Analysis of Longitudinal Observational Data Using Marginal Structural Models

Douglas E. Faries
Zbigniew A. Kadziola

Abstract

Assessing treatment effectiveness in longitudinal, observational data can be complex because in observational treatment patients can change medications at any time. In addition to the need to control for selection bias at baseline due to the lack of randomization, time-varying confounders can influence treatment changes over time and, thus, affect treatment group effectiveness comparisons. One approach to producing causal treatment effect estimates—even in the presence of treatment switching, missing data, and time-varying confounders—is to use marginal structural models. To illustrate, simulated data based on an observational schizophrenia study were analyzed using a marginal structural model approach. SAS code for performing the analysis is provided, and output using data from the schizophrenia study is examined.

9.1   Introduction

Assessing the causal effect of medications in longitudinal, observational (naturalistic) data presents analytical challenges—including the need to address selection bias; missing data; and switching, stopping, and augmenting medications. Addressing the issue of selection bias is critical because treatment groups likely differ in aspects other than treatment choice, and adjustment in the analysis is necessary (Rosenbaum and Rubin, 1983; Grimes and Schulz, 2002; Haro et al., 2006). In addressing data that are both longitudinal and observational, the issue of selection bias also extends to treatment switching over time (Robins et al., 2000, Hernán et al., 2000) as patients may switch, stop, augment, or otherwise not comply for a variety of reasons. In addition, such patient/physician choices are typically based upon stochastic and/or time-varying factors that may well differ among treatments. Because of such issues, statistical methods commonly used for longitudinal analyses of randomized clinical trial data, such as intent-to-treat

(ITT) last observation carried forward (LOCF) or repeated measures models, may not be appropriate.

ITT analyses group patients based only on their initial treatment assignment and ignore all information on other medications prescribed or taken. Patient dropout in such studies is often addressed by utilizing a LOCF approach. Clearly, such a technique does not directly address treatment effectiveness when there has been a substantial amount of switching among treatments. ITT analyses certainly have their place in longitudinal, observational research, such as in studies to compare policies or treatment strategies where one is not primarily interested in the effects of individual medications (Tunis et al., 2006).

While utilizing repeated measures models with treatment as a time-dependent variable may seem to provide a simple solution, Hernán and colleagues (2004, 2005) explain that such an approach does not provide estimates with a causal interpretation (see the following) in the presence of time-dependent confounders (a predictor of subsequent outcome and subsequent treatment) that are also affected by prior treatment. For instance, any longitudinal measure of disease severity would likely be problematic because it could be associated with the outcome measure, it could predict subsequent treatment, and it could have been affected by prior treatment. Thus, even if treatment is randomized at the beginning of a study, the result of usual-care treatment over time will ultimately result in imbalance in key patient characteristics among treatment groups. To address the switching of treatments, one could ignore the data after the medication switch and use standard repeated measures mixed models that have proven very useful in longitudinal data analyses (Verbeke and Molenberghs, 2000; Mallinckrodt et al., 2003). Such an approach treats the data after the switch as missing data but clearly does not make use of the information gathered after the medication switch.

In this chapter, we examine the use of marginal structural models (MSMs) for longitudinal, observational data. To explain the potential benefits of the MSM approach, we first must briefly review the notions of counterfactual outcomes and causal effect. We will follow the notation provided by Hernán and colleagues (2002). Let $\bar{a}$ denote the treatment history for a patient over a period of time (for example, $\bar{a} = [a(1), a(2), \ldots, a(t)]$, where $a(1)$ denotes the treatment used at time 1). A counterfactual outcome for patient i on treatment sequence $\bar{a}$ denotes that patient's outcome if, possibly contrary to fact, the patient received treatment $\bar{a}$. It is denoted by $Y_{\bar{a},i}(t)$. Each patient has an unknown counterfactual outcome for each treatment he did not receive, plus an observed outcome for the treatment actually received. On an individual basis, treatment is said to have a causal effect on a patient's outcome if $Y_{\bar{a},i}(t) \neq Y_{\bar{a}',i}(t)$ for some time point t and treatment patterns $\bar{a}$ and $\bar{a}'$. That is, the outcome for the patient differs based on the treatment taken. On a population basis, treatment is said to have a causal effect on outcome if the mean outcome had all patients followed a particular treatment pattern ($\bar{a}$ for example) differs from the mean outcome had all patients followed a different treatment pattern ($\bar{a}'$ for example) (that is, $E[Y_{\bar{a}}(t)] \neq E[Y_{\bar{a}'}(t)]$ for some time point and treatment pattern).

Robins and colleagues (1999) demonstrated that MSMs, under a set of assumptions discussed here, produce consistent estimates of the average causal treatment effects—even in the presence of treatment changes, time-dependent confounders, and missing at random study dropout. In this chapter, we first describe the MSM approach (Section 9.2) and then present the MSM analysis of a longitudinal schizophrenia study (Section 9.3). SAS code is provided, and SAS output is discussed to allow readers to understand the implementation of the analysis and to modify the

code for their own use. Faries and colleagues (2007) also summarize data from this study using a variety of methods, including MSMs. Some other applications of MSMs in the literature include Hernán and colleagues (2000, 2002), Ko and colleagues (2003), Brumback and colleagues (2004), and Cole and colleagues (2005), who examined time to event outcomes for HIV patients; and Bodnar and colleagues (2004), Yamaguchi and Ohashi (2004a and 2004b), Mortimer and colleagues (2005), Suarez and colleagues (2006), Peterson and colleagues (2007), and Vansteeldandt and colleagues (2009), who assessed other applications.

9.2 MSM Methodology

An MSM analysis is basically a weighted repeated measures approach – using treatment as a time-varying covariate. Weights, based on inverse probability of treatment weighting, control for time-dependent confounders and essentially produce a pseudo-population with balance in both time-invariant and time-varying covariates allowing for causal treatment comparisons using standard repeated measure models. The weighting can also be adjusted to incorporate adjustments for missing data—providing validity under missing at random (MAR; missing data may depend upon observed but not unobserved measures) and missing completely at random (MCAR; missing data does not depend upon observed or unobserved measures) data.

Conducting an MSM analysis is a two-step process. First, one estimates two weights for each observation (patient visit): one adjusting for treatment selection and one adjusting for study discontinuation. Computation of these estimated weights can incorporate time-independent and time-dependent factors. The stabilized weight is recommended by Hernán (2002), and we use the notation from that manuscript (here for the treatment selection weight),

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

where $A(k)$ represents the treatment at time k and $\overline{A}(k-1)$ represents the treatment history prior to time k, V represents a vector of time-independent variables (baseline covariates), and $\overline{L}(k)$ represents a vector of time-varying covariates through time k—which includes baseline variables V. The numerator of the weight is the probability a patient is on the observed treatment at time k, given the prior treatment history and baseline covariates. The denominator is basically the same factor, except it incorporates time-varying covariates as predictors. Thus, one can see that observations where the time-varying factors are strong predictors of the current treatment selection are down-weighted in the analyses (because such observations are over-represented in the observed data).

To incorporate adjustment for early patient dropout, the same stabilized weight approach is used—except the outcome is not treatment selection but a flag variable denoting whether the patient remained in the study. The final weight for each patient's observation is obtained by multiplying the treatment selection weights and the censoring weights.

For the second step of the MSM analysis, one simply conducts a weighted repeated measures model analysis using generalized estimating equations. In this second stage, time-dependent confounders are not included in the repeated measures model—as their effects have been incorporated into the weights. Treatment is included as a time-dependent factor, and time-invariant covariates may also be included as appropriate (just as in a cross-sectional analysis where variables may be included in a propensity model and in the analysis model).

As mentioned previously, MSMs can produce consistent estimates of the average causal treatment effects—even in the presence of treatment changes, time-dependent confounders, and missing at random study dropout. The assumptions necessary for causal inference from an MSM correspond to the same assumptions necessary for common cross-sectional bias control methods such as propensity scoring:

1. no unmeasured confounders—that is, all variables that relate to treatment assignment and outcome were collected and utilized in the analysis; called *conditional exchangeability*; formally: $Y_{\bar{a}}(t+1) \coprod A(k) \big| \bar{A}(k-1), \bar{L}(k)$, for all $\bar{a}$ and $t \geq k$.

2. positivity—there is a positive probability of each treatment for each set of covariates (no perfect confounding); formally $f[\bar{a}(k-1), \bar{l}(k)] > 0 \Rightarrow f[a(k)\big|\bar{a}(k-1), \bar{l}(k)] > 0$.

3. use of the correct models (weighting and analysis models).

As strong assumptions are necessary, assessing the appropriateness of the assumptions and performing sensitivity analyses are critical to a quality analysis. The no unmeasured confounders and correct models assumptions can never be fully proven and are discussed here. The positivity assumption basically says that all treatment options are possible given any combination of covariate values. Mortimer and colleagues (2005) recommended assessing this by computing predicted probabilities of treatment selection using the covariates from the models across the entire study (looking for 0 or 1 predicted probabilities). Mortimer and colleagues (2005) also provide an example of assessing the correctness of the models used in an MSM analysis using test and training data sets.

To limit the possibility of unmeasured confounding, every effort should be made to identify and collect data on potential confounders by searching the literature, examining relevant data, having discussions with experts, and utilizing potential confounding variables in the analysis. Such diligence will still never allow one to completely conclude no bias is unaccounted for, but an analyst first needs to make sure that all known confounders are addressed. Robins and colleagues (1999) and Brumback and colleagues (2004) have also provided a more formal method to study the sensitivity of an MSM analysis to unmeasured confounding. They quantify such confounding through a sensitivity parameter (alpha) and confounding function and assess the amount of unmeasured confounding that can be present before inferences would change. The confounding function (or alpha itself when a simple constant function is used) represents the difference in potential outcomes between patients in the different treatment groups. We use a simple constant function in our analysis, referring the reader to Brumback and colleagues (2004) for more options.

9.3 Example: MSM Analysis of a Simulated Schizophrenia Trial

9.3.1 Study Description

To illustrate the implementation of an MSM analysis, we simulated data based on a study of the effectiveness of medications for patients with schizophrenia in usual-care settings. A brief description of the design for the actual study follows, though the reader is referred to Tunis and colleagues (2006) for details. This was a one-year study of patients with schizophrenia, schizoaffective disorder, or schizophreniform disorder who were randomized to one of three

different treatment regimens. After randomization, the remainder of the study was observational in the sense that physicians/patients were allowed to stop or switch medications as deemed necessary in usual practice. Data on a variety of domains were captured at five post-baseline visits (approximately 2 weeks and 2, 5, 8, and 12 months post-baseline). The outcome measure of interest for this analysis was the Brief Psychiatric Rating Scale (BPRS) total score, a measure of schizophrenia symptom severity where lower scores indicate lesser symptom severity. For demonstrative purposes, we simulated data for this analysis (see Tunis et al., 2006, for actual data results) maintaining the design and data structure and focused the final comparison between two groups rather than three. These groups are referred to as *treatment* and *control* during this example analysis. However, each treatment is considered as a separate treatment in the analytical steps until the final model to demonstrate how one can handle more than two groups with the MSM approach.

9.3.2 Data Analysis

9.3.2.1 Data Overview
Before conducting the MSM analysis, we provide a brief summary of the simulated data pertaining to medication changes and the steps taken to prepare the data set for analysis. The treatment groups were balanced with respect to demographics and baseline patient characteristics due to the randomization. After randomization, study discontinuation was similar across the treatment groups though rates of switching medication differed, with almost half (47.7%) of the control group (treatment C) switching medications during the study with a lower rate for the treatment group (treatment groups A and B pooled; 20.4%). At each visit, patients were considered to be on a particular medication if they had been treated with that medication for at least the previous 14 days. Given this definition, on 41.6% of the 2,548 patient visits during the study, patients were taking treatment A, 25.1% were taking treatment B, 23.8% were taking treatment C, 4.8% were taking both A and C, and 4.8% were not taking any antipsychotic medication. A total of 17 patient visits were excluded from the analysis due to small sample sizes for certain treatment combinations (A and B n=16; A and B and C n=1). In addition, approximately 4% of the patient visits had missing covariate information (see the list of covariates later) that was imputed using a LOCF approach. Outcome data were not imputed, only covariate data and only when the outcome measure was available. The analysis data set, INPDS, used a one observation per patient per visit format. A description of the key variables follows.

Table 9.1 Description of Key Variables in MSM Analysis

Variable	Description	Variable	Description
INVSC	Investigator number	BPRS	BPRS at this visit
PATSC	Patient number	GAFC	GAF at this visit
AGEYRS	Age in years	EVNT	Events during visit
GENDER	Gender	HOSP	Hosp. during visit
ORIGIN2	Race	PR1TRTA	Trt A previous visit
THERAPY	Randomized trtmnt	PR1TRTB	Trt B previous visit
VIS	Visit number	PR1TRTC	Trt C previous visit
BAVAR	Baseline BPRS total	PR1BPRS	Previous vis BPRS
BGAF	Baseline GAF score	PR1GAF	Previous vis GAF
BEVNT	Baseline adv events	PR1EVNT	Previous vis event
BHOSP	Baseline hospitaliz.	PR1HOSP	Previous vis hosp

(*continued*)

Table 9.1 (*continued*)

Variable	Description	Variable	Description
TRTA	On Trt A during vis		
TRTB	On Trt B during vis		
TRTC	On Trt C during vis		
TRT	Treatment during vis		

9.3.2.2 Computation of Inverse Weights

Step one in conducting the MSM analysis requires estimation of the treatment selection and censoring weights (using the formula for stabilized weights [SW] from Section 9.2). For estimating the treatment selection weights, separate multinomial models for the numerator and denominator were implemented using PROC LOGISTIC with the LINK= GLOGIT option (see Program 9.1). Treatment was the dependent variable for both models and the GLOGIT option was used because the treatment choice at each visit had five potential outcomes (three individual treatments, no treatment, and one combination). Output from the LOGISTIC procedure contains the predicted probabilities of treatment selection (data sets PREDTRT0 and PREDTRT1). In addition to previous treatment, the following time-varying covariates were included in this model:

- BPRS total

- global assessment of functioning (GAF)

- events (the presence or absence of at least one moderate or severe adverse event)

- hospitalization

These variables were chosen a priori to cover the domains of symptom severity, functioning, tolerability, and resource utilization. Time-independent variables included in this initial model included age, gender, ethnicity, and baseline value for each of the time-dependent variables. Macro variables could be utilized to input all model parameters at the beginning of the code; however, we chose to simplify the understanding of the process by simply showing the models directly in the LOGISTIC statements. In addition, output from the weight models is suppressed for simplification of the output. However, one can easily remove the restriction to evaluate the weight model in more detail.

PROC GENMOD was used to compute the estimated stabilized weights to adjust for censoring using a logistic regression model (see Program 9.1). The dependent variable for the censoring weight model was a binary flag for remaining in the study. The independent variables for this model were the same as for the treatment selection weight model—though the time-varying covariates were offset by one visit due to the structure of the data as censoring looked forward (did the patient return for a following visit?) relative to treatment (what was the treatment assigned in the previous time period?). The Logit LINK function was used here because the outcome measure was binomial. The GENMOD approach could be used for both weight calculations in studies where only two treatment groups are assessed.

Partial output from the multinomial and logistic regression denominator weights models is provided in the Output from Program 9.1. Previous treatment was the strongest predictor of present treatment. None of the time-varying covariates were strong predictors of treatment changes—suggesting bias in treatment selection over time may not be particularly strong in these data. Along with previous treatment, higher (time-varying) symptom severity was found to be a predictor of censoring. Patients with more severe symptoms were more likely to discontinue from the study.

Program 9.1 Computing Treatment Selection and Censoring Weights

```
/* This section of code computes the treatment selection and censoring
weights. This is accomplished in 4 steps:
1) multinomial model to compute numerator of treatment selection weights;
2) multinomial model to compute denominator of treatment selection weights;
3) binomial model to compute numerator of censoring adjustment weights;
4) binomial model to compute denominator of censoring adjustment weights.*/

/* treatment selection weights:  numerator calculation
(probability of treatment using only baseline covariates) */
PROC LOGISTIC DATA = INPDS;
  CLASS VIS THERAPY PR1TRTA PR1TRTB PR1TRTC GENDER ORIGIN2 BHOSP BEVNT;
  MODEL TRT = VIS THERAPY PR1TRTA PR1TRTB PR1TRTC GENDER ORIGIN2 BHOSP
  BEVNT AGEYRS BGAF BBPRS
    /LINK=GLOGIT;
  OUTPUT OUT=PREDTRT0(WHERE=(TRT=_LEVEL_)) PRED=PREDTRT0;
run;

/* treatment selection weights:  denominator calculation
(probability of treatment with baseline covariates and time-dependent
covariates) */
PROC LOGISTIC DATA = INPDS;
  CLASS VIS THERAPY PR1TRTA PR1TRTB PR1TRTC GENDER ORIGIN2 BHOSP BEVNT
        PR1EVNT PR1HOSP;
  MODEL TRT = VIS THERAPY PR1TRTA PR1TRTB PR1TRTC GENDER ORIGIN2 BHOSP
            BEVNT AGEYRS BGAF BBPRS PR1EVNT PR1HOSP PR1BPRS PR1GAFC
    /LINK=GLOGIT;
  OUTPUT OUT=PREDTRT1(WHERE=(TRT=_LEVEL_)) PRED=PREDTRT1;
run;

/* censoring adjustment weights:  numerator calculation
(probability of censoring using only baseline covariates) */
ODS LISTING EXCLUDE OBSTATS;
PROC GENMOD DATA = INPDS;
  CLASS PATSC   VIS THERAPY TRTA TRTB TRTC GENDER ORIGIN2 BHOSP BEVNT;
  MODEL CFLAG = VIS THERAPY TRTA TRTB TRTC GENDER ORIGIN2 BHOSP BEVNT
                AGEYRS BGAF BBPRS
    /DIST = BIN LINK = LOGIT TYPE3 OBSTATS;
  REPEATED SUBJECT = PATSC / TYPE = EXCH;
  ODS OUTPUT OBSTATS = PREDCEN0(RENAME=(PRED=PREDCEN0));
run;
ODS LISTING SELECT ALL;

/* censoring adjustment weights:  denominator calculation
(probability of censoring using baseline covariates and time-dependent
covariates) */
ODS LISTING EXCLUDE OBSTATS;
PROC GENMOD DATA = INPDS;
  CLASS PATSC VIS THERAPY TRTA TRTB TRTC GENDER ORIGIN2 BHOSP BEVNT
  EVNT HOSP;
  MODEL CFLAG = VIS THERAPY TRTA TRTB TRTC GENDER ORIGIN2 BHOSP BEVNT
          EVNT HOSP AGEYRS BGAF BBPRS BPRS GAFC
    /DIST = BIN LINK = LOGIT TYPE3 OBSTATS;
  REPEATED SUBJECT = PATSC / TYPE = EXCH;
  ODS OUTPUT OBSTATS = PREDCEN1(RENAME=(PRED=PREDCEN1));
run;
ODS LISTING SELECT ALL;
```

Output from Program 9.1

```
                    weight denominator model: treatment weights

                        Type 3 Analysis of Effects

                                      Wald
                Effect        DF   Chi-Square    Pr > ChiSq

                VIS           16    124.2743       <.0001
                THERAPY        8     11.8349       0.1587
                PR1TRTA        4    229.7185       <.0001
                PR1TRTB        4    100.6526       <.0001
                PR1TRTC        4    269.8939       <.0001
                GENDER         4      3.3971       0.4937
                ORIGIN2        8     10.4911       0.2322
                BHOSP          4      0.4611       0.9772
                BEVNT          4      3.0900       0.5429
                AGEYRS         4      1.4473       0.8359
                BGAF           4      1.9736       0.7406
                BBPRS          4      2.2390       0.6919
                PR1EVNT        4      7.5905       0.1078
                PR1HOSP        4      4.6487       0.3253
                PR1BPRS        4      2.5962       0.6275
                PR1GAFC        4      2.5983       0.6271


                weight denominator model: censoring weights

                 Score Statistics For Type 3 GEE Analysis

                                      Chi-
                Source          DF   Square    Pr > ChiSq

                VIS              4    13.21       0.0103
                THERAPY          2     9.04       0.0109
                TRTA             1     3.95       0.0468
                TRTB             1     3.74       0.0532
                TRTC             1     0.39       0.5346
                GENDER           1     0.15       0.6964
                ORIGIN2          2     0.99       0.6087
                BHOSP            1     0.88       0.3473
                BEVNT            1     0.93       0.3338
                EVNT             1     0.61       0.4338
                HOSP             1     0.05       0.8163
                AGEYRS           1     4.89       0.0270
                BGAF             1     0.04       0.8431
                BBPRS            1     0.67       0.4115
                BPRS             1    12.21       0.0005
                GAFC             1     0.32       0.5726
```

To produce the overall weights for each observation (patient visit) in this analysis, the inverse probability weights for treatment selection and censoring computed here were multiplied together cumulatively in a DATA step based on the formula for SW in Section 9.2 (see Program 9.2). In the SAS code, the SQL procedure gathers data from the four output data sets from the treatment selection and weight models (PREDTRT0, PREDTRT1, PREDCENS0, and PREDCENS1) and produces the stabilized weights. Variable STABWT is the final estimate of SW. Output from Program 9.2 displays the distribution of the final weights across all patient visits in box plot form. The mean is near 1 (mean = 1.002, SD = 0.1555), as would be expected from the average of weights, and no major outliers were noted in the box plot.

Program 9.2 Merging Output from Program 9.1

```
/* This section of code performs the steps necessary to merge the output
from the weight models (Program 9.1) to allow for computation of a single
adjustment for each observation in the analysis data set (stabilized
weight). This is followed by code to produce summaries of the final
weights.   */

PROC SQL;
  /*ratio of probabilities for treatment*/
  CREATE TABLE PREDTRT AS
    SELECT *,PREDTRT0/PREDTRT1 AS PREDTRT
    FROM PREDTRT1(KEEP=PATSC VIS PREDTRT1)
         NATURAL FULL JOIN
         PREDTRT0(KEEP=PATSC VIS PREDTRT0)
    ORDER PATSC,VIS
  ;
  /*ratio of probabilities for censoring*/
  CREATE TABLE PREDCEN AS
    SELECT *,PREDCEN0/PREDCEN1 AS PREDCEN
    FROM (SELECT INPUT(PATSC,BEST.) AS PATSC, INPUT(VIS,BEST.) AS VIS,
PREDCEN0 FROM PREDCEN0)
         NATURAL FULL JOIN
         (SELECT INPUT(PATSC,BEST.) AS PATSC, INPUT(VIS,BEST.) AS VIS,
PREDCEN1 FROM PREDCEN1)
    ORDER PATSC,VIS;
QUIT;

/*calculate stabilized weight*/
PROC SORT DATA=INPDS OUT=WEIGHTS;
  BY PATSC VIS;
RUN;

DATA WEIGHTS;
  MERGE WEIGHTS PREDTRT PREDCEN;
  BY PATSC VIS;
  VWT=PREDTRT*PREDCEN;
  IF FIRST.PATSC THEN STABWT=VWT;
                 ELSE STABWT=VWT*DUM;
  RETAIN DUM;
  DROP DUM;
  DUM=STABWT;
RUN;

/*diagnostic plot for weights*/
DATA GRPH;
  SET WEIGHTS; /*assignment of months to visits is study-specific*/
  IF VIS = 3 THEN MONTH = 0.5;
  IF VIS = 4 THEN MONTH = 2;
  IF VIS = 5 THEN MONTH = 5;
  IF VIS = 6 THEN MONTH = 8;
  IF VIS = 7 THEN MONTH = 12;

  IF MONTH = 0.5 THEN DELETE; /*simplify plot by focusing on months
     with greater switching and thus greatest variability in weights */
RUN;

PROC SORT DATA = GRPH;
  BY MONTH;
RUN;
```

```
ODS RTF FILE="%SYSFUNC(PATHNAME(WORK))\FIG1.RTF";

FILENAME FIGURE "%SYSFUNC(PATHNAME(WORK))\SASGRAPH.EMF";

GOPTIONS RESET=ALL TARGET=SASEMF DEVICE=SASEMF FTEXT=DUPLEX HTEXT=.75
    CBACK=WHITE XMAX=6IN XPIXELS=1200 YMAX=5IN YPIXELS=1000
    GSFNAME=FIGURE GSFMODE=REPLACE;

SYMBOL1 COLOR=BLACK INTERPOL=JOIN
        WIDTH=2 VALUE=SQUARE
        HEIGHT=1;

AXIS1 MINOR = NONE COLOR = BLACK LABEL=("STABILIZED WEIGHT" ANGLE=90
ROTATE=0);

PROC BOXPLOT DATA=GRPH;
  PLOT STABWT*MONTH / CFRAME = WHITE
                      CBOXES = DAGR
                      CBOXFILL = WHITE
                      VAXIS = AXIS1;
  TITLE "SUMMARY OF VISITWISE WEIGHT VALUES";
  TITLE2 "(box and whiskers: min, 1st quartile, median, 3rd quartile, max;
square: mean)";
RUN;

GOPTIONS RESET=ALL;

ODS RTF CLOSE;
```
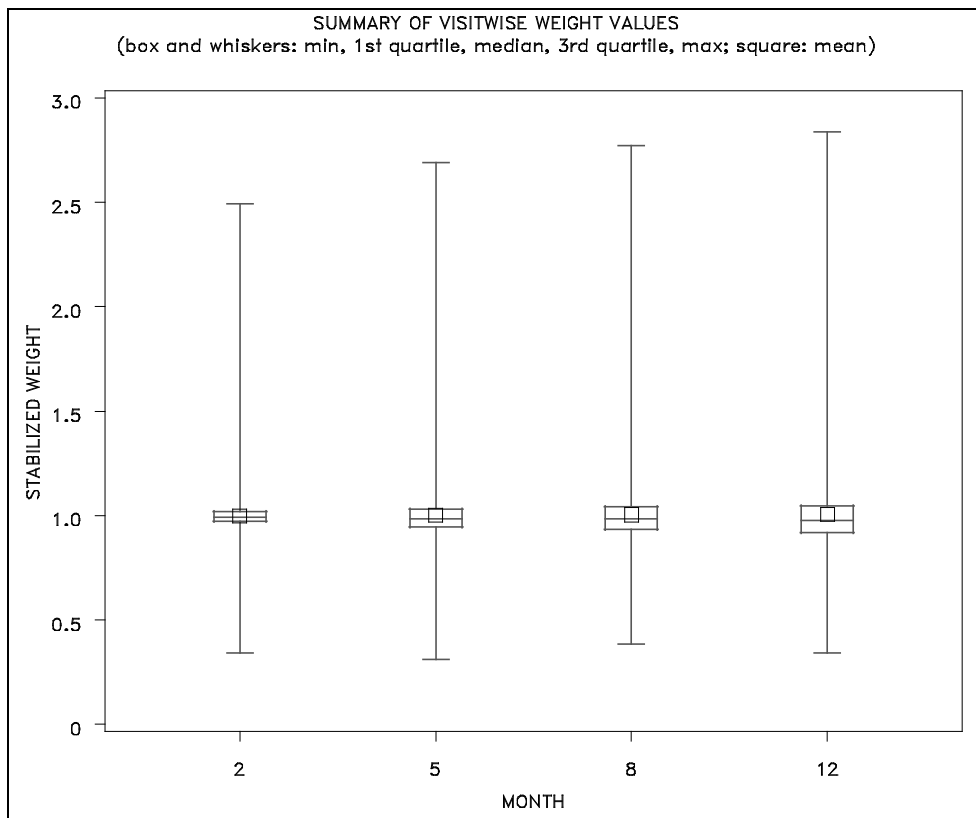
Output from Program 9.2

9.3.2.3 Treatment Effectiveness Analysis Model

The final step of the MSM analysis is to estimate causal treatment effects using a weighted repeated measures model with generalized estimating equations (PROC GENMOD—see Program 9.3) and an exchangeable correlation matrix. Change from baseline BPRS total score was the dependent measure in the analysis model. Independent variables for the analysis model were investigational site, age, gender, race, baseline BPRS, visit, time-varying treatment, and the treatment-by-visit interaction. The WEIGHT statement in PROC GENMOD incorporates the inverse probability weighting, which allows for the causal treatment effect estimates. The ESTIMATE statement utilized in PROC GENMOD pooled individual treatments together and produced estimated mean group differences for pooled groups (as opposed to comparing individual treatment groups). This portion of the code is not necessary for many applications. Output from Program 9.3 displays a summary of the final model results and a figure displaying the least squares means at each visit. Results showed a statistically significant treatment difference favoring the treatment group (pooled treatments A and B) relative to the control group, with an estimated average treatment difference in BPRS changes of 1.8 [0.4, 3.2], *p*=.015 across the 1-year period. Though treatment differences grew numerically over time, the treatment-by-visit interaction term was not statistically significant (*p*=.158).

Program 9.3 Running Final Analysis Model

```
/* This section of code runs the final analysis model using a weighted
repeated measures approach. The results are presented graphically using
PROC GPLOT.*/

/*final analysis model*/
PROC GENMOD DATA = WEIGHTS;
  CLASS VIS PATSC GENDER ORIGIN2 INVSC TRT;
  WEIGHT STABWT;
  MODEL CAVAR = INVSC BAVAR VIS AGEYRS GENDER ORIGIN2 TRT VIS*TRT
                / DIST=NORMAL LINK=ID TYPE3;
  REPEATED SUBJECT = PATSC / TYPE=EXCH;
  LSMEANS TRT VIS*TRT / PDIFF;
  TITLE 'MSM FINAL ANALYSIS MODEL';
  ESTIMATE 'A+B VS C'
    TRT 0 .5 .5 -1 0;
  ESTIMATE 'A+B VS C AT VIS 3'
    TRT 0 .5 .5 -1 0
    VIS*TRT 0 .5 .5 -1 0     0 0 0 0 0     0 0 0 0 0
            0  0  0  0 0     0 0 0 0 0;
  ESTIMATE 'A+B VS C AT VIS 4'
    TRT 0 .5 .5 -1 0
    VIS*TRT 0 0 0 0 0         0 .5 .5 -1 0      0 0 0 0 0
            0 0 0 0 0         0  0  0  0 0;
  ESTIMATE 'A+B VS C AT VIS 5'
    TRT 0 .5 .5 -1 0
    VIS*TRT 0 0 0 0 0         0 0 0 0 0         0 .5 .5 -1 0
            0 0 0 0 0         0 0 0 0 0;
  ESTIMATE 'A+B VS C AT VIS 6'
    TRT 0 .5 .5 -1 0
    VIS*TRT 0 0 0 0 0         0 0 0 0 0         0 0 0 0 0
            0 .5 .5 -1 0   0 0 0 0 0;
  ESTIMATE 'A+B VS C AT VIS 7'
    TRT 0 .5 .5 -1 0
    VIS*TRT 0 0 0 0 0         0 0 0 0 0         0 0 0 0 0
            0 0 0 0 0         0 .5 .5 -1 0;
  ODS OUTPUT LSMEANS=LSMEANS;
RUN;
```

```
/*LS means plot for the final model*/
DATA LSMEANS2;
  SET LSMEANS;
  WHERE TRT IN ('A  ',' B ','  C');
  IF TRT IN ('A  ',' B ') THEN TRT2='A+B';
                          ELSE TRT2='C  ';
RUN;

PROC SQL;
  CREATE TABLE LSMEANS3 AS
    SELECT TRT2 AS TRT, VIS, MEAN(ESTIMATE) AS ESTIMATE
    FROM LSMEANS2
    GROUP TRT2, VIS;
QUIT;

ODS RTF FILE="%SYSFUNC(PATHNAME(WORK))\FIG2.RTF";

FILENAME FIGURE "%SYSFUNC(PATHNAME(WORK))\SASGRAPH.EMF";

GOPTIONS RESET=ALL TARGET=SASEMF DEVICE=SASEMF FTEXT=DUPLEX HTEXT=.75
CBACK=WHITE
         XMAX=6IN XPIXELS=1200 YMAX=5IN YPIXELS=1000 GSFNAME=FIGURE
GSFMODE=REPLACE;

AXIS1 MINOR = NONE COLOR = BLACK LABEL=(ANGLE=90 ROTATE=0 "CHANGE IN BPRS
TOTAL SCORE");

SYMBOL1 I=JOIN W=2 L=1 C=RED   V=SQUARE;
SYMBOL2 I=JOIN W=2 L=2 C=BLACK V=CIRCLE;

PROC GPLOT DATA=LSMEANS3;
  PLOT ESTIMATE*VIS=TRT/VAXIS=AXIS1;
  TITLE "MSM ESTIMATED MEAN CHANGE FROM BASELINE BPRS SCORES";
  LABEL VIS="VISIT";
  LABEL TRT="TREATMENT";
RUN;

GOPTIONS RESET=ALL;

ODS RTF CLOSE;
```

Output from Program 9.3

```
                        MSM FINAL ANALYSIS MODEL
                    Score Statistics For Type 3 GEE Analysis


                                       Chi-
                  Source          DF   Square    Pr > ChiSq

                  INVSC           19   84.25       <.0001
                  BAVAR            1   71.12       <.0001
                  VIS             4   37.69       <.0001
                  AGEYRS           1    0.80       0.3725
                  GENDER           1    0.09       0.7624
                  ORIGIN2          2    0.86       0.6500
                  TRT              4   11.70       0.0197
                  VIS*TRT         16   21.57       0.1576



                          Contrast Estimate Results


                        Standard                            Chi-
  Label                 Estimate  Error   Alpha  Confidence Limits    Square   Pr > ChiSq

  A+B VS C               -1.7852  0.7321  0.05   -3.2201   -0.3504     5.95      0.0147
  A+B VS C AT VIS 3      -0.7336  0.9615  0.05   -2.6181    1.1509     0.58      0.4455
  A+B VS C AT VIS 4      -1.1225  0.9417  0.05   -2.9682    0.7232     1.42      0.2333
  A+B VS C AT VIS 5      -1.7746  1.2532  0.05   -4.2307    0.6815     2.01      0.1567
  A+B VS C AT VIS 6      -2.1785  1.2752  0.05   -4.6778    0.3208     2.92      0.0876
  A+B VS C AT VIS 7      -3.1169  1.3185  0.05   -5.7012   -0.5327     5.59      0.0181
```


MSM ESTIMATED MEAN CHANGE FROM BASELINE BPRS SCORES
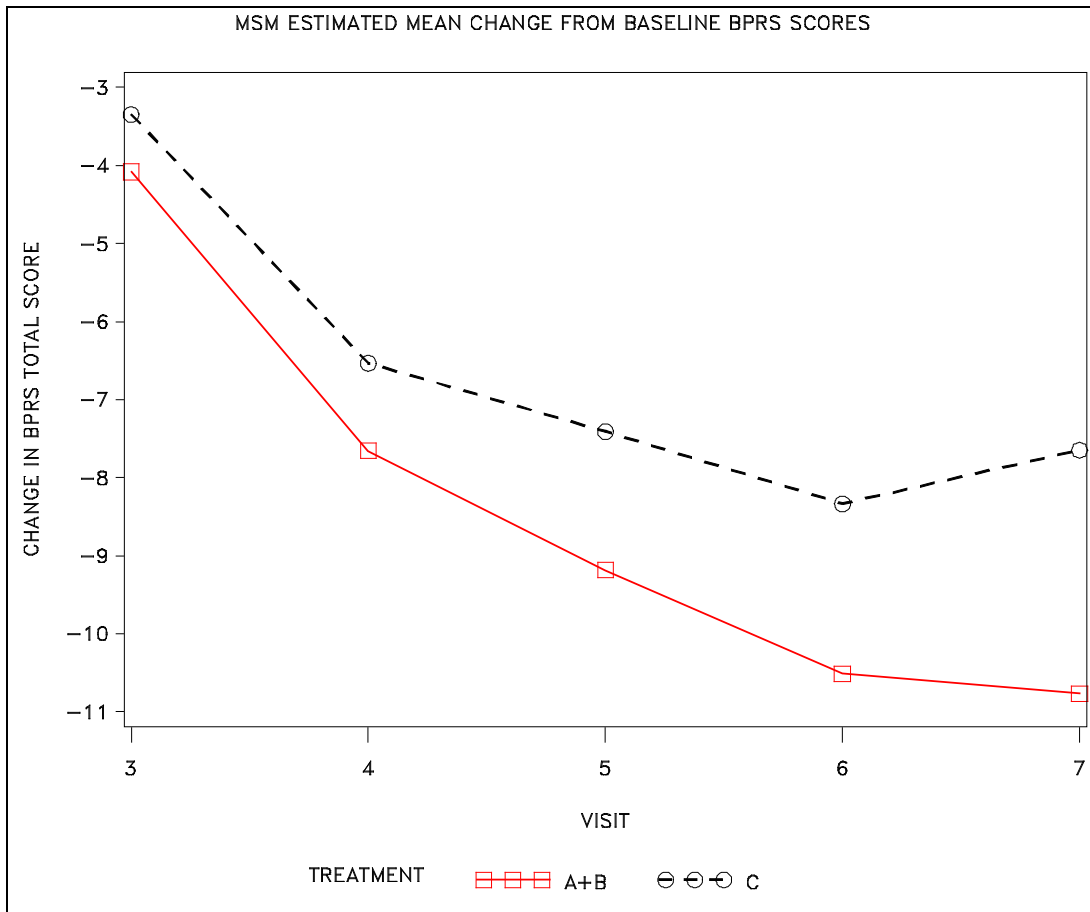
9.3.2.4 Sensitivity Analysis

To assess the robustness of the findings, one should evaluate the plausibility of the three main assumptions for causality (no unmeasured confounding, positivity, correct model) as well as implementing other statistical methods that are based on different assumptions. Regarding the unmeasured confounding assumption, effort was made to collect and incorporate information from experts and the literature on potential confounders prior to the analysis. The variables included in the model were selected in order to cover the domains of symptom severity, functioning, tolerability, and resource use burden. To assess the potential impact of unmeasured confounding on the results, we implemented a simple, unmeasured confounding function (we chose a constant function alpha) based on Brumback and colleagues (2004). Using this function, a missing confounder resulting in a shift of 0.45 (in BPRS total score) in potential outcomes on the BPRS scale favoring patients with high probability of being in the treatment group would result in a loss of the statistically significant finding. In other words, our finding of a significant difference between groups depends upon the assumption that such a confounder (or a group of confounders combining to have the same effect) does not exist. Multiple constants were evaluated using the ALPHA=c statement in the SAS code (see Program 9.4). The value of 0.45 was retained because this was the smallest value producing a p-value of approximately 0.05. While it is challenging to interpret the implications of a particular alpha, such an effect (<1 point on BPRS) does not appear to be extremely large, and the existence of such a confounder is certainly possible in an actual application. However, it is greater than the observed confounding (the difference between unweighted and weighted analyses) of 0.29 in the opposite direction observed in this study. Regardless of the sensitivity analysis results, one can simply not dismiss the possibility of unaccounted for factors that would result in this analysis failing to produce an estimate of a causal treatment effect. Thus, one must view the results with some caution. Different sensitivity functions can easily be tested using the provided SAS code by altering the calculation of the variables CAVAR_SENS and SAPT.

Program 9.4 Finding Level of Confounding That Would Eliminate Treatment Difference

```
/* This section of code examines the sensitivity of the results to
unmeasured confounding by finding the level of confounding that would
eliminate the observed treatment difference.*/

/*sensitivity analysis per Brumback et al. (2004) – constant
function alpha is used*/
DATA WEIGHTSS;
  SET WEIGHTS;
  BY PATSC VIS;
  IF TRT IN ('A__','_B_') THEN SNSTRT = 1;
                         ELSE SNSTRT = -1;
  SAVT = -SNSTRT*(1 - PREDTRT1);
  IF FIRST.PATSC THEN SAPT = SAVT;
               ELSE SAPT = SAVT + DUM;
  RETAIN DUM;
  DROP DUM;
  DUM=SAPT;
  ALPHA = 0.45;
  CAVAR_SENS = CAVAR - ALPHA*SAPT;
RUN;

PROC GENMOD DATA = WEIGHTSS;
  CLASS VIS PATSC GENDER ORIGIN2 INVSC TRT;
  WEIGHT STABWT;
  MODEL CAVAR SENS = INVSC BAVAR VIS AGEYRS GENDER ORIGIN2 TRT VIS*TRT
               / DIST=NORMAL LINK=ID TYPE3;
  REPEATED SUBJECT = PATSC / TYPE=EXCH;
  LSMEANS TRT VIS*TRT / PDIFF;
  TITLE 'MSM SENSITIVITY ANALYSIS: ALPHA = 0.45';
```

```
    ESTIMATE 'A+B VS C'
      TRT 0 .5 .5 -1 0;
    ESTIMATE 'A+B VS C AT VIS 3'
      TRT 0 .5 .5 -1 0
      VIS*TRT 0 .5 .5 -1 0      0 0 0 0 0      0 0 0 0 0
              0  0  0  0 0      0 0 0 0 0;
    ESTIMATE 'A+B VS C AT VIS 4'
      TRT 0 .5 .5 -1 0
      VIS*TRT 0 0 0 0 0          0 .5 .5 -1 0          0 0 0 0 0
              0 0 0 0 0          0  0  0  0 0;
    ESTIMATE 'A+B VS C AT VIS 5'
      TRT 0 .5 .5 -1 0
      VIS*TRT 0 0 0 0 0          0 0 0 0 0          0 .5 .5 -1 0
              0 0 0 0 0          0 0 0 0 0;
    ESTIMATE 'A+B VS C AT VIS 6'
      TRT 0 .5 .5 -1 0
      VIS*TRT 0 0 0 0 0          0 0 0 0 0          0 0 0 0 0
              0 .5 .5 -1 0       0 0 0 0 0;
    ESTIMATE 'A+B VS C AT VIS 7'
      TRT 0 .5 .5 -1 0
      VIS*TRT 0 0 0 0 0          0 0 0 0 0          0 0 0 0 0
              0 0 0 0 0          0 .5 .5 -1 0;
  RUN;
```

Output from Program 9.4

```
                    MSM SENSITIVITY ANALYSIS: ALPHA = 0.45

                          Contrast Estimate Results

                              Standard                                Chi-
  Label                 Estimate   Error   Alpha    Confidence Limits  Square   Pr > ChiSq

  A+B VS C               -1.3911  0.7303   0.05    -2.8224    0.0402    3.63      0.0568
  A+B VS C AT VIS 3      -0.6203  0.9615   0.05    -2.5048    1.2642    0.42      0.5188
  A+B VS C AT VIS 4      -0.8879  0.9417   0.05    -2.7335    0.9577    0.89      0.3457
  A+B VS C AT VIS 5      -1.3184  1.2534   0.05    -3.7750    1.1383    1.11       0.292
  A+B VS C AT VIS 6      -1.6479  1.2726   0.05    -4.1422    0.8464    1.68      0.1954
  A+B VS C AT VIS 7      -2.4810  1.3143   0.05    -5.0569    0.0950    3.56      0.0591
```

To assess the positivity assumption, we followed the ideas of Mortimer and colleagues (2005) and estimated the probability of selection of each of the five treatment possibilities using all possible covariates across all study visits. The CATMOD procedure is used here in order to generate predicted probabilities for all possible treatment options. While theoretically there were no issues with the positivity assumption in this study (all patients had the opportunity of being switched from and to any combination of treatments at any time), the goal was to assess whether any observed set of covariates produced a predicted probability of 0 or 1 for this set of data. This was done as shown with Program 9.5 where we computed the predicted probabilities of treatment for each observed set of covariates and then summarized the predicted probabilities (summary statistics using PROC MEANS by treatment). We observed that the smallest (nonzero but <.00015) probabilities were for the theoretical switch from no treatment to the combination of treatments A and C at earlier study visits. In general, the smaller probabilities were associated with the no treatment and combination treatment groups—as would be expected given their observed frequencies. Sensitivity analyses were then performed without the more extreme records and then with combination patients re-assigned to the initial randomized single treatment (for example, a patient randomized to treatment A but treated with A and C would be counted as being treated only with treatment A). No major changes in the inferences were observed from these sensitivity analyses and these results are not shown.

Program 9.5 Examining Positivity Assumption for Causal Inference

```
/* This section of code examines the positivity assumption for causal
inference by examining the predicted values of all possible treatment
changes. Summary statistics are presented as well as a listing to allow
examination of outliers.*/

/*positivity check Mortimer (2005) - generate predicted probabilities*/
ODS LISTING CLOSE;
PROC CATMOD DATA = INPDS;
  DIRECT AGEYRS BGAF BBPRS PR1BPRS PR1GAFC;
  MODEL TRT = VIS THERAPY PR1TRTA PR1TRTB PR1TRTC GENDER ORIGIN2 BHOSP
            BEVNT AGEYRS BGAF BBPRS PR1EVNT PR1HOSP PR1BPRS PR1GAFC
                /PRED=PROB;
  ODS OUTPUT PREDICTEDPROBS=PREDTRT1POS(KEEP=VIS TRT THERAPY PR1TRTA
              PR1TRTB PR1TRTC GENDER ORIGIN2 BHOSP BEVNT AGEYRS BGAF
          BBPRS PR1EVNT PR1HOSP PR1BPRS PR1GAFC
              SAMPLE OBSFUNCTION PREDFUNCTION
               RENAME=(PREDFUNCTION=PREDTRT1));
RUN;
ODS LISTING;

ODS RTF FILE="%SYSFUNC(PATHNAME(WORK))\FIG3.RTF";

FILENAME FIGURE "%SYSFUNC(PATHNAME(WORK))\SASGRAPH.EMF";

GOPTIONS RESET=ALL TARGET=SASEMF DEVICE=SASEMF FTEXT=DUPLEX HTEXT=.75
        CBACK=WHITE XMAX=6IN XPIXELS=1200 YMAX=5IN YPIXELS=1000
GSFNAME=FIGURE GSFMODE=REPLACE;

PROC UNIVARIATE DATA = PREDTRT1POS;
  VAR PREDTRT1;
  HISTOGRAM;
  PROBPLOT;
  TITLE 'SUMMARY STATS ON PREDICTED VALUES';
RUN;

GOPTIONS RESET=ALL;

ODS RTF CLOSE;

PROC MEANS DATA = PREDTRT1POS;
  CLASS TRT;
  VAR PREDTRT1;
  TITLE 'Positivity Check: Summary Stats on Predicted Values by Observed
Treatment';
RUN;

PROC SORT DATA = PREDTRT1POS;
  BY PREDTRT1;
RUN;

PROC PRINT DATA = PREDTRT1POS;
  TITLE 'Positivity Check: Sorted Listing of Predicted Values';
RUN;

   * (PROC PRINT output not shown but is used to identify individual
patients with combinations of covariates with extreme values *;
```

Output from Program 9.5

```
        Positivity Check: Summary Stats on Predicted Values by Observed Treatment

                            The MEANS Procedure

                Analysis Variable : PREDTRT1 Predicted: Probability

  TRT    N Obs      N          Mean         Std Dev          Minimum          Maximum
─────────────────────────────────────────────────────────────────────────────────────
  A_C     2547     2547     0.0475067     0.1254780                0        0.9758886

  A__     2547     2547     0.4157827     0.4165088     4.0755123E-7        0.9999826

  _B_     2547     2547     0.2508929     0.3999598     1.8471036E-8        0.9999993

  __C     2547     2547     0.2383188     0.3699969     3.4082211E-7        0.9998542

  ___     2547     2547     0.0474988     0.1179316                0        0.9817690
─────────────────────────────────────────────────────────────────────────────────────
```

In addition to the assessment of the no unmeasured confounders and positivity assumptions, the appropriateness of the models was assessed by adding interactions, quadratic terms, and other potential confounder variables to the models. No significant changes to the outcome were noted. The model assessment was done by adjusting the appropriate models in the SAS code—which could be automated by putting the model variables as macro variables at the top of the code. In addition, we analyzed the data using other methods: an ITT LOCF analysis (using all data), a repeated measures model (excluding data after a medication switch) and an epoch analysis (see Chapter 8). Analysis of these simulated data using a simple ITT LOCF analysis failed to show a significant difference between treated and control groups (estimated difference of 1.16, p=.334). The repeated measures mixed model (discarding the data after switching) and epoch analyses results were fairly similar to the MSM approach. The disagreement between the ITT analysis and other analyses appears to be a result of ignoring treatment information for the data after a treatment switch, as discussed by Faries and colleagues (2007). Switchers from treatment C performed well after the switch to treatments A and B, information not attributed to treatments A and B in the ITT analysis.

9.4 Discussion

This chapter has presented the issue of assessing the effects of treatment in longitudinal, observational data—with a focus on addressing treatment switching using MSM. We were interested in the performance of MSM because this approach utilizes all of the study data and produces consistent estimates of the causal effect of treatments, even when there are treatment switching, missing data, and time-varying confounders. Validity of the MSM analysis rests on three key assumptions:

1. no unmeasured confounding
2. positivity
3. correct models

Also, the missing data are assumed to follow a missing completely at random (MCAR) or missing at random (MAR) pattern. Thus, well-planned sensitivity analyses are important. This should include an assessment of the assumptions supporting the causal effect estimation by the MSM as well as use of other methodology supported by differing assumptions. In addition, presentation of

results should make it clear to the reader that causal interpretation of the results rests on unverifiable assumptions as well as being transparent about which steps were taken to assess the robustness of the results.

In summary, marginal structural models are a promising approach for estimating the causal effects of treatment in longitudinal, observational data. Other chapters in this book provide details on alternative methodologies.

References

Bodnar, L. M., M. Davidian, A. M. Siega-Riz, A. A. Tsiatis. 2004. "Marginal structural models for analyzing causal effects of time-dependent treatments: an application in perinatal epidemiology." *American Journal of Epidemiology*. 159: 926–934.

Brumback, B. A., M. A. Hernán, S. J. Haneuse, and J. M. Robins. 2004. "Sensitivity analyses for unmeasured confounding assuming a marginal structural model for repeated measures." *Statistics in Medicine* 23(5): 749–767.

Cole, S. R., M. A. Hernán, J. B. Margolick, M. H. Cohen, and J. M. Robins. 2005. "Marginal structural models for estimating the effect of highly active antiretroviral therapy initiation on CD4 cell count." *American Journal of Epidemiology* 162: 471–478. Epub 2005 Aug 2.

Faries, D., H. Ascher-Svanum, and M. Belger. 2007. "Analysis of treatment effectiveness in longitudinal observational data." *Journal of Biopharmaceutical Statistics* 17(5): 809–826.

Grimes, D. A., and K. F. Schulz. 2002. "Bias and causal associations in observational research." The *Lancet* 359: 248–252.

Haro, J. M., S. Kontodimas, M. A. Negrin, M. Ratcliffe, D. Suarez, and F. Windmeijer. 2006. "Methodological aspects in the assessment of treatment effects in observational health outcomes studies." *Applied Health Economics & Health Policy* 5(1): 11–25.

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: 561–570.

Hernán, M. A., B. 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.

Hernán, M. A., J. M. Robins, and L. A. Garcia Rodrigues. 2005. Comment on: Prentice, R. L., M. Pettinger, and G. L. Anderson. Discussion on "Statistical issues arising in the women's health initiative." *Biometrics* 61: 922–941.

Hernán, M. A., S. Hernández-Díaz, and J. M. Robins. 2004. "A structural approach to selection bias." *Epidemiology* 15(5): 615–625.

Ko, H., J. W. Hogan, and K. H. Mayer. 2003. "Estimating causal treatment effects from longitudinal HIV natural history studies using marginal structural models." *Biometrics* 59: 152–162.

Mallinckrodt, C. H., T. M. Sanger, S. Dubé, D. J. DeBrota, G. Molenberghs, R. J. Carroll, W. Z. Potter, and G. D. Tollefson. 2003. "Assessing and interpreting treatment effects in longitudinal clinical trials with missing data." *Biological Psychiatry* 53(8): 754–760.

Mortimer, K. M., R. Neugebauer, M. van der Laan, and I. B. Tager. 2005. "An application of model-fitting procedures for marginal structural models." *American Journal of Epidemiology* 162(4): 382–388. Epub 2005 Jul 13.

Petersen, M. L., S. G. Deeks, J. N. Martin, M. J. van der Laan. 2007. "History-adjusted marginal structural models for estimating time-varying effect modification. *American Journal of Epidemiology* 166: 985–993.

Robins, J. M., A. Rotnitzky, and D. O. Scharfstein. 1999. *Statistical Models in Epidemiology: The Environment and Clinical Trials.* Edited by M.E. Halloran and D. Berry. IMA Volume 116. New York: Springer-Verlag, 1–92.

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

Rosenbaum, P. R. 2005. "Observational Study." In *Encyclopedia of Statistics in Behavioral Science* 3: 1451–1462. Edited by B. S. Everitt and D. C. Howell. Chichester: John Wiley & Sons, Ltd.

Rosenbaum, P., and D. B. Rubin. 1983. "The central role of the propensity score in observational studies for causal effects." *Biometrika* 70: 41–55.

Suarez, D., J. M. Haro, D. Novick, S. Ochoa. 2008. "Marginal structural models might overcome confounding when analyzing multiple treatment effects in observational studies." Journal of Clinical Epidemiology 61: 525–530.

Tunis, S. L., D. E. Faries, A. W. Nyhuis, B. J. Kinon, H. Ascher-Svanum, and R. Aquila. 2006. "Cost-effectiveness of olanzapine as first-line treatment for schizophrenia: results from a randomized, open-label, 1-year trial." *Value in Health* 9(2): 77–89.

Vansteelandt, S., K. Mertens, C. Suetens, E. Goetghebauer. 2009. "Marginal structural models for partial exposure regimes. *Biostatistics* 10(1): 46–59.

Verbeke, G., and G. Molenberghs. 2000. *Linear Mixed Models for Longitudinal Data*. New York: Springer-Verlag.

Yamaguchi, T., and Y. Ohashi. 2004a. "Adjusting for differential proportions of second-line treatment in cancer clinical trials. Part I: structural nested models and marginal structural models to test and estimate treatment arm effects." *Statistics in Medicine* 23: 1991–2003.

Yamaguchi, T., and Y. Ohashi. 2004b. "Adjusting for differential proportions of second-line treatment in cancer clinical trials. Part II: an application in a clinical trial of unresectable non-small-cell lung cancer." *Statistics in Medicine* 23: 2005–2022.