Even a “treatment policy” estimand may have missing data.

How can we take account of this in practice?
A report on work in progress.

Michael O’Kelly and Sylvia Li

Scope of Presentation

+ The “treatment policy” approach under the new estimands guideline ICH E9 (R1).
  - …and its missing data.
  - Problems: sparsity, bias, variance.
+ Case study: trial data in major depressive disorder (MDD).
+ Some existing options for taking account of the missing data.
+ A selected set of models for missingness under the “treatment policy” estimand.
+ Simulation set up and some results
+ Findings/Limitations/Discussions – not prescriptive – still investigating sparsity problems and potential bias when implementing “treatment policy”

Acknowledgements

- DIA Scientific Working Group for Estimands and Missing Data
  - James Bell (Elderbrook Solutions, Boehringer Ingleheim) and
  - James Roger (Livedata, LSHTM)
- Elena Polverejan (Janssen R&D)

“Treatment policy” approach, often requested by regulators

Reason for discontinuation?

- ICH E9 (R1) suggests that *where these is evidence for a reason,* subject discontinuation of study could be well handled by a compound approach (e.g. assigning a poor outcome if reason for discontinuing study reflects badly on the study treatment). So pure “treatment policy” may not be favoured by regulators.
  - This topic addressed by David Wright in earlier presentation
- But if there is not good evidence regarding reason for discontinuation, ICH E9 (R1) has a suggestion.

ICH E9 (R1) suggestion for the “treatment policy” approach

- ICH E9 (R1), “Section A.5.1. Main estimation”, p. 16

> “For example, for subjects who discontinue treatment without further data being collected, a model may use data from other subjects who discontinued treatment but for whom data collection has continued.”

ICH E9 (R1) suggestion for the “treatment policy” approach

- Recommends modelling missings on the distribution of a particular set of data.
  - i.e. ICH E9 (R1) is recommending a form of “reference-based” inference – could be implemented e.g. via reference-based multiple imputation (MI).
    - => MI outcomes of those who discontinue the study, basing the model on subjects who discontinued the **treatment**, but did not discontinue the **study**.
      - “Off-treatment only” approach described in earlier talk by James Bell.
- Thus, for regulatory purposes, ICH E9 (R1) sanctions a particular assumption about missing outcomes – we are allowed to take a distribution as “true” for the missing data.
  - Our task becomes, not justifying a modelled outcome as credible, but finding a way to model a distribution as best we can.
  - But there are (of course) problems, as outlined by James Bell earlier.

Problems with modelling missings based on subset

Sparsity!..

- …and, because of sparsity, potential bias/variance issues in model estimates.
- See James Bell's formula to estimate impact of "off-treatment only" approach on the standard error of potential estimates of treatment effect, when usual methods are applied.

Scope of this presentation

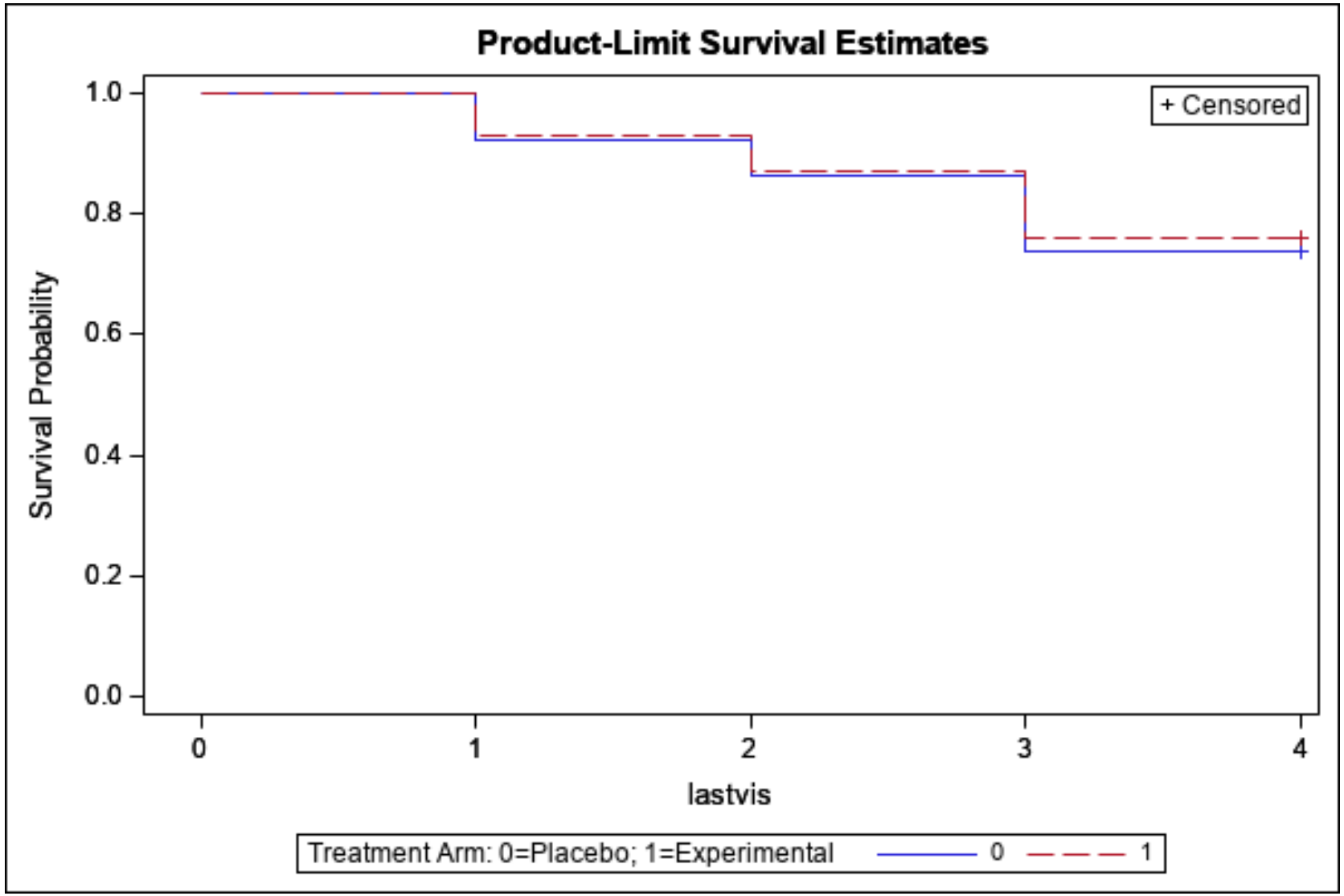
- A selection of modelling approaches for ICH E9 (R1)'s suggestion, exploring some of the questions posed by James Bell.
- Emphasis on practical concerns (e.g. convergence issues and standard error), with preliminary notes on bias.

Overview of case study

+ A case selected as likely to throw up issues and problems of interest
  - -moderate size, so sparsity problems expected
+ Trial in major depressive disorder (DIA Estimands and Missing Data group)*
  - 172 subjects, 43 stopped treatment early.
  - Primary efficacy endpoint: HAMDTL17 depression score
  - Baseline and 4 post baseline visits
  - Pooled site ID available, but not used in this experiment.
+ 23/88 in the control group discontinued treatment – 26%.
  - 7, 5, and 11 discontinued after Visit 1, 2, 3, respectively.
+ 20/84 in the experimental arm discontinued treatment – 24%.
  - 6, 5, and 9 discontinued after Visit 1, 2, 3, respectively.

*available at https://www.lshtm.ac.uk/research/centres-projects-groups/missing-data, data donated by Eli Lilly

Kaplan-Meier plot of probability of discontinuing treatment



Trajectory of on-treatment completers

Likely some deterioration after stopping treatment

Possible trajectories of experimental arm, early discontinuation

Likely some deterioration after stopping treatment

?

?

?

Possible trajectories of placebo, early discontinuation

Likely some deterioration after stopping treatment

Often, we posit post withdrawal trajectories, with uncertainty

With ICH E9 (R1), for regulatory purposes:
we may take the distribution of the missing outcomes as known, based on the distribution of subjects who stopped study treatment but stayed in the study

No question marks needed

Simulation setup

- For each of 4 scenarios, created “Oracle” dataset, missing outcomes for each subject calculated as mean of 1000 multiply imputed datasets. P-value <0.0001 for all 4 scenarios.
- For each of the four scenarios, 500 instances of the trial created;
  - each instance with outcomes of random sample of proportion of the 43 treatment-discontinued subjects rendered missing, stratified by treatment group.
  - Only difference in simulated datasets: which discontinued subjects had missings.
  - Proportion rendered missing: 30%, 50%, 75%, 85% & 90%.
  - 29, 22, 11, 7 & 5 subjects to estimate the model for missing outcomes!
- Thus 500 instances * 4 scenarios * 5 proportions missing.
- If discontinuing study, subjects did so immediately on discontinuing treatment.

Problem in treatment policy missing data

A revisit after the exploratory tour in the case study

Sparsity

43 withdrawals, 2 treatment groups, 4 timepoints

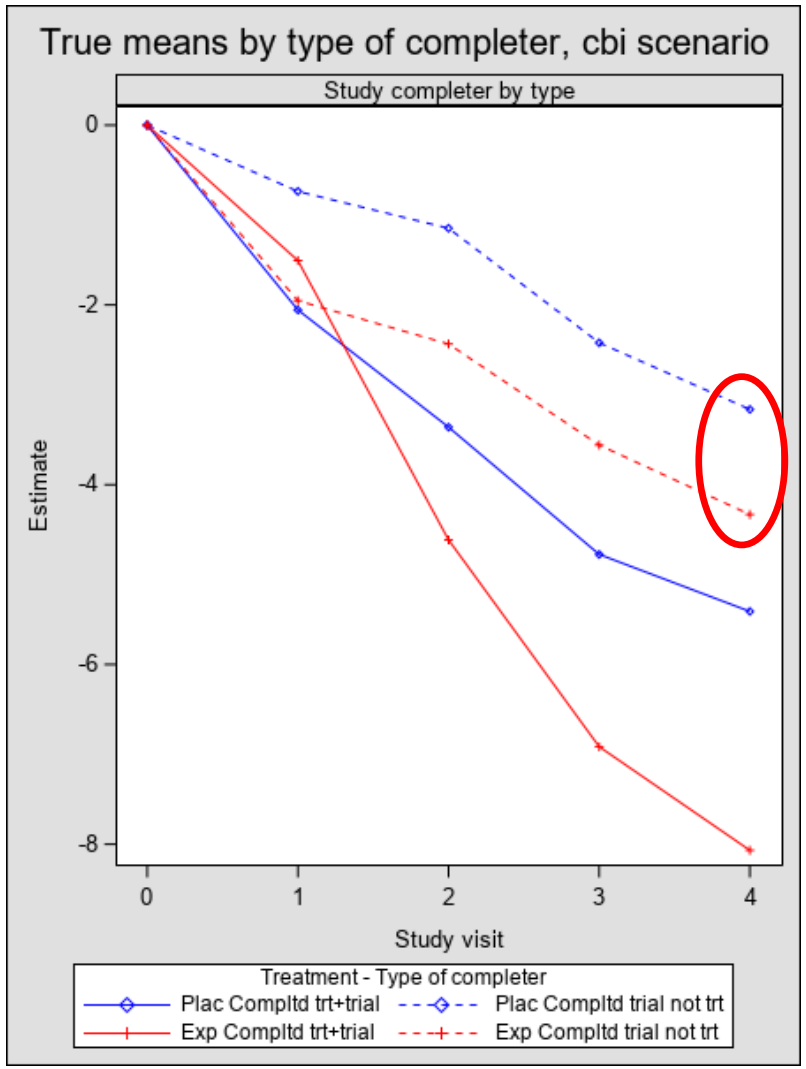
A “proper” model may not converge

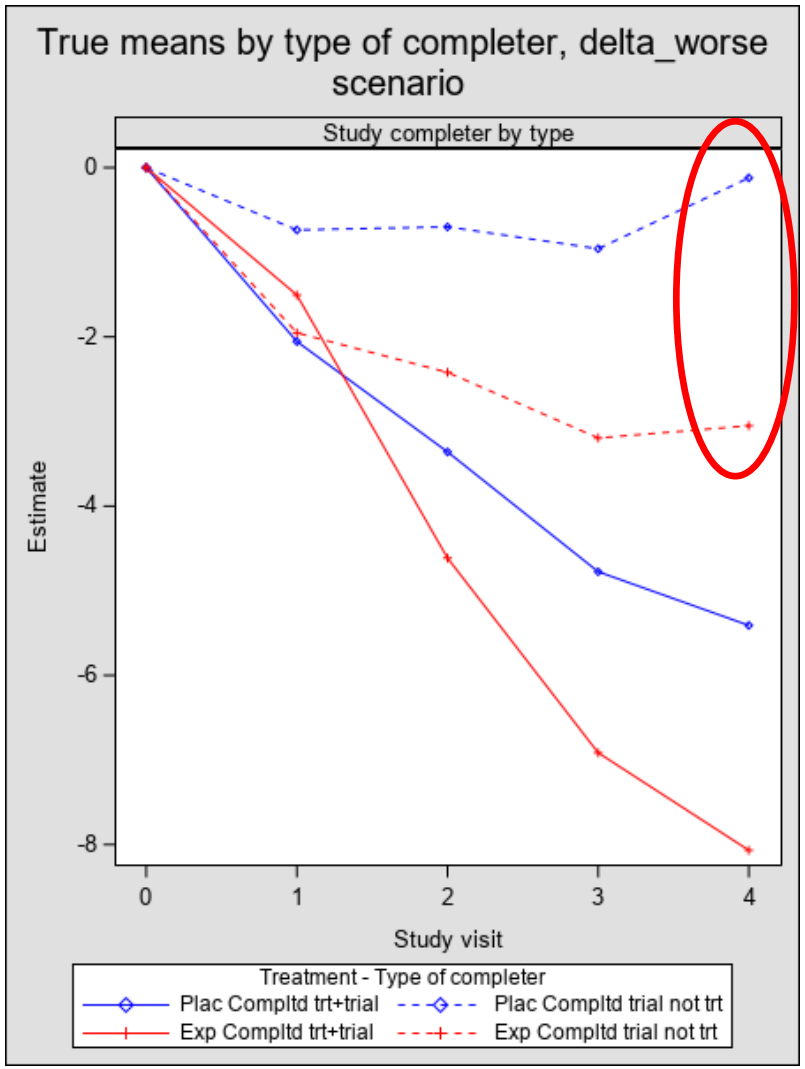
If model is reduced so as to converge it risks bias and/or high variance.

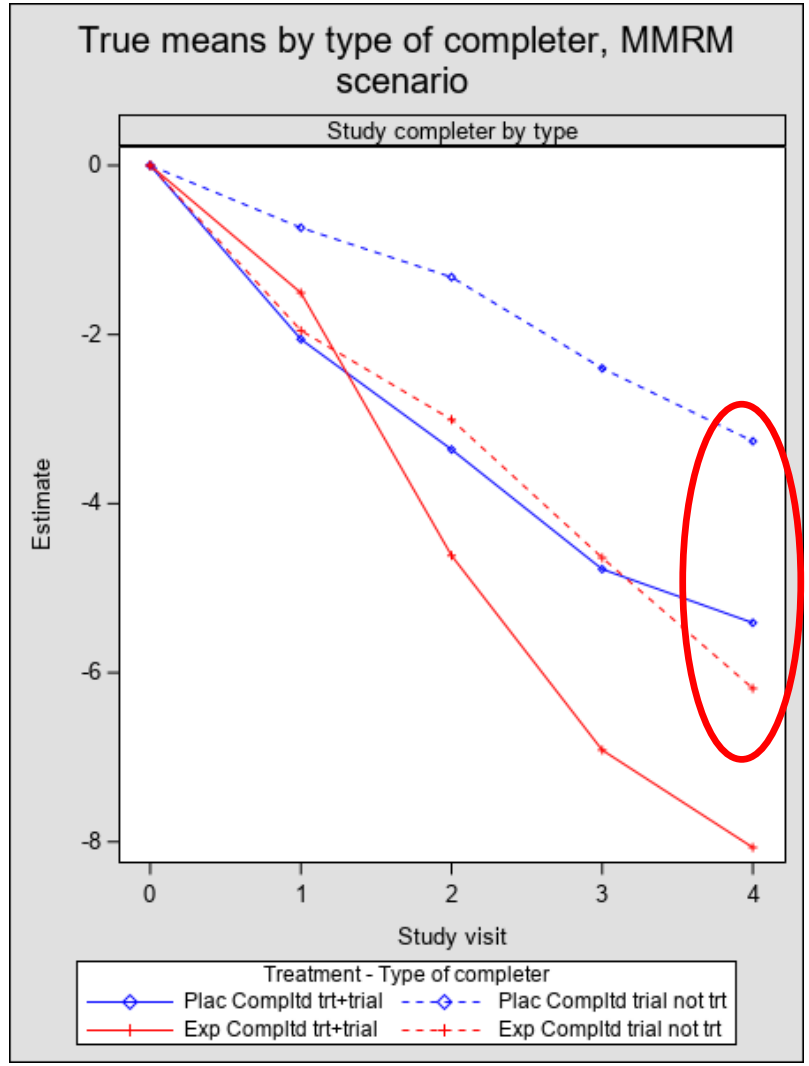
Scenarios

Trajectory of outcomes after treatment discontinuation	Short name	Treatment effect relative to MMRM
As for MMRM, i.e. ignoring treatment discontinuation	MMRM	
Copy reference – based on the placebo arm	CBI	Attenuated
Worsen vs MMRM by delta=3 each visit	Delta-worse	Increased
Trajectory drawn at random for each subject from the above	Random	

Three trajectories simulated for post-treatment outcomes: means by visit, completers vs. discontinued treatment.







Approaches investigated

Description	Short name	Model (SAS-speak)	Population to estimate model	Comment
"MAR" imputation but ignoring treatment discontinuation	MMRM	CFB(k) = baseline treat change(1)…change(k-1)	FAS	Missings a weighted average of completers and discontinuers
As above but add on/off treatment **indicator**	Standard (Std)	…add Ontreat(2)…Ontreat(k)	FAS	Requires a moderate N in discontinuers at each visit.
Model change from treatment discontinuation, time as linear	Off-treatment	CFLT = baseline treat **time CFBTL**	Subjects off treatment	Linear time covariate not usually accepted by regulators, but some leeway here.

MAR = missing at random; **MMRM**= mixed model for repeated measures; **CFB** = Change from baseline;
CFLT = Change from last treated visit; **CFBTL**=Change from baseline to last visit on treatment

How results were calculated

- Mean outcomes by treatment and their differences at the primary endpoint were estimated via LSMEANS.
  - Standard error was summarised at the mean standard error across the 500 instances.
  - Bias was calculated as difference vs. the “Oracle” LSMEANS estimate from the same model as that used for each of the 500 instances of the simulated trial.

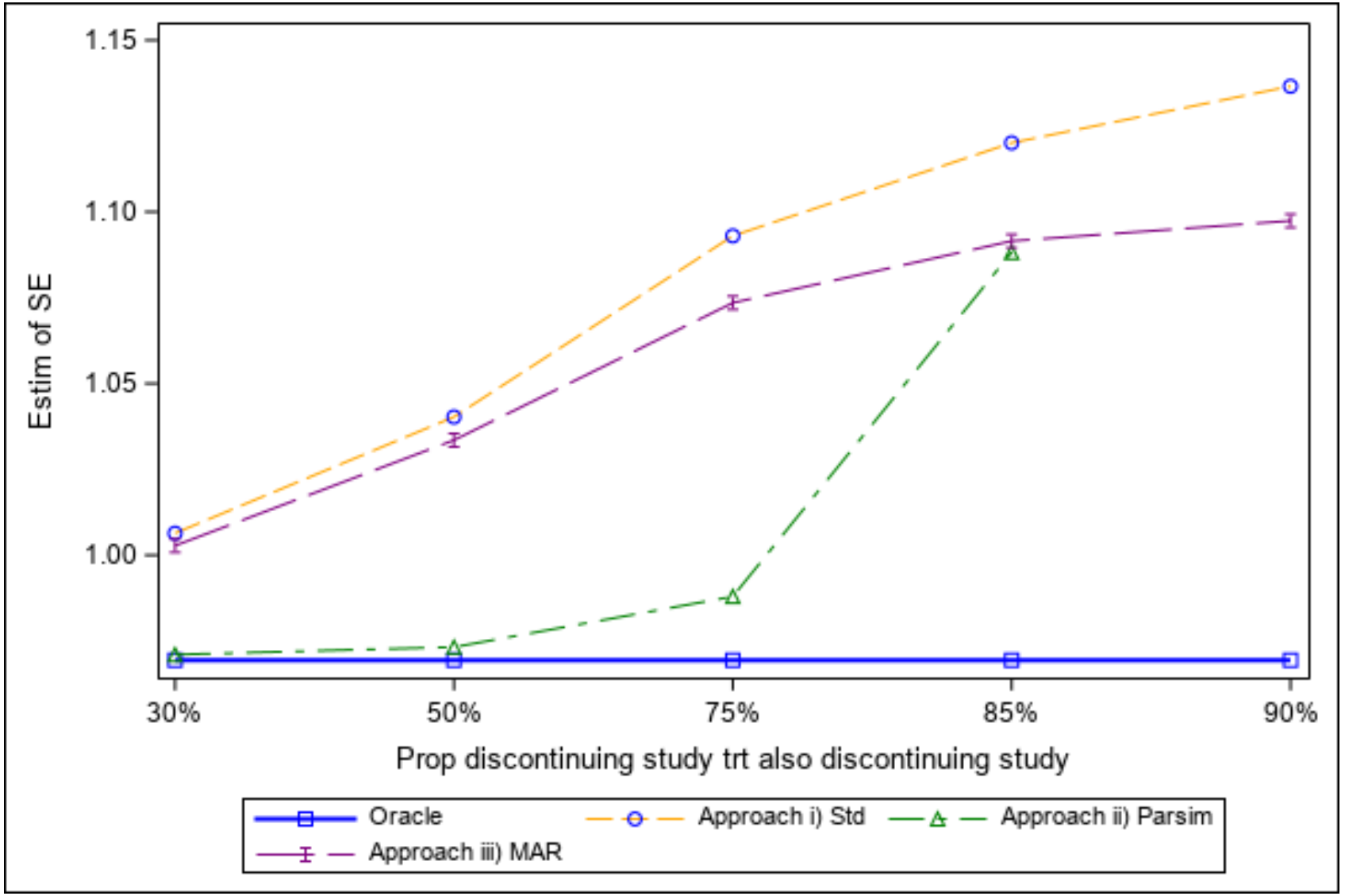
- Caveat: we do not as yet have a full understanding of the sources of bias in the four approaches; we are still researching the results.

Summary of results from the approaches

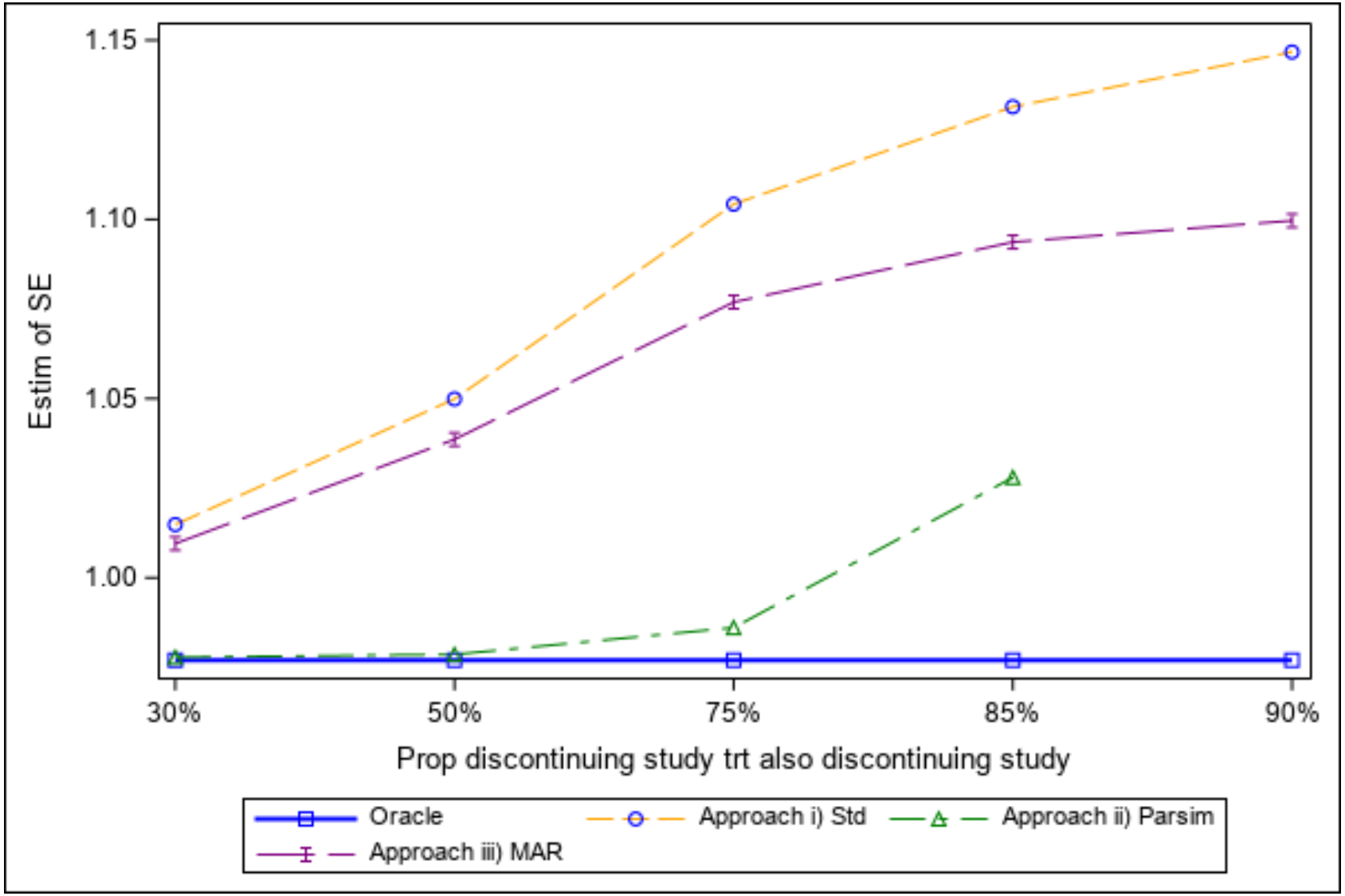
Short name	Feasible for sample moderate-sized trial (172 subjects)?	SE of estimate of treatment difference
MMRM	Yes	Fairly stable across the four scenarios
Standard (Std)	Up to 20% simulated trials had sparsity-related errors	Stable and slightly larger than MMRM
Off-treatment	Fails at <=5 subjects	Varied with scenario; for all scenarios, lower than MMRM and Std when >=11 subjects with post-treatment outcomes

Bias: Off-treatment approach appears to compare well with the other approaches when there are >= 11 observed post-treatment-discontinuation subjects; worse with <11 such subjects.

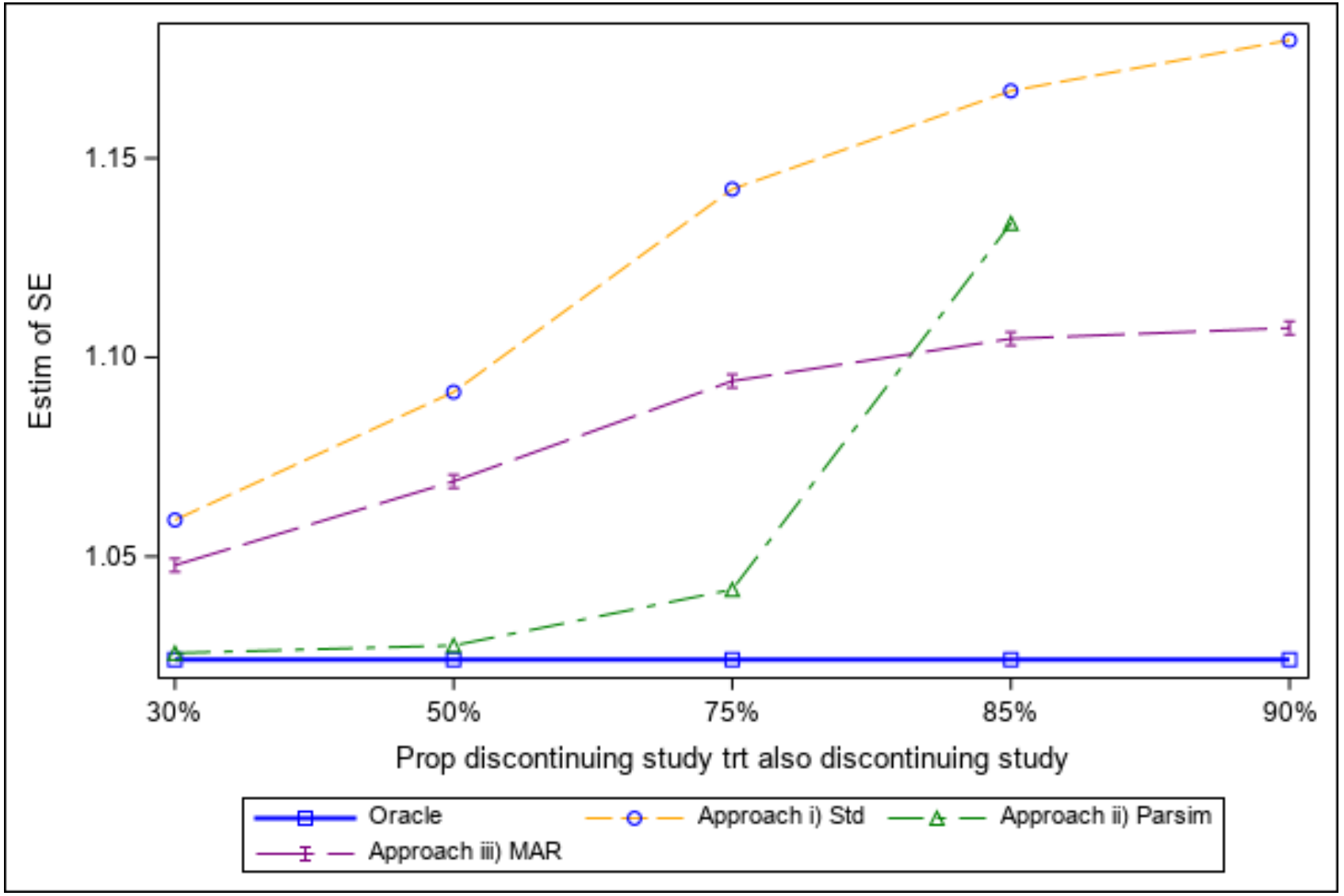
Standard error of treatment estimate, true post treatment outcomes MMRM



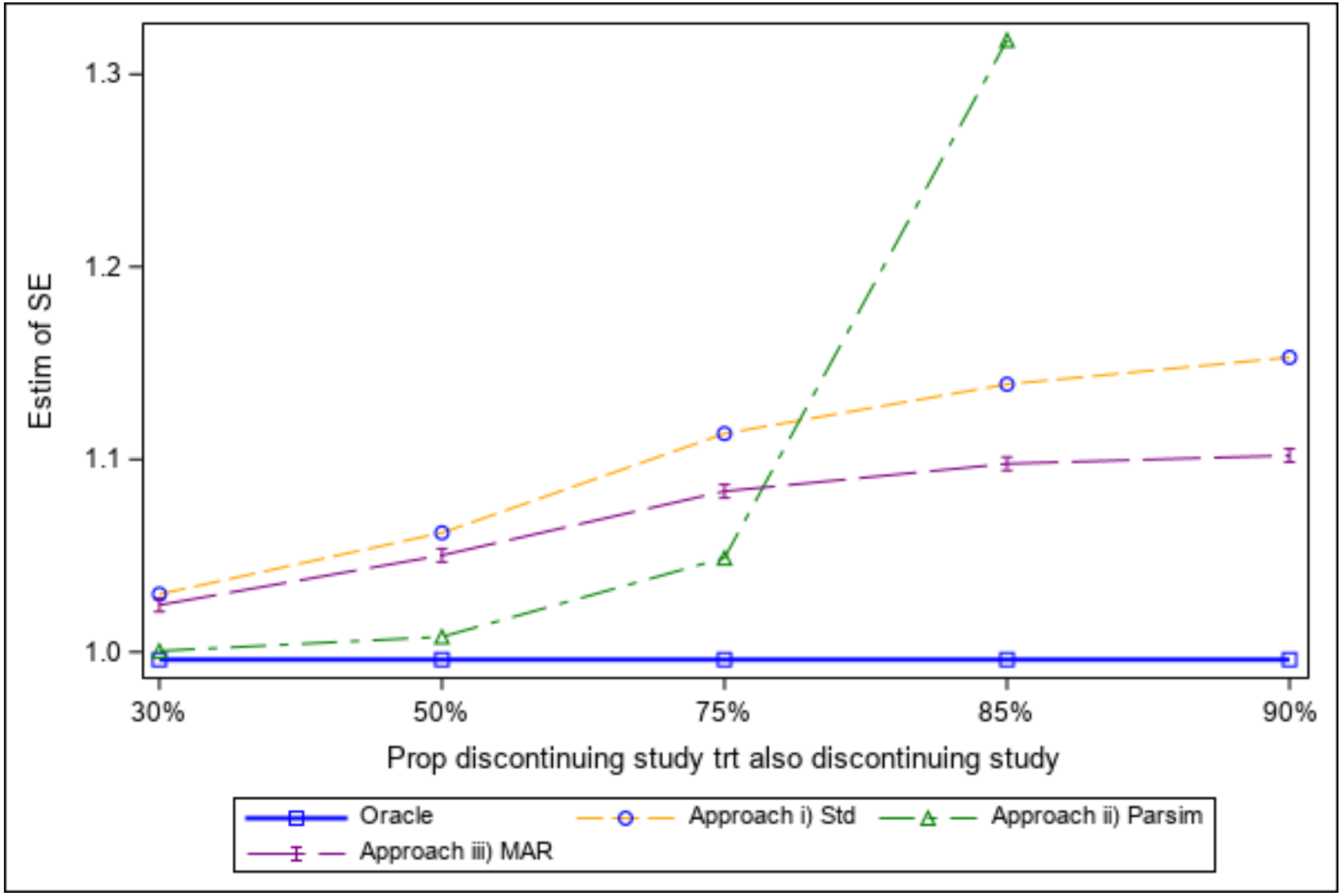
Standard error of treatment estimate, true post treatment outcomes all with distribution of control group



Standard error of treatment estimate, true post treatment outcomes worsened by delta=2 at each visit



Standard error of treatment estimate, true post treatment outcomes a mixture of MMRM, CBI and MMRM+delta



Limitations of the experiment

- Scenarios
  - Small trial (total N=172); one proportion of trt discontinuations (25%) explored.
  - Relatively short trial.
  - Continuous endpoint, only approximately normally distributed.
  - Distribution of outcomes missing in the source data simulated for this experiment – e.g. even in MMRM scenario, distribution after treatment discontinuation not guaranteed to be identical to that of the observed outcomes.
  - Patterns/covariates predictive of discontinuation not investigated.
  - Only 500 simulated datasets used (due to time constraints);
    - › Only 50 imputations used in the implementation approaches.

So…

- Under “treatment policy” estimand, for moderately-sized trials (<200 subjects),
  - “Standard” MI approach allowing for on/off treatment by visit, estimation may not always be possible, given the multiple degrees of freedom required to estimate the model, even though the model is based on all subjects in the trial.
  - Where many (>85%) subjects discontinuing treatment are also expected to discontinue the study and N available to estimate model post-treatment-discontinuation is small (<7), may not be feasible to base model for missings solely on post-treatment subjects, due to sparsity, even with time reduced to a linear covariate.
  - Where >=11 subjects discontinuing treatment are expected to stay in the study, ICH E9 (R1)’s suggestion to use only these subjects to model post-treatment-discontinuation missings seems feasible.

Questions?