Multiple imputation to account for measurement error in marginal structural models

Supplementary material

A. Standard marginal structural model

We estimate the parameters of the marginal structural model in equation 2 using a 2-step process: First, we estimate inverse probability of exposure weights $W_i^*(t)$ and then we maximize the weighted partial likelihood

$$L(\theta) = \prod_{i=1}^{N} \left\{ \frac{\exp[\theta_1 s^* + \theta_2 a(y_i) + \theta_3 s^* a(y_i)]}{\sum_{j=1}^{N} I(y_j \geq y_i) \times \exp[\theta_1 s^* + \theta_2 a(y_i) + \theta_3 s^* a(y_i)] \times w_j^*(y_i)} \right\}^{\delta_i \times w_i^*(y_i)}$$

(3)

where $I(y_j \geq y_i)$ is an indicator of patient $j$ being in the risk set at time $y_i$.

Weights in the standard analysis were estimated for each month as $W^*(t) = W_S \cdot W_{A(S^*)}(t)$, or the product of time-fixed stabilized inverse-probability of (provider-reported) smoking weights and time-varying stabilized inverse-probability of treatment weights. Specifically, the smoking weight for each individual was estimated as

$$W_{S^*} = \frac{f(S^*)}{f(S^*|L)}$$

(4)

where $L$ is a vector time-fixed covariates including age, sex, race, ethnicity, and year, HIV transmission risk factor (indicators of being a man who has sex with men and injection drug use), CD4 cell count, and viral load at study entry. Numerator and denominator were estimated using logistic regression.

Each patient’s weight for therapy initiation was the inverse probability receiving the treatment history that patient received by time $t$, where time is measured in months since
study entry for each patient. Stabilized weights for therapy initiation were estimated for each patient at each time $t$ in the standard analysis as

$$W_{A(S^*)}(t) = \prod_{k=0}^{\lfloor t \rfloor} \frac{f\{A(k) | \bar{A}(k - 1), S^*\}}{f\{A(k) | \bar{A}(k - 1), L, \bar{Z}(k), S^*\}}$$  \hspace{1cm} (5)$$

where $\bar{A}(k - 1)$ represents treatment history, $L$ represents the time-fixed covariates described above, $\bar{Z}(k)$ represents the history of a vector of time-varying covariates, including CD4 cell count and an indicator of detectable viral load (>500 copies/mL). To summarize treatment history, we let $\bar{A}(k - 1) = 1$ if the patient had initiated treatment through time $k - 1$, and $\bar{A}(k - 1) = 0$ otherwise. Covariate history was summarized as the value of each covariate at time $k$. Note that the treatment weights include the patient’s error-prone smoking status $S^*$. The numerator and denominator of the treatment weights were estimated using pooled logistic regression. Continuous variables in both sets of weights were modeled using restricted quadratic splines with 4 knots placed at the 5th, 35th, 65th, and 95th percentiles.
B. SAS code to implement multiple imputation to account for measurement error in an inverse-probability-weighted marginal structural model

The SAS code below is adapted from Cole et al (2006)\textsuperscript{25}.

```sas
%let m=20;
%let _null_; %run;

data m; set null;
proc logistic data=p descending covout outest=b(keep=_name_ intercept w delta wd logy z dlogy) noprint; model x = w d wd logy z dlogy;

data bb; set b; if _name_="x";
bh0=intercept; bh1=w; bh2=delta; bh3=wd; bh4=logy; bh5=z; bh6=dlogy;
keep bh0-bh6;

data cov; set b; if name_ ^= "x"; keep intercept w delta wd logy z dlogy;

proc iml;
use cov; read all into cov; *variance-covariance matrix;
use bb; read all into mu; mu=mu`; *means;
v=nrow(cov); *number of variables;
n=&m; *number of imputations;
seed=222;
1=t(root(cov)); *cholesky root of cov matrix;
z=normal(j(v,n,seed)); *generate nvars*samplesize normals
x=1*z;
x=repeat(mu,1,n)+x; *premultiply by cholesky root;
tx=t(x);
create m from tx; *write out sample data to sas dataset;
append from tx;
quit;

data m ; set m; array col {7} col1-col7; array b {7} b0-b6;
retain _imputation_ 0;
_imputation_=_imputation_+1;
do u=1 to &m; b[u]=col[u]; end;
drop col1-col7 u;
run;

data aa; set p; do _imputation_=1 to &m; output; end; run;
proc sort data=m; by _imputation_; run;
proc sort data=aa; by _imputation_; run;
data mi; merge aa m; by _imputation_;
```
data mi; set mi; call streaminit(320);
   if x=. then x=rand("bernoulli",1/(1+exp(-(b0+b1*w+b2*delta+b3*wd+
b4*logy+b5*z+b6*dlogy))));
run;

proc logistic data=mi desc noprint; by _imputation_; model x= ; output
out=num(keep=imputation_ j n) p=n;
proc logistic data=mi desc noprint; by _imputation_; model x = z; output
out=den(keep=imputation_ j d) p=d;
data w; merge mi num den; by _imputation_ id;
   num=(n*(x=1)+(1-n)*(x=0));
   den=(d*(x=1)+(1-d)*(x=0));
   sw=num/den;
run;

proc phreg data=w covs covout outest=imp noprint; by _imputation_; model
y*delta(0)=x; weight sw;
proc mianalyze data=imp; modeleffects x; run;
C. Simulation details

In each scenario, let \( i \) index hypothetical patients from 1 to 2,000 in each of 2,000 simulated cohorts. Measured confounder \( Z = 1 \) for 30% of patients, and patients with \( Z = 1 \) were half as likely to be exposed to the time-fixed exposure \( (X = 1) \) as patients with \( Z = 0 \). The outcome time \((T_i)\) was generated for each patient based on \( X_i \) and \( Z_i \), and time to censoring for each patient \((C_i)\) was generated based on \( X_i \). The observed time for each patient was represented by \( Y_i = T_i \wedge C_i \). A marginal structural Cox model estimated by inverse probability weighting of the true exposure \( X \) yielded a hazard ratio of 1.75.

A misclassified version of the exposure, \( V_i \), was generated for each patient based on \( X_i \) and misclassification probabilities reflecting the misclassification probabilities for smoking status in the application described above. When \( X_i = 1 \), \( V_i \) was drawn from a Bernoulli distribution with probability equal to the sensitivity, which was set to 70%. When \( X_i = 0 \), \( V_i \) was drawn from a Bernoulli distribution with probability \( 1 - \) specificity, where specificity was 80%. For each patient, \( R_i \) was an indicator of being included in the validation subgroup. \( R_i \) was drawn from a Bernoulli distribution with probability of 0.3 to represent a 30% validation subgroup.

In each simulated cohort, the standard analysis fit a marginal structural Cox model using inverse probability weighting of exposure \( V_i \). Weights were stabilized by the marginal probability of exposure so that they took the form \( w_i^* = P(V_i = v_i) / P(V_i = v_i | Z_i) \). The hazard ratio in the standard analysis was estimated as \( \exp(\zeta) \) in the marginal structural Cox model \( \lambda_{T^*}(t) = \lambda_0(t) \exp(\zeta v) \), where \( \zeta \) was obtained by maximizing the weighted partial likelihood given in text as equation 3.
The analysis accounting for exposure measurement error first imputed the gold standard exposure for individuals outside the validation subgroup based on the observed exposure \( V_i \) and covariate \( Z_i \) in each of \( M \) imputations. Weights were stabilized by the marginal probability of exposure and were estimated in each imputation \( m \) as \( w_i^m = P(X_i^m = x_i^m) / P(X_i^m = x_i^m | Z_i) \). The hazard ratio in each imputation was estimated as \( \exp(\psi^m) \) in the marginal structural Cox model \( \lambda_{T|x}^m(t) = \lambda_0^m(t) \exp(\psi^m x^m) \), where \( \psi \) was obtained by maximizing the weighted partial likelihood given in the text as equation 7.

The results were combined across imputations using Rubin’s rules so that the summary hazard ratio was \( \exp(\overline{\psi}) = \exp\left(M^{-1} \sum_{m=1}^{M} \hat{\psi}^m\right) \) and the variance was given by \( V(\overline{\psi}) = M^{-1} \sum_{m=1}^{M} V(\hat{\psi}^m) + (1 + M^{-1})(M - 1)^{-1} \sum_{m=1}^{M} (\hat{\psi}^m - \overline{\psi})^2 \).

Bias was defined as the average difference between the \( \ln(HR) \) estimated using the standard marginal structural model or the marginal structural model using multiple imputation to account for measurement error and the “true” \( \ln(HR) \), as given by the marginal structural model fit using the true exposure. Coverage was the proportion of simulated cohorts in each scenario in which the 95% confidence intervals included the true HR, and power was the proportion of cohorts in which the 95% confidence interval excluded the null value. The bias-variance tradeoff was evaluated using mean squared error, which was the sum of the squared bias and the squared standard deviation of the bias.