Global Sensitivity Analysis of Randomized Trials with Missing Data

Novartis

Daniel Scharfstein
Johns Hopkins University
dscharf@jhu.edu

December 5, 2016

Funding Acknowledgments

- FDA
- PCORI

Case Study: Quetiapine Bipolar Trial

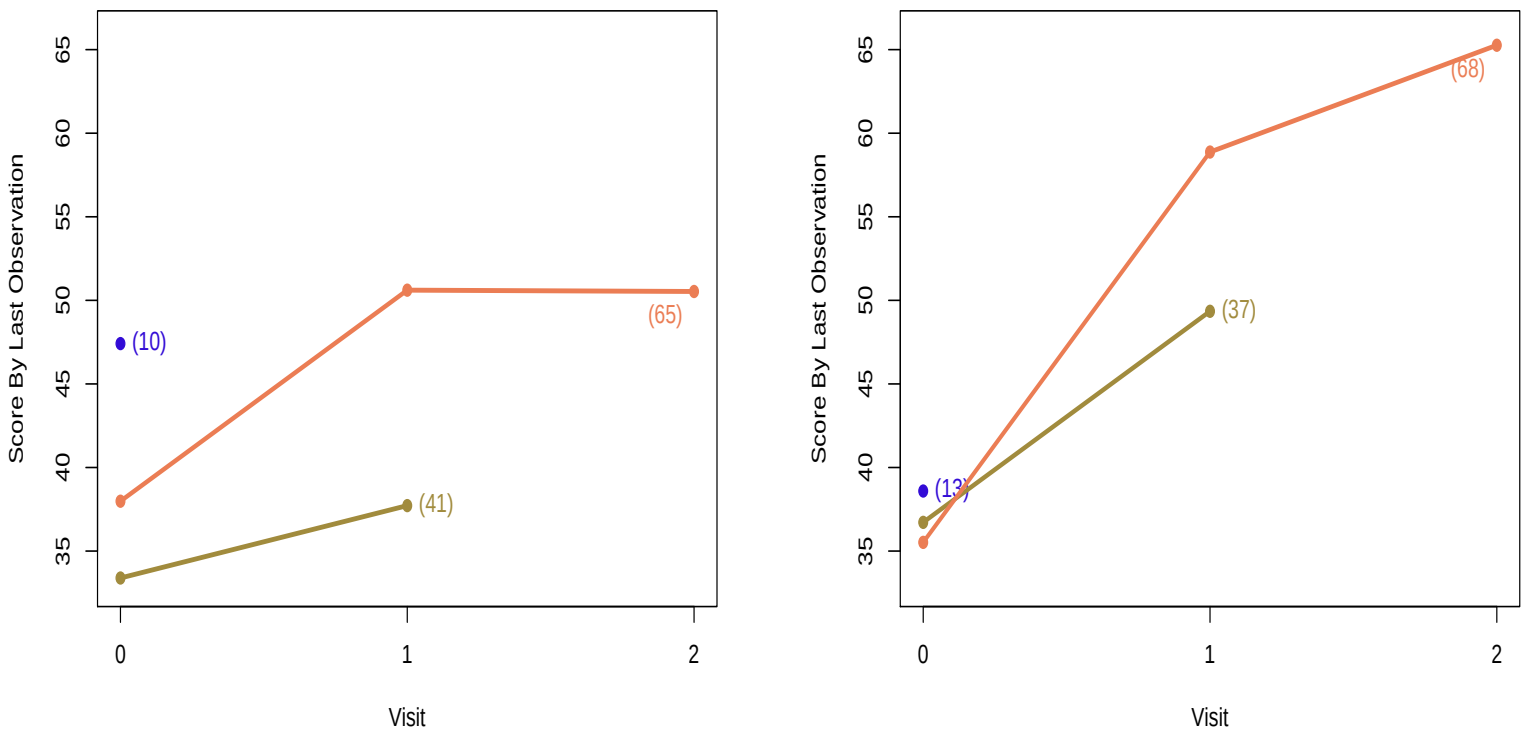
- Patients with bipolar disorder randomized equally to one of three treatment arms: placebo, Quetiapine 300 mg/day or Quetiapine 600 mg/day (Calabrese *et al.*, 2005).
- Randomization was stratified by type of bipolar disorder.
- Short-form version of the Quality of Life Enjoyment Satisfaction Questionnaire (QLESSF, Endicott *et al.*, 1993), was scheduled to be measured at baseline, week 4 and week 8.

Quetiapine Bipolar Trial

- Focus on the subset of 234 patients with bipolar 1 disorder who were randomized to either the placebo (n=116) or 600 mg/day (n=118) arms.
- Only 65 patients (56%) in placebo arm and 68 patients (58%) in the 600mg/day arm had a complete set of QLESSF scores.
- Patients with complete data tend to have higher average QLESSF scores, suggesting that a complete-case analysis could be biased.

Observed Data

Figure: Treatment-specific (left: placebo; right: 600 mg/day Quetiapine) trajectories of mean QLESSF scores, stratified by last available measurement.



Central Question

What is the difference in the mean QLESSF score at week 8 between Quetiapine 600 mg/day and placebo in the counterfactual world in which all patients were followed to that week?

Global Sensitivity Analysis

- Inference about the treatment arm means requires two types of assumptions:
  - (i) *unverifiable* assumptions about the distribution of outcomes among those with missing data and
  - (ii) additional testable assumptions that serve to increase the efficiency of estimation.

Global Sensitivity Analysis

- Type (i) assumptions are necessary to identify the treatment-specific means.
- By *identification*, we mean that we can write it as a function that depends only on the distribution of the observed data.
- When a parameter is identified we can hope to estimate it as precisely as we desire with a sufficiently large sample size,
- In the absence of identification, statistical inference is fruitless as we would be unable to learn about the true parameter value even if the sample size were infinite.

Global Sensitivity Analysis

- To address the identifiability issue, it is essential to conduct a sensitivity analysis, whereby the data analysis is repeated under different type (i) assumptions, so as to investigate the extent to which the conclusions of the trial are dependent on these subjective, unverifiable assumptions.
- The usefulness of a sensitivity analysis ultimately depends on the plausibility of the unverifiable assumptions.
- It is key that any sensitivity analysis methodology allow the formulation of these assumptions in a transparent and easy to communicate manner.

Global Sensitivity Analysis

- There are an infinite number of ways of positing type (i) assumptions.
- Ultimately, however, these assumptions prescribe how missing outcomes should be "imputed."
- A reasonable way to posit these assumptions is to
  - stratify individuals with missing outcomes according to the data that we were able to collect on them and the occasions at which the data were collected
  - separately for each stratum, hypothesize a connection (or link) between the distribution of the missing outcome with the distribution of the outcome among those with the observed outcome and who share the same recorded data.

Global Sensitivity Analysis

- Type (i) assumptions will not suffice when the repeated outcomes are continuous or categorical with many levels. This is because of *data sparsity*.
- For example, the stratum of people who share the same recorded data will typically be small. As a result, it is necessary to draw strength across strata by "smoothing."
- Without smoothing, the data analysis will rarely be informative because the uncertainty concerning the treatment arm means will often be too large to be of substantive use.
- As a result, it is necessary to impose type (ii) smoothing assumptions.
- Type (ii) assumptions should be scrutinized with standard model checking techniques.

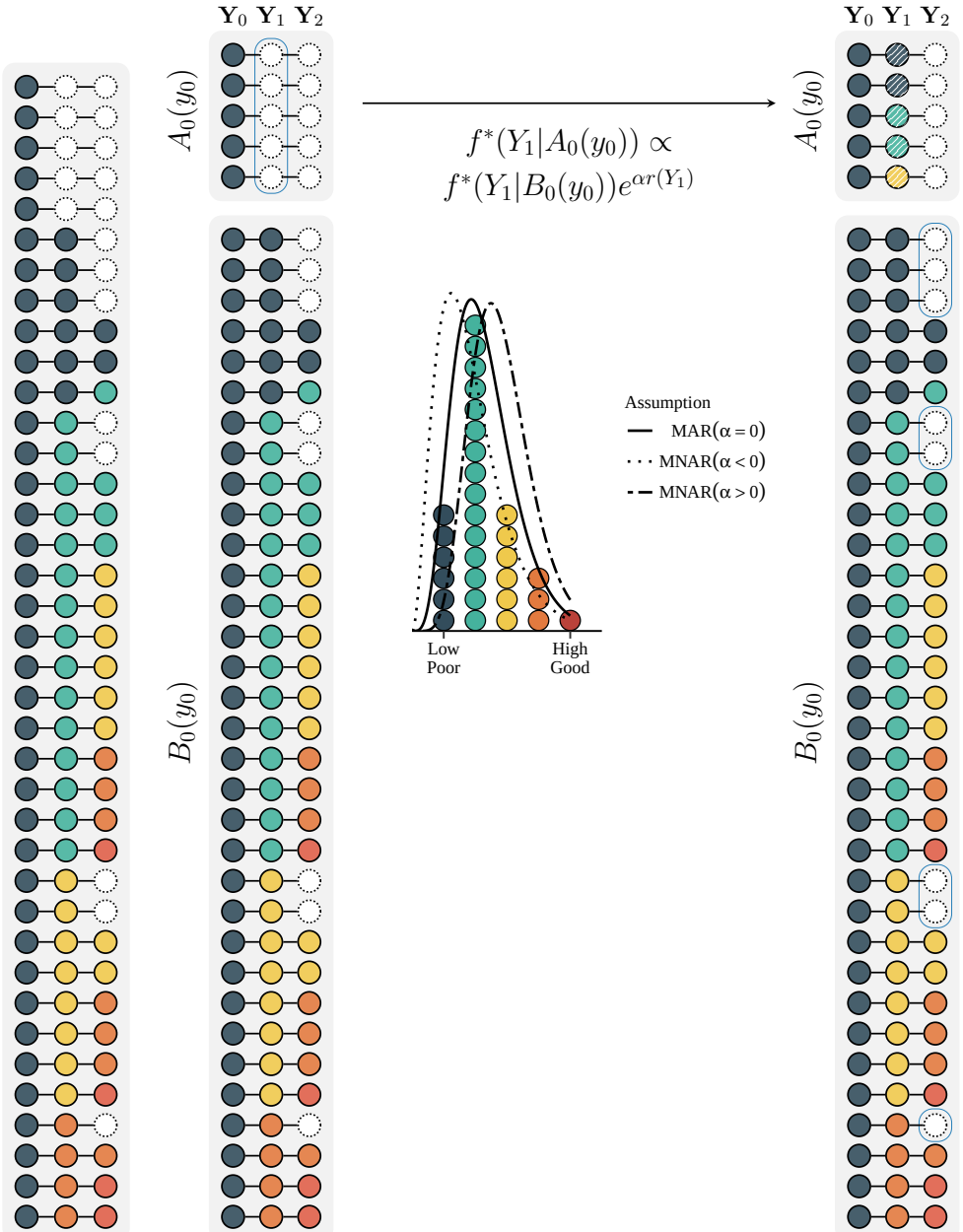
Global Sensitivity Analysis

- The global sensitivity framework proceeds by parameterizing (i.e., indexing) the connections (i.e., type (i) assumptions) via sensitivity analysis parameters.
- The parameterization is configured so that a specific value of the sensitivity analysis parameters (typically set to zero) corresponds to a benchmark connection that is considered reasonably plausible and sensitivity analysis parameters further from the benchmark value represent more extreme departures from the benchmark connection.

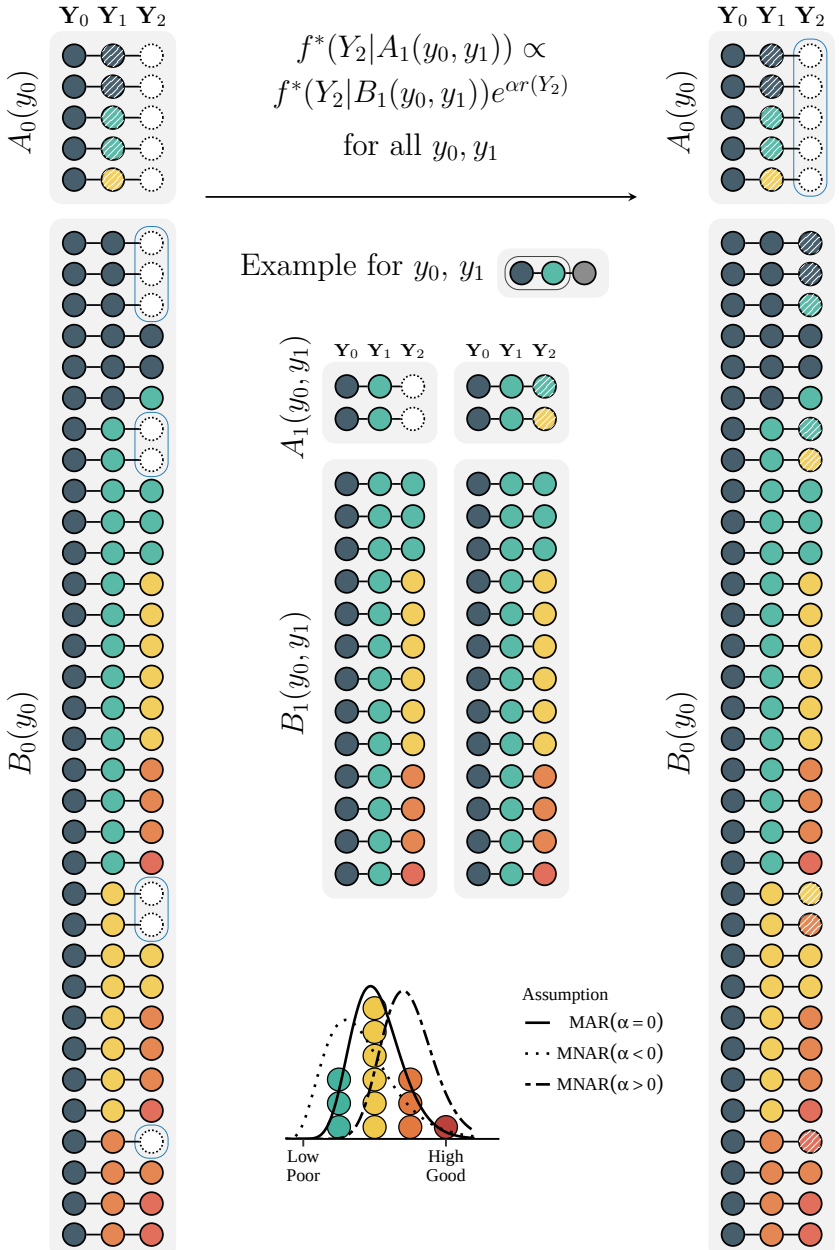
Global Sensitivity Analysis

- The global sensitivity analysis strategy that we propose is focused on separate inferences for each treatment arm, which are then combined to evaluate treatment effects.

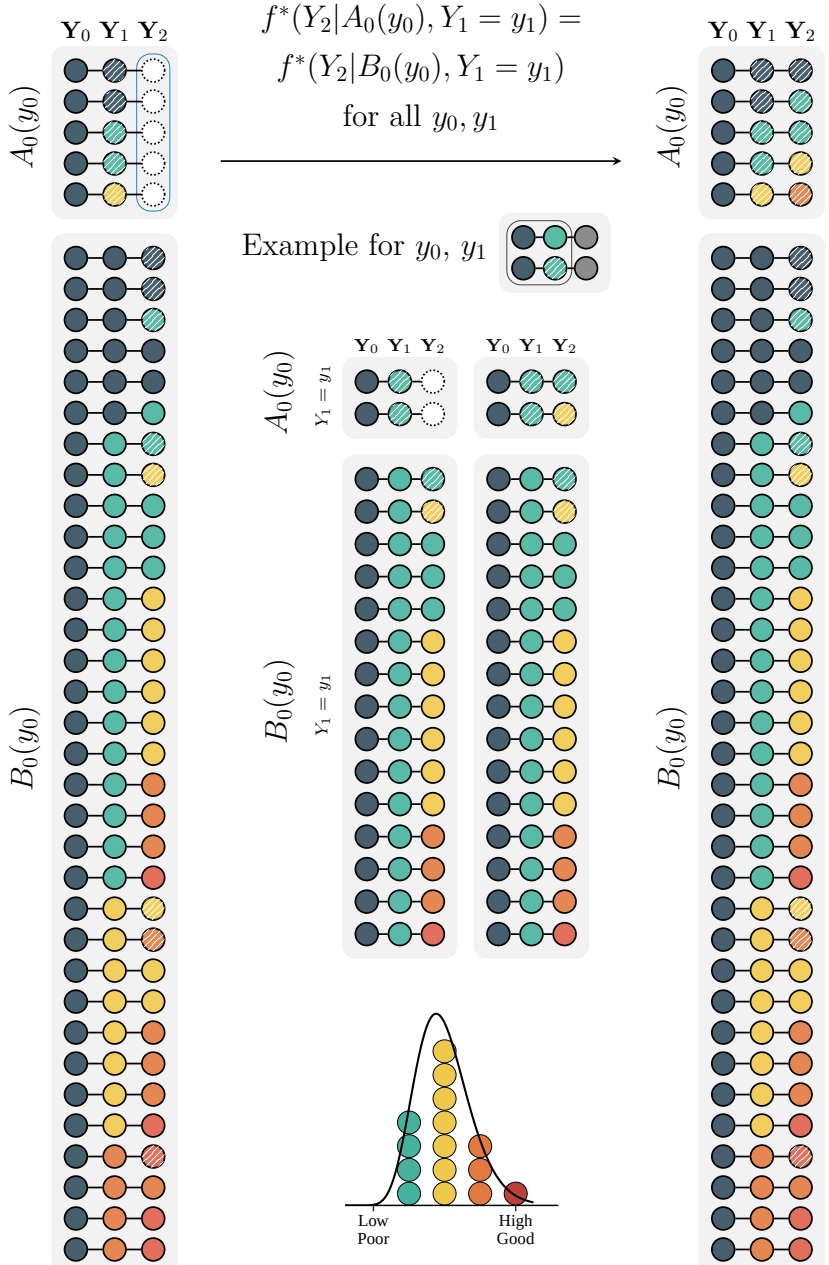
Global Sensitivity Analysis



Global Sensitivity Analysis



Global Sensitivity Analysis



Inference

For given α, inference depends on:

- $F_2(y_2|y_1, y_0) = P[Y_2 \leq y_2|R_2 = 1, Y_1 = y_1, Y_0 = y_0]$
- $F_1(y_1|y_0) = P[Y_1 \leq y_1|R_1 = 1, Y_0 = y_0]$
- $H_2(y_1, y_0) = P[R_2 = 0|R_1 = 1, Y_1 = y_1, Y_0 = y_0]$
- $H_1(y_0) = P[R_1 = 0|R_0 = 1, Y_0 = y_0]$
- $F_0(y_0) = P[Y_0 \leq y_0]$

Inference

- With the exception of $F_0(y_0)$, it is tempting to think that we can use non-parametric procedures to estimate these quantities.
- Curse of Dimensionality
- Make first-order Markovian type (ii) assumptions and estimate using non-parametric smoothing.
- Estimate mean as a "corrected" plug-in estimator.
- Confidence intervals - use symmetric studentized bootstrap with jackknife standard errors.
- SAMON 3.0 is available at `www.missingdatamatters.org`.

Quetiapine Bipolar Trial - Fit

- Estimated smoothing parameters for the drop-out model are 11.54 and 9.82 for the placebo and 600 mg arms.
- Estimated smoothing parameters for the outcome model are 6.34 and 8.05 for the placebo and 600 mg arms.
- In the placebo arm, the observed percentages of last being seen at visits 0 and 1 among those at risk at these visits are 8.62% and 38.68%. Model-based estimates are 7.99% and 38.19%.
- For the 600 mg arm, the observed percentages are 11.02% and 35.24% and the model-based estimates are 11.70% and 35.08%.

Quetiapine Bipolar Trial - Fit

- In the placebo arm, the Kolmogorov-Smirnov distances between the empirical distribution of the observed outcomes and the model-based estimates of the distribution of outcomes among those on-study at visits 1 and 2 are 0.013 and 0.033.
- In the 600 mg arm, these distances are 0.013 and 0.022.
- These results suggest that our model for the observed data fits the observed data well.

Quetiapine Bipolar Trial - MAR

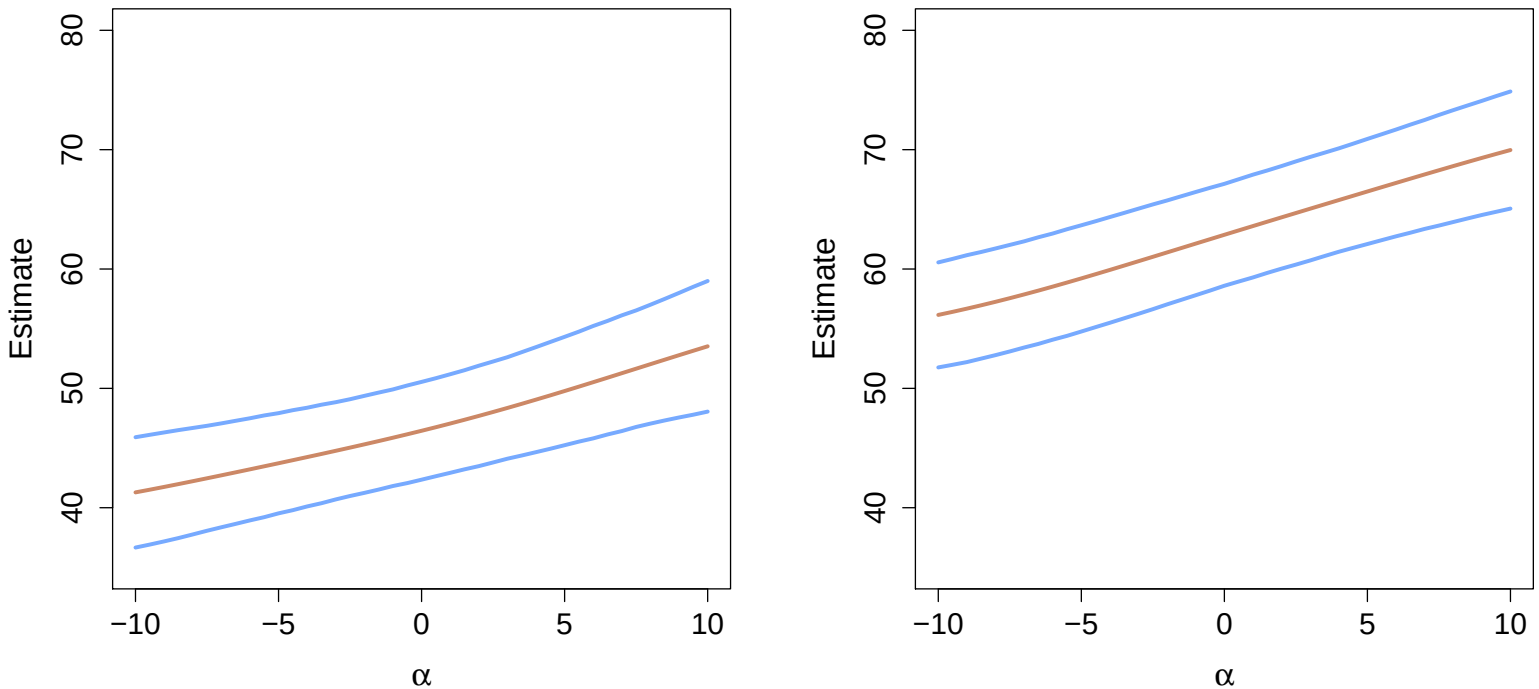
- Under MAR, the estimated values of μ^* are 46.45 (95% CI: 42.35,50.54) and 62.87 (95% CI: 58.60,67.14) for the placebo and 600 mg arms.
- The estimated difference between 600 mg and placebo is 16.42 (95% 10.34, 22.51)
- Statistically and clinically significant improvement in quality of life in favor of Quetiapine.

Quetiapine Bipolar Trial - Sensitivity Analysis

- We set $r(y) = y$ and ranged the sensitivity analysis parameter from -10 and 10 in each treatment arm.
- According to experts, there is no evidence to suggest that there is a differential effect of a unit change in QLESSF on the hazard of drop-out based on its location on the scale.

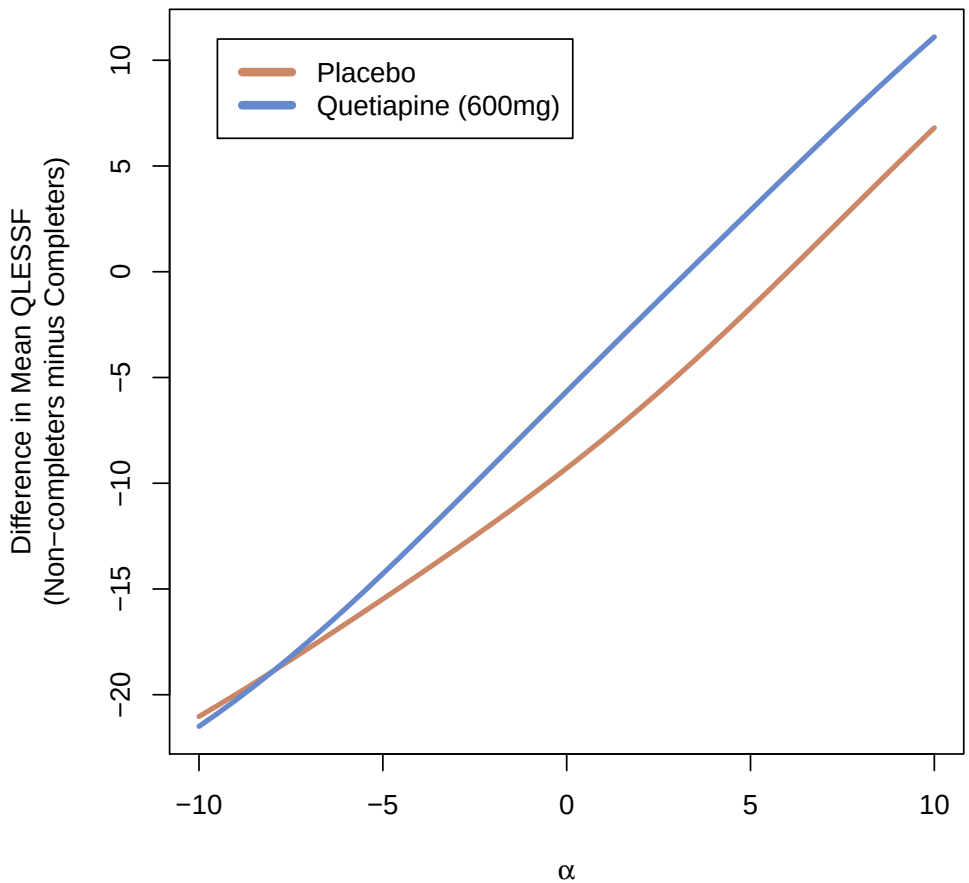
Quetiapine Bipolar Trial - Sensitivity Analysis

Figure: Treatment-specific (left: placebo; right: 600 mg/day Quetiapine) estimates (along with 95% pointwise confidence intervals) of μ^* as a function of α.



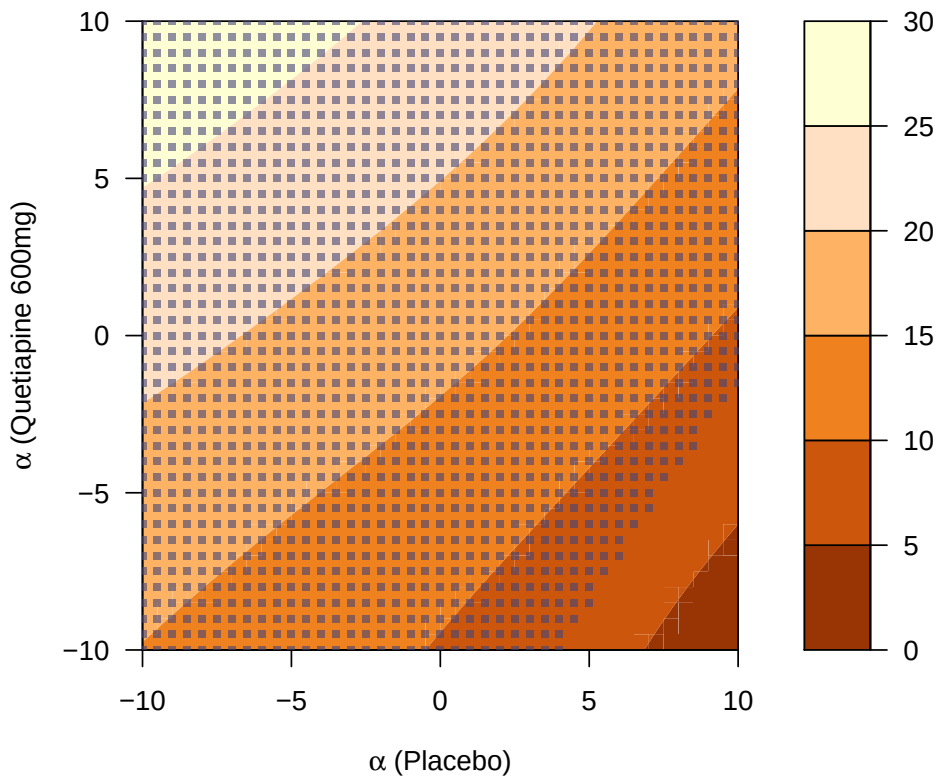
Quetiapine Bipolar Trial - Sensitivity Analysis

Figure: Treatment-specific differences between the estimated mean QLESSF at Visit 2 among non-completers and the estimated mean among completers, as a function of α.



Quetiapine Bipolar Trial - Sensitivity Analysis

Figure: Contour plot of the estimated differences between mean QLESSF at Visit 2 for Quetiapine vs. placebo for various treatment-specific combinations of the sensitivity analysis parameters.



Quetiapine Bipolar Trial - Sensitivity Analysis

- Only when the sensitivity analysis are highly differential (e.g., $\alpha(\text{placebo}) = 8$ and $\alpha(\text{Quetiapine}) = -8$) are the differences no longer statistically significant.
- Conclusions under MAR are highly robust.

Discussion

- Global vs. ad-hoc vs. local sensitivity analysis
- Mixed models
- Dissemination
- Collaboration