Statistical modelling with missing data using multiple imputation

Session 4: Sensitivity Analysis after Multiple Imputation

James Carpenter
London School of Hygiene & Tropical Medicine
Email: james.carpenter@lshtm.ac.uk
www.missingdata.org.uk

Support from MRC, ESRC & German Research Foundation

November 17, 2009

Overview 2 / 31

Outline

- Sensitivity analysis
  - Motivation
  - 'Pattern mixture' approaches
  - 'Selection' approaches
- Course Summary

2 / 31

Motivation

So far, we have assumed missing data are MAR, and discussed various methods of analysis.

Unfortunately, data may well not be MAR.

Therefore, we need to see if our conclusions are sensitive to *plausible* MNAR models.

They will almost certainly be sensitive to *implausible* MNAR models; but this tells us little.

Often, methods for sensitivity analysis are problem specific and opaque.

To be acceptable to collaborators, they need to be fairly general and transparent.

But not necessarily technically simplistic.

3 / 31

Note

We can do sensitivity analysis

1. 'congenially' — i.e. there exists a likelihood equivalient that could be fitted by other means
2. 'uncongenially' — i.e. the imputation model is more general, so stricly not compatible, with the model of interest.

The interesting result is, in common with MI under MAR, remarkably good inference results from mildly non-congenial models (subject to the asymptotic normality of the full data estimators).

4 / 31

Motivating example: peer review trial

The data come from a single blind randomised controlled trial among reviewers for a general medical journal. The aim was to see if training improved the quality of peer review. The study compared two different types of training (face to face training, or a self taught package) with no training[6].

Reviewers were sent three manuscripts (baseline, 1 month and c5 months post intervention) in a similar style to the standard British Medical Journal request for a review, but were told these articles were part of the study and were not paid. The three articles were based on three previously published papers, with original author names, titles and location changed. In addition, nine major and five minor errors were introduced.

5 / 31

Aim of analysis

The outcome is the Review Quality Index, a numerical assessment of the quality of the review averaged over 8 items each scored between 1 and 5.

The analysis showed that the only statistically significant difference was in the quality of the review of paper 2 (immediately post intervention), where the self-taught group did significantly better than the no-training group.

We therefore focus on participants in the self-taught group and no-training group, and focus on the mean review quality index recorded from paper 1 and paper 2 only.

6 / 31

More details

173 reviewers completed the baseline review and were randomised to ‘no training’ and 166 completed the baseline review and were randomised to the ‘self-taught package’.

Completed second review	
No training	162
self-taught package	120

7 / 31

MAR analysis

Suppose the review of paper 2 is MAR given baseline review and intervention.

Then the regression on the observed cases is valid:

```
regress resp base inter

------------------------------------------------------------------------------
        resp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        base |   .3639737   .0470663     7.73   0.000     .2713236    .4566238
       inter |   .2373983   .0703005     3.38   0.001     .0990115     .375785
       _cons |    1.59234   .1328671    11.98   0.000      1.33079    1.853889
------------------------------------------------------------------------------
```

8 / 31

But

		Control group	Self taught group
Returned review of paper 2	n	162	120
	mean	2.65	2.80
	SD	0.81	0.62
Did not return review of paper 2	n	11	46
	mean	3.02	2.55
	SD	0.50	0.75

9 / 31

Analyses

We will compare:

1. MAR analysis (given baseline and treatment)
2. MNAR analysis with a pattern mixture model approach using MI
3. MNAR analysis with a selection model approach using MI.

10 / 31

Pattern mixture models

11 / 31

Key idea

Let Y be a partially observed variable, and X, Z be fully observed.

Let $R = 1$ if Y observed, and $R = 0$ otherwise.

The MAR assumption means $[Y|X, Z, R = 0] = [Y|X, Z, R = 1]$.

In other words the statistical distribution of Y given X, Z is the same whether or not Y is seen.

This is directly exploited by MI: we estimate this distribution from the observed data and use it to impute the missing data.

It follows that if data are MNAR, $[Y|X, Z, R = 0] \neq [Y|X, Z, R = 1]$.

11 / 31

Pattern mixture model

This suggests the following:

1. Multiply impute missing data under MAR
2. Change the imputations to represent likely differences in conditional distributions between observed and missing data.
3. Fit the model of interest to each imputed data set, and combine the results using Rubin's rules in the usual way.

This is an example of the 'pattern mixture' approach: the imputed data is a mixture of potentially different imputations for different patterns.

If the assumptions in Step 2 are valid, inference is valid.

The approach is potentially very general.

12 / 31

Example revisited:

We apply this strategy to the example.

Step 1: impute under MAR, and analyse.
We do 500 imputations (note: no gain here over CC analysis)

Results:

```
. mim: regress resp base inter

Multiple-imputation estimates (regress) Imputations = 500
Linear regression Minimum obs = 339
Minimum dof = 261.0

 ------------------------------------------------------------------------------
 resp |  Coef.   Std. Err. t     P>|t|   [95% Conf. Int.]   FMI
 -----+------------------------------------------------------------------------
 base | .362203  .04721    7.67  0.000   .269258 .455147    0.175
inter | .239395  .070703   3.39  0.001   .100174 .378616    0.202
_cons | 1.59685  .133218   11.99 0.000   1.33461 1.85908    0.149
------------------------------------------------------------------------------
```

As expected, inference similar to regression above: there the treatment effect was 0.24, s.e.$= 0.07$, $p = 0.001$.

13 / 31

Step 2

Using a questionnaire, I asked editorial staff at the BMJ by how much *on average* they expected the quality of missing reviews to differ from observed ones.

I received eighteen responses. For simplicity here, we summarise these using a normal distribution, mean -0.21, variance 0.46^2

We then proceed as follows. For each imputation

1. Draw $\Delta \sim N(-0.21, 0.46^2)$
2. Add Δ to each imputed response

14 / 31

Step 3

We now

1. Fit the model of interest to each imputed data set
2. Combine the results using Rubin's rules

15 / 31

Results

Analysis	Estimate	S.E.	p-value
Regression	0.237	0.070	0.001
MI (MAR)	0.239	0.071	0.001
MI (MNAR)	0.187	0.119	0.120

The prior information (from the editors) quantifies how much, on average, they believe the review quality differs between missing and observed reviews. Using MI, we have incorporated this MNAR information into the analysis.

We conclude there is no benefit of the self taught intervention.

16 / 31

Further comments

- The approach is relatively straightforward to communicate.
- If we can get information from experts about the distribution of Δ this can be directly incorporated.
- We can extend to more complex (realistic) settings: e.g. a different distribution of Δ in different arms.
- In general the distribution of Δ is not independent across arms (i.e. expert's values of Δ in different arms will be correlated).
- The approach generalises to more complex settings: a start has been made with longitudinal data[3].
- We can also include information on different reasons for dropout when deciding how to move imputed values away from MAR.

17 / 31

Selection model

18 / 31

Selection Model analysis

The alternative general approach is to fit a *selection model*. Essentially, within this framework we now have two models (i) our model of interest, and (ii) a model for the chance of observations being missing.

We fit these jointly for inference about our questions of interest.

We now illustrate with our asthma example.

18 / 31

Reviewers example re-visited

One possible selection model (many would fit the data equally well) is:

$$\text{Quality of review 2}_i = \beta_0 + \beta_1 \text{inter}_i + \beta_2 \text{base}_i + \epsilon_i, \quad \epsilon_i \overset{iid}{\sim} N(0, \sigma^2).$$

Selection model:

$$\begin{aligned}\text{logit}\Pr\{\text{observe review 2}_i\} =& \alpha_0 + \alpha_1 \text{inter}_i \\ &+ \alpha_2 \text{base}_i + \delta(\text{quality of review 2})_i.\end{aligned}$$

If $\delta = 0$ then data are MAR and we can fit the models separately.

If $\delta \neq 0$ then data are MNAR, and the models must be fitted jointly.

We cannot test if data are MNAR, because this depends on uncheckable modelling assumptions (we would need to see the missing reviews).

19 / 31

Model fitting

Implicit in fitting models like this is numerical integration over the missing data, to obtain a likelihood in terms of the observed data which is then maximised [5].

In practice, the easiest way to do this is using MCMC methods, in a package like `WinBUGS`. See [1] for a trials example; examples with missing covariates can be downloaded from www.missingdata.org.uk

There is usually little information about parameters relating the chance of observing a variable to the potentially unseen value of that variable (i.e. the parameter δ above).

We can either (i) explore different values or (ii) try and obtain expert opinion about possible values

20 / 31

Approximate fitting using MI

It turns out it is possible to approximate the fit of a selection model after MI under MAR [4].

Essentially, we impute data assuming MAR, and fit the model of interest to each imputed data set.

We then use a weighted version of Rubin's rules, up-weighting estimates more likely under the MNAR mechanism.

21 / 31

Asthma example revisited

The model of interest is

$$\text{Quality of review 2}_i = \beta_0 + \beta_1 \text{inter}_i + \beta_2 \text{base}_i + \epsilon_i, \quad \epsilon_i \overset{iid}{\sim} N(0, \sigma^2).$$

and the selection model:

$$\begin{aligned}\text{logit}\,\Pr\{\text{observe review 2}_i\} =& f(\text{observed data}_i) \\ &+ \delta(\text{quality of review 2}_i).\end{aligned}$$

Above we had a linear model in terms of the observed covariates. However, it turns out that the part of the linear predictor involving the observed covariates does not have to be specified. So it could be any function, which we denote f.

We need to choose δ.

We can use logistic regression to inform the choice of δ, as illustrated below.

Algorithm

In the context of the asthma study, with n_1 missing response values, we proceed as follows:

1. Use multiple imputation, under MAR, to create K imputed data sets.
   Denote the k^{th} set of imputed responses $y_1^k, \ldots, y_{n_1}^k$, and the parameter from the model of interest $\hat{\theta}_k$, with variance $\hat{\sigma}_k^2$.
2. For imputation $k = 1, \ldots, K$ calculate $\gamma_k = -\delta \sum_{i=1}^{n_1} y_i^k$, and calculate $\bar{\gamma} = \sum_{k=1}^{K} \gamma_k / K$.
3. Set
$$w_k = \frac{e^{\gamma_k - \bar{\gamma}}}{\sum_{k=1}^{K} e^{(\gamma_k - \bar{\gamma})}}.$$
4. Our estimate of θ under MNAR is
$$\sum_{k=1}^{K} w_k \hat{\theta}_K$$

The weights can also be applied to estimate the variance [4].
We need at least $K = 100$ imputations for this approach.

Example: reviewers trial

To select a plausible δ, perform a logistic regression of the chance of observing the response on the observed data:

```
* r is 1 if second review observed observed
logit r base inter
------------------------------------------------------------------------------
imputed |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
------------+-----------------------------------------------------------------
   base |   .2097852   .2195739    0.96   0.339    -.2205717    .6401421
  inter | -1.750147    .358082   -4.89   0.000    -2.451975   -1.048319
  _cons |   2.140697   .642787    3.33   0.001     .8808579    3.400537
------------------------------------------------------------------------------
```

The log-odds of observing response is not strongly related to baseline, but if anything the association is positive.

Suppose the association with the unseen response is the same in the first instance, $\delta = 0.2$.

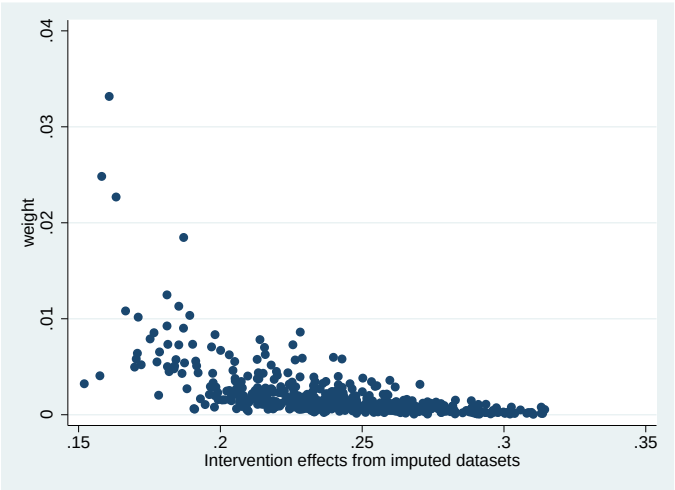
We create 500 imputations under MAR (exactly as for the pattern mixture approach above)

Results

Applying the algorithm above gives the following results (again, Stata code is given in the practical notes):

Missing data Assumption	Estimate	S.E.	Z/T	p-value
MAR	0.237	0.070	3.38	0.001
MNAR ($\delta = 0.2$)	0.214	0.071	3.02	0.003

Diagnostic plot



There is reasonable support for MNAR distribution from MAR imputations.

Some comments

- Although this method is approximate, generating a lot of imputations and re-weighting is often quicker and less prone to error than a fitting a selection model. Cf the results here and [1].
- At the least it provides a check on the results of such models.
- As we chose a particular δ, we did not inflate the standard error, as we did with the pattern mixture approach.
- This method can be applied when the selection mechanism depends on a partially observed covariate.
- Including additional covariates in the mechanism is more tricky.

27 / 31

Summary

28 / 31

Summary of this session

- We have sketched how sensitivity analysis can be framed using
  - a pattern mixture model, and
  - a selection model.
- We have outlined how sensitivity analysis can be done under both models following multiple imputation under the missing at random assumption.
- We have illustrated the approaches with data from a clinical trial with patient withdrawal.
- We have discussed how the approach can be used with partially observed covariates.

28 / 31

What we haven't covered in this workshop

There is a vast literature on the analysis of partially observed data. Important strands we have not touched on are:

- Inverse probability weighting. Simply, we fit a logistic regression to estimate the probability of a partially observed variable been seen, and then weight cases in the complete case analysis by the inverse of these probabilities.
  See [2] for a comparison with MI and [7] for an accessible introduction; also course notes on www.missingdata.org.uk
- Hot-deck imputation, where we impute from a donor pool identified using some distance metric.
- Joint modelling methods for imputation.
- Imputation of longitudinal and hierarchical data.

29 / 31

Course summary

- It is important to understand—as far as possible—the mechanism(s) causing the missing data
- Avoid ad-hoc methods
- Undertake preliminary analyses before MI to (i) identify auxiliary variables to include in MI and (ii) get an idea of the likely gains of MI
- Think carefully about the MI model: is it consistent with the model of interest?
- Be cautious: methods such as MI are not a panacea. They depend on assumptions you cannot verify from your data. Without care, you may on occasion get worse results than from a CC analysis.
- Consider sensitivity analysis after MI to check the robustness of key conclusions
- Report the MI analysis carefully, as suggested in the guidelines at the end of Lecture 3.
- When designing studies, don't forget prevention is better than cure!

30 / 31

References

[1] J Carpenter, S Pocock, and C J Lamm. Coping with missing data in clinical trials: a model based approach applied to asthma trials. *Statistics in Medicine*, 21:1043–1066, 2002.

[2] J R Carpenter, M G Kenward, and S Vansteelandt. A comparison of multiple imputation and inverse probability weighting for analyses with missing data. *Journal of the Royal Statistical Society, Series A (Statistics in Society)*, 169:571–584, 2006.

[3] James Carpenter, James Roger, and Mike Kenward. Relevant, Accessible Sensitivity Analyses Using Multiple Imputation *submitted*, 2009.

[4] James R Carpenter, Michael G Kenward, and Ian R White. Sensitivity analysis after multiple imputation under missing at random — a weighting approach. *Statistical Methods in Medical Research*, 16:259–275, 2007.

[5] P J Diggle and M G Kenward. Informative dropout in longitudinal data analysis (with discussion). *Journal of the Royal Statistical Society Series C (applied statistics)*, 43:49–94, 1994.

[6] S Schroter, N Black, S Evans, J Carpenter, F Godlee, and R Smith. Effects of training on quality of peer review: randomised controlled trial. *British Medical Journal*, 328:673–675, 2004.

[7] S Vansteelandt, J R Carpenter, and M G Kenward. Analysis of incomplete data using inverse probability weighting and doubly robust estimators, 2009.

31 / 31