Enhanced Tipping-Point Displays

Victoria Liublinska* Donald B. Rubin*

Abstract

Assumptions about the missingness mechanism often cannot be assessed empirically, which calls for the sensitivity analyses. However, few studies with missing values are subjected to such analyses due to the lack of clear guidelines on a systematic exploration of alternative assumptions as well as the difficulty of formulating plausible missing not at random (MNAR) models. We present graphical displays, based on the "tipping-point" analysis first introduced in Yan et al. (2009), that help us visualize the results of a set of sensitivity analyses for missing outcomes in studies that compare two treatments. The resulting "enhanced tipping-point displays" provide compact summaries of conclusions drawn from different alternative assumptions about the missingness mechanism simultaneously. A recent use of these enhanced displays in a medical device clinical trial has helped lead to FDA approval.

Key Words: Missing data, MNAR, sensitivity analysis, multiple imputation.

1. Introduction

An assumption is *unassessable* if there is no statistical procedure that can be applied to available data that would support the assumption, without adding more restrictions. Missing data modeling is one area of statistics where unassessable assumptions are essentially inevitable. For most of the 20th century, the presence of missing values in the data was largely handled by editing or case deletion (see Schafer and Graham, 2002), and only with the formalization of a framework of inference from incomplete data developed in Rubin (1976), did research on methods to handle missing data begin to gain momentum (Molenberghs, 2007). Rubin (1976) introduced the notion of "missing data mechanism" and formulated three types of missingness mechanisms: *missing completely at random* (MCAR), *missing at random* (MAR) and *missing not at random* (MNAR).

It is difficult to assesses assumptions about the missing data mechanism empirically, without imposing additional structure on the data model or introducing other auxiliary information (Little and Rubin, 2002). Moreover, it is virtually impossible to test whether the mechanism is MNAR because such models state that the missingness depends on the *unobserved* values. Therefore, it is strongly recommended to perform sensitivity checks when drawing conclusions from data with missing values.

Here we introduce a set of displays that reveal the effects of all possible combinations of the values of missing data in the treatment and control groups, typically on p-values and point estimates. The displays are based on the idea of "tipping-point" (TP) analysis, first introduced in Yan et al. (2009) to assess the impact of missing data on the conclusions of a study. Yan et al. (2009) defined the *tipping points* of a study to be the particular combinations of missing data values that would change the study's conclusions and presented a simple way to display this information. We enhance this initial idea by adding more details onto the display, including smooth changes in the quantities of interest, the output from multiple MAR and MNAR models, and, when available, historical estimates.

*Harvard University, 1 Oxford Street, Cambridge, MA 02138

2. Notation and Definitions

For N independent units in the data set, let $X = (x_{ik}) = (X_1, X_2, \ldots, X_K)$ be the NxK matrix of K predictors (or covariates) and let $Y = (y_{ij}) = (Y_1, Y_2, \ldots, Y_J)$ be the NxJ matrix of outcomes. We define the matrix of missingness indicators for J outcomes, $D = (d_{ij})$, such that $d_{ij} = 1$ if unit i is missing the j'th outcome. Let the set $Y_{obs} = \{y_{ij}|d_{ij} = 0\}$ represent observed values among outcomes and set Y_{mis} represent the missing elements of the matrix, $Y = \{Y_{obs}, Y_{mis}\}$. Also, let $f(D|X, Y, \phi)$ be the conditional distribution of missingness indicators given all data values, observed and missing, and unknown vector-parameter ϕ. We assume that the conditional distribution of outcomes Y given predictors X has a probability model $f(Y|X, \theta)$, governed by a vector-parameter θ. The missingness mechanism is called *missing completely at random* (MCAR) if

$$f(D|X, Y, \phi) = f(D, \phi),$$

for all D, X and Y, and for each possible value of ϕ. The missing data are *missing at random* (MAR) if

$$f(D|X, Y, \phi) = f(D|X, Y_{obs}, \phi),$$

for the observed D, X and Y_{obs}, and for each possible value of ϕ.

Suppose we are interested in estimating the unknown vector-parameter θ in the distribution of complete data $f(Y|X, \theta)$. The missing data are said to be *ignorable* for the purpose of direct likelihood (or Bayesian) inference on θ if the MAR assumption is satisfied and parameters that govern the missingness mechanism, ϕ, do not carry any information about θ (i.e., ϕ and θ are *distinct*, Rubin, 1976; Little and Rubin, 2002). If either the distinctness of ϕ and θ or MAR is not met, missing data are considered *nonignorable*. Violation of the former is less consequential because the likelihood-based inference will still produce consistent, although inefficient, estimates. If the missingness mechanism does not satisfy MAR, it is regarded as *missing not at random* (MNAR) and requires specifying a full-data likelihood $f(Y, D|X, \theta, \phi)$, including the model for the missingness mechanism $f(D|X, Y, \phi)$, to produce a generally valid likelihood-based inference with consistent estimates. In practice, such a model involves making assumptions about the distribution of missing values that usually cannot be assessed empirically, and, therefore, the obtained results should be subjected to sensitivity analyses.

One general approach to handling missing data is multiple imputation (MI, Rubin, 1987), used to create multiple completed datasets by imputing missing values from their posterior predictive distribution $f(Y_{mis}|Y_{obs}, X, \theta, \phi)$. The advantage of MI is that it enables practitioners to use widely-available complete-data methods on each imputed dataset separately and incorporate the uncertainty due to the presence of missing data by pooling the results using Rubin's Combining Rules (Rubin, 1987; 2004). More important, this method is very suitable for performing sensitivity analyses because one can use multiple models to generate imputations and compare conclusions across the models.

3. Sensitivity Analysis Using Enhanced TP Displays

Several guidelines on handling missing data issued lately (National Research Council, 2010; CHMP, 2009; Burzykowski et al, 2010) agree on the importance of performing sensitivity analyses. However, they also note that there is a shortage of practical recommendations as to how one should go about conducting the sensitivity analyses; no clear consensus among statisticians has been achieved on this issue yet. TP analysis was first proposed in Yan et al. (2009) to aid clinical reviewers when judging the impact of missing

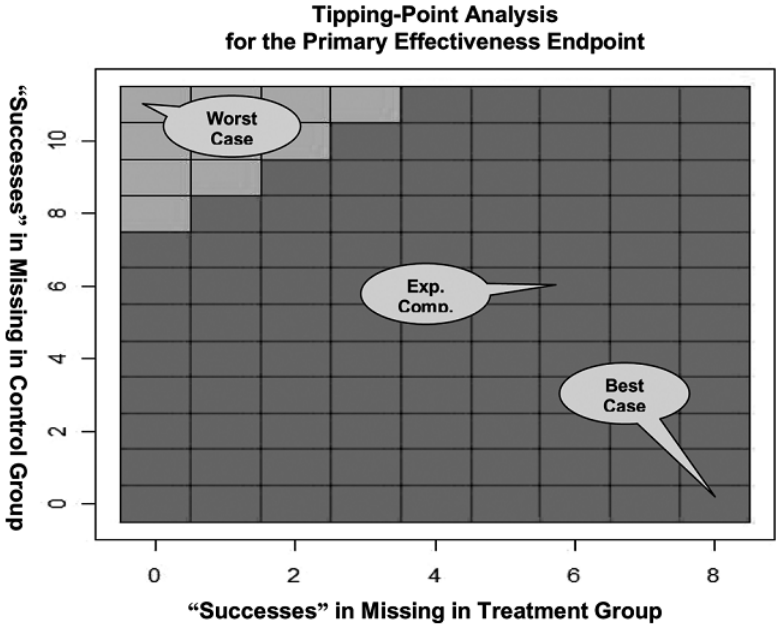

Figure 1: Basic tipping-point display presented in Campbell et al. (2011).

data on a binary outcome on the estimation of a treatment effect. Figure 1 from Campbell et al. (2011) illustrates the initial idea of TP analysis described in Yan et al. (2009). It results in the matrix of all possible combinations of a number of "successes" among non-respondents in the treated group (horizontal axes) and the control groups (vertical axes). Each combination is categorized based on whether the corresponding missing data would change the conclusion about the estimated treatment effect's statistical significance: the staircase region marks the tipping points of the study, i.e., the combinations of the number of successes among nonrespondents in the treated group and the control group that alter the conclusion. One fundamental issue with this basic depiction is that the display does not inform us about the likelihood of each individual combination. Therefore, unless we discover that none of possible missing data patterns change the study conclusion, we cannot utilize these displays to their fullest potential.

We generalize the initial idea of TP analysis and propose a visual aid that allows for a systematic execution of sensitivity analyses. The impact of missing values may be illustrated by a TP display with horizontal and vertical axes representing a summary $g(\cdot)$ of values of missing outcomes for treated and control groups, $g(Y^T_{mis})$ and $g(Y^C_{mis})$. The analyst may choose any summary of interest as long as it is easily interpretable for the intended audience. For example, for a continuous outcome, the axes can represent average outcomes among nonrespondents in treated and control groups. Here is the list of the proposed enhancements to the basic TP displays:

- Ticks that represent historical estimates of the number of successes in each group, when available.
- The results from the current basic modeling procedure, i.e., the posterior distribution of missing data under the chosen model $f(Y, D|X, T)$, possibly, represented by MIs.
- A colored heat-map, i.e., a matrix of colors, that illustrates a *gradual* change of a certain quantity of interest, e.g., an estimated average treatment effect, a p-value from a hypothesis test or confidence interval bound, that corresponds to each pair of summaries $\{g(Y^T_{mis}), g(Y^C_{mis})\}$.

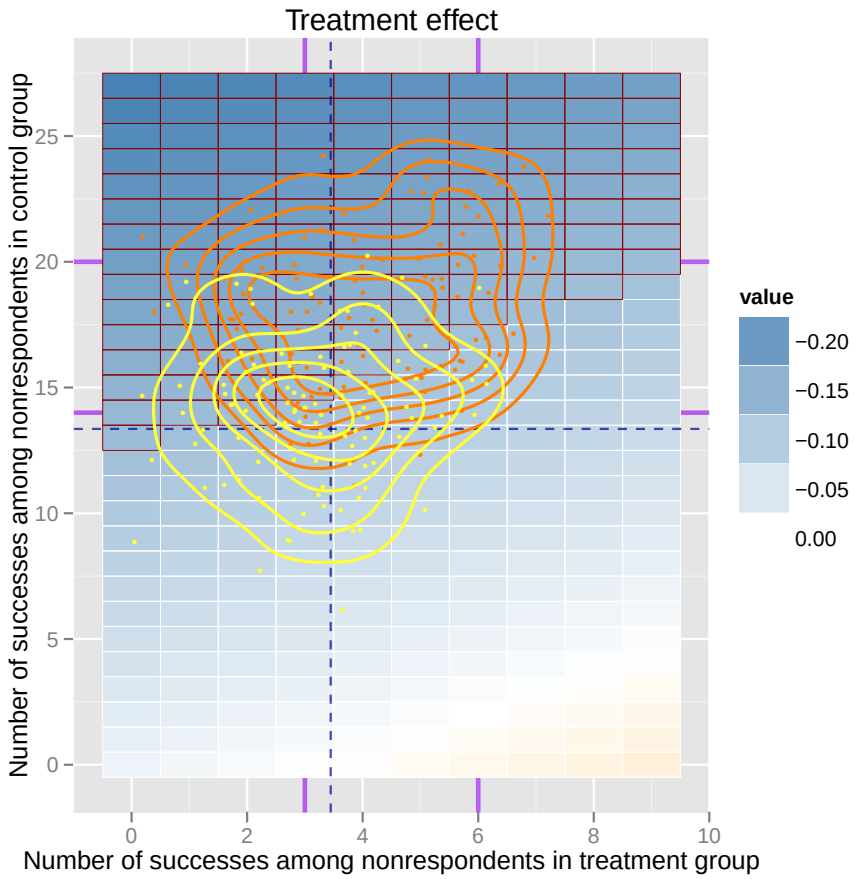

Figure 2: Example of the enhanced TP display for a binary outcome.

- The posterior distributions of missing data obtained under a collection of models with alternative assumptions, that can also be represented by MIs. Section 4 below describes a transparent procedure to explore alternative models using the TP displays.

Figure 2 illustrates an example of the enhanced TP display produced for a simulated dataset with two fully-observed predictors, dichotomous treatment and a binary outcome that is partially missing (not-at-random) in each treatment group. Out of $N = 270$ simulated subjects, 150 were assigned to the treatment group and 120 to the control group, with 9 and 27 subjects missing the outcome in each group, respectively. The axes represent the number of successes that could be observed among nonrespondents in the treated and control groups. Each combination corresponds to a value of the estimated average treatment effect. Its magnitude and sign are represented using a color palette that changes from dark blue (large negative value) to dark orange (large positive values), with white representing zero estimated effect. The red grid highlights combinations that correspond to a significant treatment effect based on a hypothesis test for the difference between two proportions, using the traditional 0.05 significance level. Vertical and horizontal dashed lines (in blue) correspond to the observed success rate among treated and control subjects. Also, the display shows two vertical and horizontal ticks (in purple), an example of historical data that may be available for the study.

The simulated data were analyzed under two assumptions about the missingness mechanism, MCAR and MAR. The posterior distribution of missing values was approximated using 100 MIs, represented by yellow (MCAR) and brown (MAR)points on Figure 2. The

contours correspond to the estimated density of the joint distribution of successes among nonrespondents in teated and control groups under each model, using bivariate normal kernels. The results obtained from the two models are noticeably different, indicating that MCAR assumption is not appropriate; otherwise both models would have produced similar imputations.

4. Systematic Exploration of MNAR Models With TP Displays

Analysis of data with missing values is often performed under the MAR assumption. It greatly simplifies the analysis itself by allowing to avoid modeling the missingness mechanism. More important, there is a shortage of standardized and agreed-upon ways to explore assumptions, alternative to MAR, systematically. In order to systematize the sensitivity analysis, we propose to utilize the pattern-mixture decomposition of the joint distribution of responses and the missingness mechanism,

$$\begin{aligned} f(Y, D|X, \theta, \phi) &= f(Y|D, X, \theta) f(D|X, \phi) \\ &= f(Y_{obs}|X, \theta) P(D = 0|X, \phi) + f(Y_{mis}|X, \theta) P(D = 1|X, \phi). \end{aligned} \quad (1)$$

Under the MAR assumptions, $f(Y_{obs}|X, \theta) = f(Y_{mis}|X, \theta)$ but, in general, the outcome models for respondents and nonrespondents may differ; and the joint model for Y and D is a mixture of these two models. Two out of three components of (1), $f(D|X, \phi)$ and $f(Y_{obs}|D, X, \theta)$, may be approximated from the observed data, and the only part that requires unverifiable model assumption is the distribution of the outcomes for nonrespondents, i.e., $f(Y_{mis}|X, \theta)$. This allows for a natural formulation of alternative models by specifying the outcome distribution of nonrespondents based on the one estimated for respondents; the idea that first appeared in Rubin (1977). Alternative models for nonrespondents may be formed by introducing various deviations to the model for respondents and defining corresponding *sensitivity parameters*. For example, for a continuous outcome we can introduce the following parameters:

- response shift, $E(Y|X, D = 1) = E(Y|X, D = 0) + \delta$.
- effect change, $E(Y|X, D = 1) = E(Y|X, D = 0) + \Delta\beta_{X_j} X_j$.
- variance scaling, $Var(Y|X, D = 1) = \upsilon Var(Y|X, D = 0)$.

Then, the general family of alternative response models for nonrespondents may be represented by a set of sensitivity parameters $(\delta, \Delta\beta_{X_1}, \ldots, \Delta\beta_{X_K}, \upsilon)$. Similar ideas may be applied to binary outcomes with corresponding modifications in the definition of sensitivity parameters.

5. Discussion

We propose an original and systematic way to perform sensitivity analyses using enhanced TP displays in studies with partially missing outcomes and dichotomous treatments. The display allows us to assess the strength of the conclusion of the study under the adopted assumptions and to inform us about the alternative models that may alter these conclusions. An intuitive way to explore MNAR models is to use a fitted model under the MAR assumption as a base-line and introduce various sensitivity parameters to construct the outcome model for nonrespondents. In fact, TP displays themselves can suggest possible directions for alternative models that will result in changes of the study conclusions. This idea provides a new collection of useful tools for the analysis of data sets plagued with missing values.

Recently, the enhanced TP displays were used to conduct a sensitivity analysis in a medical device clinical trial, that was subsequently reviewed and approved by the FDA. The software that creates TP displays is under development. A basic package that allows drawing TP displays using existing imputations will appear in R in the near future. In addition, Statistical Solutions is considering implementing a sensitivity analysis capability, based on TP displays and the procedure described in Section 4, into their SOLAS product.

REFERENCES

Burzykowski, T., Carpenter, J., Coens, C., Evans, D., France, L., et al. (2010), "Missing data: discussion points from the PSI missing data expert group." *Pharmaceutical Statistics*, 9(4), 288-297.

Campbell, G., Pennello, G., Yue, L. (2011)"Missing data in the regulation of medical devices." *Journal of Biopharmaceutical Statistics*, 21(2), 180-195.

European Medicines Evaluation Agency. (2009) "Guideline on Missing Data in Confirmatory Clinical Trials." Committee for Medical Products for Human Use. (Available at: `http://www.ema.europa.eu/pdfs/human/ewp/177699endraft.pdf`; accessed March 11, 2012).

Little, R. J. A. (1994) "A class of pattern-mixture models for normal incomplete data". *Biometrika*, 81(3), 471-483.

Little, R. J. A., Rubin, D. B. (2002) *Statistical analysis with missing data* (2nd ed.). New York: Wiley.

Liublinska, V., Rubin, D.B. (2012). "Re: Dealing With Missing Outcome Data In Randomized Trials And Observational Studies." *American Journal of Epidemiology*, 176(4), 357–358.

Molenberghs, G. (2007) "Editorial: What to do with missing data?" *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 170(4), 861-863.

National Research Council. (2010) "The Prevention and Treatment of Missing Data in Clinical Trials." Washington, DC: The National Academies Press.

Rubin, D. B. (1976) "Inference and missing data." *Biometrika*, 63(3), 581-592.

Rubin, D. B. (1977) "Formalizing subjective notions about the effect of nonrespondents in sample surveys." *Journal of the American Statistical Association*, 72, 538-543.

Rubin, D. B. (1987) "Multiple Imputation for Nonresponse in Surveys." (1st ed.) New York: Wiley.

Rubin, D. B. (2004) "Multiple Imputation for Nonresponse in Surveys." (2nd ed.) New York: Wiley-Interscience

Schafer, J. L., Graham, J. W. (2002) "Missing data: our view of the state of the art." *Psychological Methods*, 7(2), 147-177.

Yan, X., Lee, S., and Li, N. (2009), "Missing data handling methods in medical device clinical trials," *Journal of Biopharmaceutical Statistics*, 19(6), 1085-1098.