

Model averaging in the analysis of leukemia mortality among Japanese A-bomb survivors

David B. Richardson · Stephen R. Cole

Received: 4 October 2011 / Accepted: 8 December 2011 / Published online: 7 January 2012
© Springer-Verlag 2012

Abstract Epidemiological studies often include numerous covariates, with a variety of possible approaches to control for confounding of the association of primary interest, as well as a variety of possible models for the exposure–response association of interest. Walsh and Kaiser (Radiat Environ Biophys 50:21–35, 2011) advocate a weighted averaging of the models, where the weights are a function of overall model goodness of fit and degrees of freedom. They apply this method to analyses of radiation–leukemia mortality associations among Japanese A-bomb survivors. We caution against such an approach, noting that the proposed model averaging approach prioritizes the inclusion of covariates that are strong predictors of the outcome, but which may be irrelevant as confounders of the association of interest, and penalizes adjustment for covariates that are confounders of the association of interest, but may contribute little to overall model goodness of fit. We offer a simple illustration of how this approach can lead to biased results. The proposed model averaging approach may also be suboptimal as way to handle competing model forms for an exposure–response association of interest, given adjustment for the same set of confounders; alternative approaches, such as hierarchical regression, may provide a more useful way to stabilize risk estimates in this setting.

Keywords Confounding · Epidemiologic methods · Regression · Shrinkage · Validity · Variable selection

In their recent paper on leukemia mortality in the Life Span Study (LSS) of Japanese atomic bomb survivors, Walsh and Kaiser (2011) employ a model averaging approach to radiation risk estimation. They recommend that this approach be used more generally in radiation risk analyses; moreover, they recommend that the model that carried the greatest weight in their model averaging is the one that should be used for application in future leukemia risk assessments that do not employ a model averaging approach. Both recommendations should be viewed cautiously.

As noted by Walsh and Kaiser, observational studies often include numerous covariates. This implies a variety of possible approaches to control for confounding of the association of primary interest, as well as a variety of possible models for the exposure–response association of interest. Faced with several different regression models, Walsh and Kaiser advocate a weighted averaging of the models, where the weights are a function of overall model goodness of fit and degrees of freedom. However, in an analysis aimed at understanding the effect of exposure on mortality in an observational cohort study, one should focus on minimizing bias due to confounding rather than on overall model goodness of fit or parsimony. One might justifiably omit covariates that are strong predictors of the outcome (and hence reduce the residual model deviance) because these covariates do not bias the risk estimates of primary interest; and, one might justifiably include covariates that contribute minimally to model fit, but by their omission would lead to bias in the estimated association of primary interest. Walsh and Kaiser recommend an approach that turns this logic on its head. The weights employed in their model averaging approach penalize the omission of covariates that are strong predictors of the outcome, but irrelevant as confounders of the association of interest, and penalize adjustment for covariates that are

D. B. Richardson (✉) · S. R. Cole
Department of Epidemiology, School of Public Health,
University of North Carolina at Chapel Hill,
Chapel Hill, NC 27599, USA
e-mail: david.richardson@unc.edu

confounders of the association of interest, but contribute little to overall model goodness of fit. Epidemiologists have long recognized that moderately strong confounders may not be statistically significantly associated with the outcome (Robins and Greenland 1986; Greenland 2008).

We specifically addressed the distinction between a modeling approach that minimizes bias in estimation of an association and an approach that focuses on overall goodness of fit in (Richardson et al. 2009). There we adjusted for proximal versus distal location as compared to the hypocenters of the atomic bomb explosions over Hiroshima and Nagasaki in our analysis of radiation–leukemia mortality associations among atomic bomb survivors, while previous analyses had not (Richardson et al. 2009). We noted that prior research suggested that location was a potential confounder of the radiation dose–leukemia association, because rural (i.e., distal) location was a determinant of estimated DS02 dose and rural cohort members may have different mortality risks than urban. Adjustment for this variable did not substantially improve overall model goodness of fit, but failure to adjust for it led to a substantial change in estimate of the association of primary interest. The averaging approach advocated in (Walsh and Kaiser 2011) would discount this model when compared to a model that omitted adjustment for location, and hence discounts the control of confounding bias. Similarly, we noted that background stratification on the selected model covariates provided the desired control for confounding by these factors. The radiation risk estimates of primary interest obtained from this background stratified model were of similar precision to the estimates obtained when using a parametric model for the covariates. The approach employed by Walsh and Kaiser discounted models that employed a background stratified approach to adjustment for confounding factors without regard to the fundamental question of whether such models are liable to greater bias or mean square error in radiation risk estimates than the other models evaluated. Indeed, a model omitting several statistically non-significant (for outcome) confounders which each bias the effect away from the null will “fit better”, and this biased reduced-parameter model may have the best Akaike information criterion (AIC) and therefore highest model averaging weight. As a simple example, consider the data in Table 1 and two logistic regression models for the binary outcome Y . Model 1 includes only the binary exposure X , while model 2 includes the exposure X and a binary confounder Z . The AIC (AIC weight) for the crude model 1 is 679.1 (45%); and the AIC (AIC weight) for the adjusted model 2 is 678.1 (55%). The crude model yields an odds ratio of 1.24, while the adjusted model yields an odds ratio of 1.01. These data were generated such that the exposure X does not cause the outcome Y .

Table 1 Hypothetical example data on 106 cases observed in 1,000 participants

Covariate Z	Exposure X	Outcome Y	Number
0	0	0	354
0	0	1	38
0	1	0	102
0	1	1	6
1	0	0	115
1	0	1	12
1	1	0	323
1	1	1	50

Therefore, the adjusted model is correct. The model-averaged odds ratio is 1.11, which makes little sense.

Walsh and Kaiser might have focused on a comparison of models that employed an identical approach to confounder control, but employed different forms for the radiation dose–response association and its effect modifiers. Models that accommodate more flexibility in description of the exposure–time–response relationship between radiation dose and leukemia tend to involve greater degrees of freedom. A simple linear term to describe effect modification by time since exposure, for example, may be adequate for some purposes; however, a more flexible model may be substantively more plausible or offer interesting insights. Model averaging is one approach to characterizing uncertainty in risk estimates in epidemiological studies in which there is low statistical power to discriminate between alternative model forms. However, other appealing alternatives include hierarchical regression approaches that nest simpler models within more complicated models and provide a way to stabilize risk estimates and reduce bias arising from model misspecification Richardson et al. (2011). The model averaging approach used in Walsh and Kaiser (2011) leads to a focus on predictive modeling of the outcome that is seldom the primary goal for researchers interested in etiologic relationships. Rather researchers are better served by focusing on issues that will strengthen causal interpretation of the radiation dose–outcome associations of primary interest and on approaches for smoothing and pattern recognition in regression modeling that will aid in understanding factors that influence radiation health effects.

Acknowledgments This project was supported by Grant R01-CA117841 from the National Cancer Institute, National Institutes of Health.

References

- Greenland S (2008) Invited commentary: variable selection versus shrinkage in the control of multiple confounders. *Am J Epidemiol* 167:523–529 (discussion 30–1)

- Richardson DB, MacLehose RF, Langholz B, Cole SR (2011) Hierarchical latency models for dose–time–response associations. *Am J Epidemiol* 173:695–702
- Richardson D, Sugiyama H, Nishi N, Sakata R, Shimizu Y, Grant EJ, Soda M, Hsu WL, Suyama A, Kodama K, Kasagi F (2009) Ionizing radiation and leukemia mortality among Japanese atomic bomb survivors, 1950–2000. *Radiat Res* 172:368–382
- Robins JM, Greenland S (1986) The role of model selection in causal inference from nonexperimental data. *Am J Epidemiol* 123:392–402
- Walsh L, Kaiser JC (2011) Multi-model inference of adult and childhood leukaemia excess relative risks based on the Japanese A-bomb survivors mortality data (1950–2000). *Radiat Environ Biophys* 50:21–35