Int J Public Health (2011) 56:117–119
DOI 10.1007/s00038-010-0198-4

HINTS & KINKS

Marginal Structural Models: unbiased estimation for longitudinal studies

Erica E. M. Moodie · D. A. Stephens

Received: 7 May 2010 / Revised: 1 August 2010 / Accepted: 12 September 2010 / Published online: 8 October 2010

Abstract
Introduction In this article, we introduce Marginal Structural Models, which yield unbiased estimates of causal effects of exposures in the presence of time-varying confounding variables that also act as mediators.
Objectives We describe estimation via inverse probability weighting; estimation may also be accomplished by g-computation (Robins in Latent Variable Modeling and Applications to Causality, Springer, New York, pp 69–117, 1997; van der Wal et al. in Stat Med 28:2325–2337, 2009) or targeted maximum likelihood (Rosenblum and van der Laan in Int J Biostat 6, 2010).
Conclusions When both time-varying confounding and mediation are present in a longitudinal setting data, Marginal Structural Models are a useful tool that provides unbiased estimates.

Keywords Bias · Confounding · Mediation · Longitudinal data · Marginal Structural Models · Inverse probability weighting

Marginal Structural Models and inverse probability weighting

Marginal Structural Models (MSMs) provide a powerful tool to assess the effects of exposures in longitudinal settings though they may also be used for cross-sectional data. These models are *marginal* because they pertain to population-average effects and *structural* because they describe causal (not associational) effects.

To understand MSMs, it helps to introduce *counterfactual* or *potential* outcomes: outcomes that would have been observed had a person been exposed to a particular treatment pattern. Consider a two-interval setting where data are collected at baseline (t_0), t_1, and t_2, with covariates L_j measured at t_{j-1} ($j = 1, 2$), treatments A_j taken between t_{j-1} and t_j ($j = 1, 2$), and outcome Y measured at t_2. The predictors of interest, A_j, may be actual treatments (clinical studies) or exposures (general epidemiology); these terms will be used interchangeably. There are four possible exposure patterns: always exposed $(A_1,A_2) = (1,1)$, never exposed $(A_1,A_2) = (0,0)$, or only exposed in one interval $(A_1,A_2) = (1,0), (0,1)$. Each person has four potential responses (one corresponding to each exposure pattern), denoted $Y(1,1)$, $Y(0,0)$, $Y(1,0)$, $Y(0,1)$. Suppose an individual is treated in both intervals: we observe outcome Y, which equals $Y(1,1)$, but we do not observe outcomes under the three other exposure patterns. Counterfactuals help to formulate causal models. In particular, a MSM is a model for $\mathrm{E}[Y(A_1,A_2)]$, the average outcome if the entire population was exposed to treatment pattern A_1,A_2, for each A_1, A_2.

Inverse probability weighting (IPW) is used to create a *pseudopopulation* in which the treatments A_j and covariates L_j are unassociated, and therefore, there exists no confounding so that effect estimates are unbiased; see Robins et al. (2000) for a worked example.

Choice of weights

There are several options available for the weights. The simplest are the *unstandardized weights*:

E. E. M. Moodie (✉)
Department of Epidemiology and Biostatistics,
McGill University, 1020 Pine Avenue West, Montreal,
QC H3A 1A2, Canada
e-mail: erica.moodie@mcgill.ca

D. A. Stephens
Department of Mathematics and Statistics, McGill University,
805 Sherbrooke Street West, Montreal, QC H3A 2K6, Canada

$$w^{u} = \{P(A_1 = a_1|L_1) \times P(A_2 = a_2|L_1, A_1, L_2)\}^{-1} = \frac{1}{P(A_1, A_2|L_1, L_2)}.$$

Each individual's weight is computed by taking the inverse of the product of the estimated probability of receiving the observed treatment in each interval, conditional on past time-varying covariates (including baseline covariates and previous treatment). Unstandardized weights typically lead to very large standard errors (Hernan et al. 2000); it is more common to use *standardized weights*:

$$w^{s} = \left\{\frac{P(A_1 = a_1) \times P(A_2 = a_2|L_1, A_1)}{P(A_1 = a_1|L_1) \times P(A_2 = a_2|L_1, A_1, L_2)}\right\}.$$

The denominators of the standardized and unstandardized weights are identical. The numerator of the standardized weights is obtained by taking the product of the estimated probability of receiving the observed treatment in each interval, conditional on baseline covariates and previous treatments only. Treatment models are usually fit by logistic regression. Standardized weights may still yield variable estimates if there are large weights, indicating that some individuals received unusual treatments given their covariates. Normalization (Xiao et al. 2010) and truncation (Cole and Hernan 2008) have been proposed to further reduce standard errors.

Estimation

The estimation of a MSM via IPW is straightforward:

1. Fit *treatment models*: fit a model for the probability of being treated at each interval. For binary treatment, this is typically a logistic regression model.
2. Use the models in (1) to predict the observed exposure pattern by taking the product of the probability of receiving the treatment that he did receive in each interval.
3. Set each individual's weight to one over the probability computed in (2). Optionally (recommended): stabilize, normalize, and/or truncate the weights.
4. Fit a *response model*, weighting each individual by the weights computed in (3). For a continuous outcome, this is typically a linear regression model. The regression model should condition on treatments and possibly baseline covariates, but not time-varying covariates.

The covariates included in the treatment models will depend on the substantive problem. All confounding covariates should be included. Automated model selection (e.g., stepwise procedures) should not be used. See Austin et al. (2006) and Lefebvre et al. (2008) for further discussion.

Censoring and missing data

Censoring may be dealt with naturally in the IPW framework. Censoring in longitudinal studies (including studies of survival time) occurs when a participant is lost to follow-up due to voluntary drop-out or moving to another geographical region. The key is to build *censoring models* and use these to predict the probability that a person remained in the study for each interval in which he did remain in the study. These probabilities are multiplied and inverted to give a censoring weight, which is multiplied by the weight in step (3) of the estimation algorithm above with the resulting value used in the response model of step (4).

Intermittently missing data may be handled in several ways; two typical approaches are to impute (preferably using multiple imputations) or to censor at the first instance of missing data and then to use censoring weights as described above (Moodie et al. 2008).

Other outcomes

Although the development in this article has focused on a continuous outcome Y, MSMs may be used for any type of outcome by following the algorithm outlined above, using the appropriate regression model in step (4). For a binary outcome, weighted logistic regression should be used. For time-to-event data, a discrete-time approximation is often used along with logistic regression (e.g. Hernan et al. 2000; Bryan et al. 2004; Westreich et al. 2009); however, Cox models also may be fitted (Xiao et al. 2010).

A simulated example

We return to the simulated study of amblyopia of our companion paper (Moodie and Stephens 2010). Traditional regression models were unable to produce unbiased estimates of the treatment effects. Using a MSM yields estimates (SE) of -0.039 (0.013) and -0.086 (0.013) for A_1 and A_2, respectively, which we compare to the true values of -0.038 and -0.086. Repeating the simulation 10,000 times with a sample size of 100 indicated bias of less than 0.65% for each coefficient.

Software

Marginal Structural Models are easily implemented with standard software. Sample code is available in SAS, Stata, and R for fitting a MSM for survival data using a discrete-time approximation (Hernan et al. 2000; Fewell et al. 2004; Bryan et al. 2004). Sample SAS code is also available for

the binary outcome case (Bodnar et al. 2004), and a number of SAS macros are available at http://www.hsph.harvard.edu/causal/software.htm.

Acknowledgments Both authors acknowledge funding from the Natural Sciences and Research Council of Canada. Moodie also acknowledges funding from the Canadian Institutes of Health Research.

References

Austin PC, Grootendorst P, Anderson GM (2006) A comparison of the ability of different propensity score models to balance measured variables between treated and untreated subjects: a Monte Carlo study. Stat Med 26:734–753

Bodnar LM, Davidian M, Siega-Riz AM, Tsiatis A (2004) Marginal structural models for analyzing causal effects of time-dependent treatments: an application in perinatal epidemiology. Am J Epidemiol 159:926–934

Bryan J, Yu Z, van der Laan MJ (2004) Analysis of longitudinal marginal structural models. Biostatistics 5:361–380

Cole SR, Hernan MA (2008) Constructing inverse probability weights for marginal structural models. Am J Epidemiol 168:656–664

Fewell Z, Hernan MA, Wolfe F, Tilling K, Choi H, Sterne JAC (2004) Controlling for time-dependent confounding using marginal structural models. Stat J 4:402–420

Hernan MA, Brumback B, Robins JM (2000) Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. Epidemiology 11:561–570

Lefebvre G, Delaney JAC, Platt RW (2008) Impact of mis-specification of the treatment model on estimates from a marginal structural model. Stat Med 27:3629–3642

Moodie EEM, Stephens DA (2010) Using directed acyclic graphs to detect limitations of traditional regression in longitudinal studies. Int J Public Health (in press)

Moodie EEM, Delaney JAC, LeFebvre G, Platt RW (2008) Missing confounding data in marginal structural models: a comparison of inverse probability weighting and multiple imputation. Int J Biostat 4:13

Robins JM (1997) Causal inference from complex longitudinal data. In: Berkane M (ed) Latent variable modeling and applications to causality. Springer, New York, pp 69–117

Robins JM, Hernan MA, Brumback B (2000) Marginal structural models and causal inference in epidemiology. Epidemiology 11:550–560

Rosenblum M, van der Laan MJ (2010) Targeted maximum likelihood estimation of the parameter of a marginal structural model. Int J Biostat 6

van der Wal WM, Prins M, Lumbreras B, Geskus RB (2009) A simple g-computation algorithm to quantify the causal effect of a secondary illness on the progression of a chronic disease. Stat Med 28:2325–2337

Westreich D, MacPhail P, Van Rie A, Malope-Kgokong B, Ive P, Rubel D, Boulme R, Eron J, Sanne I (2009) Effect of pulmonary tuberculosis on mortality in patients receiving HAART. AIDS 23:707–715

Xiao Y, Abrahamowicz M, Moodie EEM (2010) Accuracy of conventional and marginal structural Cox model estimators: a simulation study. Int J Biostat 6