Why missing data should not be ignored and Bayesian methods are good

Alexina Mason

with thanks to my PhD supervisors, Nicky Best, Sylvia Richardson and Ian Plewis, and the BIAS team

8 April 2011

Outline

Why missing data should not be ignored

Why Bayesian methods are good

Antidepressant Clinical Trial Example

Introduction

- Missing data are common
- But usually handled inadequately
- Potentially distort scientific investigation
- Many methods of dealing with missing data exist
- The usefulness/validity of these methods depends on:
  - the cause of the missing data (missing data mechanism)
  - the pattern and extent of the missingness
  - whether outcomes or explanatory variables are missing

We now introduce missing data ideas using a clinical trial example

Motivating example: antidepressant clinical trial

- 6 centre clinical trial, comparing 3 treatments of depression
- 367 subjects randomised to one of 3 treatments
- Subjects rated on Hamilton depression score (HAMD) on 5 weekly visits
  - week 0 before treatment
  - weeks 1-4 during treatment
- HAMD score takes values 0-50
  - the higher the score, the more severe the depression
- Subjects drop out from week 2 onwards (246 complete cases)
- Data were previously analysed by Diggle and Kenward (1994)

Study objective: are there any differences in the effects of the 3 treatments on the change in HAMD score over time?

HAMD example: implications of drop-out

- Before analysing the data, we should consider
  - Why did some subjects drop out of the study?
  - Do the subjects who dropped out have similar characteristics to individuals who remained in the study?
  - How many subjects dropped out of the study each week?
  - Is the level and pattern of drop out consistent across treatments?
- Simple tables and plots can help elucidate
- Experts may provide background information and rationale

HAMD example: missingness level and pattern

Percentage of missingness by treatment and week

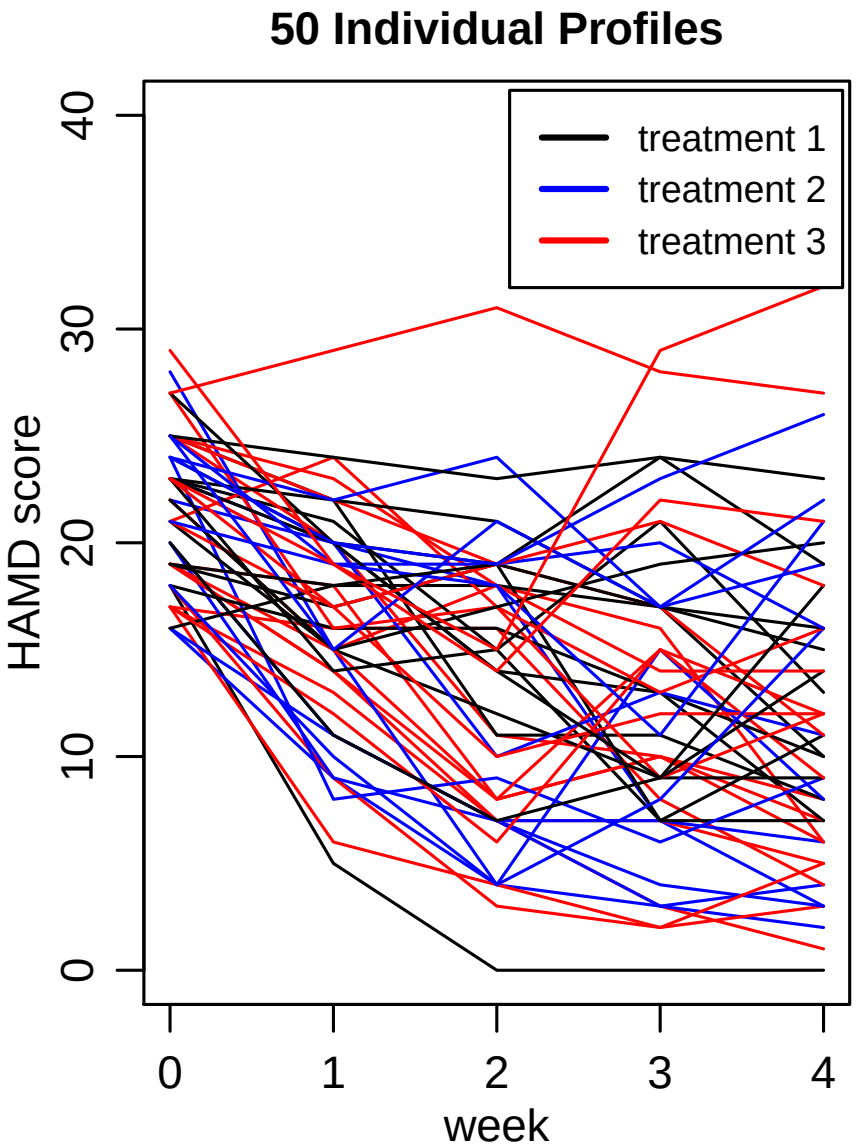
	treat. 1	treat. 2	treat. 3	all treatments
week 0	0.0	0.0	0.0	0.0
week 1	0.0	0.0	0.0	0.0
week 2	11.7	22.0	9.3	14.2
week 3	19.2	29.7	16.3	21.5
week 4	36.7	35.6	27.1	33.0

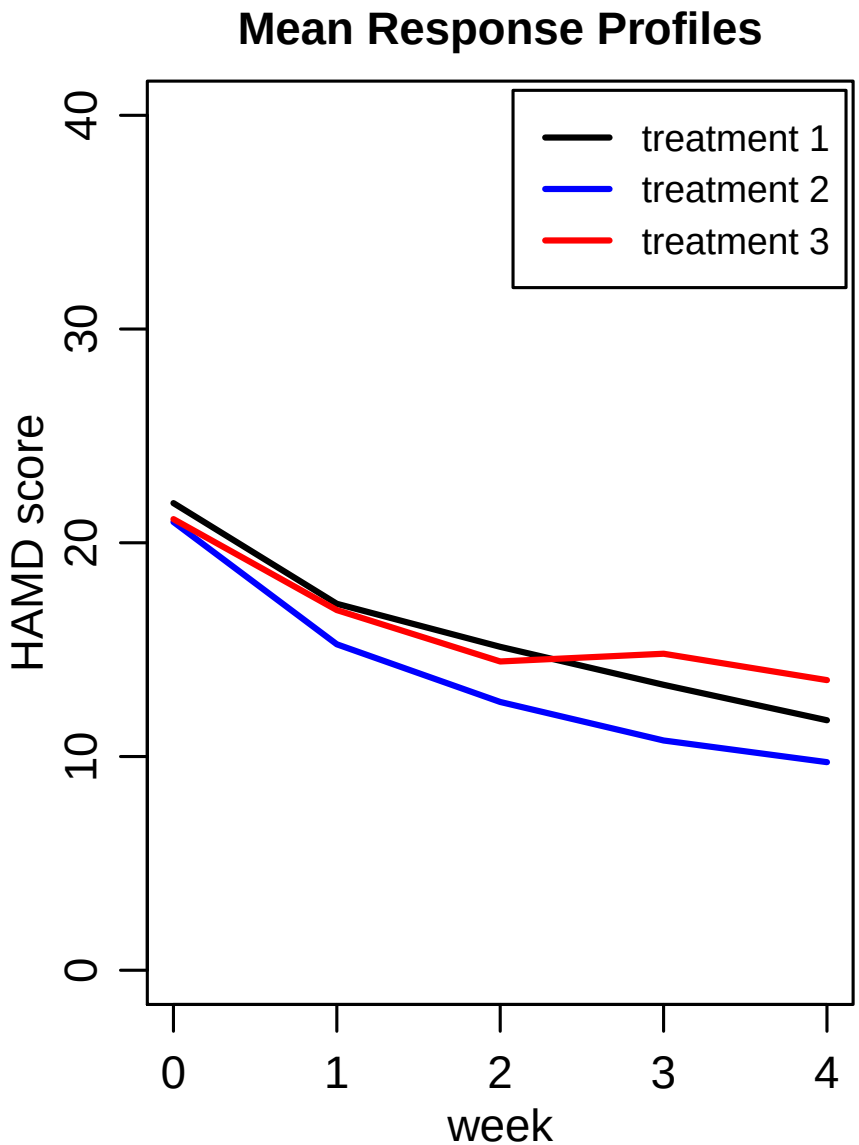
- By the end of the study, one third of subjects drop-out
- Fewer subjects drop out of treatment 3
- Drop-out occurs earlier for treatment 2

Complete Case analysis (CC)

- If we had complete data, we could just fit an appropriate regression model
- However, the missing data complicates this analysis
- One approach is to
  - discard individuals with incomplete information
  - analyse complete cases only
- Advantage: simple
- Disadvantages:
  - loss of precision, as not using all available information
  - often introduces bias (see below)
- Many computer packages do this by default
- Frequently used

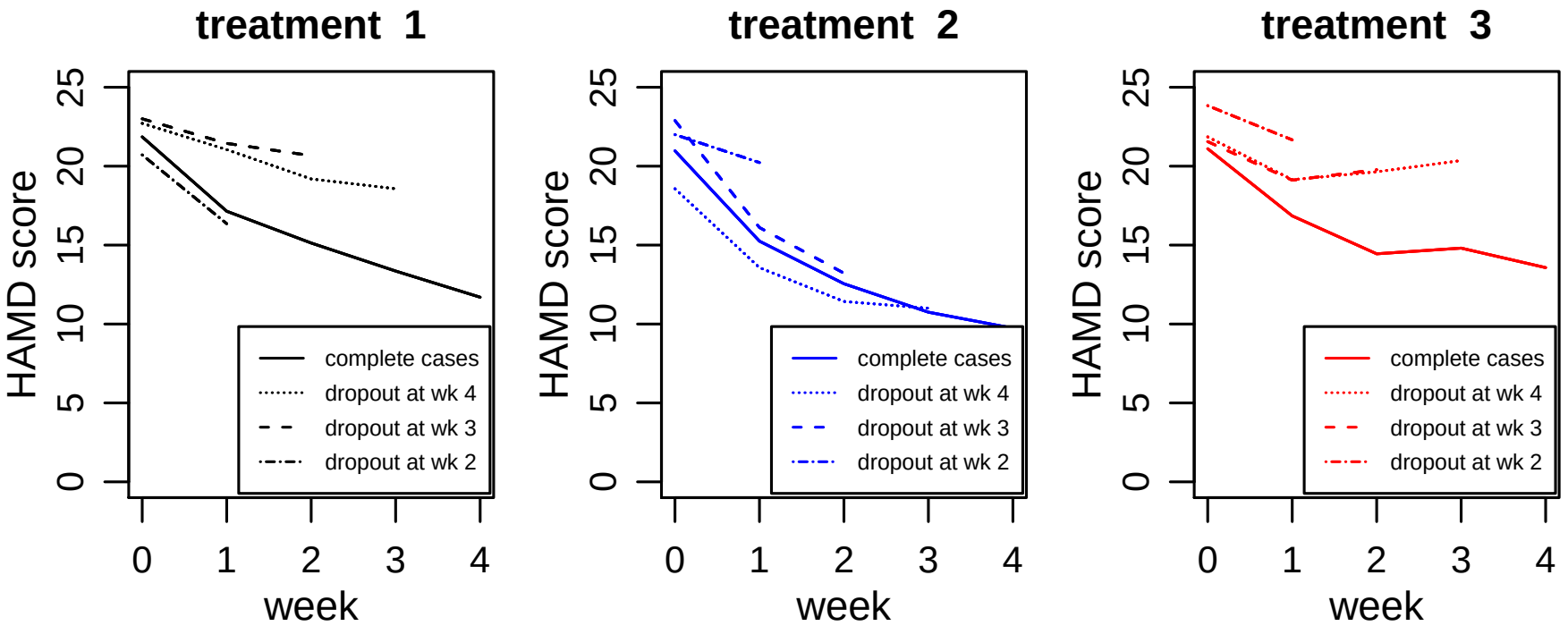
HAMD example: complete cases





HAMD example: all cases

Mean response profiles by drop-out pattern



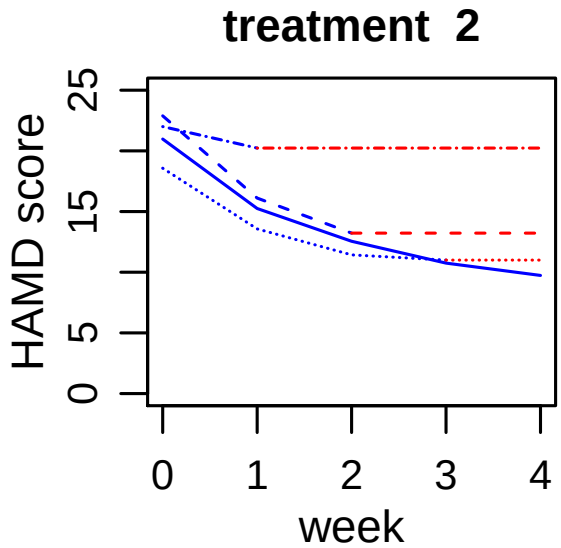
- Individuals allocated to treatments 1 and 3 generally have higher profiles if they dropped out rather than remained in the study.
- But the drop-out and CC profiles are similar for treatment 2

HAMD example: is CC sensible?

- CC loses valuable information from 33% of individuals
- The plots suggest drop-out more likely for treatments 1 and 3 if the score is high
- Hence if we only analyse complete cases we will overestimate the effectiveness of the treatment
- In general, with regression analysis on data with missing outcomes, CC is biased (misleading) if
  - the missingness is related to the outcome values

If CC is not sensible, what are the alternatives?

Last observation carried forward (LOCF)



- Widely used in clinical trial settings
- But makes strong and unrealistic assumptions
- Assumption: all unseen measurements = last seen measurement
- In the HAMD example
  - If the treatment is working but individuals drop out early, we will underestimate the treatment effect
  - LOCF changes the shape of the profile, as drop-outs likely to have improved at least a little
  - This effect will vary by treatment, and hence inference about treatment difference will be misleading
- Also, LOCF overestimates precision as information is 'made up' with no acknowledgement of the uncertainty

CC and LOCF not recommended

To quote Molenberghs et al. (2004)

> LOCF typically produces bias of which the direction and magnitude depend on the true but unknown treatment effects

and

> we have used both formal derivations and case studies to show that there is little justification for analyzing incomplete data from longitudinal clinical trials by means of such simple methods as LOCF and CC

Or as Carpenter (2004) succinctly puts it, we have

the bad - CC; the ugly - LOCF

So what is the good?

Good methods for missing data

- CC and LOCF are examples of 'ad hoc' methods
- 'ad hoc' methods are frequently used but not recommended
- Good methods are 'statistically principled'
- In contrast to ad-hoc methods, principled methods are:
  - based on a well-defined statistical model for the complete data and explicit assumptions about the missing value mechanism
  - the subsequent analysis, inferences and conclusions are valid under these assumptions
  - does not mean the assumptions are necessarily true, but it does allow the dependence of the conclusions on these assumptions to be investigated
- Principled methods include multiple imputation and Bayesian full probability modelling

We now introduce Bayesian analysis

Schools of inference

- There are 3 major schools of inference
  - Frequentist (Neyman-Pearson)
  - Likelihood (Fisher)
  - Bayesian (Thomas Bayes)
- The first two are often combined and the theory taught in basic statistics courses ('classical' statistics)
- Historically, philosophical debates between classical and Bayesian statisticians have sometimes been heated

Bayesian inference

- Bayesian inference distinguishes between
  - observable quantities, i.e. observed data
  - unobserved quantities (e.g. statistical parameters, missing data)
- Unobserved quantities are viewed as unknown with an associated probability distribution
- So Bayesian methods are a very natural way of handling missing data
  - a probability distribution is estimated for each missing value
  - allows uncertainty to be adequately captured
- The reverse viewpoint is that the data is random, but the generating process is fixed and unknown
  - cannot make probability statements about parameters

Credibility vs confidence intervals

- Bayesian inference uses credible intervals to provide a measure of the variability of a parameter (e.g. regression coefficient, θ)
- A 95% credible interval for θ is interpreted as:

  The probability that θ lies in this interval is 95%
- Tells us what we want to know
- By contrast a typical description of a 95% confidence interval from a classical textbook is:

  If we were able to reproduce the experiment under the exact same conditions a large number of times, then the true value of θ would be included in this interval 95% of the time
- Difficult to interpret!

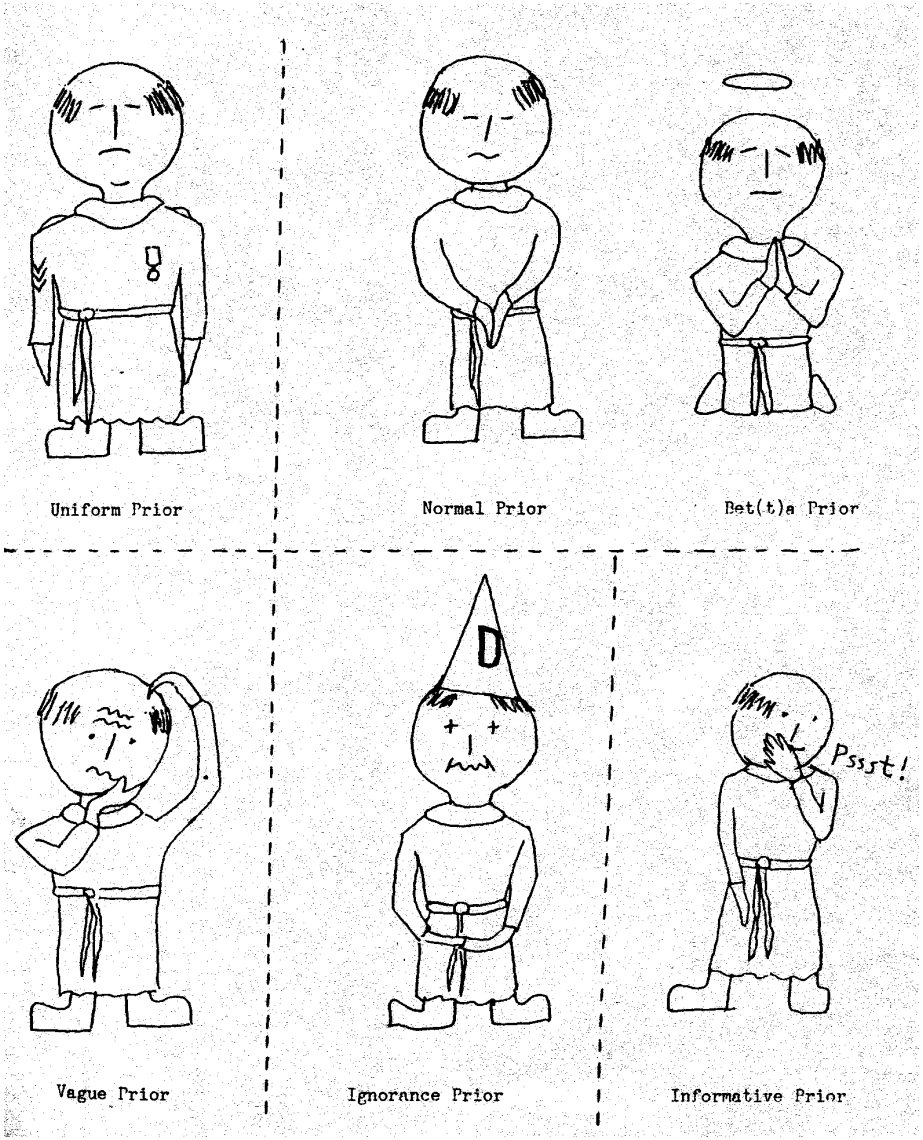
3 components of Bayesian inference

- The prior distribution
  - reflects the plausibility of different values of the unknowns before the data is seen
  - often referred to as the prior
- The likelihood
  - expresses support for different values of the unknowns based solely on the data
  - also used in classical inference
- The posterior distribution
  - combines information in the prior distribution with the likelihood using Bayes theorem
  - is the basis of all Bayesian inference
  - expresses uncertainty about the unknowns after seeing the data

Priors

- The prior can be used to incorporate additional information/knowledge into Bayesian models
- Additional information can come from
  - other studies
  - experts - elicitation
- If extra information exists, it is helpful to use it
- If not, ‘vague’ priors can be used to reflect no knowledge
- Specifying a good prior is non-trivial, but invaluable
- Sensitivity to different priors should always be explored

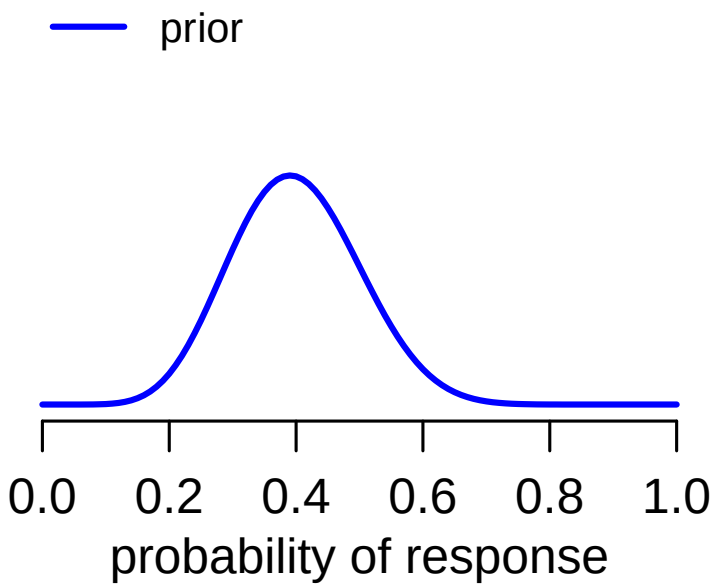
A range of priors



D M Titterington. Bulletin in Applied Statistics 9 (1982) (1), 16-18

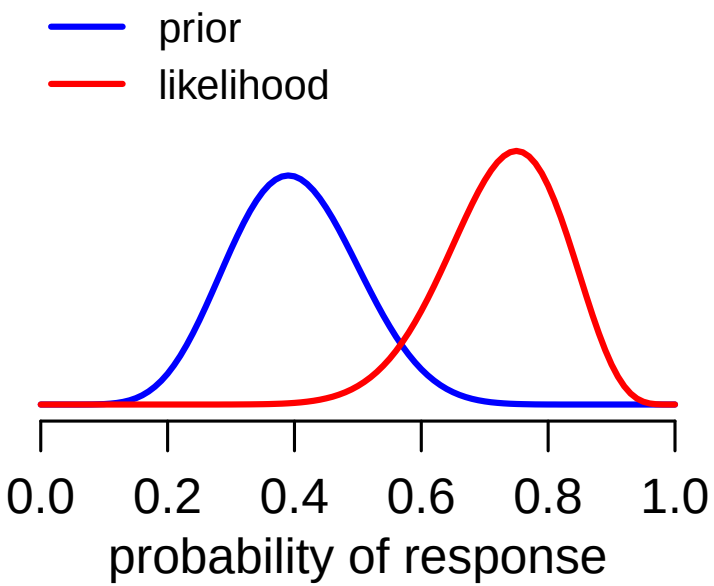
Combining the prior and likelihood

- Suppose we are at an early stage of investigating a new drug
- We can use experience with similar compounds, which suggest response rates between 0.2 and 0.6 are feasible, to create an informative prior



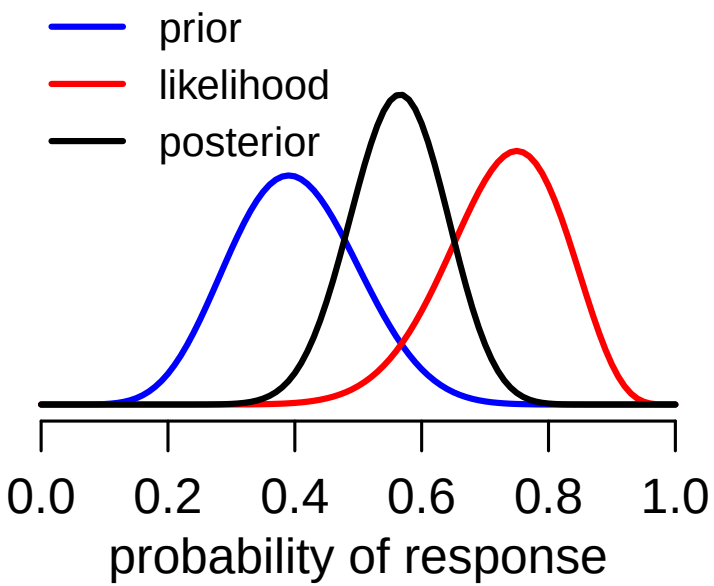
Combining the prior and likelihood

- Suppose we are at an early stage of investigating a new drug
- We can use experience with similar compounds, which suggest response rates between 0.2 and 0.6 are feasible, to create an informative prior
- We then treat 20 volunteers with the new drug and observe 15 positive responses



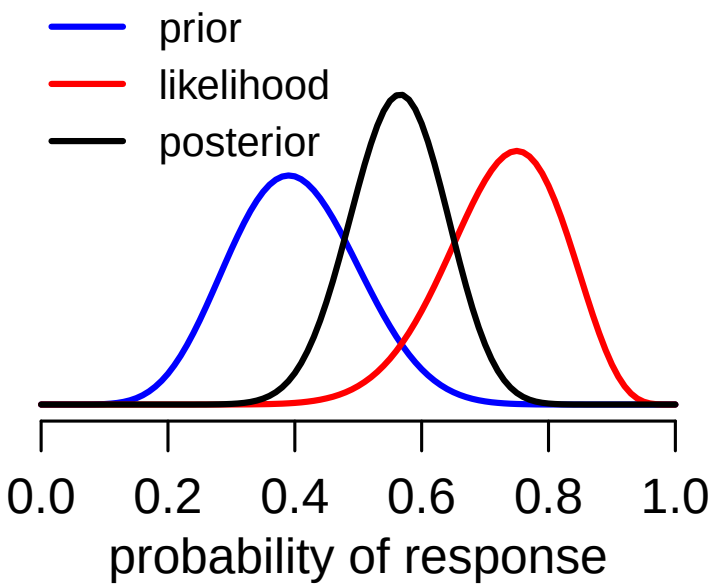
Combining the prior and likelihood

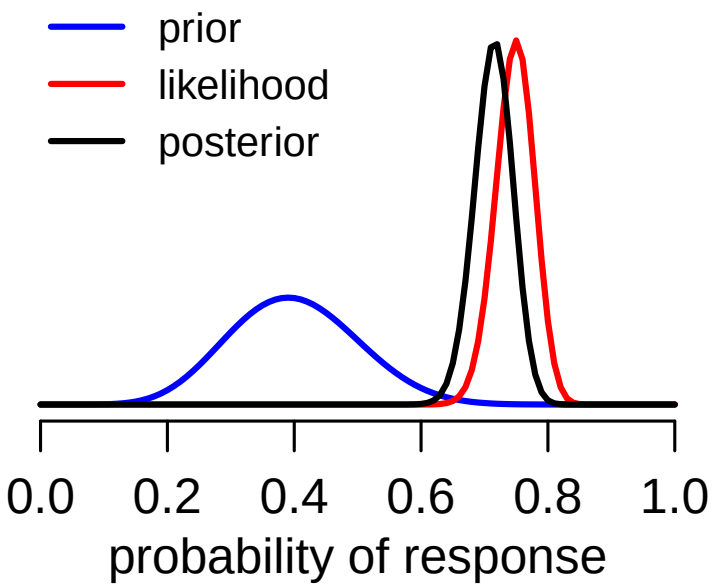
- Suppose we are at an early stage of investigating a new drug
- We can use experience with similar compounds, which suggest response rates between 0.2 and 0.6 are feasible, to create an informative prior
- We then treat 20 volunteers with the new drug and observe 15 positive responses



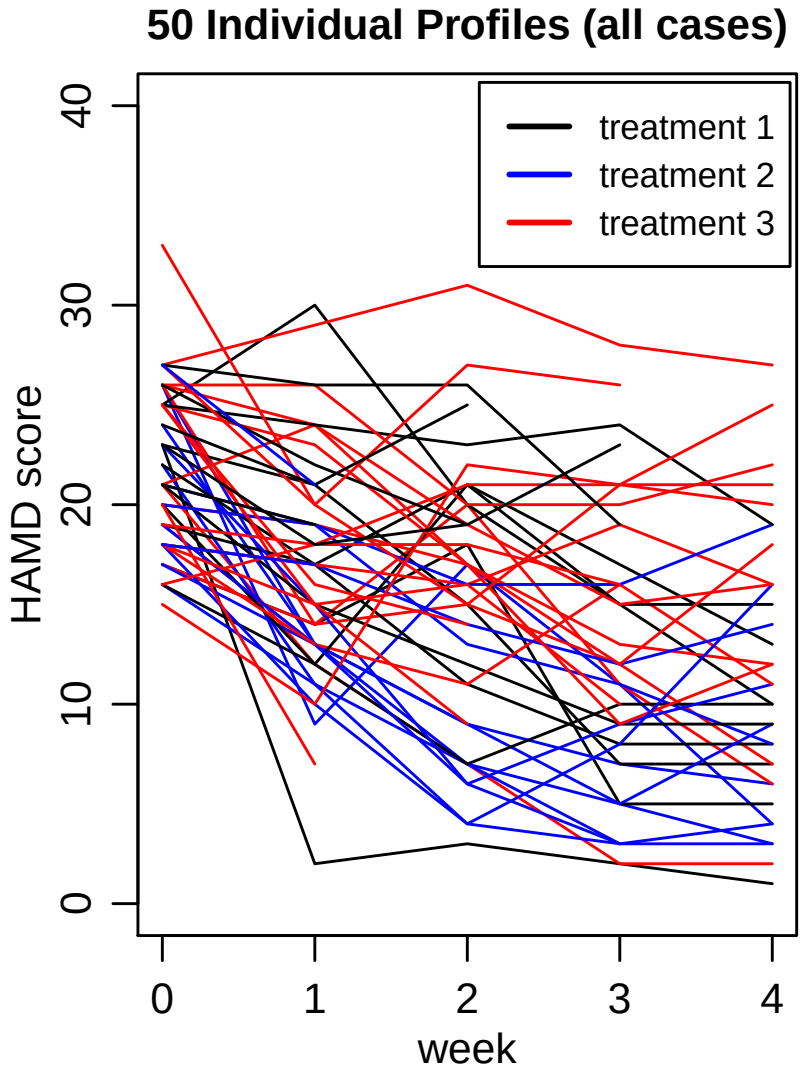
Combining the prior and likelihood

- Suppose we are at an early stage of investigating a new drug
- We can use experience with similar compounds, which suggest response rates between 0.2 and 0.6 are feasible, to create an informative prior
- We then treat 20 volunteers with the new drug and observe 15 positive responses
- For a larger trial (200 individuals), the prior will have less effect





HAMD example: recap



- Clinical trial, comparing 3 treatments of depression
- Study objective: are there any differences in the effects of the 3 treatments on the change in HAMD score over time?
- Subjects drop out from week 2 onwards

HAMD example: analysis model

- To investigate the study objective, we need to regress the HAMD score against week
- For the purposes of exposition, we use a simple model that
  - assumes a linear relationship
  - uses different slope parameters for each treatment
  - takes account of the repeated structure in the data
- Options for allowing for the repeated structure include:
  - incorporating random effects
  - modelling the autocorrelation between weekly visits directly
- The covariates (week and treatment) are fully observed
- Only the outcome, HAMD score, has missing values

HAMD example: model of missingness

- The specification of the model of missingness depends on our assumptions about the missingness mechanism
- We could assume that the probability of drop-out depends on:
  1. HAMD score in the previous week
  2. current HAMD score
  3. change in the HAMD score between previous and current week
- Option 1 depends only on observed data
  - this is ignorable missingness
- Options 2 and 3 depend on observed and missing data
  - this is non-ignorable missingness
  - now we must include a model of missingness - model the HAMD score missingness indicator
- Further variants can be created by allowing different parameters for each treatment

HAMD example: how the Bayesian approach helps

- Although we are not using informative priors in this example, the Bayesian approach still has advantages
- The analysis model and model of missingness are fitted simultaneously as a joint model, ensuring
  - estimation is internally consistent
  - uncertainty is properly taken into account
- Adapting the model to test different assumptions is easy
- All this is possible in the classical framework, but harder

HAMD example: interpretation of results

- Study objective: are there any differences in the effects of the 3 treatments on the change in HAMD score over time?
- So we are particularly interested in the differences in the slope parameters
  - we will call these contrasts
- Interpretation of contrast for treatment A v treatment B:
  - negative values favour treatment A
  - positive values favour treatment B

HAMD example: results from Bayesian Joint Model

- Mimicking classical analysis, we could present a point estimate and credible interval

Table: posterior mean (95% credible interval) for the contrasts from a Bayesian joint model

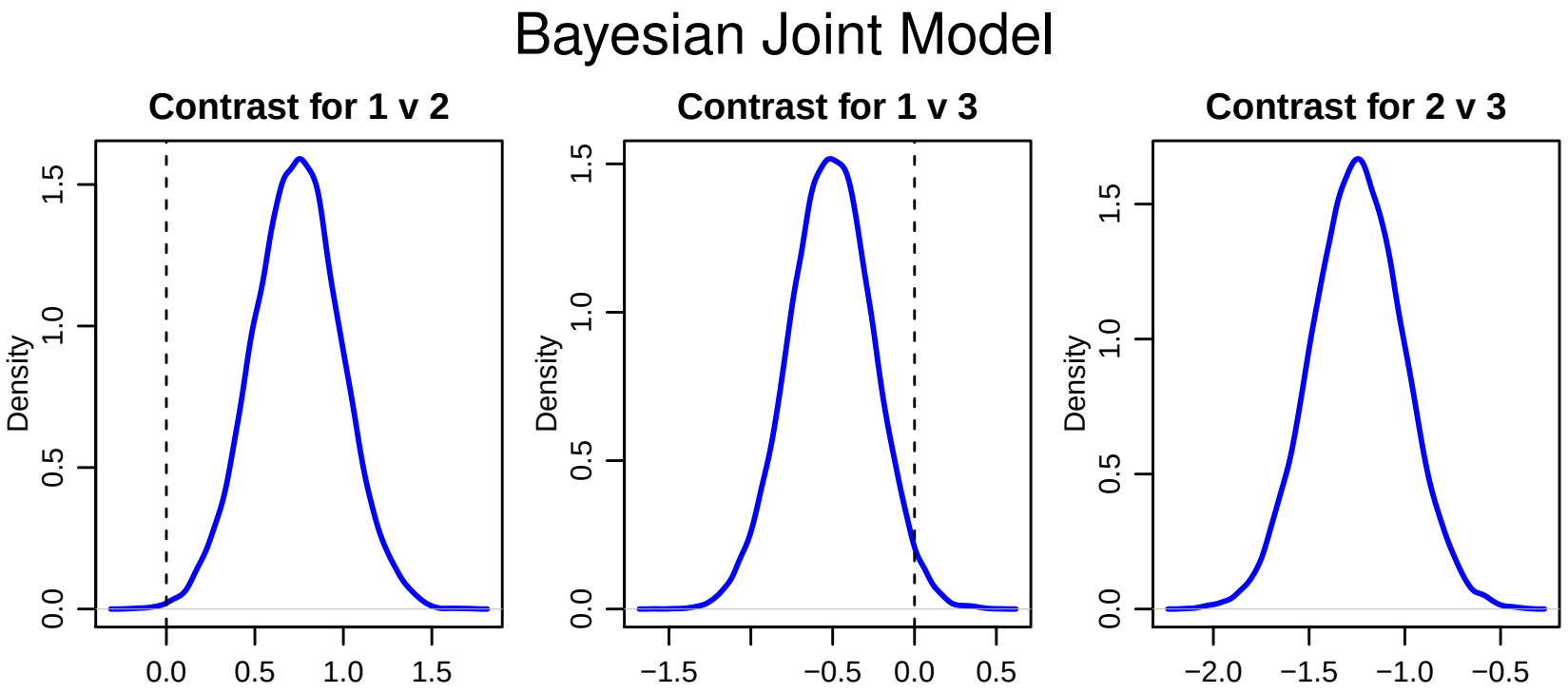
treatment comparison	Bayesian Joint Model*	
1 v 2	0.74	(0.25,1.23)
1 v 3	-0.51	(-1.02,-0.01)
2 v 3	-1.25	(-1.72,-0.77)

* analysis model - random effects; model of missingness - assumes dropout depends on change in HAMD score with different parameters for each treatment

- treatment 2 is more effective than treatments 1 and 3
- treatment 1 is more effective than treatment 3

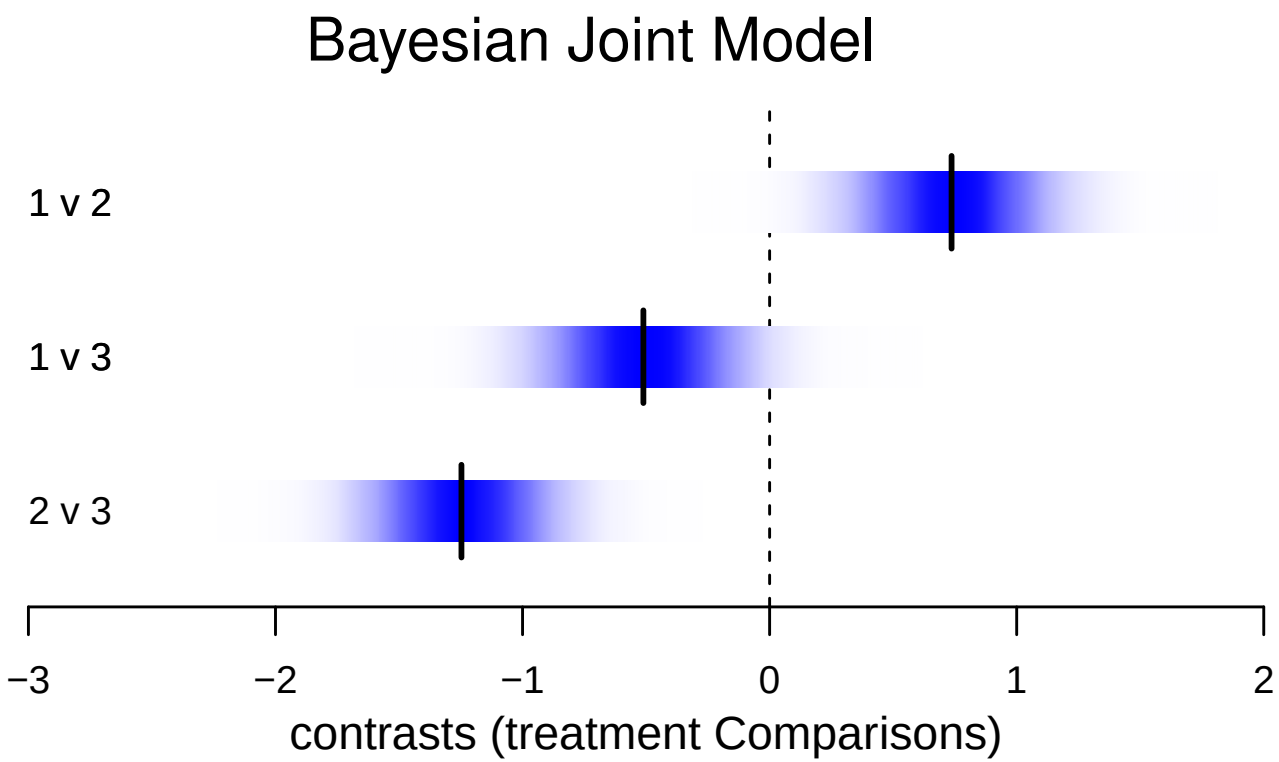
HAMD example: results from Bayesian Joint Model

- However, the reporting of Bayesian analysis is not restricted to point estimates and credible intervals
- A complete probability distribution for each quantity of interest can be presented



HAMD example: results from Bayesian Joint Model

- However, the reporting of Bayesian analysis is not restricted to point estimates and credible intervals
- A complete probability distribution for each quantity of interest can be presented



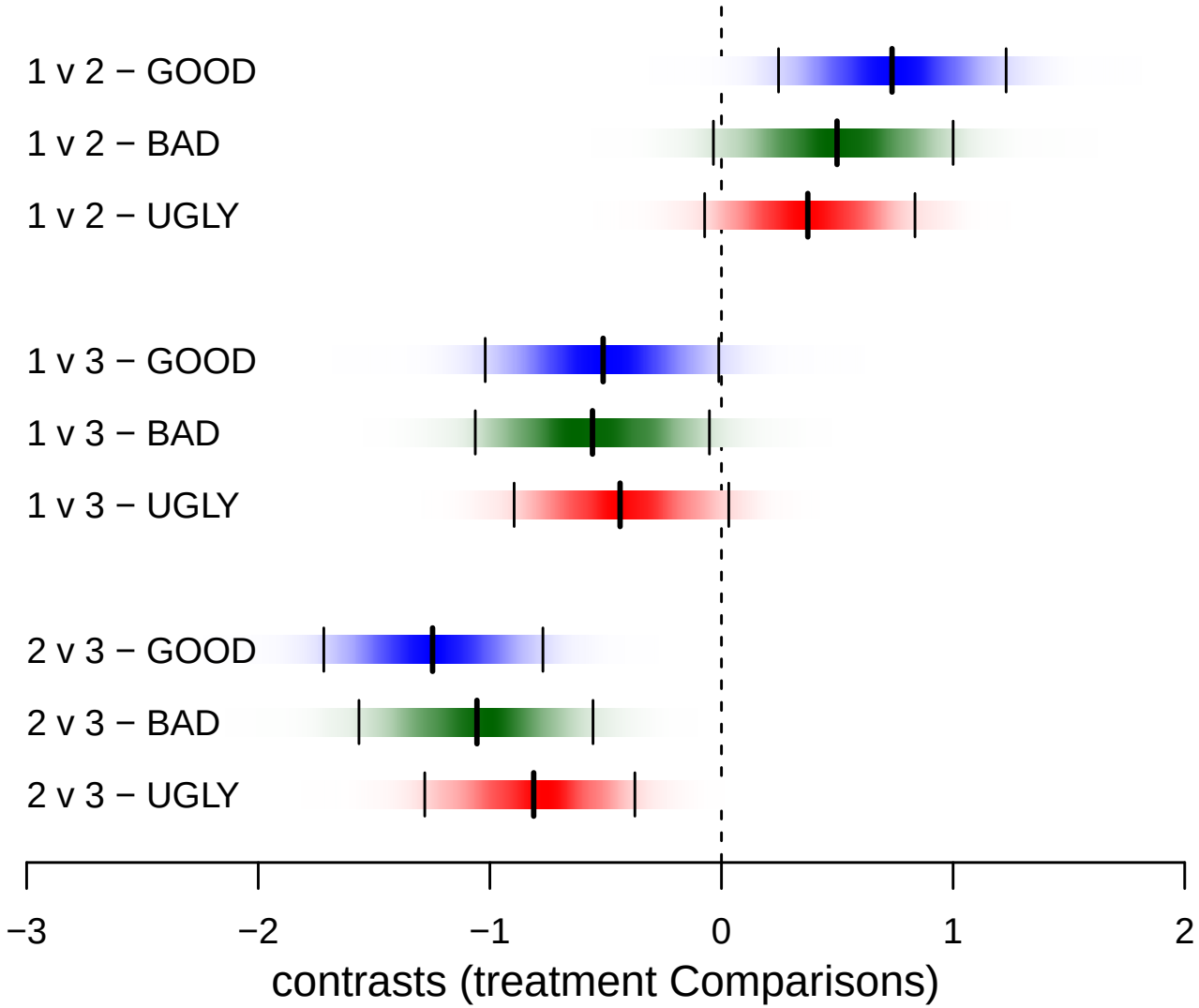
HAMD example: results from Bayesian Joint Model

We can also directly calculate the probability that one treatment is better than another

- Probability treatment 2 more effective than treatment 1 = 0.999
- Probability treatment 1 more effective than treatment 3 = 0.977
- Probability treatment 2 more effective than treatment 3 = 1.000

HAMD example: does the model make a difference?

The Good (Bayesian Joint Model), Bad (CC) and Ugly (LOCF)



Summary

- Ignoring missing data and performing CC implies strong (usually unrealistic) assumptions
- Think carefully about the missing data process
- Bayesian methods provide a natural way of incorporating realistic assumptions
- Missing data adds uncertainty - acknowledge this by presenting results from a range of models based on plausible assumptions

Useful website on missing data:

`www.missingdata.org.uk`

For recent research on Bayesian methods for missing data:

`www.bias-project.org.uk/research.htm`

Sequel

How to deal with missing explanatory variables: application to low birth weight data

References

- Carpenter, J. (2004).
  LOCF - time to stop carrying it forward.
  www.missingdata.org.uk.
  Talk given given to PSI scientific meeting on missing data, London, 20th October 2004.
- Daniels, M. J. and Hogan, J. W. (2008).
  Missing Data In Longitudinal Studies Strategies for Bayesian Modeling and Sensitivity Analysis. Chapman & Hall.
- Diggle, P. and Kenward, M. G. (1994).
  Informative Drop-out in Longitudinal Data Analysis (with discussion).
  Journal of the Royal Statistical Society, Series C (Applied Statistics), **43**, (1), 49–93.
- Little, R. J. A. and Rubin, D. B. (2002).
  Statistical Analysis with Missing Data, (2nd edn). John Wiley and Sons.
- Molenberghs, G. and Kenward, M. G. (2007).
  Missing Data in Clinical Studies, (1st edn). John Wiley and Sons.
- Molenberghs, G., Thijs, H., Jansen, I., Beunckens, C., Kenward, M. G., Mallinckrodt, C., and Carroll, R. J. (2004).
  Analysing incomplete longitudinal clinical trial data.
  Biostatistics, **5**, (3), 445–64.