Analysis Strategies for Clinical Trials with Treatment Non-Adherence

Bohdana Ratitch, PhD

Acknowledgments:
Michael O'Kelly, James Roger, Ilya Lipkovich, DIA SWG On Missing Data

Outline

- Non-adherence to treatment, treatment effect/estimand types, missingness
- Strategies with explicit imputations of missing values under MAR and deviations from MAR
  - Imputation strategies with adjustments, tipping point analysis
  - Imputation strategies based on reference group
- Illustration with an example in Major Depressive Disorder

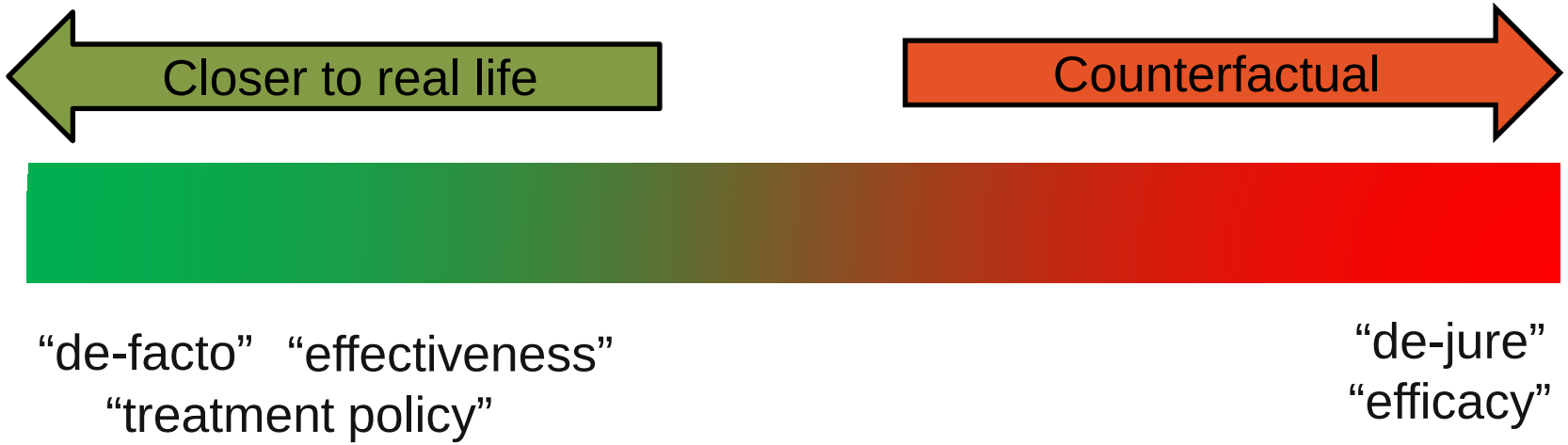
Non-adherence to treatment

- Adherence: completion of randomized treatment for planned duration of the trial, without rescue
- Non-adherence is a deviation from an ideal treatment administration plan
  - Early (permanent) discontinuation of randomized treatment, for any reason
  - Initiating rescue therapy in addition to the randomized treatment

Treatment effect / estimand types

- Effect, if treatment is administered as directed with full adherence/under ideal conditions: efficacy, a.k.a. *de jure*
  - Perfect adherence by all subjects can almost never be achieved, and efficacy effect can almost never be fully observed

- Effect, given the actual adherence to treatment/under realistic conditions: effectiveness, a.k.a. *de facto*
  - Some consider effectiveness to be close to an effect expected in the future patient population, although a clinical trial is still not the same as clinical practice for many reasons



Non-adherence to treatment ≠ missing data

- Data may be collected during the period of non-adherence if usable for the chosen estimand.
- Data is missing when usable data for an estimand of interest cannot be obtained for some subjects

  E.g.,
  - For *de jure* estimand (if all subjects adhered), impossible to obtain data for subjects actually not adhering
  - For *de facto* estimand, some subjects withdraw from study overall – planned "retrieved dropout" data unobserved
  - Data may be missing during adherence, but often less problematic (typically intermittent)
- Estimand
  - Specifies how treatment effect is defined for all subjects, including non-adhering subjects
  - Implies what data would be usable and should be collected during adherence and non-adherence
- Analysis method must deal with missing data in a manner consistent with the chosen estimand.

Missingness mechanism

- At analysis stage, we postulate assumptions about mechanism of missingness, including whether missingness can be considered independent of unobserved outcomes, conditional on observed data
  - Missing Completely at Random (MCAR)
  - Missing at Random (MAR)
  - Missing Not at Random (MNAR)

- Missingness mechanism can only be considered in the context of a specific analytical model
  – it depends on how observed factors are accounted by the model

Analysis methods

Research Article

Statistics in Medicine

Received 9 December 2014, Accepted 14 September 2015 Published online in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.6753

Sensitivity analysis for missing data in regulatory submissions

Thomas Permutt*†‡

The Prevention and Treatment of Missing Data in Clinical Trials

"The NRC report recommends modeling away the not-completely-at-randomness, so that the primary analysis treats missingness as missing at random (MAR) with respect to an explanatory model. They recognize limitations, however, and recommend sensitivity analyses …"

- We will start with MAR as a "base case" and then look at some MNAR scenarios as deviations from MAR

Missing at random (MAR)

- Probability of missingness is independent of unobserved outcomes given observed data

This implies:

- Subjects with missing data can be modeled based on similar subjects with available data if we account in the model for relevant observed factors.

Missing at random (MAR)

- Probability of missingness is independent of unobserved outcomes given observed data

This implies:

- Subjects with missing data can be modeled based on similar subjects with available data if we account in the model for relevant observed factors.

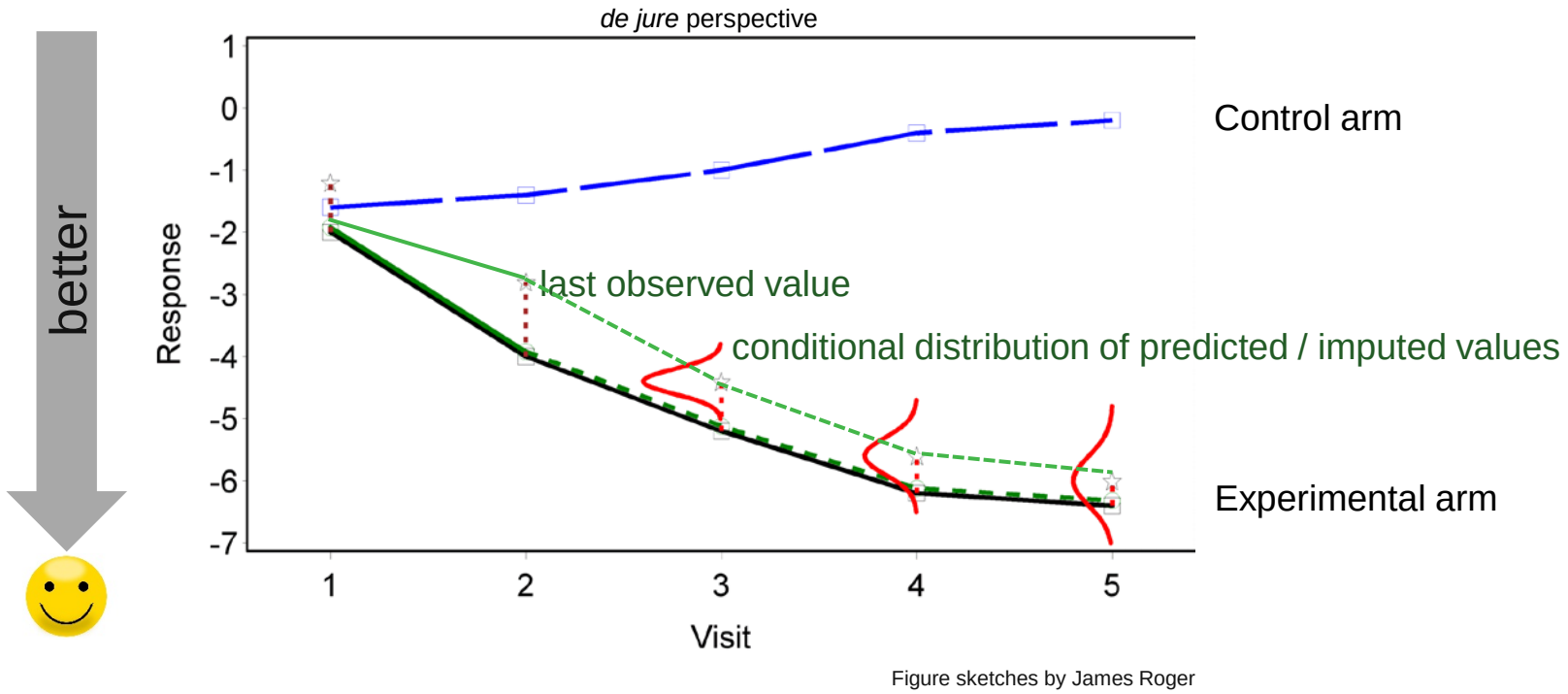


Figure sketches by James Roger

Methods assuming MAR

- Mixed Models with Repeated Measures (MMRM) assumes MAR implicitly without imputing missing values
- Standard Multiple Imputation (MI) imputes missing values explicitly under MAR

- MAR-based methods are often used to model missing data based on observed data of adhering subjects from the same treatment group, which produces a “best case” estimate

however

- MAR-based methods may also be used for a *de facto* estimand if:
  - Retrieved dropout data are collected and usable for the estimand,
  - MAR may be assumed for non-adhering subjects without planned “retrieved dropout data” - overall study withdrawals,
  - Missing data may be modeled using data from adhering subjects and/or retrieved dropout data from non-adherers

MAR need not imply *de jure* or “best case” estimand – depends on data utilized and the model

Missing not at random (MNAR)

- Probability of missingness depends on unobserved outcomes even after accounting for observed data

This implies:

- Subjects with missing data can NOT be modeled based on subjects with available data accounting only for observed data – additional assumptions are needed

Pattern-mixture models (PMM)

- Pattern-mixture models – a framework for MNAR data

- Y_{obs} – observed data
- Y_{mis} – missing data
- R – response/missingness indicators
- X - covariates

- Factor joint probability of observed data, missing data, and missingness into two components:

$$p(Y_{obs}, Y_{mis}, R|X) = p(R|X) \times p(Y_{obs}, Y_{mis}|R, X)$$

- Parameters of $p(Y_{obs}, Y_{mis} | R, X)$ cannot be estimated from the available data alone

Pattern-mixture models (PMM)

- Y_{obs} – observed data
- Y_{mis} – missing data
- R – response/missingness indicators
- X - covariates

- Factor joint probability of observed data, missing data, and missingness into two components:

$$p(Y_{obs}, Y_{mis}, R|X) = p(R|X) \times p(Y_{obs}, Y_{mis}|R, X)$$

- Parameters of $p(Y_{obs}, Y_{mis}|R, X)$ cannot be estimated from the available data alone
- Separate observed data distribution and a predictive distribution of missing data given observed data

$$p(Y_{obs}, Y_{mis}, R|X) = p(R|X) \times p(Y_{obs}|R, X) \times p(Y_{mis}|Y_{obs}, R, X)$$

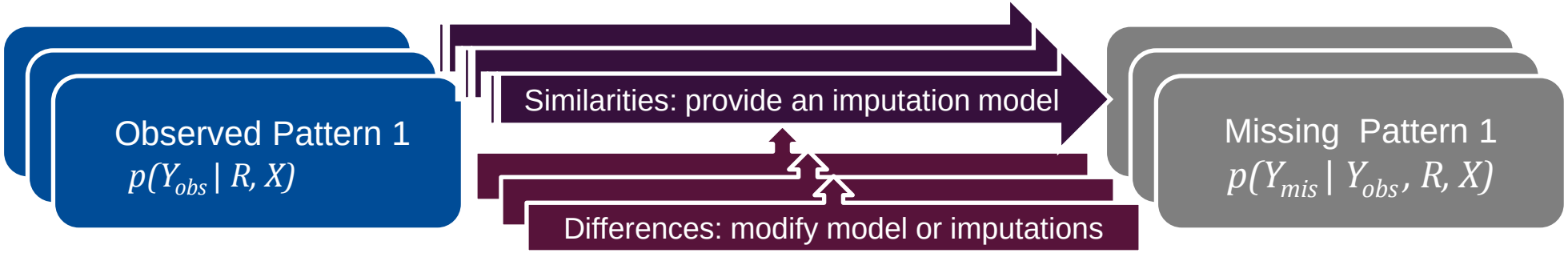
- Impose explicit restrictions/assumptions on $p(Y_{mis}|Y_{obs}, R, X)$

Pattern-mixture models – identifying restrictions

$$p(Y_{obs}, Y_{mis}, R|X) = p(R|X) \times p(Y_{obs}|R,X) \times p(Y_{mis}|Y_{obs}, R, X)$$

- Restrictions/assumptions on $p(Y_{mis} \mid Y_{obs}, R, X)$ can be expressed in terms of similarities and differences between $p(Y_{mis} \mid Y_{obs}, R, X)$ and $p(Y_{obs} \mid R, X)$ within different patterns R

- Multiple patterns can be defined and different links between them postulated



Thijs H, Molenberghs G. Strategies to fit pattern-mixture models. Biostatistics 2002; 3(2):245–265.

Multiple imputation (MI)

- MI (Rubin, 1987) is a principled method that accounts for uncertainty of imputations
- Imputation model can be different from analysis model
  - useful to assess sensitivity to assumptions behind imputation
- "Standard" MI operates under MAR, but MI can also be used under departures from MAR

Multiple imputation (MI)

- MI (Rubin, 1987) is a principled method that accounts for uncertainty of imputations
- Imputation model can be different from analysis model
  - useful to assess sensitivity to assumptions behind imputation

- "Standard" MI operates under MAR, but MI can also be used under departures from MAR

E.g., in the PMM framework:

- For each pattern with missing data, identify a pattern with observed data from which imputation model can be estimated
- Estimate (pattern-specific) imputation model(s) and use them to impute missing values, e.g.,:
  - Using predictions from the imputation model "as is"; or
  - Modify parameters of the imputation model according to specific postulated assumptions; or
  - Modify predictions from the imputation model according to specific postulated assumptions
- Analyze multiply-imputed data and combine results using Rubin's rule.

Departures from MAR

(Permutt, 2015):

"… *we think of unobserved data in de facto rather than de jure terms; that is, what might have been observed notwithstanding discontinuation of treatment, rather than what might have been observed if patients who discontinued had not done so.* "

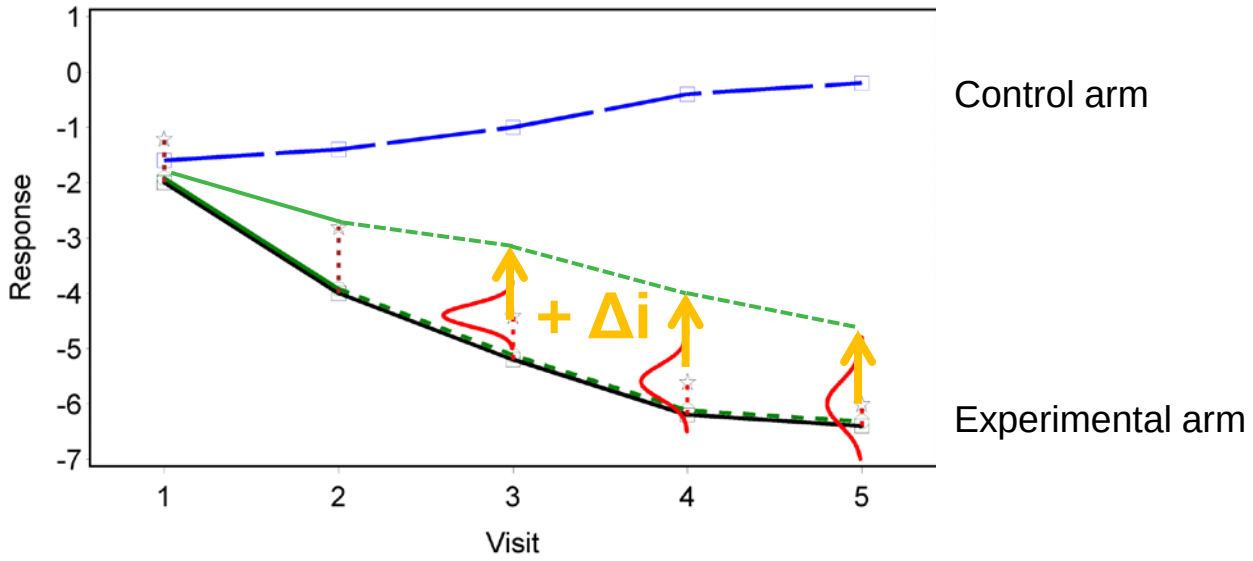
"… *If missing data are not like observed data, what matters is whether they are more on average or less on average than the observed data, in each treatment group.*

… Basically, the method is to predict the missing outcomes and then add values Δi *to the predictions in group i, varying the* Δi *over a plausible range.*

We think this is the most appropriate kind of sensitivity analysis for the missing data problem."

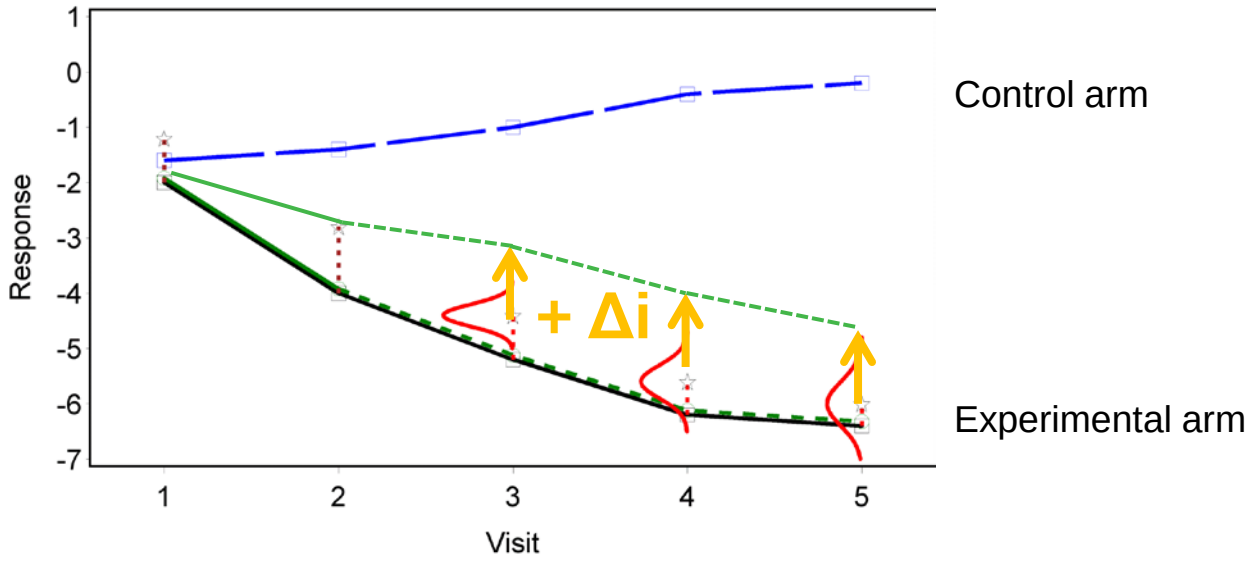
Delta-adjustment of imputations for continuous outcomes

- To investigate deviations from MAR, obtain imputations under MAR (e.g., standard multiple imputation) and adjust these values by Δi.



Delta-adjustment of imputations for continuous outcomes

- To investigate deviations from MAR, obtain imputations under MAR (e.g., standard multiple imputation) and adjust these values by Δi.



- This strategy follows PMM paradigm
- Delta may be pattern-specific, e.g.,
  - Delta may be applied to one treatment arm only, e.g., to experimental arm but not placebo
  - Delta may vary depending on reason for discontinuation

Delta-adjustment at the first visit vs. all visits

- Delta-adjustment only at the first visit after discontinuation
  - Recall: MNAR – missingness depends on unobserved outcome
  - Non-future dependence - idea: If we knew the value at the time of / immediately after discontinuation, the rest could be modeled based on that value, just like with MAR
  - Adjust imputed values only at the first visit after discontinuation, then use them as predictors for the following visit(s) without additional adjustment
  - Assumes that the correlations between visits are the same in completers and drop-outs after discontinuation and that the strength of these correlations in the MAR-based model will propagate a worsened value through predictions to future visits

Delta-adjustment at the first visit vs. all visits

- Delta-adjustment at all visits after discontinuation
  - Appropriate when correlations between visits estimated from completers cannot be assumed to adequately model continuing worsening after treatment discontinuation
  - Variant 1: Impute all visits under MAR first, then apply delta adjustments at each visit
  - Variant 2: Impute and delta-adjust one visit at a time; use adjusted values as predictors for next visit.
    - Produces a cumulative adjustment effect; larger adjustment at the last visit compared to Variant 1
  - Magnitude of delta may be visit-specific or may depend on timing of discontinuation

Tipping point analysis

(Permutt, 2015):

> *"Tests of the hypothesis of no treatment effect should be carried out, we suggest, for a range of deviations ($\Delta 1, \Delta 2$) from missingness at random, whatever the model is with respect to which missing at random is assumed in the primary analysis.*
>
> …
>
> *The decision problem then becomes a matter of judgment about the plausibility of values of ($\Delta 1, \Delta 2$) outside the range of significance."*

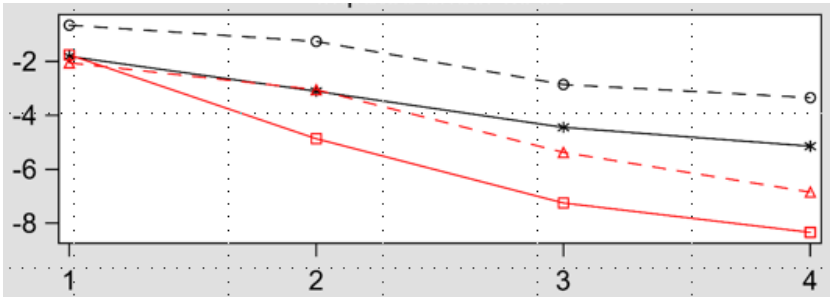
- Tipping point: settings of *($\Delta 1, \Delta 2$)* on the divide between significant and nonsignificant results.

Example - Major Depressive Disorder

Mallinckrodt et al. (2013), ***Missing Data: Turning Guidance Into Action***. Stat. in Biopharm. Research, 5:4, 369-382

Change from Baseline to Week 8 in HAM-D-17 Total Score		
Analysis	LS Mean Difference of Experimental vs. Placebo (SE)	p-value
Standard MI (MAR-based)	-2.54 (1.12)	0.024
Delta=2 – first visit	-2.38 (1.04)	0.022
Delta=2 – all visit	-2.02 (1.05)	0.054

Mean Change from Baseline
Imputation under MAR / de jure



- Completion rates: 76% for experimental and 74% for placebo
- Treatment difference often assumed at design stage is 2 points
- Change of 1/2 SD (~2 points) in HAM-D-17 total score is considered meaningful for individual subject

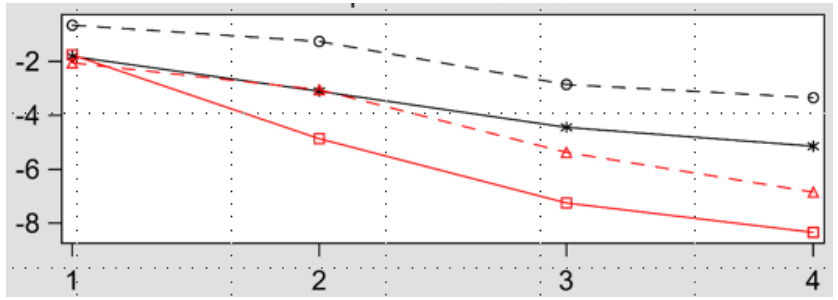
Example - Major Depressive Disorder

Mallinckrodt et al. (2013), ***Missing Data: Turning Guidance Into Action***. Stat. in Biopharm. Research, 5:4, 369-382

Change from Baseline to Week 8 in HAM-D-17 Total Score		
Analysis	LS Mean Difference of Experimental vs. Placebo (SE)	p-value
Standard MI (MAR-based)	-2.54 (1.12)	0.024
Delta=2 – first visit	-2.38 (1.04)	0.022
Delta=2 – all visit	-2.02 (1.05)	0.054

- Completion rates: 76% for experimental and 74% for placebo
- Treatment difference often assumed at design stage is 2 points
- Change of 1/2 SD (~2 points) in HAM-D-17 total score is considered meaningful for individual subject
- Tipping point (based on loss of statistical significance) is reached at δ=2 applied sequentially to all visits after discontinuation in experimental arm
  - Is it plausible?

Mean Change from Baseline
Imputation under MAR / de jure



Imputation under MAR + delta at all visits

Example - Major Depressive Disorder

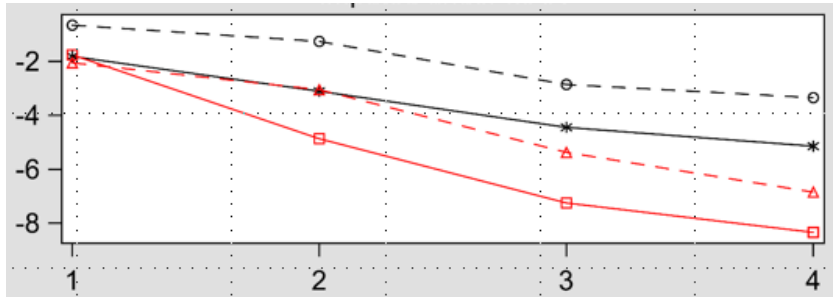
Mallinckrodt et al. (2013), ***Missing Data: Turning Guidance Into Action***. Stat. in Biopharm. Research, 5:4, 369-382

Change from Baseline to Week 8 in HAM-D-17 Total Score		
Analysis	LS Mean Difference of Experimental vs. Placebo (SE)	p-value
Standard MI (MAR-based)	-2.54 (1.12)	0.024
Delta=2 – first visit	-2.38 (1.04)	0.022
Delta=2 – all visit	-2.02 (1.05)	0.054

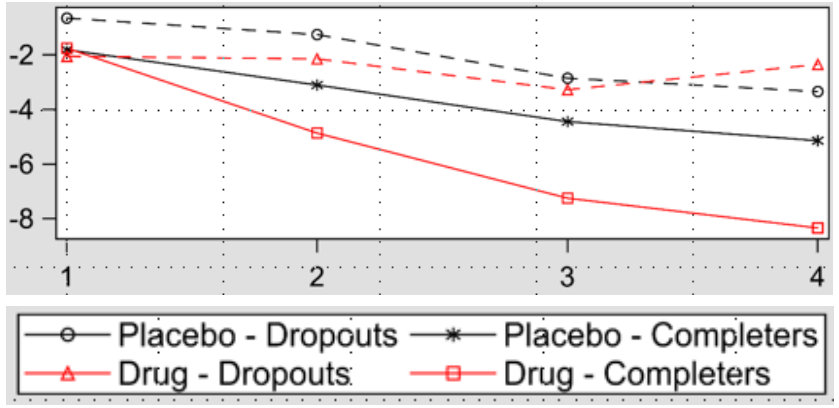
- Completion rates: 76% for experimental and 74% for placebo
- Treatment difference often assumed at design stage is 2 points
- Change of 1/2 SD (~2 points) in HAM-D-17 total score is considered meaningful for individual subject
- Tipping point (based on loss of statistical significance) is reached at δ=2 applied sequentially to all visits after discontinuation in experimental arm
  - Is it plausible?

    Experimental drug dropouts would need to have worse outcomes than placebo dropouts to overturn primary analysis conclusions…

Mean Change from Baseline
Imputation under MAR / de jure



Imputation under MAR + delta at all visits



Delta-adjustment for various types of endpoints

- Binary (responder): δ can represent the odds ratio of response for completers versus withdrawals
  - Multiple imputation model: logistic regression
  - Delta-adjustment: impute response with $\Pr(response) = \frac{e^{\hat{\alpha}+x'\hat{\beta}+\delta^*}}{e^{\hat{\alpha}+x'\hat{\beta}+\delta^*}+1}$, $\delta^* = log\frac{1}{\delta}$

Delta-adjustment for various types of endpoints

- Binary (responder): δ can represent the odds ratio of response for completers versus withdrawals
  - Multiple imputation model: logistic regression
  - Delta-adjustment: impute response with $\Pr(response) = \frac{e^{\hat{\alpha}+x'\hat{\beta}+\delta^*}}{e^{\hat{\alpha}+x'\hat{\beta}+\delta^*}+1}, \quad \delta^* = log\frac{1}{\delta}$
- Time-to-event: δ can represent the ratio of subject-specific hazard at any given time point *t* following withdrawal compared to that same subject's hazard at the same time t if s/he had continued the study
  - Multiple imputation model: estimated survival function $\hat{S}(t|\boldsymbol{x}_i, \hat{\boldsymbol{\beta}})$ (piecewise exponential, Cox regression, Kaplan-Meier, piecewise logistic)
  - Delta-adjusted imputed time of event: solution for t from $u_i = 1 - \hat{S}(t|\boldsymbol{x}_i, \hat{\boldsymbol{\beta}})^{\boldsymbol{\delta}}$, u_i ~ uniform $[1 - \hat{S}(c_i|\boldsymbol{x}_i, \hat{\boldsymbol{\beta}})^{\boldsymbol{\delta}}, 1]$

Lipkovich I, Ratitch B, O'Kelly M (2016) Sensitivity to censored-at-random assumption in the analysis of time-to-event endpoints. Pharmaceut. Statist., 15: 216–229

Delta-adjustment for various types of endpoints

- Binary (responder): δ can represent the odds ratio of response for completers versus withdrawals
  - Multiple imputation model: logistic regression
  - Delta-adjustment: impute response with $\Pr(response) = \frac{e^{\hat{\alpha}+x'\hat{\beta}+\delta^*}}{e^{\hat{\alpha}+x'\hat{\beta}+\delta^*}+1}, \qquad \delta^* = log\frac{1}{\delta}$
- Time-to-event: δ can represent the ratio of subject-specific hazard at any given time point *t* following withdrawal compared to that same subject's hazard at the same time t if s/he had continued the study
  - Multiple imputation model: estimated survival function $\hat{S}(t|\boldsymbol{x}_i, \widehat{\boldsymbol{\beta}})$ (piecewise exponential, Cox regression, Kaplan-Meier, piecewise logistic)
  - Delta-adjusted imputed time of event: solution for t from $u_i = 1 - \hat{S}(t|\boldsymbol{x}_i, \widehat{\boldsymbol{\beta}})^{\delta}$, $u_i \sim$ uniform $[1 - \hat{S}(c_i|\boldsymbol{x}_i, \widehat{\boldsymbol{\beta}})^{\delta}, 1]$
    - Lipkovich I, Ratitch B, O'Kelly M (2016) Sensitivity to censored-at-random assumption in the analysis of time-to-event endpoints. Pharmaceut. Statist., 15: 216–229
- Recurrent events: δ can represent a multiplicative adjustment for expected number/rate of events for withdrawals compared to completers
  - Multiple imputation model: Poisson, negative binomial

    Keene ON, Roger JH, Hartley BF, Kenward MG (2014) Missing data sensitivity analysis for recurrent event data using controlled imputation. Pharmaceut. Statist., 13: 258–264

Reference-based imputations

(Permutt, 2015):

> *"… we think the plausible values to consider for unobserved data in the active treatment group may be values near observed values in the placebo group, not observed values in the active group."*

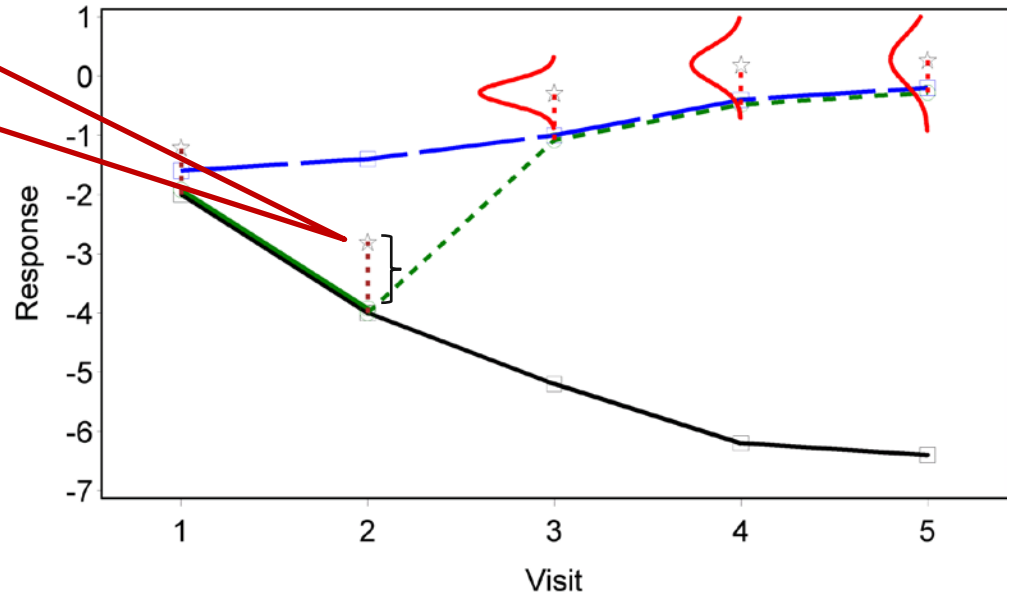
- This approach can be implemented by imputing missing data with an appropriate choice of delta or using placebo/reference-based imputation methods, where imputation model is estimated from the placebo arm.

Reference-based imputation strategies

- Main idea: for subjects with missing data in the experimental arm, their mean response distribution is assumed to be that of a `reference' (e.g., control arm) group of subjects
- Variants differ depending on how subject's previous (pre-withdrawal) outcomes are accounted for:
  - Via subject's difference from the mean of his/her own treatment arm before withdrawal
  - Via subject's difference from the mean of the reference group before withdrawal

Jump to reference (J2R)

This subject will be determined to be worse than mean of his own treatment group based on available data



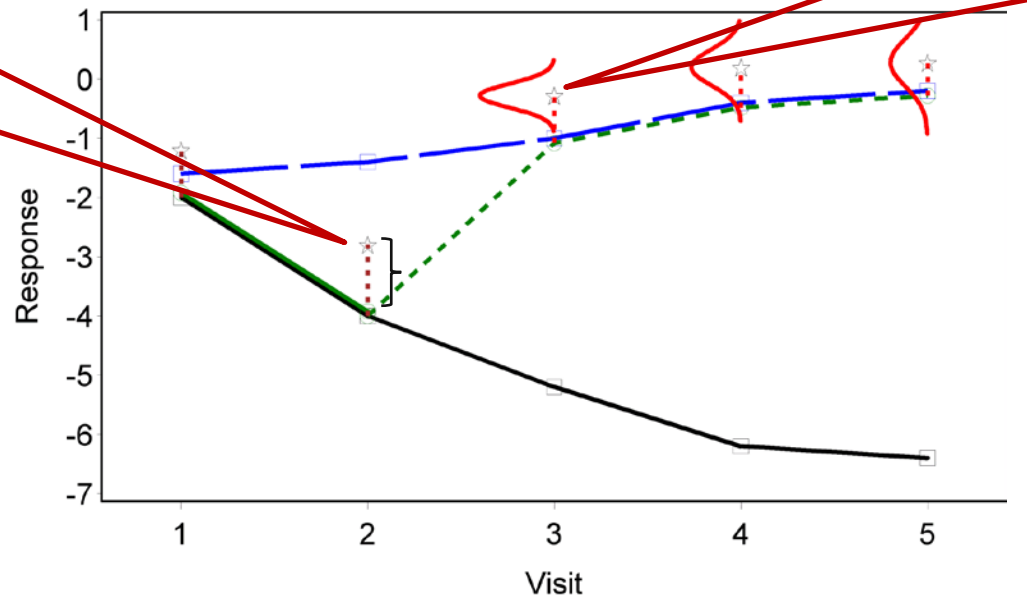
Reference group

Experimental arm

Jump to reference (J2R)

This subject will be determined to be worse than mean of his own treatment group based on available data

Then this subject will be predicted as worse than mean of the reference group for future values

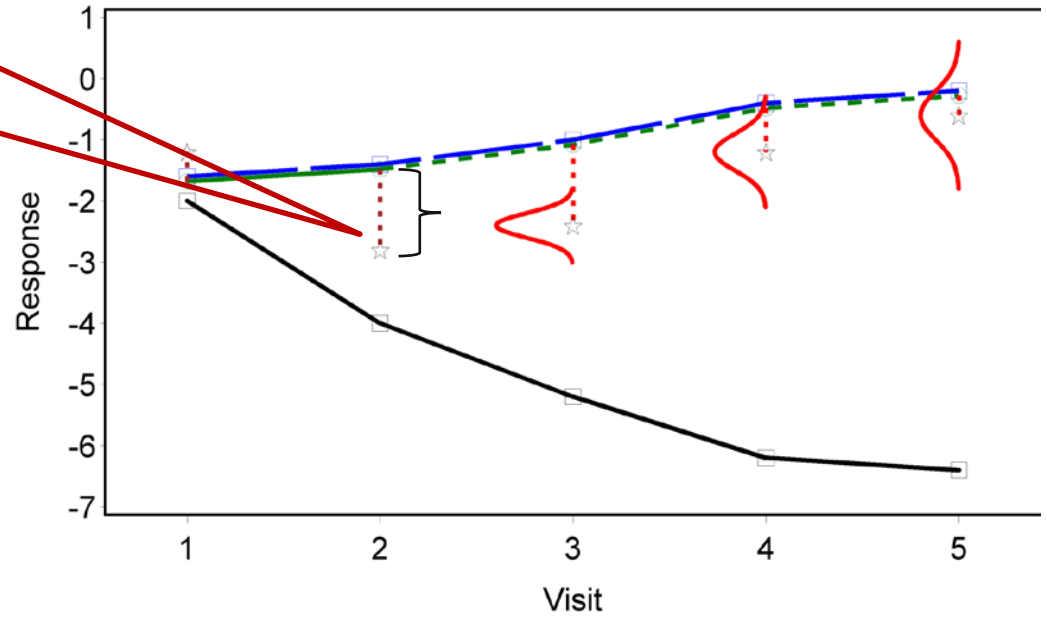


Reference group

Experimental arm

Copy reference (CR)

This subject will be determined to be better than mean of the reference group based on available data



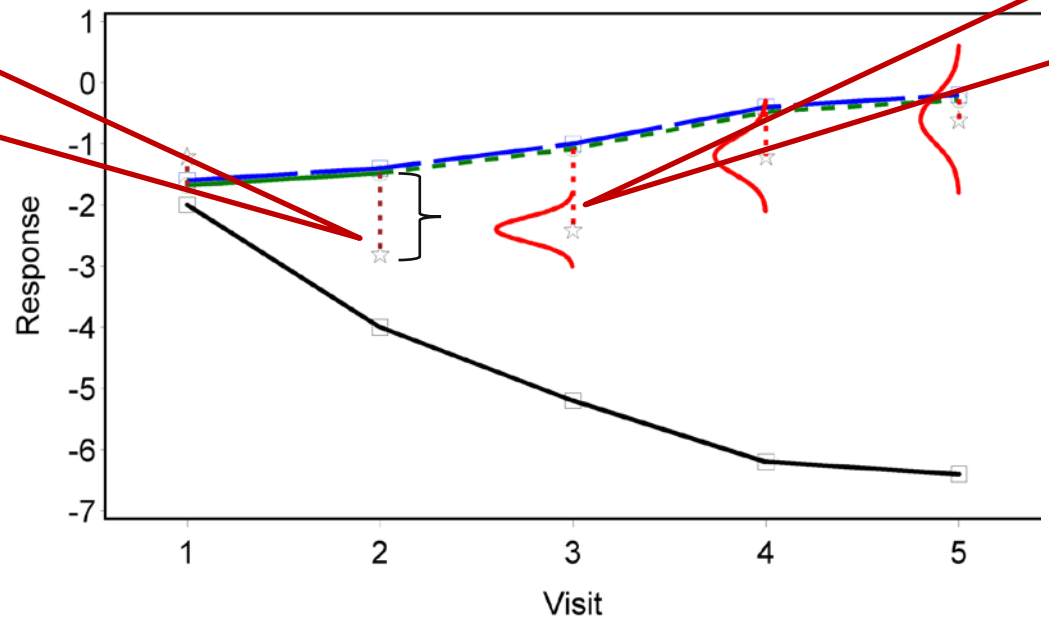
Reference group

Experimental arm

Copy reference (CR)

This subject will be determined to be better than mean of the reference group based on available data

Then this subject will be predicted as better than mean of the reference group for future values



Reference group

Experimental arm

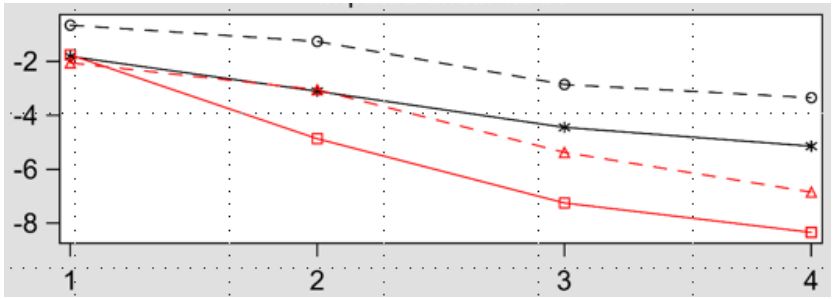
Example - Major Depressive Disorder

Mallinckrodt et al. (2013), ***Missing Data: Turning Guidance Into Action***. Stat. in Biopharm. Research, 5:4, 369-382

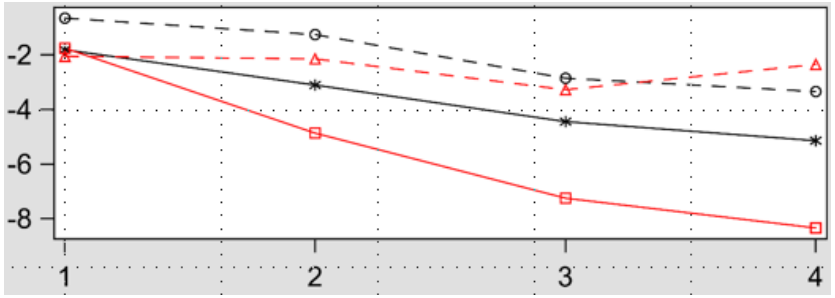
Change from Baseline to Week 8 in HAM-D-17 Total Score		
Analysis	LS Mean Difference of Experimental vs. Placebo (SE)	p-value
Standard MI (MAR-based)	-2.54 (1.12)	0.024
Delta=2 – first visit	-2.38 (1.04)	0.022
Delta=2 – all visit	-2.02 (1.05)	0.054
CR	-2.20 (0.99)	0.028
J2R	-1.98 (1.01)	0.051

- CR remains statistically and clinically significant
- J2R results in marginal loss of significance

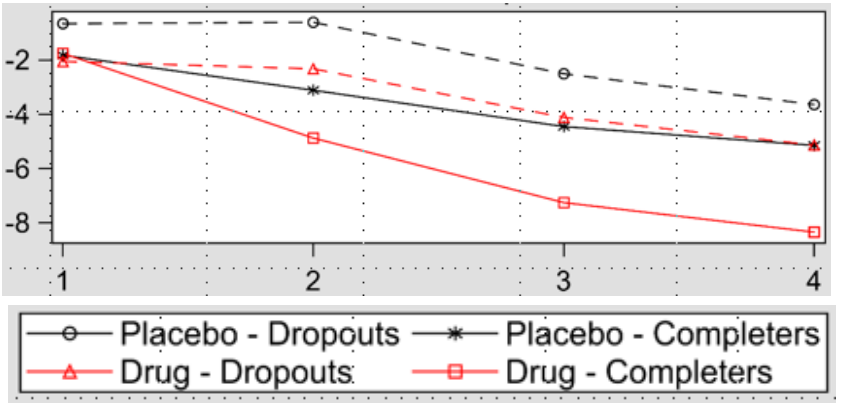
Mean Change from Baseline
Imputation under MAR / de jure



Imputation under MAR + delta at all visits



Imputation under CR



QuintilesIMS™

Reference-based imputations - caveats

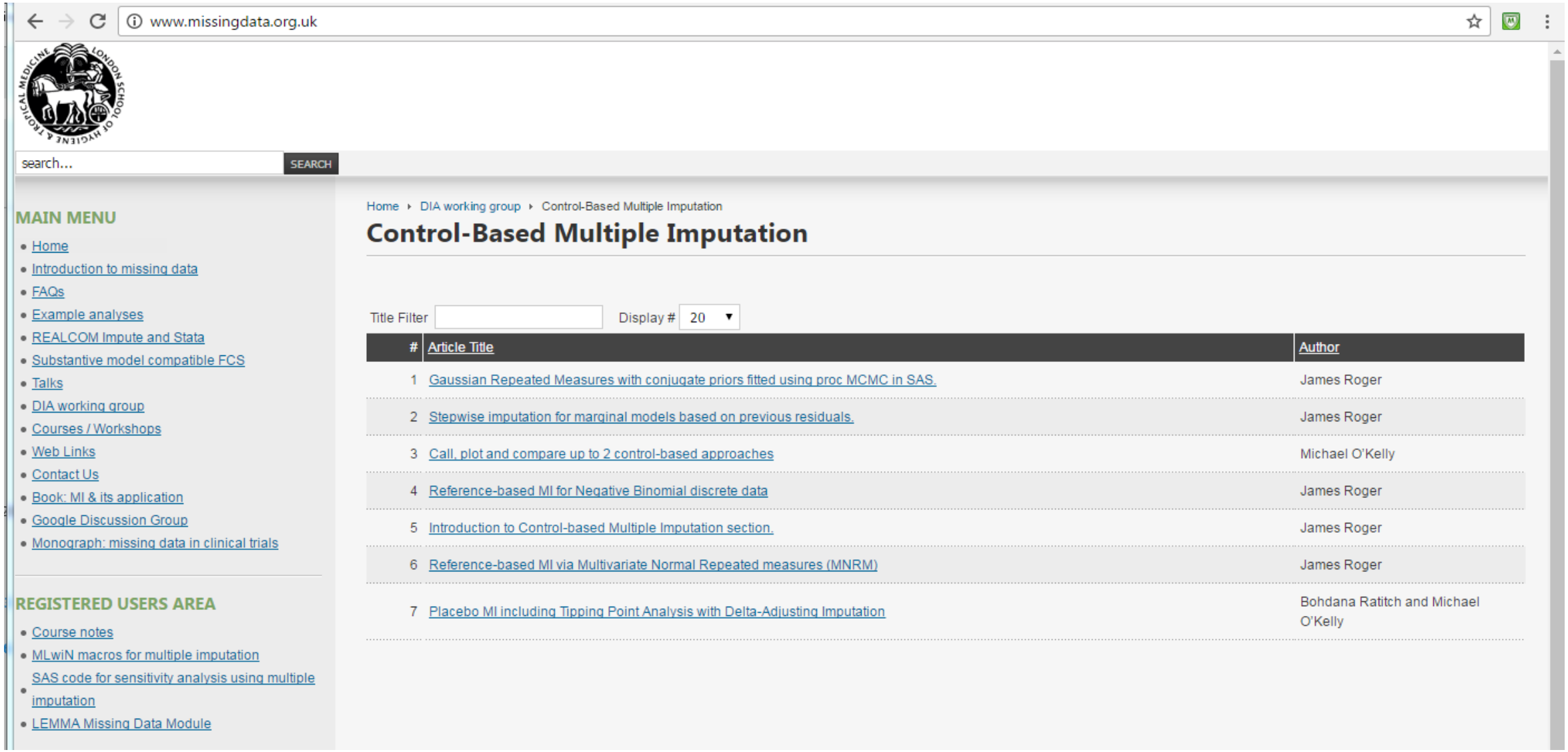
- This strategy is sometimes said to be “worst-case plausible”
- May be still too optimistic if reference group is an “elite” group – doing much better than any dropouts
- Appropriateness needs to be justified, given
  - Natural disease trajectory
  - Mechanism of action of experimental treatment and reference treatment
  - Reasons for non-adherence in all arms

Implementation tools

- SAS
  - PROC MI, new MNAR statement (v9.4):
    - delta adjustment for continuous and categorical outcomes
    - CR
    - J2R – with additional user programming for sequential regression MI on residuals
    - Also (with additional programming), MI variants of BOCF and LOCF
  - With additional programming:
    - PROC MCMC – substantial additional programming required
      - Macros by James Roger are available on www.missingdata.org.uk
    - PROC PHREG with BAYES statement for time to event with piecewise exponential survival imputation model
    - PROC LIFETEST or PROC PHREG with bootstrap for time to event with Kaplan-Meier or Cox regressions imputation model
    - PROC GENMOD with BAYES statement for count data

- www.missingdata.org.uk
  - Code, macros
  - Manuscripts
  - Training materials
  - Template SAP text

Implementation tools



Summary and Discussion

- Crucial to start with a clearly defined estimand and determination of what constitutes usable data for non-adherers.
- When usable data cannot be collected, multiple imputation in combination with clearly interpretable imputation strategies is a valuable tool in analysis of missing data under MAR and MNAR mechanisms
- Other strategies available, e.g., selection models, shared parameter models, MMRM with explicitly adjusted LS means, etc.
- Still facing difficult decisions…
  - MNAR is criticized as non-verifiable and thus potentially not suitable for the primary analysis…
  - MAR in a de jure sense is criticized as farfetched and not relevant…
  - “de facto” is criticized when confounded with the effect of post-discontinuation alternative treatments and/or rescue…
  - How to protect against “false negatives” when interpreting multiple sensitivity analyses?

References

- National Research Council. (2010). The Prevention and Treatment of Missing Data in Clinical Trials. Panel on Handling Missing Data in Clinical Trials. Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.
- European Medicines Agency. Guideline on Missing Data in Confirmatory Clinical Trials. EMA/CPMP/EWP/1776/99 Rev. 1. 2 July 2010.
- LaVange LM, Permutt T. (2015). A regulatory perspective on missing data in the aftermath of the NRC report. Statistics in Medicine (on-line at wileyonlinelibrary.com DOI: 10.1002/sim.6840)
- Permutt T. (2015a). A taxonomy of estimands for regulatory clinical trials with discontinuations. Statistics in Medicine (on-line at wileyonlinelibrary.com DOI: 10.1002/sim.6841)
- Permutt T. (2015b). Sensitivity analysis for missing data in regulatory submissions. Statistics in Medicine (on-line at wileyonlinelibrary.com DOI: 10.1002/sim.6841)
- Rubin DB. Multiple imputation for nonresponse in surveys. John Wiley and Sons, Inc.: New York, 1987.
- Mallinckrodt C, Roger J, Chuang-Stein C, Molenberghs G, Lane PW, O'Kelly M, Ratitch B, Xu L, Gilbert S, Mehrotra DV, Wolfinger R, Thijs H. (2013) Missing Data: Turning Guidance Into Action, Statistics in Biopharmaceutical Research, 5:4, 369-382
- Carpenter JR, Roger JH, Kenward MG. (2013). Analysis of longitudinal trials with protocol deviation: a framework for relevant, accessible assumptions and inference via multiple imputation. Journal of Biopharmaceutical Statistics. 23(6):1352-1371.
- Ratitch B, O'Kelly M, Tosiello R. (2013) , Missing data in clinical trials: from clinical assumptions to statistical analysis using pattern mixture models. Pharmaceutical Statistics, Volume 12, Issue 6, Pages 337-347
- Keene ON, Roger JH, Hartley BF, Kenward MG (2014) Missing data sensitivity analysis for recurrent event data using controlled imputation. Pharmaceut. Statist., 13: 258–264
- Lipkovich I, Ratitch B, O'Kelly M (2016) Sensitivity to censored-at-random assumption in the analysis of time-to-event endpoints. Pharmaceut. Statist., 15: 216–229
- www.missingdata.org.uk