The Stata Journal (2004)
4, Number 4, pp. 402–420

Controlling for time-dependent confounding using marginal structural models

Zoe Fewell
University of Bristol, UK
Zoe.Fewell@bristol.ac.uk

Miguel A. Hernán
Harvard School of
Public Health, USA
mhernan@hsph.harvard.edu

Frederick Wolfe
National Data Bank for
Rheumatic Diseases, USA
fwolfe@arthritis-research.org

Kate Tilling
University of Bristol, UK
Kate.Tilling@bristol.ac.uk

Hyon Choi
Harvard Medical School, USA
hchoi@partners.org

Jonathan A. C. Sterne
University of Bristol, UK
Jonathan.Sterne@bristol.ac.uk

Abstract. Longitudinal studies in which exposures, confounders, and outcomes are measured repeatedly over time have the potential to allow causal inferences about the effects of exposure on outcome. There is particular interest in estimating the causal effects of medical treatments (or other interventions) in circumstances in which a randomized controlled trial is difficult or impossible. However, standard methods for estimating exposure effects in longitudinal studies are biased in the presence of time-dependent confounders affected by prior treatment.

This article describes the use of marginal structural models (described by Robins, Hernán, and Brumback [2000]) to estimate exposure or treatment effects in the presence of time-dependent confounders affected by prior treatment. The method is based on deriving *inverse-probability-of-treatment* weights, which are then used in a pooled logistic regression model to estimate the causal effect of treatment on outcome. We demonstrate the use of marginal structural models to estimate the effect of methotrexate on mortality in persons suffering from rheumatoid arthritis.

Keywords: st0075, marginal structural models, causal models, weighted regression, survival analysis, logistic regression, confounding

1 Introduction

Observational studies in which the recruited subjects are followed over time are called *cohort studies* by epidemiologists and *panel studies* by social scientists. Values of the characteristics of interest that are recorded when the subjects enter the study are known as the *baseline measurements*. Subjects are *followed up* until the *outcome event* occurs or they are *censored*. *Censoring* occurs when the study ends or the subject withdraws from the study. In the context of epidemiological research, the *outcome event* might be

death or the occurrence of a particular disease or illness. The time between the start of follow-up and occurrence of the outcome event is called the *failure time*.

The problem addressed here is estimation of the causal effect of exposure to *treatment* on the outcome event, but the method applies equally to *exposures* or *risk factors* that are not assigned by a health professional. In observational studies, the control of *confounding* is a fundamental problem in analyzing data and interpreting results. A *confounder* or *confounding variable* is associated with both the occurrence of the outcome event and the treatment of interest. Confounding can be controlled by stratification, for example by using Mantel–Haenszel methods, or by using regression models in which both the treatment and the confounders are included as covariates.

Here we will consider observational studies in which measurements of exposures and confounders are made (*updated*) on a number of different occasions. Use of standard regression models for the analysis of cohort studies with time-updated measurements may result in biased estimates of treatment effects if *time-dependent* confounders affected by prior treatment are present. A covariate is a time-dependent confounder if it predicts

1. future treatment and
2. future outcome, conditional on past treatment.

If past treatment predicts the current covariate value (e.g., if the covariate is on the causal pathway between treatment and the outcome), standard survival analyses with time-updated treatment effects will give biased treatment effect estimates.

For example, consider a study to estimate the effect of methotrexate on mortality among rheumatoid arthritis sufferers. Methotrexate is a disease-modifying antirheumatic drug (DMARD) commonly prescribed to people suffering from rheumatoid arthritis. The health assessment questionnaire disability index, a measure of a patient's level of functional ability, predicts whether a person suffering from rheumatoid arthritis will be treated with methotrexate and also predicts survival of that person. Therefore, disability index is a time-dependent confounder. Additionally, treatment with methotrexate lowers disability index. This is illustrated in figure 1.

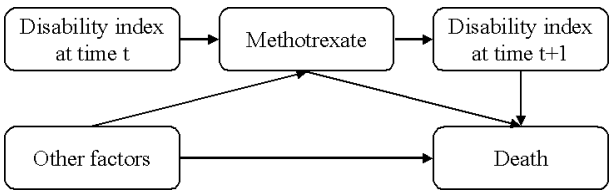

Figure 1: Time-dependent confounding

There are several standard methods that might be used in this setting to estimate the causal effect of treatment on the outcome, but all will produce biased results.

1. The crude estimate, in which there is no control for confounding, will produce biased estimates because methotrexate will tend to be given to people whose disability index is higher and who therefore experience greater severity of rheumatoid arthritis and higher death rates.

2. Control for the baseline values of confounders such as disability index will give biased estimates because this ignores the fact that people who started methotrexate therapy after the start of the study will be those whose disability index worsened.

3. Control for time-updated measurements of time-dependent confounders affected by prior treatment such as disability index will give biased estimates because methotrexate acts partly by lowering the disability index score. This is illustrated in figure 2.

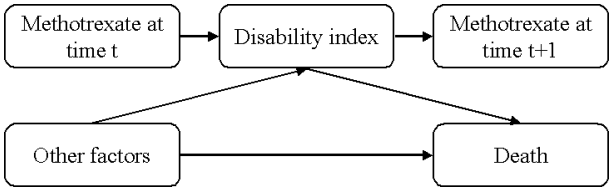

Figure 2: Time-dependent confounder affected by prior treatment

In figure 2, methotrexate therapy at time t predicts the subsequent value of disability index. Additionally, disability index predicts the outcome (death) and future methotrexate therapy. The health assessment questionnaire disability index lies on the causal pathway between exposure to methotrexate at time t and the outcome. In this case, controlling for disability index will bias the estimation of treatment effect, as any effect of methotrexate acting via disability index will be lost. If there is no causal pathway between methotrexate and death, or between disability index and death, standard methods may still produce biased results when disability index is affected by previous treatment. For further information, see Robins et al. (1992).

Marginal structural models aim to appropriately control for the effects of time-dependent confounders affected by prior treatment. We describe how to fit these models here, using estimation of the effect of methotrexate on mortality among rheumatoid arthritis patients as an illustration.

2 Example

The analysis described in this paper is performed on data from the Wichita Arthritis Center database, which is described in detail in Choi et al. (2002), and was done using Stata version 8.2. The data consist of observations on 1,240 patients with rheumatoid arthritis who were seen at the Wichita Arthritis Center between 1 January 1981 and 31 December 1999, contains 91,007 person-months of observation, and is arranged so that there is one observation per person per month that they remained in the study.

Censoring occurred two years after a subject last visited the clinic, and missing data is filled in using the last known values of the covariates. The aim of the analysis is to estimate the effect of methotrexate on all-cause mortality among rheumatoid arthritis patients.

For the purposes of the analysis, we will assume that, once a patient starts taking methotrexate, they will remain on it until the end of follow-up. This provides a conservative estimate of the treatment hazard ratio, analogous to intention-to-treat analysis in an unblinded randomized controlled trial, and can be considered as an intention to continue treatment analysis.

2.1 Variables

The outcome variable in these analyses will be `dead`, and the treatment variable will be `mtxspan`. Variable `cens` is an indicator for whether follow-up for a patient was censored during the current month.

variable name	storage type	display format	value label	variable label
dead	byte	%8.0g		Death status (1/0)
mtxspan	byte	%9.0g		On methotrexate
cens	byte	%9.0g		Censored

As described below, we will model patients' probabilities of being treated or uncensored over time. To do this, we will include variables measured at times t (the current month) and $t - 12$ (one year ago) in logistic regression models. This means that we are assuming that only the current value and the value one year previously of each covariate are useful for predicting the probability of being treated or uncensored. For this analysis, the variables included in these models record use of other DMARDs, erythrocyte sedimentation rate, patient's global assessment of disease severity, health assessment questionnaire score, tender-joint count, prednisone use, and smoking, measured at month t and month $t-12$, if applicable. For the first twelve months of follow-up, the lagged variables are defined to be equal to the baseline value of that variable.

(*Continued on next page*)

variable name	storage type	display format	value label	variable label
dmrd	byte	%9.0g	othdmrd	On other DMARDs
dmrd_nd	byte	%9.0g	othdmrd	On other DMARDs 12 months before
esrc	byte	%9.0g	esrc	Erythrocyte sedimentation rate
esrc_nd	byte	%9.0g	esrc	Erythrocyte sedimentation rate 12 months before
gs	byte	%14.0g	gs	Patient's assessment of global disease severity
gs_nd	byte	%14.0g	gs	Patient's global assessment of disease severity 12 months before
haq	byte	%9.0g	haq	Health assessment questionnaire score
haq_nd	byte	%9.0g	haq	Health assessment questionnaire score 12 months before
jc	byte	%9.0g	jc	Tender joint count
jc_nd	byte	%9.0g	jc	Tender joint count 12 months before
onprd2	byte	%8.0g		On prednisone
onprd2_nd	byte	%9.0g		On prednisone 12 months before
smokenow	byte	%11.0g	smoking	Smoking by code

In addition to the variables described above, we will estimate the probabilities of being treated and uncensored conditionally on variables measured at baseline. These are age, use of other DMARDs, rheumatoid arthritis duration, education, erythrocyte sedimentation rate, patient's global assessment of disease severity, health assessment questionnaire score, tender-joint count, prednisone use, presence of rheumatoid factor, sex, smoking, and calendar year.

variable name	storage type	display format	value label	variable label
age_0	byte	%12.0g	age	Baseline value of age
dmrd_0	byte	%9.0g	othdmrd	On other DMARDs at baseline
duration_0	byte	%9.0g	duration	Baseline value of duration
edu_0	byte	%26.0g	education	Education level at baseline
esrc_0	byte	%9.0g	esrc	Erythrocyte sedimentation rate at baseline
gs_0	byte	%14.0g	gs	Patient's assessment of global disease severity at baseline
haq_0	byte	%9.0g	haq	Health assessment questionnaire score at baseline
jc_0	byte	%9.0g	jc	Tender joint count at baseline
onprd2_0	byte	%9.0g		On prednisone at baseline
rapos	byte	%26.0g	rapos	Rheumatoid factor
sex	byte	%8.0g		Sex of patient
smoke_0	byte	%11.0g	smoking	Smoking at baseline by code
year_0	byte	%21.0g	year	Baseline value of year

Many of these variables are categorical. The categories are described below.

```
age:
            1 <=50
            2 [51, 60]
            3 [61, 70]
            4 >70
duration:
            1 [0, 3]
            2 (3, 10]
            3 >10
education:
            1 <12
            2 =12
            3 >12

rapos:
            0 Rheumatoid factor negative
            1 Rheumatoid factor positive
smoking:
            0 Never
            1 Now
            2 In the past
esrc:
            1 <=30
            2 >30
gs:
            1 [0, 3.333]
            2 (3.333, 6.666]
            3 (6.666, 10]
jc:
            1 [0, 3]
            2 (3, 6]
            3 (6, 18]
othdmrd:
            1 None
            2 >=1
year:
            0 Before 1981
            1 1981-1990
            2 1991-2000
```

In the following display, we list some example data. The variable `patkey` records each subject's unique identity number, and `cummonth` records a person's number of months in the study. Person 10521 was censored in month 80, which is recorded by `cens` = 1 at `cummonth` = 80. The outcome, death, cannot be observed, and `dead` is recorded as missing. Person 10541 died between month 4 and month 5 of follow-up, which is recorded as `dead` = 1 at `cummonth` = 4. Also shown are the current tender-joint count values, the values 12 months previously, and the baseline values.

```
. list patkey cummonth dead mtxspan cens jc jc_nd jc_0 in 87791/87800
```

	patkey	cummonth	dead	mtxspan	cens	jc	jc_nd	jc_0
87791.	10521	75	0	1	0	1	2	1
87792.	10521	76	0	1	0	1	2	1
87793.	10521	77	0	1	0	1	2	1
87794.	10521	78	0	1	0	1	2	1
87795.	10521	79	0	1	0	1	2	1
87796.	10521	80	.	1	1	1	2	1
87797.	10541	1	0	1	0	2	2	2
87798.	10541	2	0	1	0	1	2	2
87799.	10541	3	0	1	0	1	2	2
87800.	10541	4	1	1	0	1	2	2

3 Marginal structural models

Marginal structural models aim to estimate the effect of treatment on outcome by appropriate control for the effects of time-dependent confounders. The model is fitted in a two-stage process in which

1. we estimate each subject's probability of having their own treatment history and use these to derive *inverse-probability-of-treatment weights* (IPTW), and

2. the treatment–outcome association is estimated in a regression model that is weighted using the IPTWs.

3.1 Notation

The dataset we are using for this analysis has been discretized by time, so that there is one observation per person per month that they remained in the study. The reason for this is that, as mentioned in the previous section, lagged variables must be used. Suppose, for example, that one subject had last visited the clinic one month ago and that another had last visited six months ago. This would make it very difficult to create a variable lagged to the last visit that would be comparable between subjects. With our time-discretized dataset, the lagged variables are comparable between subjects.

Using the same notation as Hernán, Brumback, and Robins (2002), let T_i denote the observed failure time of subject i, and let $A_i(t)$ denote treatment of subject i at time t. Throughout, $A_i(t)$ is a dichotomous variable, taking the value 1 if subject i receives treatment in month t and 0 otherwise. Let V_i denote a subset of the baseline values of all covariates and $L_i(t)$ denote the values of the covariates at month t for subject i. In this particular example, we will use $V_i = L_i(0)$, the baseline values of all covariates. $\overline{A}_i(t)$ denotes treatment history (i.e., the vector of values of $A_i(k)$ from $k = 0$ to $k = t - 1$), and similarly the matrix $\overline{L}_i(t)$ denotes history of time-dependent confounders for subject i. We often suppress the i subscript denoting individual in the notation because we assume that the random vector for each subject is drawn independently from a distribution common to all subjects. We will also use standard statistical notation, in which an uppercase letter denotes a random variable and the corresponding lowercase letter denotes a particular realization of that random variable.

3.2 Counterfactuals

At each month t, $A(t)$ can be either 1 or 0. If the study length is K months, there will therefore be 2^K different possible values for $\overline{a}(K)$.

We denote by $T_{\overline{a}}$ a subject's failure time had they received treatment history $\overline{a}$. Only one value of $\overline{a}$ is observed for each subject, and the only failure time we observe is that for which $T_{\overline{a}} = T$. All other values of $T_{\overline{a}}$ occur contrary to fact and so are called *counterfactual* variables. (For further discussion of counterfactuals, see Rothman and Greenland [1998].) For each $\overline{a}$, we will specify the marginal structural Cox proportional hazards model

$$\lambda_{T_{\overline{a}}}(t|V) = \lambda_0(t) \exp\{\beta_1 a(t) + \beta_2 V\} \tag{1}$$

where $\lambda_{T_{\overline{a}}}(t|V)$ is the hazard of death at time t among subjects with baseline covariates V had they all followed treatment history $\overline{a}$, β_1 and the vector β_2 are unknown parameters to be estimated, and λ_0 is an unspecified baseline hazard function. Our focus of

interest is on parameter β_1, which is an estimate of the causal log hazard ratio for the effect of treatment on mortality, comparing subjects' hazard of mortality if they were continuously treated with their hazard of mortality if they were never treated.

This model is known as a *marginal structural* model because

1. it is a model for the *marginal* distribution of the counterfactual variables $T_{\overline{a}}$ (the counterfactual survival times associated with each treatment history $\overline{a}$) conditional on the baseline variables V, rather than a model for the joint distribution of the $T_{\overline{a}}$, and
2. causal models, or models for counterfactual variables, are often referred to as *structural* in the econometric and social sciences literature (Hernán, Brumback, and Robins 2002).

4 Inverse-probability-of-treatment weights (IPTW)

Each observation on an individual will be weighted using an inverse-probability-of-treatment weight. For the time being, we assume that each subject is followed-up until the outcome event occurs or the study ends, whichever comes first. This means that there will be no censoring by competing risks. In the following equations, $\overline{A}(-1)$ is defined to be 0.

The inverse-probability-of-treatment weight is

$$W(t) = \prod_{k=0}^{t} \frac{1}{f\{A(k)|\overline{A}(k-1), \overline{L}(k)\}} \tag{2}$$

where $f(\ldots)$ denotes the conditional probability mass function. However, in practice these weights tend to be highly variable and fail to be approximately normally distributed. Therefore, the stabilized version

$$SW(t) = \prod_{k=0}^{t} \frac{f\{A(k)|\overline{A}(k-1), V\}}{f\{A(k)|\overline{A}(k-1), \overline{L}(k)\}} \tag{3}$$

is preferable, due to its smaller variance, and yields 95% confidence intervals that are narrower and have better coverage rates (Hernán, Brumback, and Robins 2000; Robins, Hernán, and Brumback 2000).

Informally, the denominator of (3) can be thought of as a subject's conditional probability of receiving his or her own observed treatment history up to time t, given past treatment and prognostic factor history. Note that V is included as $L(0)$. The numerator can be thought of, informally, as a subject's conditional probability of receiving his or her own observed treatment history up to time t, given past treatment history.

For the IPTW estimates to perform well, our estimate of $SW(t)$ cannot be exceedingly variable (Hernán, Brumback, and Robins 2002). To guarantee this, we reduce the

number of free parameters in the model estimating $SW(t)$. Instead of estimating a separate intercept for each month, we assume that the intercept is a smooth function and estimate it using cubic splines. Alternative methods would be to assume that the intercept is constant in windows of, for example, 3 months, or to use other smoothing techniques, such as kernel regression (Hernán, Brumback, and Robins 2000).

4.1 Deriving IPTWs using Stata

We use the command `spbase` (Sasieni 1994) to generate a truncated power basis for a natural cubic spline. The variable used to create the basis is `cummonth`. The `knots` option tells `spbase` where the knots in the spline should be: here months 4, 25, 52, 90, and 165 were chosen. These correspond to the 5th, 27.5th, 50th, 72.5th, and 95th centiles of `cummonth`. Finally, we give a name to the basis variables by using `gen(spline)`.

```
. spbase cummonth, knots(4, 25, 52, 90, 165) gen(spline)
```

This basis can now be used in our regression models to adjust for `cummonth`.

We assume that, once a subject starts treatment, they remain on treatment until the end of follow-up, and we can therefore view $A(t)$ as a failure-time variable. Hence, we can model the probability of being treated at month t by a Cox proportional hazards model. However, because the dataset has been discretized into one observation per person per month, we will use a pooled logistic regression model in which we model the probability that each individual is treated in each month. This is equivalent to a Cox model (D'Agostino, Lee, and Belanger 1990) because the hazard of treatment in any single month is small.

To estimate the denominator of (3), we first fit a logistic regression in which we model the association of the outcome variable `mtxspan` with the covariates measured at times t, $t-12$, and baseline, described in section 2.1. We also include as covariates `cummonth` and the spline variables `spline*`. The regression is performed only for months up to and including a subject's first month on methotrexate. The variable `mtx1stcu` records the month in which each person started methotrexate therapy, so we use the qualifier `if cummonth<=mtx1stcu | mtx1stcu==.`.

```
. xi: logistic mtxspan onprd2 i.dmrd i.haq i.gs i.esrc i.jc i.smokenow
> onprd2_nd i.dmrd_nd i.haq_nd i.gs_nd i.esrc_nd i.jc_nd onprd2_0
> i.duration_0 i.age_0 i.year_0 i.dmrd_0 i.haq_0 i.gs_0 i.esrc_0
> i.jc_0 i.smoke_0 sex i.edu_0 rapos cummonth spline*
> if cummonth<=mtx1stcu | mtx1stcu==.
  (output omitted)
```

Following the logistic regression, we use the `predict` command to estimate the probability of receiving methotrexate for each subject-month included in the regression.

```
. predict pmtx if e(sample)
(option p assumed; Pr(mtxspan))
(37006 missing values generated)
```

The 37,006 missing values generated are for the months following a subject's methotrexate initiation. As we assume that, once a subject begins taking methotrexate, they remain on it until the end of follow-up, the probability of receiving methotrexate in the months following initiation is equal to 1.

```
. replace pmtx=1 if cummonth>mtx1stcu
(37006 real changes made)
```

We have now estimated the probability of each person receiving methotrexate therapy in each month, given their covariate history. We now derive the probability of a person's *observed* methotrexate treatment in each month, so that for months in which they are not taking methotrexate, we need to subtract the estimated probability of treatment from 1.

```
. replace pmtx=pmtx*mtxspan+(1-pmtx)*(1-mtxspan)
(53413 real changes made)
```

To estimate each subject's probability of their complete treatment history up to each month (the denominator in [3]), we multiply the estimated probabilities of their observed treatment during each month cumulatively over time. The first estimated probability for each subject is left as it is. For all others, the estimated probability at the current time point is multiplied by the estimated probability at the previous time point. As Stata works from the first observation on a subject to the last, the result is the estimated probability of a subject's observed treatment history.

```
. sort patkey cummonth
. by patkey: replace pmtx=pmtx*pmtx[_n-1] if _n!=1
(89767 real changes made)
. rename pmtx mtxdenom
. summ mtxdenom, detail
```

Pr(mtxspan)

Percentiles		Smallest		
1%	.0008458	.0002996		
5%	.0034857	.0002996		
10%	.0064226	.0002996	Obs	91007
25%	.0372735	.0002996	Sum of Wgt.	91007
50%	.5785145		Mean	.4944946
		Largest	Std. Dev.	.3907536
75%	.8852679	.999525		
90%	.9649602	.9995492	Variance	.1526884
95%	.982233	.9995544	Skewness	-.1142317
99%	.9948645	.9998104	Kurtosis	1.301617

Variable `mtxdenom` now contains the required denominator for (3). Using these values to create the unstabilized version of the weight (2) would result in weights ranging between 1.0 and 3337.8. Using stabilized inverse-probability-of-treatment weights (3) will reduce this range of values.

The numerator of (3) is estimated in a similar way to the denominator. The only difference is that the initial logistic regression does not include covariates measured at time t and time $t - 12$.

```
. xi: logistic mtxspan onprd2_0 i.duration_0 i.age_0 i.year_0 i.dmrd_0
> i.haq_0 i.gs_0 i.esrc_0 i.jc_0 i.smoke_0 sex i.edu_0 rapos cummonth
> spline* if cummonth<=mtx1stcu | mtx1stcu==.
```

(*output omitted*)

```
. predict pmtx if e(sample)
(option p assumed; Pr(mtxspan))
(37006 missing values generated)

. replace pmtx=1 if cummonth>mtx1stcu
(37006 real changes made)

. replace pmtx=pmtx*mtxspan+(1-pmtx)*(1-mtxspan)
(53413 real changes made)

. sort patkey cummonth

. by patkey: replace pmtx=pmtx*pmtx[_n-1] if _n!=1
(89767 real changes made)

. rename pmtx mtxnum
```

Variable `mtxnum` now contains the required numerator for (3).

We derive the stabilized weight (3) by dividing the numerator by the denominator.

```
. gen stabweightmtx=mtxnum/mtxdenom
. summ stabweightmtx, detail
```

stabweightmtx

Percentiles		Smallest		
1%	.0353742	.0067794		
5%	.0893559	.0067794		
10%	.1619441	.0067794	Obs	91007
25%	.579341	.0067794	Sum of Wgt.	91007
50%	.8884353		Mean	1.019517
		Largest	Std. Dev.	1.339553
75%	.9990956	24.85306		
90%	1.494492	24.89879	Variance	1.794403
95%	2.348431	24.94612	Skewness	7.473572
99%	6.083678	30.61479	Kurtosis	80.63894

The values of the stabilized weights are centered around 1.02 and show a much narrower range ([0.007, 30.615]) than the unstabilized weights.

5 Censoring

In a cohort study, some subjects will drop out before the outcome event occurs or the study ends. We say that these subjects are *censored*. Let $C(t)$ be a dichotomous variable taking the value 1 if a subject is censored in month t and 0 otherwise. $\overline{C}(t)$ will denote censoring history (that is, the vector of values of $C(k)$ from $k = 0$ to $k = t - 1$). To deal with censoring in the marginal structural model, we again derive weights, this time for the probability of remaining uncensored up to time t, and again use the stabilized version of the weight. Each observation on an individual will then be weighted by the IPTW (see section 4) multiplied by the inverse-probability-of-censoring weight, which we will derive below. In the following equations, $\overline{A}(-1)$, $\overline{C}(-1)$, and $\overline{L}(-1)$ are defined to be 0.

The stabilized version of the inverse-probability-of-censoring weight is

$$SW^{\dagger}(t) = \prod_{k=0}^{t} \frac{\Pr\{C(k) = 0 | \overline{C}(k-1) = 0, \overline{A}(k-1), V, T > k\}}{\Pr\{C(k) = 0 | \overline{C}(k-1) = 0, \overline{A}(k-1), \overline{L}(k-1), T > k\}} \tag{4}$$

If we want to additionally adjust for censoring in our analysis, the conditioning event in (3) for the IPTW changes to include $C(k) = 0$ so that the stabilized IPTW is now estimated by

$$SW(t) = \prod_{k=0}^{t} \frac{f\{A(k) | \overline{A}(k-1), V, C(k) = 0\}}{f\{A(k) | \overline{A}(k-1), \overline{L}(k), C(k) = 0\}}$$

In practice, changing the conditioning event does not affect the commands used to estimate $SW(t)$. The final weight for each subject-month is

$$SW(t) \times SW^{\dagger}(t)$$

as shown at the end of section 5.1 below.

When dealing with censoring, we must also pay attention to the setup of the dataset. In order to perform a pooled logistic regression, time, a continuous variable, must be discretized. This requires some decisions to be made about how certain variables are recorded. The two variables of concern here are the outcome, `dead`, and the censoring variable, `cens`.

Consider a minidataset with only one subject. The subject is followed-up for three months. Each month, all relevant covariates are measured and recorded in `covar`, following which the decision whether to treat or not is made. This is recorded in the variable `treat`. Following the third month, the subject sends a letter to the study organizers informing them that he no longer wishes to participate in the study. The subject is censored at month 4, and the outcome variable, `dead`, is not observed and so is set to missing. This example dataset is displayed below. (Note that, in our example of the effect of methotrexate on mortality in rheumatoid arthritis patients, letters were not sent to drop out of the study. Censoring occurred 24 months after a subject's last clinic visit.)

	month	treat	covar	cens	dead
1.	1	a1	l1	0	0
2.	2	a2	l2	0	0
3.	3	a3	l3	0	0
4.	4	.	.	1	.

To estimate the weights due to censoring, we use (4). The probability of not being censored at time k is estimated using treatment history and other covariate history at time $k-1$. Similarly, to estimate the mortality hazard ratio, we use the outcome at time k and all other covariates at time $k-1$. As our dataset stands, the values of each variable that we want to use are not aligned. In order to analyze the data, it is convenient to shift the values of `cens` and `dead` up by one, as shown.

	month	treat	covar	cens	dead
1.	1	a1	l1	0	0
2.	2	a2	l2	0	0
3.	3	a3	l3	1	.
4.	4	.	.	.	.

Suppose that we want to estimate the denominator of (4). We will fit a logistic regression for `cens` based on treatment and covariate history and use it to predict values for $\Pr\{C(k) = 0\}$, which are recorded in the variable `pnotcens`.

	month	treat	covar	cens	pnotcens	dead
1.	1	a1	l1	0	p1	0
2.	2	a2	l2	0	p2	0
3.	3	a3	l3	1	p3	.
4.	4	.	.	.	.	.

At this point, we could replace the value `p3` with `1-p3` and then multiply the estimated probabilities together to obtain the subject's estimated probability of observed censoring history. This would be analogous to the procedure used to calculate the treatment weights. In practice, however, there is no point in replacing the value `p3` with `1-p3`. This value will never be used in the final weighted model because the outcome, `dead`, is recorded as missing. We therefore multiply the estimated probabilities together as they are and record the results in the variable `censden`.

	month	treat	covar	cens	pnotcens	censden	dead
1.	1	a1	l1	0	p1	p1	0
2.	2	a2	l2	0	p2	p1*p2	0
3.	3	a3	l3	1	p3	p1*p2*p3	.
4.	4	.	.	.	.	.	.

5.1 Deriving censoring weights using Stata

The weights due to censoring are estimated in a similar way to the treatment weights. To estimate the denominator of (4), we first perform a logistic regression in which we model the association of the outcome variable `cens` with the covariates measured at times t, $t - 12$, and baseline, described in section 2.1. We additionally adjust for methotrexate therapy using `mtxspan`. We also include as covariates `cummonth` and the spline variables `spline*`. The regression is performed for all person-months.

```
. xi: logistic cens mtxspan onprd2 i.dmrd i.haq i.gs i.esrc i.jc i.smokenow
> onprd2_nd i.dmrd_nd i.haq_nd i.gs_nd i.esrc_nd i.jc_nd onprd2_0
> i.duration_0 i.age_0 i.year_0 i.dmrd_0 i.haq_0 i.gs_0 i.esrc_0 i.jc_0
> i.smoke_0 sex i.edu_0 rapos cummonth spline*
```

(*output omitted*)

Following the logistic regression, we use the `predict` command to estimate the probability of being censored for each month of observation.

```
. predict pcens if e(sample)
(option p assumed; Pr(cens))
```

We have now estimated the probability of each person being censored in each month, given their covariate history. Subtracting these estimated probabilities from 1 results in an estimate of the probability of a person being *uncensored* in each month.

```
. replace pcens=1-pcens
(91007 real changes made)
```

To derive each subject's estimated probability of their complete censoring history up to each month (the denominator in [4]), we multiply the estimated probabilities of being uncensored for each month cumulatively over time. The first estimated probability for each subject is left as it is. For all others, the estimated probability at the current time point is multiplied by the estimated probability at the previous time point. As Stata works from the first observation on a subject to the last, the result is an estimate of the probability of a subject's observed uncensored history.

```
. sort patkey cummonth
. by patkey: replace pcens=pcens*pcens[_n-1] if _n!=1
(89767 real changes made)
. rename pcens censdenom
. summ censdenom, detail
```

Pr(cens)

Percentiles		Smallest		
1%	.0712816	.0034389		
5%	.1937141	.0036158		
10%	.2856901	.0039224	Obs	91007
25%	.4672786	.0040208	Sum of Wgt.	91007
50%	.6622012		Mean	.6308459
		Largest	Std. Dev.	.2353487
75%	.822933	.9974812		
90%	.9229034	.9974812	Variance	.055389
95%	.9598724	.9975808	Skewness	-.4834176
99%	.9881712	.9978794	Kurtosis	2.395454

Variable `censdenom` now contains the required denominator for (4). Using these values to create the unstabilized version of the weight would result in weights ranging between 1.0 and 290.8. Using the stabilized weights (4) will reduce this range of values.

The numerator of (4) is estimated in a similar way to the denominator, except we do not include covariates measured at times t and $t-12$ in the initial logistic regression.

```
. xi: logistic cens mtxspan onprd2_0 i.duration_0 i.age_0 i.year_0 i.dmrd_0
> i.haq_0 i.gs_0 i.esrc_0 i.jc_0 i.smoke_0 sex i.edu_0 rapos cummonth spline*
```

(*output omitted*)

```
. predict pcens if e(sample)
(option p assumed; Pr(cens))
. replace pcens=1-pcens
(91007 real changes made)
```

```
. sort patkey cummonth
. by patkey: replace pcens=pcens*pcens[_n-1] if _n!=1
(89767 real changes made)
. rename pcens censnum
```

Variable `censnum` now contains the required numerator for (4).

The censoring weight is generated by dividing the numerator by the denominator.

```
. gen censweight=censnum/censdenom
. summ censweight, detail
```

censweight

Percentiles		Smallest		
1%	.7237051	.3170074		
5%	.8562644	.3278275		
10%	.9037426	.3361159	Obs	91007
25%	.9569152	.3386074	Sum of Wgt.	91007
50%	.995711		Mean	.9992465
		Largest	Std. Dev.	.1126509
75%	1.026015	3.472285		
90%	1.088238	3.660231	Variance	.0126902
95%	1.160644	3.725618	Skewness	3.091838
99%	1.405247	3.794446	Kurtosis	43.95382

The values of the stabilized weights are centered around 0.999 and show a much narrower range ([0.317, 3.794]) than the unstabilized weights.

The overall weight $SW(t) \times SW^{\dagger}(t)$ is calculated by multiplying `stabweightmtx` by `censweight`.

```
. gen stabweightcens=stabweightmtx*censweight
```

6 Marginal structural model

We can now use model (1) to estimate the causal hazard ratio for the effect of treatment on mortality. This is done by weighting the observation for each subject-month by $SW(t) \times SW^{\dagger}(t)$. In the weighted analysis, each subject's probability of being treated at each time point is unrelated to their time-updated covariates.

Because `stcox` does not allow for time-varying, subject-specific weights, we will again fit a pooled logistic regression. When we additionally weight each subject, we introduce within subject correlation, which must then be adjusted for by deriving robust variance estimators (Hernán, Brumback, and Robins 2000). The model is

$$\text{logit pr}\{D(t) = 1 | D(t-1) = 0, \overline{A}(t), V\} = \beta_0(t) + \beta_1 A(t-1) + \beta_2 V$$

where $D(t) = 1$ if the subject dies in month t and $D(t) = 0$ otherwise.

6.1 Fitting the marginal structural model using Stata

The following shows the result fitting an unweighted pooled logistic regression. Note that, as explained earlier, the odds ratios are equivalent to the hazard ratios that would be obtained from the equivalent Cox model.

```
. xi: logistic dead mtxspan onprd2_0 i.duration_0 i.age_0 i.year_0 i.dmrd_0
> i.haq_0 i.gs_0 i.esrc_0 i.jc_0 i.smoke_0 sex i.edu_0 rapos onprd2_nd
> i.dmrd_nd i.haq_nd i.gs_nd i.esrc_nd i.jc_nd onprd2 i.dmrd i.haq i.gs
> i.esrc i.jc i.smokenow cummonth spline*
  (output omitted)
Logistic regression                               Number of obs   =      89958
                                                  LR chi2(49)     =     452.35
                                                  Prob > chi2     =     0.0000
Log likelihood = -1140.1952                       Pseudo R2       =     0.1655

------------------------------------------------------------------------------
        dead | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     mtxspan |   .5574524   .1108437    -2.94   0.003     .3775331     .823115
  (output omitted)
------------------------------------------------------------------------------
```

This estimate of the mortality hazard ratio of methotrexate therapy in rheumatoid arthritis sufferers is biased (as explained in section 1), although it does suggest that methotrexate has a beneficial effect.

To fit the marginal structural model, we weight the observation for each subject-month by $SW(t) \times SW^{\dagger}(t)$ by including in the command `[pw=stabweightcens]`. This weighting means that observations on the same subject will be correlated: we therefore use the `cluster` option to derive robust standard errors allowing for clustering.

```
. xi: logistic dead mtxspan onprd2_0 i.duration_0 i.age_0 i.year_0 i.dmrd_0
> i.haq_0 i.gs_0 i.esrc_0 i.jc_0 i.smoke_0 sex i.edu_0 rapos cummonth
> spline* [pw=stabweightcens], cluster(patkey)
  (output omitted)
Logistic regression                               Number of obs   =      89958
                                                  Wald chi2(29)   =     200.31
                                                  Prob > chi2     =     0.0000
Log pseudo-likelihood = -1359.8394                Pseudo R2       =     0.1590
                                  (standard errors adjusted for clustering on patkey)

------------------------------------------------------------------------------
             |               Robust
        dead | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     mtxspan |   .4299443   .1233882    -2.94   0.003     .2449791    .7545629
  (output omitted)
------------------------------------------------------------------------------
```

This estimates that the mortality hazard ratio of methotrexate therapy on rheumatoid arthritis sufferers is 0.4 (95% CI; [0.2, 0.8]), as reported in Choi et al. (2002).

7 Discussion

There are several assumptions that must be made when fitting a marginal structural model. The first is that there are *no unmeasured confounders*; i.e., we have measured all variables that are associated with both mortality and the probability of being treated. Note that this assumption cannot be tested using the data. Secondly, we assume that the marginal structural model for the effect of methotrexate on mortality among rheumatoid arthritis patients is correctly specified. Finally, we assume that the models for initiation of treatment and censoring used to estimate $SW(t)$ and $SW^{\dagger}(t)$ are correctly specified. These are strong assumptions, but they are the same assumptions required to make a causal interpretation when estimating the effect of a time-independent treatment using standard statistical methods. However, to make causal inferences about time-varying treatments from standard statistical models, we must assume that there is no time-dependent confounding by covariates that are affected by previous treatment. This assumption is not required when using a marginal structural model with IPTWs because the method adjusts appropriately for this type of confounding.

We have shown how to control for time-dependent confounders affected by prior treatment using a marginal structural model in which the treatment is a dichotomous variable. The method will also work with an ordinal treatment variable, for example, if the variable records a subject's dose of methotrexate in units of 5mg. Similarly, a continuous treatment variable can be used. For more details on fitting a marginal structural model with ordinal or continuous treatments, see Robins, Hernán, and Brumback (2000). As stated in the introduction, marginal structural models can also be used to estimate the exposure–outcome association for exposures or risk factors that are not assigned by a health professional.

Although useful to unbiasedly estimate the effect of time-dependent exposures in the presence of time-dependent confounders affected by prior treatment, marginal structural models do have their limitations. They cannot be used when, at time t, a subset of the population defined by the variables in $\overline{L}$ is certain to have a particular exposure, for example in an occupational cohort if people at work on day t were guaranteed exposure. Hence, marginal structural models should not be used in occupational cohort studies.

Marginal structural models also cannot be used to estimate the effects of dynamic treatment regimes. An example of a dynamic treatment regime is to treat when a patient's tender-joint count reaches a certain value. They can, however, be used to estimate the effect of a nondynamic treatment regime when the data are derived from a cohort study in which the treatment regime is dynamic. In this example, we estimated the hazard ratio of always taking methotrexate compared with never taking methotrexate on mortality in rheumatoid arthritis patients.

For cases in which marginal structural models cannot be used, G-estimation of structural nested models can be used instead. For information on G-estimation in Stata and a description of the `stgest` command, see Sterne and Tilling (2002).

Marginal structural models have some advantages over G-estimation. The major advantage of using marginal structural models over G-estimation to control for time-

dependent confounding is their close resemblance to standard modeling techniques. Further, fitting a marginal structural model is usually less computationally demanding. The example we have shown implements standard logistic regression techniques which makes marginal structural models extremely intuitive to use and easy to interpret. Whether the study is using G-estimation or marginal structural models to control for time-dependent confounding, it is crucial that the dataset contains comprehensive information on the variables used by the physician to make the decision to initiate treatment.

For more information on the advantages and disadvantages of marginal structural models versus structural nested models, see Robins (1999).

8 References

Choi, H. K., M. A. Hernán, J. D. Seeger, J. M. Robins, and F. Wolfe. 2002. Methotrexate and mortality in patients with rheumatoid arthritis: a prospective study. *Lancet* 359: 1173–1177.

D'Agostino, R. B., M.-L. Lee, and A. J. Belanger. 1990. Relation of pooled logistic regression to time dependent Cox regression analysis: The Framingham Heart Study. *Statistics in Medicine* 9: 1501–1515.

Hernán, M. A., B. Brumback, and J. M. Robins. 2000. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology* 11(5): 561–570.

Hernán, M. A., B. A. Brumback, and J. M. Robins. 2002. Estimating the causal effect of zidovudine on CD4 count with a marginal structural model for repeated measures. *Statistics in Medicine* 21: 1689–1709.

Robins, J. M. 1999. Marginal structural models versus structural nested models as tools for causal inference. In *Statistical Models in Epidemiology: The Environment and Clinical Trials*, ed. M. E. Halloran and D. Berry, 95–134. New York: Springer.

Robins, J. M., D. Blevins, G. Ritter, and M. Wulfsohn. 1992. G-estimation of the effect of prophylaxis therapy for Pneumocystis carinii pneumonia on the survival of AIDS patients. *Epidemiology* 3: 319–336.

Robins, J. M., M. A. Hernán, and B. Brumback. 2000. Marginal structural models and causal inference in epidemiology. *Epidemiology* 11(5): 550–560.

Rothman, K. and S. Greenland. 1998. *Modern Epidemiology*. 2nd ed. Philadelphia: Lippincott–Raven.

Sasieni, P. 1994. snp7: Natural cubic splines. *Stata Technical Bulletin* 22: 19–22. In *Stata Technical Bulletin Reprints*, vol. 4, 171–174. College Station, TX: Stata Press.

Sterne, J. A. C. and K. Tilling. 2002. G-estimation of causal effects, allowing for time-varying confounding. *Stata Journal* 2(2): 164–182.

About the Authors

Zoe Fewell is a PhD student in the Department of Social Medicine, University of Bristol, UK. Her thesis topic is causal modeling in epidemiology with specific interests in compliance in randomized controlled trials and confounding in observational epidemiological studies.

Miguel Hernán is Assistant Professor of Epidemiology at the Harvard School of Public Health. He conducts research on the development and application of analytic methods for drawing causal inferences from complex longitudinal data with time-varying exposures.

Fred Wolfe is a rheumatologist interested in long-term outcomes of rheumatic diseases. He is the director of the National Data Bank for Rheumatic Diseases (NDB). The NDB collects, manages, analyzes, and publishes data using NDB-written Stata programs.

Kate Tilling is a Senior Lecturer in Medical Statistics in the Department of Social Medicine, University of Bristol, UK. Her research interests include statistical methods for epidemiology and health services research, causal models, longitudinal analysis, and the epidemiology of stroke and cardiovascular disease.

Hyon Choi, MD, DrPH is the Director of Outcomes Research at the Rheumatology Unit of Massachusetts General Hospital, Harvard Medical School, Boston, MA. He conducts clinical and epidemiologic research on rheumatic disorders including the application of causal modeling in rheumatic disease research.

Jonathan Sterne is a Reader in Medical Statistics and Epidemiology in the Department of Social Medicine, University of Bristol, UK. His research interests include statistical methods for epidemiology and health services research, causal models, meta-analysis and systematic reviews, and the epidemiology of HIV and other sexually transmitted infections.