

Differences Between Marginal Structural Models and Conventional Models in Their Exposure Effect Estimates

A Systematic Review

David Suarez,^{a,b} Roger Borràs,^a and Xavier Basagaña^{c,d,e}

Background: Marginal structural models were developed to address time-varying confounding in nonrandomized exposure effect studies. It is unclear how estimates from marginal structural models and conventional models might differ in real settings.

Methods: We systematically reviewed the literature on marginal structural models since 2000.

Results: Data to compare marginal structural models and conventional models were obtained from 65 papers reporting 164 exposure-outcome associations. In 58 (40%), estimates differed by at least 20%, and in 18 (11%), the 2 techniques resulted in estimates with opposite interpretations. In 88 papers, marginal structural models were used to analyze real data; only 53 (60%) papers reported the use of stabilized inverse-probability weights and only 28 (32%) reported that they verified that the mean of the stabilized inverse-probability weights was close to 1.0.

Conclusions: We found important differences in results from marginal structural models and from conventional models in real studies. Furthermore, reporting of marginal structural models can be improved.

(*Epidemiology* 2011;22: 586–588)

Although well-conducted randomized experiments are in principle the preferred study design for evaluating the causal effects of an exposure on an outcome, randomization is often unfeasible, leaving observational studies as the best alternative. However, observational studies are prone to confounding and selection bias. Several analytic methods can be

used to overcome these potential biases, such as multivariable regression, propensity scores, and the instrumental variable approach.¹

These techniques may be inadequate in longitudinal observational studies when more complex biases are present, such as when exposure is time-dependent and time-varying confounders are present. A time-varying confounder is a variable affected by prior exposure that predicts both the subsequent outcome and subsequent exposure. To correct the effects of time-varying confounding in observational studies, Robins in 1997 developed a new class of causal models called marginal structural models.² These models use inverse-probability (of exposure) weights to create an artificial population in which covariate imbalances are removed and causal effects can be estimated. The publication of 2 companion papers^{3,4} in 2000 facilitated the practical application of this new method.

We aimed to determine whether there were differences in exposure effect estimates between marginal structural models and conventional models in real applications, and to examine the completeness of reporting of marginal structural models.

METHODS

We systematically reviewed published studies that used marginal structural models to estimate exposure effects potentially affected by time-varying confounding. Citations from 2000 to October 2009 were retrieved from both PubMed and ISI Web of Knowledge databases following this search strategy: first, we retrieved all papers from PubMed selected by the query: (marginal structural model) OR (“marginal structural model”) OR (marginal structural models) OR (“marginal structural models”) OR (“inverse probability weighting”) OR (“inverse probability weighted”) OR (“inverse probability weights”); second, we retrieved from ISI all papers selected by the query: Topic = (“marginal structural”) AND Topic = (model*); third, we retrieved from ISI all papers selected by the query: Topic = (“inverse probability”) AND Topic = (weight*); and fourth, we retrieved all papers that cited one of the 2 seminal epidemiologic papers on marginal structural models.^{3,4} Methodologic papers and non-English language papers were included, but not abstracts, editorials, or letters.

Submitted 11 October 2010; accepted 20 January 2011; posted 3 May 2011.

From the ^aEpidemiology and Assessment Unit, Fundació Parc Taulí, Universitat Autònoma de Barcelona, Sabadell, Spain; ^bDepartment of Studies and Statistics, Comisión del Mercado de las Telecomunicaciones, Barcelona, Spain; ^cCentre for Research in Environmental Epidemiology (CREAL), Barcelona, Spain; ^dMunicipal Institute of Medical Research (IMIM-Hospital del Mar), Barcelona, Spain; and ^eCIBER Epidemiología y Salud Pública (CIBERESP), Spain.

Supported by research grants from the Instituto de Salud Carlos III, Ministerio de Sanidad y Consumo, Spain and from the Departament de Salut, Generalitat de Catalunya (FIS ECA07/041) (to D.S.).

SDC Supplemental digital content is available through direct URL citations in the HTML and PDF versions of this article (www.epidem.com).

Correspondence: David Suarez, Epidemiology and Assessment Unit Fundació Parc Taulí, Parc Taulí, 1, 08208 Sabadell (Barcelona), Spain. E-mail: david.suarez.lamas@gmail.com.

Copyright © 2011 by Lippincott Williams & Wilkins

ISSN: 1044-3983/11/2204-0586

DOI: 10.1097/EDE.0b013e31821d0507

The selected articles were abstracted independently by 2 reviewers. From each article with results from both marginal structural models and conventional analyses, we abstracted the measure of association, confidence intervals, and standard errors for the exposure–outcome associations. We defined conventional models as the alternative model reported in the paper, if any (eg, a multivariable regression model). If more than one conventional model was reported, we chose the one providing the highest degree of adjustment. Only results from the exposure of interest were abstracted, even if a paper reported estimates on other covariates. In addition, for all papers using marginal structural models (including those that did not provide a comparison with conventional methods), we recorded 2 good practices in marginal structural modeling: whether the paper stated that stabilized inverse-probability weights had been used, and whether it stated that the mean of the stabilized inverse-probability weights was close to one. Robins et al³ recommended stabilized inverse-probability weights because they have good convergence properties, are efficient, and are easy to compute. Stabilized inverse-probability weights with a mean far from one may indicate a violation of some of the hypotheses of the model, eg, a misspecification of the weight model or a violation of the positivity assumption.⁵

RESULTS

The search yielded 661 unique papers; 543 of these were excluded because they did not use marginal structural models to estimate the effect of an exposure on outcome while accounting for time-dependent confounding. The Figure shows the distribution of the resulting 118 papers by year. The number of articles that used marginal structural models to control for time-varying confounding increased from 2 in 2002 to 29 in 2008 (almost 15-fold); in comparison, the total number of publications listed in PubMed increased only 46%

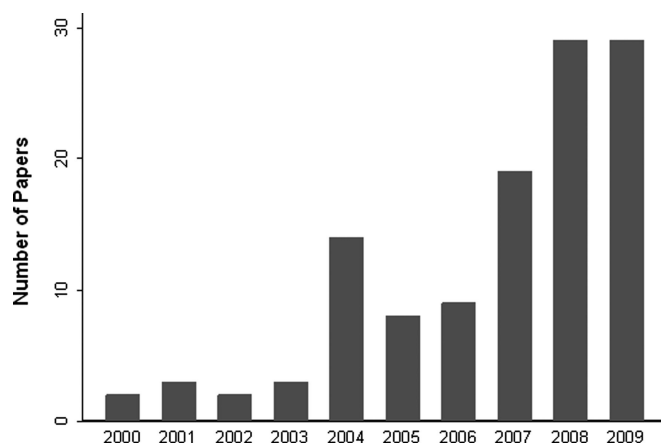


FIGURE. Number of papers using marginal structural models to control for time-varying confounding, by year of publication.

in the same period. Of the 118 papers, 65 (55%) compared the numerical results of the marginal structural model and a conventional model, and an additional 5 papers (4%) reported the results qualitatively.

The 65 papers that provided data to compare marginal structural models and the conventional models reported 164 exposure–outcome associations. Of these, 18 (11%) yielded a marginal structural model (MSM) estimate and a conventional model estimate with opposite interpretations, ie, one model considered the exposure a risk factor and the other considered it a protective factor. For the remaining associations (in which the direction of the effect was the same), the MSM estimate differed by at least 20% from the conventional estimate on the usual scale (eg, odds ratio or coefficient of linear regression) in 58 (40%) of the exposure–outcome associations. Similar results were obtained when restricting the comparisons to linear regression coefficients (43 of 164 estimates, 26%), where 6 (14%) yielded estimates with opposite interpretations, and, the MSM estimates differed from the conventional estimates by at least 20% in 17 of the remaining 37 associations (46%).

The standard errors for the 2 methods were available for 156 associations. The standard errors of the MSM associations were in median 19% greater than the respective conventional standard errors (IQR: 2% to 48%).

Of the original 118 papers, 30 were purely methodological. The remaining 88 used marginal structural models to analyze a real dataset. Only 53 (60%) of these papers reported that stabilized inverse-probability weights were used, and only 28 (32%) reported that they verified that the mean of the stabilized inverse-probability weights was close to one. Finally, almost half of the 88 papers were about HIV treatment (46%). The eAppendix (<http://links.lww.com/EDE/A481>) provides a list of the 118 publications that used marginal structural models.

DISCUSSION

This review of the literature shows that the number of published papers using marginal structural models has increased rapidly since the publication of 2 seminal papers on marginal structural models in 2000.^{3,4} Some reports from single studies have provided evidence that, compared with conventional models, marginal structural models yield results closer to those results of randomized experiments.^{6,7} However, no systematic review of this difference in the literature has been undertaken. Our review shows that there are substantial differences between marginal structural models and conventional analyses in real studies in which time-varying confounding was suspected.

Controlling for time-varying confounding increased the median standard error of the estimates nearly 20% compared with conventional standard errors. Cole and Hernán⁵ have described this increase in standard error as a trade-off between bias and precision.

Our review also shows that the reporting of marginal structural models can be improved. Researchers should routinely report whether they used stabilized inverse-probability weights, and verify that the mean of these weights was close to 1.0.

The implications of the results of this review are not limited to observational studies with time-varying confounding. As noted by Toh and Hernán,⁸ randomized studies can also suffer from biases similar to those affecting observational studies if the study is randomized at baseline but postbaseline changes in treatment are not randomized. Thus, the use of marginal structural models is to be encouraged in that context as well.

Our review has some limitations. First, we missed any papers that used marginal structural models without explicitly reporting doing so in the title, keywords, or abstract. However, marginal structural models are still viewed as an advanced and novel approach to data analysis, so it is likely that few papers would use this method without clearly acknowledging it. Second, marginal structural models provide marginal estimates while conventional estimates are conditional. In the case of odds ratios and hazard ratios, the 2 approaches can differ even in the absence of confounding.⁹ However, this is also true for the comparison of propensity scores versus multivariable regression models, and the differences found in the present review are stronger than those previously reported for propensity scores.^{10,11} In addition, restricting our comparison to the linear regression coefficient estimates, where marginal and conditional estimates are coincident,^{12,13} resulted in differences similar to those found in the overall analysis.

Third, violations of the positivity assumption could lead to biased marginal structural model estimates.¹⁴ This point underlines the importance of checking that the mean of the stabilized inverse-probability weights is close to 1.0 and of exploring the impact of extreme weights. Fourth, if papers with similar results from the marginal structural models and conventional analyses were more likely to report the results of only one of the 2 methods, there might be a publication bias.

In conclusion, the findings of this review suggest that there are important differences between marginal structural model and conventional estimates in real settings and that

reporting of marginal structural models can be improved. Applied researchers should use marginal structural models in their analyses whenever time-varying confounding is likely to occur, taking care to provide a complete description of their results.

ACKNOWLEDGMENTS

We thank Samuel Rolfe for his help and an anonymous reviewer for the helpful comments.

REFERENCES

1. Haro JM, Kontodimas S, Negrin MA, Ratcliffe M, Suarez D, Windmeijer F. Methodological aspects in the assessment of treatment effects in observational health outcomes studies. *Appl Health Econ Health Policy*. 2006;5:11–25.
2. Robins JM. Marginal structural models. In: 1997 Proceedings of the Section on Bayesian Statistical Science, Alexandria, VA: American Statistical Association, 1998;1–10.
3. Robins JM, Hernán MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology*. 2000;11:550–560.
4. Hernán MA, Brumback B, Robins JM. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology*. 2000;11:561–570.
5. Cole SR, Hernán MA. Constructing inverse probability weights for marginal structural models. *Am J Epidemiol*. 2008;168:656–664.
6. Cook NR, Cole SR, Hennekens CH. Use of a marginal structural model to determine the effect of aspirin on cardiovascular mortality in the Physicians' Health Study. *Am J Epidemiol*. 2002;155:1045–1053.
7. Suarez D, Haro JM, Novick D, Ochoa S. Marginal structural models might overcome confounding when analyzing multiple treatment effects in observational studies. *J Clin Epidemiol*. 2008;61:525–530.
8. Toh S, Hernán MA. Causal inference from longitudinal studies with baseline randomization. *Int J Biostat*. 2008;4:Article 22.
9. Kaufman JS. Marginalia: comparing adjusted effect measures. *Epidemiology*. 2010;21:490–493.
10. Shah BR, Laupacis A, Hux JE, Austin PC. Propensity score methods gave similar results to traditional regression modeling in observational studies: a systematic review. *J Clin Epidemiol*. 2005;58:550–559.
11. Sturmer T, Joshi M, Glynn RJ, Avorn J, Rothman KJ, Schneeweiss S. A review of the application of propensity score methods yielded increasing use, advantages in specific settings, but not substantially different estimates compared with conventional multivariable methods. *J Clin Epidemiol*. 2006;59:437–447.
12. Gail M, Wieand S, Piantadosi S. Biased estimates of treatment effect in randomized experiments with nonlinear regression and omitted covariates. *Biometrika*. 1984;71:431–444.
13. Vellaisamy P, Vijaya V. Collapsibility of regression coefficients and its extensions. *J Stat Plan Inference*. 2008;138:982–994.
14. Kurth T, Walker AM, Glynn RJ, et al. Results of multivariable logistic regression, propensity matching, propensity adjustment, and propensity-based weighting under conditions of nonuniform effect. *Am J Epidemiol*. 2006;163:262–270.