Global Sensitivity Analysis of Randomized Trials with Missing Data: From the Software Development Trenches

Daniel Scharfstein
Johns Hopkins University
dscharf@jhu.edu

November 13, 2015
Interested in comparing treatment groups with respect to the mean outcome at the last scheduled study visit.

Some patients prematurely drop out of the study.

The set of possible assumptions about the drop out mechanism is very large and cannot be fully explored.

Sensitivity analysis:
- Ad-hoc
- Local
- Global - “Tipping point”
Inference requires two types of assumptions:

(i) *unverifiable* assumptions about the distribution of outcomes among those who dropped out and

(ii) additional testable assumptions that serve to increase the efficiency of estimation.
Global Sensitivity Analysis

Restrictions on Distribution of Observed Data

- None
- Type (ii)

Type (i) Assumptions

Treatment-Specific Mean
- \( K \) scheduled post-baseline assessments.
- There are \((K + 1)\) patterns representing each of the visits an individual might last be seen, i.e., 0, \ldots, \( K \).
- The \((K + 1)^{st}\) pattern represents individuals who complete the study.
- Let \( Y_k \) be the outcome scheduled to be measured at visit \( k \), with visit 0 denoting the baseline measure (assumed to be observed).
- Let \( Y_k^- = (Y_0, \ldots, Y_k) \)
Let $R_k$ be the indicator of being on study at visit $k$

$R_0 = 1; R_k = 1$ implies that $R_{k-1} = 1$.

Let $C$ be the last visit that the patient is on-study.

We focus inference separately for each treatment arm.

The observed data for an individual is $O = (C, Y_C^-)$.

We want to estimate $\mu^* = E[Y_K]$. 
logit \( P[R_{k+1} = 0|R_k = 1, Y_{k+1}, Y_K] = h_{k+1}(Y_k^-) + \alpha r(Y_{k+1}) \)

where

\[ h_{k+1}(Y_k^-) = \text{logit } P[R_{k+1} = 0|R_k = 1, Y_k^-] - \log \{E[\exp\{\alpha r(Y_{k+1})]\}|R_{k+1} = 1, Y_k^-]\]

- \( r(Y_{k+1}) \) is a specified function of \( Y_{k+1} \)
- \( \alpha \) is a sensitivity analysis parameter
- Each \( \alpha \) is type (i) assumption.
Inference

- Inference will rely on models for either
  - \( f(Y_{k+1}|R_{k+1} = 1, Y_k^-) \)
  - \( P(R_{k+1} = 0 | R_k = 1, Y_k^-) \)
- Impose first-order Markov assumption (Type (ii) assumption)
- Non-parametric smoothing using cross-validation
- Corrected plug-in estimator using efficient influence function
- Confidence intervals using t-based bootstrap
Alternative Model

\[
\text{logit } P[R_{k+1} = 0| R_k = 1, Y_k^- , Y_K] = l_{k+1}(Y_k^-) + \alpha q(Y_K)
\]

where

\[
l_{k+1}(Y_k^-) = \text{logit } P[R_{k+1} = 0| R_k = 1, Y_k^-] - \log\{E[\exp\{\alpha r(Y_K)\}| R_{k+1} = 1, Y_k^-]\}
\]

- \( q(Y_K) \) is a specified function of \( Y_K \)
- \( \alpha \) is a sensitivity analysis parameter
- Each \( \alpha \) is type (i) assumption.
Major Challenges

- Wald confidence intervals with influence function-based standard errors perform poorly in sample sizes seen in registration trials.
  - Is it our simulation procedure?
- Intermittent missing data
  - Impute to a monotone structure?
Software, Papers, Presentations

www.missingdatamatters.org

- Funded by FDA and PCORI