Software for unbiased estimation of attributable risk.

Jean-François Hamel, Natacha Fouquet, Catherine Ha, Marcel Goldberg, Yves Roquelaure

In the eAppendix, we discuss implications of our findings for weighted analyses of marginal structural mean models for repeated measures. Specifically, to avoid bias, we generally recommend that a possibly incorrect independence correlation structure be used in such analyses; we also discuss, in the eAppendix, an alternative approach that allows for use of a non-independence working correlation structure when estimating marginal structural models.

Eric J. Tchetgen Tchetgen
Departments of Epidemiology and Biostatistics
Harvard School of Public Health
Boston, MA
etchetge@hsph.harvard.edu

M. Maria Glymour
Department of Society, Human Development and Health
Harvard School of Public Health
Boston, MA

Jennifer Weuve
Rush Institute for Healthy Aging
Department of Internal Medicine
Rush University Medical Center
Chicago, IL

James Robins
Departments of Epidemiology and Biostatistics
Harvard School of Public Health
Boston, MA

REFERENCES

1. Robins JM. Marginal structural models. In: *1997 Proceedings of the American Statistical Association. Section on Bayesian Statistical Science*. Alexandria, VA: American Statistical Association; 1998:1–10.
2. Robins JM, Hernán M, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology*. 2000; 11:550–560.
3. Robins JM, Rotnitzky A. Semiparametric efficiency in multivariate regression models with missing data. *J Am Stat Assoc*. 1995;90:122–129.
4. Robins JM, Rotnitzky A, Zhao LP. Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *J Am Stat Assoc*. 1995;90:106–121.
5. SAS Institute Inc. *SAS/STAT 9.2 User's Guide*. 2nd ed. Cary, NC: SAS Institute Inc.; 2009.
6. Weuve J, Tchetgen Tchetgen EJ, Glymour MM, et al. Accounting for bias due to selective attrition: the example of smoking and cognitive decline. *Epidemiology*. 2012;23:119–128.

Software for Unbiased Estimation of Attributable Risk

To the Editor:

Attributable risk (AR) provides a measurement of the population burden of a disease associated with a particular exposure. AR can be interpreted as the proportion of disease cases (eg, carpal tunnel syndrome) that could be avoided if the effects of the exposure (eg, work activity) were totally eliminated from the population. Such an index can help in the design of public health prevention strategies by establishing the relative importance of various exposures.[1] However, few studies report ARs, and those are usually Levin's unadjusted estimators, which are suitable only for binary exposure factors. A more consistent AR assessment would take into account both the multifactorial nature of diseases and confounding factors.[2]

Bruzzi et al[3] proposed a method based on logistic regression for computing n-dimensional ARs. This method provides estimates of adjusted ARs for combinations of exposures, but not separately for each exposure. Several methods have subsequently been proposed to compute an AR estimate for each exposure from these combined ARs[4]:

- The "sequential AR method" is based on the assessment of exposure-specific effects by successively removing each exposure from the analysis while simultaneously calculating their respective contribution to the combined AR. This approach depends on the choice of the exposure permutation.[5]
- The "average AR method" averages the sequential ARs over all the set of possible permutations, and thus allows estimation of AR independent of the order in which exposures are removed.

We briefly illustrate these various methods using data from the Surveillance Program for carpal tunnel syndrome in the Pays de la Loire Region (France). We focus on the main carpal tunnel syndrome exposures for AR estimations: age, sex, obesity, diabetes mellitus, and occupational category for comparing unadjusted, sequential, and average ARs. For the sequential AR estimations, we chose 2 removal sequences: (1) sex, obesity, occupational category, diabetes mellitus, age; and (2) occupational category, diabetes mellitus, obesity, age, sex.

The ARs were highly dependent on the computation method. Crude ARs computed with the Levin formula were overvalued, as demonstrated by the fact that the sum of these ARs (160%) was higher than the possible maximum of 100%. Sequential ARs were slightly dependent on the order of exposures chosen for removal. The sequential ARs were overvalued for the first exposures removed and undervalued for the last exposures removed (Table). (The 2 sequences presented here are just 2 examples among the $5 = 120$ possible permutations of the 5 exposures.) In the final method, the average ARs were constant regardless of the order of removal of the exposures, and their sum was less than 100%, which allows them to be interpreted as the proportion of disease cases that could be avoided in the population if the exposure of interest was eliminated.

Although the average AR method is clearly advantageous, there is no convenient software that allows estimation of average ARs. Standard statistical software, such as The R Project for Statistical Computing, SAS (SAS Institute, Cary, NC) and Stata (StataCorp, College Station, TX), allows AR estimations only for dichotomous exposures.[2,6,7] We provide in the appendix a Stata program for estimating average ARs for dichotomous, polytomous, and quantitative exposures

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: 1044-3983/12/2304-0646
DOI: 10.1097/EDE.0b013e318259c31c

TABLE. AR Estimated for Carpal Tunnel Syndrome Exposures Possibly Considered as Polytomous or Quantitative Variables, Depending on the Attributable Risk Estimation Approach

	OR OR (95% CI)	Sequential AR I AR (95% CI)	Sequential AR II AR (95% CI)	Average AR AR (95% CI)
Age	1.08 (1.07 to 1.09)	0.17 (0.10 to 0.25)	0.40 (0.25 to 0.57)	0.50 (0.44 to 0.56)
Sex				
Men	1.00	0.56 (0.50 to 0.62)	0.01 (0.01 to 0.02)	0.24 (0.21 to 0.27)
Women	4.07 (3.47 to 4.78)			
Diabetes mellitus				
No	1.00	0.00 (−0.01 to 0.01)	0.00 (−0.02 to 0.01)	0.00 (−0.01 to 0.00)
Yes	1.13 (0.70 to 1.85)			
Obesity				
No	1.00	0.01 (−0.01 to 0.023)	0.01 (0.00 to 0.03)	0.02 (0.01 to 0.03)
Yes	6.33 (2.61 to 15.37)			
Occupational category				
Craftsmen, shopkeepers, managers	1.00	0.25 (0.18 to 0.33)	0.57 (0.40 to 0.74)	0.24 (0.17 to 0.31)
Executives, higher intellectual professions	1.01 (0.60 to 1.69)			
Farmers	1.42 (0.83 to 2.41)			
Intermediate occupations	1.46 (0.95 to 2.25)			
Low-grade white collar workers	2.37 (1.57 to 3.57)			
Blue collar workers	3.99 (2.66 to 5.99)			

ARs are presented for each variable and ORs for each qualitative variable modality.
Sequential AR I: removal sequence was gender, obesity, occupational category, diabetes mellitus, and age.
Sequential AR II: removal sequence was occupational category, diabetes mellitus, obesity, age, and sex.
CI indicates confidence interval.

(eAppendix, http://links.lww.com/EDE/A593).

Public health prevention strategies should not only highlight risk factors for a given disease but also the consequences of exposure to these risk factors at the population level. These effects should be evaluated in terms of attribution of risk through reliable and unbiased estimators, such as the average AR. We believe that the additional Stata program provided here will facilitate such estimates.

Jean-François Hamel
Laboratory of Ergonomics and Epidemiology in Health at Work
University of Angers
Angers, France
jean-francois.hamel@etud.univ-angers.fr

Natacha Fouquet
Laboratory of Ergonomics and Epidemiology in Health at Work
University of Angers
Angers, France
Work Health Department
Institute of Health Surveillance
Saint-Maurice, France

Catherine Ha
Work Health Department
Institute of Health Surveillance
Saint-Maurice, France

Marcel Goldberg
Inserm U1018
Population-based Cohorts Research Platform
Centre for Research in Epidemiology and Population Health
Villejuif, France
Versailles-Saint Quentin University
UMRS, France

Yves Roquelaure
Laboratory of Ergonomics and Epidemiology in Health at Work
University of Angers
Angers, France

REFERENCES

1. Davis CH, MacKinnon DP, Schultz A, Sandler I. Cumulative risk and population attributable fraction in prevention. *J Clin Child Adolesc Psychol*. 2003;32:228–235.
2. Rückinger S, von Kries R, Toschke AM. An illustration of and programs estimating attributable fractions in large scale surveys considering multiple risk factors. *BMC Med Res Methodol*. 2009;9:7.
3. Bruzzi P, Green SB, Byar DP, Brinton LA, Schairer C. Estimating the population attributable risk for multiple risk factors using case-control data. *Am J Epidemiol*. 1985;122:904–914.
4. Eide GE, Gefeller O. Sequential and average attributable fractions as aids in the selection of preventive strategies. *J Clin Epidemiol*. 1995; 48:645–655.
5. Land M, Vogel C, Gefeller O. Partitioning methods for multifactorial risk attribution. *Stat Methods Med Res*. 2001;10:217–230.
6. Raemsch C. pARccs: estimation of attributable and partial attributable risks (AR and PAR) and visualization of attributable risks from case-control data. R package version 0.2-2. 2010. Available at: http://CRAN.R-project.org/package=pARccs.
7. Lehnert-Batar A. pARtial: pARtial package. R package version 0.1. 2006. Available at: http://cran.r-project.org/src/contrib/Archive/pARtial/.

Cell Phone Use and Crash Risk

To the Editor:

In a recent report,[1] Young argues that prior studies indicating harmful effects of cell phone use are confounded by driving time, and "corrects" his estimates to claim that no association exists. However, driving time does not confound the association—it is a requirement for the occurrence of the outcome, just as person-time with a uterus is necessary for uterine cancer.

In both the density-sampled case-control and case-crossover designs, the exposure distribution in the controls is meant to estimate the exposure distribution in the person-time at risk for the outcome. In general, one samples directly from at-risk person-time. However, it is also possible to obtain valid estimates of the exposure distribution in