Revista Colombiana de Estadística
July 2018, Volume 41, Issue 2, pp. 191 to 233
DOI: http://dx.doi.org/10.15446/rce.v41n2.63332

Construction of the Design Matrix for Generalized Linear Mixed-Effects Models in the Context of Clinical Trials of Treatment Sequences

Construcción de la matriz de diseño en modelos lineales de efectos mixtos generalizados en un contexto de ensayos clínicos de secuencias de tratamientos

Francisco J. Diaz[a]

Department of Biostatistics, The University of Kansas Medical Center, Kansas City, Kansas, United States

Abstract

The estimation of carry-over effects is a difficult problem in the design and analysis of clinical trials of treatment sequences including cross-over trials. Except for simple designs, carry-over effects are usually unidentifiable and therefore nonestimable. Solutions such as imposing parameter constraints are often unjustified and produce differing carry-over estimates depending on the constraint imposed. Generalized inverses or treatment-balancing often allow estimating main treatment effects, but the problem of estimating the carry-over contribution of a treatment sequence remains open in these approaches. Moreover, washout periods are not always feasible or ethical. A common feature of designs with unidentifiable parameters is that they do not have design matrices of full rank. Thus, we propose approaches to the construction of design matrices of full rank, without imposing artificial constraints on the carry-over effects. Our approaches are applicable within the framework of generalized linear mixed-effects models. We present a new model for the design and analysis of clinical trials of treatment sequences, called Antichronic System, and introduce some special sequences called Skip Sequences. We show that carry-over effects are identifiable only if appropriate Skip Sequences are used in the design and/or data analysis of the clinical trial. We explain how Skip Sequences can be implemented in practice, and present a method of computing the appropriate Skip Sequences. We show applications to the design of a cross-over study with 3 treatments and 3 periods, and to the data analysis of the STAR*D study of sequences of treatments for depression.

[a]PhD. E-mail: fdiaz@kumc.edu

Key words: Augmented regression; Carry-over effects; Vross-over design; Design matrix; Estimability; Generalized inverses; Generalized least squares; Identifiability; Maximum likelihood; Placebo; Quasi-likelihood; Random effects linear models; Robust fixed-effects estimators.

Resumen

La estimación de los efectos de arrastre es un problema difícil en el diseño y análisis de ensayos clínicos de secuencias de tratamientos, incluyendo ensayos cruzados. Excepto por diseños simples, estos efectos son usualmente no identificables y, por lo tanto, no estimables. La imposición de restricciones a los parámetros es a menudo no justificada y produce diferentes estimativos de los efectos de arrastre dependiendo de la restricción impuesta. Las inversas generalizadas o el balance de tratamientos a menudo permiten estimar los efectos principales de tratamiento, pero no resuelven el problema de estimar la contribución de los efectos de arrastre de una sequencia de tratamiento. Además, los períodos de lavado no siempre son factibles o éticos. Los diseños con parámetros no identificables comúnmente tienen matrices de diseño que no son de rango completo. Por lo tanto, proponemos métodos para la construcción de matrices de rango completo, sin imponer restricciones artificiales en los efectos de arrastre. Nuestros métodos son aplicables en un contexto de modelos lineales mixtos generalizados. Presentamos un nuevo modelo para el diseño y análisis de ensayos clínicos de secuencias de tratamientos, llamado Sistema Anticrónico, e introducimos secuencias de tratamiento especiales llamadas Secuencias de Salto. Demostramos que los efectos de arrastre son identificables sólo si se usan Secuencias de Salto apropiadas. Explicamos como implementar en la práctica estas secuencias, y presentamos un método para calcular las secuencias apropiadas. Presentamos aplicaciones al diseño de un estudio cruzado con 3 tratamientos y 3 períodos, y al análisis del estudio STAR*D de secuencias de tratamientos para la depresión.

Palabras clave: Cuasi-verosimilitud; diseño cruzado; efectos de arrastre; estimabilidad; estimadores robustos de efectos fijos; identificabilidad; inversas generalizadas; matriz de diseño; máxima verosimilitud; mínimos cuadrados generalizados; modelos lineales de efectos aleatorios; placebo.

1. Introduction

The estimation of carry-over effects is a challenging problem in the design and analysis of clinical trials (CTs) of treatment sequences. Parametrizations of generalized linear mixed-effects models (GLMMs) that include carry-over effects usually produce design matrices that are not of full rank (Jones & Kenward 2015). When the design matrix is not of full rank, not all linear contrasts of parameters are identifiable (Christensen 2011). In particular, except for very simple designs and parametrizations, carry-over effects are usually not identifiable, which creates difficulties in the quantification of the contribution of carry-over effects to the subjects' responses. The most common solution when the design matrix is not of full rank is to impose linear constraints that produce full rank. For instance, in cross-over trials, one assumption is that carry-over effects add up to zero (Jones

& Kenward 2015). Such constraints are usually not justified, however, in terms of the clinical phenomenon under study, and this limits parameter interpretation (Fleiss 1989). Moreover, as pointed out by Christensen (2011), different computer packages may use different linear constraints, giving the impression that programmers employ some constraints for the sole purpose of delivering a computer output. The problem is that although different linear constraints on unidentifiable parameters produce identical estimates of an identifiable parameter, they lead to different estimates of the unidentifiable parameters (Christensen 2011, p. 97).

Another common approach when the design matrix is not of full rank is to use generalized inverses (Christensen 2011, Bronson 1989). Usually, data analysts can use this approach to estimate overall differences among treatment effects. Only identifiable contrasts can be estimated in this way, however, and further elaborations on how treatments administered in earlier stages of the clinical trial affect the final response are often difficult, except for simple parametrizations that are not always justified (Fleiss 1989, Senn 2002, Jones & Kenward 2015). With this approach, therefore, the problem of examining the contribution of carry-over to the subjects' responses remains open.

The main objective of this article is to propose an approach to constructing a design matrix of full rank for clinical trials (CTs) of treatment sequences (including cross-over trials) without imposing unjustified constraints on the parameters representing carry-over effects. The goal is to achieve the identifiability of model parameters including the parameters corresponding to carry-over effects. The proposed approach is applicable within the framework of GLMMs and allows assessing the contribution of carry-over effects to the subjects' responses. With this approach, data analysts can assess not only first-order carry-over effects but also effects from treatments administered two or more stages (or periods) earlier.

Cross-over trials are the most common CTs of treatment sequences (Jones & Kenward 2015, Senn 2002, Diaz, Berg, Krebill, Welty, Gidal, Alloway & Privitera 2013, Berg, Welty, Gidal, Diaz, Krebill, Szaflarski, Dworetzky, Pollard, Elder Jr, Jiang, Jiang, Switzer & Privitera 2017, Grajales & Lopez 2006). In this article, however, the author is not only concerned with cross-over trials but also with CTs of treatment sequences in general. A representative example of a CT of treatment sequences, whose study design is not usually treated in the literature of cross-over trials, is the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) study (Fava, Rush, Trivedi, Nierenberg, Thase, Sackeim, Quitkin, Wisniewski, Lavori, Rosenbaum & Kupfer 2003). In that study, the researchers investigated adult subjects who were treated in outpatient settings for non-psychotic major depressive disorder. The primary purpose of the study was to determine which alternative treatments worked best if Citalopram monotherapy did not produce an acceptable response. The investigators examined a large number of treatments and treatment sequences including antidepressants and/or cognitive therapy and combined standard randomization procedures and a detailed protocol with a naturalistic approach that modeled real-life clinical practice. For instance, if a treatment fails, the subject and physician may choose the next treatment from a prescribed list of alternative treatments that are allowed at that stage, or the subject may be randomly assigned a treatment from a subset of the list determined by the subject

and/or the physician. The idea was to repeat this procedure until reaching an acceptable treatment.

The difficulty of modeling carry-over effects is well documented in the cross-over trials literature (Senn & Lambrou 1998, Senn, D'Angelo & Potvin 2004, Senn 2002, Abeyasekera & Curnow 1984, Fleiss 1989). For pharmacokinetic studies, the modern consensus is that minimizing carry-over effects during the design and implementation of the study is better than attempting to model these effects during the data-analysis phase (Senn 2002, Privitera, Welty, Gidal, Diaz, Krebill, Szaflarski, Dworetzky, Pollard, Elder, Jiang, Jiang & Berg 2016). In particular, specialists recommend washout periods in order to rationally exclude the possibility of carry-over effects, followed by regression modeling that does not include carry-over effect parameters [Senn(2002), Center for Drug Evaluation and Research(2001, 2003)]. In pharmacodynamic or behavioral studies, however, the situation is more complex. Although washout periods of sufficient length may help eliminate the carry-over effects of some therapies, the permanence of carry-over effects cannot always be ruled out when cognitive or learning processes are involved in the therapy (Hofmann, Wrobel, Kessner & Bingel 2014). Moreover, washout periods may not be ethical or convenient in some cases. For instance, patients may not agree to stop receiving treatment temporarily because they believe that doing this may exacerbate their symptoms.

In some cases, in cross-over trials, a balance of treatments across periods can be used to eliminate the influence of carry-over effects on treatment comparisons (Senn 2002). Even in these cases, however, an evaluation of how one treatment administered in one period affects the responses measured in subsequent periods cannot be conducted if the carry-over effects are not identifiable. When the presence of carry-over effects cannot be rationally ruled-out, another common solution is to include only first-order carry-over effects in the modeling equations. However, this type of carry-over effect, despite its widespread use in data analyses of pharmacological trials, may not be consistent with standard pharmacokinetic theory (Senn 2002).

Mixed-effects regression modeling is a well-established approach in statistical practice (Frees 2004, Diaz et al. 2013, Berg et al. 2017, Diaz 2016). In particular, the approach is frequently used in the analysis of clinical trials collecting repeated measures and/or longitudinal data. This article presents new developments in GLMMs that have not been published before. These developments are applicable to continuous and discrete responses, including dichotomous and count responses. Our approach is applicable to pharmacological and non-pharmacological therapies and combinations of these. It can also accommodate the same treatment being administered at different stages (or periods) of the trial to some subjects as well as different treatment durations at a particular stage. In addition, it allows the examination of whether clinical decisions made at a particular sequence stage affect the subjects' responses in future stages as well as the investigation of delayed treatment effects. For instance, for the STAR*D study, we can answer the question: Is the benefit of cognitive therapy administered in the first stage still present in the second stage which consisted of only antidepressant drugs?

GLMMs are a combination of two general modeling frameworks, which are very popular in statistical practice: random effects linear regression models (RELMs) and generalized linear models (GLMs) (see the review by Rabe-Hesketh & Skrondal 2009). RELMs (Henderson 1953, Laird & Ware 1982) consist of the classic linear regression model with the additional assumption that some of the regression coefficients are random variables. RELMs have had a dramatic impact on the statistical analysis of repeated measures and longitudinal data (Frees 2004, Fitzmaurice & Molenberghs 2004, Hooks, Marx, Kachman & Pedersen 2009, Diaz et al. 2013, Diaz 2017). A GLM (Nelder & Wedderburn 1972) is essentially a linear regression model whose response has a distribution that belongs to an exponential family of distributions with a possibly transformed mean response that depends linearly on covariates. The richness of GLMs comes of the profound statistical properties of exponential families (see Shao 2003) and the fact that many useful families are exponential (e.g. normal, Poisson and Bernoulli). Popular GLMMs are the Poisson regression model with random effects (Long, Preisser, Herring & Golin 2015), the logistic regression model with random effects (Vermunt 2005), and RELMs.

The most popular estimation methods for GLMMs are maximum likelihood, quasi-likelihood, and generalized least squares (GLS) (Breslow & Clayton 1993, Rabe-Hesketh, Skrondal & Pickles 2005, Frees 2004), but robust methods such as the generalized method of moments (Kim & Frees 2007) and robust fixed-effects estimation (Hausman 1978, Mundlak 1978, Frees 2001, Kim & Frees 2006, Ebbes, Bockenholt & Wedel 2004, Frees 2004) also exist. These methods can estimate only identifiable model parameters (Christensen 2011). Thus, if the design matrix is not of full rank, users cannot estimate all parameters. Therefore, ideally, the design matrix should be of full rank to ensure the estimability of all parameters and contrasts of interest. As illustrated in this article, the definition of covariates in the context of clinical trials of treatment sequences is not a trivial algebraic problem if we want to obtain a design matrix of full rank that includes all carry-over effects of interest. In this article, we propose a new model, called Antichronic System (AS), along with new experimental design and analysis strategies that assist in the construction of design matrices of full rank. Once users implement the proposed tools, they can utilize the usual estimation methods.

In addition to the difficulties of building an appropriate design matrix, we face another challenge in the application of GLMMs to clinical trials of treatment sequences. The treatments administered at some stages may be determined by (unmeasured or measured) subjects' characteristics and/or by their responses to treatments administered in previous stages. Problems such as self-selection, simultaneity, and omitted variables may occur (Kim & Frees 2006, Ebbes et al. 2004). Such problems can occur, for instance, when the subjects are not randomly assigned to the sequences. They cause correlations between random effects and covariates and, consequently, if data analysts employ inappropriate estimation methods, they bias the estimators of GLMM parameters. Fortunately, the econometrics literature contains well-established solutions for analogous problems. In the case of RELMs, the most popular solutions are robust fixed-effect estimation, which provides unbiased estimates of regression coefficients for stage dependent

covariates (Hausman 1978, Kim & Frees 2006), and augmented regression, which provides unbiased coefficient estimates for time invariant covariates such as demographics (Mundlak 1978, Frees 2001, 2004). These are the estimation methods we will adopt for the analysis of the STAR*D data. On the other hand, if reasons exist to believe that these issues did not occur, standard maximum likelihood, quasi-likelihood or GLS may be appropriate.

Due to the novelty of our approach and to ensure mathematical coherence, rigorous definitions of the new concepts will be necessary. In sections 2-3, we introduce some special sequences of treatments, called skip sequences (SSs), and propose that SSs be incorporated in the experimental design and analysis of trials because they guarantee the existence of design matrices of full rank. Section 4 introduces antichronic systems (ASs), which are models of treatment sequences. Section 5 connects ASs with GLMMs. Section 6 presents our main results, which show the important role of SSs in the construction of design matrices. Theorem 2 shows that the identifiability of all model parameters is guaranteed only if SSs are implemented. Theorems 3-4 establish methods for implementing SSs in practice. Section 7 presents a test to examine whether the responses obtained in a treatment stage (or period) are influenced by a previous stage. Section 8 shows an application to the design of cross-over trials with 3 treatments and 3 periods. Sections 9-10 analyze the STAR*D data using ASs. See proofs of theorems in Appendix A. To understand some of the proofs, the reader must have a familiarity with advanced algebraic treatments of GLMMs (for instance, Frees 2004).

2. Introduction to the Strategies to Obtain a Design Matrix of Full Rank

This and the next section present a convenient notation representing the process of administering a sequence of medical or behavioral treatments (MBTs) to a subject (or a patient) suffering from a chronic disease. Let Y be a (possibly transformed) variable measuring or indicating some aspect of the state of a particular chronic disorder of a subject (or patient). Assume the goal of a MBT is to modify the value of Y. Examples of Y are: 1) a dichotomous indicator of whether or not a child with autism has appropriate expressive language skills; 2) the positive or negative subscales of the PANSS syndrome scale, which measure schizophrenia severity; 3) blood glucose or cholesterol concentrations, which partially assess the metabolic syndrome; 4) the 17-item Hamilton Rating Scale for Depression (HAM-D17); or 5) the total number of abstemious days (NADs) in a month, which assesses chronic alcoholism.

Assume the CT consists of *at most* q stages, $q \geq 2$. (In a cross-over trial, a stage corresponds to a period.) At Stage i, l_i different MBTs (or "decisions") $A_{i,1}, \ldots, A_{i,l_i}$ are available for administration to the subject, where $l_i \geq 2$, $i = 1, \ldots, q$; the clinician must administer one and only one of these decisions to the subject. The set of all decisions, $\{A_{1,1}, \ldots, A_{1,l_1}, \ldots, A_{q,1}, \ldots, A_{q,l_q}\}$ is called the *treatment pool,* and $A_{i,1}, \ldots, A_{i,l_i}$ are called *i-stage decisions.* Here, $A_{i,j} \neq A_{i,j'}$

for all $j \neq j'$, but the same decision can be used in two or more different stages, that is, we can have that $A_{i,j} = A_{i',j'}$ for some $i \neq i'$ and some pair (j, j'). In other words, our methods can accommodate the same treatment occurring at different stages for some subjects. Also, in our formulation, the MBTs of a particular stage can have different durations. For instance, $A_{1,1}$ can represent "administer cognitive therapy for only 1 month", $A_{1,2}$ can represent "administer cognitive therapy for 3 months", and so on.

Let $Y_{0,\omega}$ be the baseline value of Y for a subject ω, measured just before the sequence of decisions is administered. Our formulation introduces a new mathematical object, symbolized as Ϙ (read "koppa") and called the *basal backdrop*. As explained in Section 4, Ϙ acts as a zero vector in computations. The symbol Ϙ is interpreted as the existing clinical context under which $Y_{0,\omega}$ is measured, that is, the context before Stage 1 begins. By definition, $A_{ij} \neq$ Ϙ for all i and j. That is, no treatment from the treatment pool can be administered at baseline.

Ϙ allows representing symbolically two proposed alternative strategies for obtaining a design matrix of full rank. In one strategy, Ϙ represents placebo and it is assumed that placebo effects are negligible. For instance, in a CT with $q = 2$ stages, the pair $(A_{1,5}, A_{2,3})$ represents a sequence in which a subject receives treatment $A_{1,5}$ at Stage 1 and treatment $A_{2,3}$ at Stage 2. In contrast, the pair $(\text{Ϙ}, A_{2,3})$ will be used to represent the following sequence of events: the subject is administered placebo at the first stage and $A_{2,3}$ at the second stage. Similarly, if $q = 3$, the triplet $(\text{Ϙ}, \text{Ϙ}, A_{2,3})$ will indicate that the subject is administered placebo in the first two stages and treatment $A_{2,3}$ at the third stage. This strategy is called *real placebo*.

In the second strategy, called *virtual placebo*, Ϙ symbolizes a virtual absence of treatment at a particular stage, but the first treatment that the patient receives corresponds to a treatment from a subsequent stage. For instance, in a CT with $q = 2$ stages, the pair $(\text{Ϙ}, A_{2,3})$ will be used to represent the following strategy: the subject is not administered any 1-stage treatment during Stage 1, but is administered $A_{2,3}$ at Stage 1 instead. That is, the subject "skips" having a 1-stage treatment, receives the 2-stage treatment $A_{2,3}$ right after measuring $Y_{0,\omega}$, and *completes the trial at the end of Stage 1*. Although $A_{2,3}$ is implemented in this subject as if it were a 1-stage treatment, $A_{2,3}$ will be treated as a 2-stage treatment in the design matrix. In particular, the response determined by $A_{2,3}$ will be treated as a response measured at the end of Stage 2. Provided $Y_{0,\omega}$ measures a stable illness state, this strategy is methodologically sound and does not compromise the conclusions of data analysis. This is a consequence of Theorem 4. As shown in Theorems 2 and 3, the above strategies help construct design matrices of full rank.

3. Admissible and Skip Sequences

For $r = 1, ..., q$, a treatment sequence up to Stage r is viewed as an element of the Cartesian product $\mathcal{A}_r = \{\text{Ϙ}, A_{1,1}, \ldots, A_{1,l_1}\} \times \{\text{Ϙ}, A_{2,1}, \ldots, A_{2,l_2}\} \times \cdots \times \{\text{Ϙ}, A_{r,1}, \ldots, A_{r,l_r}\}$. For fix $S \in \mathcal{A}_r$, denote $S = (S_1, \ldots, S_r)$, where $S_i \in \{\text{Ϙ}, A_{i,1}, \ldots, A_{i,l_i}\}$, $i \leq r$. A sequence $S \in \mathcal{A}_r$ is called *inadmissible* if there exist i and j with

$1 \leq i < j \leq r$ such that $S_i \neq \text{Ø}$ but $S_j = \text{Ø}$, or if $S_i = \text{Ø}$ for all $i = 1, \ldots, r$. A sequence that is not inadmissible is called *admissible*. The subset of admissible sequences in $\mathcal{A}_r$ is denoted as $\mathcal{A}_r^*$. An admissible sequence $S = (S_1, \ldots, S_r)$ is called a *skip sequence* (SS) if $S_i = \text{Ø}$ for some i. An inadmissible sequence $S \neq (\text{Ø}, \ldots, \text{Ø})$ can be interpreted as a protocol violation in which a subject was *returned* at some stage ≥ 2 to baseline conditions, which are not part of the treatment pool.

Thus, any admissible sequence can be written as either

$$(A_{1,j_1}, A_{2,j_2}, \ldots, A_{r,j_r}),\ (\text{Ø}, A_{2,j_2},\ \ldots, A_{r,j_r}), \ldots,\ \text{or}\ \ (\text{Ø}, \text{Ø}, \ldots, \text{Ø}, A_{r,j_r}),$$

for some $r = 1, ..., q$ and some j_i's such that $j_i \in \{1, ..., l_i\}$. Only admissible sequences will be used in our model. Examples of admissible sequences in $\mathcal{A}_3$ are $(A_{1,2}, A_{2,2}, A_{3,1})$, $(\text{Ø}, A_{2,2}, A_{3,1})$, and $(\text{Ø}, \text{Ø}, A_{3,1})$. All these except the first are skip. In the second sequence, under a virtual-placebo strategy, the subject "skipped" receiving any 1-stage treatment, but received $A_{2,2}$ at Stage 1 and $A_{3,1}$ at Stage 2, completing the trial at the end of Stage 2. In the third sequence, the subject "skipped" receiving 1- and 2-stage treatments, but received $A_{3,1}$ at Stage 1 and completed the trial at the end of Stage 1. Examples of inadmissible sequences are $(A_{1,2}, \text{Ø}, A_{3,1}) \in \mathcal{A}_3$, $(A_{1,2}, A_{2,2}, \text{Ø}) \in \mathcal{A}_3$, $(A_{1,2}, \text{Ø}) \in \mathcal{A}_2$ and $S = (\text{Ø}, \ldots, \text{Ø})$.

4. Antichronic Systems

This section presents a new model of sequences of decisions. We assume that there is a clearly defined population of subjects to whom admissible sequences of possibly different lengths will be administered. Suppose a subject experienced (or was administered) a sequence $S \in \mathcal{A}_r^*$. Under a real placebo strategy, we assume that the subject's response Y was measured at the end of stages $1, \ldots, r$. Alternatively, under a virtual placebo strategy, we assume that the response was measured at the end of stages $m(S), m(S) + 1, \ldots, r$, where $m(S) = \min\{i; 1 \leq i \leq r, S_i \neq \text{Ø}\}$. Measured responses from subject ω are denoted as $Y_{k,\omega}$ with $k \in \{0, 1, \ldots, r\}$ under real placebo, or $k \in \{0, m(S), m(S) + 1, \ldots, r\}$ under virtual placebo. Thus, after administering S, the subject's response vector is $\boldsymbol{Y}_\omega^{(r)} = \boldsymbol{Y}_\omega^{(r)}(S) = (Y_{0,\omega}, Y_{1,\omega}, \ldots, Y_{r,\omega})^T$ under real placebo, but it is $= \big(Y_{0,\omega}, Y_{m(S),\omega}, Y_{m(S)+1,\omega}, \ldots, Y_{r,\omega}\big)^T$ under virtual placebo. Under virtual placebo, if $S_i = \text{Ø}$ for some particular i, $Y_{i,\omega}$ is undefined. However, $Y_{0,\omega}$ must be measured in both approaches.

Importantly, our model identifies decision $A_{i,j}$ with a *row vector* of dimension l_i which has 1 in its j-th component but 0 elsewhere. (That is, $A_{i,j} = (0, ..., 0, 1, 0, ..., 0) \in \mathbb{R}^{l_i}$). Also, if Ø occurs at Stage i, we identify Ø with the zero row vector of dimension l_i. [That is, $\text{Ø} = \mathbf{0} \in \mathbb{R}^{l_i}$.] Thus, a sequence $S = (S_1, \ldots, S_r)$ is viewed as a sequence of row vectors $S_1, \ldots, S_r$ such that S_i has dimension l_i.

Definition 1. A set $\mathcal{C}$ of admissible sequences is called an antichronic system (AS) if, for subject ω, there exist a unique number α_ω and unique column vectors

$\boldsymbol{\beta}_{ij,\omega}$ of dimension l_i, $i = 1, \ldots, j$, $j = 1, \ldots, q$, such that, for all $r = 1, \ldots, q$, if the subject experiences sequence $S = (S_1, \ldots, S_r) \in \mathcal{C}$, then

$$E_\omega[Y_{0,\omega}] = g^{-1}(\Lambda_\omega)\,, \quad E_\omega[Y_{k,\omega}] = g^{-1}(\Lambda_\omega + \sum_{i=1}^{k} S_i \boldsymbol{\beta}_{ik,\omega})\,, \quad \text{and}$$

$$\Lambda_\omega = \alpha_\omega + \boldsymbol{\lambda}^T \mathbf{X}_\omega \tag{1}$$

where $k = 1, \ldots, r$ if real placebo is implemented, or $k = m(S), \ldots, r$ under virtual placebo. We assume $\mathcal{C}$ includes at least one sequence in $\mathcal{A}_q^*$.

Here, g is a known link function and $E_\omega[\,\cdot\,]$ represents conditional expectation given subject ω, that is, given both Λ_ω and $\boldsymbol{\beta}_{ij,\omega}$, $i = 1, \ldots, j$, $j = 1, \ldots, q$. In addition, $\mathbf{X}_\omega$ is a $p \times 1$ vector of baseline clinical, environmental, genetic, biological or demographic covariates, and $\boldsymbol{\lambda} \in \mathbb{R}^p$. The number α_ω and the vectors $\boldsymbol{\beta}_{ij,\omega}$ are all characteristic constants of subject ω (they do not change over time). However, some of them may vary from subject to subject. Thus, at the subject population level, α_ω and $\boldsymbol{\beta}_{ij,\omega}$ are viewed as particular realizations of a random variable α and a random vector $\boldsymbol{\beta}_{ij}$, which may possibly take on constant values. It is possible to assume that $\boldsymbol{\lambda}$ varies from subject to subject, but the simpler model that assumes a fixed $\boldsymbol{\lambda}$ will suffice to introduce our ideas. For fixed ω, under real placebo, the conditional distributions of $Y_{0,\omega}, Y_{1,\omega}, \ldots, Y_{r,\omega}$ belong to the same exponential family (e.g. normal, Poisson or Bernoulli), although with possibly different parameter values. Similarly, under virtual placebo and fixed ω, the conditional distributions of $Y_{0,\omega}$ and $Y_{m(S),\omega}, \ldots, Y_{r,\omega}$ belong to the same exponential family. Importantly, observe $S_i \boldsymbol{\beta}_{ik,\omega} = 0$ if $S_i = \text{Ø}$, because Ø acts as a *zero vector* in computations (that is, $S_i = \mathbf{0}$ in that case).

To interpret equation (1), denote $\beta_{k,\omega}^S = \sum_{i=1}^{k} \theta_{ik,\omega}$ with $\theta_{ik,\omega} = S_i \boldsymbol{\beta}_{ik,\omega}$. The effect $\beta_{k,\omega}^S$ of subsequence $(S_1, \ldots, S_k)$ on $Y_{k,\omega}$ is equal to the sum: "Effect of 1-stage treatment on $Y_{k,\omega}$ (that is, $\theta_{1k,\omega}$)" plus "effect of 2-stage treatment on $Y_{k,\omega}$" ($\theta_{2k,\omega}$) plus ...plus "effect of k-stage treatment on $Y_{k,\omega}$" ($\theta_{kk,\omega}$). Thus, by denoting $\boldsymbol{\beta}_{ik,\omega} = (\beta_{ik,\omega,1}, \ldots, \beta_{ik,\omega,l_i})^T$, we see that $\beta_{ik,\omega,j}$ is the effect of decision $A_{i,j}$ on response $Y_{k,\omega}$, if decision $A_{i,j}$ was actually taken at stage $i, i \leq k$. Note that we can easily model *delayed treatment effects*. For instance, if higher values of Y represent a less severe illness, and if $\beta_{ik,\omega,j} > 0$, but $\beta_{ir,\omega,j} \leq 0$ for all r with $i \leq r < k$, then $\beta_{ik,\omega,j}$ can be interpreted as a *delayed beneficial effect* of treatment $A_{i,j}$ on response $Y_{k,\omega}$. Finally, note we are measuring all these effects with respect to baseline conditions.

The virtual placebo approach requires an additional assumption. To present the assumption, we need the concept of interchangeable sequences. Let r and t be such that $1 \leq r < t \leq q$. We say that a sequence $S = (S_1, \ldots, S_r) \in \mathcal{A}_r^*$ is interchangeable with a sequence $T \in \mathcal{A}_t^*$ if $T = (\text{Ø}, \ldots, \text{Ø}, S_1, \ldots, S_r)$. As an interpretation, the skip sequence T may represent the hypothetical situation in which the administration of S was postponed $t - r$ stages but the baseline conditions did not change during the first $t - r$ stages.

Interchangeability assumption for virtual placebo: Suppose $S \in \mathcal{A}_r^*$ and $T \in \mathcal{A}_t^*$ are interchangeable sequences. Then, for a particular subject ω, $\mathbf{Y}_\omega^{(t)}(T)$ has

the same distribution as $\mathbf{Y}_{\omega}^{(r)}(S)$. Thus, interchangeable sequences S and T will potentially produce the same responses in a subject. Consequently, the effect of treatment sequence S will not change in the hypothetical situation that we delay the administration of the sequence $t-r$ stages, provided the subject continues on the baseline clinical backdrop ($\Qoppa$) during the delay.

The practical consequence of the Interchangeability Assumption is that the effect of $T = (\Qoppa, \ldots, \Qoppa, S_1, \ldots, S_r)$ on a subject ω can be examined by administering $S = (S_1, \ldots, S_r)$ right after measuring $Y_{0,\omega}$, and labeling the obtained responses as $Y_{t-r+1,\omega}$, $Y_{t-r+2,\omega}$, $\ldots$, $Y_{t,\omega}$. That is, we do not need to delay the administration of S by administering placebo during the first $t-r$ stages, and we can use the responses to S to incorporate T in the design matrix. Thus the name "virtual placebo". We demonstrate in Section 6 that this approach is valid under the Interchangeability Assumption and that we need sequences like T in order to build a design matrix of full rank. This approach is recommendable when placebo is unfeasible or unethical, but other applications are possible as illustrated in Sections 9-10. Note that the Interchangeability Assumption is an assumption about particular sequences S and T. Not all sequences of an AS have to satisfy interchangeability assumptions in order for the virtual placebo approach to be applicable.

Finally, note that both equation (1) and the Interchangeability Assumption implicitly require that $Y_{0,\omega}$ measures a stable illness state (relative to sequence duration). That is, they are applicable if the subject's baseline state will not change rapidly in the hypothetical situation that the subject is not intervened. That subjects must enter the study in a stable condition is an inclusion criterion of many clinical trials.

5. Matrix Representation of Responses

Next we exhibit the connection between ASs and the algebraic machinery that underlies GLMMs. To motivate the connection using $q = 2$, suppose subject ω was administered $S = (S_1, S_2) \in \mathcal{A}_2^*$ with $S_1 \neq \Qoppa$. Then, by equation (1),

$$
\begin{aligned}
g\Big(E_\omega[Y_{0,\omega}]\Big) &= \Lambda_\omega + \vec{\mathbf{0}}\boldsymbol{\beta}_{11,\omega} + \vec{\mathbf{0}}\boldsymbol{\beta}_{12,\omega} + \vec{\mathbf{0}}\boldsymbol{\beta}_{22,\omega} \\
g\Big(E_\omega[Y_{1,\omega}]\Big) &= \Lambda_\omega + S_1\boldsymbol{\beta}_{11,\omega} + \vec{\mathbf{0}}\boldsymbol{\beta}_{12,\omega} + \vec{\mathbf{0}}\boldsymbol{\beta}_{22,\omega} \\
g\Big(E_\omega[Y_{2,\omega}]\Big) &= \Lambda_\omega + \vec{\mathbf{0}}\boldsymbol{\beta}_{11,\omega} + S_1\boldsymbol{\beta}_{12,\omega} + S_2\boldsymbol{\beta}_{22,\omega}
\end{aligned}
$$

$$
= \begin{bmatrix} 1 & \vec{\mathbf{0}} & \vec{\mathbf{0}} & \vec{\mathbf{0}} \\ 1 & S_1 & \vec{\mathbf{0}} & \vec{\mathbf{0}} \\ 1 & \vec{\mathbf{0}} & S_1 & S_2 \end{bmatrix}_{3\times(m_2+1)} \begin{bmatrix} \Lambda_\omega \\ \boldsymbol{\beta}_{11,\omega} \\ \boldsymbol{\beta}_{12,\omega} \\ \boldsymbol{\beta}_{22,\omega} \end{bmatrix}_{(m_2+1)\times 1} \tag{2}
$$

where the $\vec{\mathbf{0}}$'s represent row vectors, not all of them with the same dimension, and $m_2 = 2l_1 + l_2$. If $S_1 = \text{Ø}$, we treat S_1 as a zero vector; and if a virtual placebo approach is followed we do not use the equation at the second row.

In general, suppose that $S = (S_1, \ldots, S_r) \in \mathcal{A}_r^*$. If the study follows a real placebo strategy, we define

$V = V(S) =$

$$\begin{pmatrix} \vec{\mathbf{0}} & \vec{\mathbf{0}} & \vec{\mathbf{0}} & \vec{\mathbf{0}} & \cdots & \vec{\mathbf{0}} & \vec{\mathbf{0}} & \vec{\mathbf{0}} & \vec{\mathbf{0}} & \vec{\mathbf{0}} & 0 & \cdots & 0 \\ S_1 & \vec{\mathbf{0}} & \vec{\mathbf{0}} & \vec{\mathbf{0}} & \cdots & \vec{\mathbf{0}} & \vec{\mathbf{0}} & \vec{\mathbf{0}} & \vec{\mathbf{0}} & \vec{\mathbf{0}} & 0 & \cdots & 0 \\ \vec{\mathbf{0}} & S_1 & S_2 & \vec{\mathbf{0}} & \cdots & \vec{\mathbf{0}} & \vec{\mathbf{0}} & \vec{\mathbf{0}} & \vec{\mathbf{0}} & \vec{\mathbf{0}} & 0 & \cdots & 0 \\ \vec{\mathbf{0}} & \vec{\mathbf{0}} & \vec{\mathbf{0}} & S_1 & \cdots & \vec{\mathbf{0}} & \vec{\mathbf{0}} & \vec{\mathbf{0}} & \vec{\mathbf{0}} & \vec{\mathbf{0}} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \vec{\mathbf{0}} & \vec{\mathbf{0}} & \vec{\mathbf{0}} & \vec{\mathbf{0}} & \cdots & S_{r-1} & \vec{\mathbf{0}} & \vec{\mathbf{0}} & \cdots & \vec{\mathbf{0}} & 0 & \cdots & 0 \\ \vec{\mathbf{0}} & \vec{\mathbf{0}} & \vec{\mathbf{0}} & \vec{\mathbf{0}} & \cdots & \vec{\mathbf{0}} & S_1 & S_2 & \cdots & S_r & 0 & \ldots & 0 \end{pmatrix}_{(r+1)\times m_q} \tag{3}$$

where $m_q = \sum_{i=1}^{q} u_i$, with $u_i = l_1 + \cdots + l_i$. If the study implements virtual placebo and S is not skip, $V = V(S)$ is defined exactly as in equation (3); but if S is skip, V is defined as the matrix obtained after eliminating rows 2 through $m(S)$ from the matrix in (3).

The number of rows of V is equal to the dimension of the subject's response $\boldsymbol{Y}_{\omega}^{(r)}(S)$. The first row of V includes only zeros because, by equation (1), $Y_{0,\omega}$ does not depend on any investigated MBT. All V matrices need to be put on top of each other to build the study design matrix. Therefore, they must have the same number of columns. For this reason, if $r < q$, the last $m_q - m_r$ columns of V must contain only zeros. If $r = q$, the additional zero columns in equation (3) are not needed. The following theorem gives a matrix representation of responses.

Theorem 1. *Let $\mathcal{C}$ be an antichronic system and suppose that, by the end of Stage r, $r \leq q$, a particular subject ω with covariate vector X_ω has been administered sequence $S = (S_1, \ldots, S_r) \in \mathcal{C}$. Then,*

$$E_\omega\left[\boldsymbol{Y}_{\omega}^{(r)}\right] = g^{-1}\left(W_\omega \boldsymbol{\gamma}_\omega\right), \textit{where}$$

$$\boldsymbol{\gamma}_\omega = \left(\alpha_\omega, \boldsymbol{\lambda}^T, \boldsymbol{\delta}_{\omega}^{(q)}\right)^T,$$

$$\boldsymbol{\delta}_{\omega}^{(q)} = \left[\left(\boldsymbol{\beta}_{11,\omega}^T\right), \left(\boldsymbol{\beta}_{12,\omega}^T, \boldsymbol{\beta}_{22,\omega}^T\right), \left(\boldsymbol{\beta}_{13,\omega}^T, \boldsymbol{\beta}_{23,\omega}^T, \boldsymbol{\beta}_{33,\omega}^T\right), \ldots, \left(\boldsymbol{\beta}_{1q,\omega}^T, \boldsymbol{\beta}_{2q,\omega}^T, \ldots, \boldsymbol{\beta}_{qq,\omega}^T\right)\right],$$

$$W_\omega = [\mathbf{1}, \boldsymbol{X}_\omega^*, V], \quad \boldsymbol{X}_\omega^* = [\boldsymbol{X}_\omega, \ldots, \boldsymbol{X}_\omega]^T \textit{ and } \mathbf{1} = (1, \ldots, 1)^T$$

($\mathbf{X}_\omega^*$ and $\mathbf{1}$ have the same number of rows as V.) For convenience, we denote $g^{-1}\left((a_1, \ldots, a_k)^T\right) = \left(g^{-1}(a_1), \ldots, g^{-1}(a_k)\right)^T$.

Theorem 1 shows that an AS is algebraically analogous to a GLMM of repeated measures of Y. Specifically, $g\left(E_\omega\left[Y_{r,\omega}\right]\right)$ depends linearly not only on the

covariates in $\mathbf{X}_\omega$ but also on stage-dependent covariates whose values for a subject who underwent sequence $S = (S_1, \ldots, S_r)$ are given in the first m_r columns of $V(S)$. Thus, at least in principle, to estimate AS parameters, we can adapt standard software for fitting GLMMs (or RELMs in the case of Gaussian response with identity link). For example, we can use SAS GLIMMIX or MIXED procedures (SAS Institute Inc, Cary, NC, USA) or Stata's meglm and mixed commands (StataCorp LP, College Station, TX, USA). Stata's xtreg command with the option fe provides fixed-effects estimators for RELMs, which are robust to self-selection, simultaneity and omitted variables (Hausman 1978, Ebbes et al. 2004, Kim & Frees 2006). The option vce(robust) additionally provides Huber/White estimators of standard errors, which are robust to covariance matrix misspecifications (Huber et al. 1967, White 1982).

Fitting an AS requires that appropriate (stage-dependent) covariates be created, using matrix $V(S)$ as a template. The supplementary material, described in Appendix C, provides a Stata ado program written by the author, called asdesign, which creates the covariates. The help file of asdesign provides additional instructions to fit an AS.

Here, we are interested in estimating the *fixed effects* vector $E[\boldsymbol{\gamma}]$, where $\boldsymbol{\gamma} = \left(\alpha, \boldsymbol{\lambda}^T, \boldsymbol{\delta}^{(q)}\right)^T$ and

$$\boldsymbol{\delta}^{(q)} = \left[\left(\boldsymbol{\beta}_{11}^T\right), \left(\boldsymbol{\beta}_{12}^T, \boldsymbol{\beta}_{22}^T\right), \left(\boldsymbol{\beta}_{13}^T, \boldsymbol{\beta}_{23}^T, \boldsymbol{\beta}_{33}^T\right), \ldots, \left(\boldsymbol{\beta}_{1q}^T, \boldsymbol{\beta}_{2q}^T, \ldots, \boldsymbol{\beta}_{qq}^T\right)\right]$$

(We interpret $\boldsymbol{\gamma}_\omega$ as a constant vector characterizing one particular subject, which is a realization of the random vector $\boldsymbol{\gamma}$.) If the experimental design does not incorporate SSs, however, the study design matrix will not be of full rank and, therefore, $E[\boldsymbol{\gamma}]$ will not be identifiable. We show this in Theorems 2-3.

Finally, note that the Interchangeability Assumption does not participate in the proof of Theorem 1. This assumption is required to prove Theorem 4, which guarantees that the introduction of SSs for the virtual placebo approach will not affect parameter interpretation.

6. Experimental Designs and Data Analysis Strategies for Antichronic Systems

This section presents the crux of our proposal: to estimate all AS model parameters, CTs of treatment sequences must incorporate SSs in their experimental design and/or data analysis. First, we show we need SSs to render $E[\boldsymbol{\gamma}]$ identifiable. Then we explain strategies for constructing SSs in practice.

In the context of a CT, let $S^{(1)}, S^{(2)}, \ldots, S^{(d)}$ be *distinct* admissible sequences. Assume that $\mathcal{C} = \left\{S^{(1)}, S^{(2)}, \ldots, S^{(d)}\right\}$ is an AS, and that we have a sample of N subjects, each of whom experienced one sequence in $\mathcal{C}$. Suppose N_i subjects completed sequence $S^{(i)}$, $i = 1, \ldots, d$, where $\sum_{i=1}^{d} N_i = N$ and $N_i \geq 1$ for all i. For $S = (S_1, \ldots, S_r) \in \mathcal{C}$, if subject ω has covariate vector $\mathbf{X}_\omega$ and underwent sequence S, the matrix $W_\omega = W\left(S, \boldsymbol{X}_\omega\right) = [\mathbf{1}, \boldsymbol{X}_\omega^*, V(S)]$ is *the subject's design matrix* (DM), $\omega = 1, \ldots, N$.

Under the usual assumptions of GLMMs, $E[\boldsymbol{\gamma}]$ will be identifiable if and only if the *study design matrix* $K = \left(W_1^T, \ldots, W_N^T\right)^T$ is of full rank, regardless of the method used to estimate $E[\boldsymbol{\gamma}]$ (for instance, quasi-likelihood, maximum likelihood, generalized method of moments or robust fixed-effect estimation) (see Christensen 2011). The following theorem indicates the importance of SSs for the estimation of $E[\boldsymbol{\gamma}]$.

Theorem 2. *Let* $D = \left(\overline{W}\left(S^{(1)}\right)^T, \ldots, \overline{W}\left(S^{(d)}\right)^T\right)^T$, *where* $\overline{W}(S) = [\mathbf{1}, V(S)]$. *A necessary condition for* K *to be of full rank is that the following two facts simultaneously occur: (1)* $\rho(K) = \rho(D)+p$, *and (2) at least one subject experiences a treatment sequence* $S \in \mathcal{C}$ *such that* S *is skip. (Here,* $\rho(A) =$ *rank of* A.*)*

The practical consequence of Theorem 2 is that, to estimate $E[\boldsymbol{\gamma}]$ through any method implemented in statistical software for GLMMs without imposing additional restrictions on $E[\boldsymbol{\gamma}]$, the clinical trial must implement SSs. Appendix B illustrates ASs with a DM of full rank. In practice, at the design stage of a CT, we must add SSs to the set of investigated treatment sequences until D is of full rank. This sometimes can also be done at the data analysis stage of existing data, as in Section 9. With existing data, however, it is not always possible to implement an adequate amount of SSs (see below). Condition 1 of Theorem 2 is achieved when the covariates in $\boldsymbol{X}_\omega$ are not perfectly correlated with each other and the final sequence allocations.

The next theorem will be useful to find the appropriate SSs at the design stage of a CT.

Theorem 3. *Suppose* $S^{\langle 1\rangle}, \ldots, S^{\langle c\rangle}$ *are* c *distinct non-skip sequences in* $\mathcal{A}_q^*$, *which will occur with positive probability in the CT. Assume that each 1-stage treatment will be administered in the first stage of at least one of these sequences. Denote*

$$\Gamma = \left[\left(S^{\langle 1\rangle}\right)^T, \ldots, \left(S^{\langle c\rangle}\right)^T\right]^T \text{ and } \tau = \sum_{i=1}^{q} l_i - \rho(\Gamma)$$

Then, there exist τ *skip sequences* $S_{\mathfrak{Q}}^{\langle 1\rangle}, \ldots, S_{\mathfrak{Q}}^{\langle \tau\rangle}$ *such that* $D = \left(D_1^T, D_2^T\right)^T$ *is of full rank, where*

$$D_1 = \left(\overline{W}(S^{\langle 1\rangle})^T, \ldots, \overline{W}(S^{\langle c\rangle})^T\right)^T \text{ and } D_2 = \left(\overline{W}(S_{\mathfrak{Q}}^{\langle 1\rangle})^T, \ldots, \overline{W}(S_{\mathfrak{Q}}^{\langle \tau\rangle})^T\right)^T$$

The practical implication of Theorem 3 is that $E[\boldsymbol{\gamma}]$ will be identifiable and therefore estimable if the CT implements all the sequences $S^{\langle 1\rangle}, \ldots, S^{\langle c\rangle}$, $S_{\mathfrak{Q}}^{\langle 1\rangle}, \ldots, S_{\mathfrak{Q}}^{\langle \tau\rangle}$. That is, adding $S_{\mathfrak{Q}}^{\langle 1\rangle}, \ldots, S_{\mathfrak{Q}}^{\langle \tau\rangle}$ to the design will produce a design matrix of full rank. The theorem's proof actually establishes a procedure to construct $S_{\mathfrak{Q}}^{\langle 1\rangle}, \ldots, S_{\mathfrak{Q}}^{\langle \tau\rangle}$ starting from $S^{\langle 1\rangle}, \ldots, S^{\langle c\rangle}$. In essence, these SSs are taken from some of the families of SSs,

$$\mathcal{F}(A_{i,j}) = \left\{(S_1, \ldots, S_q) \in \mathcal{A}_q^* \,;\, S_i = A_{i,j} \text{ and } S_k = \mathfrak{Q} \text{ for } k < i\right\}, \quad i = 2, \ldots, q \text{ and } j = 1, \ldots, l_i$$

There are a total of $\sum_{i=2}^{q} l_i$ families, but only τ of them need to be used. To illustrate, for $q = 4$, any member of the family $\mathcal{F}(A_{2,3})$ can be written as $(\mathfrak{Q}, A_{2,3}, B, C)$, where B and C are 3-stage and 4-stage treatments.

Specifically, the procedure to construct $S_{\mathfrak{Q}}^{\langle 1\rangle}, \ldots, S_{\mathfrak{Q}}^{\langle \tau\rangle}$ consists of the following easy steps: (1) Using the treatment names, label the columns of Γ. With our notation, the labels are

$$A_{1,1}, \ldots, A_{1,l_1}, A_{2,1}, \ldots, A_{2,l_2}, \ldots, A_{q,1}, \ldots, A_{q,l_q},$$

from the left to the right column of Γ. (2) Convert Γ into an echelon matrix using only elementary row operations, and then identify the columns of the echelon that do not have a pivot. Let $A^{\langle 1\rangle}, \ldots, A^{\langle \tau\rangle}$ denote their labels. (3) Lastly, for each $k = 1, \ldots, \tau$, choose $S_{\mathfrak{Q}}^{\langle k\rangle}$ from the family $\mathcal{F}\left(A^{\langle k\rangle}\right)$. We recommend choosing the sequence in $\mathcal{F}\left(A^{\langle k\rangle}\right)$ that both can be implemented in practice and will cause the lightest burden for the subjects. For each k, the CT has to implement at least one SS from the family $\mathcal{F}\left(A^{\langle k\rangle}\right)$. Otherwise, $E[\boldsymbol{\gamma}]$ will not be estimable. Section 8 illustrates this procedure for a cross-over design.

As explained in Section 2, two general strategies to implement SSs in practice are possible: real and virtual placebo. In the real placebo strategy, the trialist implements sequence S by administering a placebo at each stage k with $k < m(S)$. The rest of this section elaborates further on virtual placebo. This strategy exploits the following theorem, which is a consequence of the Interchangeability Assumption.

Theorem 4. *Let $\mathcal{C}$ be an antichronic system and $S, T \in \mathcal{C}$. Suppose that $S = (S_1, \ldots, S_r) \in \mathcal{A}_r^*$ is interchangeable with $T \in \mathcal{A}_t^*$, where $r < t$. Then, for a specific subject ω,*

$$g^{-1}\left(W\left(S, \boldsymbol{X}_\omega\right)\boldsymbol{\gamma}_\omega\right) = g^{-1}\left(W\left(T, \boldsymbol{X}_\omega\right)\boldsymbol{\gamma}_\omega\right)$$

The practical consequence of Theorem 4 is that we can use $W(T, \boldsymbol{X}_\omega)$ in place of $W(S, \boldsymbol{X}_\omega)$ in order to code the covariates of a subject who experienced S. The theorem guarantees that doing this will not affect the interpretation of model parameters contained in $E[\boldsymbol{\gamma}]$.

As suggested by Theorem 4, $T = (\mathfrak{Q}, \ldots, \mathfrak{Q}, S_1, \ldots, S_r)$ can be implemented in a CT by administering $S = (S_1, \ldots, S_r)$ without delay after measuring $Y_{0,\omega}$, labeling the obtained r responses as $Y_{t-r+1,\omega}, Y_{t-r+2,\omega}, \ldots, Y_{t,\omega}$. For instance, suppose that $q = 3$ and that $(\mathfrak{Q}, \mathfrak{Q}, A)$ is needed, where A is a 3-stage treatment. To implement this sequence, some subjects are administered treatment A right after measuring $Y_{0,\omega}$, without a previous administration of 1-stage or 2-stage treatments. Then, the response produced by A is labeled as $Y_{3,\omega}$. Responses to 1- or 2-stage treatments will not be needed from these subjects, because the DMs of these subjects will not have rows corresponding to these stages.

Under some circumstances, we can use virtual placebo in the analysis of data from studies that did not implement SSs in their protocol. The method constructs SSs at the data-analysis phase of the study and is also based on Theorem 4. To

code the covariates of a subject who experienced S, we can use $W(T, \boldsymbol{X}_\omega)$ in place of $W(S, \boldsymbol{X}_\omega)$ if the skip sequence T is needed to construct a study design matrix of full rank and S is interchangeable with T. For instance, for $q = 2$, suppose that A is *simultaneously* a 1-stage and 2-stage treatment, and that, after providing $Y_{0,\omega}$, a particular subject was administered A at Stage 1, then the subject provided $Y_{1,\omega}$ at the end of this stage, but then he/she dropped out of the study right after that. These events can be represented as $S = (A) \in \mathcal{A}_1^*$. Matrix $V(S)$ corresponding to $S = (A)$ can be used to construct K. But if we need SSs to estimate $E[\boldsymbol{\gamma}]$, we can use matrix $V(T)$ instead, where $T = (\varnothing, A) \in \mathcal{A}_2^*$, and relabel $Y_{1,\omega}$ as $Y_{2,\omega}$. We followed this method in the analysis of the STAR*D data in Sections 9-10. Clearly, this approach to data analysis is not always applicable to studies that were not designed with SSs in mind, because it requires that some treatments be *simultaneously* administered at two or more different stages, and appropriate SSs may not always be available. This also highlights the importance of incorporating SSs at the planning stage of a CT.

7. Effect of a Stage on Future Responses

ASs allow testing naturally the null hypothesis that clinical decisions made at a particular stage do not affect the subjects' responses in future stages. Note that $\boldsymbol{\beta}_{ik}$ includes all the effects of i-stage treatments on $Y_{k,\omega}$. If $\boldsymbol{\beta}_{ik} \neq 0$, we say that Stage i *affects* $Y_{k,\omega}$. The importance of this concept is that if $\boldsymbol{\beta}_{ik} = 0$ for $i < k$, then clinical decisions made at Stage i will not affect the response measured at Stage k, and *there are no delayed effects* of Stage i on $Y_{k,\omega}$. To examine whether Stage i does not affect $Y_{k,\omega}$ we test the null hypothesis, $\boldsymbol{\beta}_{ik} = 0$. This test can be implemented with standard statistical software for fitting GLMMs, examining the joint significance of the covariates corresponding to $\boldsymbol{\beta}_{ik}$. These covariates are called i-to-k *transition covariates.* For a particular subject, the values of these (stage-dependent) covariates are in columns c_1 through c_2 of matrix V, where $c_1 = m_{k-1} + \sum_{j=1}^{i-1} l_j + 1$ and $c_2 = m_{k-1} + \sum_{j=1}^{i} l_j$. For convenience, the covariates corresponding to $\boldsymbol{\beta}_{ii}$ are called i-stage covariates; these are also stage-dependent.

For instance, for $q = 2$, the third terms of the equations in (2) contain the 1-to-2 transition covariates. There are l_1 of these covariates, one for each 1-Stage treatment. Similarly, first and second stage covariates are in the second and fourth terms, respectively. Tables 2 and 5 illustrate these types of covariates for the cross-over example (Section 8) and the STAR*D data (Section 9), respectively.

8. Application: Improving a Cross-Over Design with 3 Treatments and 3 Periods

This Section illustrates an application of the proposed methodology to the design of a cross-over study of 3 treatments administered in 3 periods. The treatments are denoted A, B and C. Specifically, we show how placebo can be implemented in order to achieve the estimability of carry-over effects. Suppose the trialist wants to implement 6 sequences arranged in 2 Latin squares, as follows (Senn 2002):

$$\begin{array}{ccc} A & B & C \\ B & C & A \\ C & A & B \\ A & C & B \\ B & A & C \\ C & B & A \end{array}$$

We assume a baseline measure is obtained prior to the administration of a sequence. We also assume a normal response Y, independent and identically distributed random errors, and a random intercept representing subjects's heterogeneity (Jones & Kenward 2015). Cross-over studies often assign the same number of subjects to the investigated sequences, but we do not need to assume this for this illustration.

Table 1 shows the expected responses by sequence and period for the above sequences, assuming the responses satisfy an AS (sequences 1-6). In Table 1, a parameter τ_{Aij} measures the effect of treatment A on Y, when A is administered in period i and Y is measured at the end of period j, for $i \leq j$ and $i, j = 0, 1, 2, 3$. This effect is measured relative to the average baseline response μ. The effects of treatments B and C are defined analogously. Table 1 also shows expected responses for some skip sequences (sequences 7-12). Here, we interpret ϙ as placebo and assume placebo has negligible effects on Y.

We built the DM corresponding to sequences 1-6 without assuming any linear constraint on model parameters. For a design with only one subject per sequence, the DM is shown in Table 2. For a study with more than one subject in a sequence, the study DM will include the rows corresponding to the sequence as many times as subjects are assigned to the sequence. As predicted from Theorem 2, the matrix in Table 2 (and the study DM in general) are not of full rank.

Here, we are interested in estimating the percent contribution of carry-over to the total effect of a treatment sequence. For instance, the percent contribution of carry over to the total effect of sequence ABC is defined as

$$\frac{\tau_{A13} + \tau_{B23}}{\mu + \tau_{A13} + \tau_{B23} + \tau_{C33}} \times 100$$

Similarly, the contribution of carry-over to the total effect of the partial sequence AB is

$$\frac{\tau_{A12}}{\mu + \tau_{A12} + \tau_{B22}} \times 100$$

Carry-over contributions for other sequences are defined analogously.

TABLE 1: Expected responses under a cross-over design with 3 treatments, 3 periods and 6 sequences (Sequences 1-6), assuming an antichronic system. Expected responses for some skip sequences are also shown (Sequences 7-12). Placebo is symbolized with ♀.

	Period			
Sequence	0 (Baseline)	1	2	3
1 ABC	μ	$\mu+\tau_{A11}$	$\mu+\tau_{A12}+\tau_{B22}$	$\mu+\tau_{A13}+\tau_{B23}+\tau_{C33}$
2 BCA	μ	$\mu+\tau_{B11}$	$\mu+\tau_{B12}+\tau_{C22}$	$\mu+\tau_{B13}+\tau_{C23}+\tau_{A33}$
3 CAB	μ	$\mu+\tau_{C11}$	$\mu+\tau_{C12}+\tau_{A22}$	$\mu+\tau_{C13}+\tau_{A23}+\tau_{B33}$
4 ACB	μ	$\mu+\tau_{A11}$	$\mu+\tau_{A12}+\tau_{C22}$	$\mu+\tau_{A13}+\tau_{C23}+\tau_{B33}$
5 BAC	μ	$\mu+\tau_{B11}$	$\mu+\tau_{B12}+\tau_{A22}$	$\mu+\tau_{B13}+\tau_{A23}+\tau_{C33}$
6 CBA	μ	$\mu+\tau_{C11}$	$\mu+\tau_{C12}+\tau_{B22}$	$\mu+\tau_{C13}+\tau_{B23}+\tau_{A33}$
Some skip sequences:				
7 ♀♀A	μ	μ	μ	$\mu+\tau_{A33}$
8 ♀♀B	μ	μ	μ	$\mu+\tau_{B33}$
9 ♀♀C	μ	μ	μ	$\mu+\tau_{C33}$
10 ♀CA	μ	μ	$\mu+\tau_{C22}$	$\mu+\tau_{C23}+\tau_{A33}$
11 ♀CB	μ	μ	$\mu+\tau_{C22}$	$\mu+\tau_{C23}+\tau_{B33}$
12 ♀CC	μ	μ	$\mu+\tau_{C22}$	$\mu+\tau_{C23}+\tau_{C33}$

In general, however, if the cross-over trial includes only sequences 1-6, the numerators of the carry-over contributions will not be estimable and, therefore, we will not be able to estimate the carry-over contributions. To demonstrate this assertion, we used Searle criterion for estimability (Searle 1966). According to this criterion, if $\boldsymbol{\tau}$ is the column vector containing the 19 parameters in Table 1 (including μ), L is a row vector, and $L\boldsymbol{\tau}$ is a contrast of these parameters, then $L\boldsymbol{\tau}$ is estimable if and only if

$$L - L(M^TM)^-M^TM = 0$$

where M is the matrix in Table 2 and $(M^TM)^-$ is the Moore-Penrose inverse of M^TM (Bronson 1989). Using Searle criterion, we found that all parameters in Table 1 are nonestimable, except μ, τ_{A11}, τ_{B11} and τ_{C11}, if only sequences 1-6 are implemented in the trial. In addition, although the denominators of all percent carry-over contributions are estimable, the numerators are not.

To assess carry-over contributions, the estimability of all model parameters is needed. By Theorem 2, we will be able to achieve estimability if we add SSs to

the design. Figure 1 illustrates how to compute appropriate skip sequences in accordance to the procedure explained after Theorem 3. The top portion of the figure shows the matrix Γ, whose six rows represent sequences 1-6. The figure also shows the reduced row echelon form of matrix Γ. Since the echelon has four columns without pivots, at least four skip sequences are necessary to achieve the estimability of all model parameters. The bottom of Figure 1 suggests that we need to add to the design the sequences $(\mathfrak{Q}, \mathfrak{Q}, A)$, $(\mathfrak{Q}, \mathfrak{Q}, B)$ and $(\mathfrak{Q}, \mathfrak{Q}, C)$, and at least one of the sequences $(\mathfrak{Q}, C, T)$ with $T = A, B,$ or C. Thus, for instance, if at least one subject is randomized to each of the sequences 1-10 in Table 1, percent carry-over contributions can be estimated. Table 3 shows the DM corresponding to the skip sequences 7-10 in Table 1, assuming only one subject per sequence and without imposing any parameter constraint. The reader can verify that if Tables 2 and 3 are combined, the resultant matrix is of full rank as predicted from Theorem 3. Therefore, the study DM of a cross-over trial including sequences 1-10 will be of full rank, provided that at least one subject undergoes each sequence. We computed the Moore-Penrose inverse with SAS PROC IML (SAS Institute Inc.) and the echelon matrix in Figure 1 with the calculator TI-Nspire CX CAS (Texas Instruments Inc.).

		Period 1			Period 2			Period 3		
	Sequence	*A*	*B*	*C*	*A*	*B*	*C*	*A*	*B*	*C*
$\Gamma =$	*ABC*	1	0	0	0	1	0	0	0	1
	BCA	0	1	0	0	0	1	1	0	0
	CAB	0	0	1	1	0	0	0	1	0
	ACB	1	0	0	0	0	1	0	1	0
	BAC	0	1	0	1	0	0	0	0	1
	CBA	0	0	1	0	1	0	1	0	0
Echelon=		1	0	0	0	0	1	0	1	0
		0	1	0	0	0	1	1	0	0
		0	0	1	0	0	1	1	1	–1
		0	0	0	1	0	–1	–1	0	1
		0	0	0	0	1	–1	0	–1	1
		0	0	0	0	0	0	0	0	0
Columns without pivots							(*C*)	(*A*)	(*B*)	(*C*)

FIGURE 1: Illustration of the computation of skip sequences for a cross-over design with 3 treatments, 3 periods and 6 sequences, according to Theorem 3.

TABLE 2: Design matrix for sequences 1-6 in Table 1, assuming an antichronic system. The parameter corresponding to a particular column is shown at the top of the column. For a study with N_i subjects on a sequence, the complete study design matrix includes the four rows of the sequence N_i times.

			Period 1			1-to-2 transition			Period 2			1-to-3 transition			2-to-3 transition			Period 3		
Sequence	Period	μ	τ_{A11}	τ_{B11}	τ_{C11}	τ_{A12}	τ_{B12}	τ_{C12}	τ_{A22}	τ_{B22}	τ_{C22}	τ_{A13}	τ_{B13}	τ_{C13}	τ_{A23}	τ_{B23}	τ_{C23}	τ_{A33}	τ_{B33}	τ_{C33}
1 *ABC*	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1 *ABC*	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1 *ABC*	2	1	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0
1 *ABC*	3	1	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	1
2 *BCA*	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2 *BCA*	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2 *BCA*	2	1	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0
2 *BCA*	3	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1	0	0
3 *CAB*	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3 *CAB*	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3 *CAB*	2	1	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0
3 *CAB*	3	1	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0
4 *ACB*	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4 *ACB*	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4 *ACB*	2	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0
4 *ACB*	3	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	1	0
5 *BAC*	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5 *BAC*	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5 *BAC*	2	1	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0
5 *BAC*	3	1	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	1
6 *CBA*	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6 *CBA*	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6 *CBA*	2	1	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0
6 *CBA*	3	1	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	1	0	0

TABLE 3: Design matrix for the four skip sequences, (♀, ♀, A), (♀, ♀, B), (♀, ♀, C) and (♀, C, A). Combining this matrix with that in Table 2 produces a matrix of full rank. The rows corresponding to a particular sequence must be included in the study design matrix as many times as subjects are assigned to the sequence.

			Period 1			1-to-2 transition			Period 2			1-to-3 transition			2-to-3 transition			Period 3		
Sequence	Period	μ	τ_{A11}	τ_{B11}	τ_{C11}	τ_{A12}	τ_{B12}	τ_{C12}	τ_{A22}	τ_{B22}	τ_{C22}	τ_{A13}	τ_{B13}	τ_{C13}	τ_{A23}	τ_{B23}	τ_{C23}	τ_{A33}	τ_{B33}	τ_{C33}
7 ♀♀A	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7 ♀♀A	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7 ♀♀A	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7 ♀♀A	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
8 ♀♀B	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8 ♀♀B	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8 ♀♀B	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8 ♀♀B	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
9 ♀♀C	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9 ♀♀C	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9 ♀♀C	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9 ♀♀C	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
10 ♀CA	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10 ♀CA	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10 ♀CA	2	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
10 ♀CA	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0

9. Application: Fitting an AS to the STAR*D Data

In this section, we use virtual placebo for the analysis of the STAR*D clinical trial database which was provided by the National Institutes of Mental Health. The author did not conduct this clinical trial or participate in it. The STAR*D study investigators have declared in their publications that the study was approved by ethical committees and that patient consents were obtained (Fava et al. 2003). The database provided to the author did not contain information identifying the subjects, and the author maintained the database in a password-protected computer system.

The STAR*D study investigated 4,041 adult subjects with non-psychotic major depressive disorder (Fava et al. 2003). We analyzed the 1,439 subjects who did not have an acceptable response to Citalopram (CIT) monotherapy but continued in the study receiving alternative treatments. Here, Y is the HAM-D17 total depression score and $Y_{0,\omega}$ is the score obtained right after CIT monotherapy ended. The HAM-D17 scale produced values from 0 to 42 inclusive, and all integer numbers in this range were observed in the patient population. Thus, we treated the HAM-D17 score as a continuous variable, as most statisticians would do in practice. An AS model with identity link and Gaussian response was fitted. The design matrix K was built using our Stata program asdesign (see Appendix C).

Alternative treatments given right after Citalopram monotherapy failure are named here "1-stage" treatments (Table 4). If a 1-stage treatment was declared to be successful by the clinician, the subject was passed to follow-up (FU). If the 1-stage treatment failed, the subject and clinician were presented with other treatment choices; these other choices, together with FU, are named here "2-stage" treatments. Although additional treatment choices were available if a 2-stage treatment failed, we analyzed here only sequences of at most 2 additional treatments given after CIT monotherapy failure; that is, $q = 2$ (Table 4).

In the STAR*D data, correlated effects were possible because 2-stage decisions may have been correlated with observed or unobserved subjects' characteristics that determined the responses to Stage-1 decisions. Thus, we assumed a model with correlated effects and used robust fixed-effects estimators to estimate regression coefficients of first and second stage covariates and transition covariates (Kim & Frees 2006, Ebbes et al. 2004, Frees 2004) (Table 5, footnote g). Additional advantages of this approach are that the actual structure and distribution of random effects do not need to be specified, and it is not necessary to conduct subject-level variable selection because these variables are not used in robust fixed-estimation. However, since we were also interested in estimating the effects of demographics, we additionally conducted a separate augmented regression to estimate their effects (Mundlak 1978, Frees 2001, 2004) (Table 5, footnote d). In the augmented regression, we treated second-stage covariates as endogenous and used their average to augment the regression equation. Since subjects were initially randomized to first-stage treatments, first-stage and 1-to-2 transition covariates were treated as exogenous. Huber/White robust standard errors were computed (Huber et al. 1967, White 1982) (Table 5).

TABLE 4: Sequences of at most two additional treatments administered to 1,439 subjects with major depression from the STAR*D study who were refractory to Citalopram monotherapy.

Sequence	Stage 1[a]	Stage 2[b]	N[c]	Sequence	Stage 1[a]	Stage 2[b]	N[c]
1	BUP	--	87[d]	20	CIT+CTH	BUP	10
2	BUP	BUP+LI	18	21	CIT+CTH	FU	45
3	BUP	BUP+THY	8	22	CIT+CTH	VEN	12
4	BUP	FU	69	23	CTH	--	20[d]
5	BUP	MIRT	28	24	CTH	BUP	5
6	BUP	NTP	29	25	CTH	FU	33
7	CIT+BUP	--	68[d]	26	CTH	VEN	4
8	CIT+BUP	CIT+LI	14	27	SER	--	75[d]
9	CIT+BUP	CIT+THY	20	28	SER	FU	87
10	CIT+BUP	FU	150	29	SER	MIRT	33
11	CIT+BUP	MIRT	10	30	SER	NTP	22
12	CIT+BUP	NTP	17	31	SER	SER+LI	11
13	CIT+BUS	--	78[d]	32	SER	SER+THY	10
14	CIT+BUS	CIT+LI	10	33	VEN	--	81[d]
15	CIT+BUS	CIT+THY	17	34	VEN	FU	101
16	CIT+BUS	FU	137	35	VEN	MIRT	21
17	CIT+BUS	MIRT	18	36	VEN	NTP	22
18	CIT+BUS	NTP	26	37	VEN	VEN+LI	10
19	CIT+CTH	--	18[d]	38	VEN	VEN+THY	15

FU: follow-up; CTH: Cognitive Therapy; BUP: Bupropion; BUS: Buspirone; CIT: Citalopram; LI: Lithium; MIRT: Mirtazapine; NTP: Nortriptyline; SER: Sertraline; TCP: Tranylcypromine; THY: Thyroid; VEN: Venlafaxine.

[a]Stage 1 treatments are the treatments from Level 2 of the STAR*D study. In Level 1 of the study, only Citalopram monotherapy was given to the subjects.

[b]Stage 2 treatments are the treatments from Levels 2A and 3 of STAR*D study. A follow-up (FU) after Level 2 of the study was treated as a "treatment" choice for Stage 2. The first HAM-D17 score available from follow-up was usually obtained at the end of the 3rd month of follow-up and was used in the fit of the AS model.

[c]Number of subjects who underwent the treatment sequence.

[d]These subjects withdrew from the study after completing Stage 1 (=Level 2 of STAR*D study).

Stata's xtreg and mixed commands were used (StataCorp LP, College Station, TX, USA). For comparison purposes, we also present GLS estimates for a random intercept model with i.i.d errors (Table 5). These estimates would be more appropriate if there were not correlated effects, but this cannot be guaranteed in the STAR*D data. Residual analyses showed both that the model fitted well and that the assumption of normality for the untransformed scores was reasonable.

TABLE 5: Parameter estimates of a 2-stage antichronic system model of HAM-D17 scores computed with 3,303 scores from 1,434 STAR*D subjects refractory to Citalopram monotherapy.[a]

Covariates	GLS estimators (random intercept)[b]		Robust estimators[c]	
	Regression coefficient	95% Confidence Interval	Regression Coefficient	95% Confidence interval
Age (y)	0.060***	[0.033, 0.087]	0.055***	[0.032, 0.078][d]
Black or African American[e]	2.52***	[1.61, 3.44]	2.01***	[1.16, 2.85][d]
Cocaine abuser or dependent[f]	–3.70*	[–7.10, –0.31]	–3.87***	[–6.15, –1.60][d]
First stage covariates				
BUP	–5.09***	[–6.17, –4.01]	–5.26***	[–6.45, –4.06][g]
CIT+BUP	–5.87***	[–6.75, –4.98]	–4.90***	[–5.87, –3.93][g]
CIT+BUS	–4.87***	[–5.72, –4.02]	–4.16***	[–5.15, –3.16][g]
CIT+CTH	–4.67***	[–6.16, –3.19]	–4.65***	[–6.50, –2.80][g]
CTH	–5.58***	[–7.47, –3.68]	–4.57***	[–7.03, –2.10][g]
SER	–4.74***	[–5.70, –3.77]	–5.56***	[–6.70, –4.43][g]
VEN	–5.89***	[–6.93, –4.85]	–6.01***	[–7.23, –4.80][g]
Transition covariates[h]				
BUP	0.73	[–2.76, 4.22]	0.28	[–3.65, 4.21][g]
CIT+BUP	0.58	[–2.80, 3.95]	0.66	[–3.17, 4.48][g]
CIT+BUS	0.48	[–2.90, 3.86]	0.60	[–3.22, 4.41][g]
CIT+CTH	–0.53	[–3.50, 2.44]	–1.29	[–4.71, 2.14][g]
CTH	–0.0021	[–3.52, 3.51]	–0.35	[–4.62, 3.91][g]
SER	–0.86	[–4.30, 2.59]	–2.05	[–5.93, 1.83][g]
VEN	–0.48	[–3.94, 2.97]	–1.66	[–5.58, 2.25][g]
Second stage covariates				
BUP	–2.53*	[–4.60, –0.46]	–2.85**	[–4.91, –0.79][g]
BUP+LI	–4.35	[–8.90, 0.21]	–5.48*	[–9.96, –0.99][g]
BUP+THY	–10.6***	[–16.0, –5.17]	–11.2***	[–17.0, –5.32][g]
CIT+LI	–3.96	[–8.29, 0.36]	–5.09*	[–9.29, –0.88][g]
CIT+THY	–5.81**	[–9.79, –1.83]	–5.58*	[–10.2, –1.01][g]
FU	–6.69***	[–9.92, –3.45]	–5.03**	[–8.70, –1.36][g]
MIRT	–5.17**	[–8.72, –1.62]	–5.95**	[–9.90, –2.01][g]
NTP	–5.25**	[–8.78, –1.72]	–6.02**	[–10.1, –1.95][g]
SER+LI	–4.91	[–10.3, 0.46]	–6.29*	[–11.3, –1.32][g]
SER+THY	–3.26	[–8.47, 1.95]	–3.49	[–8.24, 1.26][g]
VEN	–4.17***	[–6.36, –1.97]	–4.68**	[–7.66, –1.70][g]
VEN+LI	–8.25**	[–13.6, –2.87]	–7.78**	[–13.2, –2.34][g]
VEN+THY	–5.71*	[–10.5, –0.96]	–5.92*	[–11.1, –0.76][g]
Constant (μ_α)	14.9***	[13.7, 16.1]	17.4***	[17.2, 17.6][g]

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Note: To estimate model parameters, skip sequences with BUP and VEN at the second stage were implemented.

FU: follow-up; CTH: Cognitive Therapy; BUP: Bupropion; BUS: Buspirone; CIT: Citalopram; Li: Lithium; MIRT: Mirtazapine; NTP: Nortriptyline; SER: Sertraline; TCP: Tranylcypromine; THY: Thyroid; VEN: Venlafaxine.

[a]The data were obtained from version 4.1 of the limited access datasets distributed from the NIH-supported "Sequenced Treatment Alternatives to Relieve Depression" (STAR*D). Before fitting the models, a few inconsistencies in the STAR*D data were found and corrected by the author. Inconsistencies and corrections were documented in a text file that is available from the author on request.

[b]For the random intercept model estimated with GLS, estimated standard deviations of intercept and error were $\hat{\sigma}_\alpha = 5.56$ (95% confidence interval, [5.26, 5.88]) and $\hat{\sigma}_\epsilon = 5.12$ ([4.95, 5.29]).

[c]To estimate regression coefficients of demographic variables, augmented regression was used. To estimate regression coefficients of first and second stage covariates and transition covariates, robust fixed-effects estimators were used.

[d]Augmented regression (Mundlak 1978; Frees 2001, 2004) and Huber/White sandwich standard errors (Huber 1967, White 1982) were used. The estimators are robust to correlated effects and covariance matrix misspecification.

[e]The dichotomous covariate was defined as 1 if the subject was self-declared black or African American, 0 otherwise.

[f]The dichotomous covariate was defined as 1 if the subject declared to be cocaine abuser or dependent, 0 otherwise.

[g]Robust fixed effects estimators (Hausman 1978, Kim and Frees 2006) and Huber/White sandwich standard errors (Huber 1967, White 1982) were computed. The estimators are robust to correlated effects and covariance matrix misspecification.

[h]1-to-2 transition covariates. These covariates were jointly significant (robust test of $H_0 : \boldsymbol{\beta}_{1,2} = 0$, p= 0.0079), suggesting that treatments at Stage 1 affected significantly the responses measured at the end of Stage 2. However, no individual transition covariate reached significance.

Data and skip sequences used are described in the rest of this section. Five of the 1,439 subjects did not provide any of the three scores $Y_{0,\omega}$, $Y_{1,\omega}$ and $Y_{2,\omega}$. Each of the other 1,434 subjects provided at least one of them. Thus, the 2-stage AS was fitted with these 1,434 subjects (Table 5). An assumption of missingness at random was made to include subjects who did not have all three values of Y. In total, the subjects provided 3,303 HAM-D17 scores for this analysis.

After CIT monotherapy failed, subjects had seven alternative treatment choices for Stage 1; thus, $l_1 = 7$ (Table 4). For Stage 2, subjects had 13 alternative choices, including follow-up (FU) ($l_2 = 13$). Note that the STAR*D protocol did not use the term "Stage" (footnotes a and b of Table 4).

Skip sequences were built as follows. Both Bupropion (BUP) and Venlafaxine (VEN) were simultaneously 1-stage and 2-stage treatments (Table 4). Also, 87 subjects who were on BUP at Stage 1 withdrew from the study after completing this stage (Sequence 1 in Table 4). Applying Theorem 4, these 87 subjects were assigned the skip sequence, $(\text{Ƴ}, \text{BUP}) \in \mathcal{A}_2^*$. Similarly, the skip sequence, $(\text{Ƴ}, \text{VEN}) \in \mathcal{A}_2^*$ was assigned to each of 81 subjects who were on VEN at Stage 1 but withdrew after completing this stage (Sequence 33 in Table 4).

In summary, we built an AS that included the 38 sequences in Table 4, but Sequences 1 and 33 were redefined to corresponding interchangeable SSs. By following this approach, we obtained a *study design matrix* K of full rank. Parameter estimates are shown in Table 5. Note that, here, Ƴ can be interpreted

as "CIT monotherapy". Also, the regression coefficient of follow-up (FU) must be interpreted with caution, because two different subjects on FU may have been under different therapies.

Note that subjects in Sequence 7 of Table 4 took CIT+BUP at Stage 1 but withdrew from the study after completing this stage. These subjects cannot be used to build SSs because CIT+BUP was not a 2-stage treatment. Thus, Sequence 7 was treated as $(\text{CIT} + \text{BUP}) \in \mathcal{A}_1^*$. Similar comments can be made for subjects in Sequences 13, 19, 23 and 27. In contrast, Sequence 35 in Table 4, say, was treated as $(\text{VEN}, \text{MIRT}) \in \mathcal{A}_2^*$.

10. Results for STAR*D Data

Hausman test comparing robust fixed-effects estimators with GLS estimators confirmed that there were correlated effects ($\chi^2 = 346.9$, df $= 27$, $p < 0.0001$) (Frees 2004). Therefore, robust fixed-effects estimators and augmented regression were necessary.

At $\alpha = 0.05$, each 1-Stage treatment reduced significantly the average HAM-D17 score during Stage 1 in subjects who received the treatment (Table 5). Using the robust fixed-effects estimates, we found that Stage 1 significantly affect $Y_{2,\omega}$ [overall test with robust estimates, $F(7, 1433) = 2.74$, $p = 0.0079$]. That is, at least one transition covariate was significant according to the overall test. However, none of the individual regression coefficients of transition covariates reached significance. Thus, the evidence is weak that treatments used at Stage 1 had a potential influence on HAM-D17 scores measured at the end of Stage 2. In particular, *there did not appear to be appreciable delayed effects.* Not even cognitive therapy (CTH), which was a 1-stage treatment, gave these difficult-to-treat subjects a significant added advantage at fighting their illness at the next stage—the regression coefficient for the transition covariate corresponding to CTH was -0.35, 95% CI, $(-4.62, 3.91)$ (Table 5). The other 1-stage treatments included drugs and it is possible that these drugs' blood levels dropped during Stage 2, which may explain in part why Stage 1 treatments did not affect substantially the 2^{nd} stage response.

Only one 2-stage treatment did not lower significantly the HAM-D17 average score in comparison with the score measured at the end of CIT monotherapy, namely SER+THY (Table 5). Also, according to the robust augmented regression, self-declared blacks or African Americans had significantly higher HAM-D17 average scores after controlling for treatment; cocaine abuser or dependent subjects had significantly lower average scores; and subjects of older age had significantly higher average scores.

The regression coefficients of 1-stage treatments were not significantly different from each other (using robust estimation, $F(6, 1433) = 1.13$, $p = 0.3398$; Table 5). Therefore, there was no evidence that 1-stage treatments had differential average effects on HAM-D17. Similarly, 2-stage treatments did not have significantly different average effects on HAM-D17 ($F(12, 1433) = 1.18$, $p = 0.2927$).

We also tested the appropriateness of the Interchangeability Assumption. This assumption implies that the regression coefficients of BUP for the first and second stage covariates are equal and that those of VEN are equal as well. Employing the robust standard errors, a simultaneous Wald test of these two null hypotheses was not significant ($F(2, 1433) = 2.06$, $p = 0.1275$). This suggests the assumption was appropriate.

With a few additions, our model allows investigating interactions between treatments administered at different stages. Transition covariates are useful in this regard. As an illustration, we examined the interaction between BUP administered at the first stage and MIRT administered at the second stage by computing the product of the transition covariate corresponding to BUP and the second-stage covariate corresponding to MIRT. Using robust estimation, this interaction was not significant ($p = 0.70$). It is also possible to compute interactions between baseline covariates and stage or transition covariates.

In general, however, some interaction terms may induce a non-full-rank design matrix. Thus, future research must investigate how the presence of interaction terms affects the rank of design matrices, and must give conditions for the identifiability of their regression coefficients. Fortunately, however, the omission of significant interaction terms should not bias robust fixed-effects estimators because they are robust to omitted variables (Kim & Frees 2006).

11. Discussion

The main contribution of this article is the introduction of skip sequences in the design and analysis of clinical trials of treatment sequences. These sequences guarantee the existence of design matrices of full rank and, therefore, guarantee the identifiability (estimability) of carry-over effects. Importantly, we did not need unjustified linear constraints in order to examine carry-over effects. Skip sequences also facilitate the implementation of many estimation methods for GLMMs.

Theorem 2 indicates the importance of SSs in the experimental design and statistical analysis of CTs of treatment sequences, and theorem 3 gives a procedure for selecting appropriate SSs at the planning stage of a particular study. This procedure is very general, being applicable to any number of study stages and treatments, and to continuous or discrete responses, including dichotomous responses.

Real placebo can be used to implement SSs in practice. Alternatively, a virtual placebo approach can be applied to the analysis of data from some CTs such as the STAR*D study which did not specifically implement SSs in their protocol. As illustrated with the STAR*D data, useful insights can also be obtained in such CTs. This approach, however, cannot always be applicable if SSs are not implemented in the study protocol. Theorem 4 supports the use of virtual placebo. A virtual placebo approach can also be applied when real placebo is unethical or unfeasible. In such cases, appropriate SSs must be selected at the design stage of the CT.

ASs are applicable if we can assume the subjects' health will not improve or deteriorate rapidly should they remain under baseline conditions. In other words, $Y_{0,\omega}$ must measure a stable illness state. Having a stable illness is an inclusion criterion of many clinical trials. Our approach is applicable to these trials. This assumption, however, may not be applicable to some conditions such as advanced cancer or mental illness in acute state.

Our model does not include stage-dependent variables other than the investigated MBTs. The reason is that CTs are highly controlled experiments that attempt maintaining other variables affecting the response constant during the study. Therefore, in a well-conducted CT, the response Y should be a function *only* of both MBTs administered during the CT and baseline covariates. For instance, it is well known that smoking affects the metabolism of some neurotropic drugs (e.g. Botts, Diaz, Santoro, Spina, Muscatello, Cogollo, Castro & de Leon 2008). In a CT investigating these drugs, a subject quitting smoking (or starting smoking) during the CT is usually considered a protocol violation. Therefore, data from this subject is not included in data analyses. Thus, although it would be easy to incorporate in our model stage-dependent covariates other than those proposed in this article, doing so would be just a mere academic exercise that is inconsistent with the principles of CTs.

Note also that variables changing as a result of the investigated MBTs should not be treated as independent variables in ASs, because ASs account for the effects of treatments applied in one stage on subsequent stages. Thus, for instance, if a 1-stage treatment changed the value of a mediating variable that ultimately affected $Y_{2,\omega}$, the corresponding 1-to-2 transition covariate will be significant. This means that we do not need to include the mediating variable as independent variable.

Our approach based on SSs poses an interesting question: when Theorem 3 guarantees that two or more alternative sets of SSs will produce a design matrix of full rank, which set will produce the lowest standard errors for the estimates of the regression coefficients? This open problem needs to be investigated by future research.

If the value of Y is missing for some stages, the corresponding rows of V can be eliminated accordingly. In such case, the assumption of missingness at random must be made, as we did to fit an AS to the STAR*D data.

Acknowledgements

The STAR*D study was supported by NIMH Contract #N01MH90003, and has the ClinicalTrials.gov Identifier, NCT00021528. The author was supported by an Institutional Clinical and Translational Science Award, NIH/NCATS Grant Number UL1TR000001 (awarded to the University of Kansas Medical Center). The author was not involved in the planning or conduction of the STAR*D study and has never received any payment or salary from its associated NIMH contract or its investigators. This manuscript may not reflect the opinions or views of the STAR*D study investigators or the NIH. The author thanks Mr. Douglas

Eikermann and Guanlin Zhang for editorial assistance and making useful comments. Mr. Jiawei Duan also made useful comments. The author also thanks the two anonymous reviewers, whose comments motivated interesting additions to the manuscript.

[Received: March 2017 — Accepted: May 2018]

References

Abeyasekera, S. & Curnow, R. N. (1984), 'The desirability of adjusting for residual effects in a crossover design', *Biometrics* **40**, 1071–1078.

Berg, M., Welty, T. E., Gidal, B. E., Diaz, F. J., Krebill, R., Szaflarski, J. P., Dworetzky, B. A., Pollard, J. R., Elder Jr, E. J., Jiang, W., Jiang, X., Switzer, R. D. & Privitera, M. D. (2017), 'Bioequivalence between generic and branded lamotrigine in people with epilepsy: the EQUIGEN randomized clinical trial', *JAMA Neurology* **74**, 919–926.

Botts, S., Diaz, F., Santoro, V., Spina, E., Muscatello, M. R., Cogollo, M., Castro, F. E. & de Leon, J. (2008), 'Estimating the effects of co-medications on plasma olanzapine concentrations by using a mixed model', *Progress in Neuro-Psychopharmacology & Biological Psychiatry* **32**, 1453–1458.

Breslow, N. E. & Clayton, D. G. (1993), 'Approximate inference in generalized linear mixed models', *Journal of the American Statistical Association* pp. 9–25.

Bronson, R. (1989), *Matrix Operations*, Schaum's Outline Series, 1 edn, McGraw-Hill, New York.

Center for Drug Evaluation and Research (2001), Us food and drug administration. guidance for industry: Statistical approaches to establishing bioequivalence. https://www.fda.gov/ downloads/drugs/guidances/ucm070244.pdf, Accessed February 22, 2018.

Center for Drug Evaluation and Research (2003), Us food and drug administration. guidance for industry: Bioavailability and bioequivalence studies for orally administered drug products - general considerations. https://www.fda.gov/ohrms/dockets/ac/03/briefing/3995B1_07_GFI-BioAvail-BioEquiv.pdf, Accessed February 22, 2018.

Christensen, R. (2011), *Plane Answers to Complex Questions: The Theory of Linear Models*, Springer, New York.

Diaz, F. J. (2016), 'Measuring the Individual Benefit of a Medical or Behavioral Treatment Using Generalized Linear Mixed-Effects Models', *Statistics in Medicine* **35**, 4077–4092.

Diaz, F. J. (2017), 'Estimating individual benefits of medical or behavioral treatments in severely ill patients', *Statistical Methods in Medical Research* (DOI: 10.1177/0962280217739033).

Diaz, F. J., Berg, M. J., Krebill, R., Welty, T., Gidal, B. E., Alloway, R. & Privitera, M. (2013), 'Random-effects linear modeling and sample size tables for two special cross-over designs of average bioequivalence studies: the 4-period, 2-sequence, 2-formulation and 6-period, 3-sequence, 3-formulation designs', *Clinical Pharmacokinetics* **52**, 1033–1043.

Ebbes, P., Bockenholt, U. & Wedel, M. (2004), 'Regressor and random-effects dependencies in multilevel models', *Statistica Neerlandica* **58**, 161–178.

Fava, M., Rush, A. J., Trivedi, M. H., Nierenberg, A. A., Thase, M. E., Sackeim, H. A., Quitkin, F. M., Wisniewski, S., Lavori, P. W., Rosenbaum, J. F. & Kupfer, D. J. (2003), 'Background and rationale for the sequenced treatment alternatives to relieve depression (STAR*D) study', *Psychiatric clinics of North America* **26**, 457–494.

Fitzmaurice, G. & Molenberghs, G. (2004), Advances in longitudinal data analysis, *in* G. Fitzmaurice, G. Molenberghs, M. Davidian & G. Verbeke, eds, 'Longitudinal data analysis', Chapman and Hall/CRC, London.

Fleiss, J. L. (1989), 'A critique of recent research on the two-treatment Crossover design', *Controlled Clinical Trials* **10**, 237–43.

Frees, E. W. (2001), 'Omitted variables in longitudinal data models', *The Canadian Journal of Statistics* **29**, 573–595.

Frees, E. W. (2004), *Longitudinal and Panel Data*, Cambridge, Cambridge University Press.

Grajales, L. F. & Lopez, L. A. (2006), 'Data imputation in switchback designs using a mixed model with correlated errors', *Revista Colombiana de Estadística* **29**, 221–238.

Hausman, J. A. (1978), 'Specification tests in econometrics', *Econometrica* **46**, 1251–1271.

Henderson, C. R. (1953), 'Estimation of variance and covariance components', *Biometrics* **9**, 226–252.

Hofmann, M., Wrobel, N., Kessner, S. & Bingel, U. (2014), 'Minimizing carryover effects after treatment failure and maximizing therapeutic outcome. Can changing the route of administration mitigate the influence of treatment history?', *Zeitschrift fur Psychologie* **222**, 171–178.

Hooks, T., Marx, D., Kachman, S. & Pedersen, J. (2009), 'Optimality criteria for models with random effects', *Revista Colombiana de Estadística* **32**, 17–31.

Huber, P. J. et al. (1967), The behavior of maximum likelihood estimates under nonstandard conditions, *in* 'Proceedings of the fifth Berkeley symposium on mathematical statistics and probability', Vol. 1, University of California Press, pp. 221–233.

Jones, B. & Kenward, M. G. (2015), *Design and Analysis of Cross-over Trials*, CRC Press, Boca Raton.

Kim, J.-S. & Frees, E. W. (2006), 'Omitted variables in multilevel models', *Psychometrika* **71**, 659–690.

Kim, J.-S. & Frees, E. W. (2007), 'Multilevel modeling with correlated effects', *Psychometrika* **72**, 505–533.

Laird, N. M. & Ware, J. H. (1982), 'Random effects models for longitudinal data', *Biometrics* **38**, 963–974.

Long, D. L., Preisser, J. S., Herring, A. H. & Golin, C. E. (2015), 'A marginalized zero-inflated Poisson regression model with random effects', *Journal of the Royal Statistical Society: Series C* **64**, 815–830.

Mundlak, Y. (1978), 'On the pooling of time series and cross-section data', *Econometrica* **46**, 69–85.

Nelder, J. A. & Wedderburn, R. W. M. (1972), 'Generalized linear models', *Journal of the Royal Statistical Society, Series A* **135**, 370–384.

Privitera, M. D., Welty, T. E., Gidal, B. E., Diaz, F. J., Krebill, R., Szaflarski, J. P., Dworetzky, B. A., Pollard, J. R., Elder, E. J., Jiang, W., Jiang, X. & Berg, M. (2016), 'Generic-to-Generic Lamotrigine Switches in People with Epilepsy: A Randomized Controlled Trial', *Lancet Neurology* **15**, 365–372.

Rabe-Hesketh, S. & Skrondal, A. (2009), Generalized linear mixed-effects models, *in* 'Longitudinal Data Analysis', Chapman and Hall/CRC, London, pp. 93–120.

Rabe-Hesketh, S., Skrondal, A. & Pickles, A. (2005), 'Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects', *Journal of Econometrics* **128**, 301–323.

Searle, S. R. (1966), *Matrix Algebra for the Biological Sciences*, Wiley, New York.

Senn, S. (2002), *Cross-over Trials in Clinical Research*, 2 edn, Wiley, Hoboken.

Senn, S., D'Angelo, G. & Potvin, D. (2004), 'Carry-over in cross-over trials in bioequivalence: Theoretical concerns and empirical evidence', *Pharmaceutical Statistics* **3**, 133–142.

Senn, S. & Lambrou, D. (1998), 'Robust and realistic approaches to carry-over', *Statistics in Medicine* **17**, 2849–2864.

Shao, J. (2003), *Mathematical Statistics*, 2 edn, Springer, New York.

Vermunt, J. K. (2005), 'Mixed-effects logistic regression models for indirectly observed discrete outcome variables', *Multivariate Behavioral Research* **40**, 281–301.

White, H. (1982), 'Maximum likelihood estimation of misspecified models', *Econometrica* **50**, 1–25.

Appendix A. Proofs of Theorems

Some proofs utilize the following convenient matrix notation. If a_i, $i = 1, \ldots, n$, is a sequence of scalars or matrices, the matrix $[a_1, \ldots, a_n]$ is denoted with $\exists_{i=1}^{n} a_i$, and the matrix $\left[a_1^T, \ldots, a_n^T\right]^T$ is denoted with $\exists_{n}^{i=1} a_i$. The symbols $\exists_i$ and $\exists^i$ are read "stack right" and "stack down", respectively. For simplicity, for a constant number or matrix a, we denote $\exists^n a = \exists_{i=1}^{n} a = [a, \ldots, a]$ and $\exists_n a = \exists_{n}^{i=1} a = \left[a^T, \ldots, a^T\right]^T$, in which a is repeated n times.

Proof of Theorem 1. Observe that, in general, if $\boldsymbol{a}_{ij}$ and $\boldsymbol{b}_{ij}$ are column vectors of the same dimension, then

$$\sum_{i=1}^{r_0}\sum_{j=1}^{r_i} \boldsymbol{a}_{ij}^T \boldsymbol{b}_{ij} = \left(\exists_{i=1}^{r_0}\exists_{j=1}^{r_i} \boldsymbol{a}_{ij}^T\right)\left(\exists_{i=1}^{r_0}\exists_{j=1}^{r_i} \boldsymbol{b}_{ij}^T\right)^T$$

Then, by Fubini's theorem,

$$\sum_{i=1}^{r}\sum_{h=i}^{r} \boldsymbol{a}_{ih}^T \boldsymbol{b}_{ih} = \sum_{h=1}^{r}\sum_{i=1}^{h} \boldsymbol{a}_{ih}^T \boldsymbol{b}_{ih} = \left(\exists_{h=1}^{r}\exists_{i=1}^{h} \boldsymbol{a}_{ih}^T\right)\left(\exists_{h=1}^{r}\exists_{i=1}^{h} \boldsymbol{b}_{ih}^T\right)^T \tag{4}$$

Fix r and k, where $r = 1, \ldots, q$, and $k \in \{0, 1, \ldots, r\}$ (real placebo) or $k \in \{0, m(S), m(S)+1, \ldots, r\}$ (virtual placebo). Denote $\tau_{ik,\omega} = \sum_{h=i}^{r} I_{\{h=k\}} S_i \boldsymbol{\beta}_{ih,\omega}$, where $I_{\{h=k\}} = 1$ if $h = k$, or $I_{\{h=k\}} = 0$ if $h \neq k$. Observe that $\tau_{ik,\omega} = 0$ if $i > k$ and $\tau_{ik,\omega} = S_i \boldsymbol{\beta}_{ik,\omega}$ if $i \leq k$. Thus,

$$\sum_{i=1}^{k} S_i \boldsymbol{\beta}_{ik,\omega} = \sum_{i=1}^{r} \tau_{ik,\omega} \tag{5}$$

Now define the row vectors $\boldsymbol{T}_k^* = \exists_{h=1}^{r}\left(\exists_{i=1}^{h}(I_{\{h=k\}} S_i)\right)$, $k = 0, \ldots, r$, and observe that the first row of V is $\left[\boldsymbol{T}_0^*, \vec{\boldsymbol{0}}\right]$, where $\vec{\boldsymbol{0}}$ is a row vector of zeros of dimension $m_q - m_r$. And, for $j \geq 2$, the j-th row of V is $\left[\boldsymbol{T}_k^*, \vec{\boldsymbol{0}}\right]$ with $k = j - 2 + m(S)$. Also note that $\boldsymbol{\delta}_{\omega}^{(q)} = \exists_{h=1}^{q}\left(\exists_{i=1}^{h} \boldsymbol{\beta}_{ih,\omega}^T\right)$ Therefore, by equations (1), (4) and (5),

$$
\begin{aligned}
E_\omega\left[Y_{k,\omega}\right] &= g^{-1}\left(\alpha_\omega + \boldsymbol{\lambda}^T\boldsymbol{X}_\omega + \sum_{i=1}^{k} S_i\boldsymbol{\beta}_{ik,\omega}\right) = g^{-1}\left(\alpha_\omega + \boldsymbol{\lambda}^T\boldsymbol{X}_\omega + \sum_{i=1}^{r} \tau_{ik,\omega}\right) \\
&= g^{-1}\left(\alpha_\omega + \boldsymbol{\lambda}^T\boldsymbol{X}_\omega + \sum_{i=1}^{r}\sum_{h=i}^{r} I_{\{h=k\}} S_i\boldsymbol{\beta}_{ih,\omega}\right) \\
&= g^{-1}\left(\alpha_\omega + \boldsymbol{\lambda}^T\boldsymbol{X}_\omega + \boldsymbol{T}_k^*(\boldsymbol{\delta}_\omega^{(r)})^T\right) \\
&= g^{-1}\left(\alpha_\omega + \boldsymbol{\lambda}^T\boldsymbol{X}_\omega + \left[\boldsymbol{T}_k^*, \vec{\mathbf{0}}\right](\boldsymbol{\delta}_\omega^{(q)})^T\right)
\end{aligned}
$$

Definition A.1 To prove Theorem 2, we need to define a new binary matrix operation. For a matrix A with rows A_i, $i = 1, \ldots, n$, and a matrix B with rows B_j, $j = 1, \ldots, k$, define the new "product" of A and B as $A * B = \exists_k^{j=1}\left(\exists_n^{i=1}\left[A_i, B_j\right]\right)$, where $[A_i, B_j]$ is a row vector that contains from left to right the elements of the i-th row of A first, and then those of the j-th row of B. Observe that $A * B$ has $n \cdot k$ rows.

Lemma A.1. (Used in proof of Theorem 2.) Let A be an $n \times p$ matrix. Let I_k be the $k \times k$ identity matrix. Let J_n be an $n \times 1$ vector with ones in all its entries. If there exists a $p \times 1$ vector c such that $Ac = J_n$, then $\rho\left(A * I_k\right) = \rho(A) + k - 1$ for all $k \geq 1$, where $*$ is the product defined in Definition A.1.

Proof of Lemma A.1. Let A_i and B_i denote the i-th rows of A and I_k, respectively. We will reduce $A * I_k$ into echelon form by using both elementary row and column transformations. Let $0_k = (0, \ldots, 0)$ denote a row with k zeros. (Here, the symbol " $\Longleftrightarrow$ " is read "has the same rank as", and numbers above this symbol refer to explanations below.) Then,

$$A * I_k =$$

$$\left(\begin{array}{cc} A_1 & B_1 \\ A_2 & B_1 \\ \vdots & \vdots \\ A_n & B_1 \\ \text{- - - - -} & \\ A_1 & B_2 \\ A_2 & B_2 \\ \vdots & \vdots \\ A_n & B_2 \\ \text{- - - - -} & \\ \vdots & \\ \text{- - - - -} & \\ A_1 & B_k \\ A_2 & B_k \\ \vdots & \vdots \\ A_n & B_k \end{array}\right) \overset{1}{\Longleftrightarrow} \left(\begin{array}{cc} B_1 & A_1 \\ B_1 & A_2 \\ \vdots & \vdots \\ B_1 & A_n \\ \text{- - - - -} & \\ B_2 & A_1 \\ B_2 & A_2 \\ \vdots & \vdots \\ B_2 & A_n \\ \text{- - - - -} & \\ \vdots & \\ \text{- - - - -} & \\ B_k & A_1 \\ B_k & A_2 \\ \vdots & \vdots \\ B_k & A_n \end{array}\right) \overset{2}{\Longleftrightarrow} \left(\begin{array}{cc} B_1 & A_1 \\ B_1 - B_1 & A_2 - A_1 \\ \vdots & \vdots \\ B_1 - B_1 & A_n - A_1 \\ \text{- - - - - - - - - - - -} & \\ B_2 & A_1 \\ B_2 - B_2 & A_2 - A_1 \\ \vdots & \vdots \\ B_2 - B_2 & A_n - A_1 \\ \text{- - - - - - - - - - - -} & \\ \vdots & \\ \text{- - - - - - - - - - - -} & \\ B_k & A_1 \\ B_k - B_k & A_2 - A_1 \\ \vdots & \vdots \\ B_k - B_k & A_n - A_1 \end{array}\right)$$

$$
= \left(\begin{array}{cc}
B_1 & A_1 \\
0_k & A_2 - A_1 \\
\vdots & \vdots \\
0_k & A_n - A_1 \\
\text{- - - - - - - - -} & \\
B_2 & A_1 \\
0_k & A_2 - A_1 \\
\vdots & \vdots \\
0_k & A_n - A_1 \\
\text{- - - - - - - - -} & \\
\multicolumn{2}{c}{\vdots} \\
\text{- - - - - - - - -} & \\
B_k & A_1 \\
0_k & A_2 - A_1 \\
\vdots & \vdots \\
0_k & A_n - A_1
\end{array}\right)
\overset{3}{\Longleftrightarrow}
\left(\begin{array}{cc}
B_1 & A_1 \\
B_2 & A_1 \\
\vdots & \vdots \\
B_k & A_1 \\
\text{- - - - - - - - -} & \\
0_k & A_2 - A_1 \\
\vdots & \vdots \\
0_k & A_n - A_1 \\
\text{- - - - - - - - -} & \\
0_k & A_2 - A_1 \\
\vdots & \vdots \\
0_k & A_n - A_1 \\
\text{- - - - - - - - -} & \\
\multicolumn{2}{c}{\vdots} \\
\text{- - - - - - - - -} & \\
0_k & A_2 - A_1 \\
\vdots & \vdots \\
0_k & A_n - A_1
\end{array}\right)
$$

$$
\overset{4}{\Longleftrightarrow}
\left(\begin{array}{cc}
B_1 & A_1 \\
B_2 & A_1 \\
\vdots & \vdots \\
B_k & A_1 \\
\text{- - - - - - - - -} & \\
0_k & A_2 - A_1 \\
\vdots & \vdots \\
0_k & A_n - A_1 \\
\text{- - - - - - - - -} & \\
0_k & 0_p \\
\vdots & \vdots \\
0_k & 0_p \\
\text{- - - - - - - - -} & \\
\multicolumn{2}{c}{\vdots} \\
\text{- - - - - - - - -} & \\
0_k & 0_p \\
\vdots & \vdots \\
0_k & 0_p
\end{array}\right)
\overset{5}{\Longleftrightarrow}
\left(\begin{array}{cc}
B_1 & A_1 \\
B_2 & A_1 \\
\vdots & \vdots \\
B_k & A_1 \\
\text{- - - - - - - - -} & \\
0_k & A_2 - A_1 \\
\vdots & \vdots \\
0_k & A_n - A_1
\end{array}\right)
$$

$$\iff \begin{pmatrix} B_1 & A_1 \\ B_1 - B_2 & 0_p \\ \vdots & \vdots \\ B_1 - B_k & 0_p \\ \hline 0_k & A_2 - A_1 \\ \vdots & \vdots \\ 0_k & A_n - A_1 \end{pmatrix} \overset{6}{\iff} \begin{pmatrix} B_1 & A_1 \\ B_1 - B_2 & 0_p \\ \vdots & \vdots \\ B_1 - B_k & 0_p \\ \hline B_1 & A_2 \\ \vdots & \vdots \\ B_1 & A_n \end{pmatrix} \overset{7}{\iff} \Delta,$$

$$\text{where } \Delta = \begin{pmatrix} B_1 - B_2 & 0_p \\ \vdots & \vdots \\ B_1 - B_k & 0_p \\ \hline B_1 & A_1 \\ B_1 & A_2 \\ \vdots & \vdots \\ B_1 & A_n \end{pmatrix} = \left(\begin{array}{cccc|c} 1 & -1 & \cdots & 0 & 0_p \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & \cdots & -1 & 0_p \\ \hline 1 & 0 & \cdots & 0 & A_1 \\ 1 & 0 & \cdots & 0 & A_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & \cdots & 0 & A_n \end{array}\right).$$

where Δ is a matrix with dimension $\{k-1+n\}\times\{k+p\}$. Above, $\overset{1}{\iff}$ is obtained by switching columns; $\overset{2}{\iff}$ is obtained by subtracting the first row of the block from the other block's rows; $\overset{3}{\iff}$ by moving the first row of each block to the upper part of the matrix; $\overset{4}{\iff}$ by subtracting the second block from the blocks below it; $\overset{5}{\iff}$ by deleting rows with zeroes; $\overset{6}{\iff}$ by adding first row to rows in second block; and $\overset{7}{\iff}$ by putting first row at top of second block.

Now observe that the last p columns of Δ constitute the matrix

$$H = \begin{pmatrix} 0_p \\ \vdots \\ 0_p \\ \hline A \end{pmatrix}_{\{k-1+n\}\times p}$$

By hypothesis, $J_{k-1+n} - Hc = J_{k-1+n} - \begin{pmatrix} 0_{k-1}^T \\ J_n \end{pmatrix} = \begin{pmatrix} J_{k-1} \\ 0_n^T \end{pmatrix}$. Also, the first column of Δ is J_{k-1+n}. By replacing the first column of Δ with the column $J_{k-1+n} - Hc$, which can be viewed as a sequence of elementary column transfor-

mations on matrix Δ, we obtain that $A * I_k$ has the same rank as

$$\left(\begin{array}{cccc|c} 1 & -1 & \cdots & 0 & 0_p \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & \cdots & -1 & 0_p \\ \hline 0 & 0 & \cdots & 0 & A_1 \\ 0 & 0 & \cdots & 0 & A_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & A_n \end{array}\right)$$

The rank is $\rho(A) + k - 1$.

Proof of Theorem 2. If subject ω underwent sequence S, denote $\overline{W}_\omega = \overline{W}(S)$, $\omega = 1, \ldots, N$. Denote $\overline{Q} = \left(\overline{W}_1^T, \ldots, \overline{W}_N^T\right)^T$. Clearly, for fixed i, all subjects who experienced sequence $S^{(i)}$ have the same matrix $\overline{W}$. Note that $\rho(D) = \rho\left(\overline{Q}\right)$ because, for any $\omega = 1, \ldots, N$, there exists $i = 1, \ldots, d$ such that $\overline{W}_\omega = \overline{W}\left(S^{(i)}\right)$ and, for any i, there exists ω such that $\overline{W}_\omega = \overline{W}\left(S^{(i)}\right)$, which imply that any row in $\overline{Q}$ is equal to some row in D, and any row in D is equal to some row in $\overline{Q}$.

Suppose first that

$$\mathcal{C} = \{A_{1,1}, ..., A_{1,l_1}\} \times \{A_{2,1}, ..., A_{2,l_2}\} \times \cdots \times \left\{A_{q,1}, ..., A_{q,l_q}\right\} \tag{6}$$

that is, that all possible decision combinations occur but Ø does not occur at any stage. Then, $d = \prod_{i=1}^{q} l_i$. Denote $a_0 = d$, $a_j = d \Big/ \prod_{i=1}^{j} l_i$, $j = 1, \ldots, q$; and $G_0 = \exists_{a_0} 1$, $G_1 = \exists_{a_1} I_{l_1}$, $G_2 = \exists_{a_2} \{I_{l_1} * I_{l_2}\}$, $G_3 = \exists_{a_3} \{(I_{l_1} * I_{l_2}) * I_{l_3}\}$, ..., and $G_q = \exists_{a_q} \left\{(\cdots((I_{l_1} * I_{l_2}) * I_{l_3}) * \cdots) * I_{l_q}\right\}$, where $*$ is the product defined in Definition A.1. Then, the rows of D can be rearranged as

$$D^* = \left(\begin{array}{cccc} G_0 & \overrightarrow{\mathbf{0}} & \ldots & \overrightarrow{\mathbf{0}} \\ \overrightarrow{\mathbf{1}} & G_1 & \ldots & \overrightarrow{\mathbf{0}} \\ \vdots & \vdots & \ddots & \vdots \\ \overrightarrow{\mathbf{1}} & \overrightarrow{\mathbf{0}} & \overrightarrow{\mathbf{0}} & G_q \end{array}\right)_{\{d(q+1)\} \times m_q}$$

where the $\overrightarrow{\mathbf{1}}$'s are column vectors with only ones, not all of the same dimension. By Lemma A.1, $\rho(G_0) = 1$, $\rho(G_1) = l_1$ and $\rho(G_j) = l_1 + \sum_{i=2}^{j}(l_i - 1)$, $j = 2, \ldots, q$. Hence, if $\mathcal{C}$ satisfies equation (6),

$$\rho\left(\overline{Q}\right) = \rho(D) = \rho(D^*) = \sum_{j=0}^{q} \rho(G_j) = 1 + m_q - q(q-1)/2 < 1 + m_q$$

(Recall that $q \geq 2$.)

Now note that if S is not skip for any $S \in \mathcal{C}$, and if there exists $S \in \mathcal{C}$ such that $S \in \mathcal{A}_r^*$ for some $r < q$, that is, if at least one subject did not receive a q-stage

treatment, then the number of rows of D is $\leq d(q+1)$; or if $S \in \mathcal{A}_q^*$ for all $S \in \mathcal{C}$ but the number of sequences in $\mathcal{C}$ is $\leq \prod_{i=1}^{q} l_i$, then the number of rows of D is also $\leq d(q+1)$. In both situations, $\rho\left(\overline{Q}\right) \leq 1 + m_q - q(q-1)/2$.

But, K is of full rank if and only if $\rho(K) = 1+p+m_q$. This implies that K will not be of full rank if $\rho(K) = \rho\left(\overline{Q}\right) + p$ but S is not skip for any $S \in \mathcal{C}$, because, in such case, $\rho(K) \leq 1 + p + m_q - q(q-1)/2$. In particular, if $\rho(K) = \rho\left(\overline{Q}\right) + p$ and $\mathcal{C}$ satisfies equation (6), then $\rho(K) = 1 + p + m_q - q(q-1)/2$, which is the maximum possible rank of K when no subject experiences a skip sequence.

Proof of Theorem 3. By reorganizing the rows of D_1 and subtracting rows from their duplicates, and then eliminating rows with only zeros, we obtain

$D_1 \iff$

$$\begin{pmatrix}
1 & \vec{\mathbf{0}} & \vec{\mathbf{0}} & \vec{\mathbf{0}} & \cdots & \vec{\mathbf{0}} \\
\vec{\mathbf{0}} & I_{l_1} & \vec{\mathbf{0}} & \vec{\mathbf{0}} & \cdots & \vec{\mathbf{0}} \\
\vec{\mathbf{0}} & \vec{\mathbf{0}} & \exists_c^{i=1}\left(S_1^{\langle i\rangle}, S_2^{\langle i\rangle}\right) & \vec{\mathbf{0}} & \cdots & \vec{\mathbf{0}} \\
\vec{\mathbf{0}} & \vec{\mathbf{0}} & \vec{\mathbf{0}} & \exists_c^{i=1}\left(S_1^{\langle i\rangle}, S_2^{\langle i\rangle}, S_3^{\langle i\rangle}\right) & \cdots & \vec{\mathbf{0}} \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
\vec{\mathbf{0}} & \vec{\mathbf{0}} & \vec{\mathbf{0}} & \vec{\mathbf{0}} & \cdots & \vec{\mathbf{0}} \\
\vec{\mathbf{0}} & \vec{\mathbf{0}} & \vec{\mathbf{0}} & \vec{\mathbf{0}} & \cdots & \exists_c^{i=1}\left(S_1^{\langle i\rangle}, \ldots, S_q^{\langle i\rangle}\right)
\end{pmatrix}$$

where I_{l_i} is the $l_i \times l_i$ identity matrix, the $\vec{\mathbf{0}}$'s represent matrices of zeros, and, as before, the " $\iff$ " is read "has the same rank as". [We obtained I_{l_1} in place of $\exists_c^{i=1} S_1^{\langle i\rangle}$ in the above matrix because of the assumption that every 1-stage treatment is implemented in the first stage of at least one of the sequences $S^{\langle 1\rangle}, \ldots, S^{\langle c\rangle}$.]

By applying elementary row operations to $\exists_c^{i=1}\left(S_1^{\langle i\rangle}, S_2^{\langle i\rangle}\right)$, we can obtain an echelon matrix E_2 which has the form

$$E_2 = \begin{pmatrix} I_{l_1} & B_{12} \\ \vec{\mathbf{0}} & C_2 \\ \vec{\mathbf{0}} & \vec{\mathbf{0}} \end{pmatrix}$$

where C_2 is also a row echelon matrix. Noting that the number of columns of C_2 is l_2, let $\tau_2 = l_2 - \rho(C_2)$, and let $w_{2,1} < \cdots < w_{2,\tau_2}$ be the indices of the columns of C_2 that do not have a pivot. Recall that the rank of a matrix is equal to the number of columns minus the number of columns without pivots. Clearly, $1 \leq w_{2,1} < \cdots < w_{2,\tau_2} \leq l_2$.

For $i = 2, \ldots, q$ and $j = 1, \ldots, l_i$, consider the family of skip sequences,

$$\mathcal{F}(i,j) = \mathcal{F}(A_{i,j}) = \left\{(S_1, \ldots, S_q) \in \mathcal{A}_q^* \,; S_i = A_{i,j} \text{ and } S_k = \text{Ø for } k < i\right\}$$

A particular sequence in $\mathcal{F}(i,j)$ will be denoted $\mathcal{F}(i,j,*)$. If $\mathcal{F}(i,j,*)=(S_1,\ldots,S_q)$, we denote $\mathcal{F}(i,j,*)_t=(S_1,\ldots,S_t)$ for $1\leq t\leq q$.

Now take τ_2 skip sequences $\mathcal{F}(2,w_{2,1},*),\ldots,\mathcal{F}(2,w_{2,\tau_2},*)$ from $\mathcal{F}(2,w_{2,1}),\ldots,\mathcal{F}(2,w_{2,\tau_2})$, respectively, and consider the matrix

$$M_2=\begin{pmatrix} \exists_c^{i=1}\left(S_1^{\langle i\rangle},S_2^{\langle i\rangle}\right) \\ \text{- - - - - - - - - - -} \\ \exists_{\tau_2}^{j=1}\mathcal{F}(2,w_{2,j},*)_2 \end{pmatrix}$$

By using elementary row operations,

$$M_2 \iff \begin{pmatrix} E_2 \\ \text{- - - - - - - - - - - -} \\ \exists_{\tau_2}^{j=1}\ \mathcal{F}(2,w_{2,j},*)_2 \end{pmatrix} \iff E_2':=\begin{pmatrix} I_{l_1} & B_{12} \\ \overset{1}{\mathbf{0}} & I_{l_2} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$$

because $\exists_{\tau_2}^{j=1}\mathcal{F}(2,w_{2,j},*)_2$ has 1's only in its columns $l_1+w_{2,1}<\cdots<l_1+w_{2,\tau_2}$, which are matched with the columns of C_2 that do not have a pivot after converting the upper part of M_2 into E_2. Then, $\rho(M_2)=l_1+l_2$. Now observe that, by starting with the same elementary row operations used to obtain E_2, we can obtain

$$\exists_c^{i=1}\left(S_1^{\langle i\rangle},S_2^{\langle i\rangle},S_3^{\langle i\rangle}\right) \iff E_3:=\begin{pmatrix} I_{l_1} & B_{12} & B_{13} \\ \overset{1}{\mathbf{0}} & C_2 & B_{23} \\ \mathbf{0} & \overset{1}{\mathbf{0}} & C_3 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix}$$

where C_3 is row echelon. Denoting $\tau_3=l_3-\rho(C_3)$, let $w_{3,1}<\cdots<w_{3,\tau_3}$ be the columns of C_3 that do not have a pivot. Now take τ_3 sequences $\mathcal{F}(3,w_{3,1},*),\ldots,\mathcal{F}(3,w_{3,\tau_3},*)$ from $\mathcal{F}(3,w_{3,1}),\ldots,\mathcal{F}(3,w_{3,\tau_3})$, respectively. By using elementary row operations, we obtain

$$M_3:=\begin{pmatrix} \exists_c^{i=1}\left(S_1^{\langle i\rangle},S_2^{\langle i\rangle},S_3^{\langle i\rangle}\right) \\ \text{- - - - - - - - - - - -} \\ \exists_{\tau_2}^{j=1}\ \mathcal{F}(2,w_{2,j},*)_3 \\ \text{- - - - - - - - - - - -} \\ \exists_{\tau_3}^{j=1}\ \mathcal{F}(3,w_{3,j},*)_3 \end{pmatrix} \iff \begin{pmatrix} E_3 \\ \text{- - - - - - - - - - - -} \\ \exists_{\tau_2}^{j=1}\ \mathcal{F}(2,w_{2,j},*)_3 \\ \text{- - - - - - - - - - - -} \\ \exists_{\tau_3}^{j=1}\ \mathcal{F}(3,w_{3,j},*)_3 \end{pmatrix}$$

$$\iff \begin{pmatrix} I_{l_1} & B_{12} & B_{13} \\ \overset{1}{\mathbf{0}} & I_{l_2} & B_{23} \\ \mathbf{0} & \overset{1}{\mathbf{0}} & C_3 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \text{- - - - - - - - - - - -} \\ \exists_{\tau_3}^{j=1}\ \mathcal{F}(3,w_{3,j},*)_3 \end{pmatrix} \iff \begin{pmatrix} I_{l_1} & B_{12} & B_{13} \\ \overset{1}{\mathbf{0}} & I_{l_2} & B_{23} \\ \mathbf{0} & \overset{1}{\mathbf{0}} & I_{l_3} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix}$$

where the second equivalence was obtained by using row operations analogous to those used to obtain E_2', and the third equivalence is because the 1's of $\exists_{\tau_3}^{j=1} \mathcal{F}(3, w_{3,j}, *)_3$ are aligned vertically with the columns of C_3 that do not have a pivot. Then, $\rho(M_3) = l_1 + l_2 + l_3$.

Thus, working iteratively in this way, we can construct $q-1$ sequences of numbers, namely $w_{i,1} < \cdots < w_{i,\tau_i}$, $i = 2, \ldots, q$, and matrices of the form

$$M_k := \begin{pmatrix} \exists_c^{i=1} \left(S_1^{\langle i \rangle}, \ldots, S_k^{\langle i \rangle} \right) \\ \text{- - - - - - - - - - - - - - -} \\ \exists_k^{h=2} \exists_{\tau_h}^{j=1} \ \mathcal{F}(h, w_{h,j}, *)_k \end{pmatrix}, \quad k = 2, \ldots, q,$$

such that $\rho(M_k) = l_1 + \cdots + l_k$. By construction, since $\Gamma = \exists_c^{i=1} \left(S_1^{\langle i \rangle}, \ldots, S_q^{\langle i \rangle} \right)$, a conversion of Γ into a row echelon matrix shows that $\rho(\Gamma) = \sum_{i=1}^{q} l_i - \tau$, where $\tau = \sum_{i=2}^{q} \tau_i$.

Now observe that, for all $i = 2, \ldots, q$ and $j = 1, \ldots, l_i$,

$\overline{W}(\mathcal{F}(i, j, *)) =$

$$\begin{pmatrix} 1 & \overrightarrow{\mathbf{0}}_{m_{i-1}} & \overrightarrow{\mathbf{0}}_{m_i - m_{i-1}} & \overrightarrow{\mathbf{0}}_{m_{i+1} - m_i} & \overrightarrow{\mathbf{0}}_{m_{i+2} - m_{i+1}} & \cdots & \overrightarrow{\mathbf{0}}_{m_q - m_{q-1}} \\ 1 & \overrightarrow{\mathbf{0}}_{m_{i-1}} & \mathcal{F}(i, j, *)_i & \overrightarrow{\mathbf{0}}_{m_{i+1} - m_i} & \overrightarrow{\mathbf{0}}_{m_{i+2} - m_{i+1}} & \cdots & \overrightarrow{\mathbf{0}}_{m_q - m_{q-1}} \\ 1 & \overrightarrow{\mathbf{0}}_{m_{i-1}} & \overrightarrow{\mathbf{0}}_{m_i - m_{i-1}} & \mathcal{F}(i, j, *)_{i+1} & \overrightarrow{\mathbf{0}}_{m_{i+2} - m_{i+1}} & \cdots & \overrightarrow{\mathbf{0}}_{m_q - m_{q-1}} \\ 1 & \overrightarrow{\mathbf{0}}_{m_{i-1}} & \overrightarrow{\mathbf{0}}_{m_i - m_{i-1}} & \overrightarrow{\mathbf{0}}_{m_{i+1} - m_i} & \mathcal{F}(i, j, *)_{i+2} & \cdots & \overrightarrow{\mathbf{0}}_{m_q - m_{q-1}} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \overrightarrow{\mathbf{0}}_{m_{i-1}} & \overrightarrow{\mathbf{0}}_{m_i - m_{i-1}} & \overrightarrow{\mathbf{0}}_{m_{i+1} - m_i} & \overrightarrow{\mathbf{0}}_{m_{i+2} - m_{i+1}} & \cdots & \mathcal{F}(i, j, *)_q \end{pmatrix}$$

where $\overrightarrow{\mathbf{0}}_k$ is a row vector of zeros of dimension k, and the dimension of $\overline{W}(\mathcal{F}(i, j, *))$ is $(q - i + 2) \times (m_q + 1)$.

Then, reordering rows,

$$D = \begin{pmatrix} D_1 \\ \text{- - - - - - - - - - - -} \\ \exists_{\tau_2}^{j=1} \overline{W} \left(\mathcal{F}(2, w_{2,j}, *) \right) \\ \exists_{\tau_3}^{j=1} \overline{W} \left(\mathcal{F}(3, w_{3,j}, *) \right) \\ \vdots \\ \exists_{\tau_q}^{j=1} \overline{W} \left(\mathcal{F}(q, w_{q,j}, *) \right) \end{pmatrix} \iff$$

$$\left(\begin{array}{ccccc}
1 & \overset{1}{\mathbf{0}} & \overset{1}{\mathbf{0}} & \overset{1}{\mathbf{0}} & \overset{1}{\mathbf{0}} \\
\overset{1}{\mathbf{0}} & I_{l_1} & \overset{1}{\mathbf{0}} & \overset{1}{\mathbf{0}} & \overset{1}{\mathbf{0}} \\
\overset{1}{\mathbf{0}} & \overset{1}{\mathbf{0}} & \exists_c^{i=1}\left(S_1^{\langle i\rangle}, S_2^{\langle i\rangle}\right) & \overset{1}{\mathbf{0}} & \overset{1}{\mathbf{0}} \\
\overset{1}{\mathbf{0}} & \overset{1}{\mathbf{0}} & \overset{1}{\mathbf{0}} & \exists_c^{i=1}\left(S_1^{\langle i\rangle}, S_2^{\langle i\rangle}, S_3^{\langle i\rangle}\right) & \overset{1}{\mathbf{0}} \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
\overset{1}{\mathbf{0}} & \overset{1}{\mathbf{0}} & \overset{1}{\mathbf{0}} & \overset{1}{\mathbf{0}} & \overset{1}{\mathbf{0}} \\
\overset{1}{\mathbf{0}} & \overset{1}{\mathbf{0}} & \overset{1}{\mathbf{0}} & \overset{1}{\mathbf{0}} & \exists_c^{i=1}\left(S_1^{\langle i\rangle}, \ldots, S_q^{\langle i\rangle}\right) \\
- & - & - - - - - - - - - - - & - - - - - - - - - - - & - - - - - - - - - - - \\
\overset{1}{\mathbf{0}} & \overset{1}{\mathbf{0}} & \exists_{\tau_2}^{j=1}\mathcal{F}\left(2, w_{2,j}, *\right)_2 & \overset{1}{\mathbf{0}} & \overset{1}{\mathbf{0}} \\
\overset{1}{\mathbf{0}} & \overset{1}{\mathbf{0}} & \overset{1}{\mathbf{0}} & \exists_3^{h=2}\exists_{\tau_h}^{j=1}\mathcal{F}\left(h, w_{h,j}, *\right)_3 & \overset{1}{\mathbf{0}} \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
\overset{1}{\mathbf{0}} & \overset{1}{\mathbf{0}} & \overset{1}{\mathbf{0}} & \overset{1}{\mathbf{0}} & \overset{1}{\mathbf{0}} \\
\overset{1}{\mathbf{0}} & \overset{1}{\mathbf{0}} & \overset{1}{\mathbf{0}} & \overset{1}{\mathbf{0}} & \exists_q^{h=2}\exists_{\tau_h}^{j=1}\mathcal{F}\left(h, w_{h,j}, *\right)_q
\end{array}\right)$$

We finally obtain

$$D \iff \left(\begin{array}{cccccc}
1 & \overset{1}{\mathbf{0}} & \overset{1}{\mathbf{0}} & \overset{1}{\mathbf{0}} & \cdots & \overset{1}{\mathbf{0}} \\
\overset{1}{\mathbf{0}} & I_{l_1} & \overset{1}{\mathbf{0}} & \overset{1}{\mathbf{0}} & \cdots & \overset{1}{\mathbf{0}} \\
\overset{1}{\mathbf{0}} & \overset{1}{\mathbf{0}} & M_2 & \overset{1}{\mathbf{0}} & \cdots & \overset{1}{\mathbf{0}} \\
\overset{1}{\mathbf{0}} & \overset{1}{\mathbf{0}} & \overset{1}{\mathbf{0}} & M_3 & \cdots & \overset{1}{\mathbf{0}} \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
\overset{1}{\mathbf{0}} & \overset{1}{\mathbf{0}} & \overset{1}{\mathbf{0}} & \overset{1}{\mathbf{0}} & \cdots & M_q
\end{array}\right)$$

and conclude that $\rho(D) = 1 + l_1 + \sum_{i=2}^{q} \rho\left(M_i\right) = 1 + m_q$, which is equal to the number of columns of D. Thus, with the selected skip sequences, D is of full rank.

Proof of Theorem 4. By Theorem 1, $E_\omega[\boldsymbol{Y}_\omega^{(r)}(S)] = g^{-1}\left(W\left(S, \boldsymbol{X}_\omega\right)\boldsymbol{\gamma}_\omega\right)$, and similarly for T. By the Interchangeability Assumption, $\boldsymbol{Y}_\omega^{(r)}(S)$ and $\boldsymbol{Y}_\omega^{(t)}(T)$ have the same distribution. Hence, $E_\omega[\boldsymbol{Y}_\omega^{(r)}(S)] = E_\omega[\boldsymbol{Y}_\omega^{(t)}(T)]$. Then,

$$g^{-1}\left(W\left(S, \boldsymbol{X}_\omega\right)\boldsymbol{\gamma}_\omega\right) = g^{-1}\left(W\left(T, \boldsymbol{X}_\omega\right)\boldsymbol{\gamma}_\omega\right).$$

Appendix B. Examples of Antichronic Systems with a Full-Rank Design Matrix

Example B.1. Consider the simplest ASs, which occur when $q = 2$ and $l_1 = l_2 = 2$. In this case, A_q^* includes only four non-skip sequences, namely: $S^{\langle 1\rangle} = (A_{1,1}, A_{2,1})$, $S^{\langle 2\rangle} = (A_{1,1}, A_{2,2})$, $S^{\langle 3\rangle} = (A_{1,2}, A_{2,1})$, and $S^{\langle 4\rangle} = (A_{1,2}, A_{2,2})$, for

which

$$\overline{W}\left(S^{\langle 1\rangle}\right) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 \end{pmatrix}$$

$$\overline{W}\left(S^{\langle 2\rangle}\right) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 \end{pmatrix}$$

$$\overline{W}\left(S^{\langle 3\rangle}\right) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 0 \end{pmatrix}$$

$$\text{and} \quad \overline{W}\left(S^{\langle 4\rangle}\right) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 \end{pmatrix}$$

Suppose that $\mathcal{C}_1 = \left\{S^{\langle 1\rangle}, \ldots, S^{\langle 4\rangle}\right\}$ is an antichronic system. By Theorem 2, an estimator of $E[\boldsymbol{\gamma}]$ cannot be computed if only the sequences in $\mathcal{C}_1$ are implemented in a CT, even if each sequence in $\mathcal{C}_1$ is administered to at least one subject, because, in this case, $m_q = 6$, $q(q-1)/2 = 1$ and $\rho(K) = p+6 < p+7$ (see proof of Theorem 2); that is, the study design matrix will never be of full rank. Only two SSs are possible, namely $S^{\langle 5\rangle} = (\Upsilon, A_{2,1})$ and $S^{\langle 6\rangle} = (\Upsilon, A_{2,2})$, for which, respectively,

$$\overline{W}\left(S^{\langle 5\rangle}\right) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

$$\text{and} \quad \overline{W}\left(S^{\langle 6\rangle}\right) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

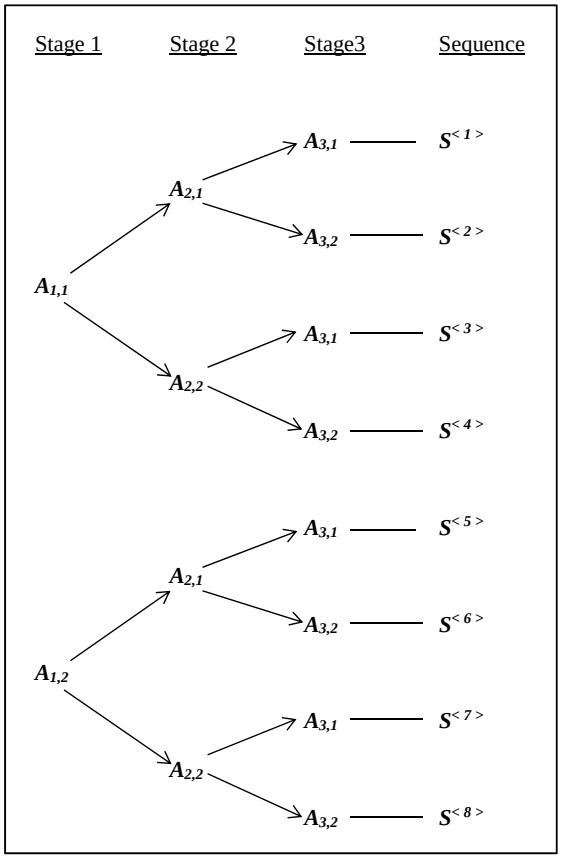
Thus, for $q = l_1 = l_2 = 2$, the design matrix will be of full rank for some systems including at least one SS, provided each sequence of the system is administered to at least one subject. Examples of such systems are

$$\mathcal{C}_2 = \mathcal{C}_1 \cup \left\{S^{\langle 5\rangle}\right\};\ \mathcal{C}_3 = \mathcal{C}_1 \cup \left\{S^{\langle 6\rangle}\right\};\ \mathcal{C}_4 = \mathcal{C}_1 \cup \left\{S^{\langle 5\rangle}, S^{\langle 6\rangle}\right\},\ \text{and}$$
$$\mathcal{C}_5 = \left\{S^{\langle 1\rangle}, S^{\langle 2\rangle}, S^{\langle 3\rangle}, S^{\langle 5\rangle}\right\}$$

To show that the design matrix is of full rank for $\mathcal{C}_5$, say, just observe that the rank of $[\overline{W}\left(S^{\langle 1\rangle}\right)^T, \overline{W}\left(S^{\langle 2\rangle}\right)^T, \overline{W}\left(S^{\langle 3\rangle}\right)^T, \overline{W}\left(S^{\langle 5\rangle}\right)^T]^T$ is 7. Mathematical software such as *Mathematica* (Wolfram Research, Inc.), or the calculator TI-Nspire CX CAS (Texas Instruments Inc.) are helpful in these rank computations.

Example B.2. Now suppose that $q = 3$ and $l_1 = l_2 = l_3 = 2$. Here, A_q^* includes 8 possible non-skip sequences (see Figure 2). For instance, in the Figure, $S^{\langle 3\rangle} = (A_{1,1}, A_{2,2}, A_{3,1})$. Consider the skip sequences $S^{\langle 9\rangle} = (\Upsilon, \Upsilon, A_{3,2})$ and $S^{\langle 10\rangle} =$

FIGURE 2: Graphical representation of all possible non-skip sequences with $q = 3$ and $l_1 = l_2 = l_3 = 2$.



$(\Upsilon, A_{2,2}, A_{3,1})$. To illustrate,

$$\overline{W}\left(S^{\langle 3\rangle}\right) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \end{pmatrix}$$

$$\overline{W}\left(S^{\langle 9\rangle}\right) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\overline{W}\left(S^{\langle 10\rangle}\right) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \end{pmatrix}$$

If $\mathcal{C}_1 = \left\{S^{\langle 1\rangle}, \ldots, S^{\langle 8\rangle}\right\}$, the design matrix will be of full rank for the systems

$$\mathcal{C}_2 = \mathcal{C}_1 \cup \left\{S^{\langle 9\rangle}, S^{\langle 10\rangle}\right\} \text{ and } \mathcal{C}_3 = \left\{S^{\langle 1\rangle}, S^{\langle 3\rangle}, \ldots, S^{\langle 7\rangle}\right\} \cup \left\{S^{\langle 9\rangle}, S^{\langle 10\rangle}\right\}$$

provided each sequence in the system is administered to at least one subject. Other systems for which the design matrix is of full rank are possible when both $q = 3$ and $l_1 = l_2 = l_3 = 2$. For instance, in the definition of $S^{\langle 10\rangle}$ in $\mathcal{C}_2$, $A_{2,1}$ could be used in place of $A_{2,2}$.

Appendix C. Description of Supplementary Material (Computer Code)

The author wrote a Stata command (asdesign) which creates the design matrix using three input variables: the subject identifier, the stage and the treatment variables. A help file for asdesign is also included, which additionally shows code for fitting antichronic systems.

The included files are:

asdesign.ado: Stata program that runs as a regular command.

asdesign.sthlp: Help file for asdesign.

Instructions for asdesign.docx: Word file with instructions to install asdesign.

The above files are available from the author on request.