Reference based multiple imputation for trials - what's the right variance and how to estimate it?

Jonathan W. Bartlett

Department of Mathematical Sciences
University of Bath, UK
www.thestatsgeek.com

ISCB
19th July 2021

Acknowledgements

This work was supported by a UK Medical Research Council grant (MR/T023953/1).

Reference-based MI, congeniality, and variance estimation

What's the right variance for reference based MI?

How to estimate the variance

Simulations

Conclusions

Outline

Randomised trials with missing data

Consider a randomised trial, with repeated measurements of outcome over follow-up.

Estimand: difference in mean outcomes (actual, not hypothetical) between treatment groups at final visit.

Randomised trials often have missing data, caused by a variety of things.

Sometimes patients may withdraw from the trial, leading to subsequent outcome data being missing.

Withdrawal from the trial often also means patients no longer being given their randomised treatment.

Missing at random (MAR) is then not plausible.

Reference-based MI

An increasingly popular alternative to MAR based analyses is reference-based multiple imputation (RBMI), proposed by Carpenter *et al* 2013 [4].

RBMI uses (in some way) reference (typically control) arm information to help impute missing data in active arm.

For analysis, we fit regression of final time point outcome with treatment and baseline as covariates.

Estimates and standard errors are combined using Rubin's rules.

Variance estimation for reference based MI

A number of researchers found that Rubin's rules variance overestimated the repeated sampling variance of RBMI point estimates of treatment effect [10].

This leads to type I error being controlled at levels below 5% under the null, and power being reduced [6].

The cause is uncongeniality (Meng 1994 [9]) between the *imputation model* and *complete/full data analysis model*.

Why are Rubin's rules biased upwards for reference based MI?

The linear regression complete data variance estimator does not recognise/know that the point estimator has been made more precise as a result of the reference based assumption.

To see the problem, consider what happens when the proportion of missingness in active arm increases towards one.

Jump to reference (J2R) MI (one type of RBMI) effect estimate goes to zero with zero repeated sample variance.

But the linear regression complete data variance will not, and so Rubin's variance will not go to zero.

See Bartlett 2021 [1] for further details.

Outline

What's the right variance?

Whether you use Rubin's rules or repeated sampling variance with RBMI can make a big difference.

Some have argued (e.g. [8, 11]) that frequentist variance should be used since this is what is needed for correct type 1 error control.

Using repeated sampling variance also gives higher statistical power.

But others have argued for Rubin's rules to be used...

Arguments against repeated sampling variance

Carpenter *et al* [3] suggested that for missing data sensitivity analyses, the variance should be no lower (on average) than the complete data variance estimator, and they showed that J2R MI with the repeated sampling variance violates this principle.

But it violates this seemingly sensible principle because it makes a very strong assumption.

And a logical consequence of this assumption is that with more missing data you are more certain about the magnitude of the treatment effect.

If this behaviour does not seem right, it probably means you do not really believe the assumption being made in the reference based MI approach.

Arguments against using the frequentist variance

Cro *et al* 2019 [5] proposed an 'information anchoring' principle.

Information is the reciprocal of the variance of the estimate.

They argue that the ratio of the information in the primary analysis given observed data (e.g. assuming MAR) to what would be obtained with full data, is the same as the ratio in the sensitivity analysis (e.g. via RBMI):

$$\frac{I(\hat{\theta}_{obs,primary})}{I(\hat{\theta}_{full,primary})} = \frac{I(\hat{\theta}_{obs,sensitivity})}{I(\hat{\theta}_{full,sensitivity})}$$

Information anchored sensitivity analysis

In the context of sensitivity analyses, they argue information anchoring is a sensible principle to preserve (which indeed seems reasonable).

Cro *et al* 2019 [5] show that for delta-based MI and RBMI, if you use Rubin's rules for inference, this information anchoring property is essentially satisfied.

But this takes information/variance to be estimated, rather than repeated sampling variance.

Using RBMI with Rubin's rules amounts to acting as if you have neither added nor taken away information, when really you have.

Outline

How to estimate true repeated sampling variance

If we want to use RBMI but estimate the true repeated sampling variance, how can we do it?

There have been a number of proposals for analytical estimators of this variance, including [8, 11].

But these (necessarily) tend to be complex and very model specific...

Tang 2017 [11]

4.2. *Alternative Variance Estimator via the Delta Method*

The asymptotic sampling variance of the MI estimator can be obtained via the delta method

LEMMA 7. $\text{var}(\bar{\alpha}_p^{\infty}) = \sum_{t=1}^{p} \frac{\partial \bar{\alpha}_p^{\infty}}{\partial \underline{\theta}_t} \text{var}(\underline{\hat{\theta}}_t) (\frac{\partial \bar{\alpha}_p^{\infty}}{\partial \underline{\theta}_t})' + V_{\delta}^{e} (X'_f X_f)^{-1} [X'_a X_a] (X'_f X_f)^{-1}$, *where* $\text{var}(\underline{\hat{\theta}}_t) = \sigma_t^2 (Z'_{o_t} Z_{o_t})^{-1}$, *and* $\frac{\partial \bar{\alpha}_p^{\infty}}{\partial \underline{\theta}_t} = l_{pt} [\Upsilon_t - n_{1+} (X'_f X_f)^{-1} (\sum_{s=0}^{p-1} \pi_{1s} \bar{x}_{1s} J'_{t_s})]$.

Let $\frac{\partial \bar{\delta}_p^{\infty}}{\partial \underline{\theta}_t} = l_{pt} (J'_{\delta_t} - \sum_{s=0}^{p-1} \pi_{1s} v_s J'_{t_s})$. The variance of the treatment effect at visit p is

$$\text{var}(\bar{\delta}_p^{\infty}) = \sum_{t=1}^{p} \left(\frac{\partial \bar{\delta}_p^{\infty}}{\partial \underline{\theta}_t} \right) \text{var}(\underline{\hat{\theta}}_t) \left(\frac{\partial \bar{\delta}_p^{\infty}}{\partial \underline{\theta}_t} \right)' + V_{\delta}^{e} \left[\frac{1}{n_{1+}} + h_x \right]. \tag{16}$$

Bootstrap variance estimation for MI under uncongeniality

A possible alternative approach for variance estimation is to use bootstrapping.

Under uncongeniality, you must bootstrap then multiply impute, and not vice-versa, for valid inferences [2].

You need quite a lot of bootstraps to get accurate inference. And then if each bootstrap is imputed lots of times, this is slow.

von Hippel bootstrap MI approach

von Hippel and Bartlett [12] proposed an approach where you bootstrap then impute each bootstrap sample a small (e.g. 2) times.

A simple one-way ANOVA model is used to estimate the between bootstrap and within bootstrap (between imputation) variances, and hence estimate the variance of the overall estimate.

Because you are not relying on Rubin's rules, the imputation does not have to be 'proper' – you can by-pass the posterior draw step, which speeds things up (sometimes considerably).

Outline

Simulation setup

Sample size $n = 1,000$. Randomised 1:1 to control and active treatments.

Outcome: recurrent event count over 365 days, generated from negative binomial model.

Control rate: 0.01, active rate: 0.005. Rate ratio (RR) under full compliance: 0.5, log rate ratio: -0.69.

Dropout completely at random, with dropout time exponentially distributed (rate 0.00025 or 0.0025).

- Copy reference MI [7] with Rubin's rules (10 imputations)
- Copy reference MI with von Hippel bootstrap standard errors (200 bootstraps, 2 imputations)

R code available at
`https://github.com/jwb133/rbmiBootSims`

Simulation results

Means over 1,000 simulations.

	Est. log RR	Emp. SE	Est. SE	SE ratio
Dropout rate 0.00025, 91% complete f/up				
Rubin	-0.66	0.049	0.053	1.08
Bootstrap	-0.66	0.049	0.050	1.01
Dropout rate 0.0025, 40% complete f/up				
Rubin	-0.45	0.040	0.060	1.51
Bootstrap	-0.45	0.039	0.038	0.98

Outline

Conclusions

- Rubin's rules variance can be materially larger than repeated sampling variance for RBMI.
- I argue that the repeated sampling variance is the 'right' one if we are operating in the frequentist paradigm.
- Combining bootstrapping with MI is an attractive approach for estimating the repeated sampling variance of RBMI.
- If the behaviour of RBMI with repeated sampling variance is deemed inappropriate, I believe the correct response is to formulate alternative assumptions & estimation methods which have the desired characteristics.
- Pre-print available here [1].
- R packages:
  - von Hippel bootstrap method: bootImpute
  - RBMI for recurrent events: dejaVu

References I

[1] J. W. Bartlett.
Reference based multiple imputation–what is the right variance and how to estimate it.
arXiv preprint arXiv:2104.14016, 2021.

[2] J. W. Bartlett and R. A. Hughes.
Bootstrap inference for multiple imputation under uncongeniality and misspecification.
Statistical Methods in Medical Research, 29(12):3533–3546, 2020.

[3] J. Carpenter, J. Roger, S. Cro, and M. Kenward.
Response to comments by seaman et al. on "analysis of longitudinal trials with protocol deviation: A framework for relevant, accessible assumptions, and inference via multiple imputation," journal of biopharmaceutical statistics 23: 1352–1371.
Journal of Biopharmaceutical Statistics, 24(6):1363–1369, 2014.

[4] J. R. Carpenter, J. H. Roger, and M. G. Kenward.
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, 2013.

[5] S. Cro, J. R. Carpenter, and M. G. Kenward.
Information-anchored sensitivity analysis: theory and application.
Journal of the Royal Statistical Society: Series A (Statistics in Society), 182(2):623–645, 2019.

[6] F. Gao, G. Liu, D. Zeng, G. Diao, J. F. Heyse, and J. G. Ibrahim.
On inference of control-based imputation for analysis of repeated binary outcomes with missing data.
Journal of Biopharmaceutical Statistics, 27(3):358–372, 2017.

[7] O. N. Keene, J. H. Roger, B. F. Hartley, and M. G. Kenward.
Missing data sensitivity analysis for recurrent event data using controlled imputation.
Pharmaceutical Statistics, 13(4):258–264, 2014.

References II

[8] K. Lu.
An analytic method for the placebo-based pattern-mixture model.
Statistics in Medicine, 33(7):1134–1145, 2014.

[9] X. L. Meng.
Multiple-imputation inferences with uncongenial sources of input (with discussion).
Statistical Science, 10:538–573, 1994.

[10] S. R. Seaman, I. R. White, and F. P. Leacy.
Comment on 'Analysis of Longitudinal Trials With Protocol Deviations: A Framework for Relevant, Accessible Assumptions, and Inference via Multiple Imputation', by Carpenter, Roger, and Kenward.
Journal of Biopharmaceutical Statistics, 24(6):1358–1362, 2014.

[11] Y. Tang.
On the multiple imputation variance estimator for control-based and delta-adjusted pattern mixture models.
Biometrics, 73(4):1379–1387, 2017.

[12] P. T. von Hippel and J. W. Bartlett.
Maximum likelihood multiple imputation: Faster imputations and consistent standard errors without posterior draws.
Statistical Science, 36:400–420, 2021.