Sensitivity Analysis for Missing Data: The “how to” for the “what if”

Neil J. Perkins, Ph.D.
Rose Radin, Ph.D.

Epidemiology Branch, Diphpr
Eunice Kennedy Shriver National Institute of Child Health and Human Development
Missing Data

- Exposure
  - (e.g. missed visits, blood draw issue)

- Confounders
  - (e.g. incomplete medical records)

- Outcome
  - (e.g. withdrawal, pregnancy loss before measurement obtained)
The Prevention and Treatment of Missing Data in Clinical Trials

Key findings (paraphrased):

“...missing data are a serious problem that undermines the scientific credibility of causal conclusions...”

“...in studies with missing data, analysis methods that are based on plausible scientific assumptions should be used... they all require unverifiable assumptions.”

“Thus, sensitivity analyses should be conducted to assess the robustness of findings to plausible alternative assumptions about the missing data.”

“...missing data are a serious problem...”

Power
• Missing data is lost efficiency, no matter what.
“…missing data are a serious problem…”

Power
• Missing data is lost efficiency, no matter what.

Potential Bias
• Missing data mechanisms (Influenced by)
  • MCAR – missing completely at random (random)
  • MAR – missing at random (observed data)
  • MNAR – missing not at random (unobserved data)
Mitigating the effects of Missing Data

1. Study Design!
   - An ounce of prevention is worth a pound of cure
Mitigating the effects of Missing Data

1. **Study Design!**
   - An ounce of prevention is worth a pound of cure

2. **Analysis Techniques**
   - Complete Case Analysis
   - Single Imputation
   - Estimating Equations
   - Multiple Imputation
     - Naïve, easy but sometimes useful
     - More work but more rigorous
Mitigating the effects of Missing Data

1. **Study Design!**
   - An ounce of prevention is worth a pound of cure

2. **Analysis Techniques**
   - Complete Case Analysis
   - Single Imputation
   - Estimating Equations
   - Multiple Imputation

   “…analysis methods that are based on plausible scientific assumptions…”
Mitigating the effects of Missing Data

1. **Study Design!**
   - An ounce of prevention is worth a pound of cure

2. **Analysis Techniques**
   - Complete Case Analysis
   - Single Imputation
   - Estimating Equations
   - Multiple Imputation

   “…analysis methods that are based on plausible scientific assumptions…”
   “…all require unverifiable assumptions.”
Mitigating the effects of Missing Data

1. **Study Design!**
   - An ounce of prevention is worth a pound of cure

2. **Analysis Techniques**
   - Complete Case Analysis
   - Single Imputation
   - Estimating Equations
   - Multiple Imputation
   “…analysis methods that are based on plausible scientific assumptions…”
   “…all require unverifiable assumptions.”

3. **Sensitivity Analysis**
   - “What if ?”
   “…sensitivity analyses should be conducted to assess the robustness of findings to plausible alternative assumptions about the missing data.”
Sensitivity Analysis for Missing Data

What if I had *observed* the *unobserved*?

1. Would my conclusions have changed?
2. What scenarios would have led to a change or not change?
3. What is the plausibility of these scenarios?
Sensitivity Analysis for Missing Outcome

- Is low dose aspirin an effective therapy for women with trying to conceive?

- Preconception Treatment: 81mg Aspirin versus Placebo

- Block randomized by Site and Eligibility Strata
  - Original and Expanded

- Follow up: 6 cycles or through pregnancy

- Primary Endpoint: Live Birth
## Sensitivity analysis: overall

<table>
<thead>
<tr>
<th></th>
<th>LDA</th>
<th>Placebo</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Randomized</td>
<td>615</td>
<td>613</td>
<td>1228</td>
</tr>
</tbody>
</table>
**Sensitivity analysis: overall**

Information available for 1088 of 1228

<table>
<thead>
<tr>
<th></th>
<th>LDA</th>
<th>Placebo</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Randomized</strong></td>
<td>615</td>
<td>613</td>
<td>1228</td>
</tr>
<tr>
<td><strong>Live Birth</strong></td>
<td>309</td>
<td>288</td>
<td>597</td>
</tr>
<tr>
<td><strong>No Live Birth</strong></td>
<td>228</td>
<td>263</td>
<td>491</td>
</tr>
</tbody>
</table>
Sensitivity analysis: overall

Missing outcome on 140 of 1228 (12%)

<table>
<thead>
<tr>
<th></th>
<th>LDA</th>
<th>Placebo</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>615</td>
<td>613</td>
<td>1228</td>
</tr>
<tr>
<td>Live Birth</td>
<td>309</td>
<td>288</td>
<td>597</td>
</tr>
<tr>
<td>No Live Birth</td>
<td>228</td>
<td>263</td>
<td>491</td>
</tr>
<tr>
<td><strong>Unknown</strong></td>
<td><strong>78</strong></td>
<td><strong>62</strong></td>
<td><strong>140</strong></td>
</tr>
</tbody>
</table>
Live Births: Overall

RD 5.1%
\(p = 0.093\)

Overall

% Live Birth

LDA

Placebo

Schisterman et. al. The Lancet 2014
Live Births: Overall

Would the conclusions have changed if we had observed all of the data?

**Overall**

- LDA: 57.8%
- Placebo: 52.7%

RD 5.1%
p = 0.093

Schisterman et al. The Lancet 2014
Sensitivity analysis: methods

• Idea: impute missing outcome to:
  • success (live birth)
  • failure (no live birth)

• Consider all possible imputations between the two randomized treatment arms (uniformly)

• For each possible imputation, calculate the difference in probability of live birth and calculate the p-value for the chi-square test (n = 1228)

(Hollis et al Stat Med 2002)
### Sensitivity analysis: overall

Success rates with various methods of allocating missing outcomes to success (live birth) or failure.

<table>
<thead>
<tr>
<th>Method</th>
<th>LDA</th>
<th>Placebo</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Complete cases</strong></td>
<td>57.54</td>
<td>52.27</td>
<td>0.0805</td>
</tr>
<tr>
<td>Allocation to poor outcome</td>
<td>50.24</td>
<td>46.98</td>
<td>0.2528</td>
</tr>
<tr>
<td>Allocation to good outcome</td>
<td>62.93</td>
<td>57.10</td>
<td>0.0370</td>
</tr>
<tr>
<td><strong>Best Case</strong></td>
<td>62.93</td>
<td>46.98</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td><strong>Worst Case</strong></td>
<td>50.24</td>
<td>57.10</td>
<td>0.0161</td>
</tr>
<tr>
<td>*Multiple Imputation</td>
<td>52.32</td>
<td>47.69</td>
<td>0.0945</td>
</tr>
</tbody>
</table>

*Model: Rx, nulliparity, age, race, married, eligibility strata, BMI, time since loss, Log CRP
Sensitivity analysis: Plot

LDA vs. Placebo on Live Birth, RR = 1.10
(P-values)

Placebo: 62 missing
LDA: 78 missing

CC p-value = 0.081
Sensitivity analysis: Plot Regions

LDA vs. Placebo on Live Birth (P-values)

Aspirin: % of missing allocated to LB
Placebo: % of missing allocated to LB

LDA Harmful
LDA No Effect
LDA Effective

Complete
Sensitivity analysis: Plot Extremes

LDA vs. Placebo on Live Birth
(P-values)

All Poor = 0.253

All Good = 0.037
Sensitivity analysis: Plot Extremes

LDA vs. Placebo on Live Birth
(P-values)

Worst Case = 0.016
All Poor = 0.253
All Good = 0.037
Best Case < .0001
Sensitivity analysis: What’s reasonable?

LDA vs. Placebo on Live Birth (P-values)

Complete

MNAR

Bounds contain even the most extreme mechanism
Sensitivity analysis: What’s reasonable?

LDA vs. Placebo on Live Birth (P-values)

- **Complete MI**
- **MNAR**: Triangle region is what is intuitively reasonable under completely random missingness.
- **MCAR**: Bounds contain even the most extreme mechanism.
Sensitivity analysis: What’s reasonable?

What is reasonable under missing at random?

LDA vs. Placebo on Live Birth (P-values)

Aspirin: % of missing allocated to LB

Placebo: % of missing allocated to LB

Complete

Triangle region is what is intuitively reasonable under missing completely at random

Bounds contain even the most extreme mechanism

MNAR

MCAR

MAR?
Sensitivity analysis: MI

Multiple Imputation RR=1.10
(500 Imputed Datasets)

- Aspirin: % of missing allocated to LB
- Placebo: % of missing allocated to LB

- CC p-value = 0.081
- MI p-value = 0.094

*Model: Rx, nulliparity, age, race, married, eligibility strata, BMI, time since loss, Log CRP
Sensitivity analysis: overall

Multiple Imputation RR=1.10
(500 Imputed Datasets)

500 imputed datasets give a region that is reasonable under missing at random?

Multiple Imputation RR=1.10
(500 Imputed Datasets)

500 imputed datasets give a region that is reasonable under missing at random?

MNAR

Bounds contain even the most extreme mechanism

MCAR

Triangle region is what is intuitively reasonable under missing completely at random
Sensitivity analysis: overall

Multiple Imputation RR=1.10
(500 Imputed Datasets)

Most show no effect

...but 40% do
Live Births: Overall

Would the conclusions have changed if we had observed all of the data?

RD 5.1%
p=0.093

Overall

LDA

Placebo

Schisterman et. al. The Lancet 2014
Sensitivity analysis: By Stratum

**Original**
- Information available for 495 of the 549 participants (90%)
- Missing Outcome:
  - 22 Placebo
  - 32 LDA

**Expanded**
- Information available for 593 of the 679 participants (87%)
- Missing Outcome:
  - 40 Placebo
  - 46 LDA
## Sensitivity analysis: By Stratum

<table>
<thead>
<tr>
<th></th>
<th>Original</th>
<th>Expanded</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LDA</td>
<td>Placebo</td>
</tr>
<tr>
<td><strong>Complete cases</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LDA</td>
<td>62.14</td>
<td>53.17</td>
</tr>
<tr>
<td>Placebo</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P-value</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>All poor</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LDA</td>
<td>54.91</td>
<td>48.91</td>
</tr>
<tr>
<td>Placebo</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P-value</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>All good</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LDA</td>
<td>66.55</td>
<td>56.93</td>
</tr>
<tr>
<td>Placebo</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P-value</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Best Case</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LDA</td>
<td>66.55</td>
<td>48.91</td>
</tr>
<tr>
<td>Placebo</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P-value</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Worst Case</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LDA</td>
<td>54.91</td>
<td>56.93</td>
</tr>
<tr>
<td>Placebo</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P-value</td>
<td></td>
<td></td>
</tr>
<tr>
<td>*<strong>MI</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LDA</td>
<td>59.33</td>
<td>50.64</td>
</tr>
<tr>
<td>Placebo</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P-value</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Model: Rx, nulliparity, age, race, married, eligibility strata, BMI, time since loss*
Sensitivity analysis: By Stratum

LDA vs. Placebo on Live Birth
(P-values)

Original

Expanded
Sensitivity analysis: By Stratum

LDA vs. Placebo on Live Birth

(P-values)

**Original**

Most show effect

**Expanded**

No evidence

Mostly significant
Sensitivity analysis: overall

LDA vs. Placebo on **Positive Pregnancy Test**

(P-values)

Consistent Effect
Sensitivity analysis: By Stratum

LDA vs. Placebo on Positive Pregnancy Test (P-values)

Original

Expanded

Overwhelming effect

No evidence
The Prevention and Treatment of Missing Data in Clinical Trials

Key finding:

“Substantial instances of missing data are a serious problem that undermines the scientific credibility of causal conclusions from clinical trials. The assumption that analysis methods can compensate for such missing data are not justified, so aspects of trial design that limit the likelihood of missing data should be an important objective. In addition to specific aspects of trial design, many components of clinical-trial conduct can limit the extent of missing data. Finally, in studies with missing data, analysis methods that are based on plausible scientific assumptions should be used. For example, this consideration often rules out simple fixes, such as imputation by the last observation carried forward. Although there are better analysis alternatives to that method, they all require unverifiable assumptions. Thus, sensitivity analyses should be conducted to assess the robustness of findings to plausible alternative assumptions about the missing data.”

Missing outcome data: “How to”
What is the impact on the risk ratio?

LDA and pregnancy with male offspring
What this section will cover

• Motivating example

• Plot: sensitivity of RR to missing outcome data

• SAS program
  • Addresses confounding, selection with weights (optional)
  • Generates data
  • Analyzes data
  • (R program)
Inflammation and implantation

- Endometrium as biosensor
  - Responds to embryonic signals
  - Appropriate regulation of inflammatory response
  - LDA may modulate overactive inflammation

Macklon and Brosens *Biol Reprod* 2014
Quenby *Human Reprod* 1999
Are male embryos more vulnerable?

- Male and female pre-implantation embryos differ on:
  - Response to maternal inflammation
  - Gene expression
  - Metabolism

References:
Perez-Crespo *Mol Rep Dev* 2005,
Dobbs *Biol Rep* 2014,
Kay *Cell* 1994,
Bermejo-Alvarez *PNAS* 2010,
Ray *J Reprod Fertil* 1995
Pregnancy Follow-up

4 weeks’ GA – end-cycle study visit

Study pill plus folic acid

Genetic testing of clinical pregnancy losses
## Intent-to-treat analysis

<table>
<thead>
<tr>
<th></th>
<th>LDA</th>
<th>Placebo</th>
</tr>
</thead>
<tbody>
<tr>
<td>No pregnancy detected</td>
<td>203 (33%)</td>
<td>228 (37%)</td>
</tr>
<tr>
<td>Pregnancy:</td>
<td>412</td>
<td>385</td>
</tr>
<tr>
<td>Pregnancy loss – no sex determined</td>
<td>73 (12%)</td>
<td>75 (12%)</td>
</tr>
<tr>
<td>female offspring</td>
<td>164 (27%)</td>
<td>173 (28%)</td>
</tr>
<tr>
<td><strong>male offspring</strong></td>
<td><strong>175 (28%)</strong></td>
<td><strong>137 (22%)</strong></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>615 (100%)</strong></td>
<td><strong>613 (100%)</strong></td>
</tr>
</tbody>
</table>
What about sex ratio at implantation?

Conception (primary sex ratio)

Unrecognized early losses

Implantation

Early pregnancy Losses (65)

Clinical Losses (83 had no sex determined)

Ultrasound confirmation

Live birth (secondary sex ratio)
Male offspring among pregnancies: LDA vs. placebo

Complete Case
52% vs. 44%
RR = 1.17 (1.02,1.33)

Multiple Imputation
51% vs. 44%
RR = 1.16 (1.02,1.34)
Male offspring among pregnancies: LDA vs. placebo
• SAS program –
  • Addresses confounding, selection with weights (optional)
  • Generates data for every scenario
  • Analyzes data
## Cross tab

<table>
<thead>
<tr>
<th></th>
<th>LDA</th>
<th>Placebo</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boy</td>
<td>175</td>
<td>137</td>
</tr>
<tr>
<td>Girl</td>
<td>164</td>
<td>173</td>
</tr>
<tr>
<td>Missing</td>
<td>73</td>
<td>75</td>
</tr>
<tr>
<td>TOTAL</td>
<td>412</td>
<td>385</td>
</tr>
</tbody>
</table>

/* distribution of girls, boys, and missings by exposure among women with PPT */
/* N = number observed, M = number missing in that exposure arm */

data mylib.crosstab;
input exposure$ outcome$ N M;
cards;
Aspirin boy  175  73
Aspirin girl 164  73
Placebo boy  137  75
Placebo girl 173  75
;run;
Apply weights (optional)

- Address confounding or selection in your data

<table>
<thead>
<tr>
<th></th>
<th>LDA</th>
<th>Placebo</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boy</td>
<td>175</td>
<td>137</td>
</tr>
<tr>
<td>Girl</td>
<td>164</td>
<td>173</td>
</tr>
<tr>
<td>Missing</td>
<td>73</td>
<td>75</td>
</tr>
<tr>
<td>TOTAL</td>
<td>412</td>
<td>385</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>LDA</th>
<th>Placebo</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boy</td>
<td>261.23</td>
<td>218.13</td>
</tr>
<tr>
<td>Girl</td>
<td>244.81</td>
<td>275.45</td>
</tr>
<tr>
<td>Missing</td>
<td>108.96</td>
<td>119.42</td>
</tr>
<tr>
<td>TOTAL</td>
<td>615</td>
<td>613</td>
</tr>
</tbody>
</table>

```r
data mylibwts.crosstabwt; set mylibwts.crosstab;

w_LDA_preg = 1 / (412/615);
w_Placebo_preg = 1 /(385/613);

if exposure = 'Aspirin' then do;
    N = N * w_lla_preg;  M = M * w_lla_preg; end;
else if exposure = 'Placebo' then do;
    N = N * w_placebo_preg;  M = M * w_placebo_preg; end;
```
• SAS program –
  • Addresses confounding, selection with weights (optional)
  • Generates data for every scenario
  • Analyzes data
  • R program
Create the outcome distribution under every scenario (0 – 100%)

<table>
<thead>
<tr>
<th>exposure</th>
<th>outcome</th>
<th>x1</th>
<th>x2</th>
<th>x10200</th>
<th>x10201</th>
</tr>
</thead>
<tbody>
<tr>
<td>aspirin</td>
<td>boy</td>
<td>261.23</td>
<td>261.23</td>
<td>370.19</td>
<td>370.19</td>
</tr>
<tr>
<td>aspirin</td>
<td>girl</td>
<td>353.77</td>
<td>353.77</td>
<td>244.81</td>
<td>244.81</td>
</tr>
<tr>
<td>placebo</td>
<td>boy</td>
<td>218.13</td>
<td>219.33</td>
<td>336.35</td>
<td>337.55</td>
</tr>
<tr>
<td>placebo</td>
<td>girl</td>
<td>394.87</td>
<td>393.67</td>
<td>276.65</td>
<td>275.45</td>
</tr>
<tr>
<td>aspirin</td>
<td>pct</td>
<td>0.00</td>
<td>0.00</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td>placebo</td>
<td>pct</td>
<td>0.00</td>
<td>1.00</td>
<td>99.00</td>
<td>100.00</td>
</tr>
</tbody>
</table>
Create the outcome distribution under every scenario (0 – 100%)

<table>
<thead>
<tr>
<th>exposure</th>
<th>outcome</th>
<th>x1</th>
<th>x2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aspirin</td>
<td>boy</td>
<td>261.23</td>
<td>261.23</td>
</tr>
<tr>
<td>Aspirin</td>
<td>girl</td>
<td>353.77</td>
<td>353.77</td>
</tr>
<tr>
<td>Placebo</td>
<td>boy</td>
<td>218.13</td>
<td>219.33</td>
</tr>
<tr>
<td>Placebo</td>
<td>girl</td>
<td>394.87</td>
<td>393.67</td>
</tr>
<tr>
<td>Aspirin</td>
<td>pct</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Placebo</td>
<td>pct</td>
<td>0.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

**Diagram:**

- **Aspirin:** % of losses allocated to boy
- **Placebo:** % of losses allocated to boy

**Legend:**

- X1: 0.00
- X2: 100.00

**Data Table:**

<table>
<thead>
<tr>
<th>exposure</th>
<th>outcome</th>
<th>x1</th>
<th>x2</th>
<th>x10200</th>
<th>x10201</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aspirin</td>
<td>boy</td>
<td>261.23</td>
<td>261.23</td>
<td>370.19</td>
<td>370.19</td>
</tr>
<tr>
<td>Aspirin</td>
<td>girl</td>
<td>353.77</td>
<td>353.77</td>
<td>244.81</td>
<td>244.81</td>
</tr>
<tr>
<td>Placebo</td>
<td>boy</td>
<td>218.13</td>
<td>219.33</td>
<td>336.35</td>
<td>337.55</td>
</tr>
<tr>
<td>Placebo</td>
<td>girl</td>
<td>394.87</td>
<td>393.67</td>
<td>276.65</td>
<td>275.45</td>
</tr>
<tr>
<td>Aspirin</td>
<td>pct</td>
<td>0.00</td>
<td>0.00</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td>Placebo</td>
<td>pct</td>
<td>0.00</td>
<td>1.00</td>
<td>99.00</td>
<td>100.00</td>
</tr>
</tbody>
</table>
data mylibwts.wide_data;set mylibwts.crosstabwt;
array new x1-x10201;
j=1;
do a=0 to 100 by 1; * do loop in aspirin *;
do p=0 to 100 by 1; * do loop in placebo *;
   if exposure in ('Aspirin') and outcome in ('boy') then new[j]=N+(a/100)*M;
   if exposure in ('Aspirin') and outcome in ('girl') then new[j]=N+((100-a)/100)*M;
   if exposure in ('Placebo') and outcome in ('boy') then new[j]=N+(p/100)*M;
   if exposure in ('Placebo') and outcome in ('girl') then new[j]=N+((100-p)/100)*M;
   if exposure in ('Aspirin') and outcome in ('pct') then new[j]=a;
   if exposure in ('Placebo') and outcome in ('pct') then new[j]=p;
   j+1;
end;
end;

<table>
<thead>
<tr>
<th>exposure</th>
<th>outcome</th>
<th>x1</th>
<th>x2</th>
<th>............</th>
<th>x10200</th>
<th>x10201</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aspirin</td>
<td>boy</td>
<td>261.23</td>
<td>261.23</td>
<td></td>
<td>370.19</td>
<td>370.19</td>
</tr>
<tr>
<td>Aspirin</td>
<td>girl</td>
<td>353.77</td>
<td>353.77</td>
<td></td>
<td>244.81</td>
<td>244.81</td>
</tr>
<tr>
<td>Placebo</td>
<td>boy</td>
<td>218.13</td>
<td>219.33</td>
<td></td>
<td>336.35</td>
<td>337.55</td>
</tr>
<tr>
<td>Placebo</td>
<td>girl</td>
<td>394.87</td>
<td>393.67</td>
<td></td>
<td>276.65</td>
<td>275.45</td>
</tr>
<tr>
<td>Aspirin</td>
<td>pct</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td>Placebo</td>
<td>pct</td>
<td>0.00</td>
<td>1.00</td>
<td></td>
<td>99.00</td>
<td>100.00</td>
</tr>
</tbody>
</table>
Alternative approach to creating the outcome distribution (whole numbers)

Alternative approach you could use when analyzing unweighted data

```r
data mylibnum.wide_data;set mylibnum.crosstab;
array new x1-x5624;
j=1;
do i=0 to 73 by 1;  * do loop in aspirin *;
do k=0 to 75 by 1;  * do loop in placebo *
   if exposure in ('Aspirin') and outcome in ('boy') then new[j]=N+i;
   if exposure in ('Aspirin') and outcome in ('girl') then new[j]=N+(73-i);
   if exposure in ('Placebo') and outcome in ('boy') then new[j]=N+k;
   if exposure in ('Placebo') and outcome in ('girl') then new[j]=N+(75-k);
   if exposure in ('Aspirin') and outcome in ('pct') then new[j]=i;
   if exposure in ('Placebo') and outcome in ('pct') then new[j]=k;
   j+1;
end;
end;
```
• SAS program –
  • Addresses confounding, selection with weights (optional)
  • Generates data for every scenario
  • Analyzes data
  • R program
Wide data set for analysis

<table>
<thead>
<tr>
<th>exposure</th>
<th>outcome</th>
<th>x1</th>
<th>x2</th>
<th>..........</th>
<th>x10200</th>
<th>x10201</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aspirin</td>
<td>boy</td>
<td>261.226</td>
<td>261.226</td>
<td>..........</td>
<td>370.194</td>
<td>370.194</td>
</tr>
<tr>
<td>Aspirin</td>
<td>girl</td>
<td>353.774</td>
<td>353.774</td>
<td>..........</td>
<td>244.806</td>
<td>244.806</td>
</tr>
<tr>
<td>Placebo</td>
<td>boy</td>
<td>218.132</td>
<td>219.327</td>
<td>..........</td>
<td>336.354</td>
<td>337.548</td>
</tr>
<tr>
<td>Placebo</td>
<td>girl</td>
<td>394.868</td>
<td>393.673</td>
<td>..........</td>
<td>276.646</td>
<td>275.452</td>
</tr>
</tbody>
</table>
Analyze tabular data

<table>
<thead>
<tr>
<th>exposure</th>
<th>outcome</th>
<th>N</th>
<th>M</th>
<th>x1</th>
<th>x2</th>
<th>..........</th>
<th>x10200</th>
<th>x10201</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aspirin</td>
<td>boy</td>
<td>175</td>
<td>73</td>
<td>261.226</td>
<td>261.226</td>
<td>..........</td>
<td>370.194</td>
<td>370.194</td>
</tr>
<tr>
<td>Aspirin</td>
<td>girl</td>
<td>164</td>
<td>73</td>
<td>353.774</td>
<td>353.774</td>
<td>..........</td>
<td>244.806</td>
<td>244.806</td>
</tr>
<tr>
<td>Placebo</td>
<td>boy</td>
<td>137</td>
<td>75</td>
<td>218.132</td>
<td>219.327</td>
<td>..........</td>
<td>336.354</td>
<td>337.548</td>
</tr>
<tr>
<td>Placebo</td>
<td>girl</td>
<td>173</td>
<td>75</td>
<td>394.868</td>
<td>393.673</td>
<td>..........</td>
<td>276.646</td>
<td>275.452</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>boy</th>
<th>girl</th>
<th>P_PCHI</th>
<th><em>RDIF1</em></th>
<th><em>RRC1</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Aspirin</td>
<td>261.226</td>
<td>353.774</td>
<td>0.013318</td>
<td>0.068913</td>
<td>1.19366</td>
</tr>
<tr>
<td>Placebo</td>
<td>218.132</td>
<td>394.868</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

```
proc freq data=mylibwts.wide_data noprint;
tables exposure*outcome
/chisq relrisk riskdiff nopercent;
weight x1;
output out=out1(keep=p_pchi _RDIF1_ _RRC1_) chisq relrisk riskdiff;
run;
```
Analyze tabular data…10,201 times

```sas
/* data mylibwts.chisq_results_w;
   input a;
cards;
1
2
3
*/

/* macro DOCHISQ_W; */
%do i=1 %to 10201;

proc freq data=mylibwts.wide_data(where=(outcome in ("boy","girl"))) noprint;
tables exposure*outcome
/chisq relrisk riskdiff nopercent;
weight x&i;
output out=out&i(keep=p_pchi_RDIF1__RRC1__) chisq relrisk riskdiff;
run;

data mylibwts.chisq_results_w;
set mylibwts.chisq_results_w out&i;
run;

proc datasets;
delete out&i;
run;
%end;
%mend DOCHISQ_W;

%DOCHISQ_W;
quit;
```
Results from 10,201 allocation scenarios

<table>
<thead>
<tr>
<th>NAME</th>
<th>Aspirin</th>
<th>Placebo</th>
<th>P_PCHI</th>
<th>RDIF1_</th>
<th>RRC1_</th>
</tr>
</thead>
<tbody>
<tr>
<td>x1</td>
<td>0</td>
<td>0</td>
<td>0.013</td>
<td>0.069</td>
<td>1.194</td>
</tr>
<tr>
<td>x2</td>
<td>0</td>
<td>1</td>
<td>0.016</td>
<td>0.067</td>
<td>1.187</td>
</tr>
<tr>
<td>x3</td>
<td>0</td>
<td>2</td>
<td>0.020</td>
<td>0.065</td>
<td>1.181</td>
</tr>
<tr>
<td>x10199</td>
<td>100</td>
<td>98</td>
<td>0.050</td>
<td>0.055</td>
<td>1.101</td>
</tr>
<tr>
<td>x10200</td>
<td>100</td>
<td>99</td>
<td>0.059</td>
<td>0.053</td>
<td>1.097</td>
</tr>
<tr>
<td>x10201</td>
<td>100</td>
<td>100</td>
<td>0.069</td>
<td>0.051</td>
<td>1.093</td>
</tr>
</tbody>
</table>
### Make wide data tall: transpose

<table>
<thead>
<tr>
<th>exposure</th>
<th>outcome</th>
<th>x1</th>
<th>x2</th>
<th>..........</th>
<th>x10200</th>
<th>x10201</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aspirin</td>
<td>boy</td>
<td>261.226</td>
<td>261.226</td>
<td>..........</td>
<td>370.194</td>
<td>370.194</td>
</tr>
<tr>
<td>Aspirin</td>
<td>girl</td>
<td>353.774</td>
<td>353.774</td>
<td>..........</td>
<td>244.806</td>
<td>244.806</td>
</tr>
<tr>
<td>Placebo</td>
<td>boy</td>
<td>218.132</td>
<td>219.327</td>
<td>..........</td>
<td>336.354</td>
<td>337.548</td>
</tr>
<tr>
<td>Placebo</td>
<td>girl</td>
<td>394.868</td>
<td>393.673</td>
<td>..........</td>
<td>276.646</td>
<td>275.452</td>
</tr>
<tr>
<td>Aspirin</td>
<td>pct</td>
<td>0</td>
<td>0</td>
<td></td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Placebo</td>
<td>pct</td>
<td>0</td>
<td>1</td>
<td></td>
<td>99</td>
<td>100</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><em>NAME</em></th>
<th>Aspirin</th>
<th>Aspirin</th>
<th>Placebo</th>
<th>Placebo</th>
<th>Aspirin</th>
<th>Placebo</th>
</tr>
</thead>
<tbody>
<tr>
<td>x1</td>
<td>261.226</td>
<td>353.774</td>
<td>218.132</td>
<td>394.868</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>x2</td>
<td>261.226</td>
<td>353.774</td>
<td>219.327</td>
<td>393.673</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>x10200</td>
<td>370.194</td>
<td>244.806</td>
<td>336.354</td>
<td>276.646</td>
<td>100</td>
<td>99</td>
</tr>
<tr>
<td>x10201</td>
<td>370.194</td>
<td>244.806</td>
<td>337.548</td>
<td>275.452</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>
Make wide data tall: transpose

```sas
proc transpose data=mylibwts.wide_data out=mylibwts.tall_data; run;

data mylibwts.tall_data2; set mylibwts.tall_data;
  * delete any row that is not an allocation scenario;
  if _NAME_ in ('N' 'M' 'j' 'i' 'k' 'a' 'p' 'w_LDA_preg' 'w_Placebo_preg') then delete;
  rename COL1=Aspirin_boy COL2=Aspirin_girl
      COL3=Placebo_boy COL4=Placebo_girl
      COL5=Aspirin COL6=Placebo;
  id+1;
```
Export results to .csv file

```sas
/* data */
data mylibwts.data_To_Plot_wt;
merge mylibwts.tall_data2(keep=_NAME_ Aspirin Placebo id)
    mylibwts.chisq_results_w;
by id;
drop id;
run;

/* proc export */
proc export data=mylibwts.data_To_Plot_wt
    outfile= "C:\Users\radinrg\Documents\Advanced Methods Workshop\data\sensitivity_results_wt.csv"
    dbms=csv replace;
dputnames=yes;
run;
```
Heat Map: all possible scenarios

Aspirin: % of losses allocated to boy
Placebo: % of losses allocated to boy

X_RRC1_
Overlay lines, dot: plausible scenarios

R Function plots the complete case and three lines to define plausible scenarios
Make the oval and dots: PROC MI

- Define plausible scenarios
- Each imputed data set represents one plausible scenario
SAS Code: PROC MI

```
proc mi data=complete seed=12345 out=MIdata nimpute=500;
  class boy1 time_last_loss1;
  fcs logistic(boy1/details) discrim(time_last_loss1/details);
  var   boy1
       is_treatment PriorLB0 bmi logcrp ← Associated with outcome
       age white married time_last_loss1 ← Associated with missingness
run;
```
SAS Log: PROC MI

- Fully conditional specification uses continuous predictors only.

```
proc mi data=preg seed=12345 out=boy1data nimpute=500;
   class boy1 time_last_loss1;
   fcs logistic(boy1/details) discr(time_last_loss1/details);
   var boy1 is_treatment PriorLB0 age white married bmi time_last_loss1 logcrp;
run;
```

WARNING: The covariates are not specified in an FCS discriminant method for variable `time_last_loss1`, only remaining continuous variables will be used as covariates with the default CLASSEFFECTS=EXCLUDE option.

NOTE: The data set WORK.BOYMIDATA has 399500 observations and 354 variables.

NOTE: PROCEDURE MI used (Total process time):
      real time  2:19.12
      cpu time   2:18.18
```
### Each imputation: % allocated to male

<table>
<thead>
<tr>
<th>Imputation</th>
<th>is_treatment</th>
<th>Imputed data set</th>
<th>Complete case</th>
<th>Number of pregnancies</th>
<th>Percent of missing allocated to male</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>286.60</td>
<td>218.13</td>
<td>119.42</td>
<td>57.33</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>277.04</td>
<td>218.13</td>
<td>119.42</td>
<td>49.33</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>272.27</td>
<td>218.13</td>
<td>119.42</td>
<td>45.33</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>265.90</td>
<td>218.13</td>
<td>119.42</td>
<td>40.00</td>
</tr>
</tbody>
</table>

\[
\text{Imputed} = \text{observed} + (\text{Missing} \times \text{percent} \times 0.01)
\]
Calculate the % allocated to male

**proc freq data=complete;**
**tables is_treatment*boy1/missing nocol norow nopercent chisq relrisk;**
**weight mypregwt;**
**ods output CrossTabFreqs=cross0;**
**run;**

* complete case: get number of boys in LDA and placebo groups *
**data male;**
**set cross0;**
**if boy1=1 and is_treatment~=. and _type_=11;**
**rename Frequency=n0;**
**label Frequency=n0;**
**keep is_treatment Frequency;**
**run;**

* complete case: get number of missing observations in LDA and placebo *
**data missing;**
**set cross0;**
**if boy1=. and is_treatment~=. and _type_=11;**
**rename Frequency=missing;**
**label Frequency=missing;**
**keep is_treatment Frequency;**
**run;**
Calculate the % allocated to male

```
proc freq data=MIdata;
tables is_treatment*boy1/norow nocol nopercent;
by _imputation_; 
weight mypregwt;
ods output CrossTabFreqs=cross;
run;
* MI data: get number of boys in LDA and placebo groups *;

data cross2;
set cross;
if boy1=1 and is_treatment=.;
keep _imputation_ is_treatment Frequency;
rename Frequency=n;
label Frequency=n;
run;

proc sort data=cross2;
by is_treatment;
run;

* Step 4: calculate % of missing allocated to male in LDA,
data cross3;
merge cross2 male missing;
by is_treatment;

percent=(n-n0)/missing*100;
label percent='percent of missing allocated to male';
run;
```
Sensitivity Plot R function
senPlot function in R

```r
senPlot=function(dat, imp=NA, obs.per=c(.5,.5), null.para=NA, null.tol=0.001, pvplot=TRUE, XY.names=list("X","Y"), missing="missing", outcome.name="Yes", obs.col="black", imp.col="black", pv.col=c("white","gray"), gradient.col=c("white","gray")){

- Data
- Plot type
- Axis labels
- Point estimate colors
- Background colors
```
senPlot function in R

```
> senPlot(dat=dat[,c(1,2,5)])
```

- Data: dat=c(X, Y, background)

```
> head(dat[,c(1,2,5)])
1    0  fisher.exact
0.0000000 0  0.12317317
0.5882353 0  0.10591367
1.1764706 0  0.09065954
1.7647059 0  0.07724975
2.3529412 0  0.06552393
2.9411765 0  0.05532477
```
**senPlot function in R**

- `senPlot(dat=dat[,c(1,2,5)], obs.per=per.obs[2:1])`

- **Data**

  ```
  > head(dat[,c(1,2,5)])
  1    0   fisher.exact
  0.0000000 0 0.123173170.5882353 0 0.105913671.1764706 0 0.090659541.7647059 0 0.077249752.3529412 0 0.065523932.9411765 0 0.05532477
  > per.obs[2:1]
  1 0
  56.95652 44.89796
  ```
senPlot function in R

```r
> senPlot(dat=dat[,c(1,2,5)],imp=imp.sen,obs.per=per.obs[2:1])
```

- **Data**: imp=c(X, Y)

```r
> head(imp.sen)
  1  0
52.35294 49.03226
50.00000 46.45161
55.29412 46.45161
54.70588 42.58065
54.11765 45.80645
57.05882 47.74194
```
senPlot function in R

```r
> senPlot(dat=dat[,c(1,2,5)],imp=imp.sen,obs.per=per.obs[2:1], pvplot=TRUE)
```

- **Plot type**

![Plot Diagram](image-url)
**senPlot function in R**

```
> senPlot(dat=dat[,c(1,2,5)], imp=imp.sen, obs.per=per.obs[2:1], pvplot=FALSE)
```

- **Plot type**

![Plot type](image)
senPlot function in R

```r
> senPlot(dat=dat[,c(1,2,5)],imp=imp.sen,obs.per=per.obs[2:1],pvplot=TRUE,
          XY.names=list("X","Y"), missing="missing",
          outcome.name="Yes")
```

- **Axis Labels**
senPlot function in R

```r
> senPlot(dat=dat[,c(1,2,5)],imp=imp.sen,obs.per=per.obs[2:1],pvplot=TRUE,
          XY.names=list("Aspirin","Placebo"), missing="losses",
          outcome.name="boy")
```

- **Axis Labels**
**senPlot function in R**

```r
> senPlot(dat=dat[,c(1,2,5)], imp=imp.sen, obs.per=per.obs[2:1], pvplot=TRUE,
          XY.names=list("Aspirin","Placebo"), missing="losses",
          outcome.name="boy", obs.col="black", imp.col="black")
```

- **Point estimate colors**
senPlot function in R

```r
> senPlot(dat=dat[,c(1,2,5)],imp=imp.sen,obs.per=per.obs[2:1],pvplot=TRUE, 
  XY.names=list("Aspirin","Placebo"), missing="losses",
  outcome.name="boy", obs.col="blue", imp.col="red")
```

- Point estimate colors
senPlot function in R

```r
> senPlot(dat=dat[,c(1,2,5)],imp=imp.sen,obs.per=per.obs[2:1],pvplot=TRUE,
    XY.names=list("Aspirin","Placebo"), missing="losses",
    outcome.name="boy", obs.col="blue", imp.col="red",
    pv.col=c("white","black"))
```

• Background colors
senPlot function in R

> senPlot(dat=dat[,c(1,2,5)], imp=imp.sen, obs.per=per.obs[2:1], pvplot=TRUE, XY.names=list("Aspirin","Placebo"), missing="losses", outcome.name="boy", obs.col="blue", imp.col="red", pv.col=c("yellow","green"))

- Background colors
senPlot function in R

> senPlot(dat=dat[,c(1,2,5)], imp=imp.sen, obs.per=per.obs[2:1], pvplot=FALSE, XY.names=list("Aspirin","Placebo"), missing="losses", outcome.name="boy", obs.col="blue", imp.col="red", Gradient.col=c("yellow","green"))

![Scatter plot example](image)
senPlot function in R

- Gradient examples
Sensitivity Analysis: Take home

• Sensitivity analysis are essential to evaluate the potential effect of missing data on study findings under various assumptions.
• Visualizing sensitivity analysis is a concise way to convey those potential effects under numerous situations.

• Thanks to:
  • Enrique F. Schisterman, PhD
  • The EAGeR Trial Team
  • SPER
  • And…

Eunice Kennedy Shriver National Institute of Child Health and Human Development