Estimands, Randomisation and Sensitivity Analysis

James Carpenter

London School of Hygiene and Tropical Medicine, London, UK, &
MRC Clinical Trials Unit, London, UK

james.carpenter@lshtm.ac.uk

European Statistical Meeting on Estimands, 28th September 2015

Acknowledgements

Suzie Cro (LSHTM & MRC Clinical Trials Unit)

Mike Kenward (LSHTM)

James Roger (LSHTM, formerly GSK)

Outline

- Example: depression trial
  - Data
  - Estimands
  - Primary analysis
  - Sensitivity analysis
  - Jump to Reference
- Randomisation justification for primary analysis:
  - In theory
  - Performance in practice
- Modelling the selection process
  - In theory
  - Performance in practice
- Example revisited
- Discussion

Antidepressant Trial

The following data come from a three-arm multicentre RCT on the treatment of depression (see [1],[2]); I have adjusted the mean for Treatment C.

The outcome is the Hamilton depression score, which takes values in $[0, 50]$.

In the original trial, 369 patients were randomised to receive one of treatments A, B, C.

Data were collected at baseline, and weeks 1, 2, 3 and 4.

Here we consider treatments A and C, and baseline and visit 4 data.

Data

Treatment A

	Mean (SD) Hamilton Score		
Pattern	Baseline	4 weeks	n
1	21.86 (3.79)	11.70 (6.65)	76 (63%)
2	22.13 (3.66)	—	44 (37%)

Treatment C

	Mean (SD) Hamilton Score		
1	21.10 (4.27)	13.27 (7.34)	94 (73%)
2	22.46 (3.64)	—	35 (27%)

Estimands

Details of follow-up criteria for this trial are unavailable, but it is likely that patients were followed up until they discontinued the treatment.

We will consider:

- A de jure estimand
- A de facto estimand (jump to reference)

In general, we wish to pre-specify a broad based population for our estimand (i.e. incorporating a range of behaviours within the 'class'). Then results are

- more likely to be generalizable without additional assumptions;
- make good use of our data, and
- more likely to be robust in sensitivity analyses.

Primary analysis

```
. regress v4 base treat

----------------------------------------------------------------
   v4 |  Coef.    SE       t       P>|t|     [95% Conf. Interval]
------+---------------------------------------------------------
 base |  0.5820  0.126    4.61    0.000     0.3328     0.8312
treat |  2.012   1.030    1.95    0.052    -0.0213     4.047
 cons | -1.023   2.862   -0.36    0.721    -6.674      4.628
----------------------------------------------------------------
```

Question: Under the null, how robust is our estimator to

- normality (questionable if the data represent a mix of behaviours)
- MAR when the data are non-normal

Sensitivity analysis

There are two broad approaches to this:

1. Maintain our primary analysis estimation procedure, but vary the assumptions about post-deviation behaviour, obtaining a valid point estimate and corresponding SE in each case
   - the primary analysis model may be incompatible—in some aspects—with some of the sensitivity scenarios.

2. Explicitly model deviation, and post-deviation behaviour, in the primary analysis, and vary the models & assumptions for the sensitivity analyses
   - each model will be compatible with its sensitivity scenario.

Sensitivity analysis: J2R

Using approach (1) from the previous slide gives the following (using Stata `mimix` program):

Assumptions	Treatment estimate	SE	p-value
MAR	2.01	1.03	0.052
J2R (reference = A)	1.49	0.976	0.128
J2R (reference = C)	1.55	0.984	0.116

Note the SE from Rubin's MI rules satisfies:

$$V_{sens,partial} \approx \frac{V_{primary,partial}}{V_{primary,full}} \times V_{sens,full},$$

as it also does for the 'Δ' method.

Analysis of Covariance

A practically important, but sometimes overlooked, property of a t-test and also the treatment test from the analysis of covariance, is that they have a randomisation justification under the null when:

- patients are sampled randomly from a (super-)population, and
- sampled patients are randomly allocated to treatment

This is an asymptotic property [3], but means that under the null the size is likely to be well preserved, even if the data are quite non-normal.

However, the power may be reduced; but this will likely be moderate for moderate non-normality.

Simulation example

Draw $X_i \sim N(0,1)$, $T_i \sim \text{Bin}(\pi = 0.5, n = 1)$, $i = 1, \ldots, n$.

Set $\beta_0 = 0, \beta_1 = 0.5, \beta_2 = 0$ and draw

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 T_i + e_i,$$

with

1. $n = 25$, $e_i \sim N(0, 0.75)$, and
2. $n = 25$, $e_i \sim \chi^2_1$
3. $n = 100$, $e_i \sim N(0, 0.75)$
4. $n = 100$, $e_i \sim \chi^2_{10}$
5. $n = 100$, $e_i \sim \chi^2_1$

Fit a linear regression of Y on X and treatment and note whether the p-value is < 0.05.

Repeat 5000 times.

Results for β_2

Scenario	$n =$	Empirical size
Normal	25	0.050
χ^2_1	25	0.041
Normal	100	0.053
χ^2_{10}	100	0.051
χ^2_1	100	0.052

With $n = 100$ results are very robust to skewness.

Now with missing data

We simulate the following scenarios, with $\beta_0 = 0$, $\beta_1 = 1$ and $\beta_2 = 0$. Let $R_i = 1$ if the outcome for patient i is observed.

Sample size	Resid dist	Selection mechanism $\text{logit}\{\Pr(R_i = 1)\} =$	mean n_{obs}	Size
$n = 100$	normal	$-3 + 2T_i$	84	0.052
$n = 100$	χ^2_{10}	$-3 + 2T_i$	84	0.053
$n = 100$	χ^2_{10}	$-3 + 2T_i + X_i$	78	0.049
$n = 100$	χ^2_{10}	$-3 + 2T_i + 4X_i$	67	0.048
$n = 100$	χ^2_1	$-3 + 2T_i + 4X_i$	67	0.049
$n = 50$	χ^2_1	$-3 + 2T_i + 4X_i$	34	0.044
$n = 100$	χ^2_{10}	$-3 + 2T_i + X_i + 0.1Y_i$	66	0.071

Type 1 error preserved under MAR.

Selection model

Now consider the selection model:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 T_i + e_i, \quad e_i \sim \chi^2_{10}$$
$$\text{logit}\{\Pr(R_i = 1)\} = \alpha_0 + \alpha_1 x_i + \alpha_2 T_i + \alpha_3 Y_i$$

Simulate $n = 100$ observations as above, and make them MAR with mechanism

$$\text{logit}\{\Pr(R_i = 1)\} = -3 + 2T_i + X_i.$$

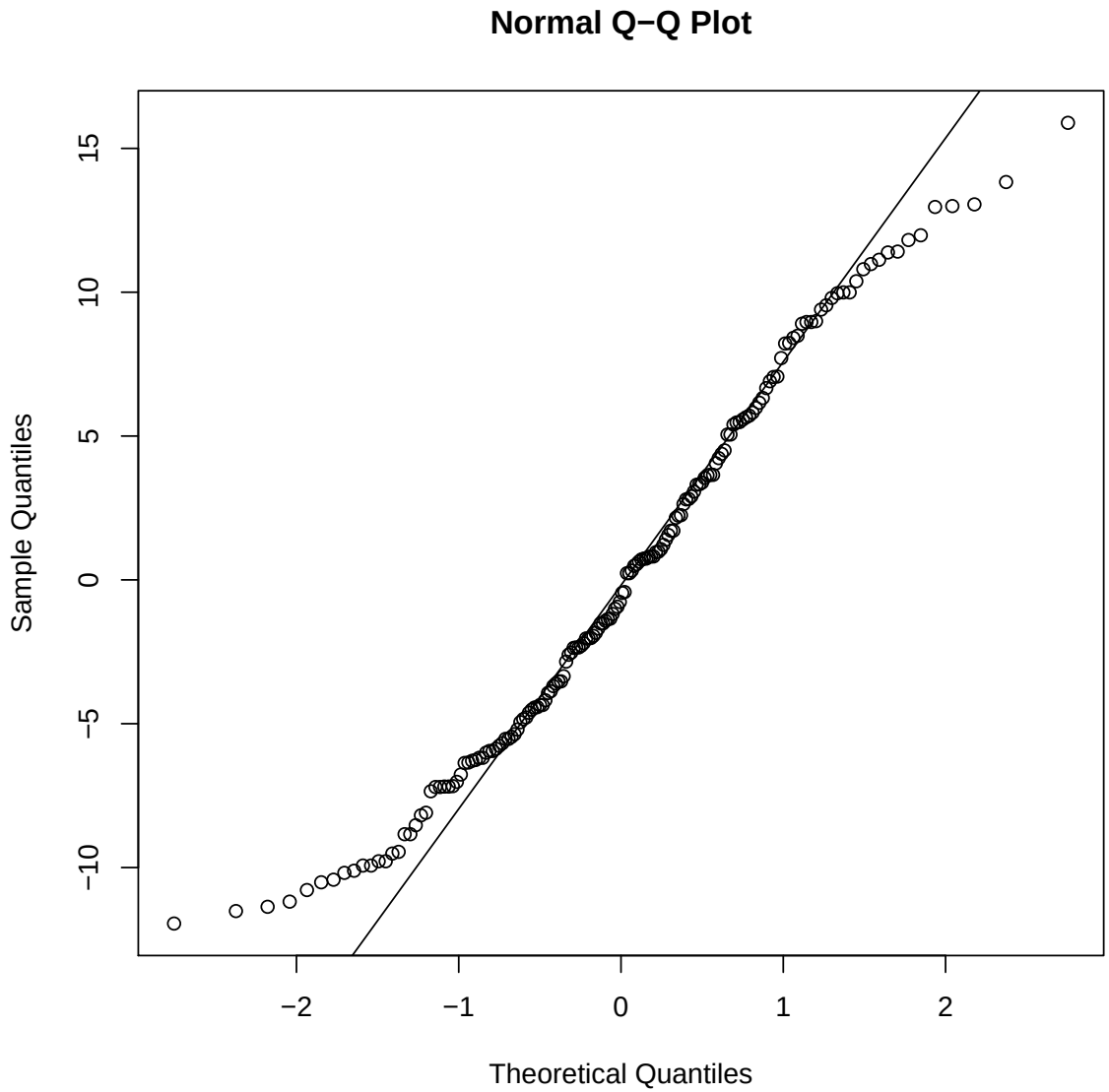
Fit the selection model above and look at the estimate and SE for β_2.

Results

Model	Sample size	Resid dist	Selection mechanism $\text{logit}\{\Pr(R_i = 1)\} =$	mean n_{obs}	Size
ANCOVA	$n = 100$	χ^2_{10}	$-3 + 2T_i + X_i$	78	0.049
Sel Mod	$n = 100$	χ^2_{10}	$-3 + 2T_i + X_i$	78	0.378

The average value of $\hat{\beta}_{2,\text{sel mod}}$ is -1.958.

Residuals


Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
15
10
5
0
−5
−10
−2
−1
0
1
2

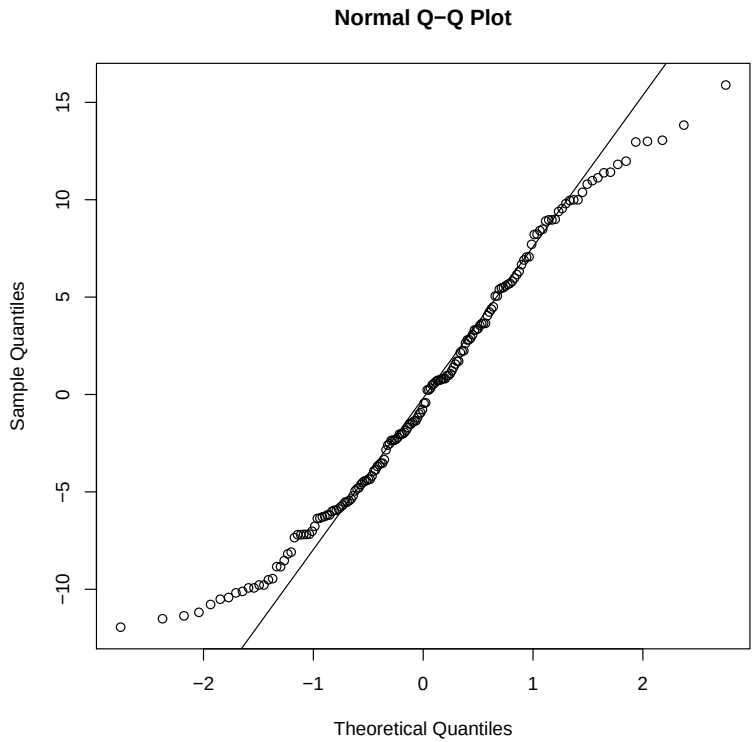
Analysis

Fit the same selection model to the depression data:

Assumptions	Treatment estimate	SE	p-value
MAR	2.01	1.03	0.052
J2R (reference = A)	1.49	0.976	0.128
J2R (reference = C)	1.55	0.984	0.116
Selection model	3.47	1.37	$\sim$ 0.011

$\hat{\alpha}_3 = 1.16$, 95% HPD $(0.45, 2.27)$.

Explanation



— Little dependence of dropout on baseline and treatment.
— Model makes selection depend on outcome: missing values put in the tail
— Results are very sensitive to the distribution tail length.

Analysis

Fit the same selection model to the depression data:

Assumptions	Treatment estimate	SE	p-value
MAR	2.01	1.03	0.052
J2R (reference = A)	1.49	0.976	0.128
J2R (reference = C)	1.55	0.984	0.116
Selection model (no constraint)	3.47	1.37	~ 0.011
Selection model (constraint)	1.44	1.13	~ 0.202

$\hat{\alpha}_{3,no\ constraint} = 1.16$, 95% HPD $(0.45, 2.27)$.
$\hat{\alpha}_{3,constraint} = -0.15$, 95% HPD $(-0.3, -0.02)$.
Both models have converged; they put missing values at opposite extremes.

Discussion

- In trials, ANCOVA inference has a randomisation justification—as well as a central limit theorem justification—when the data are non-normal.
- This holds up well under MAR.
- Inference for our primary estimand should have this protection, where possible.
- Sensitivity analysis then explores the robustness of inference from the primary analysis model as the assumptions vary.
- If our primary analysis model includes a selection model (or uses inverse probability weighting), results can be *very* sensitive to distributional/modelling assumptions.
- If we wish to do this, we should be aware that the protection of randomisation inference no longer holds, and take care!

References

[1] 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.

[2] A Heyting, J G A Essers, and J T B M Tolboom. A practical application of the patel-kenward analysis of covariance to data from an anti-depressant trial with drop-outs. *Statist. Appl.*, 2:295–307, 1990.

[3] O Kempthorne. *The Designs and Analysis of Experiments*. New York, Wiley, 1956.