

# Estimation of intention-to-treat effect: Model misspecification sensitivity analysis

Booil Jo\*  
Stanford University

Prevention Science Methodology Group  
October 27, 2005

\* The study was supported by MH066319, MH066247, and MH40859. I thank participants of the Prevention Science Methodology Group for insightful feedback. I also appreciate helpful comments from Constantine Frangakis, Keisuke Hirano, Nick Ialongo, Amiram Vinokur, Liz Ginexi, and George Howe.

# Underlying Subpopulations and Complications in Randomized Trials

- Underlying subpopulations may benefit differently from the treatment.
- Underlying subpopulations can be characterized by various **posttreatment variables (mediators)** such as compliance types.
- If these posttreatment variables interact with common complications in randomized trials, such as **missing outcomes** and **clustering** of data, intention-to-treat (ITT) effect estimates may be biased.
- In other words, even if we are interested in the overall effect of the treatment (i.e., ITT effect), considering the relationship between posttreatment variables and complications in randomized trials is helpful for better estimation of ITT effect.

## Co-presence of noncompliance and nonresponse

- Randomized trials involving human participants often suffer from both intervention noncompliance (no-show) and nonresponse (missing outcome information) at followup assessments.
- Individuals' compliance and outcome information is available only under the condition they were assigned to, and *unavailable* under other conditions - these values can be considered as latent, missing, or potential values.
- However, based on random assignment, the causal inference is possible at the average level - Intention to treat analysis as gold standard.
- ITT analysis is a reasonable method for estimating overall effectiveness of the intervention programs in the presence of noncompliance (since we are not particularly interested in differential intervention effects for individuals with different compliance types).
- However, robustness of ITT analysis can be threatened in the co-presence of noncompliance and nonresponse (missing outcomes).

## Modeling Choices for ITT Analysis

- All three estimators assume that the probability of response is unassociated with the outcome given compliance – latent ignorability (*LI*: Frangakis & Rubin, 1999). It is also assumed that the outcome is unassociated with intervention assignment for noncompliers – outcome exclusion restriction (*OER*).
- **MCAR** assumes that the probability of response (having outcome data information) is unassociated with compliance status – missing completely at random (*MCAR*: Little & Rubin, 2002).
- **MAR** assumes that the probability of response is unassociated with the outcome itself given observed compliance – missing at random (*MAR*: Little & Rubin, 2002).
- **RER** assumes the exclusion restriction not only on outcomes (*OER*), but also on outcome missing indicators (*RER*: response exclusion restriction). The combined assumption is called the compound exclusion restriction (*CER*: Frangakis & Rubin, 1999).

# Estimator Comparison

- All three estimators can be affected by violations of *LI* and *OER*.
- In addition, MCAR estimator is affected by violations of *MAR* and *MCAR*.
- In addition, MAR estimator is affected by violation of *MAR*.
- In addition, RER estimator is affected by violation of *RER*.
- The difference between ITT estimates from MCAR and MAR estimators indicates additional bias when *MCAR* is assumed instead of *MAR*.
- The difference between ITT estimates from MAR and RER estimators indicates the difference between biases due to violation of *MAR* and *RER* (therefore, we cannot identify violations from either *MAR* or *RER*).
- Which assumption is more plausible? Which assumption leads to less biased estimates?

# Common Setting

- Randomized trials, where successful placebo control is unlikely.
- 2 conditions: intervention ( $Z = 1$ ) and control ( $Z = 0$ )
- 2 compliance types ( $C_i$ )
  - 1) complier (c) - receives the intervention treatment if assigned, and does not if not assigned.  $\pi_c =$  compliance rate.
  - 2) noncomplier (n) - does not receive the intervention treatment even if assigned to receive it.  $1 - \pi_c = \pi_n =$  noncompliance rate.
- 2 observed average outcomes in  $Z = 1$ :  $\mu_{c1}$  and  $\mu_{n1}$ .
- 2 observed average responses in  $Z = 1$ :  $\pi_{c1}^R$  and  $\pi_{n1}^R$ .
- 2 unobserved average outcomes in  $Z = 0$ :  $\mu_{c0}$  and  $\mu_{n0}$ .
- 2 unobserved average responses in  $Z = 0$ :  $\pi_{c0}^R$  and  $\pi_{n0}^R$ .
- 1 observed outcome and 1 observed response rate in  $Z = 0$ :  $\mu_0$  and  $\pi_0^R$ .

# Common Setting

- Assuming  $LI$ , the true ITT effect is defined as

$$ITT = \pi_c (\mu_{c1} - \mu_{c0}) + (1 - \pi_c) (\mu_{n1} - \mu_{n0}), \quad (1)$$

where  $\mu_{c0}$  and  $\mu_{n0}$  do not have corresponding sample statistics.

- Observed average outcome in the control condition can be written as

$$\mu_0^{obs} = \frac{\pi_{c0}^R}{\pi_0^R} \pi_c \mu_{c0} + \frac{\pi_{n0}^R}{\pi_0^R} (1 - \pi_c) \mu_{n0}. \quad (2)$$

- The average response rate in the control condition can be written as

$$\pi_0^R = \pi_{c0}^R \pi_c + \pi_{n0}^R (1 - \pi_c). \quad (3)$$

- The  $\mu_{c0}$  can be derived from (2) as

$$\mu_{c0} = \frac{\mu_0^{obs} \pi_0^R - \mu_{n0} \pi_{n0}^R (1 - \pi_c)}{\pi_0^R - \pi_{n0}^R (1 - \pi_c)}. \quad (4)$$

# Identification of ITT Effect

- All estimators assume that *LI* and *OER* ( $\mu_{n1} - \mu_{n0} = 0$ ) hold.
- In respondent-based MCAR estimator, under *MCAR* ( $\pi_{c0}^R = \pi_{n0}^R$ ,  $\pi_{c1}^R = \pi_{n1}^R$ ),  $\mu_{c0}$  is identified as

$$\hat{\mu}_{c0}^{MCAR} = \frac{\hat{\mu}_0 - \hat{\mu}_{n1} \left(1 - \frac{\hat{\pi}_{c1}^R}{\hat{\pi}_1^R} \hat{\pi}_c\right)}{\frac{\hat{\pi}_{c1}^R}{\hat{\pi}_1^R} \hat{\pi}_c}. \quad (5)$$

- In MAR estimator, under *MAR* ( $\pi_{c0}^R = \pi_{n0}^R$ ),  $\mu_{c0}$  is identified as

$$\hat{\mu}_{c0}^{MAR} = \frac{\hat{\mu}_0 - \hat{\mu}_{n1} (1 - \hat{\pi}_c)}{\hat{\pi}_c}. \quad (6)$$

- In RER estimator, under *RER* ( $\pi_{n1}^R = \pi_{n0}^R$ ),  $\mu_{c0}$  is identified as

$$\hat{\mu}_{c0}^{RER} = \frac{\hat{\mu}_0 \hat{\pi}_0^R - \hat{\mu}_{n1} \hat{\pi}_{n1}^R (1 - \hat{\pi}_c)}{\hat{\pi}_0^R - \hat{\pi}_{n1}^R (1 - \hat{\pi}_c)}. \quad (7)$$

- In each estimator, ITT effect is identified as  $\hat{\pi}_c (\hat{\mu}_{c1} - \hat{\mu}_{c0})$ .

## Motivating Example: JHU PIRC Family-School Partnership (FSP) Intervention Study

- The Johns Hopkins Public School Preventive Intervention Study was conducted by the Johns Hopkins University Preventive Intervention Research Center (JHU PIRC) in 1993-1994 (Ialongo et al., 1999).
- The study was designed to improve academic achievement and to reduce early behavioral problems of school children.
- Teachers and first-grade children were randomly assigned to intervention conditions. The control condition and the Family-School Partnership (FSP) intervention condition are compared in this example (221 students in the intervention, and 219 students in the control condition).
- In the intervention condition, parents were asked to implement 66 take-home activities related to literacy and mathematics over a six-month period.

## JHU PIRC FSP Intervention Study

- In this example, individuals are categorized into low and high compliers based on the level of completeness in 66 home learning activities. Parents who completed at least 45 activities are categorized as high compliers (upper 50% of parents) and the rest (lower 50% of parents) are categorized as low compliers. Let  $C_i = c$ , if individual  $i$  is a high complier, and  $C_i = n$ , if individual  $i$  is a low complier.
- Shy behavior is the outcome of focus in this example. Change scores were calculated by subtracting the baseline score from the scores measured 1 year and 2 years after the intervention. Negative values of ITT effect estimates can be interpreted as desirable effects of the intervention, meaning that shy behavior increased less among individuals in the intervention condition.
- To illustrate different patterns of missing data and resulting biases, ITT analysis was separately conducted with each change score as a univariate outcome.

## JHU PIRC Study: Sample statistics

- Outcome: change score of shy behavior.

---

<i>Followup</i>	$\hat{\mu}_0^{obs}$	$\hat{\mu}_{c1}$	$\hat{\mu}_{n1}$	$\hat{\pi}_0^R$	$\hat{\pi}_{c1}^R$	$\hat{\pi}_{n1}^R$	$\hat{\pi}_c$
2nd Grade	0.319	0.177	-0.248	0.781	0.911	0.833	0.457
3rd Grade	0.066	0.047	-0.197	0.744	0.792	0.708	0.457

---

## ITT effect estimates

- Method of moments estimator (instrumental variable approach).

---

<i>Followup</i>	$ITT^{MCAR}$	$ITT^{MAR}$	$ITT^{RER}$
2nd Grade	-0.363 (0.140)	-0.373 (0.140)	-0.422 (0.160)
3rd Grade	-0.145 (0.152)	-0.152 (0.152)	-0.137 (0.145)

---

- In general, the effect of FSP intervention is much stronger at the second grade followup than at the third grade followup.
- At second grade followup,  $ITT^{MCAR}$  presents the smallest and  $ITT^{RER}$  presents the largest effect of the intervention.
- At third grade,  $ITT^{RER}$  presents the smallest and  $ITT^{MAR}$  presents the largest effect of the intervention.

## Deviations From Key Model Assumptions

- Whether *MCAR* is violated ( $\alpha = \pi_{c1}^R - \pi_{n1}^R$ ) can be easily learned from sample statistics. That is,  $\alpha = \hat{\pi}_{c1}^R - \hat{\pi}_{n1}^R = 0.078$  at the second grade and 0.084 at the third grade followup. High compliers responded at a higher rate at both followups.
- Whether *MAR* is violated ( $\delta = \pi_{c0}^R - \pi_{n0}^R$ ) cannot be learned from observed sample statistics. Therefore the resulting bias is unknown.
- Whether *RER* is violated ( $\beta = \pi_{n1}^R - \pi_{n0}^R$ ) cannot be learned from observed sample statistics. Therefore the resulting bias is unknown.

# Plausibility of Assumptions in the JHU Trial

- Deviation from *MCAR* is directly estimable.
- Possible deviation from *MAR*: poor compliance of parents is a good indicator of family instability, meaning that these families are more likely to move from place to place due to financial stress or other reasons related to drug or alcohol problems, therefore it is harder to locate these parents and their children at followup assessments (i.e.,  $\delta > 0$ ).
- Possible deviation from *RER*: poorly complying families in the intervention might have felt somewhat benefited from the intervention and felt more obliged to respond than families in the control who would have complied poorly if the intervention had been offered (i.e.,  $\beta > 0$ ).

Or, poorly complying families in the intervention condition might have been demoralized from failing to comply with the intervention activities and respond less at followups than their counterparts in the control condition (i.e.,  $\beta < 0$ ).

## Comparing plausibility of *MAR* and *RER*

- Although one assumption may intuitively seem more plausible than the other, degrees of deviation from the two assumptions cannot be compared unless they can be viewed from the same assumption (E.g., a small deviation from one assumption might be equivalent to a much bigger deviation from the other assumption).
- Translation between different assumptions can be done by simple calculations, which may reveal a quite surprising relationship between the two assumptions.
- The two assumptions are connected through the same parameter  $\pi_{n0}^R$ . Therefore, imposing any restrictions on plausibility of one assumption immediately affects the other assumption.

## Connectivity Between *MAR* and *RER*

- If  $\beta$  is fixed at a certain value,  $\pi_{n0}^R$  can be solved for (i.e.,  $\hat{\pi}_{n0}^R = \hat{\pi}_{n1}^R - \beta$ ). Then,  $\pi_{c0}^R$  can be identified from the mixture  $\pi_0^R = \pi_c \pi_{c0}^R + (1 - \pi_c) \pi_{n0}^R$  as

$$\hat{\pi}_{c0}^R = \frac{\hat{\pi}_0^R - (\hat{\pi}_{n1}^R - \beta)(1 - \hat{\pi}_c)}{\hat{\pi}_c}, \quad (8)$$

where  $\pi_{n1}^R$  is directly estimable from the observed data.

- Given that  $\pi_{c0}^R$  and  $\pi_{n0}^R$  are identified, deviation from *MAR* also can be identified as

$$\hat{\delta} = \hat{\pi}_{c0}^R - \hat{\pi}_{n0}^R = \frac{\hat{\pi}_0^R - \hat{\pi}_{c1}^R + \beta}{\hat{\pi}_c}, \quad (9)$$

which is the degree of deviation from *MAR* that can be compared to the degree of deviation from *RER* ( $\beta$ ).

- Similarly, if  $\delta$  is fixed at a certain value,  $\pi_{n0}^R$ ,  $\pi_{c0}^R$  and  $\beta$  can be identified.

## JHU PIRC Grade 2

### Some possible combinations of $\delta$ and $\beta$

---

$\pi_{c0}^R$	$\pi_{n0}^R$	$\hat{\delta}$	$\hat{\beta}$
0.520	<b>1.000</b>	-0.480	-0.167
0.718	0.833	-0.115	0.000
0.781	0.781	<b>0.000</b>	<b>0.053</b>
<b>1.000</b>	0.596	<b>0.404</b>	<b>0.237</b>

---

- We can look at plausibility of  $MAR$  in terms of  $RER$ , and plausibility of  $RER$  in terms of  $MAR$  (e.g., successful double blinded trials).
- Natural bounds of response rates and plausibility bounds of  $MAR$  (i.e.,  $\delta \geq 0$ ) can be applied. It is also possible to establish reasonable, but still conservative plausibility bounds for  $RER$ .
- Based on connectivity between  $MAR$  and  $RER$ , plausibility bounds of  $MAR$  also applies to  $RER$  (i.e.,  $\beta \geq 0.053$ ).

## JHU PIRC Grade 3

### Some possible combinations of $\delta$ and $\beta$

---

$\pi_{c0}^R$	$\pi_{n0}^R$	$\hat{\delta}$	$\hat{\beta}$
0.440	<b>1.000</b>	-0.560	-0.292
0.744	0.744	<b>0.000</b>	<b>-0.036</b>
0.787	0.708	<b>0.079</b>	<b>0.000</b>
0.906	0.608	<b>0.298</b>	<b>0.100</b>
<b>1.000</b>	0.529	<b>0.470</b>	<b>0.179</b>

---

- Our ultimate interest is usually in comparing sensitivity of causal effect estimates rather than plausibility of assumptions.
- Comparable  $\delta$  and  $\beta$  values do not necessarily result in the same bias, and the assumption with higher plausibility (less deviation) may not lead to smaller bias because each assumption follows its own bias mechanism.

## Bias Mechanisms

- If *MCAR* is violated,

$$MCAR_{bias} = \frac{\alpha (1 - \pi_c) \pi_c (\mu_{c1} - \mu_{n1})}{\pi_1^R}, \quad (10)$$

where  $\alpha$  indicates deviation from *MCAR* (i.e.,  $\pi_{c1}^R - \pi_{n1}^R$ ).

- If *MAR* is violated,

$$MAR_{bias} = \frac{\delta (1 - \pi_c) (\mu_0^{obs} - \mu_{n1})}{\pi_{c0}^R}, \quad (11)$$

where  $\delta$  indicates deviation from *MAR*<sup>E</sup> (i.e.,  $\pi_{c0}^R - \pi_{n0}^R$ ).

- If *RER* is violated,

$$RER_{bias} = \frac{\beta \pi_0^R (1 - \pi_c) (\mu_0^{obs} - \mu_{n1})}{\pi_{c0}^R [\pi_0^R - \pi_{n1}^R (1 - \pi_c)]}, \quad (12)$$

where  $\beta$  indicates deviation from *RER* (i.e.,  $\pi_{n1}^R - \pi_{n0}^R$ ).

## Some Possible Combinations of Deviations From Missing Data Assumptions and Resulting Biases at Grade 2

$\pi_{c0}^R$	$\pi_{n0}^R$	$\hat{\delta}$	$\widehat{MAR}_{bias}$	$\hat{\beta}$	$\widehat{RER}_{bias}$
0.520	<b>1.000</b>	-0.480	0.339	-0.167	0.280
0.718	0.833	-0.115	0.059	0.000	0.000
0.781	0.781	0.000	<b>0.000</b>	0.053	<b>-0.059</b>
<b>1.000</b>	0.596	0.404	<b>-0.148</b>	0.237	<b>-0.207</b>

- When  $\beta = 0$ , the  $\delta$  estimate is  $-0.115$ , meaning that the response rate of noncompliers is 11.5% higher than that of compliers in the control condition, which is very unlikely given the observation in the intervention condition and given the circumstances of the JHU trial. It is very likely that  $\delta \geq 0$ .
- With this quite general restriction in the possible range of deviations from  $\delta$  and  $\beta$ , we can conclude that both models assuming  $MAR$  and  $RER$  overestimate ITT effect and the model assuming  $RER$  overestimates more.

## Some Possible Combinations of Deviations From Missing Data Assumptions and Resulting Biases at Grade 3

$\pi_{c0}^R$	$\pi_{n0}^R$	$\hat{\delta}$	$\widehat{MAR}_{bias}$	$\hat{\beta}$	$\widehat{RER}_{bias}$
0.440	<b>1.000</b>	-0.560	0.182	-0.292	0.196
0.744	0.744	0.000	<b>0.000</b>	-0.036	<b>0.014</b>
0.787	0.708	0.079	<b>-0.014</b>	0.000	<b>0.000</b>
0.906	0.608	0.298	<b>-0.047</b>	0.100	<b>-0.033</b>
<b>1.000</b>	0.529	0.470	<b>-0.067</b>	0.179	<b>-0.053</b>

- The *RER* assumptions seems more realistic at grade 3 than at grade 2, since it is very unlikely that the effect of poorly completed FSP intervention will last for a long period of time. Also, at grade 3, if  $\beta = 0$ , the  $\delta$  estimate is 0.079, meaning that the response rate of high compliers is 7.9% higher than that of low compliers in the control condition. This is a very realistic situation, further supporting plausibility of *RER*.

## Some Possible Combinations of Deviations From Missing Data Assumptions and Resulting Biases at Grade 3

$\pi_{c0}^R$	$\pi_{n0}^R$	$\hat{\delta}$	$\widehat{MAR}_{bias}$	$\hat{\beta}$	$\widehat{RER}_{bias}$
0.440	<b>1.000</b>	-0.560	0.182	-0.292	0.196
0.744	0.744	0.000	<b>0.000</b>	-0.036	<b>0.014</b>
0.787	0.708	0.079	<b>-0.014</b>	0.000	<b>0.000</b>
0.906	0.608	0.298	<b>-0.047</b>	0.100	<b>-0.033</b>
<b>1.000</b>	0.529	0.470	<b>-0.067</b>	0.179	<b>-0.053</b>

- If we maintain only the general restriction that  $\delta \geq 0$ , we can conclude that, if  $\beta > 0$ , both estimators overestimate ITT effect and the  $MAR$  estimator overestimates more. If  $\beta < 0$ , the  $MAR$  estimator slightly overestimates and the  $RER$  estimator underestimates ITT effect. Therefore, taking a more conservative side, it seems reasonable to prefer  $ITT^{RER}$  estimator to  $ITT^{MAR}$  estimator in assessing ITT effect at grade 3 followup.

## Bias adjusted ITT estimates at Grade 2

$\widehat{ITT}_{adj}$	$\hat{\delta}$	$\widehat{MAR}_{bias}$	$\hat{\beta}$	$\widehat{RER}_{bias}$
-0.712	-0.480	0.339	-0.167	0.280
-0.432	-0.115	0.059	0.000	0.000
-0.373	0.000	<b>0.000</b>	0.053	<b>-0.059</b>
-0.225	0.404	<b>-0.148</b>	0.237	<b>-0.207</b>

- $\widehat{ITT}_{adj} = \widehat{ITT}^{MAR} - \widehat{MAR}_{bias} = \widehat{ITT}^{RER} - \widehat{RER}_{bias}$ .
- Within the natural bounds of  $\delta$  and  $\beta$ , the bounds of ITT effect are  $[-0.712, -0.225]$ .
- With the general restriction that  $\delta \geq 0$ , the bounds of ITT effect are  $[-0.373, -0.225]$ .
- Both bounds indicate positive overall effect of FSP intervention.

## Bias adjusted ITT estimates at Grade 3

$\widehat{ITT}_{adj}$	$\hat{\delta}$	$\widehat{MAR}_{bias}$	$\hat{\beta}$	$\widehat{RER}_{bias}$
-0.334	-0.560	0.182	-0.292	0.196
-0.152	0.000	<b>0.000</b>	-0.036	<b>0.014</b>
-0.138	0.079	<b>-0.014</b>	0.000	<b>0.000</b>
-0.105	0.298	<b>-0.047</b>	0.100	<b>-0.033</b>
-0.085	0.470	<b>-0.067</b>	0.179	<b>-0.053</b>

- Within the natural bounds of  $\delta$  and  $\beta$ , the bounds of ITT effect are  $[-0.334, -0.085]$ .
- With the general restriction that  $\delta \geq 0$ , the bounds of ITT effect are  $[-0.152, -0.085]$ .
- Both bounds indicate positive overall effect of FSP intervention.

# Conclusions

- JHU PIRC example calls attention to the importance of focusing on each case in investigating relative sensitivity of causal effect estimates with different identifying assumptions, instead of pursuing a general conclusion that applies to every occasion (i.e., one assumption always works better).
- It is important to translate competing assumptions and to understand actual bias mechanisms in making choices among different models.
- Bounding plausibility of each assumption based on plausibility bounds of several competing assumptions seems to be helpful for model comparison and parameter bounding.
- One may choose an assumption or a level of plausibility bounds that better fits the purpose of an analysis (more or less conservative).
- The proposed method is quite easy to implement. And can be used with more than one identifying assumptions (e.g., OER and RER).

# Limitations

- The common biases (i.e., biases due to deviation from  $LI$  and simultaneous deviations from  $MAR$  and  $OER$ ) in MAR and RER estimators may reverse the size and the direction of the total bias through accumulation or cancellation of biases being combined.
- Very detailed information is needed to predict bias due to deviation from  $LI$ , which is not a realistic option.
- Plausibility of exclusion restriction on outcomes ( $OER$ ) can be examined to some extent, but this is still harder to gauge than that on response behavior because  $OER$  violation tends to have wider natural bounds than  $\beta$ , especially with nonbinary outcomes, making it harder to narrow down the possible range of the resulting bias.