LONGITUDINAL REGRESSION CONDITIONING ON CONTINUATION

Eric J. B. Daza

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Public Health in the Department of Biostatistics in the Gillings School of Global Public Health.

Chapel Hill
2015

Approved by:

Michael G. Hudgens

Amy H. Herring

Linda S. Adair

Stephen R. Cole

John S. Preisser, Jr.

ABSTRACT

Eric J. B. Daza: Longitudinal Regression Conditioning on Continuation.
(Under the direction of Michael G. Hudgens and Amy H. Herring)

Individuals in a longitudinal study may have missing data for multiple reasons, including intermittent missed visits or permanent study drop out. Additionally, individuals may experience a truncating event, such as death, past which the outcomes of interest are no longer meaningful. Kurland and Heagerty (2005) developed a method to conduct regression conditioning on being alive (RCA), which constructs inverse-probability weights (IPWs) of the dropout probability among continuing individuals used to fit generalized estimating equations (GEE). RCA has since been extended to allow for intermittent missingness (IM) of outcomes (Shardell and Miller, 2008). We further extend these methods to simultaneously accommodate different mechanisms for dropout and IM, and call our method regression conditioning on continuation (RCC). RCC is illustrated using data from a recent study of mother-to-child transmission of HIV to draw inference on mean infant weights subject to truncation from infant death or HIV infection.

Currently, there is no widely available software for conducting RCA. We present the `xtrccipw` command in Stata, which can estimate the dropout IPWs required by RCC if there is no IM. These IPWs estimated using `xtrccipw` are then used as weights in a GEE estimator using the `glm` command, completing the RCC method. In the absence of truncation, the `xtrccipw` and `glm` commands can also be used in a weighted GEE analysis. The `xtrccipw` command is demonstrated by analyzing two example datasets and the original Kurland and Heagerty (2005) data.

A fundamental weakness of most non-sampling IPW methods is that the missing-data model is unknown and yet must be correctly specified to obtain consistent mean-outcome estimates. We extend the RCC approach to use augmented estimating equations (AEE) in what we call the augmented RCC (ARCC) method. In addition to the missing-data model specified by IPW-GEE methods, AEE approaches specify a model for the outcome joint probability. However, only one

of these two models need be correct for the corresponding mean-outcome estimator to be consistent, making such techniques doubly robust to model misspecification. The empirical bias of the ARCC and RCC estimators are characterized and compared in a simulation study, and the ARCC method is applied to the mother-to-child HIV transmission study.

To Mommy (Celia Alvarez).

ACKNOWLEDGEMENTS

My greatest thanks go to my Committee, comprised of Linda S. Adair, Stephen R. Cole, Amy H. Herring, Michael G. Hudgens, and John S. Preisser, Jr., for their ongoing motivation, patience, and critical guidance while we built up my biostatistical brain here at UNC. Their tutelage and camaraderie were instrumental in helping me achieve this first big step into a productive and rewarding career. My thanks also go to Ashley Buchanan, Valerie Flax, Janaki Parthasarathy, Amanda J. Selin, and Patrick J. Smith for their helpful comments in editing the manuscripts. I also charge E. Michael Foster and Mark J. van der Laan with early inspiration, mentorship, and faith in my ability to realize the potential outcome of becoming a causal-inference-oriented biostatistician. (Mike, you were truncated too soon. I hope to make you proud as a small part of your continuing counterfactual.) My thanks go to Chirayath "Suchi" Suchindran for his warm guidance and targeted counsel on how to get into UNC Biostatistics, and I also thank Gil Fine, Wayne Davis, and Henrik "Hank" Kulmala at Supergen, Inc., for their career guidance and recommendation letters. From my days at Loyola High School, I thank Thomas Goepel for teaching me how to argue logically and concisely through writing. I also thank the many NC Triangle Area cafes and coffeeshops that caffeinated, housed, and powered me through my many hours of doctoral work and play, most notably Bean Traders, Guglhupf, Looking Glass, Open Eye, Straw Valley, and Cocoa Cinnamon. Finally, I send all my love and gratitude out to my amazing family and friends for supporting me through the years while completing my doctoral program, including Celia Alvarez, Elvira V. Daza, Nilo C. Daza, Patricia A. Daza-Luu, Vincent S. Y. Lee, Billy Luu, Amanda J. Selin, and Patrick J. Smith. Maraming salamat sa inyong lahat, mga kapamilya at kaibigan ko.

This work was supported in part by grants SIP 13-01 U48-CCU409660-09, SIP 26-04 U48-DP000059-01, and SIP 22-09 U48-DP001944-01 from the Prevention Research Centers Special Interest Project of the US Centers for Disease Control and Prevention (CDC), grant P30-AI50410 from the University of North Carolina Center for AIDS Research (UNC CFAR), grants DHHS/

NIH/FIC 2-D43 Tw01039-06 and R24 Tw00798 of the American Recovery and Reinvestment Act from the National Institutes of Health Fogarty AIDS International Training and Research Program (NIH AITRP), grant OPP53107 from the Bill and Melinda Gates Foundation (BMGF), grant R24 HD050924 from the Carolina Population Center (CPC), grant R01-AI029168, the US National Institute of Allergy and Infectious Diseases (NIAID) grant R01-AI085073, and the National Institute of Environmental Health Sciences (NIEHS) grant R01ES020619. The content is solely the responsibility of the authors, and does not necessarily represent the official views of CDC, NIAID, UNC, NIH, BMGF, or CPC. I would also like to thank the BAN investigators for access to the data from their study; Brenda Kurland for access to the data from the PEP study, and for her help and guidance in working with the datasets; and the BAN and PEP study participants.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

Longitudinal studies in developing regions that investigate nutritional outcomes such as faltering or micronutrient deficiency sometimes suffer from high rates of attrition. Subject data becomes missing either intermittently, or from dropout or loss to follow-up (i.e., when study participants drop out of the study before its conclusion, and do not return). Dropout is sometimes referred to as attrition, or monotonic missingness, and intermittent missingness (IM) is sometimes called arbitrary or non-monotonic missingness, respectively. While standard methods like multiple imputation (MI) and inverse-probability weights (IPWs) can respectively adjust for these particular types of data missingness, investigators may rely on incorrect default methods for handling missingness.

Unless specified otherwise, standard generalized estimating equations (GEE) procedures will assume that any missing outcomes are missing completely at random (MCAR), which will produce biased estimates if the observability of an outcome at any given study time point depends on any of the outcomes themselves. If a more severely HIV-compromised study participant is less likely to make it into the clinic because of her weakened immune system, then her chances of missing a clinical visit for a viral load assay depend on her viral load; viral load measurements cannot therefore be justified as being MCAR. However, if we can at least assume that her chances of coming in for a viral load measurement are completely predictable using only her previously observed viral loads, we can use this information to impute participant outcomes past their point of dropout.

In the latter situation, the outcomes are considered to be merely missing at random (MAR) once previous outcomes have been taken into account. Standard maximum-likelihood (ML) based mixed-effects models, also known as random-effects models or simply just mixed models, assume MAR by design. Similarly, standard GEE approaches can be adjusted to account for previous outcomes by incorporating IPWs that are constructed based on the probability of being observed at a given time point given observed outcomes. The IPWs are used to balance the contribution

of each complete-case observation, and these methods are frequently called "weighted estimating equations" (IPWGEE).

Unfortunately, these MAR-based imputations are sensible only if we can reasonably expect a subject's unobserved post-dropout outcomes to exist. This expectation no longer makes sense, for instance, for those study subjects who die before the end of the study—an example of a notably different type of data missingness that is defined by a precluding or "truncating" event, after which a subject cannot have outcomes that are meaningfully defined. (Under attrition and dropout, we can at least assume that unobserved outcomes exist.) In a study of average trends in weight, we do not observe the weights for a living participant who prematurely drops out, assuming they remain alive for the duration of the study. A person who dies before the end of the study, however, cannot subsequently have any sensible weight measurements. The problem with many longitudinal analysis methods that adjust for MAR is that they tacitly envisage weights past the point of death because they consider death to be a missingness event similar to dropout.

Researchers have since developed methods that can correctly differentiate between dropout and "truncation by death". Kurland and Heagerty (2005) devised one such IPWGEE technique to estimate the dropout probability at each time point, but only among the surviving subjects at each of those time points. They tested their method on a binary longitudinal outcome in a prospective cohort study of elderly disability, compared trends across three groups defined by disability risk, and demonstrated small empirical bias in mean estimates relative to those of other standard methods. However, their approach cannot handle IM. Shardell and Miller (2008) developed a solution that extends the Kurland and Heagerty (2005) method to allow for IM. Neither approach, however, can handle outcomes subject to both dropout and IM simultaneously.

We propose to 1.) extend this IPWGEE method to account for truncation while also distinguishing between dropout and IM, and to then apply this new approach to draw inference about a continuous longitudinal nutritional outcome in a mother-to-child HIV transmission study, 2.) create a general programming implementation of the original method in Stata, and 3.) augment our extended method to produce a doubly robust estimator.

CHAPTER 2: LITERATURE REVIEW

2.1 Motivation from Nutritional Studies

2.1.1 On (Not) Addressing Missing Data

A number of nutrition studies have adjusted for mortality through simple exclusion of subjects with missing outcomes, or by assuming data were MCAR. A follow-up study of the effects of dietary supplements randomized to then-pregnant women in rural Gambia employed GEE/GLS methods that assumed the data were MCAR (Hawkesworth et al., 2008). The authors defended this assumption by asserting that those recruited and those lost to follow-up were similar with respect to the empirical distributions of other covariates, so that both observed and unobserved outcomes could likewise have been similarly distributed. In a randomized trial of micronutrient supplements in rural Nepal, Christian et al. (2003) assumed data were MCAR by running a standard, unadjusted GEE analysis. A randomized study conducted in Zambia investigating the effect of breastfeeding cessation on the growth faltering of uninfected infants born to HIV-infected mothers also assumed their data were MCAR (Arpadi et al., 2009). Culnane et al. (1999) examined the long-term effects of ART randomized to uninfected infants of HIV-infected mothers, but only the survivors were analyzed.

Reporting descriptive statistics of observable characteristics between those included and excluded from an analysis is a good first step at dealing with missing data, also known as missingness. However, these findings may not justify an MCAR analysis if the observed data are in fact associated with missingness, and such an analysis does not use of all of the data that are actually available. Subjects that are similar with respect to a particular set of secondary characteristics may nonetheless differ systematically in primary characteristics (i.e., outcomes or covariates) depending on whether or not those primary characteristics are observed. This is a strong possibility in longitudinal analyses, wherein the primary characteristics of the subjects at risk for becoming missing may change over time.

2.1.2 The BAN Study

The Breastfeeding, Antiretrovirals, and Nutrition (BAN) Study (van der Horst et al., 2009; Chasela et al., 2010) was a clinical trial that randomized 2369 pairs of mothers infected with human immunodeficiency virus (HIV) and their uninfected infants. Mother-infant pairs in the trial were scheduled to be followed for 48 weeks, with women receiving counseling for exclusive breastfeeding during the first 24 weeks, hereafter the "study period." During the study period, the trial's objectives were to assess the benefit of nutritional supplementation taken by women, the safety and efficacy of antiretroviral treatment (ART) given either to mothers or their infants to prevent HIV transmission, and the feasibility of exclusive breastfeeding followed by early, rapid breastfeeding cessation. BAN employed a 3×2 factorial design involving a 2-arm maternal nutritional intervention component crossed with a 3-arm ART component. Maize flour was given to all enrolled women for family consumption, and half of these mothers were randomized to receive a lipid-based nutrient supplement (LNS). Initially, ARTs were assigned either to mothers, infants, or neither; this last control group was eliminated by the data and safety monitoring board (DSMB) in March 2008 after 78% of participants had already received treatment.

For the first objective, BAN study investigators wanted to draw inference about the mean infant anthropometric outcomes weight, length, and body mass index (BMI) over time among infants who were alive and HIV-uninfected. Death and infection could therefore be said to "truncate" the trajectory of outcomes, leaving these outcomes undefined after truncation. Among infants who were alive and HIV-uninfected, some outcomes may have been unobserved because mothers intermittently missed visits, or because they dropped out of the study altogether. An infant was considered eligible for an intent-to-treat analysis if during the study period s/he had available outcome measurements for more than one visit, and was a singleton at birth. Out of 2238 eligible infants, by the end of the study period, 307 had dropped out (14% of all infants), and 187 infants who had not dropped out had either become infected with HIV or had died (8% of all infants; 162 alive and infected, 25 dead). While they were alive, uninfected, and still in the study, 645 infants missed 973 scheduled visits intermittently (5% of all scheduled visits before truncation or dropout). Scheduled visits were irregularly spaced, and mother-infant pairs adhered closely to the visit schedule.

The original analyses were carried out using random-intercept mixed-effects models that conceptualized truncated outcomes for the aforementioned 187 infants as well-defined but unobserved MAR outcomes, thereby conflating truncation with missingness. Using such an analysis, Flax et al. (2012) concluded that intervention was not significantly associated with infant weight. Classifying the truncation and missingness events together may have masked a true association among infants who survived uninfected at any given visit. In addition, infants may have dropped out or missed visits intermittently (before dropout) for entirely different reasons. Hence, the Flax et al. (2012) analysis must be corrected to separately account for death or infection, dropout, and IM.

2.2 Methods for Handling Missing Data

Rubin (1976), followed by Little and Rubin (2002), laid out the comprehensive framework of missingness mechanisms used to characterize how outcomes that always exist are related to their observation status. When the reasons for observing an outcome are statistically independent of any outcomes (observed or not), the outcomes are said to be missing completely at random. If outcome missingness is no longer statistically associated with unobserved outcomes after conditioning on observed outcomes, then outcomes are described as missing at random. If missingness is associated with unobserved (and possibly also observed) outcomes, the outcomes are said to be not missing at random, or missing not at random (MNAR). An equivalent way of defining MNAR is that outcomes are MNAR if they are neither MCAR nor MAR. These three cases are exhaustive.

Maximum likelihood estimation (MLE) can be used to estimate model parameters of longitudinal outcomes when missingness is present. These methods often factorize the outcome and missingness joint probability in one of three ways (Little, 1993, 1995; Diggle et al., 2002). For example, Little and Rubin's three missingness mechanisms can be defined using a selection factorization, with the estimand of inferential interest being the marginal, or unconditional, mean. Selection models define a conditional distribution that allows the "selection" of an outcome (i.e., to be measured or observed) to depend on some subset of the outcomes; this conditional distribution is paired with a marginal distribution for all outcomes (i.e., the complete outcome data).

They are particularly useful under MAR, where the missingness mechanism is often said to be ignorable or non-informative. In this case, inference carried out using only observed outcomes will apply to all outcomes. MLE is efficient if, in addition, the missingness-mechanism parameters are distinct from those of the outcomes (Rubin, 1976; Little, 1995; Diggle et al., 2007). Conversely, an MNAR missingness mechanism is called non-ignorable or informative.

The two other factorizations are the pattern-mixture and shared-parameter factorizations. Pattern-mixture models allow the outcomes to depend on missingness status. These models are particularly suited to outcomes that can be stratified (or otherwise categorized) by subgroups defined by their missingness "patterns". (For example, the outcome distributions of participants who become missing at time points 2 and 5 may differ.) This conditional distribution is paired with a marginal distribution for the missingness mechanism. Shared-parameter models, by contrast, attribute any association between the outcomes and missingness to underlying subject-specific latent variables; missingness and outcomes "share" a participant's random-effect coefficients (Ten Have et al., 1998). They are most useful when the estimands of interest are the subject-specific mean outcomes. These are also known as random-effect models (Wu and Carroll, 1988; Hogan and Laird, 1997; Diggle et al., 2002), random-coefficient selection models (Little, 1995), random-effect selection models (Ribaudo et al., 2000), and joint models for longitudinal measurements and event-time data (Henderson et al., 2000; Guo and Carlin, 2004; Hu et al., 2015).

The method of generalized estimating equations (GEE) is a commonly used non-MLE approach when one wishes to draw inference about the marginal means of an outcome for a large sample. The three joint-probability factorizations may also be applied to GEE. While standard, unweighted GEE is consistent only under MCAR (Liang and Zeger, 1986; Diggle et al., 2002), adjustment using IPWs extends their usefulness to MAR and MNAR settings, where each IPW is the reciprocal of a probability of not being missing (Robins et al., 1995; Scharfstein et al., 1999a). In this IPWGEE approach, weighting balances (i.e., adjusts) the sample contribution of observed

outcomes across all observed values of variables associated with missingness; the goal is to consistently estimate the population parameter. This technique has notable origins in the survey-sampling literature (Horvitz and Thompson, 1952), wherein each IPW corresponds to a sample-selection probability that is known by design.[1]

Under MAR, the MLE and IPWGEE approaches possess somewhat complementary advantages and disadvantages. While both approaches rely on correct specification of the mean outcome model, MLE additionally relies on correct modeling of the joint distribution of outcomes; however, MLE does not require knowledge of the missingness model (Kurland et al., 2009; Laird, 1988). Conversely, IPWGEE does not specify the joint-outcome distribution, but does rely on correct modeling of the missingness model. MLE can be considerably more efficient than IPWGEE if the joint-outcome model is correct, and is well-suited to handle outcomes unbalanced over time (i.e., occurring at arbitrary time points, or with varying frequency). What IPWGEE lacks in efficiency and temporal flexibility, it makes up for through inference about mean outcomes that is robust to misspecification of the covariance structure (which likelihood-based inference requires to be correct), provided the missingness model is correct. Furthermore, IPWGEE is more flexible than MLE by allowing missingness patterns to be modeled on other auxiliary covariates; MLE assumes outcomes are MAR conditional only on the covariates specified in the mean outcome model. Doubly robust estimators leverage both approaches by modeling both the joint outcomes and missingness patterns. These estimators are "doubly robust" because they are consistent if just one of either the missingness or joint-outcome model is correct. MNAR outcomes are more complicated to analyze using either approach because they require some knowledge of the unobserved outcomes. The usual way of handling such outcomes is to impose a set of

[1]Caution should be exercised when intuiting the relationship between IPWs and the sample size in an IPWGEE analysis. In survey sampling, the Horvitz-Thompson (HT) estimator of the population total uses IPWs defined as the inverse of the sample-inclusion probability of each individual in the sample (Lohr, 2010). Because the sum of the IPWs is greater than or equal to the sample size, these IPWs might be understood to inflate the sample size. Hence, the IPWGEE approach is frequently compared to the related method of imputation, which literally increases the sample size by adding estimated outcomes to the sample where none were originally observed. However, it is important to remember that many IPW estimators in survey sampling do not actually "increase" the sample size in this way. In fact, the commonly used HT ratio estimator of the population mean uses sampling-probability IPWs that are normalized or standardized over the sum of the IPWs, such that the normalized weights sum to one (Kish, 1965; Basu, 1971; Hájek, 1971). This HT ratio estimator is identical to the IPWGEE estimator for a continuous mean outcome at one time point with independent errors and an intercept covariate; i.e., no sample inflation has occurred in balancing the observed-outcome contributions.

plausible assumptions or distributions regarding the unobserved outcomes. A sensitivity analysis is then conducted to characterize any effects on inference that result from varying these assumptions.

Until now, we have assumed that all unobserved outcomes are well-defined. However, in a number of missing-data settings, conceptualizing outcomes unobserved due to death similarly to those unobserved due to missingness may not make sense substantively. Many authors have acknowledged that in such cases, "missing" outcomes may be ill-defined after death, even though death and true missingness both lead to unobserved outcomes that are easily (and therefore, often) handled within the Little and Rubin framework. To address this, Ribaudo et al. (2000) jointly modeled outcomes and censored survival times by applying methods originally developed by Schluchter (1992) to handle informatively censored survival data in longitudinal studies. In a review of such joint-modeling approaches, Billingham and Abrams (2002) expounded upon the relationship between missingness and a quality-adjusted life-years outcome that accounts for death. Pauler et al. (2003) used mixed-effects models to handle MNAR outcomes, grouping outcomes by different patterns of times-to-event for dropout and death. However, many of these approaches may still fundamentally treat death as a missingness event. For example, Pauler et al. (2003) point out that marginalization over all outcome patterns at all times is still possible using their approach, thereby implying the existence of outcomes after death; these authors espouse reporting of outcomes that resemble those of Billingham and Abrams (2002).

More recent methods have avoided this tacit conflation of death with missingness. Dufouil et al. (2004) explored the use of IPWs that distinguish death and dropout in a study of cognitive decline in the elderly. Borrowing the survey-sampling analogy that weighting increases the effective sample size, they argued that it did not make sense to implicitly impute outcomes for elderly participants who had died because this immortal sample population would not represent the target population with its higher mortality rate. Following this, Kurland and Heagerty (2005) developed an IPWGEE method they called regression conditioning on being alive (RCA) for an analysis involving an elderly population that experienced both dropout and death. RCA was applied to estimate disability status over time, but—crucially—only among living participants at each time point. Shardell and Miller (2008) subsequently extended RCA to handle missing covariates,

and to allow either outcomes or covariates to be unobserved intermittently. le Cessie et al. (2009) adapted RCA for use with a multi-state model of disease, whereby both dropout and a continuous quality-of-life outcome could vary depending on disease stage; i.e., alive and disease-free, or alive but relapsed. Basu and Manning (2010) extended the Lin et al. (1997) survival-adjusted estimator through the use of RCA to separate out covariate effects on continuous dropout and survival times, and on their continuous outcome.

2.3 Regression Conditioning on Continuation

2.3.1 Overview

Kurland and Heagerty (2005) defined two problematic types of mean-outcome models: Unconditional models assume that mean outcomes at each time point exist for both living and deceased subjects, and fully conditional models, while accounting for survival, require that the survival time of each subject be known. Clinical use of the latter could be problematic because it requires the clinician to specify the time of death, which would not be known in advance, in order to provide counsel about the expected outcome at a given time point. As a remedy, Kurland and Heagerty (2005) proposed RCA, an IPWGEE mean model that only partly conditions on survival. Inferences about outcomes from the partly conditional model of RCA can answer the more general question of what outcome to expect among the living at a given time point; i.e., one need not know or posit a specific time of death. However, it should be emphasized that RCA must first be conducted in order to draw inference on mean outcomes for subsequent use in clinical practice. To conduct RCA, the time of death for each participant in the original (i.e., pre-practice) study must be known. Hence, the real utility of RCA is that the resulting inferences can more easily be applied towards future out-of-sample predictions for the same target population, as would be the case in clinical practice.

Unlike standard MLE methods, RCA can produce consistent mean-outcome estimates for MAR outcomes subject to death. This is because the RCA missingness model conditions on being alive, while the MLE joint-outcome likelihood model is marginalized over survival. Suppose death happens completely at random such that death itself is not associated with any of the outcomes, and that outcomes are either MCAR or MAR. If the likelihood model for observed

outcomes is correct, then MLE produces consistent mean-outcome estimates because this likelihood is independent of survival. Now suppose that death occurs at random such that death is associated with observed past outcomes. Because the joint distribution of observed outcomes is associated with survival, conducting MLE on this marginal likelihood no longer estimates the same parameters that are estimated by RCA in general, even if missingness is ignorable. One immediate solution is to specify the likelihood model to condition on survival, as is done in joint-modeling approaches. In a recent example of this approach, Hu et al. (2015), developed a general competing-risks, multiple-imputation framework that handles truncation and intermittent missingness requiring specification of models for both outcome and survival.

The key technical contribution of RCA is its definition of the IPW as the reciprocal of the probability of non-dropout, but only for individuals with continuing trajectories at a given time point. Like Kurland and Heagerty (2005), we are interested in drawing inference on longitudinal mean outcomes for subjects who are alive. However, subjects in the BAN study must be both alive and HIV-uninfected; i.e., BAN infant outcomes continue to be well-defined over time unless they are truncated due to either infant death or HIV infection. To accommodate this more general concept of outcome-trajectory continuation and truncation, we call our RCA extension regression conditioning on continuation (RCC). Like the BAN infant outcomes, some of the disability outcomes of Kurland and Heagerty (2005) were intermittently missing, which they addressed through imputation before the point of dropout (i.e., before the last IM time point). A more principled approach was taken by Shardell and Miller (2008), who built an RCA method that not only allows missingness to be intermittent, but that can also handle both outcome and covariate missingness.

Our main contribution in developing RCC will be the ability to distinctly adjust for outcome dropout and IM. Most IPWGEE methods do not consider dropout to reflect an underlying process, such that arbitrarily missing outcomes before dropout are treated the same as monotonically missing outcomes that necessarily occur after dropout. Notable exceptions are found in the work of Yang and Shoptaw (2005) and Yang et al. (2008), who recognized the substantive importance of distinguishing between the dropout and IM processes. To do so, these authors developed the likelihood-based method of multiple partial imputation, which first imputes intermittently

missing outcomes in order to subsequently address dropout. Following the developments in Kurland and Heagerty (2005) and Shardell and Miller (2008), and applying the concepts of Yang, Li, and Shoptaw, we will define IM as being conditional on non-dropout.

2.3.2 Notation & Assumptions

Suppose we have a random sample of $i = 1, \ldots, n$ participants or subjects. Each participant can be measured at up to $j = 1, \ldots, m$ scheduled study time points. Dependence on i is suppressed for notational ease when it is not ambiguous. Let Y_j denote the outcome at time point j. Let $C_j = 1$ if the truncating event has not occurred by time point j, and let $C_j = 0$ otherwise. We assume the outcome Y_j is well defined if and only if $C_j = 1$, and that truncation is a permanent state transition such that $C_j = 0$ implies $C_k = 0$ for all $k > j$. The opposite of truncation is referred to as continuation. If $C_j = 1$, let $R_j = 1$ if the outcome is observed at time point j; otherwise, let $R_j = 0$. If $C_j = 1$, let $R_j^D = 1$ if a participant has not dropped out by time point j; otherwise, let $R_j^D = 0$. If all continuing outcomes are missing at and beyond time point j, then we define dropout to have occurred by time point j; i.e., dropout is monotonic such that $R_k = 0$ for all $k \geq j$ if $C_k = 1$. Let $*$ denote all undefined values. We adopt the convention that $Y_j = *$, $R_j = *$, and $R_j^D = *$ if $C_j = 0$.

For clarity, the following shorthand notation is used when applicable. For a quantity A that can be either a random variable or a constant, let A_j denote the value of A at time point j, and let $\bar{A}_j = \big(A_1, \ldots, A_j\big)$ so that $\bar{A}_{j-1}$ represents an individual's history of A prior to time point j, where $\bar{A}_{j-1} = \emptyset$ at $j = 1$. For a random variable B, if B is discrete then let $p(b)$ denote $\Pr(B = b)$, the probability mass of B at b. Likewise, if B is continuous then let $p(b)$ denote $f(b)$, the probability density of B at b. Let $p(\cdot|b)$ denote $p(\cdot|B = b)$.

Kurland and Heagerty (2005) describe regressions that incorrectly assume that truncation and dropout both result in extant but unobserved outcomes as "unconditional regression models ... that do not account for survival status" (Kurland and Heagerty, 2005). We henceforth refer to any method that does not account for truncation by conditioning on continuation as an unconditional regression (UR) method or approach.

2.3.3 Estimand of Interest

In conducting regression, the mean outcome at time point j is often denoted as a function of a $p \times 1$ covariate vector $\boldsymbol{x}_j$. This can be written as $g\{E(Y_j|\boldsymbol{x}_j)\} = \eta(\boldsymbol{x}_j)$, where $\eta(\boldsymbol{x}_j)$ is commonly specified as $\boldsymbol{x}_j'\boldsymbol{\beta}$ with parameters $\boldsymbol{\beta}$ that are linear in $\boldsymbol{x}_j$, and $g(a)$ is a link function for a quantity a. Hereafter, we consider $\boldsymbol{x}_j$ to be fixed, and hence suppress the notation for dependence on $\boldsymbol{x}_j$ in all expressions. Let $\mu_j^{RCC} = E(Y_j|C_j = 1)$ denote the mean outcome for individual i whose trajectory is still continuing at time point j. The quantity μ_j^{RCC} is our partly conditional estimand of interest (Kurland and Heagerty, 2005; Kurland et al., 2009).

The UR definition of truncation as missingness can result in estimation of a quantity not equal to μ_j^{RCC}. The expected outcome $E(Y_j)$ for subject i at time point j can be expanded as

$$\mu_j = \mu_j^{RCC} \Pr(C_j = 1) + E(Y_j|C_j = 0) \Pr(C_j = 0). \tag{2.1}$$

If there is no truncation such that $\Pr(C_j = 0) = 0$, then (2.1) reduces to $\mu_j = \mu_j^{RCC}$. Definitional problems arise when $\Pr(C_j = 0) > 0$. Specifically, $E(Y_j|C_j = 0)$ is ill-defined if $\Pr(C_j = 0) > 0$ because $Y_j = *$ if $C_j = 0$. Hence,

$$\mu_j = \mu_j^{RCC} \Pr(C_j = 1) + E(* |C_j = 0) \Pr(C_j = 0)$$

is also ill-defined. The UR estimand of interest is μ_j itself, and UR considers truncation to be a missingness event that can be handled via IPWs or imputation. Specifically, UR defines $C_j = R_j^D$ so that $\mu_j^{RCC} = E(Y_j|R_j^D = 1)$ and

$$\mu_j = \mu_j^{RCC} \Pr(R_j^D = 1) + E(Y_j|R_j^D = 0) \Pr(R_j^D = 0).$$

Recalling our discussion in Section 2.3 of MLE conducted using a marginal likelihood, it can now be seen that such an approach is a UR method with survival defined as non-dropout, where the estimand of interest is $\bar{\mu}_m$.

2.3.4 Estimation and Inference

In this section, we give a brief overview of the RCC theory covered extensively in Chapter 2.5. The generalized estimating equations of RCC are introduced first to precisely define μ_{ij}^{RCC}. The dropout and IM mechanisms are then defined more precisely, and we show that μ_{ij}^{RCC} can be consistently estimated under MAR if these mechanisms are correctly specified. For continuous and unbounded outcomes, the identity link $g(a) = a$ is commonly specified. Because the estimand of interest in the BAN study was a continuous outcome, the theory was developed using an identity link.

Details on the general class of GEE methods can be found elsewhere (Liang and Zeger, 1986; Diggle et al., 2002, for example). Let $\mu_{ij}^{RCC} = \boldsymbol{x}_{ij}'\boldsymbol{\beta}^{RCC}$ denote the model for the mean outcome as a linear function of covariates $\boldsymbol{x}_{ij}$ with corresponding parameters $\boldsymbol{\beta}^{RCC}$. The GEE expression relevant to RCC is the vector estimating equation

$$U\left(\boldsymbol{\beta}^{RCC}\right) = \sum_{i=1}^{n}\sum_{j=2}^{m} \boldsymbol{x}_{ij} C_{ij} \frac{R_{ij}R_{ij}^{D}}{\pi_{ij}}\left(Y_{ij} - \mu_{ij}^{RCC}\right), \tag{2.2}$$

where

$$\pi_{ij} = \Pr\left(R_{ij} = 1, R_{ij}^{D} = 1 \middle| \bar{r}_{i(j-1)}, \bar{r}_{i(j-1)}^{D}, \bar{y}_{im}, \bar{c}_{im}\right)$$

is the joint probability of not being missing and not having dropped out, conditional on the history of missingness and dropout, on all outcomes, and on the full truncation vector. We sometimes refer to π_{ij} as the missingness model. Recalling that $Y_{ij} = *$ if $C_{ij} = 0$, we adopt the convention that the summand in (2.2) for individual i at time point j equals 0 rather than being undefined if $C_{ij} = 0$. It can be shown that the estimating equations (2.2) are unbiased for zero such that $E\{U(\boldsymbol{\beta}^{RCC})\} = \mathbf{0}$. Hence, the solution to $U(\boldsymbol{\beta}^{RCC}) = \mathbf{0}$ is an unbiased estimate of $\boldsymbol{\beta}^{RCC}$. Unfortunately, this estimation procedure is generally intractable because π_{ij} cannot be calculated if $\bar{y}_{im}$ is not fully observed (i.e., if $R_{ij} = 0$ is true at least once in the sample). This would not be a problem, however, if in reality π_{ij} only depends on observed outcomes.

One way of formulating such a π_{ij} presents itself through the expansion

$$\pi_{ij} = r_{i(j-1)}^{D} \Pr\left(R_{ij} = 1 \middle| R_{ij}^{D} = 1, \bar{r}_{i(j-1)}, \bar{y}_{im}, \bar{c}_{im}\right) \Pr\left(R_{ij}^{D} = 1 \middle| \bar{r}_{i(j-1)}, R_{i(j-1)}^{D} = 1, \bar{y}_{im}, \bar{c}_{im}\right).$$

The IM mechanism is defined as $1 - \Pr\left(R_{ij} = 1 \middle| R_{ij}^D = 1, \bar{r}_{i(j-1)}, \bar{y}_m, \bar{c}_m\right)$, and the dropout mechanism is defined as $1 - \Pr\left(R_{ij}^D = 1 \middle| \bar{r}_{i(j-1)}, R_{i(j-1)}^D = 1, \bar{y}_{im}, \bar{c}_{im}\right)$. As stated earlier, IM is defined as missingness conditional on non-dropout, which differs from the usual definition of IM as arbitrary or non-monotonic missingness at any time point. Under certain assumptions described in Section 3.2.2, these probabilities are seen to depend only on observed outcomes, and to thereby imply that π_{ij} likewise depends only on observed outcomes. These assumptions are also shown to imply that outcomes are MAR, which is broadly defined as

$$p\left(\bar{r}_{im} \middle| \bar{y}_{im}, \bar{c}_{im}\right) = p\left(\bar{r}_{im} \middle| \bar{y}_{im}^{\text{obs}}, \bar{c}_{im}\right), \tag{2.3}$$

where $\bar{y}_{ij}^{\text{obs}} = \left\{y_{ik} : R_{ik} = 1, k \leq j\right\}$ denotes the observed values of $\bar{y}_{ij}$.

Nonetheless, the functional form of π_{ij} must still be known in order to perform estimation and inference using (2.2). That is, knowledge of the dropout and IM mechanisms coupled with a true MAR assumption is not sufficient for estimation and inference. Non-parametric approaches can be used to place minimal constraints or assumptions on the mechanism distributions. However, to keep our initial development of RCC focused, we modeled the distributions instead. This parametric approach allowed for straightforward, consistent estimation of π_{ij}. The RCC estimator is the solution to $U\left(\boldsymbol{\beta}^{RCC}\right) = \mathbf{0}$ using consistent estimates of π_{ij}, and is both consistent for $\boldsymbol{\beta}^{RCC}$ and asymptotically multivariate normal (Robins et al., 1995). This estimator can therefore be used to conduct inference on $\boldsymbol{\beta}^{RCC}$ using the empirical sandwich estimator of estimator variance available in standard software, which treats each estimated π_{ij} as fixed, and is generally conservative in constructing 95% Wald confidence intervals (Robins et al., 2000; Robins, 2000; Preisser et al., 2002).

2.3.5 Methods Comparison

In Section 3.3, we conduct a simulation study to illustrate the performance of RCC alongside the following three methods, which are not expected to be consistent for $\boldsymbol{\beta}^{RCC}$. To reflect common analyses that assume outcomes are MCAR, GEE estimation is conducted using a lag-1 autoregressive (AR-1) working correlation structure and no IPWs. A UR approach is also implemented, via IPWGEE with an independence working correlation, that nonetheless accounts for

IM. To illustrate the effect of incorrectly treating all arbitrary missingness as IM even while correctly conditioning on continuation (i.e., incorrect application of the Shardell and Miller (2008) method), RCC is conducted with IPWs that only specify an IM model; a dropout model is not specified.

Simulations are used to compare the empirical relative bias and coverage probability of these four estimators under 12 data-generation scenarios defined by varying the truncation, dropout, and IM mechanisms, where outcome correlation followed an AR-1 structure. Kurland and Heagerty (2005) assessed RCA performance using simulated data generated with only one truncation and dropout mechanism. Specifically, they generated monotonically MCAR outcomes subject to truncation completely at random; i.e., truncation that was not associated with any of the outcomes. In our simulations, outcomes were generated subject to a combination of one of two truncation mechanisms, one of two dropout mechanisms, and one of three IM mechanisms. These mechanisms are defined using "completely at random" and "at random" concepts.

2.4 Augmented Regression Conditioning on Continuation

In Section 2.3.4, it was shown that the RCC estimator is consistent for μ_j^{RCC}, but only if the missingness model π_j was correctly specified. In practice, π_j is usually not known, and hence can be wrongly specified in general. In a comprehensive simulation study, Preisser et al. (2002) reported circumstances under which IPWGEE with a misspecified dropout model actually performs worse than unweighted GEE. To mitigate the bias incurred by such misspecification, a clever, technical GEE method has since been developed that involves specifying an additional modeled component that "augments" the original IPWGEE.

In a series of seminal papers, Robins, Rotnitzky, and Zhao (RRZ) described a class of IPW-based semi-parametric estimators that includes the RCC estimator (Robins et al., 1994, 1995; Robins and Rotnitzky, 1995). They described methods for augmenting each IPWGEE summand with a conditional expectation term. The resulting augmented IPW (AIPW) estimator has the desirable property of being doubly robust in that misspecification of either the missingness model

or the conditional expectation model (but not both simultaneously) still yields a consistent estimator of the parameters, if the mean-outcome model is correct. In these articles, the RRZ methods were applied to MAR data with both monotonically and arbitrarily missing (i.e., IM) outcomes, allowing for intermittently missing covariates. They were later extended to account for MNAR data (Vansteelandt et al., 2007) and intermittently missing longitudinal MAR outcomes and covariates (Chen and Zhou, 2011). The missingness and causal-inference literatures have both been enriched by the development and use of these augmented estimating equations (AEE) techniques (Robins et al., 1995; Rotnitzky et al., 1998; Scharfstein et al., 1999a; van der Laan and Robins, 2003; Lunceford and Davidian, 2004; Bang and Robins, 2005; Kang and Schafer, 2007; Wooldridge, 2007).

Recent longitudinal AEE methods have been developed that differentiate between truncation, dropout, and IM. The AEE method developed by Chen and Zhou (2011) does not address truncation, but does handle both outcome and covariate IM. Shardell et al. (2015) extended an IPWGEE principal-stratification technique developed by Tchetgen Tchetgen et al. (2012) for drawing causal inference on a continuous longitudinal outcome subject to death, thereby addressing truncation. Shardell et al. (2015) augmented the original IPWGEE expressions, in the process allowing for separate specification of the dropout and truncation mechanisms. However, their approach does not handle IM. Both Chen and Zhou (2011) and Shardell et al. (2015) developed their methods for application to MAR data with time-varying covariates. Hence, their techniques are well-suited for use in augmenting RCC for MAR outcomes subject to truncation, dropout, and IM. We use a pattern-mixture approach similar to that of le Cessie et al. (2009) to develop a doubly robust RCC method that we call augmented RCC (ARCC).

2.5 Summary and Research Outline

Significant advances have been made in developing IPWGEE techniques that properly distinguish truncation from missingness. In particular, regression conditioning on being alive produces consistent estimates of longitudinal mean outcomes by consistently estimating the probability of dropout only for living participants at a given time point, then calculating the reciprocal of this estimated probability as the inverse-probability weight that is used to balance the contribution of each non-missing observation in weighted-GEE estimation. However, no such methods we know

of can adjust for both dropout and IM. Our RCC method accomplishes this through specification of distinct mechanisms for dropout and IM that reflect different underlying processes, wherein IM is defined as arbitrary or non-monotonic missingness occurring only before dropout.

In this dissertation, we create software to make the original RCA method available for widespread use, and extend RCA to accommodate distinct dropout and IM mechanisms. We call this extended-RCA method regression conditioning on continuation. In Chapter 2.5, we develop the RCC theory, compare its performance to three other common estimators under various scenarios, and apply RCC to an analysis of the BAN study data. Using the more general language and definitions of RCC, a Stata implementation of RCA is developed in Chapter 3.8 for general application to any dataset. Finally, in Chapter 4.6 we augment the RCC estimating equations to construct the ARCC estimator. The performance of ARCC and RCC is compared under various scenarios, and ARCC is applied to the previous BAN study analysis, with results compared to the those from RCC. Additional details are provided in the Appendix.

CHAPTER 3: REGRESSION CONDITIONING ON CONTINUATION

3.1 Introduction

The Breastfeeding, Antiretrovirals, and Nutrition (BAN) study (van der Horst et al., 2009; Chasela et al., 2010) was a clinical trial in which 2369 mothers infected with human immunodeficiency virus (HIV), along with their infants, were randomized in a 3×2 factorial design to one of three antiretroviral treatment (ART) arms, and to one of two lipid-based nutrient supplement (LNS) arms. Mother-infant pairs in the trial were scheduled to be followed for 48 weeks, with women receiving counseling for exclusive breastfeeding during the first 24 weeks, hereafter the "study period." BAN's objectives during the study period were to assess (i) the benefit of nutritional supplementation taken by women, (ii) the safety and efficacy of ART given either to mothers or their infants to prevent HIV transmission to the infant, and (iii) the feasibility of exclusive breastfeeding followed by early, rapid breastfeeding cessation. For the first objective, BAN study investigators wanted to draw inference about the mean infant anthropometric outcomes weight, length, and body mass index (BMI) over time among infants who were alive and HIV-uninfected. Infant outcomes were not of interest after infection or, of course, death. Therefore, death and infection are events that "truncate" the trajectory of infant outcomes over time. Additionally, some infant outcomes were not observed during the study period either because mother-infant pairs still in the study missed scheduled visits intermittently, or because they permanently dropped out of the study for reasons other than death or infection. In an analysis of the association of LNS with infant growth, Flax et al. (2012) included an infant in the analysis sample if s/he was a singleton at birth, who during the study period was alive and uninfected for the first two weeks, with available infant outcome data for more than one visit. The intent-to-treat (ITT) analysis included infants in the analysis regardless of death or infection in the first two weeks. Out of 2238 ITT-sample infants, by the end of the study period 307 had dropped out (14% of all infants), and 187 infants who had not dropped out had either become infected with HIV or had died (8% of all infants; 162 alive and infected, 25 dead). In addition, while they were alive, uninfected, and

still in the study, 645 infants missed 973 scheduled visits intermittently (5% of all scheduled visits before truncation or dropout).

Typical approaches to handling missing data such as weighted estimating equations (WEE) or maximum likelihood based on mixed-effects models frequently do not distinguish truncation from dropout, in essence envisaging infant outcomes past the point of death or infection. Kurland and Heagerty (2005) describe such approaches that implicitly assume the existence of infant outcomes after truncation as "unconditional regression" (UR) models because they estimate the mean outcome averaged over individuals who have and have not been truncated. Kurland et al. (2009) consider both standard selection models and conditional submodels of pattern-mixture models to be UR models. Mean infant outcomes for alive and uninfected infants may be estimated indirectly with these two types of UR models, with additional modeling assumptions (Kurland et al., 2009). As an alternative to UR models, joint modeling of longitudinal measurements and time to truncation might be employed (Henderson et al., 2000; Guo and Carlin, 2004; Kurland et al., 2009).

In order to estimate mean outcomes without relying on additional assumptions or joint modeling, Kurland and Heagerty (2005) developed a regression method conditional on being alive for a population with monotonic dropout, treating death as a truncating event. Because we are interested in outcomes for infants who continue on in the study not just alive but also uninfected, we will call this approach regression conditioning on continuation (RCC), where continuation is the complement of truncation. Shardell and Miller (2008) subsequently extended Kurland and Heagerty (2005) to handle missing covariates in addition to missing primary outcomes, and to allow intermittent missingness (IM) via the same mechanism as dropout. RCC consistently estimates longitudinal mean infant outcomes by utilizing weights based on the inverse of the estimated probability of dropout only for subjects with a continuing outcome at any given time point.

To date, RCC methods such as those of Kurland and Heagerty (2005) and Shardell and Miller (2008) do not allow for different reasons for dropout and IM. Such differences may be important in settings such as the BAN study, in which the reason a participant misses a study visit may differ from the reason for dropout. Yang and Shoptaw (2005) and Yang et al. (2008) recognized the substantive importance of distinguishing between dropout and IM in such instances, and hence

developed the likelihood-based method of multiple partial imputation (MPI) that adjusts for IM before handling dropout. Another limitation of RCC methods is that the empirical properties of these estimators have only been characterized in a few settings. For example, Kurland and Heagerty (2005) examined the empirical bias of the UR and RCC estimators for RCC parameters of interest when the truncation time was correlated with outcome, and when dropout depended on the truncation time but not on the outcome. They did not, however, investigate scenarios wherein dropout and IM depend on outcomes via different mechanisms, which is likely to be the case in practice.

Motivated by the BAN study, we develop an RCC method that defines IM as being conditional on non-dropout (similar to the MPI approach) in order to adjust for both IM and dropout. In Section 3.2, we introduce notation and key assumptions, motivate the use of RCC, and extend the method to allow for different IM and dropout mechanisms. In Section 3.3, the empirical bias and variance of the RCC and UR estimators are characterized in a simulation study. BAN study data are subsequently analyzed using RCC in Section 3.4, and some concluding remarks are given in Section 3.5. Additional details are provided in Sections 3.6 and 3.8.

3.2 Methods

3.2.1 Notation & Assumptions

Consider a random sample of $i = 1, \ldots, n$ subjects, each of whom is scheduled to be measured at fixed study time points $j = 1, \ldots, m$. Where it is not ambiguous, the dependence on i will be suppressed for notational ease. Let Y_j denote the outcome at time point j. Let $C_j = 1$ if the truncating event, i.e., death or infection, has not occurred by time point j, and let $C_j = 0$ otherwise. Assume the outcome Y_j is well defined if and only if $C_j = 1$. Assume that the truncated state is irreversible such that $C_j = 0$ implies $C_k = 0$ for all $k > j$. If $C_j = 1$, let $R_j = 1$ if the outcome is observed at time point j; otherwise, let $R_j = 0$. If $C_j = 1$, let $R_j^D = 1$ if an individual has not dropped out by time point j; otherwise, let $R_j^D = 0$. Dropout is defined to occur by time point j if all non-truncated outcomes are missing at and beyond time point j; i.e., $R_k = 0$ for all $k \geq j$ if $C_k = 1$. Note that $R_j^D = 0$ implies $R_k^D = 0$ for all $k > j$ such that $C_k = 1$. We use $*$ to denote all undefined values, and adopt the convention that $Y_j = *$, $R_j = *$, and $R_j^D = *$ if $C_j = 0$. Let $S = \sum_{j=1}^{m} C_j$ denote the number of visits before a trajectory was truncated, with

$S = m$ indicating that the trajectory was not truncated. The following shorthand notation will be used for clarity wherever possible. For any random variable A, let A_j denote the value of A at time point j, and let $\bar{A}_j = \big(A_1, \ldots, A_j\big)$ so that $\bar{A}_{j-1}$ represents an individual's history of A prior to time point j, where $\bar{A}_{j-1} = \emptyset$ at $j = 1$. For any random variable A, if A is discrete then let $p(a)$ denote $\Pr(A = a)$, the mass of A at a. Likewise, if A is continuous then let $p(a)$ denote $f(a)$, the density of A at a. Let $p(\cdot|a)$ denote $p(\cdot|A = a)$.

3.2.2 Dropout & Intermittent Missingness Mechanisms

In this section, we describe different assumptions regarding the probabilities of dropout and IM. The following conditional probabilities are needed for making our assumptions about dropout and IM. Let

$$\lambda_j^D(c_j) = \Pr\big(R_j^D = 1 \big| \bar{r}_{j-1}, R_{j-1}^D = 1, \bar{y}_{j-1}^{\text{obs}}, c_j\big),$$

where $\bar{y}_j^{\text{obs}} = \big\{y_k : R_k = 1, k \leq j\big\}$ denotes the observed values of $\bar{y}_j$. Thus, the probability of dropping out conditional on the history of missingness, on past non-dropout, on the history of observed outcomes, and on current truncation status is $1 - \lambda_j^D(c_j)$. Let

$$\lambda_j^{IM}(c_{j+1}, c_j) = \Pr\big(R_j = 1 \big| R_j^D = 1, \bar{r}_{j-1}, \bar{y}_{j-1}^{\text{obs}}, c_{j+1}, c_j\big),$$

where $\big\{c_{j+1}, c_j\big\} = \big\{c_j\big\}$ at $j = m$. Thus, for $j \leq m$ the probability of IM conditional on current non-dropout, on the histories of missingness and observed outcomes, on truncation status at the next time point if $j < m$, and on current truncation status is $1 - \lambda_j^{IM}(c_{j+1}, c_j)$.

We now define analogous assumptions regarding the dropout and IM mechanisms. We say dropout is at random (DAR) if

$$\Pr\big(R_j^D = 1 \big| \bar{r}_{j-1}, R_{j-1}^D = 1, \bar{y}_m, \bar{c}_m\big) = \lambda_j^D(c_j),$$

and dropout is completely at random (DCAR) if

$$\Pr\big(R_j^D = 1 \big| \bar{r}_{j-1}, R_{j-1}^D = 1, \bar{y}_m, \bar{c}_m\big) = \Pr\big(R_j^D = 1 \big| \bar{r}_{j-1}, R_{j-1}^D = 1, c_j\big).$$

Dropout is not at random if it is neither DAR nor DCAR. Likewise, we say IM is at random (IMAR) if

$$\Pr\left(R_j = 1 \middle| R_j^D = 1, \bar{r}_{j-1}, \bar{y}_m, \bar{c}_m\right) = \lambda_j^{IM}(c_{j+1}, c_j),$$

and IM is completely at random (IMCAR) if

$$\Pr\left(R_j = 1 \middle| R_j^D = 1, \bar{r}_{j-1}, \bar{y}_m, \bar{c}_m\right) = \Pr\left(R_j = 1 \middle| R_j^D = 1, \bar{r}_{j-1}, c_{j+1}, c_j\right).$$

IM is not at random if it is neither IMAR nor IMCAR.

These assumptions will be shown to imply that outcomes are missing at random (MAR). Outcomes are said to be MAR if

$$p\left(\bar{r}_m \middle| \bar{y}_m, \bar{c}_m\right) = p\left(\bar{r}_m \middle| \bar{y}_m^{\text{obs}}, \bar{c}_m\right). \tag{3.4}$$

Outcomes are missing completely at random (MCAR) if $p\left(\bar{r}_m \middle| \bar{y}_m, \bar{c}_m\right) = p\left(\bar{r}_m \middle| \bar{c}_m\right)$. Outcomes that are neither MAR nor MCAR are missing not at random. The left side of (3.4) can be expanded to

$$p\left(\bar{r}_m \middle| \bar{y}_m, \bar{c}_m\right) = p\left(\bar{r}_m, \bar{r}_m^D \middle| \bar{y}_m, \bar{c}_m\right) = \prod_{j=1}^{m} p\left(r_j, r_j^D \middle| \bar{r}_{j-1}, \bar{r}_{j-1}^D, \bar{y}_m, \bar{c}_m\right)$$

because $\bar{r}_m$ implies $\bar{r}_m^D$. Because $R_j^D = 1$ and $\bar{R}_j^D = (1, \ldots, 1)$ are equivalent statements, we have

$$\begin{aligned}
&\Pr\left(R_j = 1, R_j^D = 1 \middle| \bar{r}_{j-1}, \bar{r}_{j-1}^D, \bar{y}_m, \bar{c}_m\right) \\
&\quad = r_{j-1}^D \times \Pr\left(R_j = 1 \middle| R_j^D = 1, \bar{r}_{j-1}, \bar{y}_m, \bar{c}_m\right) \times \Pr\left(R_j^D = 1 \middle| \bar{r}_{j-1}, R_{j-1}^D = 1, \bar{y}_m, \bar{c}_m\right),
\end{aligned}$$

where we refer to $1 - \Pr\left(R_j = 1 \middle| R_j^D = 1, \bar{r}_{j-1}, \bar{y}_m, \bar{c}_m\right)$ as the IM mechanism, and where we refer to $1 - \Pr\left(R_j^D = 1 \middle| \bar{r}_{j-1}, R_{j-1}^D = 1, \bar{y}_m, \bar{c}_m\right)$ as the dropout mechanism. Hence, outcomes are MAR if DAR and IMAR are true.

Two common regression approaches misclassify truncation, dropout, or IM. The first is the UR approach mentioned earlier, which equates truncation with missingness. One may also define all missingness as IM and therefore only model missingness and not dropout, as was done in

Shardell and Miller (2008) with respect to the outcomes; call this the IMRCC approach. Define the UR analogues to $\lambda_j^D(c_j)$ and $\lambda_j^{IM}(c_{j+1}, c_j)$ as

$$\lambda_j^{D\dagger} = \Pr\left(R_j^D = 1 \middle| \bar{r}_{j-1}, R_{j-1}^D = 1, \bar{y}_{j-1}^{\text{obs}}\right),$$
$$\lambda_j^{IM\dagger} = \Pr\left(R_j = 1 \middle| R_j^D = 1, \bar{r}_{j-1}, \bar{y}_{j-1}^{\text{obs}}\right),$$

respectively, and let

$$\lambda_j^{IM\ddagger}(c_j) = \Pr\left(R_j = 1 \middle| \bar{r}_{j-1}, \bar{y}_{j-1}^{\text{obs}}, C_j = c_j\right)$$

represent the IMRCC analogue to $\lambda_j^{IM}(c_{j+1}, c_j)$. These quantities will be used in Section 3.2.3 to construct the UR and IMRCC estimators.

3.2.3 Estimators and Inference

In this section, we describe the RCC, UR, and IMRCC estimators used to draw inference about our estimand of interest, the mean outcome conditional on continuation at time point j for individual i, denoted $\mu_{ij}^{RCC} = E\big(Y_{ij} \big| C_{ij} = 1\big)$. In the regression setting, for a continuous and unbounded outcome we might posit a linear model of the form $\mu_{ij}^{RCC} = \boldsymbol{x}_{ij}'\boldsymbol{\beta}^{RCC}$, where $\boldsymbol{x}_{ij}$ is an observed $p \times 1$ vector of (possibly time-dependent) covariates with first element 1 for the intercept, and where $\boldsymbol{\beta}^{RCC}$ is the corresponding parameter vector. Following Kurland and Heagerty (2005), consider the vector estimating equation

$$U\left(\boldsymbol{\beta}^{RCC}\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} \boldsymbol{x}_{ij} C_{ij} \frac{R_{ij} R_{ij}^D}{\pi_{ij}} \left(Y_{ij} - \mu_{ij}^{RCC}\right), \tag{3.5}$$

where

$$\pi_{ij} = \Pr\left(R_{ij} = 1, R_{ij}^D = 1 \middle| \bar{r}_{i(j-1)}, \bar{r}_{i(j-1)}^D, \bar{y}_{im}, \bar{c}_{im}\right)$$

is the joint probability of not being missing and not having dropped out, conditional on the history of missingness and dropout, on all outcomes, and on the full truncation vector. We adopt the convention that if $C_{ij} = 0$, then the summand in (3.5) for individual i at time point j equals 0 rather than being undefined because of $Y_{ij} = *$. The probability π_{ij} is unknown in practice, but can be consistently estimated if the dropout and IM mechanism models are correctly specified.

The quantity $1/\pi_{ij}$ is the corresponding inverse-probability weight (IPW). Suppose DAR and IMAR hold such that $\pi_{ij} = \lambda_{ij}^{IM}(c_{i(j+1)}, c_{ij})\,\lambda_{ij}^{D}(c_{ij})$. Let $\hat{\pi}_{ij} = \hat{\lambda}_{ij}^{IM}(c_{i(j+1)}, c_{ij})\,\hat{\lambda}_{ij}^{D}(c_{ij})$ denote a consistent estimator of π_{ij}, and let $\hat{\beta}$ denote the solution to $U\big(\boldsymbol{\beta}^{RCC}\big) = \mathbf{0}$ when $\hat{\pi}_{ij}$ is substituted for π_{ij}. The estimator $\hat{\beta}$ is consistent and asymptotically multivariate normal for β^{RCC} (Robins et al., 1995). Standard software packages allow the empirical sandwich estimator of the variance of $\hat{\beta}$ to be computed as if the IPWs are known and fixed. This estimator is expected to be conservative (Robins et al., 2000; Robins, 2000; Preisser et al., 2002). Thus, 95% Wald confidence intervals (CIs) constructed using the empirical sandwich estimator should have a coverage probability for β^{RCC} of at least 95%.

Compared to RCC, the UR and IMRCC approaches each produce a different estimator that is generally not consistent for $\boldsymbol{\beta}^{RCC}$ when outcomes are DAR and IMAR. Comparing UR to RCC, outcomes for individuals with $C_{ij} = 0$ are used to estimate $\lambda_{ij}^{D\dagger}$ but not $\lambda_{ij}^{D}(1)$. And outcomes for individuals with $C_{i(j+1)} = 0$ are used to estimate $\lambda_{ij}^{IM\dagger}$ whereas $\lambda_{ij}^{IM}(0,1) = 1$ is set. The UR estimator is the solution to $U\big(\boldsymbol{\beta}^{RCC}\big) = \mathbf{0}$ with π_{ij} replaced by $\hat{\pi}_{ij}^{\dagger} = \hat{\lambda}_{ij}^{IM\dagger}\,\hat{\lambda}_{ij}^{D\dagger}$. If there is no truncation, then the RCC and UR estimators are identical. Otherwise, the estimating equations (3.5) with π_{ij} replaced by $\hat{\pi}_{ij}^{\dagger}$ will generally be biased for zero. Comparing IMRCC to RCC, outcomes for individuals with $R_{ij}^{D} = 0$ are used to estimate $\lambda_{ij}^{IM\ddagger}(1)$ but not $\lambda_{ij}^{IM}(1,1)$. At $j = S_i$ in particular, $\lambda_{ij}^{IM\ddagger}(1)$ is estimated while $\lambda_{ij}^{IM}(0,1) = 1$ is set. The IMRCC estimator is the solution to $U\big(\boldsymbol{\beta}^{RCC}\big) = \mathbf{0}$ with π_{ij} replaced by $\hat{\lambda}_{ij}^{IM\ddagger}$. If there is no dropout, then the RCC and IMRCC estimators are identical. Otherwise, the estimating equations (3.5) with π_{ij} replaced by $\hat{\lambda}_{ij}^{IM\ddagger}$ will generally be biased for zero. Hence, the UR and IMRCC estimators will generally not be consistent for $\boldsymbol{\beta}^{RCC}$, and therefore μ_{ij}^{RCC}.

3.3 Simulation Study

A simulation study was conducted to characterize the finite-sample performance of RCC and UR estimators. The continuous outcome of infant weight Y_{ij} was simulated for $n = 2238$ infants, reflecting the 2238 analysis sample infants of the BAN study. The mean outcome of interest was $\mu_{ij}^{RCC} = E\big(Y_{ij}\big|C_{ij} = 1\big)$ at visits $j = 1, \ldots, 10$. Outcomes were generated for $m = 11$ visits to reflect the fact that BAN outcomes (and therefore, missingness) were observed past the last visit of analytical interest, i.e., visit 10. Recall that the IM mechanism at visit j depends on whether

or not j is the last visit. Hence, estimation of the IM IPW at the last analytical visit depends on whether or not this is also the last visit with outcomes available for IM IPW estimation.

Simulations were conducted under a variety of scenarios with different truncation and missingness mechanisms. Truncation was not at random (TNAR) if truncation was generated conditional on possibly unobserved outcomes by specifying $\bar{c}_{im}$ associated with $\bar{y}_{im}$ whereby lighter infants were more likely to be truncated. Truncation was completely at random (TCAR) if truncation was generated independently of the outcomes by specifying $\bar{c}_{im}$ independent of $\bar{y}_{im}$. Simulation parameters were chosen so that heavier infants were more likely to drop out under DAR. Infant weight trajectories were simulated according to combinations of the following mechanisms: TCAR or TNAR; DCAR or DAR; and IMCAR, IMAR similar to the truncation mechanism (IMART), or IMAR similar to the dropout mechanism (IMARD). For each of the 12 resulting scenarios, we generated and analyzed $\ell = 1, \ldots, 1000$ simulated data sets. All parameter values for generating outcomes and truncation indicators were derived from the BAN data, and can be found in Tables 3.2 and 3.3.

3.3.1 Data Generation Procedure

Outcomes and truncation were generated first. For all infants and all visits, age at visit j was set equal to $\mu_{age(j)}$. Let T_i represent the natural logarithm of truncation time from birth for individual i. Let $\tau_j = \log\big(\mu_{age(j)}\big)$, and define $C_{ij} = I\big(T_i > \tau_j\big)$. Let $\rho_{y_j y_k}$ represent the autoregressive (lag-1) correlation between outcomes at time points j and k, and let $\rho_{y_j t}$ represent the correlation between outcome at time point j and T. Let $\mathbf{\Psi}$ represent the symmetric $(m+1) \times (m+1)$ matrix where $\big\{\rho_{y_j y_k}\big\}$ comprise the first corresponding $m \times m$ elements, and where $\big\{\rho_{y_j t}\big\}$ comprise the corresponding elements of row $m+1$ and column $m+1$, with element $m+1, m+1$ equal to 1. Outcomes and logged truncation time were generated from the multivariate normal distribution of $\big(Y_1, \ldots, Y_m, T\big)$ with means $\big(\mu_{y_1}, \ldots, \mu_{y_m}, \mu_t\big)$, variances $\big(\sigma^2_{y_1}, \ldots, \sigma^2_{y_m}, \sigma^2_t\big)$, and correlations $\mathbf{\Psi}$. TNAR outcomes were generated by setting $\rho_{y_j t} \equiv \gamma \sigma_{y_j} \sigma_t^{-1}$ where γ was a constant derived from the BAN data, while TCAR outcomes were generated by setting $\rho_{y_j t} \equiv 0$. Note that realizations t_i, c_{ij}, and s_i were thereby generated simultaneously.

Dropout and IM were subsequently generated. Note that $S_i \geq j$ and $C_{ij} = 1$ are equivalent statements, as are $S_i = j$ and $\big\{C_{i(j+1)} = 0, C_{ij} = 1\big\}$ where $\big\{C_{i(j+1)}\big\} = \emptyset$ at $j = m$. We

proceeded as follows.

1. If $s_i \geq 1$, then r_{i1}^D was generated using $\lambda_{i1}^D(1)$.

2. If $s_i > 1$, the following was done for $j < s_i$. If $r_{ij}^D = 0$, then $r_{ij} \equiv 0$ was set. Otherwise, if $r_{ij}^D = 1$ then r_{ij} was generated using $\lambda_{ij}^{IM}(1,1)$. Subsequently, if $r_{ij} = 0$ then $r_{i(j+1)}^D \equiv r_{ij}^D$ was set. Otherwise, if $r_{ij} = 1$ then $r_{i(j+1)}^D$ was generated using $\lambda_{i(j+1)}^D(1)$.

3. At $j = s_i$, $r_{ij} \equiv r_{ij}^D$ was set.

4. For all $j > s_i$, r_{ij}^D and r_{ij} were left undefined.

For a quantity b, let $g_j(b) = (j-1)^{-1} b \sum_{k=1}^{j-1} k R_{ik} Y_{ik}$ for $j > 1$. Dropout was generated using the probit model $\lambda_{ij}^D(1) = \Phi\{\eta_0^D + I(j > 1) g_j(\eta_1^D)\}$, where $\Phi(\,\cdot\,)$ is the standard normal cumulative distribution function, $I(a) = 1$ if statement a is true and $I(a) = 0$ otherwise, and $\eta_0^D = \Phi^{-1}(p_D^{1/m})$ where a fixed value was assigned to p_D. IM was generated using a model identical to the dropout model, with $\lambda_{ij}^D(1)$, η_0^D, η_1^D, and $p_D^{1/m}$ replaced with $\lambda_{ij}^{IM}(1,1)$, η_0^{IM}, η_1^{IM}, and p_{IM}, respectively, where a fixed value was assigned to p_{IM}. For each of the seven mechanisms, the simulation values p_D, η_1^D, p_{IM}, and η_1^{IM} are listed in Table 3.4. Values for p_D and p_{IM} were chosen so that truncation, dropout, and IM occurred at rates of approximately 8%, 14%, and 5%, respectively, by visit 10.

3.3.2 Results

For each simulated data set, the estimand of interest was parametrized as $\mu_{ij}^{RCC} = \beta_0^{RCC} + I(j > 1)\beta_{j-1}^{RCC}$. Let $\boldsymbol{\beta}^{RCC} = (\beta_0^{RCC}, \ldots, \beta_9^{RCC})'$, the parameter vector that was estimated for each data set using the following four regression methods.

1. URAR was estimated using GEE with autoregressive (AR) working correlation and no IPWs. This assumes truncation, dropout, and IM occur completely at random, and is not generally expected to be consistent.

2. UR was estimated using GEE with independence working correlation using IPWs. This allows truncation, dropout, and IM to occur at random, but assumes that all three events mask unobserved data. It is not generally expected to be consistent.

3. IMRCC was estimated using GEE with independence working correlation using IPWs. This allows truncation and IM to occur at random, and assumes that only IM masks unobserved data. However, this method assumes there is no dropout, and only specifies an IM model. Estimates of $\lambda_{ij}^{IM\ddagger}(1)$ were calculated by fitting the probit model for $\lambda_{ij}^{IM}(1,1)$ defined in Section 3.3.1. IMRCC is not generally expected to be consistent.

4. RCC was estimated using GEE with independence working correlation using IPWs. This allows truncation, dropout, and IM to occur at random, and assumes that only dropout and IM mask unobserved data. It is expected to be consistent.

Empirical bias and coverage were then calculated for each method as follows. Let $\hat{\beta}_{q\ell}$ denote the estimate of β_q^{RCC} for data set ℓ. For each of the $q = 0, \ldots, 9$ parameters, the empirical bias of $\hat{\beta}_q$ was calculated as $1000^{-1} \sum_{\ell=1}^{1000} \hat{\beta}_{q\ell} - \beta_q^{RCC}$. The estimate $\hat{\beta}_\ell$ and its empirical sandwich variance estimator were used to construct 95% Wald CIs. The empirical coverage probability of each $\hat{\beta}_q$ was calculated as the proportion of CIs over all 1000 data sets that contained β_q^{RCC}. The true value of μ_{ij}^{RCC}, a non-trivial function of μ_{y_j}, was calculated as detailed in Section 3.7.

The results are summarized as follows, where good performance was defined as an absolute empirical bias of less than or equal to 0.005. RCC and UR performed as good as or better than URAR and IMRCC in all 12 scenarios, and RCC generally performed slightly better than UR. URAR and IMRCC each performed worst in 10 out of 12 scenarios. Most importantly, IMRCC performed worst even when all events occurred completely at random because at any visit j, the IPW for an individual who had dropped out earlier was generally larger than the IPW for an individual who had not yet dropped out. This occurred because past (and therefore smaller) observed outcomes of individuals who had previously dropped out were overrepresented in estimating $\lambda_{ij}^{IM\ddagger}(1)$. The results for two scenarios are shown in Figure 3.1 for illustration. The increasing bias at later visits was a result of sample attrition, and was shown to decrease with samples of size 10000 (results not shown). The over-coverage of RCC was an expected result of using the empirical sandwich estimator to conservatively estimate the variance.

3.4 Analysis of the BAN Study

We applied the RCC method to the BAN data in our sample of $n = 2238$ infants. Our goal was to estimate mean infant outcome at each of nine scheduled follow-up visits for those infants who were alive and uninfected (i.e., who had continuing outcome trajectories) at that visit, while accounting for 307 dropouts, 187 truncations, and 973 IM observations.

The mean outcome for infant i at visit j conditional on continuation, $\mu_{ij} = E\big(Y_{ij}\big|C_{ij} = 1\big)$, was modeled separately for boys and girls because their growth patterns were considered to be different a priori. We modeled the mean outcome as a linear function of 1.) drug assignment to no ART (the reference), maternal ART, or infant ART, 2.) supplement assignment to no LNS (the reference) or LNS, 3.) dummy indicator variables for visit with visit 1 as the reference, and 4.) interactions between drug assignment, supplement assignment, and visit. The conditional probability of non-dropout, $\lambda_{ij}^{D}(1)$, was modeled as a probit function with observed past infant outcomes $\bar{Y}_{i(j-1)}^{\text{obs}}$ and drug/supplement group assignments and their interactions as predictors. The conditional probability of non-IM, $\lambda_{ij}^{IM}(1,1)$, was modeled likewise. The corresponding IPWs were used to estimate the mean outcomes at each visit, and standard errors were estimated using the empirical sandwich variance estimator.

We report the primary results for the infant outcomes of length, BMI, and weight. Infant length was not significantly associated with drug or supplement in either boys (Wald test $p = 0.72$) or girls ($p = 0.82$). Infant BMI was also not significantly associated with drug or supplement in either boys ($p = 0.60$) or girls ($p = 0.18$). Likewise, infant weight was not significantly associated with drug or supplement in either boys ($p = 0.46$) or girls ($p = 0.23$). Figure 3.2 depicts the estimated means and 95% CIs separately for girls and boys, for each treatment group at study period weeks 6, 12, 18, and 24 (i.e., visits 5, 7, 8, and 10, respectively).

These findings are consistent with those in Flax et al. (2012), who concluded that LNS was not significantly associated with infant weight. These authors also posited that LNS may have been associated with fewer adverse effects of ART on the weight of male infants. This result is generally supported by our findings. While Flax et al. (2012) considered all unobserved outcomes to be intermittently missing, our RCC approach distinguished truncation, dropout, and IM as different events by separating truncation from dropout and IM, and modeling dropout

and IM distinctly. Because our findings of statistical significance agree with those of the original analyses, they provide reassurance that the intermittent missingness, dropout, and truncation mechanisms may not have been notably different. However, the proportions of these events may also have been small enough to mask any true differences between mechanisms. Like Flax et al. (2012), we concluded that no treatments were substantively important in explaining the trend in mean infant weight over time, and instead report the reduced-model parameters in Table 3.1.

The BAN mean infant weight trajectories differed from the WHO median standard growth curves for children. These are plotted alongside their respective WHO growth curves (WHO Multicentre Growth Reference Study Group, 2006) in Figure 3.3. Exact values of WHO infant weights were not available for visits 8 through 10 (BAN study weeks 18, 21, and 24), and Figure 3.3 instead plots WHO weights at months 4, 5, and 6, which correspond to weeks 17.39, 21.74, and 26.09, respectively. The following discussion only applies to visits 2 through 7. At these visits, the BAN infant weights were significantly lower than those of girls and boys in the WHO standard population of children; i.e., the 95% confidence intervals of the BAN estimates did not overlap those of the WHO median infant weights at any visit. However, the BAN mean infant weights for both girls and boys fell above their respective WHO 25th percentiles (not shown). On average, the BAN girls were roughly 0.2 kg lighter than girls in the WHO standard population of children, while the BAN boys were roughly 0.3 kg lighter than boys in the WHO standard population of children.

3.5 Discussion

In this paper, the method of RCC for continuous longitudinal outcomes was extended to accommodate different dropout and IM mechanisms. The empirical performance of estimators using each of three IPW variants was characterized for 12 scenarios with different mechanisms for truncation, dropout, and intermittent missingness. Our simulation study demonstrated that RCC achieved much smaller empirical bias and better coverage than did IMRCC in general, even when truncation, dropout, and IM had occurred completely at random. The simulations also illustrated that RCC may be applied to samples with various truncation patterns without having to model the actual truncation probabilities.

Analysis of the BAN study data using our RCC method supported the findings of the original

study by Flax et al. (2012). Drug and supplement were not found to be significantly associated with weight in either HIV-negative infant girls or boys. Even after distinguishing truncation from dropout and IM, and modeling dropout and IM separately, mean model estimates did not substantially differ from the original findings that considered all unobserved outcomes to be intermittently missing at random. However, our investigation was limited to models that assumed at-random mechanisms, while actual mechanisms may have been not-at-random. Additional studies of the sensitivity of the mean infant weight estimates to such mechanism misspecifications could be undertaken to assess the robustness of the results presented here.

3.6 Simulation Parameter Values

All BAN-derived parameters for generating outcomes and truncation are listed in Supporting Tables 3.2 and 3.3. All parameters for generating truncation, dropout, and IM for each of the 12 simulation scenarios are listed in Supporting Table 3.4.

Parameters for generating outcomes and truncation were derived from the BAN data as follow. We first fit the mean model

$$Y_{ij} = \alpha'_0 + \alpha'_{j-1} I\big(j > 1\big) + \varepsilon_{ij}$$

to the BAN data, where $\varepsilon_{ij} \sim N\big(0, \sigma'^2_{\varepsilon_j}\big)$ was measurement error with $\text{cov}\big(\varepsilon_{ij}, \varepsilon_{i'j'}\big) = 0$ for all i, j, $i' \neq i$, and j'. We assumed a lag-1 autoregressive correlation structure for the errors. Let $\sigma'^2_{y_j} = V\big(Y_{ij}\big) = V\big(\varepsilon_{ij}\big) = \sigma'^2_{\varepsilon_j}$ Let $\rho'_{y_j y_k} = \text{corr}\big(Y_j, Y_k\big)$ for $k = 1, \ldots, m$, and let $\boldsymbol{\Psi}' = \big\{\rho'_{y_j y_k}\big\}$ represent the symmetric $m \times m$ correlation matrix. Let

$$\mu'_{y_j} = E\big(Y_{ij}\big) = \alpha'_0 + \alpha'_{j-1} I\big(j > 1\big).$$

We thereby obtained estimates of coefficients $\hat{\alpha}'_0, \ldots, \hat{\alpha}'_9$, variances $\hat{\sigma}'^2_{y_1}, \ldots, \hat{\sigma}'^2_{y_m}$, correlations $\hat{\boldsymbol{\Psi}}'$, and means $\hat{\mu}'_0, \ldots, \hat{\mu}'_9$. Let

$$T_i = \log\big(TD_i - BD_i\big),$$

where TD_i and BD_i are the truncation date and birth date, respectively, for individual i. Let μ'_t and σ'^2_t represent the empirical mean and variance, respectively, of logged truncation times for

Table 3.1: BAN infant weight RCC parameter estimates for reduced models.

Sex	Covariate	Estimate (95% CI)
male	intercept	3.07 (3.05, 3.10)
	visit 2	0.13 (0.12, 0.15)
	visit 3	0.37 (0.36, 0.39)
	visit 4	1.00 (0.98, 1.03)
	visit 5	1.56 (1.54, 1.59)
	visit 6	2.05 (2.02, 2.08)
	visit 7	2.85 (2.81, 2.89)
	visit 8	3.71 (3.66, 3.76)
	visit 9	4.05 (4.00, 4.10)
	visit 10	4.33 (4.27, 4.38)
female	intercept	2.98 (2.95, 3.00)
	visit 2	0.13 (0.12, 0.15)
	visit 3	0.34 (0.32, 0.36)
	visit 4	0.90 (0.87, 0.92)
	visit 5	1.39 (1.36, 1.42)
	visit 6	1.81 (1.78, 1.84)
	visit 7	2.52 (2.48, 2.56)
	visit 8	3.30 (3.26, 3.35)
	visit 9	3.61 (3.56, 3.66)
	visit 10	3.88 (3.82, 3.93)

Note: All estimates were statistically significant at $\alpha = 0.001$.

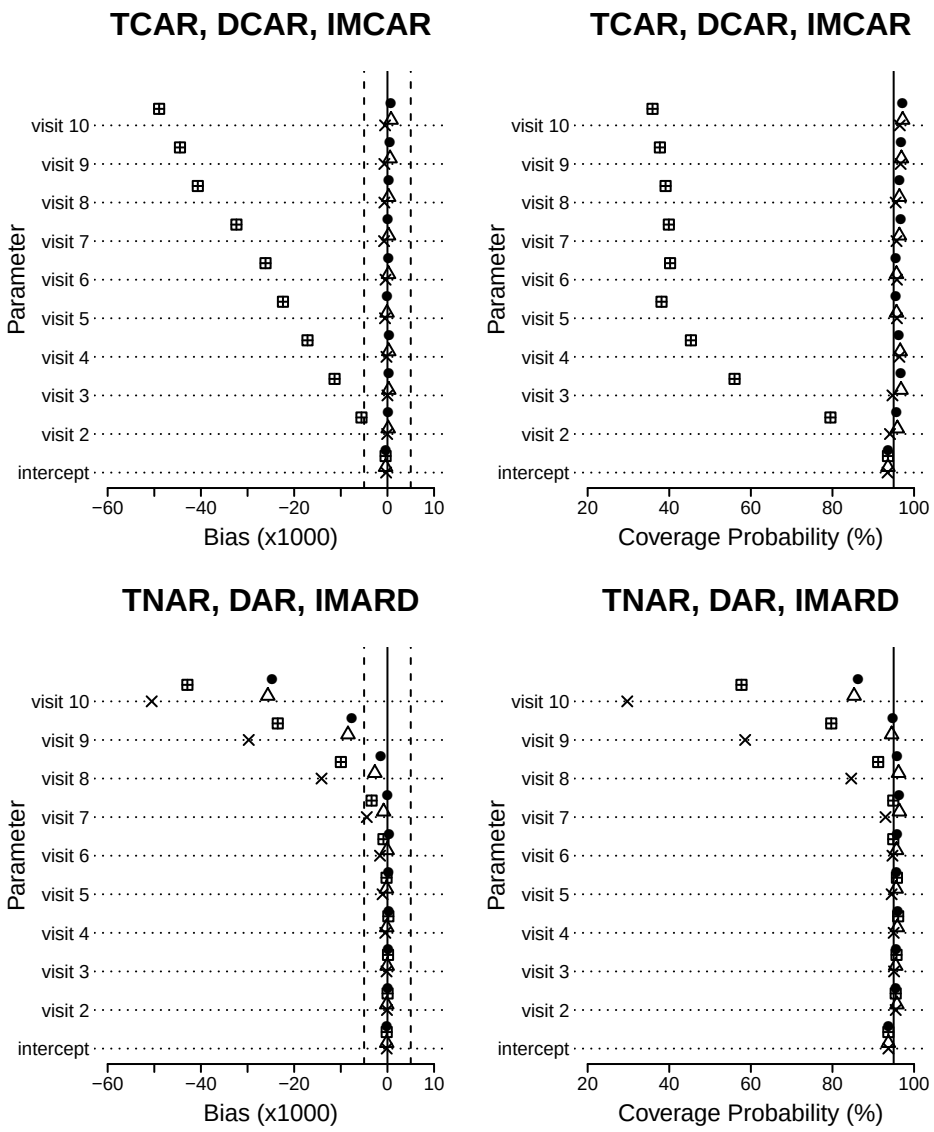


Figure 3.1: Simulation study: Empirical biases ($\times 1000$) and coverage probabilities (%) under TCAR, DCAR, IMCAR and TNAR, DAR, IMARD. (1000 simulated datasets, 2,238 subjects; URAR $\times$, UR $\triangle$, IMRCC $\boxplus$, RCC $\bullet$. Dotted lines are marked at ± 5 on bias figures.)

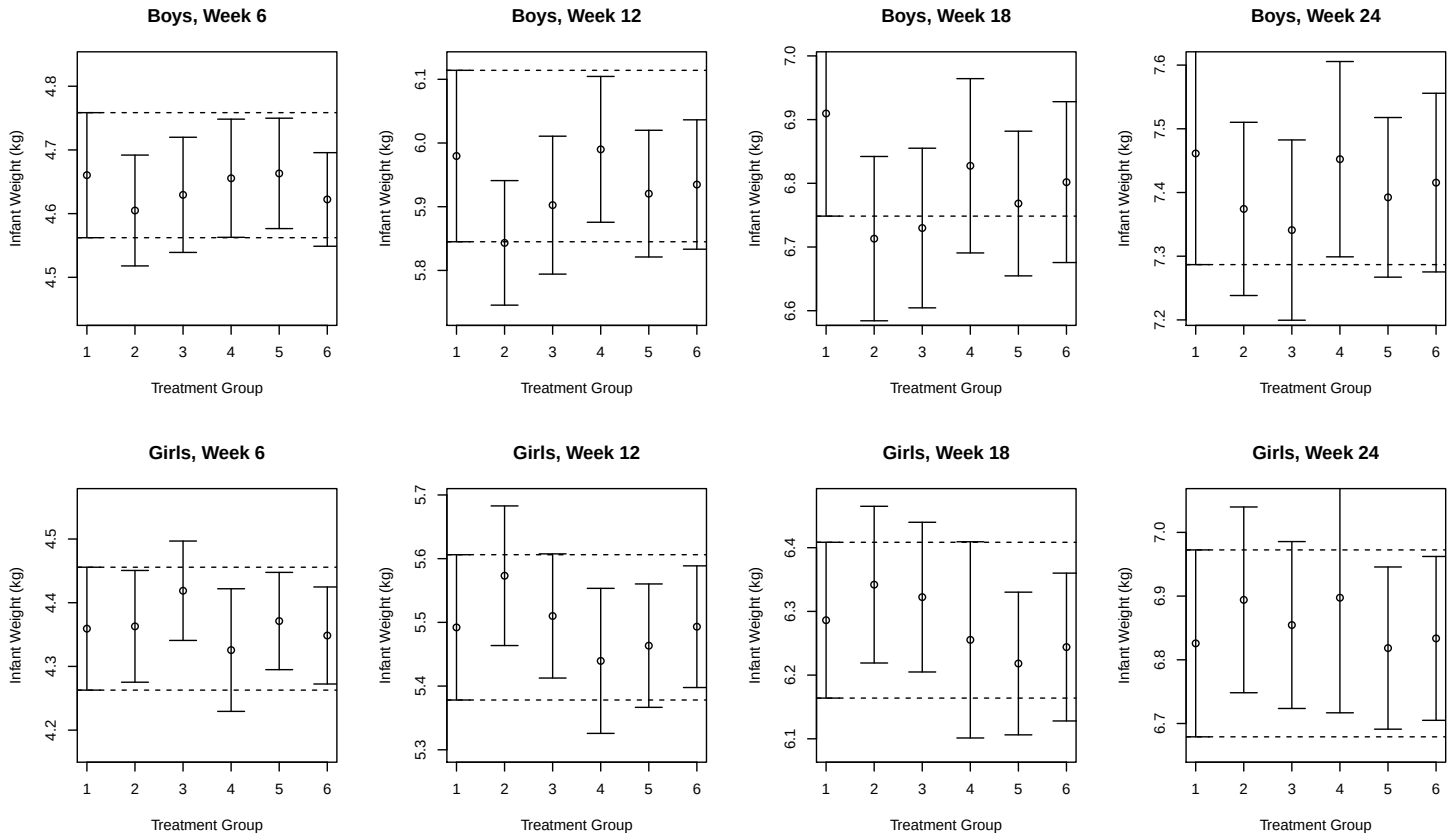


Figure 3.2: Estimates and 95% CIs of mean weight for HIV-negative, alive infants at study period weeks 6, 12, 18, and 24 using data from the BAN study. These correspond to study visits 5, 7, 8, and 10, respectively. The y-axis scales are identical. (Treatment Group: 1 = control and no LNS, 2 = maternal ART and no LNS, 3 = infant ART and no LNS, 4 = control and LNS, 5 = maternal ART and LNS, 6 = infant ART and LNS. Dashed lines correspond to CIs of the reference group, Treatment Group 1.)

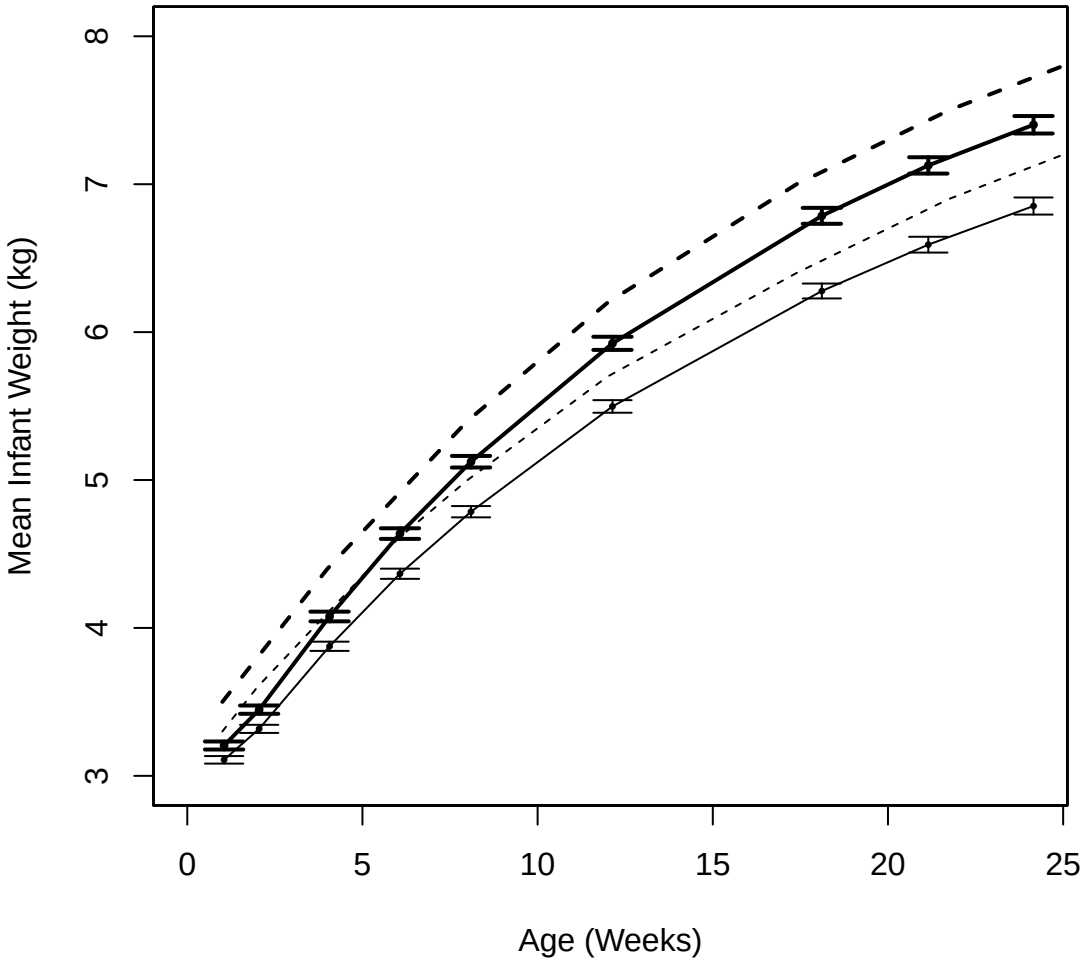


Figure 3.3: BAN study estimated mean weight trajectories and WHO median standard growth curves. (— BAN boys; **- - -** WHO boys; — BAN girls; - - - WHO girls)

individuals whose outcomes were eventually truncated, i.e., $\{t_i : c_{im} = 0\}$. Let $\mu'_{age(j)}$ denote the empirical mean age at visit j. At each visit $j = 1, \ldots, m$, we fit the mean model

$$T_{ij} = \gamma'_0 + \gamma'_j Y_{ij} + \delta_{ij},$$

where $T_{ij} = T_i$, and $\delta_{ij} \sim N(0, \sigma'^2_{\delta_j})$ was measurement error with $\text{cov}(\delta_{ij}, \delta_{i'j}) = 0$ for all i and $i' \neq i$ at each j. Let

$$\tilde{\gamma} = m^{-1} \sum_{j=1}^{m} \gamma'_j.$$

We assumed $\text{cov}(\delta_{ij}, \varepsilon_{i'j'}) = 0$ for all i, j, i', and j'. We thereby calculated the parameters μ'_t, σ'^2_t, and $\mu'_{age(1)}, \ldots, \mu'_{age(m)}$, and the estimate $\hat{\tilde{\gamma}}$.

The parameters used for outcome and truncation generation were set equal to their corresponding BAN-derived estimate or parameter values; e.g., $\sigma_{y_4} \equiv \hat{\sigma}'_{y_4}$ and $\mu_{age(1)} \equiv \mu'_{age(1)}$. There were two important exceptions. We set $\mu_t = \zeta_\mu \mu'_t$, where $\zeta_\mu \equiv 2.72$ was set to ensure the desired truncation rate of about 8% by visit 10. We also set $\gamma = \zeta_\gamma \hat{\tilde{\gamma}}$, where $\zeta_\gamma \equiv 5$ was set to ensure a desired magnitude of the association between outcome Y_{ij} and truncation time T_i.

3.7 Simulation Joint Distribution Properties

We will prove that $\mu_{ij}^{RCC} = \mu_{y_j} - \omega_j$ where

$$\omega_j = \frac{\phi\left(\frac{\tau_j - \mu_t}{\sigma_t}\right) \sigma_{y_j} \rho_{y_j t}}{\Phi\left(\frac{\tau_j - \mu_t}{\sigma_t}\right) - 1}.$$

First, we will derive a skew-normal-type (SNT) distribution. We will then use properties of the SNT to complete the proof.

3.7.1 Skew-Normal-Type Distribution

Identity between a Cumulative Distribution Function and an Expectation

We first define some notation that only applies to this subsection. Let S and X be continuous random variables, and let a and b be constants. Define $W = S - (aX + b)$, so that $F_S(aX + b) = \Pr(S < aX + b) = \Pr\{S - (aX + b) < 0\} = \Pr(W < 0)$. For any distinct random variables A and B, let $f_A(a)$ denote the probability density function (PDF) of A, and let $f_{A|B}(a|b)$ denote the PDF

of A conditioned on B. Let $F_A(a)$ and $F_{A|B}(a|b)$ denote the corresponding cumulative distribution functions (CDFs). Let $E_A(A)$ and $E_{A|B}(A|B)$ denote the corresponding expectations. It can be shown that

$$F_S(aX+b) = E_X\left\{F_S\left(aX+b\right)\right\}. \tag{3.6}$$

In particular, if $S \sim N(0,1)$, $X \sim N(\mu, \sigma^2)$, and S and X are independent, then $W \sim N\big(-\{a\mu + b\}, 1 + a^2\sigma^2\big)$. Let $\Phi(\cdot)$ represent the standard normal CDF. In this case, we have

$$F_S(aX+b) = \Phi\left(\frac{a\mu+b}{\sqrt{1+a^2\sigma^2}}\right),$$

and from (3.6), we write

$$E_X\left\{\Phi(aX+b)\right\} = \Phi(aX+b) = \Phi\left(\frac{a\mu+b}{\sqrt{1+a^2\sigma^2}}\right). \tag{3.7}$$

Similar results have been proven elsewhere (Ellison, 1964; Azzalini, 1985).

Relationship to Skew-Normal Distribution

From (3.7), note that

$$1 = \Phi\left(\frac{a\mu+b}{\sqrt{1+a^2\sigma^2}}\right)^{-1} \int \Phi(ax+b) f_X(x) dx.$$

That is,

$$f_Y(y) = \Phi\left(\frac{a\mu+b}{\sqrt{1+a^2\sigma^2}}\right)^{-1} \Phi(ay+b) f_X(y) \tag{3.8}$$

is itself a probability density function with moment-generating function

$$M_Y(t) = E\left(e^{tY}\right) = \Phi\left\{\frac{a\mu+b}{\sqrt{1+a^2\sigma^2}}\right\}^{-1} \exp\left\{\frac{t}{2}\left(t\sigma^2+2\mu\right)\right\} \Phi\left\{\frac{a\left(t\sigma^2+\mu\right)+b}{\sqrt{1+a^2\sigma^2}}\right\}$$

and mean

$$E(Y) = \frac{d}{dt} M_Y(t)\Big|_{t=0} = \mu + \Phi\left(\frac{a\mu + b}{\sqrt{1 + a^2\sigma^2}}\right)^{-1} \phi\left(\frac{a\mu + b}{\sqrt{1 + a^2\sigma^2}}\right) \frac{a\sigma^2}{\sqrt{1 + a^2\sigma^2}}, \tag{3.9}$$

where $\phi(\cdot)$ represents the standard normal PDF. The case when $b = 0$ and $X \sim N(0, 1)$ corresponds to the skew-normal distribution (Azzalini, 1985) with density

$$f_Y(y) = 2\Phi(ay)\phi(y)$$

and mean

$$E_Y[Y] = \sqrt{\frac{2}{\pi}}\left(\frac{a}{\sqrt{1 + a^2}}\right).$$

We therefore say a random variable Y with density (3.8) has a skew-normal-type (SNT) distribution with mean (3.9).

3.7.2 Proof

We now use the notation defined in the main article. The notation in this subsection is distinct from that in subsection 3.7.1.

Recall that outcomes and logged truncation time were generated from the multivariate normal distribution of $\big(Y_1, \ldots, Y_m, T\big)$ with means $\big(\mu_{y_1}, \ldots, \mu_{y_m}, \mu_t\big)$, variances $\big(\sigma^2_{y_1}, \ldots, \sigma^2_{y_m}, \sigma^2_t\big)$, and correlations $\mathbf{\Psi}$. Also recall that $\rho_{y_j y_k} = \text{corr}\big(Y_j, Y_k\big)$ represents the autoregressive (lag-1) correlation between outcomes at time points $j = 1, \ldots, m$ and $k = 1, \ldots, m$, and that $\rho_{y_j t} = \text{corr}\big(Y_j, T\big)$ represents the correlation between outcome at time point $j = 1, \ldots, m$ and T. Let $\phi(\cdot)$ represents the standard normal probability density function. Let $f_A(a)$ and $f(a)$ both denote the PDF of a continuous random variable A. Note that $\rho_{y_j t} = \sigma_{y_j t}\big(\sigma_{y_j}\sigma_t\big)^{-1}$.

We write $\big\{T\big|Y_j\big\} \sim N\big(\mu_{t|y_j}, \sigma^2_{t|y_j}\big)$ where

$$\mu_{t|y_j} = \mu_t + \frac{\sigma_t}{\sigma_{y_j}}\rho_{y_j t}\left(y_j - \mu_{y_j}\right), \qquad \sigma^2_{t|y_j} = \sigma_t\sqrt{1 - \rho^2_{y_j t}}.$$

Note that $\Pr\left(T \leq \tau_j \middle| y_j\right) = \Phi\left(b_{0j} + b_{1j} y_j\right)$, where $b_{0j} = a_{0j} - b_{1j}\mu_{y_j}$ and

$$a_{0j} = \frac{\tau_j - \mu_t}{\sigma_t \sqrt{1 - \rho_{y_j t}^2}}, \qquad\qquad b_{1j} = -\frac{\frac{\sigma_t}{\sigma_{y_j}} \rho_{y_j t}}{\sigma_t \sqrt{1 - \rho_{y_j t}^2}}.$$

Let $\xi_j = E\left\{\Phi\left(b_{0j} + b_{1j} Y_j\right)\right\}$, and note that $\xi_j = \Phi\left\{\left(b_{0j} + b_{1j}\mu_{y_j}\right) g_j\right\}$ by (3.7), where $g_j = 1/\sqrt{1 + b_{1j}^2 \sigma_{y_j}^2}$. Let $h_j = b_{1j}\sigma_{y_j}^2 g_j = -\sigma_{y_j}\rho_{y_j t}$. We then have

$$\begin{aligned}
\mu_{ij}^{RCC} &= E\left(Y_{ij} \middle| C_j = 1\right) \\
&= E\left(Y_j \middle| T > \tau_j\right) \\
&= \frac{\int y_j \left\{1 - \Phi\left(b_{0j} + b_{1j} y_j\right)\right\} f\left(y_j\right) dy_j}{\int \left\{1 - \Phi\left(b_{0j} + b_{1j} y_j\right)\right\} f\left(y_j\right) dy_j} \\
&= \frac{\mu_{y_j} - \int y_j \Phi\left(b_{0j} + b_{1j} y_j\right) f\left(y_j\right) dy_j}{1 - \xi_j} && \text{by (3.7)} \\
\xi_j^{-1}\left\{\mu_{ij}^{RCC}\left(\xi_j - 1\right) + \mu_{y_j}\right\} &= \int w_j \xi_j^{-1} \Phi\left(b_{0j} + b_{1j} w_j\right) f_Y\left(w_j\right) dw_j \\
&= E\left[W_j\right] && \text{by (3.8)} \\
&= \mu_{y_j} + \xi_j^{-1} \phi\left\{\left(b_{0j} + b_{1j}\mu_{y_j}\right) g_j\right\} h_j && \text{by (3.9)} \\
\mu_{ij}^{RCC} &= \mu_{y_j} + \frac{\phi\left\{\left(b_{0j} + b_{1j}\mu_{y_j}\right) g_j\right\} h_j}{\xi_j - 1} \\
&= \mu_{y_j} + \frac{\phi\left(a_{0j} g_j\right) h_j}{\Phi\left(a_{0j} g_j\right) - 1}.
\end{aligned}$$

Noting that $a_{0j} g_j = \left(\tau_j - \mu_t\right)/\sigma_t$, we therefore conclude

$$\mu_{ij}^{RCC} = \mu_{y_j} - \frac{\phi\left(\frac{\tau_j - \mu_t}{\sigma_t}\right) \sigma_{y_j} \rho_{y_j t}}{\Phi\left(\frac{\tau_j - \mu_t}{\sigma_t}\right) - 1}.$$

■

For additional algebraic details, please email the primary author at ericjaydaza@unc.edu.

3.8 Detailed Simulation Results

The empirical bias and coverage probability results for all 12 simulation scenarios are listed in Supporting Table 3.5. A scenario name is written in the format XXYYZZ, where XX, YY, and

ZZ represent abbreviations that indicate the truncation, dropout, and IM mechanisms, respectively. TCAR and TNAR are abbreviated TC and TN, respectively. DCAR and DAR are abbreviated DC and DA, respectively. IMCAR, IMART, and IMARD are abbreviated MC, MT, and MD, respectively.

Supporting Table 3.2: Simulation parameters

Parameter	Values					
$\mu_{age(1)} \cdots \mu_{age(6)}$	0.87	7.87	14.87	28.87	42.87	56.87
$\mu_{age(7)} \cdots \mu_{age(11)}$	84.87	126.87	147.87	168.87	196.87	
$\mu_{y_1} \cdots \mu_{y_6}$	3.03	3.16	3.39	3.99	4.52	4.97
$\mu_{y_7} \cdots \mu_{y_{11}}$	5.72	6.55	6.88	7.14	7.44	
μ_t	8.55					
$\sigma_{y_1} \cdots \sigma_{y_6}$	0.41	0.42	0.44	0.51	0.55	0.61
$\sigma_{y_7} \cdots \sigma_{y_{11}}$	0.70	0.83	0.86	0.90	0.94	
σ_t	2.38					
γ	0.11					

Supporting Table 3.3: Simulation correlation matrix $\mathbf{\Psi}$

1											
0.92	1										
0.84	0.92	1									
0.78	0.84	0.92	1								
0.71	0.78	0.84	0.92	1							
0.66	0.71	0.78	0.84	0.92	1						
0.60	0.66	0.71	0.78	0.84	0.92	1					
0.56	0.60	0.66	0.71	0.78	0.84	0.92	1				
0.51	0.56	0.60	0.66	0.71	0.78	0.84	0.92	1			
0.47	0.51	0.56	0.60	0.66	0.71	0.78	0.84	0.92	1		
0.43	0.47	0.51	0.56	0.60	0.66	0.71	0.78	0.84	0.92	1	
$\rho_{y_1 t}$	$\rho_{y_2 t}$	$\rho_{y_3 t}$	$\rho_{y_4 t}$	$\rho_{y_5 t}$	$\rho_{y_6 t}$	$\rho_{y_7 t}$	$\rho_{y_8 t}$	$\rho_{y_9 t}$	$\rho_{y_{10} t}$	$\rho_{y_{11} t}$	1

Note: Values for $\rho_{y_j t}$ depend on whether outcomes are TCAR or TNAR, and are derived in Table 3.4 using values from Table 3.2.

Supporting Table 3.4: Truncation, dropout, and IM settings for 12 scenarios

Scenario	$\rho_{y_j t}$	p_D	η_1^D	p_{IM}	η_1^{IM}
TCAR, DCAR, IMCAR	0	0.150	0	0.140	0
TCAR, DCAR, IMART	0	0.150	0	0.260	γ
TCAR, DCAR, IMARD	0	0.142	0	0.031	$-\gamma$
TCAR, DAR, IMCAR	0	0.014	$-\gamma$	0.140	0
TCAR, DAR, IMART	0	0.013	$-\gamma$	0.260	γ
TCAR, DAR, IMARD	0	0.013	$-\gamma$	0.031	$-\gamma$
TNAR, DCAR, IMCAR	$\gamma\sigma_{y_j}\sigma_t^{-1}$	0.150	0	0.140	0
TNAR, DCAR, IMART	$\gamma\sigma_{y_j}\sigma_t^{-1}$	0.150	0	0.260	γ
TNAR, DCAR, IMARD	$\gamma\sigma_{y_j}\sigma_t^{-1}$	0.142	0	0.031	$-\gamma$
TNAR, DAR, IMCAR	$\gamma\sigma_{y_j}\sigma_t^{-1}$	0.014	$-\gamma$	0.140	0
TNAR, DAR, IMART	$\gamma\sigma_{y_j}\sigma_t^{-1}$	0.013	$-\gamma$	0.260	γ
TNAR, DAR, IMARD	$\gamma\sigma_{y_j}\sigma_t^{-1}$	0.013	$-\gamma$	0.031	$-\gamma$

Note: Values for γ, σ_{y_j}, and σ_t are listed in Table 3.2.

Supporting Table 3.5: Detailed simulation results. Values are listed per coefficient as "empirical bias $\times 1000$ (coverage probability)."

Scenario	Regression Method	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9
TCDCMC	URAR	−0.29 (0.94)	−0.06 (0.94)	0.00 (0.95)	−0.21 (0.96)	−0.52 (0.96)	−0.40 (0.96)	−0.72 (0.96)	−0.66 (0.95)	−0.66 (0.97)	−0.51 (0.96)
	UR	−0.39 (0.94)	0.16 (0.96)	0.29 (0.97)	0.28 (0.97)	−0.09 (0.96)	0.23 (0.96)	0.23 (0.96)	0.26 (0.96)	0.56 (0.97)	0.78 (0.97)
	IMRCC	−0.39 (0.94)	−5.57 (0.80)	−11.34 (0.56)	−17.14 (0.45)	−22.41 (0.38)	−26.15 (0.40)	−32.39 (0.40)	−40.67 (0.39)	−44.50 (0.38)	−48.97 (0.36)
	RCC	−0.39 (0.94)	0.17 (0.96)	0.27 (0.97)	0.32 (0.96)	−0.10 (0.95)	0.22 (0.95)	0.03 (0.97)	0.31 (0.96)	0.51 (0.97)	0.74 (0.97)
TCDCMT	URAR	−0.18 (0.94)	−0.06 (0.95)	−0.41 (0.96)	−0.06 (0.96)	−0.42 (0.95)	−0.14 (0.95)	−0.24 (0.95)	−0.04 (0.96)	−0.42 (0.95)	−0.44 (0.96)
	UR	−0.35 (0.93)	0.24 (0.95)	0.20 (0.96)	0.05 (0.96)	−0.36 (0.95)	−0.06 (0.95)	−0.01 (0.97)	−0.10 (0.96)	−0.45 (0.96)	−0.55 (0.97)
	IMRCC	−0.35 (0.93)	−1.72 (0.93)	−6.05 (0.88)	−11.73 (0.75)	−16.76 (0.63)	−20.40 (0.64)	−26.75 (0.61)	−32.94 (0.65)	−34.75 (0.69)	−32.75 (0.75)
	RCC	−0.35 (0.93)	0.25 (0.95)	0.18 (0.96)	0.08 (0.96)	−0.39 (0.95)	−0.06 (0.95)	−0.21 (0.96)	0.02 (0.96)	−0.52 (0.96)	−0.53 (0.97)
TCDCMD	URAR	−0.24 (0.94)	−0.12 (0.95)	−0.12 (0.96)	−0.26 (0.95)	−0.68 (0.95)	−0.97 (0.95)	−2.65 (0.94)	−8.15 (0.92)	−17.45 (0.83)	−30.17 (0.70)
	UR	−0.24 (0.94)	−0.10 (0.97)	−0.10 (0.97)	−0.20 (0.96)	−0.52 (0.96)	−0.44 (0.96)	−0.45 (0.96)	−0.09 (0.96)	0.21 (0.97)	0.43 (0.96)
	IMRCC	−0.24 (0.94)	−8.30 (0.45)	−8.71 (0.68)	−10.17 (0.77)	−13.17 (0.73)	−18.05 (0.65)	−30.58 (0.42)	−56.22 (0.12)	−89.54 (0.01)	−122.19 (0.00)
	RCC	−0.24 (0.94)	−0.09 (0.96)	−0.11 (0.97)	−0.16 (0.96)	−0.55 (0.96)	−0.45 (0.96)	−0.63 (0.96)	−0.00 (0.96)	0.08 (0.97)	0.33 (0.97)
TCDAMC	URAR	−0.10 (0.94)	−0.09 (0.94)	−0.22 (0.94)	−0.24 (0.95)	−0.61 (0.94)	−0.65 (0.94)	−1.43 (0.95)	−3.60 (0.94)	−9.94 (0.90)	−25.33 (0.71)
	UR	−0.18 (0.94)	−0.05 (0.96)	0.11 (0.96)	0.12 (0.97)	−0.18 (0.95)	−0.09 (0.95)	−0.84 (0.96)	−2.39 (0.96)	−7.73 (0.94)	−25.26 (0.78)
	IMRCC	−0.18 (0.94)	−0.04 (0.95)	0.06 (0.96)	0.04 (0.96)	−0.47 (0.95)	−0.77 (0.95)	−2.86 (0.95)	−7.32 (0.94)	−19.78 (0.81)	−47.84 (0.38)
	RCC	−0.18 (0.94)	−0.04 (0.95)	0.10 (0.96)	0.15 (0.97)	−0.21 (0.95)	−0.09 (0.95)	−0.99 (0.96)	−2.24 (0.96)	−7.56 (0.94)	−24.89 (0.79)
TCDAMT	URAR	0.06 (0.95)	−0.10 (0.95)	−0.39 (0.94)	−0.12 (0.95)	−0.45 (0.95)	−0.43 (0.95)	−1.12 (0.95)	−3.30 (0.94)	−9.97 (0.91)	−26.27 (0.71)
	UR	−0.43 (0.94)	0.10 (0.95)	0.24 (0.95)	0.23 (0.96)	−0.09 (0.96)	−0.02 (0.94)	−0.57 (0.95)	−2.18 (0.95)	−7.73 (0.94)	−26.04 (0.77)
	IMRCC	−0.43 (0.94)	0.10 (0.95)	0.21 (0.95)	0.22 (0.96)	−0.25 (0.95)	−0.56 (0.95)	−2.59 (0.94)	−7.59 (0.93)	−24.26 (0.71)	−69.25 (0.08)
	RCC	−0.43 (0.94)	0.10 (0.95)	0.22 (0.95)	0.26 (0.96)	−0.11 (0.95)	−0.02 (0.95)	−0.72 (0.95)	−2.04 (0.95)	−7.59 (0.94)	−25.67 (0.79)
TCDAMD	URAR	−0.20 (0.93)	−0.07 (0.95)	−0.11 (0.95)	−0.24 (0.95)	−0.62 (0.95)	−1.06 (0.95)	−3.45 (0.94)	−12.54 (0.88)	−27.58 (0.64)	−47.77 (0.34)
	UR	−0.20 (0.93)	−0.07 (0.96)	−0.06 (0.95)	−0.16 (0.96)	−0.36 (0.96)	−0.38 (0.95)	−0.79 (0.97)	−2.69 (0.96)	−8.69 (0.95)	−25.60 (0.86)
	IMRCC	−0.20 (0.93)	−0.07 (0.96)	−0.11 (0.96)	−0.22 (0.96)	−0.73 (0.95)	−1.55 (0.94)	−4.32 (0.94)	−11.11 (0.91)	−24.34 (0.77)	−43.38 (0.58)
	RCC	−0.20 (0.93)	−0.07 (0.96)	−0.08 (0.96)	−0.13 (0.96)	−0.38 (0.95)	−0.38 (0.95)	−0.94 (0.96)	−2.56 (0.96)	−8.60 (0.96)	−25.46 (0.86)
TNDCMC	URAR	−0.24 (0.94)	−0.04 (0.95)	−0.12 (0.94)	−0.23 (0.95)	−0.70 (0.96)	−0.93 (0.95)	−1.40 (0.94)	−2.16 (0.95)	−2.59 (0.96)	−3.15 (0.96)
	UR	−0.28 (0.93)	0.11 (0.96)	0.35 (0.97)	0.38 (0.96)	0.38 (0.96)	0.58 (0.96)	0.42 (0.96)	0.22 (0.96)	0.93 (0.96)	0.82 (0.96)
	IMRCC	−0.28 (0.93)	−5.49 (0.81)	−11.06 (0.59)	−16.76 (0.47)	−21.44 (0.41)	−25.54 (0.43)	−31.24 (0.44)	−39.44 (0.44)	−43.41 (0.41)	−48.60 (0.39)
	RCC	−0.27 (0.93)	0.30 (0.96)	0.56 (0.97)	0.84 (0.96)	0.79 (0.96)	0.98 (0.95)	1.20 (0.96)	1.37 (0.96)	1.48 (0.97)	1.38 (0.97)
TNDCMT	URAR	−0.29 (0.94)	0.09 (0.96)	−0.27 (0.96)	−0.16 (0.94)	−0.59 (0.95)	−0.71 (0.95)	−1.30 (0.94)	−2.03 (0.95)	−2.70 (0.95)	−3.14 (0.95)
	UR	−0.20 (0.93)	−0.05 (0.95)	0.21 (0.95)	−0.14 (0.96)	−0.34 (0.95)	−0.03 (0.95)	−0.43 (0.95)	−0.81 (0.95)	−0.49 (0.96)	−0.57 (0.95)
	IMRCC	−0.20 (0.93)	−1.83 (0.94)	−5.83 (0.90)	−11.60 (0.76)	−16.38 (0.64)	−20.16 (0.65)	−25.89 (0.63)	−32.20 (0.66)	−34.98 (0.69)	−32.21 (0.77)
	RCC	−0.16 (0.93)	0.13 (0.95)	0.42 (0.95)	0.30 (0.96)	0.02 (0.95)	0.34 (0.95)	0.28 (0.95)	0.39 (0.95)	0.05 (0.95)	−0.07 (0.96)
TNDCMD	URAR	−0.19 (0.94)	−0.16 (0.95)	−0.21 (0.95)	−0.48 (0.96)	−1.02 (0.96)	−1.47 (0.94)	−3.47 (0.94)	−9.87 (0.91)	−19.67 (0.80)	−32.55 (0.65)
	UR	−0.19 (0.94)	−0.18 (0.96)	−0.10 (0.97)	−0.20 (0.96)	−0.18 (0.96)	0.07 (0.95)	−0.17 (0.96)	0.11 (0.96)	0.53 (0.97)	0.89 (0.98)
	IMRCC	−0.19 (0.94)	−8.18 (0.46)	−8.64 (0.68)	−9.76 (0.78)	−12.28 (0.77)	−17.27 (0.69)	−29.26 (0.44)	−55.18 (0.13)	−88.44 (0.00)	−120.90 (0.00)
	RCC	−0.19 (0.94)	0.01 (0.96)	0.11 (0.97)	0.27 (0.96)	0.22 (0.96)	0.48 (0.96)	0.59 (0.96)	1.30 (0.95)	1.09 (0.97)	1.48 (0.97)
TNDAMC	URAR	−0.06 (0.94)	−0.10 (0.93)	−0.29 (0.94)	−0.49 (0.95)	−0.89 (0.94)	−1.15 (0.94)	−2.28 (0.94)	−5.39 (0.93)	−12.17 (0.90)	−28.02 (0.69)
	UR	−0.18 (0.94)	−0.09 (0.96)	0.03 (0.97)	0.20 (0.97)	0.07 (0.95)	0.22 (0.96)	−0.63 (0.96)	−2.50 (0.95)	−7.58 (0.94)	−25.12 (0.78)
	IMRCC	−0.18 (0.94)	0.10 (0.96)	0.21 (0.97)	0.56 (0.96)	0.20 (0.95)	−0.06 (0.95)	−1.73 (0.95)	−6.41 (0.93)	−18.92 (0.81)	−47.10 (0.39)
	RCC	−0.17 (0.94)	0.10 (0.96)	0.24 (0.97)	0.66 (0.96)	0.47 (0.95)	0.64 (0.96)	0.18 (0.96)	−1.20 (0.95)	−6.73 (0.95)	−24.10 (0.81)
TNDAMT	URAR	0.23 (0.95)	−0.17 (0.95)	−0.62 (0.96)	−0.51 (0.95)	−0.97 (0.96)	−1.10 (0.95)	−2.23 (0.95)	−5.35 (0.94)	−12.43 (0.88)	−29.50 (0.64)
	UR	−0.34 (0.94)	−0.07 (0.95)	0.17 (0.95)	0.16 (0.95)	−0.06 (0.96)	0.21 (0.95)	−0.59 (0.95)	−2.49 (0.96)	−7.80 (0.93)	−26.54 (0.78)
	IMRCC	−0.34 (0.94)	0.11 (0.95)	0.37 (0.96)	0.58 (0.95)	0.18 (0.95)	0.06 (0.95)	−1.79 (0.95)	−6.86 (0.93)	−23.76 (0.73)	−69.79 (0.08)
	RCC	−0.30 (0.94)	0.10 (0.95)	0.38 (0.96)	0.60 (0.95)	0.29 (0.95)	0.59 (0.95)	0.17 (0.95)	−1.25 (0.96)	−7.02 (0.94)	−25.62 (0.79)
TNDAMD	URAR	−0.16 (0.94)	−0.10 (0.95)	−0.19 (0.95)	−0.46 (0.95)	−1.05 (0.94)	−1.59 (0.95)	−4.43 (0.93)	−14.13 (0.85)	−29.74 (0.59)	−50.56 (0.30)
	UR	−0.16 (0.94)	−0.15 (0.96)	−0.06 (0.95)	−0.16 (0.96)	−0.21 (0.96)	−0.07 (0.96)	−0.82 (0.96)	−2.71 (0.96)	−8.46 (0.94)	−25.65 (0.85)
	IMRCC	−0.16 (0.94)	0.03 (0.95)	0.12 (0.96)	0.19 (0.96)	−0.19 (0.96)	−0.85 (0.95)	−3.40 (0.95)	−9.98 (0.91)	−23.52 (0.80)	−42.90 (0.58)
	RCC	−0.16 (0.94)	0.04 (0.95)	0.15 (0.96)	0.29 (0.96)	0.18 (0.96)	0.33 (0.96)	−0.03 (0.96)	−1.41 (0.96)	−7.68 (0.95)	−24.77 (0.86)

CHAPTER 4: THE XTRCCIPW COMMAND

4.1 Introduction

The method of generalized estimating equations (GEE) is frequently used to estimate the marginal means of a longitudinal outcome. When outcomes are missing completely at random (MCAR), a standard GEE estimator is consistent for these marginal means (Liang and Zeger, 1986; Diggle et al., 2002). When outcomes are either missing at random (MAR) or missing not at random, inverse-probability weights (IPWs) may be used to ensure consistency of the GEE estimator provided that the data missingness model is correctly specified (Robins et al., 1995; Scharfstein et al., 1999a). We refer to this approach as the IPW-GEE method.

An individual's outcomes over time form an outcome trajectory. Events such as death can truncate the trajectory, rendering the outcome at and after truncation undefined. The opposite of truncation is referred to as continuation. Death is a common truncating event in most biomedical studies (Ribaudo et al., 2000; Billingham and Abrams, 2002; Pauler et al., 2003; Dufouil et al., 2004; Shardell and Miller, 2008; Basu and Manning, 2010). For example, the Precipitating Events Project (PEP) is an ongoing longitudinal study of 754 community-living individuals aged 70 or older who were scheduled to be followed monthly for two years (Gill et al., 2001). Kurland and Heagerty (2005) considered inference about the probability of activities-of-daily-living (ADL) disability conditioning on being alive, treating death as a truncating event in the PEP data. Other events, such as disease relapse and HIV infection, have also been defined as truncating events. For instance, investigators of the Breastfeeding, Antiretrovirals, and Nutrition (BAN) study (van der Horst et al., 2009) wanted to draw inference about a target population of infants at high risk of HIV infection, but only while they were alive and uninfected (Flax et al., 2012). In this case, HIV infection and death are truncating events. In le Cessie et al. (2009), the target population consisted of patients with advanced breast cancer who had undergone chemotherapy. The authors wanted to draw inference about patients who were alive and disease-free, such that death and relapse are truncating events.

For all of the aforementioned examples of truncated longitudinal data, outcomes were also missing for some individuals. Dropout events occur when an individual leaves the study permanently. For study dropout, the corresponding outcomes are unobserved but, unlike truncation, they are well-defined. Typical approaches to analyzing longitudinal outcomes with missing data include weighted GEE (WEE) or maximum likelihood based on mixed-effects models. These approaches generally do not distinguish truncation from dropout, in essence envisaging outcomes past the point of truncation. Kurland and Heagerty (2005) described such approaches that implicitly assume the existence of outcomes after truncation as "unconditional regression" (UR) models because they estimate the mean outcome averaged over individuals who have and have not been truncated. Kurland et al. (2009) consider both standard selection models and conditional submodels of pattern-mixture models to be UR models. Mean outcomes among continuing trajectories may be estimated indirectly with these two types of UR models, with additional modeling assumptions (Kurland et al., 2009). As an alternative to UR models, joint modeling of longitudinal measurements and time to truncation might be employed (Henderson et al., 2000; Guo and Carlin, 2004; Kurland et al., 2009).

In order to estimate mean outcomes without relying on additional assumptions or joint modeling, Kurland and Heagerty (2005) developed a method for regression conditioning on continuation (RCC), i.e., not being truncated. The RCC method consistently estimates continuing longitudinal mean outcomes by first modeling and estimating IPWs at each time point based on the probability of dropout, but only for subjects with a continuing outcome at that time point. RCC then applies these IPWs in a WEE framework. In the absence of truncation, the usual WEE method is therefore a special case of RCC. When there is truncation, WEE is a UR approach that will generally not produce consistent estimates for RCC estimands (Kurland and Heagerty, 2005). Unfortunately, there is currently no widely available Stata command for estimating the IPWs used in either RCC or WEE. The `teffects` commands `aipw`, `ipw`, and `ipwra` estimate IPWs with the goal of making causal inferences by estimating average treatment effects (StataCorp, 2013). However, these `teffects` commands cannot handle longitudinal or panel

data, nor can they properly account for truncation directly. Therefore, in this paper we introduce the `xtrccipw` command to allow Stata users to estimate the IPWs used by RCC in analyzing longitudinal data subject to missingness or truncation. The user then specifies these IPWs as the weights used by the `glm` command, which then performs WEE estimation. When there is no truncation, `xtrccipw` can also be used to estimate the IPWs used by the `glm` command in a WEE analysis.

The remainder of this paper is organized as follows. In Section 4.2, we introduce some notation and the assumptions behind the RCC method, detail the logit and probit link functions for modeling the dropout mechanism, and note some large-sample properties of the RCC estimator. The `xtrccipw` command is explained in Section 4.3. We conduct RCC in Section 4.4 on both a binary outcome and a continuous outcome using an example dataset. In Section 4.5, we reanalyze the original Kurland and Heagerty (2005) data. Finally, the command is summarized in Section 4.6, and its applications and future extensions are briefly discussed.

4.2 Background/Methods

4.2.1 Notation & Assumptions

Consider a random sample of $i = 1, \ldots, n$ individuals, each of whom is scheduled to be measured at fixed study time points $j = 1, \ldots, m$. For any random variable A, let A_j denote the value of A at time point j, and let $\bar{A}_j = \left(A_1, \ldots, A_j\right)$ so that $\bar{A}_{j-1}$ represents an individual's history of A prior to time point j. Define a_j and $\bar{a}_j$ likewise for any fixed variable a. Where it is not ambiguous, the dependence on i will be suppressed for notational ease. Let Y_j denote the outcome at time point j. Let $C_j = 1$ if the truncating event has not occurred by time j, and let $C_j = 0$ otherwise. Thus, the outcome Y_j is well defined if and only if $C_j = 1$. Assume that truncation is an irreversible state such that $C_j = 0$ implies $C_{j'} = 0$ for all $j' > j$. Define $S = \sum_{j=1}^{m} C_j$ to be the number of time points before a trajectory is truncated, with $S = m$ indicating that the trajectory is not truncated. If truncation occurs at time point j, then outcomes at that time point and beyond (i.e., $Y_j, \ldots, Y_m$) are undefined. We use $*$ to denote all undefined values, which extends the support of both the outcome Y and the binary dropout event R defined below. If truncation does not occur by time point j, and if the individual drops out, then the outcome is still defined but is not observed. Next, we define the indicator variable for dropout. If $C_j = 1$, let

$R_j = 1$ if an individual has not dropped out by time point j; otherwise, let $R_j = 0$. Assume there is no dropout at time point $j = 1$ ($R_1 = 1$), and that dropout is monotonic such that $R_j = 0$ implies $R_{j'} = 0$ for all $j' > j$. If $C_j = 0$, then we adopt the convention that $R_j = *$.

Let $\pi_j^{(r)} = \Pr\left(R_j = r \middle| \bar{Y}_m, \bar{C}_m\right)$. Thus, the probability of dropping out conditional on the history of dropout, on all outcomes, and on the full truncation vector is denoted $\pi_j^{(0)}$. Assume $\pi_1^{(r)} = \Pr\left(R_1 = r \middle| C_1\right)$. We refer to outcomes as MAR if $\pi_j^{(r)} = \Pr\left(R_j = r \middle| \bar{Y}_{j-1}^{\text{obs}}, \bar{C}_j\right)$ for $j > 1$, where $\bar{Y}_j^{\text{obs}} = \left\{Y_k : R_k = 1, k \leq j\right\}$ denotes the observed values of $\bar{Y}_j$. We refer to outcomes as MCAR if $\pi_j^{(r)} = \Pr\left(R_j = r \middle| \bar{C}_j\right)$ for $j > 1$. Outcomes that are neither MAR nor MCAR are missing not at random. Under MAR, $\pi_j^{(1)} = \prod_{k=1}^{j} \lambda_k$ where $\lambda_k = \Pr\left(R_k = 1 \middle| R_{k-1} = 1, \bar{Y}_{k-1}^{\text{obs}}, \bar{C}_k\right)$ for $k > 1$ and $\lambda_1 = \pi_1^{(1)}$. The `xtrccipw` command lets the user specify a model for λ_k.

4.2.2 The Full and Reduced Dropout Models

In the presence of dropout, the RCC method requires specifying a parametric dropout model. The `xtrccipw` command allows the user to choose between two models. In particular, let $g(\cdot)$ represent the logit or probit link function. The default dropout mechanism modeled by `xtrccipw` is

$$g\left(\lambda_{ik}\right) = \alpha_{0k} + \alpha_{1k} z_{ik} + \alpha_{2k} \bar{Y}_{i(k-1)}^{\text{obs}} I\left(k > 1\right), \tag{4.10}$$

where α_{0k} is the intercept; z_{ik} represents the vector of both time-dependent and time-independent fixed covariates, with conformable parameter vector α_{1k}; and α_{2k} represents the conformable parameter vector corresponding to lagged outcome values $\bar{Y}_{i(k-1)}^{\text{obs}}$. Equation (4.10) is referred to as the full dropout model. Note that the parameters α_{0k}, α_{1k}, and α_{2k} are time point-specific; i.e., the dropout model is estimated at each time point by default. If dropout is assumed or known to happen completely at random, but truncation is present, the user has the option to specify an MCAR model instead, which sets $\alpha_{2k} = 0$.

The user may want to estimate a reduced model with fewer lags, with possible values $lag = 1, \ldots, m-1$. In this case, the dropout mechanism is instead modeled as

$$g\left(\lambda_{ik}\right) = \begin{cases} \alpha_{0k} + \alpha_{1k} z_{ik} + \alpha_{2k} L_{ik} & \text{if } k \leq lag \\ \alpha_0 + \alpha_1 z_{ik} + \alpha_2 L_{ik} & \text{if } k > lag \end{cases}, \tag{4.11}$$

where $L_{ik} = 0$ at $k = 1$, and $L_{ik} = \big(\max\{Y_{i1}, Y_{i(k-lag)}\}, \ldots, Y_{i(k-1)}\big)$ at $k > 1$. Equation (4.11) is referred to as the reduced dropout model. This model is time point-specific for time points $k \leq lag$, but shares the same parameters for time points $k > lag$. This approach allows `xtrccipw` to estimate fewer parameters by assuming a common dropout model once all of the requested lagged outcomes potentially become available for estimation (i.e., for time points $k > lag$). The user has the option to specify a reduced MCAR model instead, which estimates the model $g\big(\lambda_{ik}\big) = \alpha_0 + \alpha_1 z_{ik}$.

Note that the full and reduced MAR models are identical when $lag = m - 1$ is set, while the full and reduced MCAR models are different. The full MCAR model specifies a model at each time point, while the reduced MCAR model specifies a common model across all time points.

4.2.3 Inference

We are now prepared to draw inference on longitudinal mean outcomes for continuing individuals, conditional on covariates. Let $\mu_{ij}^{RCC} = E\big(Y_{ij}\big|C_{ij} = 1\big)$ denote the mean outcome for individual i whose trajectory is still continuing at time point j. In the regression setting, we might posit a generalized linear model of the form $h\big(\mu_{ij}^{RCC}\big) = \boldsymbol{x}_{ij}'\boldsymbol{\beta}^{RCC}$, where $h(\cdot)$ is a link function, $\boldsymbol{x}_{ij}$ is an observed $p \times 1$ vector of (possibly time-dependent) covariates that includes a column of ones for the intercept, and $\boldsymbol{\beta}^{RCC}$ is the corresponding parameter vector. Let $\boldsymbol{d}_{ij}' = \partial\mu_{ij}^{RCC}/\partial\boldsymbol{\beta}^{RCC}$ denote the Jacobian of partial derivatives of μ_{ij}^{RCC} with respect to $\boldsymbol{\beta}^{RCC}$.

Following Kurland and Heagerty (2005), consider the vector estimating equation

$$U\left(\boldsymbol{\beta}^{RCC}\right) = \sum_{i=1}^{n}\sum_{j=2}^{m} \boldsymbol{d}_{ij} C_{ij} \frac{R_{ij}}{\pi_{ij}^{(1)}} \left(Y_{ij} - \mu_{ij}^{RCC}\right). \tag{4.12}$$

We adopt the convention that if $C_{ij} = 0$, the summand for individual i at time point j equals 0 rather than being undefined. The IPW probability π_{ij} is unknown in practice, but can be consistently estimated if the dropout mechanism model is correctly specified. Let $\hat{\pi}_{ij}$ represent a consistent estimator of π_{ij}, and let $\hat{\beta}$ denote the solution to $U\big(\boldsymbol{\beta}^{RCC}\big) = \mathbf{0}$ under MAR when $\hat{\pi}_{ij}$ is substituted for π_{ij}. The estimator $\hat{\beta}$ is consistent and asymptotically multivariate normal for β_{RCC} (Robins et al., 1995). The `glm` command is ideal for calculating $\hat{\beta}$ because by default it assumes the independence working correlation structure required by RCC, and it allows the user

to specify time-varying IPWs through the `pweight` option. The empirical sandwich estimator of the variance of $\hat{\beta}$ is readily available by specifying the `glm` command option `vce(cluster` *clustvar*`)`, where *clustvar* is the variable that identifies individuals. When computed as if the IPWs are known and fixed, the empirical sandwich estimator is expected to be conservative (Robins et al., 2000; Robins, 2000). Thus, 95% Wald confidence intervals (CIs) constructed using the empirical sandwich estimator should have a coverage probability for β^{RCC} of at least 95%.

4.3 The xtrccipw command

4.3.1 Description

The `xtrccipw` command estimates time-specific weights equal to the inverse of the dropout probability conditioning on continuation. This command uses either the `logit` or `probit` commands to estimate IPWs. The user must then run `glm` while specifying the `pweight` and `vce(cluster` *clustvar*`)` options in order to calculate RCC estimates of the mean-model parameters, along with variance estimates constructed using the empirical sandwich estimator. The `xtrccipw` command runs under Stata 13.

The rest of this section is organized as follows. Input dataset requirements will be described and illustrated in an example. The command syntax will then be presented, along with definitions of all relevant variables and options. Finally, we will describe the displayed outputs and saved results, and instruct the user on subsequent inference using the `glm` command.

4.3.2 Input Datasets

The `xtrccipw` command accepts datasets in Stata long format (i.e., each row corresponds to one observation at one measurement time point). It then creates indicator variables for truncation and dropout based on the supplied variables for measurement time, truncation time, and mean-model outcome.

The dataset must include the following variables: unique individual identifiers, measurement time, measurement time index, outcome, and dropout-model covariates. Each row must provide values for unique individual identifiers, measurement time, and measurement time index. For each individual, unique individual identifier values must be identical on all rows, and rows for all possible measurement times and time indices must be included in order to create truncation and dropout indicators, regardless of outcome value being missing or not missing on any given row.

At the current time index, values for all dropout model covariates (except for past outcomes) must be provided if an individual had not dropped out by the previous time index (i.e., if an outcome value was provided at the previous time index) and had not been truncated by the current time index. The dataset must additionally include a variable for truncation time if truncation occurred for any individual, in which case truncation time must be identical on all rows for each individual with a truncation time. Truncation time must be left missing on all rows for each individual without a truncation time.

An example dataset is illustrated in Table 4.1. The variable names correspond to a unique individual identifier `idvar`, measurement time `timevar`, measurement time index `timeidxvar`, continuous outcome `outcomevar`, dropout-model time-dependent continuous covariate `dtdcovar`, dropout-model time-independent binary covariate `dticovar`, and truncation time `trtimevar`.

4.3.3 Syntax

`xtrccipw` *outcomevar* [*if*] `,` `idvar(`*varlist*`)` `timevar(`*varname*`)` `timeidxvar(`*varname*`)`

`generate(`*name*`)` [`timeidxf(`#`)` `timeidxl(`#`)` `trtimevar(`*varname*`)` `dlinkfxn(`*string*`)`

`dtdcovars(`*varlist*`)` `dticovars(`*varlist*`)` `dcar` `reducedlag(`#`)`]

outcomevar is the mean-model outcome variable that is used as a covariate in the dropout probability mechanism model. If *outcomevar* is an indicator/categorical factor variable, it must be preceded with "`i.`". The other unary operators "`c.`" and "`o.`" are not allowed.

4.3.4 Options

`idvar(`*varlist*`)` defines variables used to uniquely identify individuals (e.g., subjects, panels). This is analogous to *panelvar* in `xtset`. This is a required option.

`timevar(`*varname*`)` defines the variable representing the measurement time (e.g., visit date). This is analogous to *timevar* in `xtset`. This is a required option.

`timeidxvar(`*varname*`)` defines the variable representing the measurement time index (e.g., visit number). All index values must be integers. This is a required option.

`generate(`*name*`)` defines the variable that will contain the estimated IPW. This is a required option.

`timeidxf(`#`)` denotes the first time index value, which must be an integer, to be used in the

mean-model analysis. This must be specified along with `timeidxl()`. The default is the first non-missing index value found in the current dataset after [*if*] is applied.

`timeidxl(#)` denotes the last time index value, which must be an integer, to be used in the mean-model analysis. This must be specified along with `timeidxf()`. The default is the last non-missing index value found in the current after [*if*] is applied.

`trtimevar(`*varname*`)` denotes the truncation time (e.g., truncation date). The default is no truncation.

`dlinkfxn(`*string*`)` specifies the dropout-model binary link function, and only accepts the values `logit` or `probit`. The default is `logit`.

`dtdcovars(`*varlist*`)` defines the dropout-model time-dependent variables, in addition to the mean-model outcome variable. Use spaces to separate multiple variables. Each indicator/categorical factor variable argument in `dtdcovars()` must be preceded with "`i.`". The other unary operators "`c.`" and "`o.`" are not allowed, and neither is variable-interaction notation (i.e., "`#`" or "`##`"). The *varlist* syntax is otherwise identical to the *indepvars* syntax for the `logit` or `probit` commands. New variables representing the interactions between variables must be created and included separately. For example, suppose we have time-dependent binary variables, x and y, and the continuous variable z. If we wish to model dropout dependent on x, y, z, the interaction between x and y, and the interaction between x and z, we would first create the interaction variables; e.g., "`gen xy = x * y`" and "`gen xz = x * z`". Then we would correspondingly type something like `dtdcovars(i.x i.y i.xy z xz)`. The default is no additional time-dependent variables.

`dticovars(`*varlist*`)` defines the dropout-model time-independent variables, in addition to the mean-model outcome variable. The same description as that for `dtdcovars(`*varlist*`)` applies. The default is no additional time-independent variables.

`dcar` defines whether to use the full MCAR model. This option cannot be specified simultaneously with `reducedlag()`. The default is the full MAR model.

`reducedlag(#)` defines whether to use the reduced dropout model. The number of lags # can range from 1 to $m-1$. However, specifying $m-1$ lags is identical to specifying the full MAR model. To specify the reduced MCAR model, type `reducedlag(0)`. This option cannot be

specified simultaneously with `dcar`. The default is the full MAR model.

4.3.5 Displayed outputs

Two outputs are shown. The first is a list of all arguments for verification by the user. The second is a tabulation of the observed values of the `regerrorcode_xtrccipwRi` variable, which indicates the number of observations at each time point for which dropout regression and probability estimation were successful, or for which there were errors. It takes on the values "success", "failure 1: outcome does not vary" when there is either no dropout or all dropout at that time point, "failure 2: collinearities and other errors" (e.g., all eligible observations dropped due to regression collinearities), or "failure 3: prediction unavailable" if the regression succeeded, but estimation was unsuccessful. In any of the failure cases, the dropout probability is estimated as the empirical mean of dropout in the risk set (i.e., among observations with $R_{i(j-1)} = 1$).

4.3.6 Saved results

The command attaches five variables to the input dataset. The outcome variable used in estimating the dropout probability while accounting for truncation is stored as `xtrccipw_`*outcomevar*. The values of this variable only differ from *outcomevar* in that in cases when a truncation event and outcome are both recorded at time point j, `xtrccipw` treats truncation as having occurred before the outcome, and sets `xtrccipw_`*outcomevar* as undefined (i.e., "."). The indicators for truncation (`xtrccipwCi`=0, =1 otherwise) and dropout (`xtrccipwRi`=0, =1 otherwise) used to estimate the IPWs are also saved, as are the estimated IPWs themselves (as the variable specified by `generate()`). Finally, the `regerrorcode_xtrccipwRi` variable is also output.

4.3.7 Relationship to glm

The `xtrccipw` command calculates IPWs, but the `glm` command must still be run on the resulting dataset. However, for the resulting GEE mean-model estimates to be consistent, it is necessary when running `glm` to include two options. The option `[pweight=`*name*`]` must be included, with *name* equal to that specified in `generate(`*name*`)`. The option `vce(cluster` *idvars*`)` must also be included, with *idvars* = *varlist* as specified in `idvar(`*varlist*`)`.

4.4 Examples

Our example data came from the National Longitudinal Survey of Young Women (NLSYW). We took a subsample of an available Stata dataset for our analysis, generated truncation, and then analyzed both a binary and a continuous outcome from this analysis sample. Example code for creating the example dataset is available in the Appendix.

We started with the dataset http://www.stata-press.com/data/r13/nlswork5.dta, a subsample of 4,711 young women ages 14-26 in 1968 that was derived to illustrate how to use the `xt` commands. These data comprised "women in years when employed, not enrolled in school and evidently having completed their education, and with wages in excess of $1/hour but less than $700/hour" (StataCorp, 2011).

The longitudinal outcomes of interest were union membership `union` (=1 if yes, =0 if no) and weeks unemployed in the previous year `wks_ue`; the former is binary, the latter, continuous. The covariates we used were `age`, ln(wage/GNP deflator) `ln_wage`, total work experience `ttl_exp`, birth year `birth_yr`, and college graduate indicator `collgrad` (=1 if yes, =0 if no). The identifier variables were NLS ID `idcode` and interview `year`.

For our analysis, we selected the `nlswork5.dta` subsample of women with non-missing values for any of these outcomes or covariates from years 70 through 73, 77, 78, and 80, which gave us 357 individuals. We then generated truncation at follow-up years, indicated by `Ci=0`. No truncation was generated for baseline year 70. Truncation was generated with probability 0.2 if union membership in the previous year was missing. Otherwise, truncation was generated with higher probability if an individual was a union member in the previous year, and with lower probability if she was not a member. The degree of increase or decrease in truncation probability itself increased over time.

4.4.1 Binary Outcome

We first regressed `union` on `age`, `ln_wage`, and `birth_yr`. Dropout was modeled on `ttl_exp` and `collgrad` using a probit link. The IPW variable was generated as `xtrccipw_ipw`.

```
. xtrccipw i.union, idvar(idcode) timevar(year) timeidxvar(yearidx)
> generate(xtrccipw_ipw) trtimevar(truncyear) dlinkfxn(probit) dtdcovars(ttl_exp)
> dticovars(i.collgrad)
```

The `xtrccipw` arguments were output to the Stata log screen for verification. Here, `timeidxf` and `timeidxl` took on values derived from the dataset because they were not specified. The dropout event regression result for each month can also be quickly scanned for errors by the automatic tabulation of errors via the `regerrorcode_xtrccipwRi` variable.

```
outcomevar = i.union
idvar = idcode
timevar = year
timeidxvar = yearidx
generate = xtrccipw_ipw
timeidxf = 1
timeidxl = 7
trtimevar = truncyear
dlinkfxn = probit
dtdcovars = ttl_exp
dticovars = i.collgrad
dcar =
reducedlag =
```

interview year	regerrorcode_xtrccipwRi success	failure 1	failure 3	Total
1	357	0	0	357
2	155	0	0	155
3	106	0	0	106
4	78	0	0	78
5	0	55	0	55
6	43	0	7	50
7	36	0	7	43
Total	775	55	14	844

At this point, the IPW `xtrccipw_ipw` has been calculated and attached to the input dataset. We then ran the main RCC regression for union membership using `glm`, which by default assumes an independence working correlation structure as required by RCC. The probability of being a union member was modeled using a logit link.

```
. glm union age ln_wage birth_yr [pweight=xtrccipw_ipw], family(binomial)
> vce(cluster id)

Iteration 0:   log pseudolikelihood = -684.98061
```

```
Iteration 1:   log pseudolikelihood = -677.41818
Iteration 2:   log pseudolikelihood =  -677.3618
Iteration 3:   log pseudolikelihood =  -677.3618

Generalized linear models                         No. of obs      =       622
Optimization     : ML                             Residual df     =       618
                                                  Scale parameter =         1
Deviance         =  1354.723594                   (1/df) Deviance =  2.192109
Pearson          =  1616.083078                   (1/df) Pearson  =  2.615021

Variance function: V(u) = u*(1-u/1)               [Binomial]
Link function    : g(u) = ln(u/(1-u))             [Logit]

                                                  AIC             =  2.190874
Log pseudolikelihood = -677.3617969               BIC             = -2620.833

                              (Std. Err. adjusted for 205 clusters in idcode)
------------------------------------------------------------------------------
             |               Robust
       union |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |  -.1633892   .0508454    -3.21   0.001    -.2630443    -.063734
     ln_wage |   1.156356   .3594872     3.22   0.001     .4517743    1.860938
    birth_yr |  -.0577466   .0794246    -0.73   0.467     -.213416    .0979228
       _cons |   3.250271   4.481489     0.73   0.468    -5.533287    12.03383
------------------------------------------------------------------------------
```

Note that while there were 844 IPW values calculated, only 622 were used by `glm`. This is because `xtrccipw` estimates dropout probabilities for all time points in the set of observations at risk for dropout; i.e., every time point with a non-missing outcome at the previous time point. Hence, this risk set includes the subset of time points immediately after an individual's last outcome is observed, and at which outcome is missing. Because `glm` only uses complete cases (i.e., non-missing outcomes), this subset is excluded from its analysis. In the example above, there were 222 time points in this subset.

Excluding `trtimevar(truncyear)` from the `xtrccipw` call resulted in truncation being treated exactly like dropout, with the following dropout regression error and UR `glm` results.

```
interview | regerrorcode_xtrccipwRi
     year |   success  failure 3 |     Total
----------+----------------------+----------
          |                      |
```

1	357	0	357
2	205	0	205
3	116	0	116
4	101	0	101
5	2	57	59
6	43	12	55
7	36	9	45
Total	860	78	938

```
Iteration 0:   log pseudolikelihood = -940.69266
Iteration 1:   log pseudolikelihood = -924.46856
Iteration 2:   log pseudolikelihood =  -924.2712
Iteration 3:   log pseudolikelihood = -924.27109
Iteration 4:   log pseudolikelihood = -924.27109

Generalized linear models                         No. of obs      =       622
Optimization     : ML                             Residual df     =       618
                                                  Scale parameter =         1
Deviance         =  1848.542185                   (1/df) Deviance =  2.991169
Pearson          =  2414.920004                   (1/df) Pearson  =  3.907638

Variance function: V(u) = u*(1-u/1)               [Binomial]
Link function    : g(u) = ln(u/(1-u))             [Logit]

                                                  AIC             =  2.984795
Log pseudolikelihood = -924.2710925               BIC             = -2127.015

                              (Std. Err. adjusted for 205 clusters in idcode)
```

union	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
age	-.2074534	.0569579	-3.64	0.000	-.3190888	-.095818
ln_wage	1.228034	.3972456	3.09	0.002	.4494474	2.006622
birth_yr	-.0845777	.0845946	-1.00	0.317	-.2503801	.0812247
_cons	5.516462	4.764426	1.16	0.247	-3.821642	14.85457

Compared to their RCC counterparts, the UR parameter estimates kept the same signs, and did not change much in magnitude. Levels of statistical significance also resembled those under RCC.

The full and reduced MCAR models produced different output, as illustrated below. The following is the output for the corresponding RCC full MCAR model.

```
outcomevar = i.union
idvar = idcode
timevar = year
timeidxvar = yearidx
generate = xtrccipw_ipw
timeidxf = 1
timeidxl = 7
trtimevar = truncyear
dlinkfxn = probit
dtdcovars = ttl_exp
dticovars = i.collgrad
dcar = dcar
reducedlag =
```

interview year	regerrorcode_xtrccipwRi success	failure 1	failure 3	Total
1	357	0	0	357
2	155	0	0	155
3	106	0	0	106
4	78	0	0	78
5	0	55	0	55
6	47	0	3	50
7	40	0	3	43
Total	783	55	6	844

```
Iteration 0:   log pseudolikelihood = -694.10882
Iteration 1:   log pseudolikelihood = -687.47893
Iteration 2:   log pseudolikelihood = -687.43162
Iteration 3:   log pseudolikelihood = -687.43162

Generalized linear models                         No. of obs      =       622
Optimization     : ML                             Residual df     =       618
                                                  Scale parameter =         1
Deviance         =  1374.863241                   (1/df) Deviance =  2.224698
Pearson          =  1574.215399                   (1/df) Pearson  =  2.547274

Variance function: V(u) = u*(1-u/1)               [Binomial]
Link function    : g(u) = ln(u/(1-u))             [Logit]

                                                  AIC             =  2.223253
Log pseudolikelihood = -687.4316203               BIC             = -2600.694

                              (Std. Err. adjusted for 205 clusters in idcode)
------------------------------------------------------------------------------
             |
```

union	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
age	-.1588589	.0498965	-3.18	0.001	-.2566542	-.0610637
ln_wage	1.158728	.3630167	3.19	0.001	.4472285	1.870228
birth_yr	-.0516636	.0795114	-0.65	0.516	-.2075031	.104176
_cons	2.909992	4.464318	0.65	0.515	-5.83991	11.65989

And here is the output for the corresponding RCC reduced MCAR model for comparison.

```
outcomevar = i.union
idvar = idcode
timevar = year
timeidxvar = yearidx
generate = xtrccipw_ipw
timeidxf = 1
timeidxl = 7
trtimevar = truncyear
dlinkfxn = probit
dtdcovars = ttl_exp
dticovars = i.collgrad
dcar =
reducedlag = 0
```

interview year	regerrorcode_xtrccipwRi success	Total
1	357	357
2	155	155
3	106	106
4	78	78
5	55	55
6	50	50
7	43	43
Total	844	844

```
Iteration 0:   log pseudolikelihood = -738.57945
Iteration 1:   log pseudolikelihood = -730.94243
Iteration 2:   log pseudolikelihood = -730.88735
Iteration 3:   log pseudolikelihood = -730.88735

Generalized linear models                         No. of obs      =       622
```

```
Optimization     : ML                                Residual df     =       618
                                                     Scale parameter =         1
Deviance         =  1461.774699                      (1/df) Deviance =  2.365331
Pearson          =  1706.799805                      (1/df) Pearson  =  2.761812

Variance function: V(u) = u*(1-u/1)                  [Binomial]
Link function    : g(u) = ln(u/(1-u))                [Logit]

                                                     AIC             =  2.362982
Log pseudolikelihood = -730.8873493                  BIC             = -2513.782

                              (Std. Err. adjusted for 205 clusters in idcode)
```

union	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
age	-.1647539	.0514639	-3.20	0.001	-.2656214	-.0638865
ln_wage	1.115535	.3651293	3.06	0.002	.3998944	1.831175
birth_yr	-.0488252	.0818688	-0.60	0.551	-.2092851	.1116348
_cons	3.01862	4.602351	0.66	0.512	-6.001822	12.03906

4.4.2 Continuous Outcome

We then regressed `wks_ue` on `age`, `ln_wage`, and `birth_yr`. Dropout was again modeled on `ttl_exp` and `collgrad` using a probit link, with the IPW variable generated as `xtrccipw_ipw`. The dropout regression error and RCC `glm` results follow below.

```
. xtrccipw wks_ue, idvar(idcode) timevar(year) timeidxvar(yearidx)
> generate(xtrccipw_ipw) trtimevar(truncyear) dlinkfxn(probit) dtdcovars(ttl_exp)
> dticovars(i.collgrad)
```

(*output omitted*)

interview year	regerrorcode_xtrccipwRi success	failure 1	failure 3	Total
1	332	0	25	357
2	256	0	20	276
3	175	0	85	260
4	171	0	62	233
5	0	228	0	228
6	223	0	0	223
7	215	0	0	215
Total	1,372	228	192	1,792

```
. glm wks_ue age ln_wage birth_yr [pweight=xtrccipw_ipw], vce(cluster id)

Iteration 0:   log pseudolikelihood = -6222.7092

Generalized linear models                         No. of obs      =      1607
Optimization     : ML                             Residual df     =      1603
                                                  Scale parameter =  45.98446
Deviance         =  73713.08961                   (1/df) Deviance =  45.98446
Pearson          =  73713.08961                   (1/df) Pearson  =  45.98446

Variance function: V(u) = 1                       [Gaussian]
Link function    : g(u) = u                       [Identity]

                                                  AIC             =  7.749482
Log pseudolikelihood = -6222.709174               BIC             =  61879.54

                              (Std. Err. adjusted for 347 clusters in idcode)
------------------------------------------------------------------------------
             |               Robust
     wks_ue |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
------------+-----------------------------------------------------------------
        age |    .086099   .0634442     1.36   0.175    -.0382493    .2104472
    ln_wage |   -.5719055   1.089603    -0.52   0.600    -2.707488    1.563677
   birth_yr |    .3239657   .1216133     2.66   0.008     .0856081    .5623233
      _cons |   -14.54912   7.865388    -1.85   0.064      -29.965    .8667553
------------------------------------------------------------------------------
```

Excluding `trtimevar(truncyear)` from the `xtrccipw` call resulted in truncation being treated exactly like dropout, with the following dropout regression error and UR `glm` results.

interview year	regerrorcode_xtrccipwRi success	failure 3	Total
1	332	25	357
2	347	0	347
3	250	20	270
4	256	0	256
5	106	126	232
6	228	0	228
7	217	0	217
Total	1,736	171	1,907

⋮

```
Iteration 0:   log pseudolikelihood = -8323.6825

Generalized linear models                         No. of obs      =      1607
Optimization     : ML                             Residual df     =      1603
                                                  Scale parameter =  62.08307
Deviance         =  99519.16617                   (1/df) Deviance =  62.08307
Pearson          =  99519.16617                   (1/df) Pearson  =  62.08307

Variance function: V(u) = 1                       [Gaussian]
Link function    : g(u) = u                       [Identity]

                                                  AIC             =  10.36426
Log pseudolikelihood = -8323.682502               BIC             =  87685.62

                              (Std. Err. adjusted for 347 clusters in idcode)
```

wks_ue	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
age	.0836737	.0669348	1.25	0.211	-.0475162	.2148636
ln_wage	-.1151565	1.260023	-0.09	0.927	-2.584757	2.354444
birth_yr	.3092506	.1330989	2.32	0.020	.0483816	.5701197
_cons	-14.6233	8.829574	-1.66	0.098	-31.92895	2.682348

As in the regressions of the binary outcome `union`, UR parameter estimates kept the same signs as the corresponding RCC estimates. However, `birth_yr` was less than half as statistically significant under UR as it was under RCC. Also, while not statistically significant in either scenario, the magnitude of `ln_wage` changed notably.

4.5 PEP Data Analysis

We now reanalyze the Kurland and Heagerty (2005) analysis data from the PEP study. Few individuals dropped out (n = 17, 2.3%), and only 62 (8.2%) died in the first two years of the study. Out of 432 low-risk individuals, 30 died (7%); 14 of 213 (6%) died in the medium-risk group; and 18 of 107 (17%) died in the high-risk group. They estimated the association of ADL disability with ADL-disability risk group (i.e., risk levels low, medium, and high), month, month^2, and the interaction between month and risk group. Their dropout model included all of these covariates in addition to sex, ADL-disability status at the previous month to reflect the MAR assumption, and a baseline depression indicator.

The `xtrccipw` and `glm` commands were called as follows, with the relevant output displayed. The variables were study ID (`studyid`), month (`month`), month index (`monthidx`), ADL disability (`adldis` = 1 if disabled; = 0 otherwise), risk group (`rgamed` = 0, `rgahigh` = 0 for low; `rgamed` = 1, `rgahigh` = 0 for medium; and `rgamed` = 0, `rgahigh` = 1 for high), month^2 (`monthsq`), medium risk interaction with month (`rgamedmonth` = `rgamed` * `month`), high risk interaction with month (`rgahighmonth` = `rgahigh` * `month`), and ADL disability status at the previous month (`reducedlag` = 1). The dropout mechanism was modeled using a logit model.

```
. xtrccipw i.adldis, idvar(studyid) timevar(month) timeidxvar(monthidx)
> generate(xtrccipw_ipw) trtimevar(deathmo) dtdcovars(month monthsq rgamedmonth
> rgahighmonth) dticovars(i.rgamed i.rgahigh i.sex i.depresbl) reducedlag(1)
outcomevar = i.adldis
idvar = studyid
timevar = month
timeidxvar = monthidx
generate = xtrccipw_ipw
timeidxf = 1
timeidxl = 24
trtimevar = deathmo
dlinkfxn = logit
dtdcovars = month monthsq rgamedmonth rgahighmonth
dticovars = i.rgamed i.rgahigh i.sex i.depresbl
dcar =
reducedlag = 1
```

	regerrorcode_xtrccipwRi		
monthidx	success	failure 1	Total
1	0	752	752
2	750	0	750
3	748	0	748
4	743	0	743
5	742	0	742
6	740	0	740
7	735	0	735
8	731	0	731
9	730	0	730
10	729	0	729
11	727	0	727
12	721	0	721

13	715	0	715
14	712	0	712
15	710	0	710
16	706	0	706
17	701	0	701
18	700	0	700
19	696	0	696
20	690	0	690
21	686	0	686
22	681	0	681
23	677	0	677
24	674	0	674
Total	16,444	752	17,196

```
. glm adldis month monthsq rgamedmonth rgahighmonth i.rgamed i.rgahigh
> [pweight=xtrccipw_ipw], family(binomial) vce(cluster studyid)

Iteration 0:   log pseudolikelihood = -4805.9074
Iteration 1:   log pseudolikelihood = -4456.9226
Iteration 2:   log pseudolikelihood = -4448.8392
Iteration 3:   log pseudolikelihood = -4448.7424
Iteration 4:   log pseudolikelihood = -4448.7424

Generalized linear models                         No. of obs      =     17177
Optimization     : ML                             Residual df     =     17170
                                                  Scale parameter =         1
Deviance         =  8897.484773                   (1/df) Deviance =  .5181995
Pearson          =  17402.11245                   (1/df) Pearson  =  1.013518

Variance function: V(u) = u*(1-u/1)               [Binomial]
Link function    : g(u) = ln(u/(1-u))             [Logit]

                                                  AIC             =  .5188033
Log pseudolikelihood = -4448.742386               BIC             = -158532.8

                              (Std. Err. adjusted for 752 clusters in studyid)
------------------------------------------------------------------------------
             |               Robust
     adldis |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      month |    .042531   .0136743     3.11   0.002     .0157298    .0693322
    monthsq |  -.0023904   .0007797    -3.07   0.002    -.0039185   -.0008622
rgamedmonth |   .0007953   .0159911     0.05   0.960    -.0305466    .0321372
rgahighmonth|   .0239548   .0186385     1.29   0.199     -.012576    .0604855
   1.rgamed |   1.869464   .2275534     8.22   0.000     1.423468     2.31546
```

1.rgahigh	2.186206	.2463283	8.88	0.000	1.703412	2.669001
_cons	-3.532125	.1850643	-19.09	0.000	-3.894844	-3.169405

These estimates were used to produce Figure 4.4. The predicted trajectories match the fitted curves for the IPCW-IEE estimator in Figure 3 of Kurland and Heagerty (2005). The fitted odds ratio comparing odds of disability in the high-risk group to that of the low-risk group at the last time point is 8.90, while Kurland and Heagerty (2005) estimated this odds ratio as 8.95. This minor difference likely results from our use of 752 individuals in the data we were provided (from Professor Kurland), compared to 754 individuals used by Kurland and Heagerty (2005).

4.6 Discussion

In this paper, we introduced the `xtrccipw` command to estimate the inverse-probability weights used to conduct weighted GEE regression, and in particular, regression conditioning on continuation. The assumed dropout probability mechanism can be specified using either a logit or probit link function. Large-sample properties of the resulting `glm` mean and empirical sandwich variance estimates were also noted, and `xtrccipw` was demonstrated using examples with binary and continuous outcomes. Finally, the command was used to reanalyze the original study findings in Kurland and Heagerty (2005). Note that the `xtrccipw` command produces IPWs that all equal 1 when there is only truncation but no dropout, so that running `glm` afterward is equivalent to unweighted GEE regression.

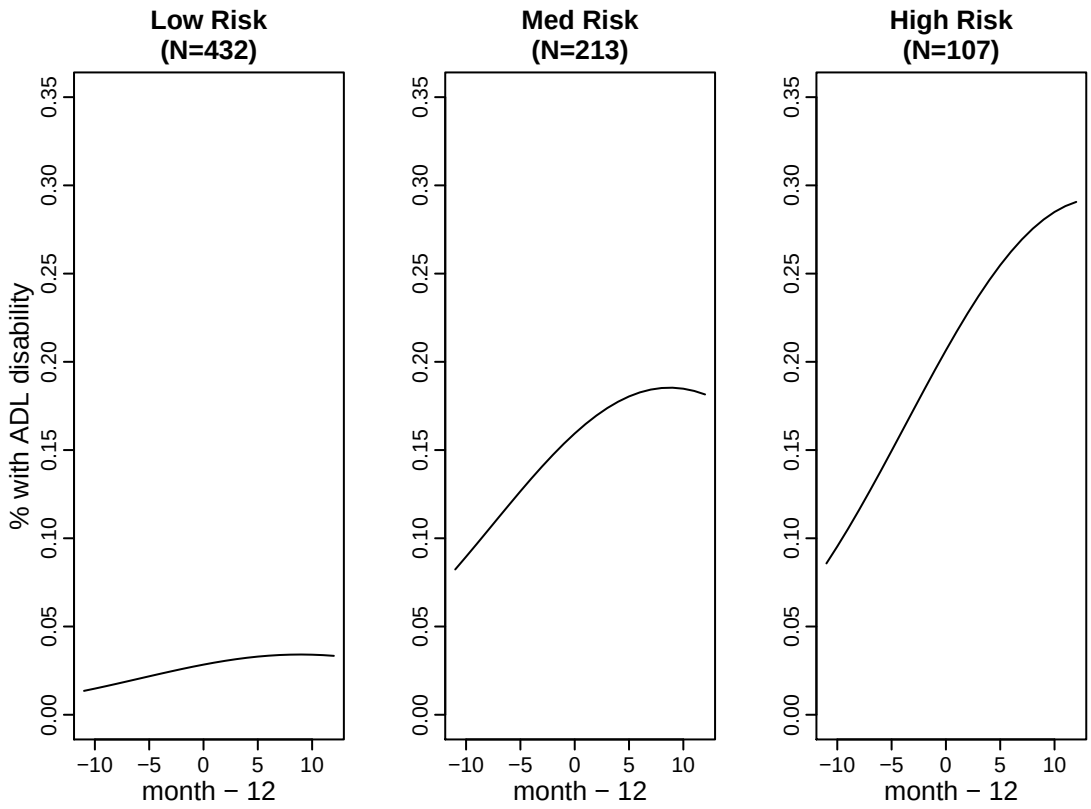
The `xtrccipw` command does have some limitations. The command can only estimate IPWs if missingness is monotonic, while many studies suffer from non-monotonic (i.e., arbitrary or intermittent) missingness. To use `xtrccipw`, an "artificial" dropout indicator that treats the first instance of missingness as dropout may be constructed, discarding any subsequent non-missing outcomes (Robins et al., 1995). One can also try imputing arbitrarily missing outcomes up until the last non-missing outcome, as done in Kurland and Heagerty (2005).

The RCC method is appropriate when one wishes to draw inference about a target population or real-world population that is itself subject to truncation, when one is interested only in the subset of the target population consisting of continuing outcomes. In the PEP study of individuals aged 70 or older, the mortality pattern seen in the study sample population can be seen

to represent that of the target population, so that up-weighting outcomes for those who had died might not be reasonable because these up-weighted outcomes may be interpreted as representing outcomes that do not exist in the target population. Regardless of the debate surrounding this interpretation of the effect of using IPWs (Chaix et al., 2012; Tchetgen Tchetgen et al., 2012), the PEP study investigators were only interested in the target population of living individuals. This is the type of research question that RCC is specifically designed to handle. The `xtrccipw` command gives the user readily available software to run a weighted GEE or RCC analysis without having to write her own code to construct IPWs.

Supporting Table 4.1: Short example dataset.

idvar	timevar	timeidxvar	outcomevar	dtdcovar	dticovar	trtimevar
1	05apr1979	1	13.2	432	yes	.
1	04may1979	2	14.3	65	yes	.
1	05jun1979	3	08.0	-5	yes	.
2	18sep1982	1	24.1	83	no	.
2	20oct1982	2	32.9	23	no	.
2	21nov1982	3	.	633	no	.
3	15sep1983	1	25.8	441	no	16nov1983
3	19oct1983	2	23.3	76	no	16nov1983
3	16nov1983	3	.	.	.	16nov1983
4	14jan1979	1	15.0	-23	no	24feb1979
4	14feb1979	2	.	455	no	24feb1979
4	16mar1979	3	.	.	.	24feb1979



Supporting Figure 4.4: Predicted trajectories for PEP data by risk group.

CHAPTER 5: AUGMENTED REGRESSION CONDITIONING ON CONTINUATION

5.1 Introduction

Previously, we introduced the Breastfeeding, Antiretrovirals, and Nutrition (BAN) study (van der Horst et al., 2009; Chasela et al., 2010), a clinical trial of 2369 mothers infected with human immunodeficiency virus (HIV) who (along with their infants) were randomized in a 3×2 factorial design of interventions. For one of the study objectives, investigators wanted to draw inference about mean infant weight, length, and body mass index (BMI) over time, as long as infants were alive and not infected with HIV. Death and infection were therefore conceptualized as "truncating" the temporal trajectory of such outcomes, rendering such outcomes undefined past the point of truncation; the opposite of truncation was called "continuation". By the end of the 24-week study period, 14% of all 2238 analysis-sample infants had dropped out, and 8% of all other infants had truncated trajectories. While infants were alive, uninfected, and still in the study, infants missed 5% of their scheduled visits intermittently. To address the difference between outcomes rendered undefined due to truncation, and outcomes that were missing but well-defined, Kurland and Heagerty (2005) developed a method based on weighted generalized estimating equations (GEE) for directly estimating continuing longitudinal mean outcomes. We extended their method to be able to accommodate different mechanisms behind dropout (i.e., monotonic missingness) and intermittent missingness (IM) using an approach similar to that of Shardell and Miller (2008), added some large-sample results to perform inference, and called this approach regression conditioning on continuation (RCC). This was achieved by specifying different probability models for the dropout and IM processes in constructing the inverse-probability weights (IPWs) used in a weighted-GEE analysis (IPWGEE). However, a fundamental weakness of most non-sampling IPWGEE methods is that while the missing-data model is unknown in practice, it must nonetheless be correctly specified to obtain consistent mean-outcome estimates (Robins et al., 1995). Simulations have shown that weighted GEE with a misspecified dropout model can

even perform worse than unweighted GEE (Preisser et al., 2002). RCC therefore performs consistent parameter estimation in general only if the models for dropout and IM are correct.

To protect against such missingness-model misspecification, methods that adjust IPW-equipped estimating equations have been widely developed. Significant progress has been made in both the missing-data and causal-inference literatures (Robins et al., 1995; Rotnitzky et al., 1998; Scharfstein et al., 1999a; van der Laan and Robins, 2003; Bang and Robins, 2005; Kang and Schafer, 2007; Wooldridge, 2007) on the development and use of these augmented estimating equations (AEE). These approaches entail adding a term with mean 0 to each summand in the estimating equation; i.e., this extra component is said to "augment" IPWGEE. The augmentation component is derived by projecting the estimating equations onto the tangent space of the missingness model, which is a Hilbert space spanned by functions of the missingness model with mean 0, and then taking the orthogonal complement of that projection (Robins and Rotnitzky, 1995; van der Laan and Robins, 2003; Carpenter and Kenward, 2006; Tsiatis, 2006; Chen and Zhou, 2011; Shardell et al., 2015). This component is often a function of the complete data (Bang and Robins, 2005). For the resulting estimator to be consistent, either the IPW or complete-data model must be correctly specified, but not necessarily both; i.e., such an estimator is "doubly robust" to model misspecification (Scharfstein et al., 1999b; Bang and Robins, 2005; Kang and Schafer, 2007; Vansteelandt et al., 2007).

Recently, the AEE approach has been applied in longitudinal settings that distinguish among truncation, dropout, or IM. Similar to work done by Shardell and Miller (2008), Chen and Zhou (2011) developed a method that allows IM in both outcomes and covariates. While they dealt with binary outcomes, they defined a framework general enough to accommodate continuous outcomes. Tchetgen Tchetgen et al. (2012) proposed a principal-stratification approach to draw causal inference from longitudinal data with a continuous outcome subject to truncation. Shardell et al. (2015) subsequently extended their work by augmenting the IPW-equipped estimating equations used to estimate the g-formulas used in the original approach, and allowed for distinct dropout and truncation mechanisms to be specified in the construction of the IPWs.

Nevertheless, these methods do not allow the AEE IPWs to be modeled with distinct mechanisms for dropout and IM, while also accounting for truncation. The method developed by Chen

and Zhou (2011) accommodates both outcome and covariate IM. However, they do not distinguish dropout from IM, and their method does not handle truncation. The approach of Shardell et al. (2015) does handle truncation and dropout, but does not allow for IM. The RCC method properly adjusts for truncation, and additionally distinguishes dropout from standard definitions of IM by instead defining IM as being conditional on non-dropout. That is, most IPW methods that allow for arbitrary or non-monotonic missingness define any such missingness as IM, while RCC restricts the definition of IM to missingness that occurs only before dropout.

In this article, we propose to extend the IPW-based RCC approach by augmenting its weighted estimating equations with expected outcomes conditional on continuation and observed data. We call this the augmented RCC (ARCC) method. In Section 5.2, we introduce notation and key assumptions, construct the augmentation component, and present the ARCC estimating equation. In Section 5.3, the empirical bias and variance of the ARCC and RCC estimators are characterized in a simulation study. The BAN data are then analyzed in Section 5.4 using ARCC. We conclude with a brief discussion in Section 5.5.

5.2 Methods

5.2.1 Notation & Assumptions

Suppose we have a random sample of $i = 1, \ldots, n$ subjects measured at up to $j = 1, \ldots, m$ fixed study time points. The dependence on i will be suppressed for notational ease when not ambiguous. Let Y_j denote the outcome at time point j. Let $C_j = 1$ if the truncating event has not occurred by time point j, and let $C_j = 0$ otherwise. Assume the outcome Y_j is well defined if and only if $C_j = 1$. Assume that truncation is irreversible such that $C_j = 0$ implies $C_k = 0$ for all $k > j$. If $C_j = 1$, let $R_j = 1$ if the outcome is observed at time point j; otherwise, let $R_j = 0$. If $C_j = 1$, let $R_j^D = 1$ if an individual has not dropped out by time point j; otherwise, let $R_j^D = 0$. Dropout is defined to occur by time point j if all non-truncated outcomes are missing at and beyond time point j; i.e., $R_k = 0$ for all $k \geq j$ if $C_k = 1$. All undefined values are denoted as $*$, and we adopt the convention that $Y_j = *$, $R_j = *$, and $R_j^D = *$ if $C_j = 0$. For any time-indexed quantity A, let A_j denote the value of A at time point j, and let $\bar{A}_j = \big(A_1, \ldots, A_j\big)$. Hence, $\bar{A}_{j-1}$ represents an individual's history of A prior to time point j, where $\bar{A}_{j-1} = \emptyset$ at $j = 1$. For any random variable A, if A is discrete then let $p(a)$ denote $\Pr(A = a)$, the mass of A at a. Likewise,

if A is continuous then let $p(a)$ denote $f(a)$, the density of A at a. Let $p(\cdot|a)$ denote $p(\cdot|A=a)$. Let $S=\sum_{j=1}^{m} C_j$ denote the number of time points before a trajectory was truncated, with $S=m$ indicating that the trajectory was not truncated. In the regression setting, we might posit a generalized linear model of the form $h\big(\mu_j(s)\big)=\boldsymbol{x}_j'\boldsymbol{\beta}_s$, where $\mu_j(s)=E\big(Y_j\big|S=s\big)$, $h(\cdot)$ is a link function, $\boldsymbol{x}_j$ is an observed $p\times 1$ vector of (possibly time-dependent) covariates that includes a column of ones for the intercept, and $\boldsymbol{\beta}_s$ is the corresponding parameter vector.

5.2.2 Dropout & Intermittent Missingness Mechanisms

In this section, we describe different assumptions regarding the probabilities of dropout and IM. The following conditional probabilities are characterized in order to make assumptions about dropout and IM. Conditioning on $\bar{\boldsymbol{x}}_m$ is assumed in all expressions, so this notation is suppressed in this section. Let

$$\lambda_j^D(c_j)=\Pr\big(R_j^D=1\big|\bar{r}_{j-1},R_{j-1}^D=1,\bar{y}_{j-1}^{\text{obs}},c_j\big),$$

where $\bar{y}_j^{\text{obs}}=\big\{y_k:R_k=1,k\le j\big\}$ denotes the observed values of $\bar{y}_j$. Let

$$\lambda_j^{IM}(c_{j+1},c_j)=\Pr\big(R_j=1\big|R_j^D=1,\bar{r}_{j-1},\bar{y}_{j-1}^{\text{obs}},c_{j+1},c_j\big),$$

where $\big\{c_{j+1},c_j\big\}=\big\{c_j\big\}$ at $j=m$. The probability of dropping out conditional on the history of missingness, on past non-dropout, on the history of observed outcomes, and on current truncation status is $1-\lambda_j^D(c_j)$. For $j\le m$ the probability of IM conditional on current non-dropout, on the histories of missingness and observed outcomes, on truncation status at the next time point if $j<m$, and on current truncation status is $1-\lambda_j^{IM}(c_{j+1},c_j)$.

Analogous assumptions regarding the dropout and IM mechanisms are now defined. If

$$\Pr\big(R_j^D=1\big|\bar{r}_{j-1},R_{j-1}^D=1,\bar{y}_m,\bar{c}_m\big)=\lambda_j^D(c_j),$$

then dropout is said to be at random (DAR). If

$$\Pr\big(R_j^D=1\big|\bar{r}_{j-1},R_{j-1}^D=1,\bar{y}_m,\bar{c}_m\big)=\Pr\big(R_j^D=1\big|\bar{r}_{j-1},R_{j-1}^D=1,c_j\big),$$

then dropout is said to be completely at random (DCAR). Likewise, if

$$\Pr\left(R_j = 1 \middle| R_j^D = 1, \bar{r}_{j-1}, \bar{y}_m, \bar{c}_m\right) = \lambda_j^{IM}(c_{j+1}, c_j),$$

then IM is said to be at random (IMAR), and if

$$\Pr\left(R_j = 1 \middle| R_j^D = 1, \bar{r}_{j-1}, \bar{y}_m, \bar{c}_m\right) = \Pr\left(R_j = 1 \middle| R_j^D = 1, \bar{r}_{j-1}, c_{j+1}, c_j\right),$$

then IM is said to be completely at random (IMCAR). Dropout is not at random if it is neither DAR nor DCAR, and IM is not at random if it is neither IMAR nor IMCAR. It can be shown that these assumptions imply that outcomes are missing at random (MAR), defined as

$$p\left(\bar{r}_m \middle| \bar{y}_m, \bar{c}_m\right) = p\left(\bar{r}_m \middle| \bar{y}_m^{\text{obs}}, \bar{c}_m\right).$$

In particular, if DAR and IMAR are true for a set of outcomes, then these outcomes are MAR.

One approach to missingness can misclassify dropout and IM. Suppose that all missingness for continuing outcomes is defined as IM, an approach akin to that of Shardell and Miller (2008) with respect to the outcomes. We denote the corresponding IPWs that consider all missing and truncated outcomes to be IM as IMRCC-IPWs, and let

$$\pi_j^{IM\dagger} = \Pr\left(R_j = 1 \middle| \bar{r}_{j-1}, \bar{y}_{j-1}^{\text{obs}}, C_j = 1\right).$$

This quantity will be used in Section 5.2.4 to construct the IMRCC-IPW-based estimators.

5.2.3 Augmentation Component

The functional component used to augment the standard IPWGEE expression is described in this section. Let $\bar{Y}_{m(-j)}^{\text{obs}} = \left\{Y_k : Y_k \in \bar{Y}_m^{\text{obs}}, k \neq j\right\}$, and define $\bar{Y}_{m(-j)}^{\text{mis}}$ likewise. In our data setting, it can be shown that the augmentation component of Chen and Zhou (2011) reduces to the expected outcome at time point j conditional on observed outcomes at any other time point, and on current continuation. Elsewhere in the literature, when similar augmentation components have been modeled and estimated, the corresponding estimator has been called an

outcome-regression (OR) estimator (Bang and Robins, 2005; Vansteelandt et al., 2010; Shardell et al., 2015). Adopting this terminology here, we let

$$\mu_j^{\mathrm{OR}}(s) = E\left(Y_j \middle| \bar{y}_{m(-j)}^{\mathrm{obs}}, \bar{\boldsymbol{x}}_m, \bar{c}_m\right) = E\left(Y_j \middle| \bar{y}_{m(-j)}^{\mathrm{obs}}, \bar{\boldsymbol{x}}_m, S = s\right) \tag{5.13}$$

denote the augmentation component with a functional form defined using OR models.

To solve the estimating equations in Section 5.2.4, $\mu_j^{\mathrm{OR}}(s)$ must be estimable. Expression (5.13) suggests taking a pattern-mixture approach that stratifies by truncation time, similar to what was done in le Cessie et al. (2009). Let $p_{s_k} = \Pr(S = k|\bar{\boldsymbol{x}}_m)$, and let $\mu_j(s) = E\big(Y_j\big|\bar{\boldsymbol{x}}_m, S = s\big)$ represent the mean outcome conditional on $S = s$ for $s > 0$. The RCC estimand of interest is

$$\mu_j^{RCC} = E\left(Y_j \middle| \bar{\boldsymbol{x}}_m, C_j = 1\right) = \frac{\sum_{k=j}^{m} \mu_j(k) p_{s_k}}{\sum_{\ell=j}^{m} p_{s_\ell}}. \tag{5.14}$$

Specifying the distribution $f\big(\bar{y}_m\big|\bar{\boldsymbol{x}}_m, S = s\big)$ would allow straightforward calculation of $\mu_j^{\mathrm{OR}}(s)$ and $\mu_j(s)$. We henceforth refer to a model for $f\big(\bar{y}_m\big|\bar{\boldsymbol{x}}_m, S = s\big)$ as an OR model.

In pattern-mixture applications, the multivariate-normal distribution is commonly used to specify an OR model for mean continuous outcomes with an identity link. The components of (5.13) are specified using a multivariate-normal distribution as follows. Let $[A]$ represent the probability distribution of a random variable A. Let $q_j = \sum_{k=1}^{m} r_k I(k \neq j)$, where $I(a) = 1$ if statement a is true and $I(a) = 0$ otherwise. The density $f\big(\bar{y}_m\big|\bar{\boldsymbol{x}}_m, S = s\big)$ is defined using $\big[\bar{Y}_m\big|\bar{\boldsymbol{x}}_m, S = s\big] = N\big(\bar{\mu}_m(s), \boldsymbol{\Sigma}\big)$, where $\bar{\mu}_m(s) = E\big(\bar{Y}_m\big|\bar{\boldsymbol{x}}_m, S = s\big)$ denotes the mean outcome vector conditioned on s, and $\boldsymbol{\Sigma}$ is an $m \times m$ covariance matrix. Hence, $\big[\bar{Y}_{m(-j)}^{\mathrm{obs}}\big|S = s\big] = N\big(\bar{\mu}(s)_{m(-j)}^{\mathrm{obs}}, \boldsymbol{\Sigma}_{q_j}\big)$ where $\bar{\mu}(s)_{m(-j)}^{\mathrm{obs}} = E\big(\bar{Y}_{m(-j)}^{\mathrm{obs}}\big|\bar{\boldsymbol{x}}_m, S = s\big)$ is a $q_j \times 1$ sub-vector of $\bar{\mu}_m(s)$, and $\boldsymbol{\Sigma}_{q_j}$ is the corresponding $q_j \times q_j$ sub-matrix of $\boldsymbol{\Sigma}$. Hence,

$$\mu_j^{\mathrm{OR}}(s) = \mu_j(s) + \boldsymbol{\sigma}_{q_j} \boldsymbol{\Sigma}_{q_j}^{-1} \boldsymbol{e}_{q_j}(s),$$

where $\boldsymbol{\sigma}_{q_j}$ is a $1 \times q_j$ subvector defined with the corresponding elements of $\boldsymbol{\Sigma}$, and $\boldsymbol{e}_{q_j}(s) = \bar{y}_{m(-j)}^{\mathrm{obs}} - \bar{\mu}(s)_{m(-j)}^{\mathrm{obs}}$.

5.2.4 Estimators and Inference

In this section, we describe the ARCC and RCC estimators for our estimand of interest, the mean outcome conditional on continuation at time point j for individual i; i.e., μ_{ij}^{RCC}. This estimand is formulated as in (5.14), so our attention is focused on the necessary intermediate quantity $\mu_{ij}(s) = E\big(Y_{ij}\big|\bar{\boldsymbol{x}}_{im}, S = s\big)$. Let $\boldsymbol{d}'_{ij} = \partial\mu_{ij}(s)/\partial\boldsymbol{\beta}_s$ denote the Jacobian of partial derivatives of $\mu_{ij}(s)$ with respect to $\boldsymbol{\beta}_s$. Let n_s denote the number of individuals with $S = s$.

Following Chen and Zhou (2011), consider the vector estimating equation

$$U\left(\boldsymbol{\beta}_s\right) = \sum_{i=1}^{n_s}\sum_{j=1}^{s} \boldsymbol{d}_{ij}\left(W_{ij} - \mu_{ij}(s)\right), \tag{5.15}$$

where

$$W_{ij} = \frac{R_{ij}}{\pi_{ij}}\left(Y_{ij} - \mu_{ij}^{\mathrm{OR}}(s)\right) + \mu_{ij}^{\mathrm{OR}}(s)$$

and

$$\pi_{ij} = \Pr\left(R_{ij} = 1\big|\bar{r}_{i(j-1)}, \bar{r}^{D}_{i(j-1)}, \bar{y}_{im}, \bar{\boldsymbol{x}}_{im}, S_i = s\right)$$

is the joint probability of not being missing, conditional on the history of missingness and dropout, on all outcomes and covariates, and on the full truncation vector. The IPW probability π_{ij} is unknown in practice, but can be consistently estimated if the dropout and IM mechanism models are correctly specified. The OR quantity $\mu_{ij}^{\mathrm{OR}}(s)$ is also unknown in practice, but can likewise be consistently estimated by correctly specifying the relevant OR model.

The RCC quantity μ_{ij}^{RCC} is estimated in two steps. The quantity $\mu_{ij}(s)$ is first estimated, and subsequently the correspondence in (5.14) is used to calculate μ_{ij}^{RCC} using a consistent estimator of p_{s_k}. Suppose DAR and IMAR hold such that $\pi_{ij} = \lambda_{ij}^{IM}(c_{j+1}, c)\lambda_{ij}^{D}(c)$. Let $\hat{\pi}_{ij}$ represent an estimator of π_{ij} with estimates $\hat{\lambda}_{ij}^{IM}$ and $\hat{\lambda}_{ij}^{D}$ substituted for λ_{ij}^{IM} and λ_{ij}^{D}, respectively, and let $\hat{\mu}_{ij}^{\mathrm{OR}}$ likewise represent an estimator of $\mu_{ij}^{\mathrm{OR}}(s)$. Let $\hat{\boldsymbol{\beta}}$ denote the solution to $U\big(\boldsymbol{\beta}_s\big) = \mathbf{0}$ when $\hat{\pi}_{ij}$ and $\hat{\mu}_{ij}^{\mathrm{OR}}$ are substituted for π_{ij} and $\mu_{ij}^{\mathrm{OR}}(s)$, respectively. If $\hat{\pi}_{ij}$ and $\hat{\mu}_{ij}^{\mathrm{OR}}$ are both consistent for their respective estimands, then $\hat{\boldsymbol{\beta}}$ is consistent and asymptotically normal for $\boldsymbol{\beta}_s$ (Robins et al., 1995). Furthermore, $\hat{\boldsymbol{\beta}}$ is doubly robust to misspecification in the sense that $\hat{\boldsymbol{\beta}}$ is consistent for $\boldsymbol{\beta}_s$ even if only one (and not both) of either $\hat{\pi}_{ij}$ or $\hat{\mu}_{ij}^{\mathrm{OR}}$ is consistent for its estimand. This can

be seen by noting that C_{ij} is an always-observed time-varying covariate in the notation of Robins et al. (1995), Chen and Zhou (2011), and Shardell et al. (2015). Hence, the large-sample consistency results in these articles apply here as well. The estimator $\hat{\boldsymbol{\beta}}$ is also asymptotically normal (Robins et al., 1995, section 6 material on $\tilde{\beta}^{\dagger}$ and $\tilde{\beta}$). Let $\hat{p}_{s_k}$ denote a consistent estimator for p_{s_k}. Let $\hat{\mu}_{ij}^{RCC}$ represent (5.14) with $\hat{\mu}_{ij}(s) = h^{-1}\big(\boldsymbol{x}_{ij}'\hat{\boldsymbol{\beta}}_s\big)$ substituted for $\mu_{ij}(s)$ and $\hat{p}_{s_k}$ substituted for p_{s_k}. By the continuous mapping theorem, if $\hat{\boldsymbol{\beta}}$ is consistent for $\boldsymbol{\beta}_s$ then $\hat{\mu}_{ij}^{RCC}$ is consistent for μ_{ij}^{RCC}. The variance of $\hat{\mu}_{ij}^{RCC}$ can be consistently estimated using the bootstrap estimator.

Compared to RCC, ARCC provides more robust estimation in cases when IPWs are incorrectly modeled as IMRCC-IPWs. When outcomes are DAR and IMAR, implementing RCC with IMRCC-IPWs produces estimators that are generally not consistent for $\boldsymbol{\beta}_s$. However, running ARCC with IMRCC-IPWs and the correct OR model will still produce a consistent estimator. Let $\hat{\pi}_{ij}^{IM\dagger}$ represent a consistent estimator of $\pi_{ij}^{IM\dagger}$, and let $\hat{\boldsymbol{\beta}}^{IM\dagger}$ denote the solution to $U\big(\boldsymbol{\beta}_s\big) = \mathbf{0}$ when $\hat{\pi}_{ij}^{IM\dagger}$ and $\hat{\mu}_{ij}^{\mathrm{OR}}$ are substituted for π_{ij} and $\mu_{ij}^{\mathrm{OR}}(s)$, respectively. If there is no truncation and no dropout, then $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}^{IM\dagger}$ are identical. If there is truncation or dropout, and if $\hat{\mu}_{ij}^{\mathrm{OR}}$ is consistent for $\mu_{ij}^{\mathrm{OR}}(s)$, then $\hat{\boldsymbol{\beta}}^{IM\dagger}$ will still be consistent for $\boldsymbol{\beta}_s$. However, if there is truncation or dropout and if $\mu_{ij}^{\mathrm{OR}}(s)$ is misspecified, then $\hat{\boldsymbol{\beta}}^{IM\dagger}$ will generally not be consistent for $\boldsymbol{\beta}_s$.

5.3 Simulation Study

A simulation study was conducted to characterize the empirical performance of ARCC and RCC estimators. The continuous outcome of infant weight Y_{ij} was simulated for $n = 1000$ infants. The mean outcome of interest was $\mu_{ij}^{RCC} = E\big(Y_{ij}\big|C_{ij} = 1\big)$ at visits 1, 4, 6, 8, and 10 that correspond to time points $j = 1, \ldots, 5$, respectively. However, outcomes were generated and estimated for up to visit 12, i.e., time point $m = 6$, because estimation of the IM IPW at the last analytical visit depends on whether or not this is also the last visit with outcomes available for IM IPW estimation. Modeling and consistent estimation of the model parameters of the OR quantity also required use of all outcomes possibly available up to time point m.

Truncation was not at random (TNAR) if outcomes were generated conditional on truncation by specifying $f\big(\bar{y}_{im}\big|s_i\big)$ as the generating probability density function (PDF). Truncation was

completely at random (TCAR) if outcomes were generated independently of truncation by specifying $f(\bar{y}_{im})$ as the PDF. Simulation parameters were chosen so that outcomes of lighter infants were more likely to be truncated under TNAR, while heavier infants were more likely to drop out under DAR. Infant weight trajectories were simulated according to combinations of the following mechanisms: TCAR or TNAR; DCAR or DAR; and IMCAR, IMAR similar to the truncation mechanism (IMART), or IMAR similar to the dropout mechanism (IMARD). For each of the 12 resulting scenarios, we generated and analyzed $\ell = 1, \ldots, 31$ simulation data sets. All parameters for generating outcomes were derived from the BAN data, and can be found in Tables 3.2 and 3.3.

5.3.1 Data Generation Procedure

For all infants and all time points, age at time point j was set to the corresponding mean BAN age, i.e., $age_{ij} \equiv \mu_{age(j)}$. About half of all infants were generated with $sex_i = 0$, and the remaining half with $sex_i = 1$. Simulated outcomes and events were generated as follows. Note that $S_i \geq j$ and $C_{ij} = 1$ are equivalent statements, as are $S_i = j$ and $\{C_{i(j+1)} = 0, C_{ij} = 1\}$ where $\{C_{i(j+1)}\} = \emptyset$ at $j = m$.

1. For each individual i, the number of continuing time points S_i was generated using a multinomial distribution with probabilities $p_{s_0} = p_{s_1} = p_{s_2} = 0$, $p_{s_3} = p_{s_4} = p_{s_5} = 0.1$, and $p_{s_6} = 0.7$ to ensure about 30% truncation by time point 5 (i.e., visit 10). Note that realizations c_{ij} and s_i were thereby generated simultaneously.

2. Set $\mu_{ij}(s_i) = w(s_i)\alpha_0 + \alpha_1 age_{ij} + \alpha_2 age_{ij}^2$ for $s > 0$, with $\alpha_0 = 2.9964$, $\alpha_1 = 0.0389$, and $\alpha_2 = -0.0001$. If $c_{ij} = 1$, then an outcome was generated as $Y_{ij} = \mu_{ij}(s_i) + b_i + \varepsilon_{ij}$, where $b_i \sim N(0, \sigma_b^2)$ was the random shift generated for individual i, and measurement error was generated as $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$, with $\sigma_b = 0.59$ and $\sigma_\varepsilon = 0.44$. We set $b_i \perp\!\!\!\perp \varepsilon_{ij}$ for all j, and $\bar{Y}_{im} \perp\!\!\!\perp \bar{Y}_{i'm}$ for all $i' \neq i$. Hence, outcomes for each individual i conditional on s_i were distributed as $N(\bar{\mu}_m(s_i), \mathbf{\Sigma})$, where $\mathbf{\Sigma}$ was a compound-symmetric (i.e., exchangeable) covariance matrix with diagonal elements $\sigma_b^2 + \sigma_\varepsilon^2$ and non-diagonal elements σ_b^2. TCAR outcomes were generated by setting $w(s_i) = 1$, while TNAR outcomes were generated by letting $w(s_i) = \zeta_0 + \zeta_1(s_i - 1)/(m - 1)$, where $\zeta_1 = 2(1 - \zeta_0)$ and $0 < \zeta_0 < 1$. In our

simulations, we set $\zeta_0 = 0.8$. Under TNAR, $w(s_{med}) = 1$ at the median value $s_{med} = (m+1)/2$, but otherwise $\mu_{ij}(s_i)$ was shifted by some distance with range $\left[\zeta_0, \zeta_0 + \zeta_1\right]$ from the median value. Hence, heavier infants tended to continue longer.

3. If $s_i \geq 1$, then r_{i1}^D was generated using $\lambda_{i1}^D(1)$.

4. If $s_i > 1$, the following was done for $j < s_i$. If $r_{ij}^D = 0$, then $r_{ij} \equiv 0$ was set. Otherwise, if $r_{ij}^D = 1$ then r_{ij} was generated using $\lambda_{ij}^{IM}(1,1)$. Subsequently, if $r_{ij} = 0$ then $r_{i(j+1)}^D \equiv r_{ij}^D$ was set. Otherwise, if $r_{ij} = 1$ then $r_{i(j+1)}^D$ was generated using $\lambda_{i(j+1)}^D(1)$.

5. At $j = s_i$, $r_{ij} \equiv r_{ij}^D$ was set.

6. For all $j > s_i$, r_{ij}^D and r_{ij} were left undefined.

For a quantity b, let $g_j\big(b\big) = (j-1)^{-1} b \sum_{k=1}^{j-1} k R_{ik} Y_{ik}$ for $j > 1$. Dropout was generated using a probit model specified as $\lambda_{ij}^D(1) = \Phi\big\{\eta_0^D + I(j > 1) g_j\big(\eta_1^D\big)\big\}$, where $\Phi\big(\,\cdot\,\big)$ is the standard normal cumulative distribution function, and $\eta_0^D = \Phi^{-1}\big(p_D^{1/m}\big)$ where a fixed value was assigned to p_D. IM was generated likewise with an identical model, but with $\lambda_{ij}^{IM}(1,1)$, η_0^{IM}, η_1^{IM}, and p_{IM} replacing $\lambda_{ij}^D(1)$, η_0^D, η_1^D, and $p_D^{1/m}$, respectively, where a fixed value was assigned to p_{IM}. For each of the seven mechanisms, the simulation values p_D, η_1^D, p_{IM}, and η_1^{IM} are listed in Table 3.4. Values for p_D and p_{IM} were chosen so that truncation, dropout, and IM occurred at rates of approximately 8%, 14%, and 5%, respectively, by time point 3.

5.3.2 Results

Six regression methods were used to estimate $E\big(Y_{ij}\big|C_{ij} = 1\big)$ for $j = 1, \ldots, 6$. The event probability π_{ij} was modeled correctly by specifying $\lambda_{ij}^{D(1)}$ and $\lambda_{ij}^{IM(1)}$ as the probit models defined in Section 5.3.1, while π_{ij} was wrongly specified as $\pi_{ij}^{IM\dagger}$. Estimates of $\pi_{ij}^{IM\dagger}$ were calculated by using all observations regardless of the corresponding $C_{i(j+1)}$, C_{ij}, or R_{ij}^D values. The correct OR model was specified by fitting $Y_{ij} = \mu_{ij}(s_i) + b_i + \varepsilon_{ij}$, where $\mu_{ij}(s_i) = \alpha_{0s_i} + \alpha_{1s_i} age_{ij} + \alpha_{2s_i} age_{ij}^2$. The wrong OR model does not account for truncation, and was specified as $Y_{ij} = \mu_j^\dagger + b_i + \varepsilon_{ij}$, where $\mu_j^\dagger = \alpha_0^\dagger + \alpha_1^\dagger age_{ij}$. For both models, b_i and ε_{ij} were specified correctly as in the generating OR model.

For each simulated data set, μ_{ij}^{RCC} was modeled as $E\big(Y_{ij}\big|C_{ij} = 1\big) = \beta_0^{RCC} + I(j > 1)\beta_{j-1}^{RCC}$.

Denote the parameter vector that was estimated for each data set using the following six regression methods as $\boldsymbol{\beta}^{RCC} = \left(\beta_0^{RCC}, \beta_1^{RCC}, \beta_2^{RCC}, \beta_3^{RCC}\right)'$. Each parameter was estimated using the correspondence in (5.14) after using the sample proportion $\hat{p}_{s_k} = 1/n \sum_{i=1}^{n} I(s_i = k)$ to consistently estimate p_{s_k}, and estimating all relevant α coefficients.

1. ARCC-11. The correct models π_{ij} and $\mu_{ij}(s_i)$ were fit. This ARCC estimator was expected to be consistent.

2. ARCC-10. The correct model π_{ij} and the wrong model $\mu_j^{\dagger}(s_i)$ were fit. This ARCC estimator was expected to be consistent.

3. ARCC-01. The wrong model $\pi_{ij}^{IM\dagger}$ and the correct model $\mu_{ij}(s_i)$ were fit. This ARCC estimator was expected to be consistent.

4. ARCC-00. The wrong models $\pi_{ij}^{IM\dagger}$ and $\mu_j^{\dagger}(s_i)$ were fit. This ARCC estimator was generally not expected to be consistent.

5. RCC-1. The correct model π_{ij} was fit. This RCC estimator was expected to be consistent.

6. RCC-0. The wrong model $\pi_{ij}^{IM\dagger}$ was fit. This RCC estimator was generally not expected to be consistent.

Empirical bias and coverage were then calculated for each method as follows. Let $\hat{\beta}_{p\ell}$ denote the estimate of β_p^{RCC} for data set ℓ. For each of the $p = 0, \ldots, 3$ parameters, the empirical bias of $\hat{\beta}_p$ was calculated as $31^{-1} \sum_{\ell=1}^{31} \hat{\beta}_{p\ell} - \beta_p^{RCC}$.

The results are summarized as follows, where good performance was defined as an absolute empirical bias of less than or equal to 0.003. As in our previous study on RCC, RCC0 performed worst in all 12 scenarios. In all six TCAR scenarios, all four ARCC methods and RCC1 performed well. The following are the results for the six TNAR scenarios. ARCC11 and RCC1 performed best, and ARCC10 generally performed well (i.e., for 15 out of 18 coefficients). ARCC01 and ARCC00 both performed decently (i.e., for 11 out of 18 coefficients). Hence, ARCC was shown to protect against misspecification of the IPW model. Figure 5.5 illustrates typical results from two scenarios.

5.4 Analysis of the BAN Study

The ARCC method was applied to our sample of $n = 2238$ infants in the BAN data, accounting for 307 dropouts, 187 truncations, and 973 IM observations. The goal was to estimate the mean infant outcome at each of nine scheduled follow-up visits for those infants who were alive and uninfected (i.e., who had continuing outcome trajectories) at that visit.

Growth patterns of girls and boys were considered to be different a priori. Hence, the mean outcome for infant i at visit j conditional on continuation, $\mu_{ij} = E\big(Y_{ij}\big|C_{ij} = 1\big)$, was modeled separately for girls and boys. The mean outcome was modeled as a linear function of 1.) drug assignment to no ART (the reference), maternal ART, or infant ART, 2.) supplement assignment to no LNS (the reference) or LNS, 3.) dummy indicator variables for visit with visit 1 as the reference, and 4.) interactions between drug assignment, supplement assignment, and visit. We modeled the conditional probability of non-dropout, $\lambda_{ij}^{D}(1)$, as a probit function with observed past infant outcomes $\bar{Y}_{i(j-1)}^{\text{obs}}$ and drug/supplement group assignments and their interactions as predictors. The conditional probability of non-IM, $\lambda_{ij}^{IM}(1,1)$, was modeled likewise. The OR model was specified as the mixed-effect model conditioned on $s_i = 1, \ldots, m$, defined as $Y_{ij} = \alpha_0(s_i) + \alpha_1(s_i)age_{ij} + \alpha_2(s_i)age_{ij}^2 + b_i + \varepsilon_{ij}$, where $b_i \sim N\big(0, \sigma_b^2\big)$ was the shift for individual i, and $\varepsilon_{ij} \sim N\big(0, \sigma_\varepsilon^2\big)$ was measurement error. The corresponding IPWs were used to estimate the mean outcomes at each visit, and standard errors were estimated using the empirical sandwich variance estimator.

The results for the infant outcomes of length, BMI, and weight are reported as follows. Infant length was not significantly associated with drug or supplement in either boys (Wald test $p = 0.60$) or girls ($p = 0.89$). Infant BMI was also not significantly associated with drug or supplement in either boys ($p = 0.56$) or girls ($p = 0.10$). Likewise, infant weight was not significantly associated with drug or supplement in either boys ($p = 0.37$) or girls ($p = 0.10$). Figure 5.6 depicts the estimated means and 95% CIs separately for girls and boys, for each treatment group at study period weeks 6, 12, 18, and 24 (i.e., visits 5, 7, 8, and 10, respectively).

Our findings are consistent with the results from our original application of RCC to the BAN study data, in which LNS was not significantly associated with infant weight. The application of

ARCC to this same analysis is additional evidence that the truncation, dropout, and intermittent missingness mechanisms may not have been different, and may have been unrelated to infant weight. The caveat still remains, however, that there may simply have been too few of these events to indicate any true variation between the mechanisms. The reduced-model parameters in Table 5.1.

5.5 Discussion

The method of regression conditioning on continuation for continuous longitudinal outcomes was extended using an augmented estimating equation approach, and the resulting augmented RCC estimator is doubly robust to misspecification of either the missingness or joint-outcome model. Truncation, dropout, and intermittent missingness mechanisms were varied to produce 12 simulation scenarios used to compare the performance of ARCC with that of RCC. The results indicated that ARCC is indeed doubly robust, and the method was applied to the BAN study data. These findings resembled the original RCC results, further providing evidence that the truncation, dropout, and intermittent missingness mechanisms may not have differed from each other, and may have been unassociated with infant weight.

As alluded to by Kurland and Heagerty (2005), the inferences acquired via ARCC may be applied towards future out-of-sample predictions without having to specify the full truncation distribution. This is advantageous in clinical practice, wherein a clinician likely will not know a patient's future continuation status, yet will nonetheless be able to use previously derived ARCC results to predict a continuing patient's current outcome. However, in our development of both RCC and ARCC, both IPW and OR-quantity estimation required future knowledge of truncation status. That is, truncation status for at least one time point beyond the current time point was needed to estimate model parameters. While this is not problematic if the ARCC analysis results will be used for future in-clinic predictions, further extensions might involve developing ARCC techniques that do not rely on such future knowledge. Such extensions would allow for real-time ARCC analysis to be conducted across a common set of patients, and across clinics, to better inform clinical decision-making.

While specification of parametric mechanisms facilitated our statistical study of RCC properties, methods that allow the functional forms of the mechanisms to vary may be more compatible with the aims of the BAN study researchers, who were not concerned with specifying models for the dropout or IM mechanisms. For example, machine learning might be applied to find the best-fitting mechanism models, since human interpretability is not required for these so-called nuisance models. If ARCC is the planned analysis, this type of approach might also be applied to the OR model. In ARCC, a model for the joint outcomes is not of primary interest, as might be the case if the planned primary analysis is a maximum likelihood estimation or multiple-imputation procedure. We also limited our investigation to at-random mechanisms, while actual mechanisms may produce data subject to not-at-random dropout or IM. Studies of the sensitivity of the mean infant weight estimates to such mechanism misspecifications would assist researchers in deciding how to explicitly model distributions of dropout, IM, and the joint outcomes.

5.6 ARCC GEE Derivation from Chen and Zhou (2011)

We derive the ARCC GEE expression (5.15) using the general framework and notation of Chen and Zhou (2011), hereafter referred to as CZ2011. Specifically, (5.15) is derived from their expression

$$\boldsymbol{S}_1(\theta) = \sum_{i=1}^{n} \boldsymbol{S}_{1i}(\theta) = \sum_{i=1}^{n} \left[\boldsymbol{D}_i \boldsymbol{M}_i (\boldsymbol{Y}_i - \mu_i) + E_{(\boldsymbol{Y}_i^m, \boldsymbol{X}_i^m | \boldsymbol{Y}_i^o, \boldsymbol{X}_i^o, \boldsymbol{Z}_i, \boldsymbol{R}_i)} \{ \boldsymbol{D}_i \boldsymbol{N}_i (\boldsymbol{Y}_i - \mu_i) \} \right]$$

as follows, specifying each quantity. In this section, "page" refers to the CZ2011 article page. The following material will be organized by article section number.

(2.1) Response Process

Refer to the following mean-outcome expression.

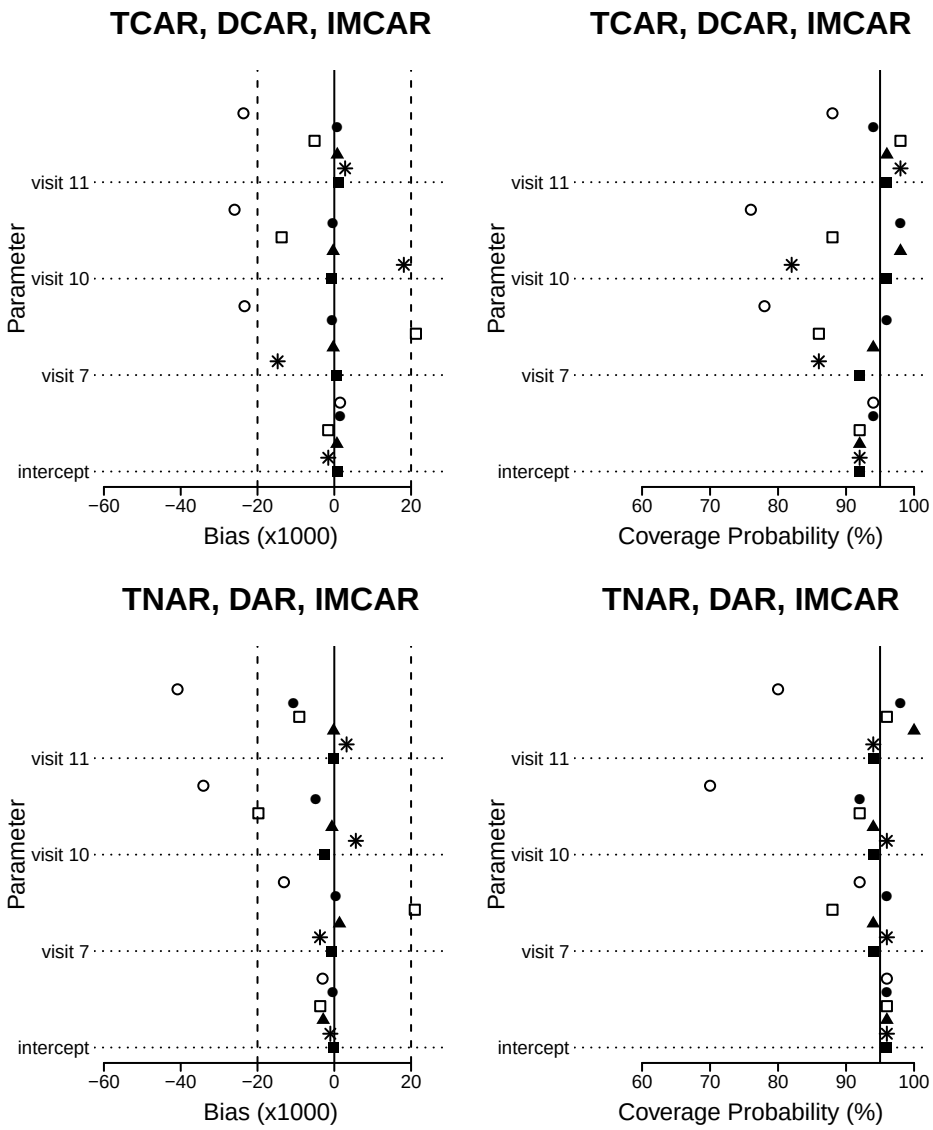
$$g(\mu_{ij}) = X_{ij}\beta_x + \boldsymbol{Z}'_{ij}\beta_z.$$

We do not consider the case with missing covariates, so we exclude X_{ij} and β_x from all expressions where relevant, exclude all results concerning X_{ij} where relevant. Let $v_{ij} = \sigma^2$ for all i and

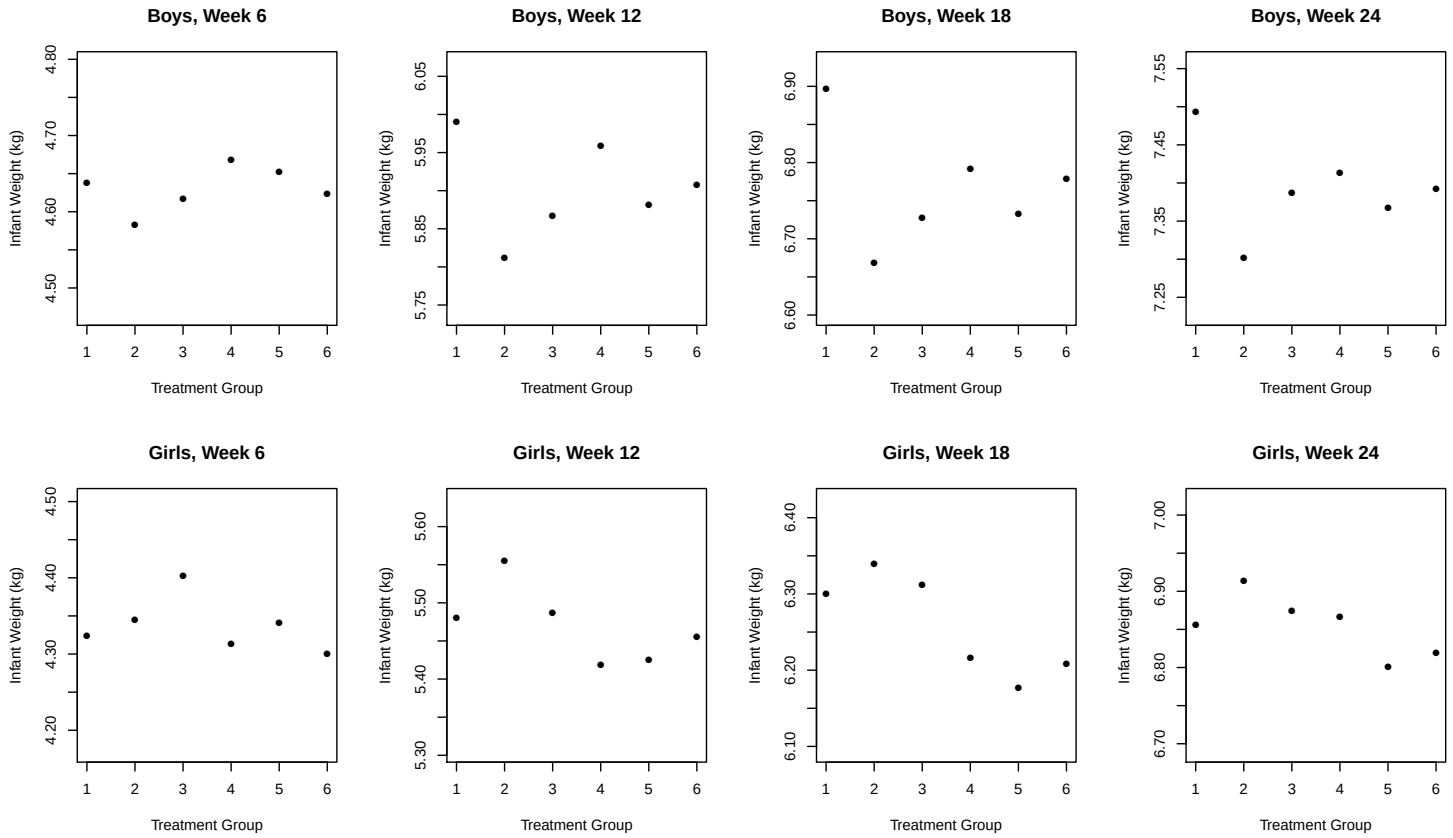
Supporting Table 5.1: BAN infant weight RCC parameter estimates for reduced models.

Sex	Covariate	Estimate (95% CI)
male	intercept	3.07 (3.05, 3.10)
	visit 2	0.14 (0.12, 0.15)
	visit 3	0.38 (0.36, 0.40)
	visit 4	1.00 (0.98, 1.02)
	visit 5	1.56 (1.53, 1.59)
	visit 6	2.04 (2.01, 2.07)
	visit 7	2.83 (2.80, 2.87)
	visit 8	3.70 (3.66, 3.74)
	visit 9	4.05 (4.00, 4.09)
	visit 10	4.33 (4.29, 4.38)
female	intercept	2.98 (2.95, 3.00)
	visit 2	0.13 (0.12, 0.15)
	visit 3	0.34 (0.32, 0.35)
	visit 4	0.89 (0.87, 0.92)
	visit 5	1.37 (1.35, 1.40)
	visit 6	1.79 (1.76, 1.82)
	visit 7	2.50 (2.47, 2.53)
	visit 8	3.29 (3.25, 3.33)
	visit 9	3.61 (3.57, 3.65)
	visit 10	3.88 (3.84, 3.93)

Note: All estimates were statistically significant at $\alpha = 0.001$.



Supporting Figure 5.5: Simulation study: Empirical biases ($\times 1000$) under TCAR, DCAR, IMCAR and TNAR, DAR, IMARD. (50 simulated datasets, 1000 subjects; ARCC-11 ■, ARCC-10 $*$, ARCC-01 ▲, ARCC-00 □, RCC-1 ●, RCC-0 ○. Dotted lines are marked at ± 20 on bias figures.)



Supporting Figure 5.6: Estimates and 95% CIs of mean weight for HIV-negative, alive infants at study period weeks 6, 12, 18, and 24 using data from the BAN study. These correspond to study visits 5, 7, 8, and 10, respectively. The y-axis scales are identical. (Treatment Group: 1 = control and no LNS, 2 = maternal ART and no LNS, 3 = infant ART and no LNS, 4 = control and LNS, 5 = maternal ART and LNS, 6 = infant ART and LNS. Dashed lines correspond to CIs of the reference group, Treatment Group 1.)

j.

(2.2) Missing Data Process

For notational consistency with the article material, let $R_{ij} = 1$ if Y_{ij} is missing, and let $R_{ij} = 3$ if Y_{ij} is observed. In our case, for MAR mechanisms we require

$$P\left(\boldsymbol{R}_i = \boldsymbol{r}_i \middle| \boldsymbol{Y}_i, \boldsymbol{Z}_i\right) = P\left(\boldsymbol{R}_i = \boldsymbol{r}_i \middle| \boldsymbol{Y}_i^o, \boldsymbol{Z}_i\right),$$

and we make the further assumption that

$$P\left(R_{ij} = r_{ij} \middle| \bar{\boldsymbol{R}}_{ij}, \boldsymbol{Y}_i, \boldsymbol{Z}_i\right) = P\left(R_{ij} = r_{ij} \middle| \bar{\boldsymbol{R}}_{ij}, \bar{\boldsymbol{Y}}_{ij}^o, \boldsymbol{Z}_i\right),$$

which implies MAR defined above. Let

$$\begin{aligned}
\pi_{ij} &= P\left(R_{ij} = 3 \middle| \boldsymbol{Y}_i, \boldsymbol{Z}_i\right) \\
&= \sum_{\bar{\boldsymbol{r}}_{ij}} P\left(R_{ij} = 3, \bar{\boldsymbol{R}}_{ij} = \bar{\boldsymbol{r}}_{ij} \middle| \boldsymbol{Y}_i, \boldsymbol{Z}_i\right) \\
&= \sum_{\bar{\boldsymbol{r}}_{ij}} P\left(R_{ij} = 3 \middle| \bar{\boldsymbol{R}}_{ij} = \bar{\boldsymbol{r}}_{ij}, \boldsymbol{Y}_i, \boldsymbol{Z}_i\right) P\left(\bar{\boldsymbol{R}}_{ij} = \bar{\boldsymbol{r}}_{ij} \middle| \boldsymbol{Y}_i, \boldsymbol{Z}_i\right) \\
&= \sum_{\bar{\boldsymbol{r}}_{ij}} P\left(R_{ij} = 3 \middle| \bar{\boldsymbol{R}}_{ij} = \bar{\boldsymbol{r}}_{ij}, \boldsymbol{Y}_i, \boldsymbol{Z}_i\right) P\left(R_{i,j-1} = r_{i,j-1}, \bar{\boldsymbol{R}}_{i,j-1} = \bar{\boldsymbol{r}}_{i,j-1} \middle| \boldsymbol{Y}_i, \boldsymbol{Z}_i\right) \\
&= \sum_{\bar{\boldsymbol{r}}_{ij}} P\left(R_{ij} = 3 \middle| \bar{\boldsymbol{R}}_{ij} = \bar{\boldsymbol{r}}_{ij}, \boldsymbol{Y}_i, \boldsymbol{Z}_i\right) P\left(R_{i,j-1} = r_{i,j-1} \middle| \bar{\boldsymbol{R}}_{i,j-1} = \bar{\boldsymbol{r}}_{i,j-1}, \boldsymbol{Y}_i, \boldsymbol{Z}_i\right) \times \\
&\qquad P\left(\bar{\boldsymbol{R}}_{i,j-1} = \bar{\boldsymbol{r}}_{i,j-1} \middle| \boldsymbol{Y}_i, \boldsymbol{Z}_i\right) \\
&= \sum_{\bar{\boldsymbol{r}}_{ij}} P\left(R_{ij} = 3 \middle| \bar{\boldsymbol{R}}_{ij} = \bar{\boldsymbol{r}}_{ij}, \boldsymbol{Y}_i, \boldsymbol{Z}_i\right) P\left(R_{i,j-1} = r_{i,j-1} \middle| \bar{\boldsymbol{R}}_{i,j-1} = \bar{\boldsymbol{r}}_{i,j-1}, \boldsymbol{Y}_i, \boldsymbol{Z}_i\right) \times \cdots \\
&\qquad \times P\left(R_{i3} = r_{i3} \middle| \bar{\boldsymbol{R}}_{i3} = \bar{\boldsymbol{r}}_{i3}, \boldsymbol{Y}_i, \boldsymbol{Z}_i\right) P\left(R_{i2} = r_{i2} \middle| R_{i1} = r_{i1}, \boldsymbol{Y}_i, \boldsymbol{Z}_i\right) P\left(R_{i1} = r_{i1} \middle| \boldsymbol{Y}_i, \boldsymbol{Z}_i\right) \\
&= \sum_{\bar{\boldsymbol{r}}_{ij}} P\left(R_{ij} = 3 \middle| \bar{\boldsymbol{R}}_{ij} = \bar{\boldsymbol{r}}_{ij}, \boldsymbol{Y}_i, \boldsymbol{Z}_i\right) \prod_{k=1}^{j-1} P\left(R_{ik} = r_{ik} \middle| \bar{\boldsymbol{R}}_{ik} = \bar{\boldsymbol{r}}_{ik}, \boldsymbol{Y}_i, \boldsymbol{Z}_i\right) \\
&= \sum_{\bar{\boldsymbol{r}}_{ij}} P\left(R_{ij} = 3 \middle| \bar{\boldsymbol{R}}_{ij} = \bar{\boldsymbol{r}}_{ij}, \bar{\boldsymbol{Y}}_{ij}^o, \boldsymbol{Z}_i\right) \prod_{k=1}^{j-1} P\left(R_{ik} = r_{ik} \middle| \bar{\boldsymbol{R}}_{ik} = \bar{\boldsymbol{r}}_{ik}, \bar{\boldsymbol{Y}}_{ik}^o, \boldsymbol{Z}_i\right) \text{ under MAR,}
\end{aligned}$$

where $\bar{\boldsymbol{R}}_{i1} = \emptyset$.

(3.1) Weighted Estimating Equation for the Response Parameters

Let $\boldsymbol{D}_{ij} = \partial\mu_{ij}/\partial\beta_z$, and note that

$$\boldsymbol{D}_i = \frac{\partial\mu_i'}{\partial\beta_z} = \left(\boldsymbol{D}_{i1}, \ldots, \boldsymbol{D}_{iJ_i}\right)_{p\times J_i}.$$

We use a working independence assumption such that the working correlation matrix is $\boldsymbol{C}_i = \boldsymbol{I}$. Let

$$\boldsymbol{F}_i = \operatorname{diag}\left(v_{ij}, j = 1, \ldots, J_i\right) = \operatorname{diag}\left(\sigma^2, j = 1, \ldots, J_i\right) = \sigma^2\boldsymbol{I},$$

where $\boldsymbol{I}$ is the $J_i \times J_i$ identity matrix. We now have

$$\begin{aligned}
\boldsymbol{M}_i &= \boldsymbol{F}_i^{-1/2}\left(\boldsymbol{C}_i^{-1} \bullet \boldsymbol{\Delta}_i\right)\boldsymbol{F}_i^{-1/2} \\
&= \boldsymbol{F}_i^{-1/2}\left(\boldsymbol{I}_i^{-1} \bullet \boldsymbol{\Delta}_i\right)\boldsymbol{F}_i^{-1/2} \text{ under working independence} \\
&= \boldsymbol{F}_i^{-1/2}\boldsymbol{\Delta}_i^*(\alpha)\,\boldsymbol{F}_i^{-1/2} \\
&= \frac{1}{\sigma^2}\boldsymbol{\Delta}_i^*(\alpha),
\end{aligned}$$

where $\bullet$ denotes the Hadamard product of matrices, and

$$\boldsymbol{\Delta}_i^*(\alpha) = \operatorname{diag}\left(\frac{I\left(R_{ij} = 3\right)}{\pi_{ij}}, 1 \le j \le J_i\right)$$

is unchanged from the article. Note that the the definition of $\delta_{ijj'}$ does not matter because the Hadamard product zeros all of these values under the working independence assumption. We also have

$$\begin{aligned}
\boldsymbol{N}_i &= \boldsymbol{F}_i^{-1/2}\left\{\boldsymbol{C}_i^{-1} \bullet \left(\boldsymbol{1}\boldsymbol{1}' - \boldsymbol{\Delta}_i\right)\right\}\boldsymbol{F}_i^{-1/2} \\
&= \boldsymbol{F}_i^{-1/2}\left\{\boldsymbol{I}^{-1} \bullet \left(\boldsymbol{1}\boldsymbol{1}' - \boldsymbol{\Delta}_i\right)\right\}\boldsymbol{F}_i^{-1/2} \text{ under working independence} \\
&= \boldsymbol{F}_i^{-1/2}\left\{\boldsymbol{I} - \boldsymbol{\Delta}_i^*(\alpha)\right\}\boldsymbol{F}_i^{-1/2} \\
&= \frac{1}{\sigma^2}\left\{\boldsymbol{I} - \boldsymbol{\Delta}_i^*(\alpha)\right\},
\end{aligned}$$

where $\mathbf{1}$ is a $J_i \times 1$ vector. Note that the Hadamard product again zeros all of these values under working independence.

The resulting estimating equations are

$$\begin{aligned}
\boldsymbol{S}_1(\theta) &= \sum_{i=1}^{n} \boldsymbol{S}_{1i}(\theta) \\
&= \sum_{i=1}^{n} \left[\boldsymbol{D}_i \boldsymbol{M}_i (\boldsymbol{Y}_i - \mu_i) + E_{(\boldsymbol{Y}_i^m | \boldsymbol{Y}_i^o, \boldsymbol{Z}_i, \boldsymbol{R}_i)} \left\{ \boldsymbol{D}_i \boldsymbol{N}_i (\boldsymbol{Y}_i - \mu_i) \right\} \right] \\
&= \sum_{i=1}^{n} \left[\boldsymbol{D}_i \frac{1}{\sigma^2} \boldsymbol{\Delta}_i^* (\alpha) (\boldsymbol{Y}_i - \mu_i) + E_{(\boldsymbol{Y}_i^m | \boldsymbol{Y}_i^o, \boldsymbol{Z}_i, \boldsymbol{R}_i)} \left\{ \boldsymbol{D}_i \frac{1}{\sigma^2} \left(\boldsymbol{I} - \boldsymbol{\Delta}_i^* (\alpha) \right) (\boldsymbol{Y}_i - \mu_i) \right\} \right] \\
&= \frac{1}{\sigma^2} \sum_{i=1}^{n} \left[\boldsymbol{D}_i \boldsymbol{\Delta}_i^* (\alpha) (\boldsymbol{Y}_i - \mu_i) + E_{(\boldsymbol{Y}_i^m | \boldsymbol{Y}_i^o, \boldsymbol{Z}_i, \boldsymbol{R}_i)} \left\{ \boldsymbol{D}_i \left(\boldsymbol{I} - \boldsymbol{\Delta}_i^* (\alpha) \right) (\boldsymbol{Y}_i - \mu_i) \right\} \right].
\end{aligned}$$

Note that

$$\begin{aligned}
P\left(\boldsymbol{Y}_i^m \middle| \boldsymbol{Y}_i^o, \boldsymbol{Z}_i, \boldsymbol{R}_i\right) &= \frac{P(\boldsymbol{Y}_i^m, \boldsymbol{Y}_i^o, \boldsymbol{Z}_i, \boldsymbol{R}_i)}{P(\boldsymbol{Y}_i^o, \boldsymbol{Z}_i, \boldsymbol{R}_i)} \\
&= \frac{P(\boldsymbol{Y}_i, \boldsymbol{Z}_i, \boldsymbol{R}_i)}{P(\boldsymbol{Y}_i^o, \boldsymbol{Z}_i, \boldsymbol{R}_i)} \\
&= \frac{P\left(\boldsymbol{R}_i \middle| \boldsymbol{Y}_i, \boldsymbol{Z}_i, \right) P(\boldsymbol{Y}_i, \boldsymbol{Z}_i)}{P\left(\boldsymbol{R}_i \middle| \boldsymbol{Y}_i^o, \boldsymbol{Z}_i, \right) P(\boldsymbol{Y}_i^o, \boldsymbol{Z}_i)} \\
&= \frac{P\left(\boldsymbol{R}_i \middle| \boldsymbol{Y}_i^o, \boldsymbol{Z}_i, \right) P(\boldsymbol{Y}_i, \boldsymbol{Z}_i)}{P\left(\boldsymbol{R}_i \middle| \boldsymbol{Y}_i^o, \boldsymbol{Z}_i, \right) P(\boldsymbol{Y}_i^o, \boldsymbol{Z}_i)} \text{ under MAR} \\
&= \frac{P(\boldsymbol{Y}_i, \boldsymbol{Z}_i)}{P(\boldsymbol{Y}_i^o, \boldsymbol{Z}_i)} \\
&= \frac{P(\boldsymbol{Y}_i^m, \boldsymbol{Y}_i^o, \boldsymbol{Z}_i)}{P(\boldsymbol{Y}_i^o, \boldsymbol{Z}_i)} \\
&= P\left(\boldsymbol{Y}_i^m \middle| \boldsymbol{Y}_i^o, \boldsymbol{Z}_i\right).
\end{aligned}$$

Hence, the estimating equations are

$$\boldsymbol{S}_1(\theta) = \frac{1}{\sigma^2} \sum_{i=1}^{n} \left[\boldsymbol{D}_i \boldsymbol{\Delta}_i^* (\alpha) (\boldsymbol{Y}_i - \mu_i) + E_{(\boldsymbol{Y}_i^m | \boldsymbol{Y}_i^o, \boldsymbol{Z}_i)} \left\{ \boldsymbol{D}_i \left(\boldsymbol{I} - \boldsymbol{\Delta}_i^* (\alpha) \right) (\boldsymbol{Y}_i - \mu_i) \right\} \right].$$

Expanding these estimating equations, we have

$$
\begin{aligned}
\boldsymbol{S}_1(\theta) &= \frac{1}{\sigma^2}\sum_{i=1}^{n}\left[\boldsymbol{D}_i\boldsymbol{\Delta}_i^*(\alpha)(\boldsymbol{Y}_i-\mu_i)+E_{(\boldsymbol{Y}_i^m|\boldsymbol{Y}_i^o,\boldsymbol{Z}_i)}\left\{\boldsymbol{D}_i(\boldsymbol{I}-\boldsymbol{\Delta}_i^*(\alpha))(\boldsymbol{Y}_i-\mu_i)\right\}\right] \\
&\propto \sum_{i=1}^{n}\left\{\boldsymbol{D}_i\boldsymbol{\Delta}_i^*(\alpha)(\boldsymbol{Y}_i-\mu_i)+\boldsymbol{D}_i(\boldsymbol{I}-\boldsymbol{\Delta}_i^*(\alpha))\left(E_{(\boldsymbol{Y}_i^m|\boldsymbol{Y}_i^o,\boldsymbol{Z}_i)}(\boldsymbol{Y}_i)-\mu_i\right)\right\} \\
&= \sum_{i=1}^{n}\boldsymbol{D}_i\left\{\boldsymbol{\Delta}_i^*(\alpha)(\boldsymbol{Y}_i-\mu_i)+(\boldsymbol{I}-\boldsymbol{\Delta}_i^*(\alpha))\left(E_{(\boldsymbol{Y}_i^m|\boldsymbol{Y}_i^o,\boldsymbol{Z}_i)}(\boldsymbol{Y}_i)-\mu_i\right)\right\} \\
&= \sum_{i=1}^{n}\boldsymbol{D}_i\left\{\boldsymbol{\Delta}_i^*(\alpha)(\boldsymbol{Y}_i-\mu_i)+\left(E_{(\boldsymbol{Y}_i^m|\boldsymbol{Y}_i^o,\boldsymbol{Z}_i)}(\boldsymbol{Y}_i)-\mu_i\right)-\boldsymbol{\Delta}_i^*(\alpha)\left(E_{(\boldsymbol{Y}_i^m|\boldsymbol{Y}_i^o,\boldsymbol{Z}_i)}(\boldsymbol{Y}_i)-\mu_i\right)\right\} \\
&= \sum_{i=1}^{n}\boldsymbol{D}_i\left\{\boldsymbol{\Delta}_i^*(\alpha)\left(\boldsymbol{Y}_i-\mu_i-E_{(\boldsymbol{Y}_i^m|\boldsymbol{Y}_i^o,\boldsymbol{Z}_i)}(\boldsymbol{Y}_i)+\mu_i\right)+\left(E_{(\boldsymbol{Y}_i^m|\boldsymbol{Y}_i^o,\boldsymbol{Z}_i)}(\boldsymbol{Y}_i)-\mu_i\right)\right\} \\
&= \sum_{i=1}^{n}\boldsymbol{D}_i\left\{\boldsymbol{\Delta}_i^*(\alpha)\left(\boldsymbol{Y}_i-E_{(\boldsymbol{Y}_i^m|\boldsymbol{Y}_i^o,\boldsymbol{Z}_i)}(\boldsymbol{Y}_i)\right)+\left(E_{(\boldsymbol{Y}_i^m|\boldsymbol{Y}_i^o,\boldsymbol{Z}_i)}(\boldsymbol{Y}_i)-\mu_i\right)\right\} \\
&= \sum_{i=1}^{n}\boldsymbol{D}_i\left\{\left(\left\{\frac{I(R_{ij}=3)}{\pi_{ij}}\left\{Y_{ij}-E\left(Y_{ij}\middle|\boldsymbol{Y}_i^o,\boldsymbol{Z}_i\right)\right\}\right\}\right)_{J_i\times 1}+\left(\left\{E\left(Y_{ij}\middle|\boldsymbol{Y}_i^o,\boldsymbol{Z}_i\right)-\mu_{ij}\right\}\right)_{J_i\times 1}\right\} \\
&= \sum_{i=1}^{n}\boldsymbol{D}_i\left(\left\{\frac{I(R_{ij}=3)}{\pi_{ij}}\left\{Y_{ij}-E\left(Y_{ij}\middle|\boldsymbol{Y}_i^o,\boldsymbol{Z}_i\right)\right\}+E\left(Y_{ij}\middle|\boldsymbol{Y}_i^o,\boldsymbol{Z}_i\right)-\mu_{ij}\right\}\right)_{J_i\times 1} \\
&= \sum_{i=1}^{n}\sum_{j=1}^{J_i}\boldsymbol{D}_{ij}(W_{ij}-\mu_{ij}),
\end{aligned}
$$

where

$$
W_{ij}=\frac{I(R_{ij}=3)}{\pi_{ij}}\left\{Y_{ij}-E\left(Y_{ij}\middle|\boldsymbol{Y}_i^o,\boldsymbol{Z}_i\right)\right\}+E\left(Y_{ij}\middle|\boldsymbol{Y}_i^o,\boldsymbol{Z}_i\right).
$$

To complete the derivation, we now redefine quantities using our own notation. Denote the CZ2011 link function $g(\cdot)$ as $h(\cdot)$, let $\boldsymbol{Z}_{ij}=\boldsymbol{x}_{ij}$ and denote the CZ2011 quantity β_z as $\boldsymbol{\alpha}_s$. Let $R_{ij}=1$ if Y_{ij} is observed, and let $R_{ij}=0$ otherwise. For a variable A, denote the CZ2011 quantity $\bar{\boldsymbol{A}}_{ij}$ as $\bar{A}_{i(j-1)}$, and denote $\bar{\boldsymbol{Y}}_{ij}^o=\bar{Y}_{j-1}^{\text{obs}}$.

5.7 Proof of ARCC Double-Robustness

We show that (5.15) is robust to misspecification of either the missingness or OR models. In this section, we suppress the i notation. Recall that

$$\begin{aligned}\pi_j &= \Pr\left(R_j = 1, R_j^D = 1 \middle| \bar{R}_{j-1}, \bar{R}_{j-1}^D, \bar{Y}_m, \bar{C}_m\right) \\ &= \Pr\left(R_j = 1, R_j^D = 1 \middle| \bar{R}_{j-1}, \bar{R}_{j-1}^D, \bar{Y}_{j-1}^{\text{obs}}, C_{j+1}, C_j\right) \text{ under DAR and IMAR.}\end{aligned}$$

If the missingness model is correct but the OR model is wrong, then

$$\begin{aligned}&E\left(W_j \middle| S = s\right) \\ &= E\left(W_j \middle| \bar{C}_m\right) \\ &= E_{\bar{Y}_m, R_j, R_j^D, \bar{R}_{j-1}, \bar{R}_{j-1}^D}\left[\frac{R_j R_j^D}{\pi_j}\left\{Y_j - E\left(Y_j \middle| \bar{Y}_{m(-j)}^{\text{obs}}, \bar{C}_m\right)\right\} + E\left\{Y_j \middle| \bar{Y}_{m(-j)}^{\text{obs}}, \bar{C}_m\right\} \middle| \bar{C}_m\right] \\ &= E_{\bar{Y}_m}\left(E_{R_j, R_j^D, \bar{R}_{j-1}, \bar{R}_{j-1}^D}\left[\frac{R_j R_j^D}{\pi_j}\left\{Y_j - E\left(Y_j \middle| \bar{Y}_{m(-j)}^{\text{obs}}, \bar{C}_m\right)\right\} + E\left\{Y_j \middle| \bar{Y}_{m(-j)}^{\text{obs}}, \bar{C}_m\right\} \middle| \bar{Y}_m, \bar{C}_m\right] \middle| \bar{C}_m\right) \\ &= E_{\bar{Y}_m}\left(E_{R_j, R_j^D, \bar{R}_{j-1}, \bar{R}_{j-1}^D}\left[\frac{R_j R_j^D}{\pi_j}\left\{Y_j - E\left(Y_j \middle| \bar{Y}_{m(-j)}^{\text{obs}}, \bar{C}_m\right)\right\} + E\left\{Y_j \middle| \bar{Y}_{m(-j)}^{\text{obs}}, \bar{C}_m\right\} \middle| \bar{Y}_m, \bar{C}_m\right] \middle| \bar{C}_m\right) \\ &= E_{\bar{Y}_m}\left\{E_{\bar{R}_{j-1}, \bar{R}_{j-1}^D}\left(E_{R_j, R_j^D}\left[\frac{R_j R_j^D}{\pi_j}\left\{Y_j - E\left(Y_j \middle| \bar{Y}_{m(-j)}^{\text{obs}}, \bar{C}_m\right)\right\} + \right.\right.\right. \\ &\qquad\qquad \left.\left.\left. E\left\{Y_j \middle| \bar{Y}_{m(-j)}^{\text{obs}}, \bar{C}_m\right\} \middle| \bar{R}_{j-1}, \bar{R}_{j-1}^D, \bar{Y}_m, \bar{C}_m\right] \middle| \bar{Y}_m, \bar{C}_m\right) \middle| \bar{C}_m\right\} \\ &= E_{\bar{Y}_m}\left(E_{\bar{R}_{j-1}, \bar{R}_{j-1}^D}\left[\frac{E_{R_j, R_j^D}\left\{R_j R_j^D \middle| \bar{R}_{j-1}, \bar{R}_{j-1}^D, \bar{Y}_m, \bar{C}_m\right\}}{\pi_j}\left\{Y_j - E\left(Y_j \middle| \bar{Y}_{m(-j)}^{\text{obs}}, \bar{C}_m\right)\right\} + \right.\right. \\ &\qquad\qquad \left.\left. E\left\{Y_j \middle| \bar{Y}_{m(-j)}^{\text{obs}}, \bar{C}_m\right\} \middle| \bar{Y}_m, \bar{C}_m\right] \middle| \bar{C}_m\right) \\ &= E_{\bar{Y}_m}\left[E_{\bar{R}_{j-1}, \bar{R}_{j-1}^D}\left\{Y_j - E\left(Y_j \middle| \bar{Y}_{m(-j)}^{\text{obs}}, \bar{C}_m\right) + E\left(Y_j \middle| \bar{Y}_{m(-j)}^{\text{obs}}, \bar{C}_m\right) \middle| \bar{Y}_m, \bar{C}_m\right\} \middle| \bar{C}_m\right] \\ &= E_{\bar{Y}_m}\left(Y_j \middle| \bar{C}_m\right) \\ &= E\left(Y_j \middle| \bar{C}_m\right) \\ &= E\left(Y_j \middle| S = s\right).\end{aligned}$$

For derivational clarity, let $R_j^M = \left\{R_j, R_j^D\right\}$, $\bar{R}_j^M = \left\{\bar{R}_j, \bar{R}_j^D\right\}$, and $R_j^{M\times} = R_j R_j^D$. Now suppose π_j is incorrectly modeled. Recall that π_j is a function of $\left\{\bar{R}_{j-1}, \bar{R}_{j-1}^D, \bar{Y}_{j-1}^{\text{obs}}, C_{j+1}, C_j\right\}$ under DAR

and IMAR.

If instead the OR model is correct but the missingness model is wrong, then

$$
\begin{aligned}
&E\left(W_j\middle|S=s\right)\\
&=E\left(W_j\middle|\bar{C}_m\right)\\
&=E_{\bar{Y}^{\text{obs}}_{m(-j)},Y_j,R_j^M,\bar{R}^M_{j-1}}\left[\frac{R_j^{M\times}}{\pi_j}\left\{Y_j-E\left(Y_j\middle|\bar{Y}^{\text{obs}}_{m(-j)},\bar{C}_m\right)\right\}+E\left\{Y_j\middle|\bar{Y}^{\text{obs}}_{m(-j)},\bar{C}_m\right\}\middle|\bar{C}_m\right]\\
&=E_{\bar{Y}^{\text{obs}}_{m(-j)}}\left(E_{Y_j,R_j^M,\bar{R}^M_{j-1}}\left[\frac{R_j^{M\times}}{\pi_j}\left\{Y_j-E\left(Y_j\middle|\bar{Y}^{\text{obs}}_{m(-j)},\bar{C}_m\right)\right\}+E\left\{Y_j\middle|\bar{Y}^{\text{obs}}_{m(-j)},\bar{C}_m\right\}\middle|\bar{Y}^{\text{obs}}_{m(-j)},\bar{C}_m\right]\middle|\bar{C}_m\right)\\
&=E_{\bar{Y}^{\text{obs}}_{m(-j)}}\left(E_{Y_j,R_j^M,\bar{R}^M_{j-1}}\left[\frac{R_j^{M\times}}{\pi_j}\left\{Y_j-E\left(Y_j\middle|\bar{Y}^{\text{obs}}_{m(-j)},\bar{C}_m\right)\right\}\middle|\bar{Y}^{\text{obs}}_{m(-j)},\bar{C}_m\right]+E\left[Y_j\middle|\bar{Y}^{\text{obs}}_{m(-j)},\bar{C}_m\right]\middle|\bar{C}_m\right)\\
&=E_{\bar{Y}^{\text{obs}}_{m(-j)}}\left\{E_{\bar{R}^M_{j-1}}\left(E_{Y_j,R_j^M}\left[\frac{R_j^{M\times}}{\pi_j}\left\{Y_j-E\left(Y_j\middle|\bar{Y}^{\text{obs}}_{m(-j)},\bar{C}_m\right)\right\}\middle|\bar{R}^M_{j-1},\bar{Y}^{\text{obs}}_{m(-j)},\bar{C}_m\right]\middle|\bar{Y}^{\text{obs}}_{m(-j)},\bar{C}_m\right)+\right.\right.\\
&\qquad\left.\left.E\left(Y_j\middle|\bar{Y}^{\text{obs}}_{m(-j)},\bar{C}_m\right)\middle|\bar{C}_m\right\}\\
&=E_{\bar{Y}^{\text{obs}}_{m(-j)}}\left[E_{\bar{R}^M_{j-1}}\left\{E_{Y_j}\left(E_{R_j^M}\left[\frac{R_j^{M\times}}{\pi_j}\left\{Y_j-E\left(Y_j\middle|\bar{Y}^{\text{obs}}_{m(-j)},\bar{C}_m\right)\right\}\middle|\bar{R}^M_{j-1},\bar{Y}^{\text{obs}}_{m(-j)},Y_j,\bar{C}_m\right]\middle|\bar{R}^M_{j-1},\bar{Y}^{\text{obs}}_{m(-j)},\bar{C}_m\right)\middle|\bar{Y}^{\text{obs}}_{m(-j)},\bar{C}_m\right\}+\right.\right.\\
&\qquad\left.\left.E\left\{Y_j\middle|\bar{Y}^{\text{obs}}_{m(-j)},\bar{C}_m\right\}\middle|\bar{C}_m\right]\\
&=E_{\bar{Y}^{\text{obs}}_{m(-j)}}\left\{E_{\bar{R}^M_{j-1}}\left(E_{Y_j}\left[\frac{\pi_j^0}{\pi_j}\left\{Y_j-E\left(Y_j\middle|\bar{Y}^{\text{obs}}_{m(-j)},\bar{C}_m\right)\right\}\middle|\bar{R}^M_{j-1},\bar{Y}^{\text{obs}}_{m(-j)},\bar{C}_m\right]\middle|\bar{Y}^{\text{obs}}_{m(-j)},\bar{C}_m\right)+E\left(Y_j\middle|\bar{Y}^{\text{obs}}_{m(-j)},\bar{C}_m\right)\middle|\bar{C}_m\right\}\\
&=E_{\bar{Y}^{\text{obs}}_{m(-j)}}\left(E_{\bar{R}^M_{j-1}}\left[\frac{\pi_j^0}{\pi_j}E_{Y_j}\left\{Y_j-E\left(Y_j\middle|\bar{Y}^{\text{obs}}_{m(-j)},\bar{C}_m\right)\middle|\bar{R}^M_{j-1},\bar{Y}^{\text{obs}}_{m(-j)},\bar{C}_m\right\}\middle|\bar{Y}^{\text{obs}}_{m(-j)},\bar{C}_m\right]+E\left[Y_j\middle|\bar{Y}^{\text{obs}}_{m(-j)},\bar{C}_m\right]\middle|\bar{C}_m\right)\\
&\quad\text{under DAR and IMAR}\\
&=E_{\bar{Y}^{\text{obs}}_{m(-j)}}\left(E_{\bar{R}^M_{j-1}}\left[\frac{\pi_j^0}{\pi_j}\left\{E_{Y_j}\left(Y_j\middle|\bar{R}^M_{j-1},\bar{Y}^{\text{obs}}_{m(-j)},\bar{C}_m\right)-E\left(Y_j\middle|\bar{Y}^{\text{obs}}_{m(-j)},\bar{C}_m\right)\right\}\middle|\bar{Y}^{\text{obs}}_{m(-j)},\bar{C}_m\right]+E\left[Y_j\middle|\bar{Y}^{\text{obs}}_{m(-j)},\bar{C}_m\right]\middle|\bar{C}_m\right),
\end{aligned}
$$

where

$$
\begin{aligned}
\pi_j^0&=E_{R_j^M}\left(R_j^{M\times}\middle|\bar{R}^M_{j-1},\bar{Y}^{\text{obs}}_{m(-j)},Y_j,\bar{C}_m\right)\\
&=E_{\bar{Y}^{\text{mis}}_{m(-j)}}\left\{E_{R_j^M}\left(R_j^{M\times}\middle|\bar{R}^M_{j-1},\bar{Y}^{\text{mis}}_{m(-j)},\bar{Y}^{\text{obs}}_{m(-j)},Y_j,\bar{C}_m\right)\middle|\bar{R}^M_{j-1},\bar{Y}^{\text{obs}}_{m(-j)},Y_j,\bar{C}_m\right\}\\
&=E_{\bar{Y}^{\text{mis}}_{m(-j)}}\left\{E_{R_j^M}\left(R_j^{M\times}\middle|\bar{R}^M_{j-1},\bar{Y}_m,\bar{C}_m\right)\middle|\bar{R}^M_{j-1},\bar{Y}^{\text{obs}}_{m(-j)},Y_j,\bar{C}_m\right\}\\
&=E_{\bar{Y}^{\text{mis}}_{m(-j)}}\left\{E_{R_j^M}\left(R_j^{M\times}\middle|\bar{R}^M_{j-1},\bar{Y}^{\text{obs}}_{j-1},C_{j+1},C_j\right)\middle|\bar{R}^M_{j-1},\bar{Y}^{\text{obs}}_{m(-j)},Y_j,\bar{C}_m\right\}\text{ under DAR and IMAR}\\
&=E_{R_j^M}\left(R_j^{M\times}\middle|\bar{R}^M_{j-1},\bar{Y}^{\text{obs}}_{j-1},C_{j+1},C_j\right).
\end{aligned}
$$

Note that $\pi_j^0/\pi_j\neq 1$ in general because π_j is modeled incorrectly. We will show that $E\left(Y_j\middle|\bar{R}^M_{j-1},\bar{Y}^{\text{obs}}_{m(-j)},\bar{C}_m\right)=E\left(Y_j\middle|\bar{Y}^{\text{obs}}_{m(-j)},\bar{C}_m\right)$. For a random variable A and corresponding fixed realization a, let $\underline{\text{A}}_j\ =$

$\left(A_j, A_{j+1}, \ldots, A_m\right)$. Note that

$$
\begin{aligned}
f\left(y_j \middle| \bar{r}_{j-1}^M, \bar{y}_{m(-j)}^{\text{obs}}, \bar{c}_m\right) &= f\left(y_j \middle| \bar{y}_{m(-j)}^{\text{obs}}, \bar{c}_m\right) \\
\frac{p\left(\bar{r}_{j-1}^M, y_j, \bar{y}_{m(-j)}^{\text{obs}}, \bar{c}_m\right)}{p\left(\bar{r}_{j-1}^M, \bar{y}_{m(-j)}^{\text{obs}}, \bar{c}_m\right)} &= \frac{p\left(y_j, \bar{y}_{m(-j)}^{\text{obs}}, \bar{c}_m\right)}{p\left(\bar{y}_{m(-j)}^{\text{obs}}, \bar{c}_m\right)} \\
\frac{\int p\left(\bar{r}_{j-1}^M, y_j, \bar{y}_{m(-j)}^{\text{obs}}, \bar{y}_{m(-j)}^{\text{mis}}, \bar{c}_m\right) d\bar{y}_{m(-j)}^{\text{mis}}}{\int\int p\left(\bar{r}_{j-1}^M, y_j, \bar{y}_{m(-j)}^{\text{obs}}, \bar{y}_{m(-j)}^{\text{mis}}, \bar{c}_m\right) d\bar{y}_{m(-j)}^{\text{mis}} dy_j} &= \frac{\int p\left(y_j, \bar{y}_{m(-j)}^{\text{obs}}, \bar{y}_{m(-j)}^{\text{mis}}, \bar{c}_m\right) d\bar{y}_{m(-j)}^{\text{mis}}}{\int\int p\left(y_j, \bar{y}_{m(-j)}^{\text{obs}}, \bar{y}_{m(-j)}^{\text{mis}}, \bar{c}_m\right) d\bar{y}_{m(-j)}^{\text{mis}} dy_j} \\
\frac{\int p\left(\bar{r}_{j-1}^M, \bar{y}_m, \bar{c}_m\right) d\bar{y}_{m(-j)}^{\text{mis}}}{\int\int p\left(\bar{r}_{j-1}^M, \bar{y}_m, \bar{c}_m\right) d\bar{y}_{m(-j)}^{\text{mis}} dy_j} &= \frac{\int p\left(\bar{y}_m, \bar{c}_m\right) d\bar{y}_{m(-j)}^{\text{mis}}}{\int\int p\left(\bar{y}_m, \bar{c}_m\right) d\bar{y}_{m(-j)}^{\text{mis}} dy_j} \\
\frac{\sum_{\underline{\mathrm{r}}_j^M} \int p\left(\bar{r}_{j-1}^M, \underline{\mathrm{r}}_j^M, \bar{y}_m, \bar{c}_m\right) d\bar{y}_{m(-j)}^{\text{mis}}}{\sum_{\underline{\mathrm{r}}_j^M} \int\int p\left(\bar{r}_{j-1}^M, \underline{\mathrm{r}}_j^M, \bar{y}_m, \bar{c}_m\right) d\bar{y}_{m(-j)}^{\text{mis}} dy_j} &= \frac{\int p\left(\bar{y}_m, \bar{c}_m\right) d\bar{y}_{m(-j)}^{\text{mis}}}{\int\int p\left(\bar{y}_m, \bar{c}_m\right) d\bar{y}_{m(-j)}^{\text{mis}} dy_j} \\
\frac{\sum_{\underline{\mathrm{r}}_j^M} \int p\left(\bar{r}_m^M, \bar{y}_m, \bar{c}_m\right) d\bar{y}_{m(-j)}^{\text{mis}}}{\sum_{\underline{\mathrm{r}}_j^M} \int\int p\left(\bar{r}_m^M, \bar{y}_m, \bar{c}_m\right) d\bar{y}_{m(-j)}^{\text{mis}} dy_j} &= \frac{\int p\left(\bar{y}_m, \bar{c}_m\right) d\bar{y}_{m(-j)}^{\text{mis}}}{\int\int p\left(\bar{y}_m, \bar{c}_m\right) d\bar{y}_{m(-j)}^{\text{mis}} dy_j} \\
\frac{\sum_{\underline{\mathrm{r}}_j^M} \int p\left(\bar{r}_m^M \middle| \bar{y}_m, \bar{c}_m\right) p\left(\bar{y}_m, \bar{c}_m\right) d\bar{y}_{m(-j)}^{\text{mis}}}{\sum_{\underline{\mathrm{r}}_j^M} \int\int p\left(\bar{r}_m^M \middle| \bar{y}_m, \bar{c}_m\right) p\left(\bar{y}_m, \bar{c}_m\right) d\bar{y}_{m(-j)}^{\text{mis}} dy_j} &= \frac{\int p\left(\bar{y}_m, \bar{c}_m\right) d\bar{y}_{m(-j)}^{\text{mis}}}{\int\int p\left(\bar{y}_m, \bar{c}_m\right) d\bar{y}_{m(-j)}^{\text{mis}} dy_j} \\
\frac{\sum_{\underline{\mathrm{r}}_j^M} \int p\left(\bar{r}_m^M \middle| \bar{y}_m^{\text{obs}}, \bar{c}_m\right) p\left(\bar{y}_m, \bar{c}_m\right) d\bar{y}_{m(-j)}^{\text{mis}}}{\sum_{\underline{\mathrm{r}}_j^M} \int\int p\left(\bar{r}_m^M \middle| \bar{y}_m^{\text{obs}}, \bar{c}_m\right) p\left(\bar{y}_m, \bar{c}_m\right) d\bar{y}_{m(-j)}^{\text{mis}} dy_j} &= \frac{\int p\left(\bar{y}_m, \bar{c}_m\right) d\bar{y}_{m(-j)}^{\text{mis}}}{\int\int p\left(\bar{y}_m, \bar{c}_m\right) d\bar{y}_{m(-j)}^{\text{mis}} dy_j} \text{ under DAR and IMAR} \\
\frac{\sum_{\underline{\mathrm{r}}_j^M} p\left(\bar{r}_m^M \middle| \bar{y}_m^{\text{obs}}, \bar{c}_m\right) \int p\left(\bar{y}_m, \bar{c}_m\right) d\bar{y}_{m(-j)}^{\text{mis}}}{\sum_{\underline{\mathrm{r}}_j^M} \int\int p\left(\bar{r}_m^M \middle| \bar{y}_m^{\text{obs}}, \bar{c}_m\right) p\left(\bar{y}_m, \bar{c}_m\right) d\bar{y}_{m(-j)}^{\text{mis}} dy_j} &= \frac{\int p\left(\bar{y}_m, \bar{c}_m\right) d\bar{y}_{m(-j)}^{\text{mis}}}{\int\int p\left(\bar{y}_m, \bar{c}_m\right) d\bar{y}_{m(-j)}^{\text{mis}} dy_j} \\
\frac{\sum_{\underline{\mathrm{r}}_j^M} p\left(\bar{r}_m^M \middle| \bar{y}_m^{\text{obs}}, \bar{c}_m\right)}{\sum_{\underline{\mathrm{r}}_j^M} \int\int p\left(\bar{r}_m^M \middle| \bar{y}_m^{\text{obs}}, \bar{c}_m\right) p\left(\bar{y}_m, \bar{c}_m\right) d\bar{y}_{m(-j)}^{\text{mis}} dy_j} &= \frac{1}{\int\int p\left(\bar{y}_m, \bar{c}_m\right) d\bar{y}_{m(-j)}^{\text{mis}} dy_j}.
\end{aligned}
$$

If $y_j \in \bar{y}_m^{\text{mis}}$ (i.e., $y_j \notin \bar{y}_m^{\text{obs}}$), then

$$
\begin{aligned}
\frac{\sum_{\underline{\mathrm{r}}_j^M} p\left(\bar{r}_m^M \middle| \bar{y}_m^{\text{obs}}, \bar{c}_m\right)}{\sum_{\underline{\mathrm{r}}_j^M} p\left(\bar{r}_m^M \middle| \bar{y}_m^{\text{obs}}, \bar{c}_m\right) \int\int p\left(\bar{y}_m, \bar{c}_m\right) d\bar{y}_{m(-j)}^{\text{mis}} dy_j^{\text{mis}}} &= \frac{1}{\int\int p\left(\bar{y}_m, \bar{c}_m\right) d\bar{y}_{m(-j)}^{\text{mis}} dy_j^{\text{mis}}} \\
\frac{\sum_{\underline{\mathrm{r}}_j^M} p\left(\bar{r}_m^M \middle| \bar{y}_m^{\text{obs}}, \bar{c}_m\right)}{\sum_{\underline{\mathrm{r}}_j^M} p\left(\bar{r}_m^M \middle| \bar{y}_m^{\text{obs}}, \bar{c}_m\right)} &= 1 \\
\sum_{\underline{\mathrm{r}}_j^M} p\left(\bar{r}_m^M \middle| \bar{y}_m^{\text{obs}}, \bar{c}_m\right) &= \sum_{\underline{\mathrm{r}}_j^M} p\left(\bar{r}_m^M \middle| \bar{y}_m^{\text{obs}}, \bar{c}_m\right),
\end{aligned}
$$

a tautology. If on the other hand $y_j \in \bar{y}_m^{\text{obs}}$, then

$$\frac{\sum_{\underline{\mathrm{r}}_j^M} p\left(\bar{r}_m^M \middle| \bar{y}_m^{\text{obs}}, \bar{c}_m\right)}{\sum_{\underline{\mathrm{r}}_j^M} \int p\left(\bar{r}_m^M \middle| \bar{y}_m^{\text{obs}}, \bar{c}_m\right) \int p\left(\bar{y}_m, \bar{c}_m\right) d\bar{y}_m^{\text{mis}} dy_j^{\text{obs}}} = \frac{1}{\int \int p\left(\bar{y}_m, \bar{c}_m\right) d\bar{y}_m^{\text{mis}} dy_j^{\text{obs}}}$$

$$\sum_{\underline{\mathrm{r}}_j^M} p\left(\bar{r}_m^M \middle| \bar{y}_m^{\text{obs}}, \bar{c}_m\right) \int \int p\left(\bar{y}_m, \bar{c}_m\right) d\bar{y}_m^{\text{mis}} dy_j^{\text{obs}} = \sum_{\underline{\mathrm{r}}_j^M} \int p\left(\bar{r}_m^M \middle| \bar{y}_m^{\text{obs}}, \bar{c}_m\right) \int p\left(\bar{y}_m, \bar{c}_m\right) d\bar{y}_m^{\text{mis}} dy_j^{\text{obs}}$$

$$\sum_{\underline{\mathrm{r}}_j^M} p\left(\bar{r}_m^M \middle| \bar{y}_m, \bar{c}_m\right) \int \int p\left(\bar{y}_m, \bar{c}_m\right) d\bar{y}_m^{\text{mis}} dy_j^{\text{obs}} = \sum_{\underline{\mathrm{r}}_j^M} \int p\left(\bar{r}_m^M \middle| \bar{y}_m^{\text{obs}}, \bar{c}_m\right) \int p\left(\bar{y}_m, \bar{c}_m\right) d\bar{y}_m^{\text{mis}} dy_j^{\text{obs}}$$

under DAR and IMAR

$$\sum_{\underline{\mathrm{r}}_j^M} \int \int p\left(\bar{r}_m^M \middle| \bar{y}_m, \bar{c}_m\right) p\left(\bar{y}_m, \bar{c}_m\right) d\bar{y}_m^{\text{mis}} dy_j^{\text{obs}} = \sum_{\underline{\mathrm{r}}_j^M} \int p\left(\bar{r}_m^M \middle| \bar{y}_m^{\text{obs}}, \bar{c}_m\right) \int p\left(\bar{y}_m, \bar{c}_m\right) d\bar{y}_m^{\text{mis}} dy_j^{\text{obs}}$$

$$\sum_{\underline{\mathrm{r}}_j^M} \int \int p\left(\bar{r}_m^M, \bar{y}_m, \bar{c}_m\right) d\bar{y}_m^{\text{mis}} dy_j^{\text{obs}} = \sum_{\underline{\mathrm{r}}_j^M} \int p\left(\bar{r}_m^M \middle| \bar{y}_m^{\text{obs}}, \bar{c}_m\right) \int p\left(\bar{y}_m, \bar{c}_m\right) d\bar{y}_m^{\text{mis}} dy_j^{\text{obs}}$$

$$\sum_{\underline{\mathrm{r}}_j^M} \int \int p\left(\bar{r}_m^M, \bar{y}_m, \bar{c}_m\right) d\bar{y}_m^{\text{mis}} dy_j^{\text{obs}} = \sum_{\underline{\mathrm{r}}_j^M} \int p\left(\bar{r}_m^M \middle| \bar{y}_m^{\text{obs}}, \bar{c}_m\right) p\left(\bar{y}_m^{\text{obs}}, \bar{c}_m\right) dy_j^{\text{obs}}$$

$$\sum_{\underline{\mathrm{r}}_j^M} \int p\left(\bar{r}_m^M, \bar{y}_m^{\text{obs}}, \bar{c}_m\right) dy_j^{\text{obs}} = \sum_{\underline{\mathrm{r}}_j^M} \int p\left(\bar{r}_m^M \middle| \bar{y}_m^{\text{obs}}, \bar{c}_m\right) p\left(\bar{y}_m^{\text{obs}}, \bar{c}_m\right) dy_j^{\text{obs}}$$

$$\sum_{\underline{\mathrm{r}}_j^M} \int p\left(\bar{r}_m^M \middle| \bar{y}_m^{\text{obs}}, \bar{c}_m\right) p\left(\bar{y}_m^{\text{obs}}, \bar{c}_m\right) dy_j^{\text{obs}} = \sum_{\underline{\mathrm{r}}_j^M} \int p\left(\bar{r}_m^M \middle| \bar{y}_m^{\text{obs}}, \bar{c}_m\right) p\left(\bar{y}_m^{\text{obs}}, \bar{c}_m\right) dy_j^{\text{obs}},$$

another tautology. Hence, working backwards from either tautology (i.e., if $y_j \in \bar{y}_m^{\text{mis}}$ or $y_j \in \bar{y}_m^{\text{obs}}$), we conclude that

$$\begin{aligned}
&E\left(W_j \middle| S = s\right) \\
&= E_{\bar{Y}_{m(-j)}^{\text{obs}}}\left(E_{\bar{R}_{j-1}^M}\left[\frac{\pi_j^0}{\pi_j}\left\{E_{Y_j}\left(Y_j \middle| \bar{R}_{j-1}^M, \bar{Y}_{m(-j)}^{\text{obs}}, \bar{C}_m\right) - E\left(Y_j \middle| \bar{Y}_{m(-j)}^{\text{obs}}, \bar{C}_m\right)\right\} \middle| \bar{Y}_{m(-j)}^{\text{obs}}, \bar{C}_m\right] + E\left[Y_j \middle| \bar{Y}_{m(-j)}^{\text{obs}}, \bar{C}_m\right] \middle| \bar{C}_m\right) \\
&= E_{\bar{Y}_{m(-j)}^{\text{obs}}}\left(E_{\bar{R}_{j-1}^M}\left[\frac{\pi_j^0}{\pi_j}\left\{E_{Y_j}\left(Y_j \middle| \bar{Y}_{m(-j)}^{\text{obs}}, \bar{C}_m\right) - E\left(Y_j \middle| \bar{Y}_{m(-j)}^{\text{obs}}, \bar{C}_m\right)\right\} \middle| \bar{Y}_{m(-j)}^{\text{obs}}, \bar{C}_m\right] + E\left[Y_j \middle| \bar{Y}_{m(-j)}^{\text{obs}}, \bar{C}_m\right] \middle| \bar{C}_m\right) \\
&= E_{\bar{Y}_{m(-j)}^{\text{obs}}}\left\{E\left(Y_j \middle| \bar{Y}_{m(-j)}^{\text{obs}}, \bar{C}_m\right) \middle| \bar{C}_m\right\} \\
&= E\left(Y_j \middle| \bar{C}_m\right) \\
&= E\left(Y_j \middle| S = s\right)
\end{aligned}$$

if outcomes are DAR and IMAR.

■

REFERENCES

S. Arpadi, A. Fawzy, G. M Aldrovandi, C. Kankasa, M. Sinkala, M. Mwiya, D. M Thea, and L. Kuhn. Growth faltering due to breastfeeding cessation in uninfected children born to HIV-infected mothers in Zambia. *American Journal of Clinical Nutrition*, 90(2):344–353, 2009.

A. Azzalini. A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, 59(305):171–178, 1985.

Heejung Bang and James M. Robins. Doubly robust estimation in missing data and causal inference models. *Biometrics*, 61(4):962–973, 2005.

Anirban Basu and Willard G. Manning. Estimating lifetime or episode-of-illness costs under censoring. *Health Economics*, 19:1010–1028, 2010.

D. Basu. An essay on the logical foundations of survey sampling, part one*. pages 167–206, 1971.

L. J. Billingham and K. R. Abrams. Simultaneous analysis of quality of life and survival data. *Statistical Methods in Medical Research*, 11:25–48, 2002.

James R. Carpenter and Michael G. Kenward. A comparison of multiple imputation and doubly robust estimation for analyses with missing data. *JRSSA*, 169:571–584, 2006.

Basile Chaix, David Evans, Juan Merlo, and Etsuji Suzuki. Weighing up the dead and missing: Reflections on inverse-probability weighting and principal stratification to address truncation by death. *Epidemiology*, 23:129–131, 2012.

Charles S. Chasela, Michael G. Hudgens, Denise Jamieson, Dumbani Kayira, Mina C. Hosseinipour, Athena P. Kourtis, Francis Martinson, Gerald Tegha, Rodney J. Knight, Yusuf I. Ahmed, Deborah D. Kamwendo, Irving F. Hoffman, Sascha R. Ellington, Zebrone Kacheche, Alice Soko, Jeffrey B. Wiener, Susan A. Fiscus, Peter Kazembe, Innocent A. Mofolo, Maggie Chigwenembe, Dorothy S. Sichali, Charles M. van der Horst, and BAN Study Group. Maternal or infant antiretroviral drugs to reduce HIV-1 transmission. *The New England Journal of Medicine*, 362:2271–2281, 2010.

Baojiang Chen and Xiao-Hua Zhou. Doubly robust estimates for binary longitudinal data analysis with missing response and missing covariates. *Biometrics*, 67:830–842, 2011.

Parul Christian, Subarna K. Khatry, Joanne Katz, Elizabeth K. Pradhan, Steven C. LeClerq, Sharada Ram Shrestha, Ramesh K. Adhikari, Alfred Sommer, and Keith P. West. Effects of alternative maternal micronutrient supplements on low birth weight in rural Nepal: Double blind randomised community trial. *BMJ : British Medical Journal*, 326(7389):571–571, 2003.

Mary Culnane, MaryGlenn Fowler, Sophia S. Lee, George McSherry, Michael Brady, Karen O'Donnell, Lynne Mofenson, Steven L. Gortmaker, David E. Shapiro, Gwendolyn Scott, Eleanor Jimenez, Ellen C. Moore, Clemente Diaz, Patricia M. Flynn, Bethann Cunningham, James Oleske, and for the Pediatric AIDS Clinical Trials Group Protocol 219/076 Teams. Lack of long-term effects of in utero exposure to Zidovudine among uninfected children born to HIV-infected women. *JAMA: The Journal of the American Medical Association*, 281(2):151–157, 1999.

Eric J. B. Daza, Timothy Cupery, Michael G. Hudgens, and Amy H. Herring. Partially imputed regression for the average treatment effect. *The ARR Journal*, August 2015.

Peter Diggle, Patrick Heagerty, Kung-Yee Liang, and Scott Zeger. *Analysis of Longitudinal Data*. Oxford University Press, USA, second edition, 2002.

Peter Diggle, Daniel Farewell, and Robin Henderson. Analysis of longitudinal data with dropout: Objectives, assumptions and a proposal. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 56(5):499–550, 2007.

Carole Dufouil, Carol Brayne, and David Clayton. Analysis of longitudinal studies with death and dropout: A case study. *Statistics in Medicine*, 23(14):2215–2226, 2004.

B. E. Ellison. Two theorems for inferences about the normal distribution with applications in acceptance sampling. *Journal of the American Statistical Association*, 101:89–95, 1964.

V. L. Flax, M. E. Bentley, C. S. Chasela, D. Kayira, M. G. Hudgens, R. J. Knight, A. Soko, D. J. Jamieson, C. M. van der Horst, and L. S. Adair. Use of lipid-based nutrient supplements by HIV-infected Malawian women during lactation has no effect on infant growth from 0 to 24 weeks. *Journal of Nutrition*, 142:1350–1356, 2012.

Thomas M. Gill, Mayur M. Desai, Evelyne A. Gahbauer, Theodore R. Holford, and Christianna S. Williams. Restricted activity among community-living older persons: Incidence, precipitants, and health care utilization. *Annals of Internal Medicine*, 135:313–321, 2001.

Xu Guo and Bradley P. Carlin. Separate and joint modeling of longitudinal and event time data using standard computer packages. *The American Statistician*, 58(1):16–24, 2004.

J. Hájek. Comment on "an essay on the logical foundations of survey sampling, part one". page 200, 1971.

Sophie Hawkesworth, Andrew M. Prentice, Anthony J. C. Fulford, and Sophie E. Moore. Dietary supplementation of rural Gambian women during pregnancy does not affect body composition in offspring at 1117 years of age. *The Journal of Nutrition*, 138(12):2468–2473, 2008.

Robin Henderson, Peter Diggle, and Angela Dobson. Joint modelling of longitudinal measurements and event time data. *Biostatistics*, 1(4):465–480, 2000.

Joseph W. Hogan and Nan M. Laird. Mixture models for the joint distribution of repeated measures and event times. *Statistics in Medicine*, 16(3):239–257, 1997.

D. G. Horvitz and D. J. Thompson. A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47(260):663–685, 1952.

Bo Hu, Liang Li, and Tom Greene. Joint multiple imputation for longitudinal outcomes and clinical events that truncate longitudinal follow-up. *Statistics in Medicine*, 2015. doi: 10.1002/sim.6590.

Joseph D. Y. Kang and Joseph L. Schafer. Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. *Statistical Science*, 22(4):523–539, 2007.

Leslie Kish. *Survey Sampling*. New York: J. Wiley, 1965.

Brenda F. Kurland and Patrick J. Heagerty. Directly parameterized regression conditioning on being alive: Analysis of longitudinal data truncated by deaths. *Biostatistics*, 6(2):241 –258, 2005.

Brenda F. Kurland, Laura L. Johnson, Brian L. Egleston, and Paula H. Diehr. Longitudinal data with follow-up truncated by death: Match the analysis method to research aims. *Statistical Science*, 24(2):211, 2009.

Nan M. Laird. Missing data in longitudinal studies. *Statistics in Medicine*, 7:305–315, 1988.

Saskia le Cessie, Elisabeth G. E. de Vries, Ciska Buijs, and Wendy J. Post. Analyzing longitudinal data with patients in different disease states during follow-up and death as final state. *Statistics in Medicine*, 28:3829–3843, 2009.

Kung-Yee Liang and Scott L. Zeger. Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13–22, 1986.

D. Y. Lin, E. J. Feuer, R. Etzioni, and Y. Wax. Estimating medical costs from incomplete follow-up data. *Biometrics*, 53(2):419–434, 1997.

Roderick J. A. Little. Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association*, 88(421):125–134, 1993.

Roderick J. A. Little. Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association*, 90(431):1112–1121, 1995.

Roderick J. A. Little and Donald B. Rubin. *Statistical Analysis with Missing Data, Second Edition*. Wiley-Interscience, second edition, 2002.

Sharon L. Lohr. *Sampling: Design and Analysis*. Boston: Brooks/Cole, second edition, 2010.

Jared K. Lunceford and Marie Davidian. Stratication and weighting via the propensity score in estimation of causal treatment effects: A comparative study. *Statistics in Medicine*, 23:2937–2960, 2004.

Donna K. Pauler, Sheryl McCoy, and Carol Moinpour. Pattern mixture models for longitudinal quality of life studies in advanced stage disease. *Statistics in Medicine*, 22:795–809, 2003.

John S. Preisser, Kurt K. Lohman, and Paul J. Rathouz. Performance of weighted estimating equations for longitudinal binary data with drop-outs missing at random. *Statistics in Medicine*, 21:3035–3054, 2002.

Heather J. Ribaudo, Simon G. Thompson, and Timothy G. Allen-Mersh. A joint analysis of quality of life and survival using a random effect selection model. *Statistics in Medicine*, 19:3237–3250, 2000.

James M. Robins. Marginal structural models versus structural nested models as tools for causal inference. In M. Elizabeth Halloran and Donald Berry, editors, *The IMA Volumes in Mathematics and its Applications: Statistical Models in Epidemiology, the Environment, and Clinical Trials*, volume 116, pages 95–133. New York: Springer, 2000.

James M. Robins and Andrea Rotnitzky. Semiparametric efficiency in multivariate regression models with missing data. *Journal of the American Statistical Association*, 90(429):122–129, 1995.

James M. Robins, Andrea Rotnitzky, and Lue Ping Zhao. Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, 89(427):846–866, 1994.

James M. Robins, Andrea Rotnitzky, and Lue Ping Zhao. Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association*, 90(429):106–121, 1995.

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

Andrea Rotnitzky, James M. Robins, and Daniel O. Scharfstein. Semiparametric regression for repeated outcomes with nonignorable nonresponse. *Journal of the American Statistical Association*, 93(444):1321–1339, 1998.

Donald B. Rubin. Inference and missing data. *Biometrika*, 63(3):581–592, 1976.

Daniel O. Scharfstein, Andrea Rotnitzky, and James M. Robins. Adjusting for nonignorable drop-out using semiparametric nonresponse models. *Journal of the American Statistical Association*, 94(448):1096–1120, 1999a.

Daniel O. Scharfstein, Andrea Rotnitzky, and James M. Robins. Rejoinder. *Journal of the American Statistical Association*, 94(448):1135–1146, 1999b.

Mark D. Schluchter. Methods for the analysis of informatively censored longitudinal data. *Statistics in Medicine*, 11:1861–1870, 1992.

M. Shardell, G. E. Hicks, and L. Ferrucci. Doubly robust estimation and causal inference in longitudinal studies with dropout and truncation by death. *Biostatistics*, 16(1):155–168, 2015.

Michelle Shardell and Ram R. Miller. Weighted estimating equations for longitudinal studies with death and non-monotone missing time-dependent covariates and outcomes. *Statistics in Medicine*, 27(7):1008–1025, 2008.

StataCorp. *Stata: Release 12. Statistical Software. Stata Longitudinal Data / Panel Data Reference Manual*. College Station, TX: StataCorp LP, 2011.

StataCorp. *Stata: Release 13. Statistical Software. Stata Treatment Effects Reference Manual: Potential Outcomes / Counterfactual Outcomes*. College Station, TX: StataCorp LP, 2013.

Eric J. Tchetgen Tchetgen, M. Maria Glymour, Ilya Shpitser, and Jennifer Weuved. To weight or not to weight? on the relation between inverse-probability weighting and principal stratification for truncation by death. *Epidemiology*, 23:132–137, 2012.

Thomas R. Ten Have, Allen R. Kunselman, Erik P. Pulkstenis, and J. Richard Landis. Mixed effects logistic regression models for longitudinal binary response data with informative dropout. *Biometrics*, 54(1):367–383, 1998.

Anastasios Tsiatis. *Semiparametric Theory and Missing Data*. New York: Springer, 2006.

C. M. van der Horst, C. Chasela, Y. Ahmed, I. Hoffman, M. Hosseinipour, R. Knight, S. Fiscus, M. G. Hudgens, P. Kazembe, M. Bentley, L. Adair, E. Piwoz, F. Martinson, A. Duerr, A. Kourtis, A. E. Loeliger, B. Tohill, S. Ellington, D. Jamieson, and BAN Study team. Modifications of a large HIV prevention clinical trial to fit changing realities: a case study of the Breastfeeding, Antiretroviral, and Nutrition (BAN) protocol in Lilongwe, Malawi. *Contemporary Clinical Trials*, 30(1):24–33, 2009.

Mark J. van der Laan and James M. Robins. *Unified Methods for Censored Longitudinal Data and Causality*. New York: Springer, 2003.

Stijn Vansteelandt, Andrea Rotnitzky, and James Robins. Estimation of regression models for the mean of repeated outcomes under nonignorable nonmonotone nonresponse. *Biometrika*, 94 (4):841–860, 2007.

Stijn Vansteelandt, James Carpenter, and Michael G. Kenward. Analysis of incomplete data using inverse probability weighting and doubly robust estimators. *Methodology*, 6(1):37–48, 2010.

WHO Multicentre Growth Reference Study Group. *WHO Child Growth Standards: Length/height-for-age, weight-for-age, weight-for-length, weight-for-height and body mass index-for-age: Methods and development*. Geneva: World Health Organization, 2006.

Jeffrey M. Wooldridge. Inverse probability weighted estimation for general missing data problems. *Journal of Econometrics*, 141:1281–1301, 2007.

Margaret C. Wu and Raymond J. Carroll. Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics*, 44(1):175–188, 1988.

Xiaowei Yang and Steven Shoptaw. Assessing missing data assumptions in longitudinal studies: an example using a smoking cessation trial. *Drug and Alcohol Dependence*, 77(3):213–225, 2005.

Xiaowei Yang, Jinhui Li, and Steven Shoptaw. Imputation-based strategies for clinical trial longitudinal data with nonignorable missing values. *Statistics in Medicine*, 27:2826–2849, 2008.