View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Carolina Digital Repository

NIH Public Access
Author Manuscript
Stat Med. Author manuscript; available in PMC 2014 September 30.

Published in final edited form as:
Stat Med. 2013 April 15; 32(8): 1383–1393. doi:10.1002/sim.5599.

An information criterion for marginal structural models

Robert W. Platt[a], **M. Alan Brookhart**[b], **Stephen R. Cole**[b], **Daniel Westreich**[b,c], and **Enrique F. Schisterman**[d]

[a]Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, QC

[b]Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC

[c]Department of Obstetrics and Gynecology and Global Health Institute, Duke University, Durham, NC

[d]Division of Epidemiology, Statistics and Prevention Research, NICHD, NIH, Bethesda, MD

Summary

Marginal structural models were developed as a semiparametric alternative to the G-computation formula to estimate causal effects of exposures. In practice, these models are often specified using parametric regression models. As such, the usual conventions regarding regression model specification apply. This paper outlines strategies for marginal structural model specification, and considerations for the functional form of the exposure metric in the final structural model. We propose a quasi-likelihood information criterion adapted from use in generalized estimating equations. We evaluate the properties of our proposed information criterion using a limited simulation study. We illustrate our approach using two empirical examples. In the first example, we use data from a randomized breastfeeding promotion trial to estimate the effect of breastfeeding duration on infant weight at one year. In the second example, we use data from two prospective cohorts studies to estimate the effect of highly active antiretroviral therapy on CD4 count in an observational cohort of HIV-infected men and women. The marginal structural model specified should reflect the scientific question being addressed, but can also assist in exploration of other plausible and closely related questions. In marginal structural models, as in any regression setting, correct inference depends on correct model specification. Our proposed information criterion provides a formal method for comparing model fit for different specifications.

Keywords

Bias; Causal inference; Marginal structural model; Regression analysis; Model specification

---

After introduction around the turn of the century [1-4] marginal structural models have diffused rapidly into the biomedical literature with numerous important applications in a range of substantive fields [5-9]. Much has been written about the identifiability assumptions (consistency, positivity, and exchangeability), which are necessary to yield consistent estimates of exposure (or treatment) effects on outcomes of interest. For marginal structural models, there are published discussions about the consistency [10-12], positivity [13-16], and exchangeability (with respect to confounding and selection bias) [14, 15]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

assumptions, as well as discussions about the specification of models used to construct inverse probability weights [15, 17, 18]. However, while others have considered the specification of regression models in general (e.g. Harrell [19]) little has been written about the choice of the functional (regression) form for the exposure of interest in the final structural model [20, 21]. Here, we discuss specification of the final functional form of the marginal structural model, extend the use of a weighted information criterion originally proposed for sampling-weighted data[22] to the generalized linear marginal structural model, and illustrate some of the challenges and implications of model fitting for marginal structural models using two examples [23, 24].

Marginal Structural Models

Marginal structural models [4] are models for the marginal expectation of a potential outcome as a function of a specified exposure plan. We refer readers to previous literature for a formal definition of potential outcomes [25, 26]. Let Y denote an outcome, X(t) a time varying exposure (for simplicity, assumed binary with X(t) = 1 denoting exposure at time t, X(t) = 0 denoting lack of exposure), and let $\bar{x}$ refer to history of time varying exposure X(t) under a specific exposure plan to the end of follow up. A marginal structural model may be stated as $g^{-1}(E[Y_{\bar{x}}]) = f(\bar{x})$ where $Y_{\bar{x}}$ is the potential outcome under exposure plan $\bar{x}$, g is a link function, and f(.) is a defined function we call the *exposure metric*, which is typically a linear combination of components of $\bar{x}$ (e.g., cumulative exposure). Exposure plans may be fixed or dynamic. Fixed (or static) exposure plans are knowable at study entry, and therefore do not depend on the evolution of participant characteristics. Examples of fixed exposure plans include: always treat (i.e., $\bar{x}$ = {1,…,1}), never treat (i.e., $\bar{x}$ = {0,…,0}), treat every other visit (i.e., $\bar{x}$ = {0,1,0,1,…,0}), or initiate treatment after four visits (i.e., $\bar{x}$ = {0,0,0,0,1, …,1}). Dynamic exposure plans are those in which the designed exposure plan depends on time varying covariates measured after baseline [27], and are not considered further here; see Hernán [28] or Cain [29] for examples of dynamic regimes and methods for estimating parameters for models for dynamic exposure plans using inverse probability weighting.

The parameters of a marginal structural model are typically estimated using inverse probability weights. One first fits a model for the probability of exposure X(t) conditional on exposure and confounder histories. This model for the probability of exposure is then used to compute weights equal to the inverse probability of observed exposure given exposure and confounder histories [4, 15]. These weights deal with confounding, and can be easily extended to handle censoring due to drop-out [30, 31]; in short, we typically view such censoring as another exposure and only want to make inference under the plan “never drop out”. Methods for constructing the weights have been considered elsewhere [15, 17, 32]. The weights are then applied to the observed data, and an unadjusted model for the outcome as a function of exposure is then fit to the weighted sample. If the functional forms of the exposure model and final structural model (i.e., the exposure metric) are correctly specified and the identifiability assumptions of the marginal structural model (mentioned above) are met, then the resulting estimate has a causal interpretation as the effect of exposure on the outcome of interest. Specifically, under assumptions the weighted model for an exposure-specific contrast in the *observed* outcomes corresponds to a model for an exposure-specific contrast in the *potential* outcomes.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Exposure Metrics

A well-specified scientific question should imply a specific form for the exposure metric f(.). For example, in the presence of an induction period one may wish to ignore recent treatment when constructing the exposure metric. Alternatively, when the exposure acts rapidly without sustained effects one may wish to ignore historic exposure when constructing the exposure metric. Different choices of the exposure metric f(.) therefore imply refinements in the scientific question and interpretation of the relationship between the exposure and outcome. We are concerned here with the selection of f(.) in a marginal structural model.

In practice, the exposure metric f(.) is often a simple function such as cumulative exposure to visit j: that is, $f(\bar{x}_j)=\sum_{k=0}^{j} x_k$. However, much more general exposure metrics are possible. For example, one might weight the cumulative exposure as $f(\bar{x}_j)=\sum_{k=0}^{j} m_k x_k$, where m_k are time-specific weights; when $m_k = 1$ for all k, f(.) simplifies to the cumulative exposure. An example of a non-trivial m_k might define the weights as a function of k, for example, $m_k = 1/(j - k + 1)$, so that the most recent exposure k = j gets weight of unity and exposure measurements get monotonically decreasing weights as they move back in time from visit j. More generally, an exposure metric that allows for a defined exposure window, as well as within-window weighting, could be defined as follows:

$$f(\bar{x}_j)=\sum_{k=\max(0,a)}^{\min(j,b)} m_k x_k,$$

where b≥0 is the end of the exposure window (i.e., lag), a≤b is the start of the exposure window, and a-b+1 is the length of the exposure window. When b=j (i.e., no lag) and a=0 this within-window weighted cumulative exposure reduces to the simple weighted cumulative exposure above. The choice of the weight function for a particular application should be guided by expert knowledge of the substantive matter and the scientific question under study.

When f(.) is not pre-determined by the scientific question at hand, or when the substantive issues at hand imply several possible functions for the exposure metric, it may be desirable to allow the observed data to influence the choice of f(.) (although such post hoc modeling necessarily compromises interpretations of formal statistical hypothesis tests). Comparisons of relative goodness of model fit often rely on graphical depictions of the data and information criteria such as Akaike's (AIC) [33], defined as −2logL + 2p, where L is the likelihood for a model and p is the number of parameters in the model (note that AIC and other criteria do not require nested models).

Such an approach can be extended naturally to the marginal structural model setting to compare different forms of the structural model assuming fixed weights. That is, the only

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

component that varies between models is the exposure metric in the final structural model. Then, because marginal structural models are fit using a weighted estimating equation, the likelihood in the AIC must be replaced with a weighted likelihood [22] or quasi-likelihood [34], depending on the functional form of the model. For a generalized linear marginal structural model the weighted information criterion AIC_w is defined as $QIC_w = -2\hat{Q_w} + 2p$, where p is the number of parameters estimated in the model and $\hat{Q_w}$ is the weighted quasi-likelihood component (the derivative of which with respect to the parameter of interest is the relevant component of the estimating function). In the linear case, (and therefore in our examples) $\hat{Q_w}$ is the weighted likelihood assuming the normal distribution evaluated at its maximum, namely $Q_w = \sum_{i=1}^{N}\sum_{j=1}^{J_i} .5 \times w_{ij} \times [y_{ij} - \hat{\mu}_{ij}]^2$. Here, $\hat{w}_{ij}$ is the estimated inverse probability weight, y_{ij} is the observed outcome, and $\hat{\mu_{ij}}$ is the estimated linear predictor.

Computation of the QIC_w is straightforward. In the linear case, QIC_w can be directly computed from observed and fitted values; $\hat{Q_w}$ is a simple sum of squares. For generalized linear models, computation using a weighted sum of the components of the log likelihood requires additional work, but many software packages provide these components directly. They are also easily computed using, for example, SAS PROC IML (sample code in online appendix A). Care must be exercised when using computed AIC's from standard modeling packages, as statistical software packages may report unweighted AIC's, even from weighted data.

Linking QIC_w and QIC in the counterfactual data

Under certain assumptions, inverse-probability methods provide an approach for mapping functions of *observed data* into functions of *counterfactual data* that are equal in expectation. Here we give a heuristic proof that in the linear marginal structural model, the expectation of the inverse-probability weighted quasi-likelihood information criterion QIC_w for the observed data equals the expectation of a quasi-likelihood information criterion applied to the counterfactual data, provided the usual assumptions hold. For simplicity of exposition, we consider the case where each subject has two regularly spaced visits during which time treatment changes are made and measurements are recorded followed by the assessment of the outcome.

Let X be a treatment that can be applied at any point in time, x_1,x_2 be a sequence of treatments, Y_{x_1,x_2} be a counterfactual outcome corresponding to a sequence of treatments, and $S = Y_{x_1,x_2},(x_1,x_2) \in X^2$ be the set of counterfactual outcomes corresponding to all possible treatment sequences. Let X(t) denote the observed treatment at time t, $\bar{L}(t)$ denote the history of all covariates up to time t, and let $V \subset L(1)$ be some baseline covariates upon which we wish to condition. We are interested in estimating the conditional expectation of the counterfactual given V: $E[Y_{x_1,x_2} \mid V]$ If for each subject, we observed all counterfactual outcomes, then we could fit a model $m(x_1,x_2,V)$ of $E[Y_{x_1,x_2} \mid V]$ directly using regression without the need for weighting or other control for confounding. For example, our models above would be given by $m(x_1,x_2,V) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + V$. Given a set of competing models that have been fit to the data, $\hat{m_i}$, i = 1,…,I, it would be reasonable to select the optimal

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

model by choosing the one with the smallest QIC (note that in the linear case the estimating function reduces to the normal distribution log-likelihood):

$$\mathrm{QIC}(\hat{\mathrm{m}})=2\mathrm{p}-\frac{1}{\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{n}}\sum_{\mathrm{x}_1,\mathrm{x}_2\in\mathrm{X}^2}(\mathrm{Y}_{\mathrm{x}_1,\mathrm{x}_2,\mathrm{i}}-\hat{\mathrm{m}}(\mathrm{x}_1,\mathrm{x}_2,\mathrm{V}_\mathrm{i}))^2,$$

where p, as before, is the number of free parameters in the model.

With only the observed data, we instead chose the model that maximizes the inverse probability weighted quasi-likelihood information criterion:

$$\mathrm{QIC}_\mathrm{w}(\hat{\mathrm{m}})=2\mathrm{p}-\frac{1}{\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{n}}\sum_{\mathrm{x}_1,\mathrm{x}_2\in\mathrm{X}^2}\frac{(\mathrm{Y}_\mathrm{i}-\hat{\mathrm{m}}(\mathrm{X}(1)_\mathrm{i},\mathrm{X}(2)_\mathrm{i},\mathrm{V}_\mathrm{i}))^2}{\mathrm{P}(\mathrm{X}_\mathrm{i}(2)=\mathrm{x}_\mathrm{i}(2)|\overline{\mathrm{L}}_\mathrm{i}(2),\mathrm{x}_\mathrm{i}(1))\mathrm{P}(\mathrm{X}_\mathrm{i}(1)=\mathrm{x}_\mathrm{i}(1)|\mathrm{L}_\mathrm{i}(1))}$$

We would like for the weighted quasi-likelihood criterion to equal our counterfactual criterion QIC(m̂) in expectation.

To demonstrate this equality, we need to show that:

$$\begin{aligned}&\mathrm{E}[\frac{(\mathrm{Y}_\mathrm{i}-\hat{\mathrm{m}}(\mathrm{X}(1)_\mathrm{i},\mathrm{X}(2)_\mathrm{i},\mathrm{V})^2}{\mathrm{P}(\mathrm{X}_\mathrm{i}(2)=\mathrm{x}_\mathrm{i}(2)|\overline{\mathrm{L}}(2),\mathrm{X}(1))\mathrm{P}(\mathrm{X}_\mathrm{i}(1)=\mathrm{x}_\mathrm{i}(1)|\mathrm{L}(1))}]\\&=\mathrm{E}[\sum_{\mathrm{x}_1\in\mathrm{X}}\sum_{\mathrm{x}_2\in\mathrm{X}}(\mathrm{Y}_{\mathrm{x}_1,\mathrm{x}_2}-\hat{\mathrm{m}}(\mathrm{x}_1,\mathrm{x}_2,\mathrm{V}))^2]\end{aligned}.$$

By the rule of double expectation, we can write

$$\begin{aligned}&\mathrm{E}[\frac{(\mathrm{Y}_\mathrm{i}-\hat{\mathrm{m}}(\mathrm{X}(1)_\mathrm{i},\mathrm{X}(2)_\mathrm{i},\mathrm{V})^2}{\mathrm{P}(\mathrm{X}_\mathrm{i}(2)=\mathrm{x}_\mathrm{i}(2)|\overline{\mathrm{L}}(2),\mathrm{X}(1))\mathrm{P}(\mathrm{X}_\mathrm{i}(1)=\mathrm{x}_\mathrm{i}(1)|\mathrm{L}(1))}]\\&=\mathrm{E}[\mathrm{E}[\frac{(\mathrm{Y}_\mathrm{i}-\hat{\mathrm{m}}(\mathrm{X}(1)_\mathrm{i},\mathrm{X}(2)_\mathrm{i},\mathrm{V}))^2}{\mathrm{P}(\mathrm{X}_\mathrm{i}(2)=1|\overline{\mathrm{L}}(2),\mathrm{X}(1))\mathrm{P}(\mathrm{X}_\mathrm{i}(1)=\mathrm{x}_\mathrm{i}(1)|\mathrm{L}(1))}.\end{aligned}$$

The only random quantity left in the inside expectation is X(2), so we re-write the inner expectation as a sum over possible values of X(2):

$$=\mathrm{E}[\sum_{\mathrm{x}_2\in\mathrm{X}}\frac{\mathrm{Y}_{\mathrm{X}(1),\mathrm{x}_2}-\hat{\mathrm{m}}(\mathrm{X}(1),\mathrm{X}(2),\mathrm{V}))^2}{\mathrm{P}(\mathrm{X}_1(2)=\mathrm{x}_\mathrm{i}(2)|\mathrm{X}(1),\overline{\mathrm{L}}(2))\mathrm{P}(\mathrm{X}_\mathrm{i}(1)=\mathrm{x}_\mathrm{i}(1)|\mathrm{L}(1))}\mathrm{P}(\mathrm{X}(2)=\mathrm{x}_\mathrm{i}(2)|\mathrm{S},\mathrm{X}(1),\overline{\mathrm{L}}(2))].$$

Suppose that the assumption of exchangeability (i.e., no unmeasured confounders) holds and, for simplicity, that the distribution of X(2) is known. Then given the measured confounders, the treatment mechanism does not depend on any of the counterfactual outcomes. Therefore the last term in the product in the denominator cancels with the probability in the numerator.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Again, we apply double expectation to the resulting equation:

$$\mathrm{E}[\sum_{x_2 \in X} \frac{(Y_{X(1),x_2} - \hat{m}(X(1),x(2),V))^2}{P(X_i(1)=x_i(1)|L(1))}]$$
$$= \mathrm{E}[\mathrm{E}[\sum_{x_2 \in X} \frac{(Y_{X(1),x_2} - \hat{m}(X(1),x(2),V))^2}{P(X_i(1)=x_i(1)|L(1))} | S, \bar{L}(1)]].$$

The only random quantity left in the inside expectation is X(1), so we re-write the inner expectation as a sum over possible values of X(1) as:

$$\mathrm{E}[\sum_{x_1 \in X} \sum_{x_2 \in X} \frac{(Y_{x_1,x_2} - \hat{m}(x_1, x_2, V))^2}{P(X_i(1)=x_i(1)|L(1))} P(X(1)=x_i|S, \bar{L}(1))].$$

Again, assuming the distribution of X(1) is known and the exchangeability assumption holds, P(X(1) | $\bar{L}$(1)) = P(X(1) | S,$\bar{L}$(1)), then we have the result that E[QIC_w($\hat{m}$)] = E[QIC($\hat{m}$)]. Note that for identifiability, we require that P(X(1) = x(1) | L(1)) ≠ 0,$x_1 \in X$ and P(X(2) = x(2) | X(1), $\bar{L}$(2)) ≠ 0,$x_2 \in X$ (i.e., experimental treatment assumption/positivity). We note that the proof naturally extends to the generalized linear model with replacement of (Y-m(x,v))^2 by Q(y,x,v).

Simulation Study

We undertook a limited simulation study to evaluate some of the properties of our proposed QIC_w measure of model fit. We generated data assuming a marginal structural model with a continuous outcome and a binary time varying treatment with four time intervals. For each simulation, we generated treatment X and a single covariate L at four time points as follows: At time 1, we generated L_1 from a normal distribution with mean 10 and variance 1 and a binary treatment X_1 with probability of treatment P=expit(-2.7+0.25*L_0) (where expit(x)=exp(x)/(1+exp(x)), the inverse logit. The intercept was chosen to give a baseline exposure of 6.2%, in line with the proportion breastfed in example 1. At time i=2,3,4, L_i was generated from a normal distribution with mean $L_{i-1}+\beta_i*X_{i-1}$ and X_i as binary with probability of treatment P=expit(-2.6 +0.25*L_i+0.1*X_{i-1}). These give rise to a marginal structural model with E[Y($\hat{x}$)] = $\beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \beta_4X_4$.

We considered 7 scenarios, described in Table A1 in the online appendix, considering a range of effect sizes and of model complexity. We then fit three models: the “full” or “saturated” model fit binary indicators for each treatment-time, the “null” model fit a model with no parameters, and the “reduced” model fit binary indicators for treatment at times 1 and 2. We considered sample sizes of 100, 1000, 5000, and 10,000, and ran 1000 repetitions per scenario. In each simulation, we correctly specified the (known) model for the inverse probability weights. We ran a second set of simulations with the same specification but a log link function (so that the outcome distribution is lognormal and therefore heteroskedastic).

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Tables B1 and B2 in the online appendix summarize the results of the simulation study. For each combination, we report the percentage of simulation runs in which the QIC_w selected each model, and compare this with the similar results for the adjusted R-square. The QIC_w usually (but not uniformly) selected the correct model more often than did the adjusted R-square. However, QIC_w tended to select models with more parameters, while the adjusted R-square tended to select models with fewer parameters. Results for the lognormal model (Table B2) were similar; however the QIC performed worse with increasing sample size, likely due to misspecification of the variance.

Example One: Breastfeeding and infant weight gain

The PROBIT study used a cluster randomized breastfeeding promotion intervention, and followed 17,044 infants for weight gain over the first year of life [35], as well as infection [36] and long-term outcomes which are not considered here [37-42]. We follow Platt et al. [43] and consider the effect of breastfeeding during the first year of life on infant weight at 12 months.

We fit marginal structural linear models with four choices for the exposure metric. In all models we use (the same) minimally stabilized weights (i.e., the numerator for all weights is the marginal probability of observed exposure at that time) to focus on the marginal estimates of the effect of the exposure in the population. The first model, following Platt et al. [43], considered indicator variables for breastfeeding in each of the time intervals between visits, which occurred at 1, 2, 3, 6, 9, and 12 months to study the effect on weight at 12 months. In this study every mother-child pair initiates breastfeeding, breastfeeding is assumed not to restart once stopped and data were only collected on these intervals. Therefore, in this model the exposure specification is “saturated”; that is, each possible exposure history completely determines the expected value of the potential outcome (infant weight at 12 months) and the model has the maximum number of free parameters. The second, third, and fourth exposure metrics were developed to examine simplifications relative to the exposure-saturated model, and considered cumulative breastfeeding; that is, weight was modeled as a linear, quadratic, and cubic function of the cumulative number of prior months breastfed. All four models are plausible given the underlying biology of breastfeeding and infant weight gain in the first year of life; our goal here is to compare these four models using the modified QIC_w described above. A graphical depiction of results from the saturated marginal structural models is provided in Figure 1. Summary statistics including QIC_w for the four models are presented in Table 1.

The exposure-saturated model suggests the effect of breastfeeding cessation on weight at 12 months was strongest for breastfeeding cessation between 6 and 9 months and relatively small before or thereafter. The simple linear cumulative breastfeeding model suggests that weight is lower by 25 g per month breastfed at 12 months. The quadratic model provided similar inferences to the linear model, except for the shortest durations, while the cubic model fit closely resembled that of the exposure-saturated model. The QIC_w was lowest for the cubic model, slightly higher for the exposure-saturated model and highest for the linear and quadratic models. Standard errors of individual coefficients (not shown) were similar

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

across the models explored. Figure B1 (online appendix) shows a plot of residuals from the four proposed models; no patterns are apparent and all models showed similar results.

Example Two: Effect of Highly Active Antiretroviral Therapy on CD4 cell count

Cole et al. reported on the effect of highly active antiretroviral therapy on the evolution of CD4-positive T-lymphocyte (CD4 cell) count among human immunodeficiency virus (HIV)-positive participants using inverse probability-of-treatment-and-censoring weighted estimation of a marginal structural repeated measures model [23]. In summary, 60% of 1,763 eligible participants from two US cohort studies followed between 1996 and 2002 initiated highly active antiretroviral therapy. For further details about study design and analysis please see reference [23]. As reported, the weighted estimate of the difference in mean CD4 cell count at 1 year among participants continuously treated versus those never treated was 71 cells/mm^3 (95% confidence limits: 48, 95) with an estimated continued increase of 29 cells/mm^3 per year (estimated effect at 6 years: 216 cells/mm^3). Here, to facilitate graphical depiction of the data we again use minimally stabilized weights, which stabilize only on the history of exposure and time, rather than additionally stabilizing on select baseline covariates, as was done in the reported analysis. Therefore, the 1 year and per year greater than 1 year differences in CD4 cell count due to treatment were 80 (95% CL: 45, 115) and 16 cells/mm^3 per year, respectively.

Figure 2 presents the difference in CD4 cell count between exposed and unexposed groups as well as a point-wise 95 percent point-wise confidence band in grey-shade. Our motivation for a 2-piece linear spline is apparent in Figure 2, where the slope of the difference is relatively steep in the first year, and then appears less steep in the second through fifth years, with some apparent instability in the final year.

A comparison of several candidate exposure specifications is provided in Table 2. We compared (1) an intercept-only model; (2) a model with no effect for exposure but an intercept, a spline for follow up time and centered baseline CD4 count; (3) model 2 with a linear effect for cumulative exposure; (4) model 2 with a curvilinear effect for cumulative exposure; (5) model 2 with a 2-piece linear spline for cumulative exposure; and (6) an unsmoothed model, which (akin to the exposure-saturated model in example one) allowed the difference in CD4 cell count to vary nonparametrically at each semiannual study visit. In the original report, based on the combination of biological knowledge [23], prior literature [44], and the apparent inflection at one year in Figure 2, the authors presented Model 5, including the two-piece linear spline for cumulative exposure. In Table 2, we demonstrate that this model (5) had the lowest QIC_W among the models considered. However, the curvilinear model (4) fit these data nearly as well as the 2-part linear spline model (5). Even a simpler linear model (3) was not wholly unreasonable; this can be checked visually by drawing a straight line through the point-wise confidence band in Figure 2. Figure B2 (online appendix) shows a plot of residuals from the six proposed models; all models showed similar results and no patterns are apparent.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Discussion

In most regression modeling settings, appropriate specification of the functional form of the regression model is essential to valid inference [45]. These standard considerations apply to model selection in marginal structural models. To date, most implementations of marginal structural models have used simple forms of the exposure (e.g., always vs. never exposed, cumulative exposure) for the final structural regression model and have not discussed model selection. Here, we have proposed an information criterion for use in assessing model fit in marginal structural models. To our knowledge this is the first use of such a criterion in this context, and these statistics may play an important role in the practical usage of marginal structural models.

In the limited simulation study, we demonstrated that our QICw performs better than does the adjusted R-squared. When the model was fully parameterized, the QICw almost always selected the correct model, while the adjusted R-squared usually selected a reduced model. When a reduced model was correct, the QICw tended to select either the full or reduced model, almost never under-specifying the model.

In the breastfeeding example, more detailed inferences were uncovered when the functional form of exposure was explored; fully-specified models, linear models, and curvilinear models fit the data relatively similarly but the fully-specified (exposure-saturated) model showed important patterns that were not evident based on simpler models, providing insight into the association between breastfeeding and growth. In the CD4 example, alternative and simpler functional forms fit the data nearly as well as the published two-part linear model.

Altering the functional form of the regression model, including the exposure metric, may change the scientific question being asked. Substantive expertise should drive selection of the functional form if possible, as biological plausibility of the functional form can lend credence to the results. However, knowledge can be gained when differing functional forms provide different fits to the data. For instance, if models specifying different exposure windows give substantially different fits this can inform the investigator about the critical windows of exposure. Exposure-saturated models similar to those fit in example one may identify thresholds or jumps in the effect. In particular, as in the breastfeeding example, the saturated model suggests a jump between 6 and 9 months of breastfeeding that was not evident in the linear or curvilinear models.

Our proposed QIC_w has important limitations. In particular, the validity of decisions based on the QIC_w is dependent on correct specification of the variance and weight models. If either of these is incorrectly specified, our QIC_w may not yield valid model comparisons. Alternative approaches to model selection exist. Here in both examples, we concentrated on a graphical depiction of a fairly flexible function form and several competing simpler models. In addition to the graphical depictions we also compared QIC_w across the specified models. Brookhart and van der Laan [20] and van der Laan and Dudoit [46] propose approaches based on cross-validation that can simultaneously select variables for the weight and structural models. These approaches depend on an initial specification of the weight model that yields consistent but potentially highly variable estimates of a parameter of

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

interest. Cross-validation is then used to optimize model specification. Here we concentrated on optimizing the structural model with weights held fixed. Neugebauer et al.[21] proposed comparisons against a highly flexible structural model, but their approach is technically demanding. Doubly-robust methods [47] are robust to misspecification of the weight model or the outcome model (though not both), and provide an opportunity for future research. More work comparing these methods to our proposed QIC_w is warranted.

Three other points should be noted regarding specification of marginal structural models. First, the specification of the functional form of the exposure in the final structural model must be equally coarse as, or coarser than, the exposure specification in the weight model. This ensures that control of confounding is as complete as possible given the observed covariates. In both examples, we estimated exposure weights based on the exposure at each time point; therefore the exposure in each structural model was specified equally as coarse (in the case of the saturated model) or more coarse (in the case of the linear and curvilinear models) than how the exposure was specified in the weight model.

Second, many previous applications of marginal structural models have made an (explicit or implicit) observational intent-to-treat assumption. Namely, it is assumed that once exposed, participants remain exposed for the remainder of their observed person-time. This assumption was made in both examples presented here (in the breastfeeding study, a converse intent-to-treat assumption was made, in which participants were assumed not to restart once stopped). The assumption almost certainly holds in the breastfeeding study for biological reasons [36]. In the CD4 example, this assumption holds for 86% of the person-time [23]. This assumption simplifies the exposure model so that only a single model for exposure onset is required, rather than separate models for onset, cessation, and resumption; likewise, it dramatically simplifies potential exposure histories, leading to simpler specification of this covariate as well. However, in cases where exposure is intermittent, this assumption may increase rather than reduce net bias. In such settings one must consider the potential tradeoff between correctly specifying a multiphase exposure process and increasing the variance of the final estimate of association. If exposure is intermittent, then a correctly specified exposure process may give estimates with high variance because relatively few participants will follow each possible exposure plan; on the other hand, an observational intent-to-treat assumption (or other simplifications that represent a coarsening of the exposure process) may provide biased estimates of the exposure effect, but have relatively low variance.

Finally, we restricted attention here to minimally stabilized models, so as to focus attention on exposure specification alone, to facilitate graphical depictions of the data, and to allow for marginal interpretations of effect estimates. In other applications of marginal structural models, however, stabilization by baseline covariates may improve the properties of the weights at a cost of complicating the final structural model [48, 49].

This paper presents a relatively simple approach to specification of exposure metrics in marginal structural models; other, more complex model fitting approaches could be considered. For instance, in both examples the outcome was a continuous measure. In such cases, one might want to explore transformations of the outcome variable [50], though this

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

may raise the issue of the re-transformation problem [51]. In cases where the outcome is a time-to-event variable, analogously, one could consider a broad family of parametric survival models [52] or a semiparametric accelerated failure time model [53, 54] as alternatives to Cox proportional hazards model [55]. In cases where the outcome is a dichotomous variable, the choice of link function (e.g., logit, log-linear, probit, identity) may provide different inferences; in part because different contrasts are made (e.g., ratio versus difference).

We illustrate the importance of correct specification of the structural form when fitting marginal structural models, and a candidate approach to this process. Additional work on structural model specification is needed, particularly with respect to simultaneous estimation of the inverse probability weights and the final structural model. In marginal structural models, as in any regression setting, correct inference often depends on correct model specification.

Acknowledgments

Dr. Platt was supported by a Chercheur-Boursier (Research Scholar) award from the Fonds de la Recherche en Sante du Quebec (FRSQ), by core support to the McGill University Health Centre Research Institute from the FRSQ, and by the Canadian Network for Observational Drug Effect Studies (CNODES). Dr. Cole was supported in part by NIH grant R01-AA-01759. Dr. Brookhart is supported by a career development award from NIH (AG027400). Dr. Westreich received support from NIH/NICHD R00-HD-06-3961. Dr Schisterman is supported by the intramural program of the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development

Online Appendix A

Here we provide a brief SAS program to estimate the QIC for the saturated linear model for example 1.

```
**Assume data as follows;

**y outcome;

**x1--x12 dummy variables for exposure to months 1, 2, 3, 6, 9, 12 (i.e.,
saturated model);

*Saturated model;

proc genmod data=a;

class i;

model y=x1 x2 x3 x6 x9 x12/d=n;

weight w;

repeated subject=i/type=ind;

title “Saturated”;
```

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

```
proc nlmixed data=a hess;

xb=b0+b1*x1+b2*x2+b3*x3+b4*x6+b5*x9+b6*x12;

logL=-.5*w*(y-xb)**2;

model y~general(logL);

predict logL out=score der;

ods output parameterestimates=parms hessian=hess;

ods select specifications fitstatistics parameterestimates;

*logL is the QIC;

*Below is code to estimate the robust variance matrix from PROC NLMIXED

* (provided automatically by PROC GENMOD with REPEATED option);

proc sort data=score; by i;

proc means data=score noprint; by i;

var der_b0 der_b1-der_b6;

output out=_out1_(keep=der_b0 der_b1-der_b6) sum=der_b0 der_b1-der_b6;

data _out1_; set; array d (*) der_b0 der_b1-der_b6; if d(1)=.then delete;

data _parms_(keep=estimate); set parms; if parameter^=“sigma”;

data _hess_(keep=b0 b1-b6); set hess; if parameter^=“sigma”;

proc iml;

use _parms_; read all into b;

use _hess_; read all into hess;

cov=inv(hess);

nse=sqrt(vecdiag(cov));

use _out1_; read all into x;

v=cov*(x`*x)*cov;
```

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

```
rse=sqrt(vecdiag(v));

wald=(b/rse)##2;

p=1-probchi(wald,1);

chi=wald||p;

c={“Chi Square” “p-value”};

reset noname fuzz=.000001;

print, “Standard Variance Matrix”,, cov;

print, “Robust Variance Matrix”,, v;

print, “Robust Standard Errors”,, rse;

print, “Wald Statistics”,, chi;

run;

quit;

run;
```

Online Appendix B

Table B1 shows the results of the simulation study for the runs in which the linear marginal structural model was correctly specified with the parameters listed.

Table B2 shows the results of the simulation study for the runs in which the linear marginal structural model was incorrectly specified (with a lognormal model) and the parameters listed.

Figure B1 shows the standardized Pearson residuals for the best fitting model for example 1, with saturated model. The mean residual was -0.001, with a standard deviation of 0.989; 95 of 17,044 observations had residuals greater than |3|, while we would expect 44.9 such outlying observations based on the standard normal distribution. No systematic patterns were apparent.

Figure B2 shows the standardized Pearson residuals for the best fitting model, with a piecewise linear spline. The mean residual was 0.007, with a standard deviation of 1.0013; 161 of 12,035 observations had residuals greater than |3|, while we would expect 31.3 such outlying observations based on the standard normal distribution. No systematic patterns were apparent.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table B1

Results of simulation study, correctly specified model. For each combination, values represent the percentage of runs in which each model-selection approach selected each model.

		Percentage of simulation runs each model selected					
		Saturated		Reduced		Simple	
True model (β1,β2,β3,β4)	Sample size	QIC_w	Adj R^2	QIC_w	Adj R^2	QIC_w	Adj R^2
(0,0,0,0)	200	70.9	29.4	11.4	56.9	17.7	13.7
	1,000	52.6	20.3	11.1	69.2	36.3	10.5
	5,000	49.4	19.8	13.3	70.6	37.3	9.6
	10,000	46.5	19.8	11.1	70.5	42.4	9.7
(1,1,1,1)	200	58.5	20.4	15.4	63.7	26.1	15.9
	1,000	12.9	0	12.3	100	74.8	0
	5,000	0.4	0	0.9	100	98.7	0
	10,000	0	0	0	100	100	0
(2,2,2,2)	200	79.5	4.3	10	93.7	10.5	2
	1,000	79.7	0	12.8	100	7.5	0
	5,000	87.6	0	10.6	100	1.8	0
	10,000	93.5	0	5.8	100	0.7	0
(3,3,3,3)	200	97.8	0	1.4	100	0.8	0
	1,000	99.5	0	0.3	100	0.2	0
	5,000	100	0	0	100	0	0
	10,000	69.5	8	11.7	43.3	18.8	48.7
(1,1,0,0)	200	53.7	0	13.1	44.1	33.2	55.9
	1,000	47.5	0	14.3	45.2	38.2	54.8
	5,000	50.2	0	16	50.6	33.8	49.4
	10,000	71.4	0.5	14.7	43.4	13.9	56.1
(2,2,0,0)	200	61.2	0	30.8	41.5	8	58.5
	1,000	56.2	0	42.1	46.3	1.7	53.7
	5,000	55.8	0	44	46.7	0.2	53.3
	10,000	77.7	0	17.2	42.5	5.1	57.5
(3,3,0,0)	200	60.7	0	38.3	46.2	1	53.8
	1,000	58.7	0	41.3	44.8	0	55.2
	5,000	55.9	0	44.1	47.9	0	52.1
	10,000	70.9	29.4	11.4	56.9	17.7	13.7

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table B2

Results of simulation study, mis-specified model. For each combination, values represent the percentage of runs in which each model-selection approach selected each model.

		Percentage of simulation runs each model selected					
		Saturated		Reduced		Simple	
True model (β1,β2,β3,β4)	Sample size	QIC_w	Adj R^2	QIC_w	Adj R^2	QIC_w	Adj R^2
(0,0,0,0)	200	68.3	25.2	12.1	51.2	19.6	23.6
	1,000	57.1	29.2	12.6	49.4	30.3	21.4
	5,000	60.6	32.3	10.2	47.9	29.2	19.8
	10,000	56.8	40.5	15	43.7	28.2	15.8
(1,1,1,1)	200	51.5	28.4	14.8	49.8	33.7	21.8
	1,000	24.5	25.9	11.8	60	63.7	14.1
	5,000	16	15	10	78.4	74	6.6
	10,000	12.2	8.8	9.1	88.8	78.7	2.4
(2,2,2,2)	200	39.7	24.5	17.7	59.1	42.6	16.4
	1,000	8.2	13.1	8	83.7	83.8	3.2
	5,000	2.1	4.6	4.2	95.1	93.7	0.3
	10,000	1.2	1.7	3.1	97.9	95.7	0.4
(3,3,3,3)	200	33.9	26.5	16.9	62.7	49.2	10.8
	1,000	2.4	7.8	5.9	90.7	91.7	1.5
	5,000	0.4	2.8	2.1	97.1	97.5	0.1
	10,000	0	0.7	1.8	99.1	98.2	0.2
(1,1,0,0)	200	67.5	18.4	12.3	54.6	20.2	27
	1,000	56.3	12.8	11.7	53.5	32	33.7
	5,000	55.3	7.6	12.8	60.8	31.9	31.6
	10,000	53.8	3.6	13.8	60.6	32.4	35.8
(2,2,0,0)	200	63.8	10.9	13.8	45.9	22.4	43.2
	1,000	56.9	3.9	13.2	49.6	29.9	46.5
	5,000	49.3	0.6	18.8	61.6	31.9	37.8
	10,000	46.7	0.1	18.9	65.1	34.4	34.8
(3,3,0,0)	200	63.3	4.7	18.2	36.1	18.5	59.2
	1,000	48	1	24.4	46	27.6	53
	5,000	37.4	0	31.9	60.4	30.7	39.6
	10,000	36.6	0	38	65.9	25.4	34.1

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

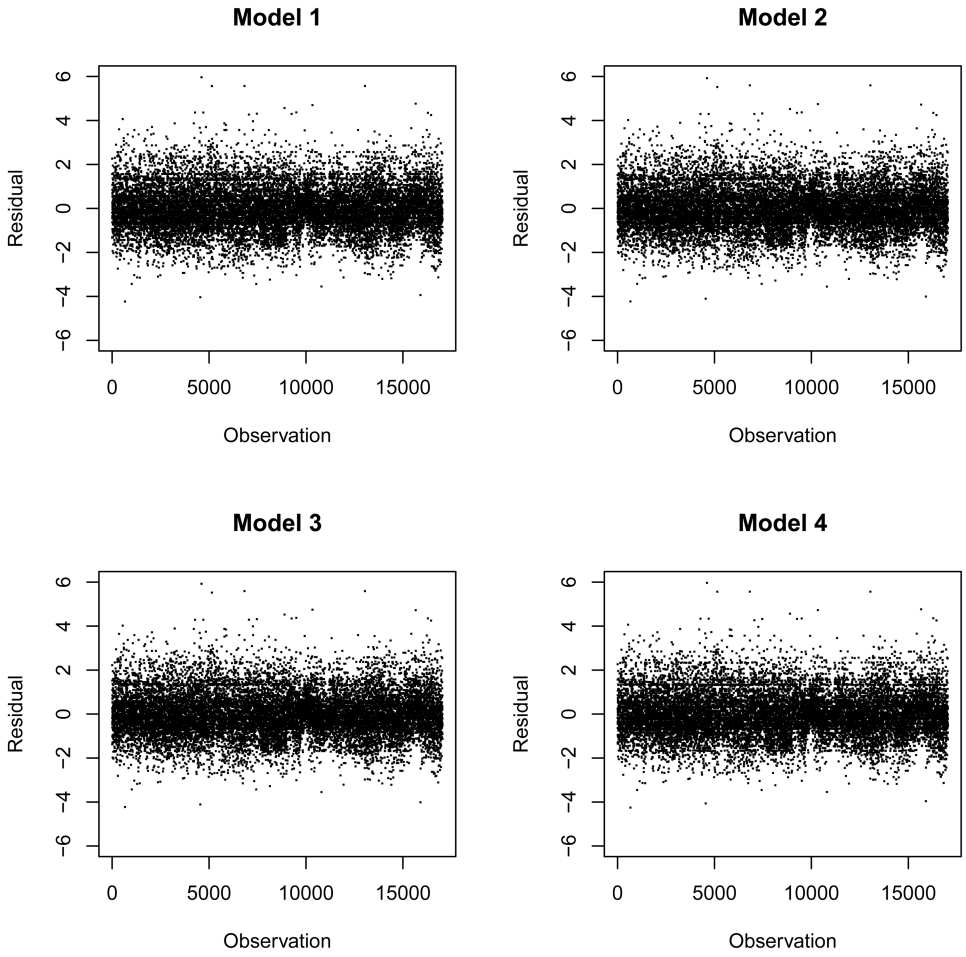

Figure B1.
Residuals for the four candidate models for Example 1, plotted against observation number.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

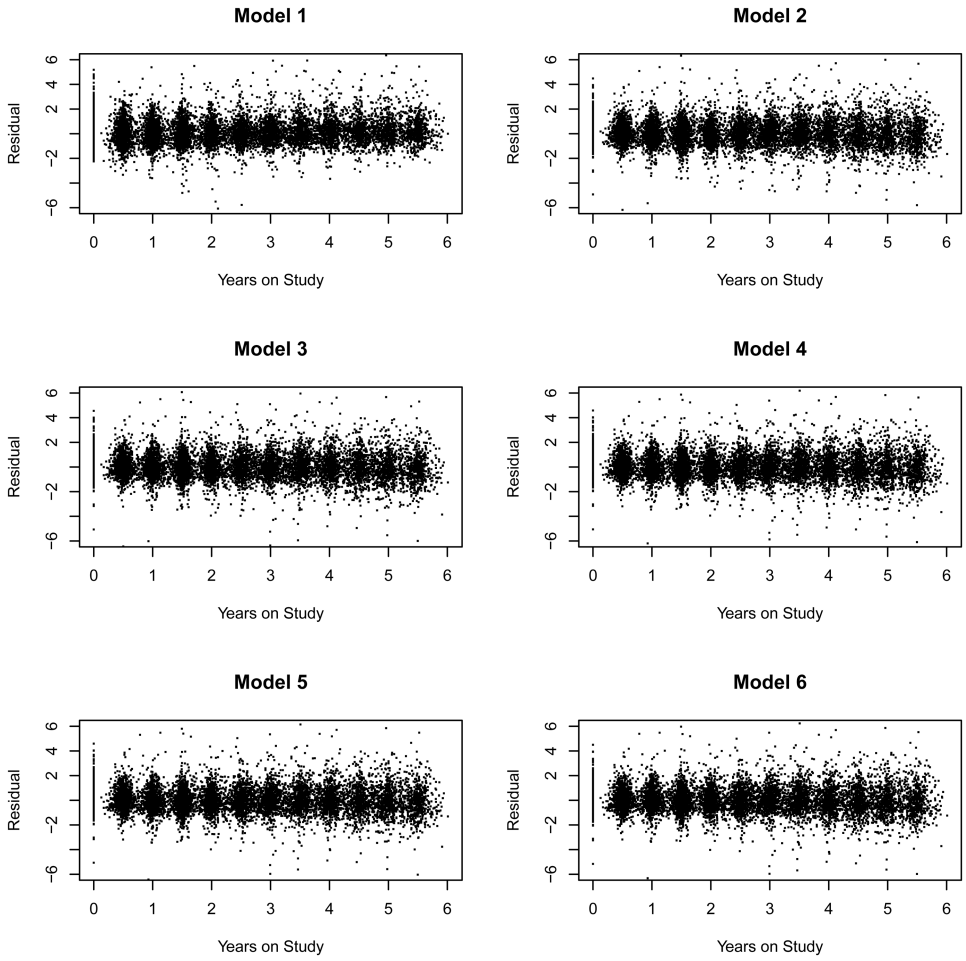

Figure B2.
Residuals for the six candidate models for Example 2, plotted against years on study.

References

1. Robins JM. Marginal structural models. 1997 Proceedings of the American Statistical Association. 1998:1–10. Section on Bayesian Statistical Science.
2. Robins, JM. Marginal structural models versus structural nested models as tools for causal inference. In: Halloran, ME.; Berry, D., editors. Marginal Structural Models Versus Structural Nested Models as Tools for Causal Inference. Springer; New York: 1999. p. 95-134.
3. Robins JM. Association, causation and marginal structural models. Synthese. 1999; 121:151–179.
4. Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. Epidemiology. 2000; 11:550–560. [PubMed: 10955408]
5. Choi HK, Hernán MA, Seeger JD, Robins JM, Wolfe F. Methotrexate and mortality in patients with rheumatoid arthritis: a prospective study. Lancet. 2002; 359:1173–1177. [PubMed: 11955534]
6. Sterne JA, Hernán MA, Ledergerber B, Tilling K, Weber R, Sendi P, Rickenbach M, Robins JM, Egger M. Long-term effectiveness of potent antiretroviral therapy in preventing AIDS and death: a prospective cohort study. Lancet. 2005; 366:378–384. [PubMed: 16054937]
7. MacKenzie EJ, Rivara FP, Jurkovich GJ, Nathens AB, Frey KP, Egleston BL, Salkever DS, Scharfstein DO. A national evaluation of the effect of trauma-center care on mortality. N Engl J Med. 2006; 354:366–378. [PubMed: 16436768]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

8. Cotter D, Zhang Y, Thamer M, Kaufman J, Hernan MA. The effect of epoetin dose on hematocrit. Kidney Int. 2008; 73:347–353. [PubMed: 18004296]
9. Palella FJ Jr, Armon C, Buchacz K, Cole SR, Chmiel JS, Novak RM, Wood K, Moorman AC, Brooks JT. The association of HIV susceptibility testing with survival among HIV-infected patients receiving antiretroviral therapy: a cohort study. Ann Intern Med. 2009; 151:73–84. [PubMed: 19620160]
10. Hernán MA, Taubman SL. Does obesity shorten life? The importance of well-defined interventions to answer causal questions. International Journal of Obesity. 2008; 32:S8–14. [PubMed: 18695657]
11. Cole SR, Frangakis CE. The consistency statement in causal inference: a definition or an assumption? Epidemiology. 2009; 20:3–5. [PubMed: 19234395]
12. VanderWeele TJ. Concerning the consistency assumption in causal inference. Epidemiology. 2009; 20:880–883. [PubMed: 19829187]
13. Mortimer KM, Neugebauer R, van der Laan M, Tager IB. An application of model-fitting procedures for marginal structural models. Am J Epidemiol. 2005; 162:382–388. [PubMed: 16014771]
14. Hernan MA, Robins JM. Estimating causal effects from epidemiological data. J Epidemiol Community Health. 2006; 60:578–586. [PubMed: 16790829]
15. Cole SR, Hernán MA. Constructing inverse probability weights for marginal structural models. Am J Epidemiol. 2008; 168:656–664. [PubMed: 18682488]
16. Westreich D, Cole SR. Invited Commentary: Positivity in Practice. American Journal of Epidemiology. 2010; 171:674–677. [PubMed: 20139125]
17. Lefebvre G, Delaney JAC, Platt RW. Impact of mis-specification of the treatment model on estimates from a Marginal Structural Model. Statistics in Medicine. 2008; 27:3629–3642. [PubMed: 18254127]
18. Moodie EE. Risk factor adjustment in marginal structural model estimation of optimal treatment regimes. Biom J. 2009; 51:774–788. [PubMed: 19816876]
19. Harrell, FE. Regression Modelling Strategies. Springer-Verlag; New York, NY: 2001.
20. Brookhart MA, van der Laan MJ. A semiparametric model selection criterion with applications to the marginal structural model. Computational Statistics & Data Analysis. 2006; 50:475–498.
21. Neugebauer R, Van der Laan M. Nonparametric causal effects based on marginal structural models. Journal of Statistical Planning and Inference. 2007; 137:419–434.
22. Hens N, Aerts M, Molenberghs G. Model selection for incomplete and design-based samples. Stat Med. 2006; 25:2502–2520. [PubMed: 16596577]
23. Cole SR, Hernán MA, Margolick JB, Cohen MH, Robins JM. Marginal structural models for estimating the effect of highly active antiretroviral therapy initiation on CD4 cell count. Am J Epidemiol. 2005; 162:471–478. [PubMed: 16076835]
24. Moodie EEM, Platt RW, Kramer MS. Estimating response-maximized decision rules with applications to breastfeeding. Journal of the American Statistical Association. 2009; 104:155–165.
25. Hernan MA. A definition of causal effect for epidemiological research. J Epidemiol Community Health. 2004; 58:265–271. [PubMed: 15026432]
26. Holland PW. Statistics and Causal Inference. Journal of the American Statistical Association. 1986; 81:945–960.
27. Robins J, Orellana L, Rotnitzky A. Estimation and extrapolation of optimal treatment and testing strategies. Stat Med. 2008; 27:4678–4721. [PubMed: 18646286]
28. Hernan MA, Lanoy E, Costagliola D, Robins JM. Comparison of dynamic treatment regimes via inverse probability weighting. Basic Clin Pharmacol Toxicol. 2006; 98:237–242. [PubMed: 16611197]
29. Cain LE, Robins JM, Lanoy E, Logan R, Costagliola D, Hernán MA. When to Start Treatment? A Systematic Approach to the Comparison of Dynamic Regimes Using Observational Data. International Journal of Biostatistics. 2010; 6

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

30. Robins JM, Finkelstein DM. Correcting for noncompliance and dependent censoring in an AIDS Clinical Trial with inverse probability of censoring weighted (IPCW) log-rank tests. Biometrics. 2000; 56:779–788. [PubMed: 10985216]
31. Hernan 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. [PubMed: 10955409]
32. Moodie EE, Delaney JA, Lefebvre G, Platt RW. Missing Confounding Data in Marginal Structural Models: A Comparison of Inverse Probability Weighting and Multiple Imputation. International Journal of Biostatistics. 2008; 4
33. Akaike, H. Information Theory and an Extension of the Maximum Likelihood Principal. Akademia Kiado; City: 1973. Information theory and an extension of the maximum likelihood principal. Editor (eds)
34. Pan W. Akaike's information criterion in generalized estimating equations. Biometrics. 2001; 57:120–125. [PubMed: 11252586]
35. Kramer MS, Guo T, Platt RW, Shapiro S, Collet JP, Chalmers B, Hodnett E, Sevkovskaya Z, Dzikovich I, Vanilovich I. Breastfeeding and infant growth: biology or bias? Pediatrics. 2002; 110:343–347. [PubMed: 12165588]
36. Kramer MS, Chalmers B, Hodnett ED, Sevkovskaya Z, Dzikovich I, Shapiro S, Collet JP, Vanilovich I, Mezen I, Ducruet T, Shishko G, Zubovich V, Mknuik D, Gluchanina E, Dombrovskiy V, Ustinovitch A, Kot T, Bogdanovich N, Ovchinikova L, Helsing E. Promotion of Breastfeeding Intervention Trial (PROBIT): a randomized trial in the Republic of Belarus. JAMA. 2001; 285:413–420. [PubMed: 11242425]
37. Kramer MS, Aboud F, Mironova E, Vanilovich I, Platt RW, Matush L, Igumnov S, Fombonne E, Bogdanovich N, Ducruet T, Collet JP, Chalmers B, Hodnett E, Davidovsky S, Skugarevsky O, Trofimovich O, Kozlova L, Shapiro S. Breastfeeding and child cognitive development: new evidence from a large randomized trial. Arch Gen Psychiatry. 2008; 65:578–584. [PubMed: 18458209]
38. Kramer MS, Fombonne E, Igumnov S, Vanilovich I, Matush L, Mironova E, Bogdanovich N, Tremblay RE, Chalmers B, Zhang X, Platt RW. Effects of prolonged and exclusive breastfeeding on child behavior and maternal adjustment: evidence from a large, randomized trial. Pediatrics. 2008; 121:e435–440. [PubMed: 18310164]
39. Kramer MS, Matush L, Vanilovich I, Platt R, Bogdanovich N, Sevkovskaya Z, Dzikovich I, Shishko G, Mazer B. Effect of prolonged and exclusive breast feeding on risk of allergy and asthma: cluster randomised trial. BMJ. 2007; 335:815. [PubMed: 17855282]
40. Kramer MS, Matush L, Vanilovich I, Platt RW, Bogdanovich N, Sevkovskaya Z, Dzikovich I, Shishko G, Collet JP, Martin RM, Davey Smith G, Gillman MW, Chalmers B, Hodnett E, Shapiro S. Effects of prolonged and exclusive breastfeeding on child height, weight, adiposity, and blood pressure at age 6.5 y: evidence from a large randomized trial. Am J Clin Nutr. 2007; 86:1717–1721. [PubMed: 18065591]
41. Kramer MS, Matush L, Vanilovich I, Platt RW, Bogdanovich N, Sevkovskaya Z, Dzikovich I, Shishko G, Collet JP, Martin RM, Smith GD, Gillman MW, Chalmers B, Hodnett E, Shapiro S. A randomized breast-feeding promotion intervention did not reduce child obesity in Belarus. J Nutr. 2009; 139:417S–421S. [PubMed: 19106322]
42. Kramer MS, Vanilovich I, Matush L, Bogdanovich N, Zhang X, Shishko G, Muller-Bolla M, Platt RW. The effect of prolonged and exclusive breast-feeding on dental caries in early school-age children. New evidence from a large randomized trial. Caries Res. 2007; 41:484–488. [PubMed: 17878730]
43. Platt RW, Schisterman EF, Cole SR. Time-modified confounding. Am J Epidemiol. 2009; 170:687–694. [PubMed: 19675141]
44. Tarwater PM, Margolick JB, Jin J, Phair JP, Detels R, Rinaldo C, Giorgi J, Munoz A. Increase and plateau of CD4 T-cell counts in the 3(1/2) years after initiation of potent antiretroviral therapy. J Acquir Immune Defic Syndr. 2001; 27:168–175. [PubMed: 11404539]
45. Maldonado G, Greenland S. Interpreting model coefficients when the true model form is unknown. Epidemiology. 1993; 4:310–318. [PubMed: 8347741]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

46. Van der Laan, M.; Dudoit, S. Unified cross-validation methodology for selection among estimators: Finite sample results, asymptotic optimality, and applications. UC Berkeley; City: 2003. Unified cross-validation methodology for selection among estimators: Finite sample results, asymptotic optimality, and applications., Editor (ed)ˆ(eds)
47. Bang H, Robins JM. Doubly robust estimation in missing data and causal inference models. Biometrics. 2005; 61:962–973. [PubMed: 16401269]
48. Hernan MA, Alonso A, Logan R, Grodstein F, Michels KB, Willett WC, Manson JE, Robins JM. Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease. Epidemiology. 2008; 19:766–779. [PubMed: 18854702]
49. Westreich D, Cole SR, Tien PC, Chmiel JS, Kingsley L, Funk MJ, Anastos K, Jacobson LP. Time scale and adjusted survival curves for marginal structural cox models. American Journal of Epidemiology. 2010; 171:691–700. [PubMed: 20139124]
50. Box GEP, Cox DR. An analysis of transformations. Journal of the Royal Statistical Society, Series B. 1964; 26:211–252.
51. Taylor JMG. The retransformATION mean after a power transformation. Journal of the American Statistical Association. 1986; 81:114–118.
52. Cox C. The generalized F distribution: an umbrella for parametric survival analysis. Stat Med. 2008; 27:4301–4312. [PubMed: 18407568]
53. Wei LJ. The Accelerated Failure Time Model: A Useful Alternative to the Cox Regression Model in Survival Analysis. Statistics in Medicine. 1992; 11:1871–1879. [PubMed: 1480879]
54. Cole SR, Chu H, Lei N. Nonparametric estimator of relative time with application to the Acyclovir Prevention Trial. Clinical Trials. 2009; 6:320–328. [PubMed: 19667028]
55. Cox DR. Regression Models and Life-Tables. Journal of the Royal Statistical Society Series B-Statistical Methodology. 1972; 34:187–&.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

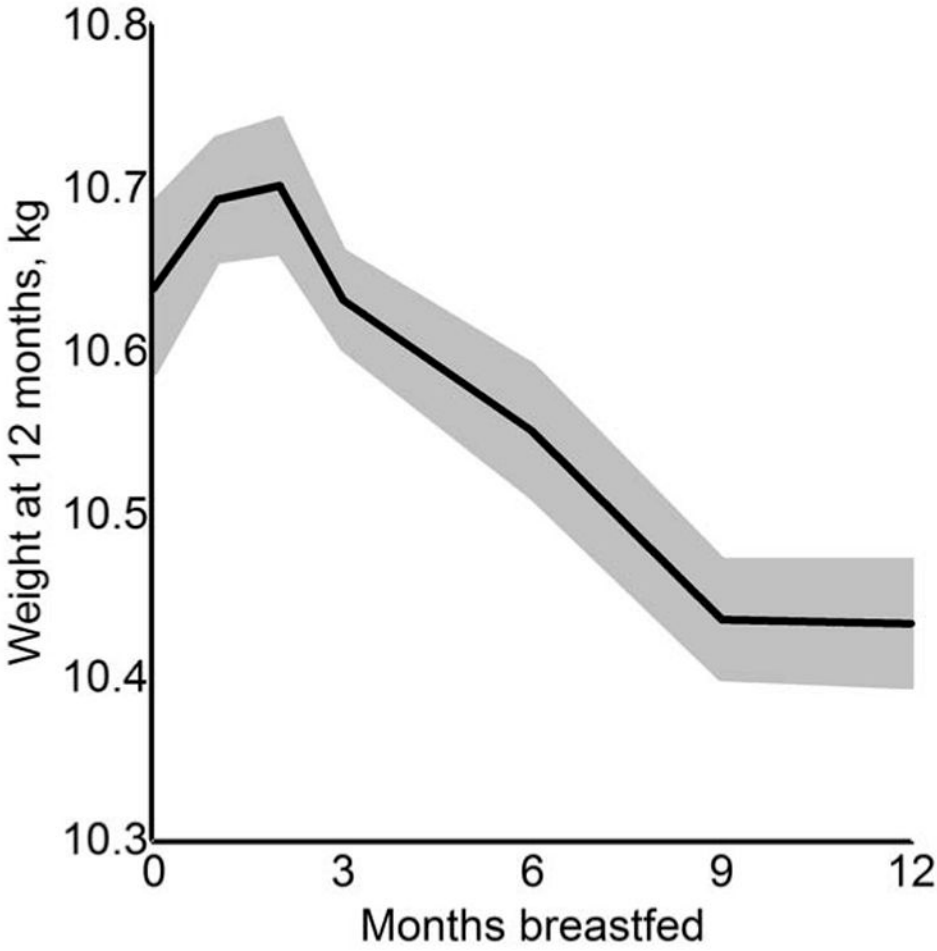

Figure 1.

Fitted mean weight at 12 months (kg) as a function of months breastfed, for exposure saturated model with 95 percent point-wise confidence bands in grey-shade for 17,044 infants in a randomized breastfeeding promotion trial conducted in Belarus.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

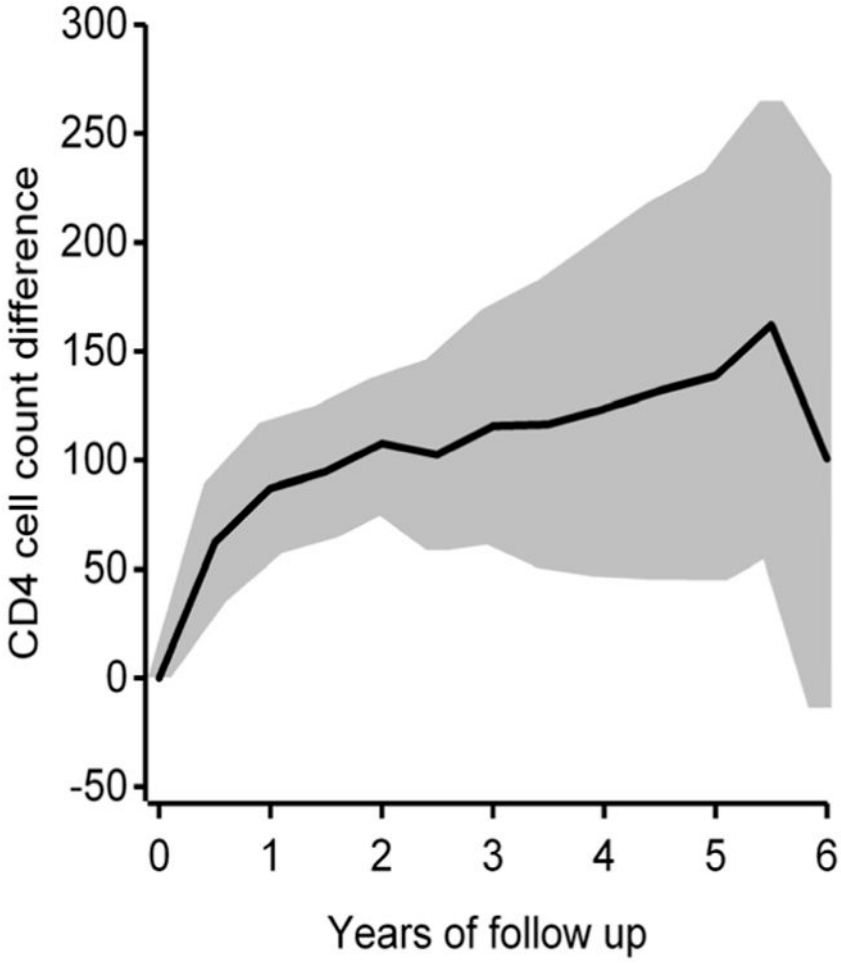

Figure 2.

Difference in CD4 count (cells/mm^3) for 5519 antiretroviral therapy exposed and 6516 unexposed visits between April 1996 and April 2002 for 1,763 HIV-positive participants from two US cohort studies, by years of follow up using under-stabilized inverse probability-of-treatment-and-censoring weights as described in text and point-wise 95 percent confidence band in grey-shade.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

Table 1

Estimated average infant weight in kilograms at 12 months and weighted QIC for four fitted marginal structural models. Estimates for 1 and 2 months not shown.

Model:	Breastfeeding until month N compared to early weaning				No. parameters	Weighted QIC
	3 months	6 m	9 m	12 m		
1. Exposure saturated	10.631	10.550	10.435	10.433	7	16,776
2. Linear exposure	10.629	10.554	10.478	10.403	2	16,784
3. Quadratic exposure	10.627	10.548	10.476	10.409	3	16,786
4. Cubic exposure	10.653	10.542	10.435	10.433	4	16,775

Stat Med. Author manuscript; available in PMC 2014 September 30.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

Table 2

Estimates from several marginal structural models of the association of antiretroviral therapy on difference in CD4 cell count among 1,763 HIV-positive participants from two US cohort studies, April 1996–April 2002.

	Model-predicted difference in CD4 cell count at:				
	1 year	3 years	5 years	No. parameters	Weighted QIC
Model:					
1. Intercept	0	0	0	1	931.77
2. Intercept and time [a]	0	0	0	5	496.94
3. Model 2 + linear exposure	34	102	171	6	482.11
4. Model 2 + curvilinear exposure	57	125	134	7	481.57
5. Model 2 + 2-part linear exposure	80	113	146	7	480.92
6. Model 2 + per visit (“Saturated” model)	88	113	138	25	516.58

[a] 5 parameters represent the intercept, a spline for follow up time, and centered baseline CD4 cell count