of risk factors is available from the very best and most representative data on these factors.

We presented these results to illustrate the concepts of modeling for external generalizability and public health impact. They should not be taken as firm

estimates of the impact of air pollution reductions on lives gained in the United States.

CONCLUSIONS

In our review of these various absolute measures of intervention effects used by policymakers and decision-makers, there is quite a bit of room for more methodological work. Existing methods tend to use simple and likely unrealistic parametric models to describe survival distributions and may not fully take into account all sources of uncertainty. Because surveys may lack data on all, or even most, risk factors for an outcome, efficient study designs need to be developed that allow estimation of both the relative effect measure and the joint distribution of target population outcome risk factors to construct a statistically efficient and externally generalizable result. New developments in targeted maximum likelihood estimation may assist in modeling efforts.²⁶ Guidance is needed on the impact of varying degrees of misalignment of the baseline risk factor distribution of the study population with the target population on the external validity of absolute effect measures. It is possible that circumstances can be delineated in which this concern sometimes turns out to be substantively relatively unimportant.

Absolute effect estimates are critical for evidence-based public health decision-making. However, the external generalizability of absolute measures thus far presented needs to be carefully examined and, in many cases, will not hold when these measures are calculated under methods and data presently available. New study designs need to be developed that allow internally valid RR estimation alongside

externally valid absolute effect estimation. *AJPH*

CONTRIBUTORS

D. Spiegelman conceptualized the article and wrote the article and the supplemental materials. D. Spiegelman, P. Khudyakov, and M. Wang contributed to data interpretation. D. Spiegelman and M. Wang contributed to data analysis. P. Khudyakov analyzed the data. P. Khudyakov and M. Wang contributed to revising the supplemental materials. P. Khudyakov, M. Wang, and T.J. Vanderweele contributed to article revision. T.J. Vanderweele contributed to article conceptualization.

ACKNOWLEDGMENTS

The writing of this commentary was supported by the National Institutes of Health (grants DP1ES025459 and R56ES017876).

REFERENCES

1. Spiegelman D, VanderWeele TJ. Evaluating public health interventions: 6. Modeling ratios or differences? Let the data tell us. *Am J Public Health*. 2017; 107(7):1087–1091.
2. Poole C. On the origin of risk relativism. *Epidemiology*. 2010;21(1):3–9.
3. Cornfield J, Haenszel W, Hammond EC, Lilienfeld AM, Shimkin MB, Wynder EL. Smoking and lung cancer: recent evidence and a discussion of some questions. *J Natl Cancer Inst*. 1959;22(1): 173–203.
4. Rothman KJ, Greenland S. *Modern Epidemiology*. 2nd ed. Philadelphia, PA: Lippincott-Raven; 1998.
5. Rothman KJ, Greenland S, Walker A. Concepts of interaction. *Am J Epidemiol*. 1980;112(4):467–470.
6. VanderWeele TJ, Knol MJ. A tutorial on interaction. *Epidemiol Methods*. 2014; 3(1):33–72.
7. Rose G. Strategy of prevention: lessons from cardiovascular disease. *Br Med J (Clin Res Ed)*. 1981;282(6279):1847–1851.
8. Mundt KA, Dell LD, Crawford L, Gallagher AE. Quantitative estimated exposure to vinyl chloride and risk of angiosarcoma of the liver and hepatocellular cancer in the US industry-wide vinyl chloride cohort: mortality update through 2013. *Occup Environ Med*. 2017;74(10):709–716.
9. Northridge ME. Public health methods—attributable risk as a link between causality and public health action. *Am J Public Health*. 1995;85(9): 1202–1204.
10. Orszag PR, Ellis P. Addressing rising health care costs—a view from the Congressional Budget Office. *N Engl J Med*. 2007;357(19):1885–1887.
11. Murray CJL, Salomon JA, Mathers CD, Lopez AD. *Summary Measures of Population Health: Concepts, Ethics, Measurement and Applications*. Geneva, Switzerland: World Health Organization; 2002.
12. *The Global Burden of Disease: Generating Evidence, Guiding Policy*. Seattle, WA: Institute for Health Metrics and Evaluation; 2013.
13. US Environmental Protection Agency. 2006 National Ambient Air Quality Standards (NAAQS) for Particulate Matter (PM_{2.5}). Available at: <https://www.epa.gov/pm-pollution/2006-national-ambient-air-quality-standards-naaqs-particulate-matter-pm25>. Accessed October 25, 2017.
14. Glymour MM, Spiegelman D. Evaluating public health interventions: 5. Causal inference in public health research—do sex, race, and biological factors cause health outcomes? *Am J Public Health*. 2017; 107(1):81–85.
15. Westreich D, Edwards JK, Rogawski ET, Hudgens MG, Stuart EA, Cole SR. Causal impact: epidemiological approaches for a public health of consequence. *Am J Public Health*. 2016;106(6):1011–1012.
16. Farland LV, Tamimi RM, Eliassen AH, Spiegelman D, Bertrand KA, Missmer SA. Endometriosis and mammographic density measurements in the Nurses' Health Study II. *Cancer Causes Control*. 2016;27(10):1229–1237.
17. Spiegelman D. Evaluating public health interventions: 4. The Nurses' Health Study and methods for eliminating bias attributable to measurement error and misclassification. *Am J Public Health*. 2016; 106(9):1563–1566.
18. *National Health and Nutrition Examination Survey Data*. Hyattsville, MD: US Department of Health and Human Services; 2016.
19. Centers for Disease Control and Prevention. National Center for Health Statistics. Available at: <http://www.cdc.gov/nchs>. Accessed November 3, 2011.
20. Spiegelman D, Hertzmark E, Wand HC. Point and interval estimates of partial population attributable risks in cohort studies: examples and software. *Cancer Causes Control*. 2007;18(5):571–579.
21. Drummond MF, Schulpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. *Methods for the Economic Evaluation of Health Care Programmes*. 3rd ed. Oxford, UK: Oxford University Press; 2005.
22. Greenland S. Interpretation and estimation of summary ratios under heterogeneity. *Stat Med*. 1982;1(3):217–227.
23. Cole SR, Stuart EA. Generalizing evidence from randomized clinical trials to target populations: the ACTG 320 trial. *Am J Epidemiol*. 2010;172(1):107–115.
24. Zucker DM. Restricted mean life with covariates: modification and extension of a useful survival analysis method. *J Am Stat Assoc*. 1998;93(442):702–709.
25. Khudyakov P, Spiegelman M, Wang M. *Estimation and Inference for the Semi-Parametric Incremental Cost Effectiveness Ratio and Related Measures*. Cambridge, MA: Harvard University; 2017. Biostatistics Working Paper no. 1225.
26. Rose S, van der Laan MJ. *Targeted Learning: Causal Inference for Observational and Experimental Data*. New York, NY: Springer; 2011.