

# Statistical methods for clinical studies with missing data: what's hot, what's cool and what's useful

James R. Carpenter

London School of Hygiene & Tropical Medicine

[james.carpenter@lshtm.ac.uk](mailto:james.carpenter@lshtm.ac.uk)

[www.missingdata.org.uk](http://www.missingdata.org.uk)

Support for JRC from UK ESRC & German Research Foundation

August 24, 2009

# Acknowledgements

Overview

● Acknowledgements

Overview

Principles

Implications

Example

What should we carry forward?

Mike Kenward (LSHTM)

James Roger (GlaxoSmithKline Research)

Many of the points covered this talk are discussed in more detail in the book 'Missing data in clinical trials — a practical guide' (joint with Mike Kenward), commissioned by the UK National Health Service, available free on-line at [www.missingdata.org.uk](http://www.missingdata.org.uk).

# Outline

Overview

Overview

● Outline

Principles

Implications

Example

What should we carry forward?

- Principles for analyses of partially observed data
- Implications
  - Assumptions → Valid methods
  - Example
- What should we carry forward?

The talk focuses on analysing repeated measures from clinical trials. However, the ideas apply more widely.

We will make some passing comments on the draft CHMP guideline on handling missing data in RCTs, available at

<http://www.emea.europa.eu/pdfs/human/ewp/43998007en.pdf>

## The ICH E9 guideline, 1999

### Overview

### Overview

### Principles

#### ● The ICH E9 guideline, 1999

- Study validity and sensible analysis
- Why there can be no universal method:
  - Form of assumptions
  - E.g. Missing At Random (MAR)
  - What's the question: 'on-treatment' vs ITT
- Key points

### Implications

### Example

What should we carry forward?

- Missing data are a potential source of bias
- Avoid if possible (!)
- With missing data, a trial may still be regarded as valid if the methods are *sensible*, and preferably predefined
- There can be no universally applicable method of handling missing data
- The sensitivity of conclusions to methods should thus be investigated, particularly if there are a large number of missing observations

These points are widely accepted. The question is, how do we apply them in practice?

# Study validity and sensible analysis

## Overview

---

## Overview

---

## Principles

---

- The ICH E9 guideline, 1999
- **Study validity and sensible analysis**
- Why there can be no universal method:
- Form of assumptions
- E.g. Missing At Random (MAR)
- What's the question: 'on-treatment' vs ITT
- Key points

## Implications

---

## Example

---

What should we carry forward?

---

The sampling process involves both the selection of the patients and the process by which observations become missing — the *missingness mechanism*.

Thus for sensible inference, we need to take account of the missingness mechanism

By *sensible* we mean the claimed properties of our estimators hold, e.g.

- consistency
- confidence interval coverage

## Why there can be no universal method:

### Overview

---

### Overview

---

### Principles

---

- The ICH E9 guideline, 1999
- Study validity and sensible analysis
- **Why there can be no universal method:**
- Form of assumptions
- E.g. Missing At Random (MAR)
- What's the question: 'on-treatment' vs ITT
- Key points

### Implications

---

### Example

---

What should we carry forward?

---

In contrast with the sampling process, which is usually known, the missingness mechanism is usually unknown.

The data alone cannot usually definitively tell us the sampling process.

Likewise, the missingness pattern, and its relationship to the observations, cannot identify the missingness mechanism.

With missing data, extra assumptions are thus required for analysis to proceed.

The validity of these assumptions cannot be determined from the data at hand.

Assessing the sensitivity of the conclusions to the assumptions should therefore play a central role.

# Form of assumptions

## Overview

---

## Overview

---

## Principles

---

- The ICH E9 guideline, 1999
- Study validity and sensible analysis
- Why there can be no universal method:
- **Form of assumptions**
- E.g. Missing At Random (MAR)
- What's the question: 'on-treatment' vs ITT
- Key points

## Implications

---

## Example

---

What should we carry forward?

---

We can either make assumptions about the selection mechanism

- ...which has implications for the distribution of the missing observations given the observed data.

Or we can make assumptions about the distribution of the missing observations given the observed data

- ...which has implications for the selection mechanism.

We should check the plausibility of our assumptions from both viewpoints.

## E.g. Missing At Random (MAR)

Overview

Overview

Principles

- The ICH E9 guideline, 1999
- Study validity and sensible analysis
- Why there can be no universal method:
- Form of assumptions
- E.g. Missing At Random (MAR)
- What's the question: 'on-treatment' vs ITT
- Key points

Implications

Example

What should we carry forward?

MAR is a *conditional independence* statement:

Given observed data, the probability of missing outcomes is independent of the underlying values.

Let  $X$  denote baseline covariates, including treatment;  $Y$  denote outcome (scalar) and  $R = 1$  if  $Y$  observed.

The MAR selection mechanism means:

$$[Y|X, R] = \frac{[Y, X, R]}{[X, R]} = \frac{[R|Y, X][Y, X]}{[R|X][X]} = [Y|X].$$

The conditional distribution of the outcome given treatment and baseline covariates is the same whether or not the underlying value is observed.

## What's the question: 'on-treatment' vs ITT

### Overview

### Overview

### Principles

- The ICH E9 guideline, 1999
- Study validity and sensible analysis
- Why there can be no universal method:
  - Form of assumptions
  - E.g. Missing At Random (MAR)
  - **What's the question: 'on-treatment' vs ITT**
- Key points

### Implications

### Example

What should we carry forward?

Consider a study in patients with chronic asthma, with scheduled FEV<sub>1</sub> measurements at 2, 4, 8, 12 weeks post-randomisation.

Suppose we only have data while patients remain 'on-treatment'.

An analysis under the MAR assumption (e.g. using a Saturated Repeated Measures (SRM) model) implicitly estimates the distribution of underlying unobserved outcomes given observed outcomes *assuming the patient remained on-treatment*.

In this setting, an ITT analysis will require a MNAR assumption.

# Key points

## Overview

## Overview

## Principles

- The ICH E9 guideline, 1999
- Study validity and sensible analysis
- Why there can be no universal method:
  - Form of assumptions
  - E.g. Missing At Random (MAR)
  - What's the question: 'on-treatment' vs ITT
- **Key points**

## Implications

## Example

What should we carry forward?

- There is no universal method
- There *are* universal principles:
  - understand as much as possible about the reasons for/causes of missing data
  - be clear about the question you are trying to answer
  - make (untestable) assumptions about the selection mechanism or distribution of missing given observed data (these should be accessible and relevant)
  - perform a statistically valid analysis under these assumptions
  - explore how robust inference is to these assumptions

## Assumptions and Methods I:

Overview

---

Overview

---

Principles

---

Implications

---

- **Assumptions and Methods I:**

- Assumptions and methods II

- Assumptions and methods III

- Assumptions and methods IV:

- So what about LOCF?

- What's cool about LOCF?

- Forecast: LOCF cool-change

- Some LOCF

References:

Example

---

What should we carry forward?

---

Assume observed patients are a random subset of randomised patients. Data are Missing Completely at Random (MCAR).

Obtain valid inference for treatment effect at the end of the study using

- ANCOVA (available data at final visit)
- SRM model (all available data — more powerful)
- Multiple imputation
- Inverse probability weighting and related methods

Note:

1. SRM models allow the variance to differ between arms
2. LOCF not valid in this setting (MCAR)

## Assumptions and methods II

Overview

---

Overview

---

Principles

---

Implications

---

- Assumptions and Methods I:

- **Assumptions and methods II**

- Assumptions and methods III

- Assumptions and methods IV:

- So what about LOCF?

- What's cool about LOCF?

- Forecast: LOCF cool-change

- Some LOCF

References:

Example

---

What should we carry forward?

---

Given treatment and baseline outcome measure, distribution of missing outcomes given observed does not depend on the underlying outcomes.

In other words, data are Missing At Random given baseline and treatment

Methods:

- ANCOVA (available data at final visit)
- SRM model (all available data — more powerful)
- Multiple imputation
- Inverse probability and related methods

Note:

1. SRM models allow the variance to differ between arms
2. LOCF not valid in this setting

## Assumptions and methods III

Overview

---

Overview

---

Principles

---

Implications

---

- Assumptions and Methods I:
  - Assumptions and methods II
  - **Assumptions and methods III**
  - Assumptions and methods IV:
  - So what about LOCF?
  - What's cool about LOCF?
  - Forecast: LOCF cool-change
  - Some LOCF
- References:

Example

---

What should we carry forward?

---

Given treatment and baseline outcome measure plus other baseline *and* post-randomisation data, the distribution of missing outcomes given those observed does not depend on the underlying outcomes.

This is more plausible, as it allows a richer dropout mechanism.

Methods:

- SRM type model (post randomisation data as additional responses; use general variance structure)
- Multiple imputation (include post-randomisation data; impute by treatment arm)
- Inverse probability and related methods (you need to work a bit harder)

Notes

1. ANCOVA does not give valid inference here
2. nor does LOCF

## Assumptions and methods IV:

Overview

---

Overview

---

Principles

---

Implications

---

- Assumptions and Methods I:

- Assumptions and methods II

- Assumptions and methods III

- **Assumptions and methods IV:**

- So what about LOCF?

- What's cool about LOCF?

- Forecast: LOCF cool-change

- Some LOCF

References:

Example

---

What should we carry forward?

---

Even given the observed data, the conditional distributions of missing and observed outcomes differ. Data are Missing Not at Random (MNAR).

Specify either:

- how they differ
  - valid inference using multiple imputation; well approximated by SRM-model calculations in some cases;

*OR*

- how, given observed data, chance of observing outcome depends on the underlying outcome value
  - valid inference using selection modelling with numerical integration; inverse probability and related methods also possible.

## So what about LOCF?

Overview

---

Overview

---

Principles

---

Implications

---

- Assumptions and

Methods I:

- Assumptions and methods II

- Assumptions and methods III

- Assumptions and methods IV:

- **So what about LOCF?**

- What's cool about LOCF?

- Forecast: LOCF cool-change

- Some LOCF

References:

Example

---

What should we carry forward?

---

LOCF gives valid inference if

- all post-withdrawal observations are exchangeable
- i.e. selection implicitly depends on the future

Are the assumptions required for LOCF as well understood as its common use might lead us to believe?

# What's cool about LOCF?

Overview

Overview

Principles

Implications

- Assumptions and Methods I:
  - Assumptions and methods II
  - Assumptions and methods III
  - Assumptions and methods IV:
  - So what about LOCF?
  - **What's cool about LOCF?**
  - Forecast: LOCF cool-change
  - Some LOCF
- References:

Example

What should we carry forward?

Cool is....

1. Fashionably attractive or impressive

# What's cool about LOCF?

Overview

Overview

Principles

Implications

- Assumptions and Methods I:
  - Assumptions and methods II
  - Assumptions and methods III
  - Assumptions and methods IV:
  - So what about LOCF?
  - **What's cool about LOCF?**
  - Forecast: LOCF cool-change
  - Some LOCF
- References:

Example

What should we carry forward?

Cool is...

1. Fashionably attractive or impressive
2. Showing no friendliness or enthusiasm...for an idea

# Forecast: LOCF cool-change

Overview

Overview

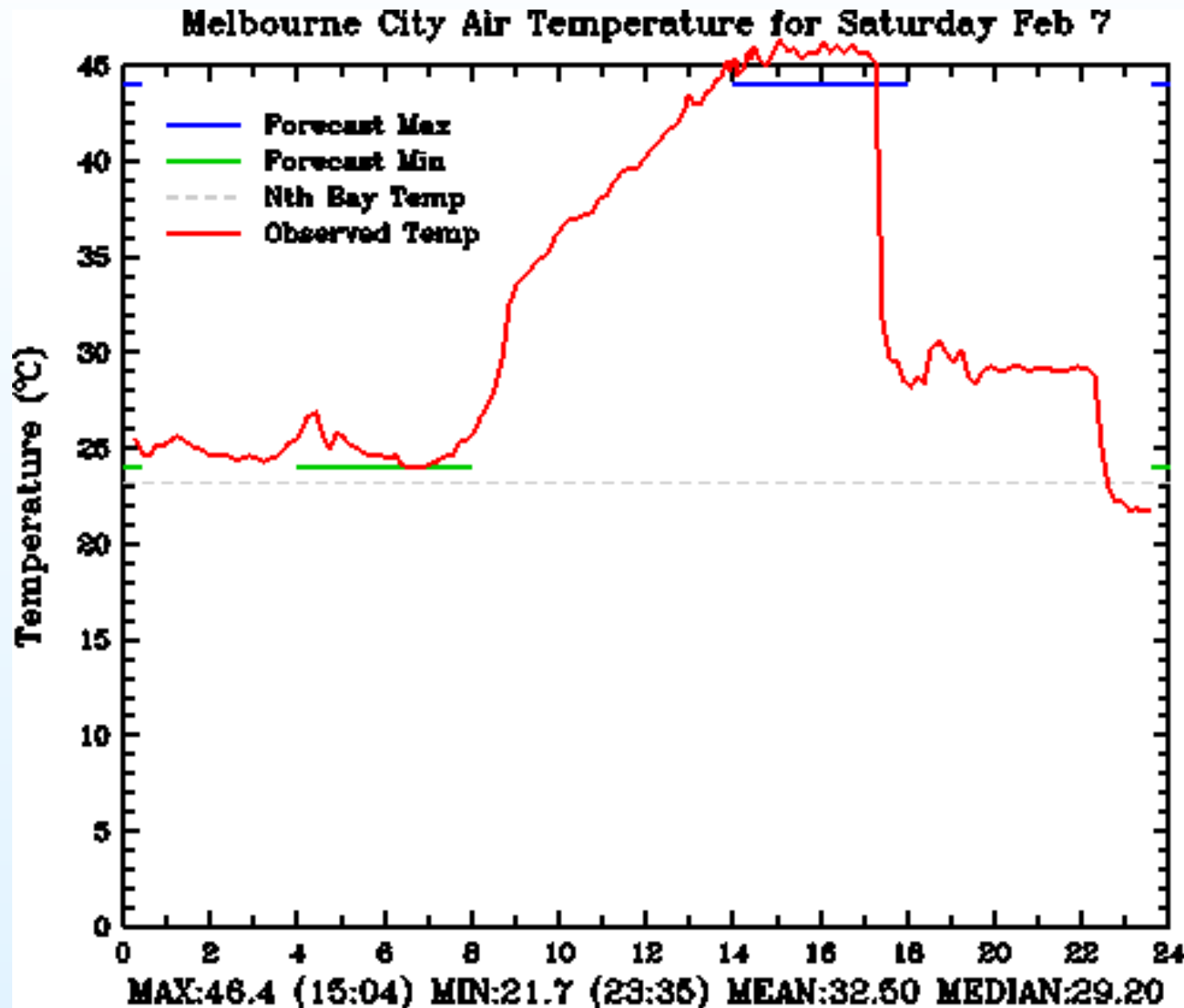
Principles

Implications

- Assumptions and Methods I:
  - Assumptions and methods II
  - Assumptions and methods III
  - Assumptions and methods IV:
  - So what about LOCF?
  - What's cool about LOCF?
  - **Forecast: LOCF cool-change**
  - Some LOCF
- References:

Example

What should we carry forward?



# Forecast: LOCF cool-change

Overview

Overview

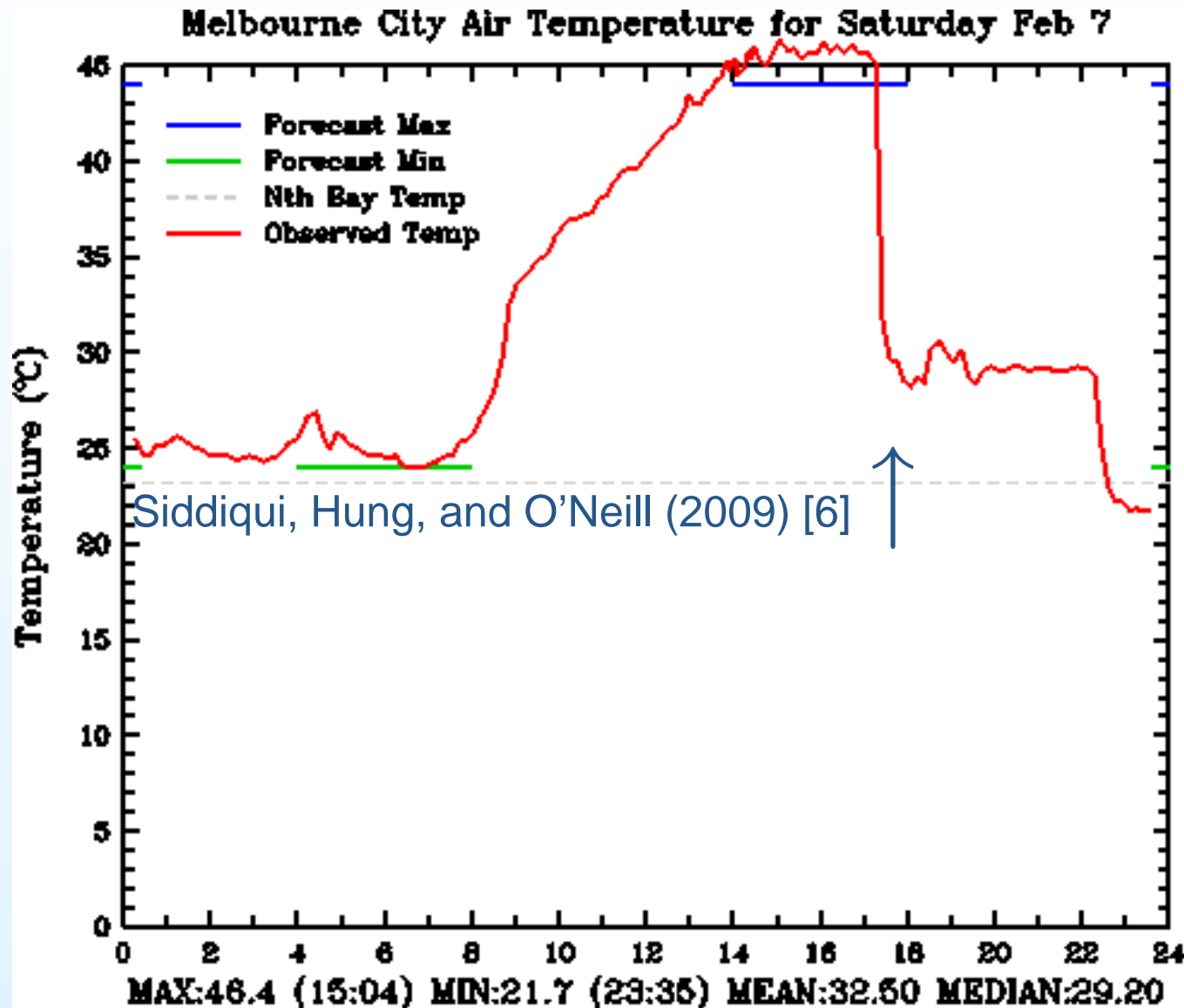
Principles

Implications

- Assumptions and Methods I:
  - Assumptions and methods II
  - Assumptions and methods III
  - Assumptions and methods IV:
  - So what about LOCF?
  - What's cool about LOCF?
  - **Forecast: LOCF cool-change**
  - Some LOCF
- References:

Example

What should we carry forward?



## Some LOCF References:

Molenberghs *et al* (2004)[5] (longitudinal data)

“In all cases LOCF typically produces bias, of which the direction and magnitude depend on the true but unknown treatment effects... note LOCF is not valid even under the strong MCAR condition”

Cook *et al* (2004)[2] (binary data)

“We find that for both of these approaches [(i) and (ii)] imputation by LOCF can lead to substantial bias in estimators of treatment effects, the type I error rates of associated tests can be greatly inflated, and the coverage probability can be far from the nominal level.”

Kenward and Molenberghs (2009) [4]

“...with the exception of certain degenerate and unrealistic special cases, these methods are incompatible with the sensible requirement that the mechanism governing dropout be independent of future, unobserved measurements, given covariates, past observed measurements, and the current possibly missing measurement. ”

Overview

Overview

Principles

Implications

- Assumptions and

Methods I:

- Assumptions and

methods II

- Assumptions and

methods III

- Assumptions and

methods IV:

- So what about LOCF?

- What's cool about

LOCF?

- Forecast: LOCF

cool-change

- **Some LOCF**

**References:**

Example

What should we carry  
forward?

## Asthma study

Overview

---

Overview

---

Principles

---

Implications

---

Example

---

- **Asthma study**

- Analyses
- Jump to active—illustration
- Results

What should we carry forward?

---

Patients with chronic asthma randomised to one of four increasing doses of an active drug, or placebo. Here, we use data from the placebo and lowest active dose arm.

After randomisation, patients were asked to attend the clinic at 2, 4, 8 and 12 weeks. The primary outcome is forced expiratory volume in 1 second ( $FEV_1$ ), recorded at each clinic visit.

91 patients were randomised to the placebo arm, of whom 38 completed. 90 were randomised to the lowest dose active arm, of whom 72 completed.

# Analyses

Overview

Overview

Principles

Implications

Example

● Asthma study

● **Analyses**

● Jump to active—illustration

● Results

What should we carry forward?

## Using on-treatment data

1. Assume MAR, On-treatment effect:
  - ANCOVA
  - SRM model
  - multiple imputation
2. Sensitivity analyses (estimation using multiple imputation)
  - assume active dropouts copy placebo increments
  - assume placebo dropouts jump to active treatment

Details in Carpenter, Roger & Kenward (2009)[1]

Other options can be specified, depending on previous studies in the area, additional data, etc. These can make use of available 'off-treatment' data.

## Jump to active—illustration

Overview

Overview

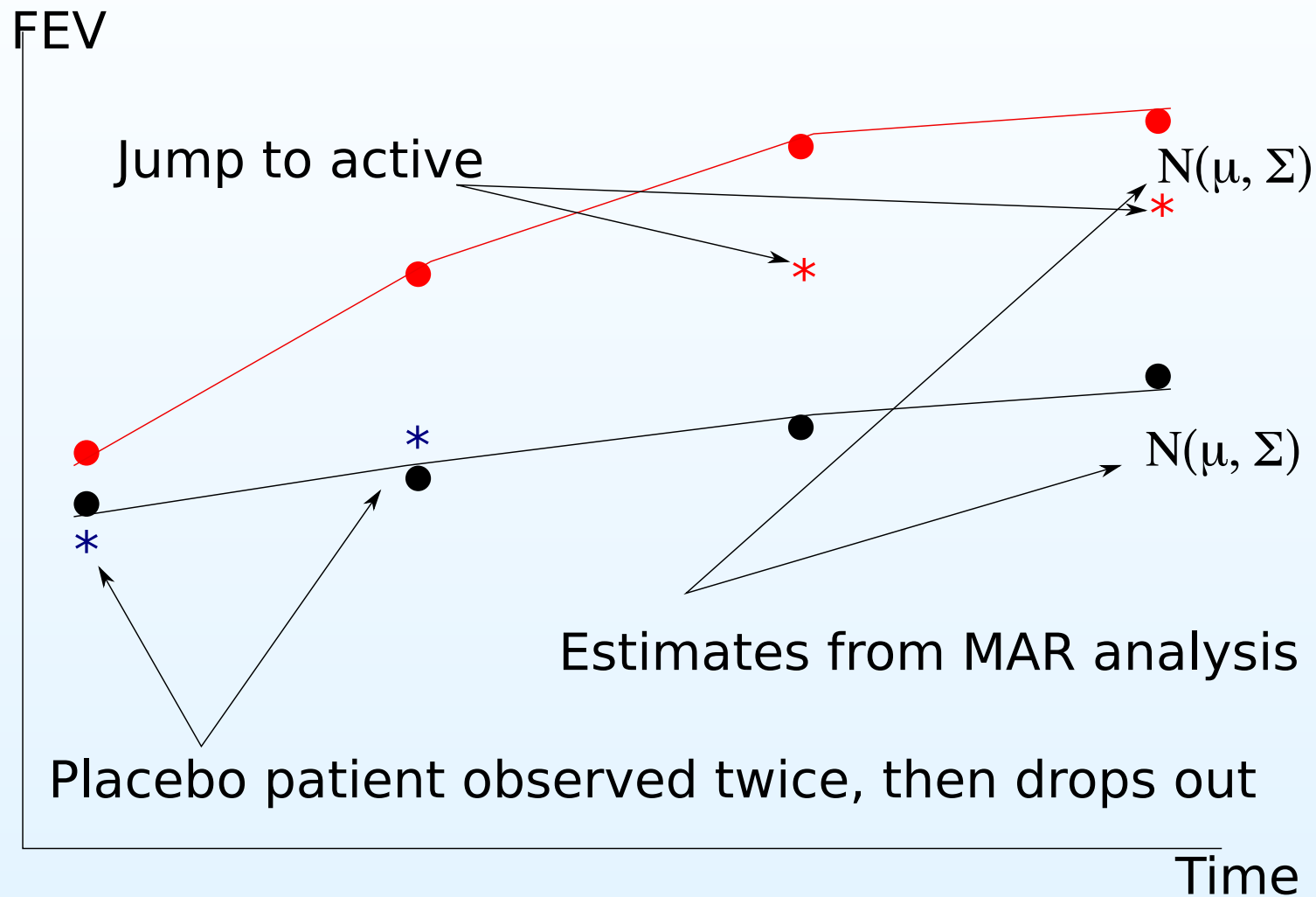
Principles

Implications

Example

- Asthma study
- Analyses
- Jump to active—illustration
- Results

What should we carry forward?



# Results

Overview

Overview

Principles

Implications

Example

- Asthma study
- Analyses
- Jump to active—illustration

● Results

What should we carry forward?

Analysis	Treat est (litres)	Std. Err.	DF (model)	DF (MI)	t-statistic	p-value
ANCOVA	0.239	0.099	106	N/A	2.42	0.0172
SRM model, (joint cov mat)	0.280	0.092	134	N/A	3.05	0.0028
SRM model, (sep cov mat)	0.340	0.101	74	N/A	3.35	0.0013
MI	0.323	0.104	N/A	3039	3.11	0.0019
Copy placebo increments	0.287	0.104	N/A	>4000	2.76	0.0058
Jump to active	0.117	0.101	N/A	996	1.15	0.2499

## General principles

Overview

Overview

Principles

Implications

Example

What should we carry forward?

● **General principles**

● Methods

● References

1. Analysing trials with partially observed data requires making untestable assumptions. Therefore:
  - design to minimise missing data
  - collect off-treatment data where possible
  - pre-specify key assumptions for patients who withdraw for various reasons
2. Apply standard statistical principles:
  - make assumptions
  - obtain valid inference under the assumptions
  - check the robustness of the conclusions as the assumptions change
3. Talking about a ‘conservatively biased’ estimator doesn’t make sense without first specifying the assumptions about the missing data.
4. Specifying ‘LOCF’ and then debating whether it is ‘conservative’ or not in a particular situation leads to unilluminating discussion.

## Methods

Overview

Overview

Principles

Implications

Example

What should we carry forward?

- General principles
- **Methods**
- References

- Think carefully about the question the analysis is addressing.
- Sensitivity analysis is key, and needs to be accessible and relevant:
  - formulating the assumptions through the considering the distribution of the missing data given the observed data (essentially the pattern mixture approach) is accessible to most collaborators
  - the more accessible the assumptions, the more likely the analyses are to be relevant
  - given the assumptions, valid estimation is straightforward via multiple imputation
- ‘Simple’ inverse probability weighting is too inefficient
- There are stimulating developments using doubly robust estimators in this area[3], including doubly robust multiple imputation. More experience is needed before they can be widely recommended in practice.

# References

Overview

Overview

Principles

Implications

Example

What should we carry forward?

- General principles
- Methods
- **References**

- [1] James Carpenter, James Roger, and Mike Kenward. Relevant, Accessible Sensitivity Analyses Using Multiple Imputation *submitted*, 2009.
- [2] R J Cook, L Zeng, and G Y Yi. Marginal analysis of incomplete longitudinal binary data; a cautionary note on locf imputation. *Biometrics*, pages 820–828, 2004.
- [3] M. Davidian, A. A. Tsiatis, and S. Leon. Semiparametric estimation of treatment effect in a pretest–posttest study with missing data (with discussion). *Statistical Science*, 20:261–301, 2005.
- [4] M G Kenward and G Molenberghs. Last Observation Carried Forward: A Crystal Ball? *To appear in the Journal of Biopharmaceutical statistics*, 2009.
- [5] G Molenberghs, H Thijs, I Jansen, C Beunkens, M G Kenward, C Mallinkrodt, and R J Carroll. Analyzing incomplete longitudinal clinical trial data. *Biostatistics*, 5:445–464, 2004.
- [6] Ohidul Siddiqui, H M James Hung, and Robert O'Neill. Mmrm vs. locf: A comprehensive comparison based on simulation study and 25 nda datasets. *Journal of Biopharmaceutical Statistics*, 19:227–246, 2009.