Published in final edited form as:

Stat Probab Lett. 2020 January ; 156: . doi:10.1016/j.spl.2019.108585.

A Note on Monotonicity in Repeated Attempt Selection Models

Seunghwan Park,

Department of Information Statistics, Kangwon National University, Chuncheon, Gangwon 24341, South Korea

Michael J. Daniels

Department of Statistics, University of Florida, Gainesville, FL 32611

Abstract

Study designs where follow-up samples are collected through multiple attempts have been called repeated attempts designs. In this note we explore the monotonicity thought to underly the models fit for these designs and show it does not always hold.

Keywords

missing data; pattern mixture model; repeated attempt design

1 Introduction

Outcomes are often collected after multiple attempts in follow-up studies (Gray et al., 2006; Müssener et al., 2016). Wood et al. (2006) referred to this type of study design as a “repeated attempt design” (RAD) and the corresponding statistical models as “repeated attempt models” (RAM). The first RAM selection model (RAM-SM) was proposed by Alho (1990) with subsequent modifications in Wood et al. (2006), Jackson et al. (2010), and Jackson et al. (2012). Unfortunately, these models do not allow sensitivity parameters as defined in Rotnitzky et al. (1998) and Daniels and Hogan (2008). This is an issue because sensitivity parameters are necessary components for the analysis of missing data. As an alternative approach that allows sensitivity parameters, Daniels et al. (2015) proposed RAM pattern mixture models (RAM-PMM). In that manuscript, they explored connections between RAM-SM and RAM-PMM including the behavior of the pattern specific means induced by the RAM-SM. The intuition behind the development of RAM-SM implied that the means, as a function of the number of attempts until the response was collected, would be monotone. Empirically, under a logit link for the ‘missing data mechanism’, these means appeared to be monotone; this behavior was consistent with the approach used for sensitivity

stat.shpark@kangwon.ac.kr.

Web Appendix

The web-based appendices contain proofs from Section 2, theoretical results for the binary outcome in Section 3.1, further exploration of the logit and probit links, and the algorithm to compute the pattern specific means.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

parameters and identification in the RAM-PMM. However, no formal proof of monotonicity was given. Here we explore monotonicity of the RAM-SM pattern specific means more carefully and shows that it only holds for certain link functions for the missing data mechanism (a somewhat surprising result that potentially has implications for using non-logit links). This has important ramifications which we discuss in Section 3.

We review RAM-SM in Section 2. We prove (and disprove) monotonicity in Section 3. Section 4 provides a re-analysis of the QUATRO data demonstrating non-monotonicity. We wrap up and provide conclusions and recommendations in Section 5.

2 Models

Let Y be the outcome of interest and let X be the set of baseline covariates. We assume that there are up to K attempts to collect the response. Here, R denotes the number of attempts until the outcome is successfully collected, with $R = K + 1$ corresponding to being unable to collect the response even after K (the maximum number of) attempts.

2.1 Repeated Attempt Selection model (RAM-SM)

RAM-SM is a model for the joint distribution of the outcome and the repeated attempts data, conditional on covariates, based on a factorization of this model into a model for the marginal distribution of the outcome given covariates, and a model for the conditional distribution of the number of attempts until the outcome is collected given the outcome and covariates. The latter distribution is often referred to as the *missing data mechanism* of the RAM-SM.

The marginal distribution of the response Y is given by the model,

$$Y|X = x \sim f(y|x), \quad (1)$$

where $f(y|x)$ is a density function of Y given X . For example, we can use a normal distribution $N\{\mu(x), \sigma^2(x)\}$, where $\mu(x) = \beta_0 + \beta x$ and $\sigma^2(x) = \sigma^2$.

The missing data mechanism of the RAM, $p_k(x, y)$ is given by the model,

$$g\{P(R = k | R \geq k, X = x, Y = y)\} = \lambda_{0k} + h(x; \lambda_k) + \delta y, \quad (2)$$

where $h(\cdot)$ is known function, e.g., $h(x; \lambda_k) = \lambda_k x$, and $R = k$ if the k th attempt is successful and g is a link function such as logit, probit, or complementary log-log. The term λ_{0k} plays the role of the intercept; if g is the logit function, λ_{0k} is the log odds, for participants with $X = 0$ and $Y = 0$. The covariate effects may depend on k through λ_k . The term δ determines whether the missing data mechanism of the RAM is MNAR or MAR; if $\delta = 0$ the missing data mechanism of the RAM is MAR.

RAM-PMM have been proposed and try to exploit a monotone pattern in the pattern-specific means which is consistent with the intuition behind the RAM-SM; we provide more details on RAM-PMMs in the web appendix. In the next section, we explore whether a monotonic pattern *always* holds in RAM-SMs.

3 Consideration of monotonicity in pattern specific means

Daniels et al. (2015) used simulated data under a RAM-SM with the logit link function to assess how the value of δ in (2) impacts the pattern specific conditional mean, $\alpha^{(k)}$ as a function of k . The results confirmed that there is an empirical connection between RAM-SM and RAM-PMM; a monotone non-decreasing pattern in the pattern specific means for $\delta > 0$ and monotone non-increasing pattern for $\delta < 0$. However, no formal proof was provided and only the logit link was considered. We provide a proof for the logit link below.

Lemma 1

For non-negative integrable functions g and h ,

$$\frac{\int y h(y) g(y) dy}{\int h(y) g(y) dy} - \frac{\int y g(y) dy}{\int g(y) dy}$$

is positive or negative if all the quantities above are finite and $h(y)$ has the corresponding monotonicity with respect to y , g -almost everywhere.

Proof

$$\begin{aligned} \frac{\int y h(y) g(y) dy}{\int h(y) g(y) dy} - \frac{\int y g(y) dy}{\int g(y) dy} &= \frac{\int y h(y) g(y) dy \int g(y) dy - \int y g(y) dy \int h(y) g(y) dy}{\int h(y) g(y) dy \int g(y) dy} \\ &= \frac{\int g(y) dy}{\int h(y) g(y) dy} \left(\frac{\int y h(y) g(y) dy}{\int g(y) dy} - \frac{\int y g(y) dy \int h(y) g(y) dy}{\int g(y) dy \int g(y) dy} \right) \\ &= \frac{1}{\int h(y) \frac{g(y)}{\int g(y) dy} dy} \left(\int y h(y) \frac{g(y)}{\int g(y) dy} dy - \int y \frac{g(y)}{\int g(y) dy} dy \int h(y) \frac{g(y)}{\int g(y) dy} dy \right) \\ &= \frac{\text{Cov}(Y, h(Y))}{E(h(Y))}, \end{aligned}$$

where the probability space in the last expression is given by the density g . If h is monotone non-decreasing with respect to y , then the above quantity will be non-negative.

End of Proof

Result 1—The pattern specific means induced by the RAM-SM (1) and (2) with any continuous conditional density function $f(y|x)$ and g as a logit function are monotone across patterns.

Proof—This follows from Lemma 1 with details in the web appendix

The pattern specific means derived from the RAM-SM with a logit link are monotone in pattern. This formally connects the pattern-specific behavior of the RAM-SM and the RAM-PMM. We consider more general links in the next subsection.

3.1 Exploration of monotonicity for non-logit links

Let's consider the case of general link function g . To use Lemma 1 as in the proof with the logit link, $\pi_{t+1}(x, y)/\pi_t(x, y)$ should always be a decreasing function of y when $\delta > 0$ (and

vice versa). To check this, without loss of generality, assume $X=0$ and $p_{\lambda}(x, y) = F(\lambda_{0t} + \delta y)$, where $F(x)$ is a c.d.f. Then, we have

$$\frac{\partial \log\{\pi_{t+1}(x, y)/\pi_t(x, y)\}}{\partial y} = \delta \left[\frac{f(\lambda_{0t} + 1 + \delta y)}{F(\lambda_{0t} + 1 + \delta y)} - \frac{f(\lambda_{0t} + \delta y)}{\{1 - F(\lambda_{0t} + \delta y)\}F(\lambda_{0t} + \delta y)} \right]$$

where $f(x)$ is a density function.

Result 2—*The pattern specific means induced by the RAM-SM (1) and (2) with any continuous conditional density function $f(y|x)$ and g as a complementary log-log link are monotone across patterns.*

Proof—This follows from Lemma 1 with details in the web appendix

However, in general, we cannot guarantee that $\log\{\pi_{t+1}(x, y)/\pi_t(x, y)\}/y$ is always negative for $\delta > 0$.

Result 3—*There is no guarantee of monotonicity for the log-log link in the pattern specific means induced by the RAM-SM (1) and (2) with any continuous conditional density function $f(y|x)$.*

Proof by counterexample in web appendix

Result 4—*There is no guarantee of monotonicity for the probit link in the pattern specific means induced by the RAM-SM (1) and (2) with any continuous conditional density function $f(y|x)$.*

Proof by counterexample—For the case of probit link, i.e., $F(x) = \Phi(x)$, $\log\{\pi_{t+1}(x, y)/\pi_t(x, y)\}/y$ has the form of

$$\delta \left[\frac{\phi(\lambda_{0t} + 1 + \delta y)\Phi(\lambda_{0t} + \delta y)(1 - \Phi(\lambda_{0t} + \delta y)) - \phi(\lambda_{0t} + \delta y)\Phi(\lambda_{0t} + 1 + \delta y)}{\Phi(\lambda_{0t} + 1 + \delta y)\Phi(\lambda_{0t} + \delta y)(1 - \Phi(\lambda_{0t} + \delta y))} \right].$$

If $\lambda_{0t} = 0$, $\lambda_{0t+1} = -1.96$, $\delta = 1$ and $y = 0$, then the numerator is given by $\phi(-1.96) \times 0.25 - \phi(0) \times 0.025$. Since $\phi(-1.96) > \phi(-2) = \{\phi(0) \times (1/e^2)\}$ and $e^2 < 10$, the numerator is positive. If $\lambda_{0t} = 0$, $\lambda_{0t+1} = -1.96$, and $y = 1.96$, then the numerator is less than $\phi(0) \times 0.25 - \phi(1.96) \times 0.5$ and it has a negative value. Since $\log\{\pi_{t+1}(x, y)/\pi_t(x, y)\}/y$ is a continuous function with respect to y , $\log\{\pi_{t+1}(x, y)/\pi_t(x, y)\}$ is not a monotone function with respect to y . The numeric counterexample using these parameter values is given in Figure 1 which illustrates that there is no guarantee of monotonicity in pattern specific mean using the probit link. **End of Proof**

This somewhat counterintuitive result, that there is no guarantee of monotonicity for the probit and log-log links, has potentially important implications. First, non-monotonicity of the pattern specific means is a behavior not desired in RAM-SM. Second, the probit link, in particular, is sometimes preferred to the logit link in terms of simpler (Bayesian)

computations using data augmentation with the ultimate inference expected to be similar to the logit link. However, given Result 4, care should be used in choosing this link in RAM-SMs.

It turns out that the same results also hold for a binary outcome. Details can be found in the web appendix. In addition, a further examination of RAM-SMs, with logit and probit links, to better understand the lack of monotonicity can also be found in the web appendix.

4 Data example

The QUATRO trial (Gray et al., 2006) was a multi-center, singled-blinded randomized controlled trial of the effectiveness of adherence therapy for participants with schizophrenia. The trial recruited a total of 409 participants from four centers in the Netherlands, Germany, the United Kingdom, and Italy. Subjects were individually randomized to receive either adherence therapy (intervention) or health education (control). The outcome was measured at baseline and at a 52 week follow-up. The main objective of this multicenter trial was to assess the impact of adherence therapy on self-reported quality-of-life. Multiple attempts were made to collect the quality-of-life outcome (as many as nine attempts) but still the outcomes from all individuals were not collected.

Here, we illustrate non-monotonicity using the data from the QUATRO trial. In what follows, X_1 is baseline quality of life score and (X_2, X_3, X_4) are the indicators of the four centers. Up to 9 attempts were made to collect the outcome variable Y , the quality of life score at a follow-up of 52 weeks. We collapse the number of attempts similar to Daniels et al. (2015), but with 5 patterns so the number of attempts R takes the values in $\{1, 2, 3, 4, 5\}$; in particular, $R = \{1, 2, 3\}$ correspond to the corresponding number of attempts until the outcome is collected. $R = 4$ corresponds to the outcome being collected in more than three attempts and $R = 5$ corresponds to the outcome never being collected. All participants who provided outcome data after more than 3 attempts were placed in the $R = 4$ category. Here, we assume a randomized trial where Z denotes randomization to the intervention of interest, which in this case is adherence therapy.

We consider the following model:

$$Y|X = x, Z = z \sim N\{\mu(x, z), \sigma^2\},$$

where $\mu(x, z) = \beta_0 + \theta_Z + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4$. The missing data mechanism $p_k(x, y) \equiv P(R = k | R \leq k, X = x, Y = y)$ is specified as

$$g\{p_k(x, y, z)\} = \lambda_{0k} + \lambda_z z + \lambda_{1k} x_1 + \lambda_{2k} x_2 + \lambda_{3k} x_3 + \lambda_{4k} x_4 + \delta_1 y + \delta_2 yz, \quad (3)$$

where $R = k$ if the k th attempt is successful and g is the link function.

For the regression parameters, $\beta_0, \theta_Z, \beta_1, \dots, \beta_4, \lambda_{0k}, \lambda_z, \lambda_{1k}, \dots, \lambda_{4k}, \delta_1$ and δ_2 , diffuse normal priors (e.g. $N(0, 100)$) are specified and for the variance components, vague inverse gamma priors are specified.

4.1 Results

We consider two link functions in (3), logit and probit. Posterior summaries of the parameters are given in Web Appendix Table 1. We compute the pattern specific means $E(Y|R = t, x_1, x_2, x_3, x_4)$, $t = 1, \dots, 4$, for $x_1 = 40$, the average of the baseline score and $(x_2, x_3, x_4) = (0, 0, 0)$ (i.e., the London site). As shown in Web Appendix Table 2 and Figure 2, the pattern specific means with the logit link are monotonic (as proven in Result 1), but they are non-monotonic with the probit link. Of note in Web Appendix Table 2 is the larger intercepts for patterns 3 and 4 for both links which is consistent with the non-monotonicity result in the counterexample for the probit link.

5 Discussion

This note confirms the conjecture in Daniels et al. (2015) about the monotonicity of pattern specific means for the RAM-SM using a *logit* link. This is both desirable and the expected behavior of that model. However, we also found out that if the logit link is replaced by the probit link, which is often done in related models for computational efficiency (without much thought), especially for Bayesian implementations (Albert and Chib, 1993), there is *no guarantee of monotonicity* and we saw that in the data example in Section 4. As such, one should be cautious in using a probit link for computational efficiency in RAM-SM. Similarly, we saw guaranteed monotonicity for the complementary log-log link but not the log-log link. We were unable to provide a general result for an arbitrary link g .

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

SP and MD were partially supported by NIH grant CA 183854. The authors would like to thank Ian White and Dan Jackson for insightful comments that improved the manuscript.

References

- Albert JH and Chib S (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American statistical Association*, 88:669–679.
- Alho JM (1990). Adjusting for nonresponse bias using logistic regression. *Biometrika*, 77:617–624.
- Daniels MJ and Hogan JW (2008). *Missing data in longitudinal studies: Strategies for Bayesian modeling and sensitivity analysis*. CRC press.
- Daniels MJ, Jackson D, Feng W, and White IR (2015). Pattern mixture models for the analysis of repeated attempt designs. *Biometrics*, 71:1160–1167. [PubMed: 26149119]
- Gray R, Leese M, Bindman J, Becker T, Burti L, David A, and Schene A (2006). Adherence therapy for people with schizophrenia european multicentre randomised controlled trial. *The British journal of psychiatry*, 189:508–514. [PubMed: 17139034]
- Jackson D, Mason D, White IR, and Sutton S (2012). An exploration of the missing data mechanism in an internet based smoking cessation trial. *BMC medical research methodology*, 12:157. [PubMed: 23067272]
- Jackson D, White IR, and Leese M (2010). How much can we learn about missing data?: an exploration of a clinical trial in psychiatry. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 173:593–612.

- Müssener U, Bendtsen M, Karlsson N, White IR, McCambridge J, and Bendtsen P (2016). Effectiveness of short message service text-based smoking cessation intervention among university students: a randomized clinical trial. *JAMA internal medicine*, 176:321–328. [PubMed: 26903176]
- Rotnitzky A, Robins JM, and Scharfstein DO (1998). Semiparametric regression for repeated outcomes with nonignorable nonresponse. *Journal of the american statistical association*, 93:1321–1339.
- Wood AM, White IR, and Hotopf M (2006). Using number of failed contact attempts to adjust for nonignorable nonresponse. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 169:525–542.

Highlights

- Repeated attempts designs make multiple attempts to collect the outcome
- Selection models are the most common model for these designs
- The intuition is that the mean outcome changes monotonically with # of attempts
- This note proves that this does not hold in general

Figure 1:
Example of monotone pattern specific means with logit link and example of non-monotone pattern specific means with probit link (counterexample from Result 4).

Figure 2:
Under RAM-SM with logit and probit links, pattern specific means for the outcome in the QUATRO example.