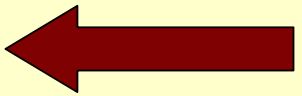
Eliciting and using expert opinions about informatively missing outcome data in clinical trials

Ian White
MRC Biostatistics Unit, Cambridge, UK

Bayes working group
German Biometric Society
Köln, 3 December 2004

Why do Bayesian analyses?

- To make computation easier / possible
  - MCMC, BUGS
- To incorporate prior beliefs
  - on parameters of interest
    - treatment effect
  - on nuisance parameters
    - characteristics of non-responders

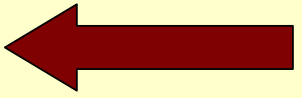


Missing data in randomised trials

Power / precision

- Loss of data → loss of power
- Inappropriate analysis may lose more power

Bias

- Missing outcomes → potential bias 
- Missing baselines → no bias
  (White & Thompson, in press)

I'll focus on RCTs, but the methods apply equally well to observational studies

Plan

1. Handling of missing outcomes in medicine
2. Missing data assumptions
3. Bayesian model allowing for informative missingness
4. QUATRO trial: elicitation
5. Peer review trial: elicitation & analysis
6. Binary outcomes and meta-analysis
7. Practicalities and discussion

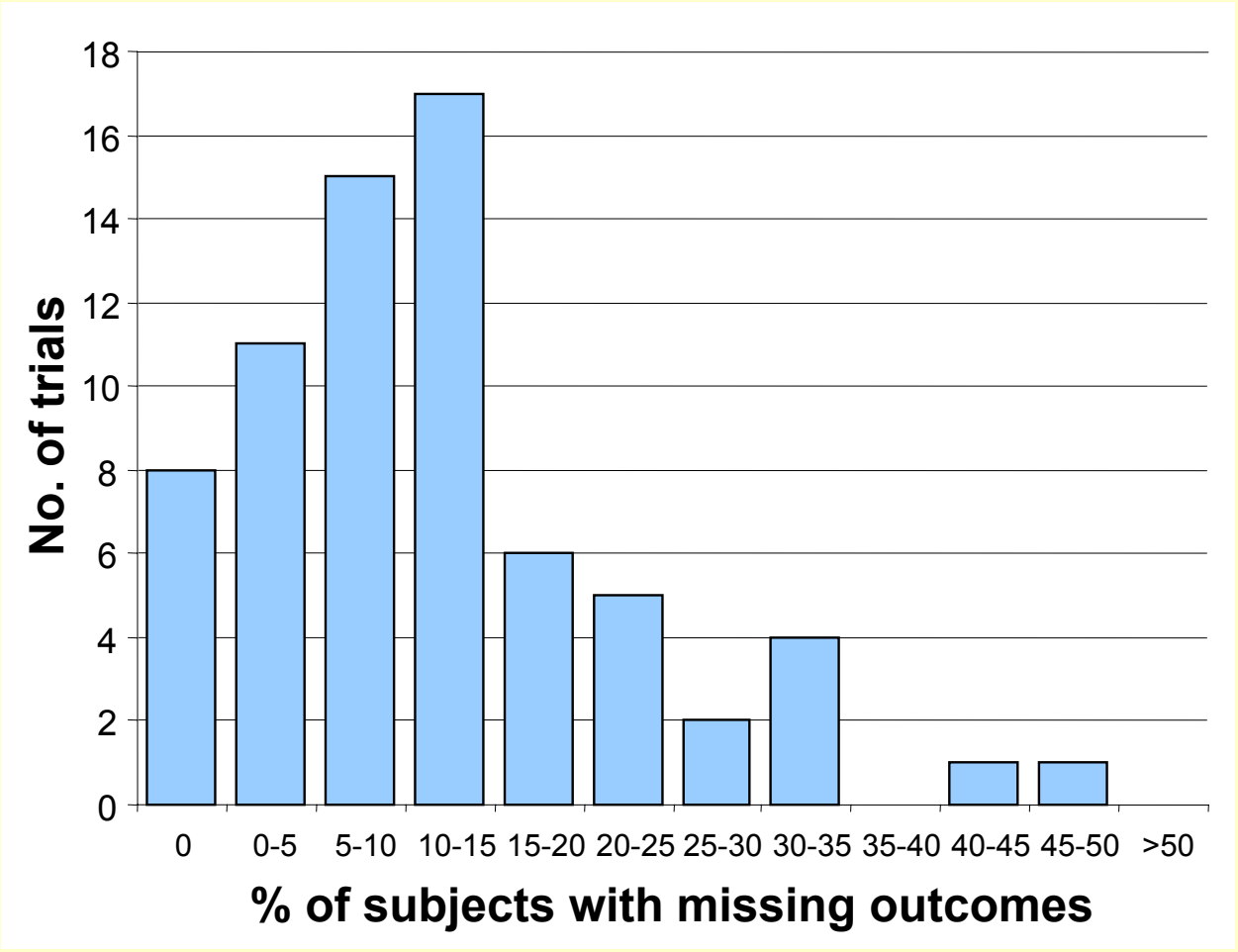
1. Handling of missing outcomes in medicine

With Angela Wood and Simon Thompson (BSU)

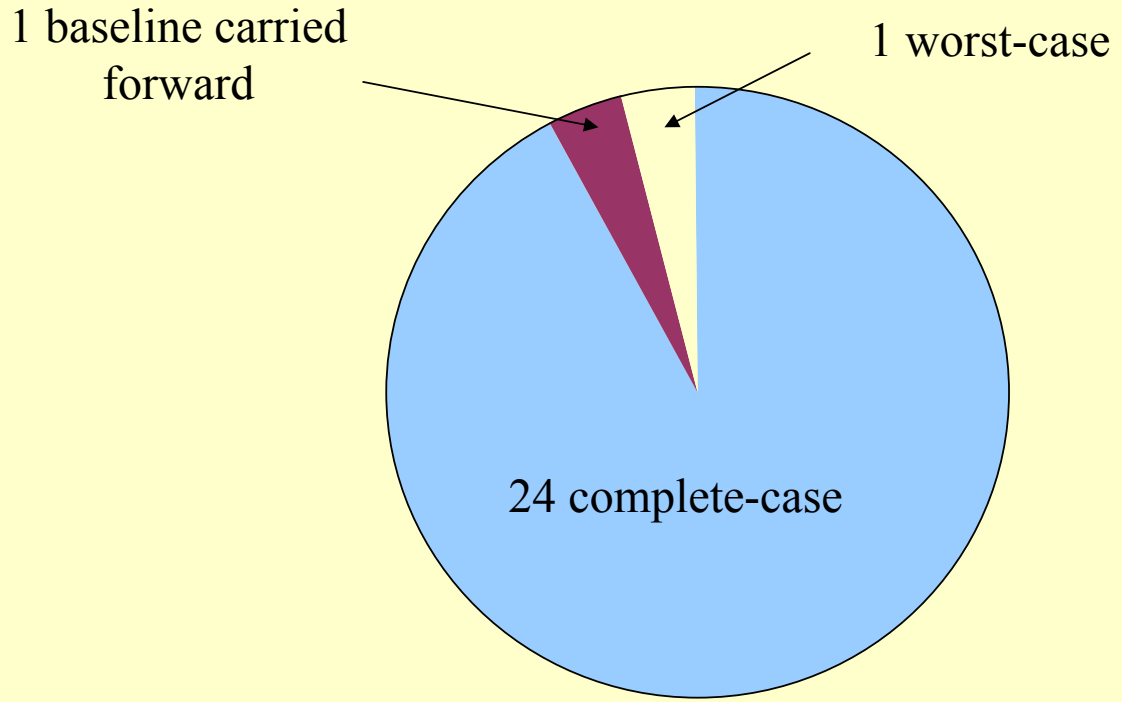
Survey of current practice

- 71 trials published in 4 major medical journals, July - December 2001.
- 63 had missing outcomes
- 61 described handling of missing data
- 35/61 had an outcome measured repeatedly
- Interest always lay in the treatment effect on the *final* outcome
- Wood et al, Clinical Trials 2004.

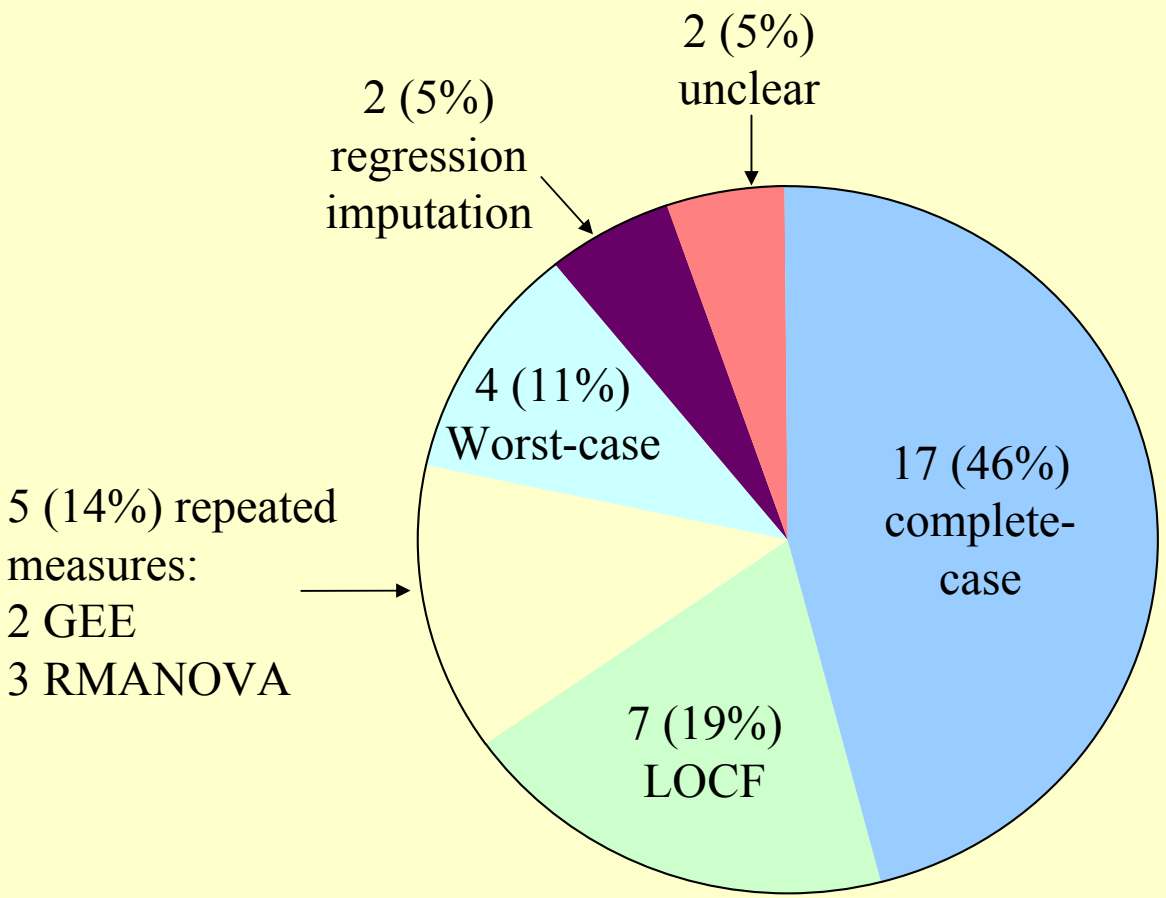
Missing data in 71 trials



26 trials with single outcome



37 trials with repeated measures



Excludes participants with intermediate outcome but no final outcome

What should be done?

3 principles:

- Intention to treat
- State and justify assumptions
- Do sensitivity analysis

Intention to treat principle

- “Subjects allocated to an intervention group should be followed up, assessed and analysed as members of that group irrespective of their compliance to the planned intervention” (ICH E9, 1999).
- Not clear what this means with missing outcomes

Comment: inclusion

- Trials aren't at present including all individuals in the analysis
- Excluding individuals with no outcome data is understandable
  - but may still cause bias
- Excluding individuals with some outcome data (in repeated measures case) is clearly wrong
  - easy to improve practice

Comment: LOCF

- Includes everyone in the analysis
- But makes an implausible assumption:
  - mean outcome after dropout = mean outcome before dropout *in those who drop out*
- Including everyone isn't enough
  - must consider what assumptions the analysis is making
- Some people argue LOCF is conservative

2. Missing data: assumptions

Missing data mechanisms
(Little, 1995)

- Outcome Y (single/repeated), missing indicator M, covariates X
- Missing completely at random (MCAR):
  $M \perp\!\!\!\perp X,Y$ $\perp\!\!\!\perp$ - is independent of
- Covariate-dependent missing completely at random (CD-MCAR): $M \perp\!\!\!\perp Y \mid X$
- Missing at random (MAR): $M \perp\!\!\!\perp Y^{miss} \mid Y^{obs}, X$
- Informative missing (IM): $M \sim Y^{miss} \mid Y^{obs}, X$

Complete Cases

same if single outcome

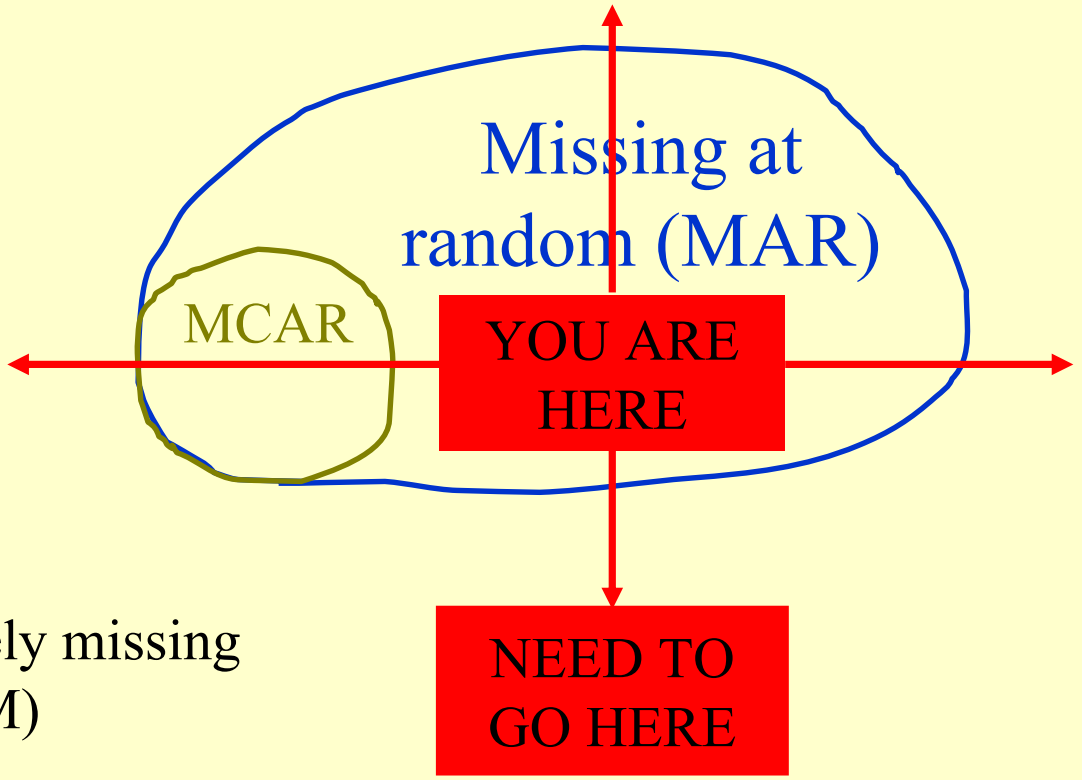
RMANOVA

Is MAR analysis enough?

- Suppose we analyse 60 individuals & find
  - treatment effect +7
  - standard error 3.
- Is this more convincing if
  - These are all 60 randomised, or
  - These are the 60 complete cases out of 80 randomised?

Equally convincing only if we ***know*** data are MAR.

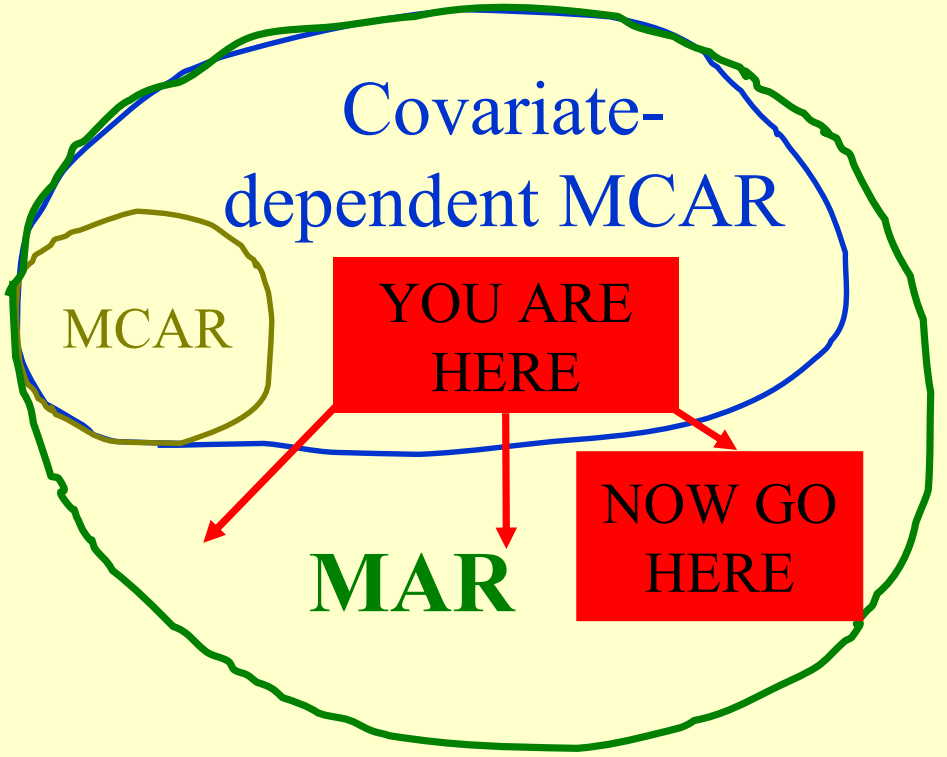
Assumptions – single outcome



Informatively missing (IM)

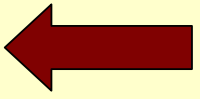
Assumptions – repeated outcome

Informatively missing (IM)



How do we go beyond MAR analysis?

1. Estimate informative missingness using number of failed attempts to collect data
   - Wood et al, submitted.
2. Model missingness and outcome jointly
   - e.g. missingness ~ outcome via random effects (Henderson et al, 2000)
3. Proxy outcomes / intensive follow-up
4. Use prior beliefs on informative missingness (Rubin, 1977)



3. Bayesian model allowing for informative missingness

With James Carpenter (LSHTM)

Quantifying informative missingness

- Focus on designs with a *single* quantitative outcome.
  - Y = outcome (possibly unobserved)
  - M = missingness
  - R = randomised group
- MAR: $M \perp\!\!\!\perp Y \mid R$
- Two approaches:
  - Selection model
  - Pattern mixture model

Selection model approach

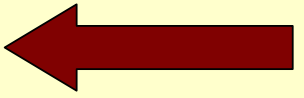
- Imagine regressing M on Y (and R)
  - examples:
  - logit P(M|Y,R) = -1+0.2R
  - logit P(M|Y,R) = -1+0.5Y
  - logit P(M|Y,R) = -1+0.5Y+0.2R–0.3YR
- Need to specify the log odds ratio for missingness for a 1-unit increase in outcome (within trial arms)

Pattern mixture model approach

- Imagine regressing Y on M (and R)
  - $E(Y|M,R) = 120+2R$
  - $E(Y|M,R) = 120+2R+7M$
  - $E(Y|M,R) = 120+2R+7M–3MR$
- Need to specify the difference between mean observed outcome and mean missing outcome
  - within trial arms

Question

- Which approach would *you* find easier to use?
- Selection model:
  - (log) odds ratio for missingness for a 1-unit increase in outcome (within trial arms)
- Pattern mixture model:
  - difference between mean observed outcome and mean missing outcome (within trial arms)



IM pattern mixture model

$r = 0/1$ indexes randomised arms.

In complete cases: $\mathrm{Y} = N(\mu_r^{CC}, \sigma^2)$

$$\Delta^{CC} = \mu_1^{CC} - \mu_0^{CC}$$

In missing cases: $\mathrm{Y} = N(\mu_r^{CC} + \delta_r, *)$

δ_r = informative missingness (unobserved)

Then true mean $\mu_r = \mu_r^{CC} + \alpha_r \delta_r$ where $\alpha_r = P(\text{missing})$

And $\Delta \equiv \mu_1 - \mu_0 = \Delta^{CC} + \alpha_1 \delta_1 - \alpha_0 \delta_0$

Note

- I allow the informative missingness, δ, to differ between arms
- e.g. dropout after health advice may be more informative than after control intervention

Bayesian analysis

Elicit informative prior for δ_0, δ_1 :

- e.g. bivariate normal.

Reference prior for $\mu_0^{CC}, \mu_1^{CC}, \alpha_0, \alpha_1$.

Easy to analyse e.g. in WinBUGS

- fit model and monitor $\Delta = \Delta^{CC} + \alpha_1\delta_1 - \alpha_0\delta_0$

Approximate bayesian analysis

Recall $\Delta = \Delta^{CC} + \alpha_1\delta_1 \text{ - } \alpha_0\delta_0$

Posterior means of $\Delta^{CC}, \alpha_0, \alpha_1 \approx$ MLEs $\hat{\Delta}^{CC}, \hat{\alpha}_0, \hat{\alpha}_1$

independent of δ_0, δ_1

So posterior mean of Δ is approximately

$$\hat{\Delta}^{CC} + \boxed{\hat{\alpha}_1 E[\delta_1] - \hat{\alpha}_0 E[\delta_0]}$$

Correction to point estimate

Posterior variance of Δ is approximately

$$\hat{\text{var}}(\hat{\Delta}^{CC}) + \boxed{\hat{\alpha}_1^2 \text{ var}(\delta_1) + \hat{\alpha}_0^2 \text{ var}(\delta_0) - 2\hat{\alpha}_0\hat{\alpha}_1 \text{ cov}(\delta_0, \delta_1)}$$

Correction to variance

Special case

- If δ's have same distribution in both arms, posterior of Δ has

$$\text{mean} = \hat{\Delta}^{CC} + E[\delta](\hat{\alpha}_1 - \hat{\alpha}_0)$$

$$\text{variance} \approx \hat{\text{var}}(\hat{\Delta}^{CC}) + \text{var}(\delta)\{(\hat{\alpha}_1 - \hat{\alpha}_0)^2 + 2(1-c)\hat{\alpha}_0\hat{\alpha}_1\}$$

- c = corr(δ_0,δ_1) in prior
- Often α's are similar, so c drives variance. Smaller c → more uncertainty.

$\alpha_r = P(\text{missing}) \text{ in arm } r$

$\delta = \text{informative missingness}$

$\Delta = \mu_1 - \mu_0$

$\Delta^{CC} = \mu_1^{CC} - \mu_0^{CC}$

What is c?

- Correlation of δ_0 and δ_1 in the prior
- c=1: you are certain that $\delta_0 = \delta_1$
- c=0: if I could tell you the value of δ_1, you wouldn't change your beliefs about δ_0.

4. Example: QUATRO

QUATRO trial: design

- Patients with schizophrenia are often on long-term anti-psychotic therapy
- Stopping therapy is a common cause of relapse
- QUATRO is evaluating the use of counselling ("adherence therapy") to improve psychotic patients' adherence to medication.
  - 4 centres: London, Leipzig, Verona, Amsterdam.
- Primary outcome: self-reported quality of life at 1 year.

QUATRO trial: missingness

- Concern that missing data may induce bias – nonresponse likely to be related to increased symptom severity
- I designed a questionnaire about informative missingness
  - completed (by email) by each of 4 centres
  - before data collection

Eliciting informativeness in QUATRO

QUATRO **adherence therapy** arm: comparing mean MCS for patients who do not respond to the final questionnaire compared with those who do respond.										
	Non-responders worse than responders by				Non-responders same	Non-responders better than responders by				TOTAL
	13 or more	9-12	5-8	1-4		1-4	5-8	9-12	13 or more	
Your answers										0
Hypothetical example	0	25	25	0	0	0	25	25	0	100

MCS: mental component score of SF36 (SD=10)

Response, pooled over centres

QUATRO **adherence therapy** arm: comparing mean MCS for patients who do not respond to the final questionnaire compared with those who do respond.

	Non-responders worse than responders by				Non-responders same	Non-responders better than responders by				TOTAL
	13 or more	9-12	5-8	1-4		1-4	5-8	9-12	13 or more	
Your answers	5	18	20	18	24	9	4	2	1	100
Hypothetical example	0	25	25	0	0	0	25	25	0	100

Mean -3.5, SD 6.2

Expect non-responders to have worse QoL than responders

Eliciting correlation c in QUATRO

Question 3: Both arms together

What I really need to know is how **similar** are your beliefs about the two arms.

You have said:		
	In the control arm:	In the adherence therapy arm:
The most likely non-responder / responder difference is	-3	-4
and the largest possible difference is about		
non-responders worse	-16	-16
non-responders better	16	16
(positive/negative values indicate non-responders having better/worse quality of life than responders)		
How closely related are your beliefs about the two arms?		
If I told you the non-responder / responder difference in the control arm really was as large as 16,		
what would be your best guess for the non-responder / responder difference in the adherence therapy arm?		
would it still be	-4	(information about one arm tells you nothing about the other arm)?
or would it change to	16	(information about one arm tells you everything about the other arm)?
or somewhere in between?		
Please enter your best guess in this case:		

QUATRO: elicited correlations

- Correlations were 0, 0.1, 0.7 and 1 in the 4 trial centres
- Does this reflect
  - genuine divergence?
  - question too hard?
  - instrument invalid?
- Will probably use an average value in analysis
- Trial is still in progress

An unanticipated result

- Centre: “Why are you asking us to guess about the missing data? Why don’t we just collect them?”
- Me: “???”
- Centre devised a short questionnaire to get patients’ QoL from their care-givers

5. Example: Peer Review Trial

Schroter et al, 2004

Peer review trial

- Does training reviewers improve the quality of their reviews?
- Reviewers for the British Medical Journal completed a “baseline” review, then randomised to
  - face-to-face training
  - postal training
  - no training
- Outcome = quality of a subsequent review (rating scale)

Results from peer review trial

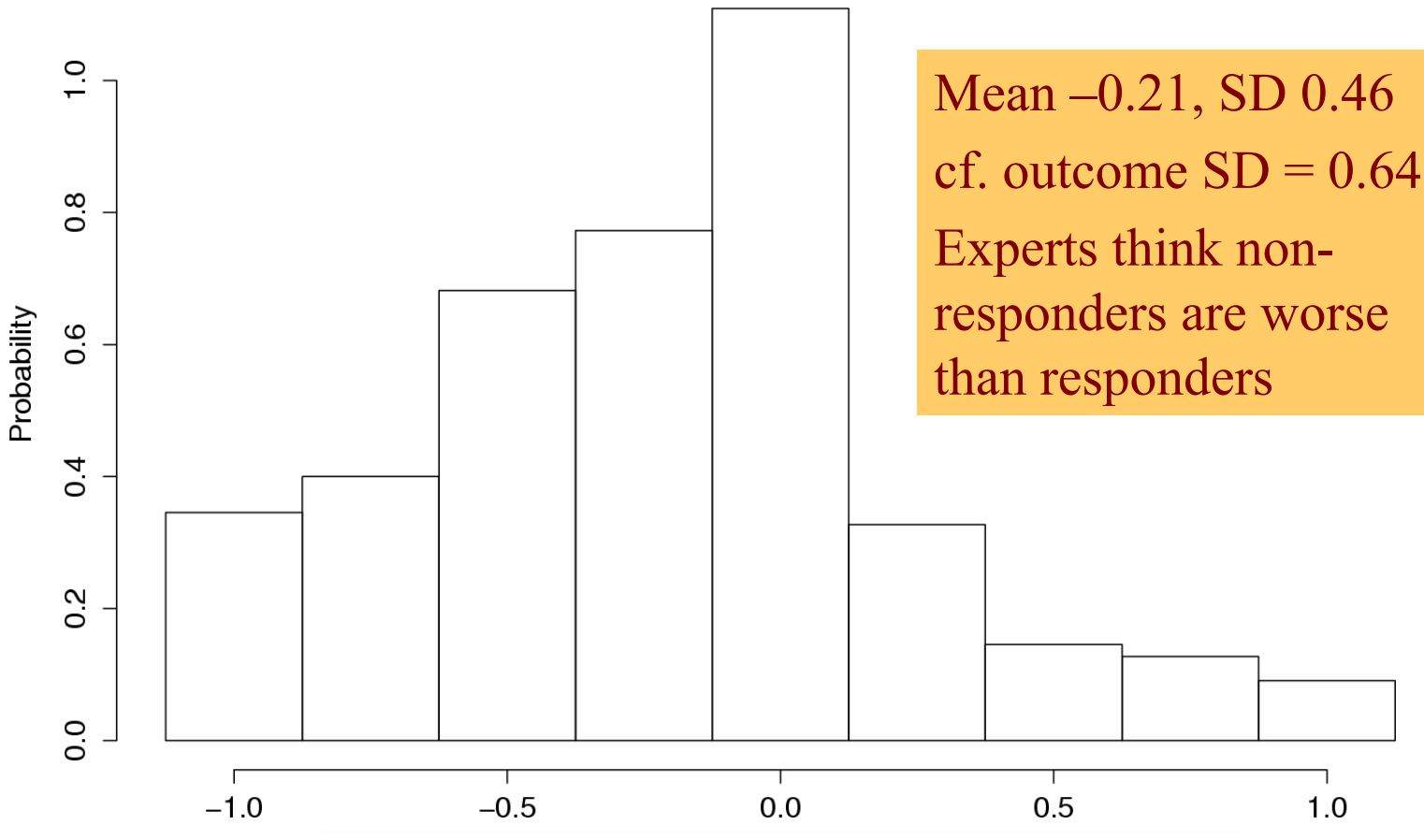
	Control	Postal	Face-to-face
Total n	173	166	183
Missing outcome	6%	28%	14%
Mean of observed outcomes	2.56	2.85	2.72
SD of observed outcomes	0.64	0.64	0.63

Imbalance in missing data led to concerns about bias

Eliciting prior

- Similar to QUATRO questionnaire
- Completed by 22 BMJ staff
  - after data collection, but blind to data
- 3 δ's (1 per arm)
  - Same prior assumed for all
- Failed to elicit correlation between δ's
  - will take values 0, 0.5, 1

Pooled prior



Mean –0.21, SD 0.46

cf. outcome SD = 0.64

Experts think non-responders are worse than responders

Analysis

1. Approximate Bayesian analysis, fitting Normal distribution to prior
2. Exact Bayesian analysis, using prior as elicited (WinBUGS)

Results from peer review trial: postal vs control

Posterior:		mean	SD	95% interval	
Complete cases		0.291	0.077	0.140	0.442
Informative missing	c=1	0.246	0.126	-0.001	0.493
	c=0.5	0.246	0.140	-0.028	0.520
	c=0	0.246	0.153	-0.053	0.545

Compare approximation with full MCMC

	Posterior:	mean	SD	95% interval	
c=1	Approximate	0.246	0.126	-0.001	0.493
	MCMC	0.246	0.126	0.004	0.505
c=0	Approximate	0.246	0.153	-0.053	0.545
	MCMC	0.246	0.151	-0.042	0.564

Approximation works very well

Extensions: covariate

- Can extend the model to allow missingness and outcome to depend on X
- Missingness varies with X → true treatment effect varies with X
  - Compute average treatment effect over X
- Modify approximate formulae:
  - complete cases analysis is ANCOVA
  - prior on δ_0, δ_1 should be conditional on X

Extensions: longitudinal data

- Need prior for missing/observed differences within previous response patterns
- Take these differences as perfectly correlated

6. Binary outcomes and meta-analysis

With Julian Higgins and Angela Wood (BSU)

Trial with binary outcome

In each arm define

- π_O = observed success fraction
- π_U = success fraction in those with missing outcome (unobserved)

Complete cases analysis: assume $\pi_U = \pi_O$

Sometimes reasonable to assume $\pi_U = 1$

e.g. trial of smoking cessation or TB treatment

Worst case analysis: assume $\pi_U = 0$ in one arm, $\pi_U = 1$ in the other.

Quantifying informativeness

π_O = observed success fraction

π_U = unobserved success fraction

Informative Missing Odds Ratio:

$$\text{IMOR} = \frac{\pi_U}{1-\pi_U} \Big/ \frac{\pi_O}{1-\pi_O} \text{ within trial arm.}$$

Can estimate π_O and α = missing fraction.

Given IMOR, can estimate π_U

& hence overall $\pi = (1-\alpha)\pi_O + \alpha\pi_U$

Model for uncertain IMOR

1 = experimental arm, 0 = control arm

α_1, α_0 = proportions missing in the two arms

δ_1, δ_0 = log(IMOR)

- here take mean 0 but don't have to

OR^{CC} = odds ratio from complete cases

OR = odds ratio allowing for non-response

Approximate results

Taylor series expansion gives

$$\log OR \approx \log OR^{CC}$$

$$\text{var}(\log OR) \approx \text{var}(\log OR^{CC}) + \alpha_1^2 \text{ var}(\delta_1) + \alpha_0^2 \text{ var}(\delta_0) - 2\alpha_0\alpha_1 \text{ cov}(\delta_0, \delta_1)$$

- Variance is inflated (Forster & Smith, 1998; Higgins et al, submitted).
- Can also work with RR – formula slightly nastier.
- Non-linear model: more approximate than before.
- Can do exact analysis.

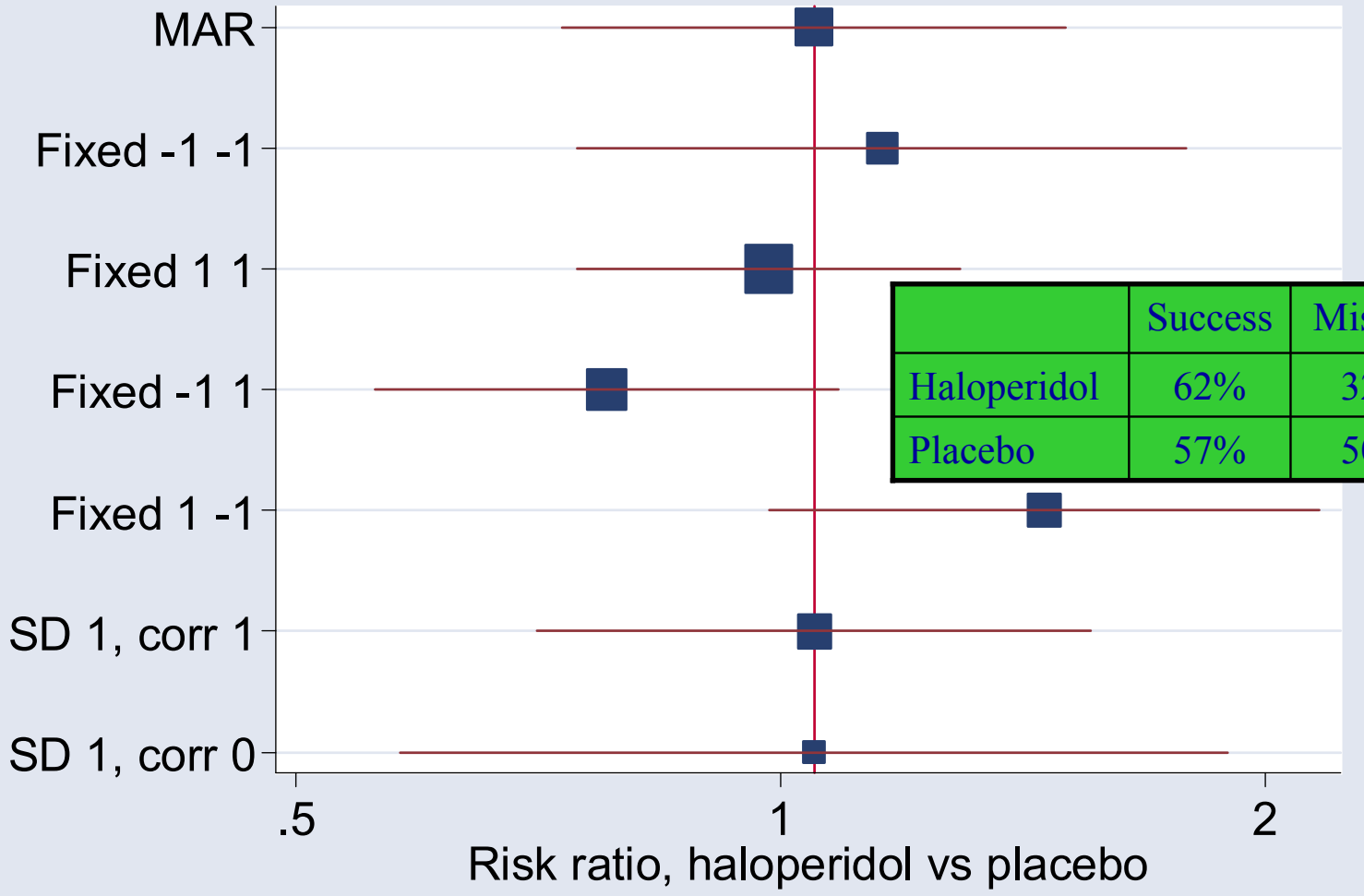
Example

Trial of haloperidol vs. placebo to treat schizophrenia (Beasley, 1996)

	Succ-ess	Fail	Miss-ing	% success (complete cases)	% missing
Haloperidol	29	18	22	29/47 = 62%	22/69=32%
Placebo	20	14	34	20/34 = 57%	34/68=50%

Aim: estimate the *risk ratio*, allowing for the missing outcomes.

Results: various priors for δ_0, δ_1



	Success	Missing
Haloperidol	62%	32%
Placebo	57%	50%

Implications

- Same IMOR in both arms → small adjustment
  - depends on imbalance in % missing
- IMOR differs between arms → often much larger adjustments
  - depends on overall degree of missingness

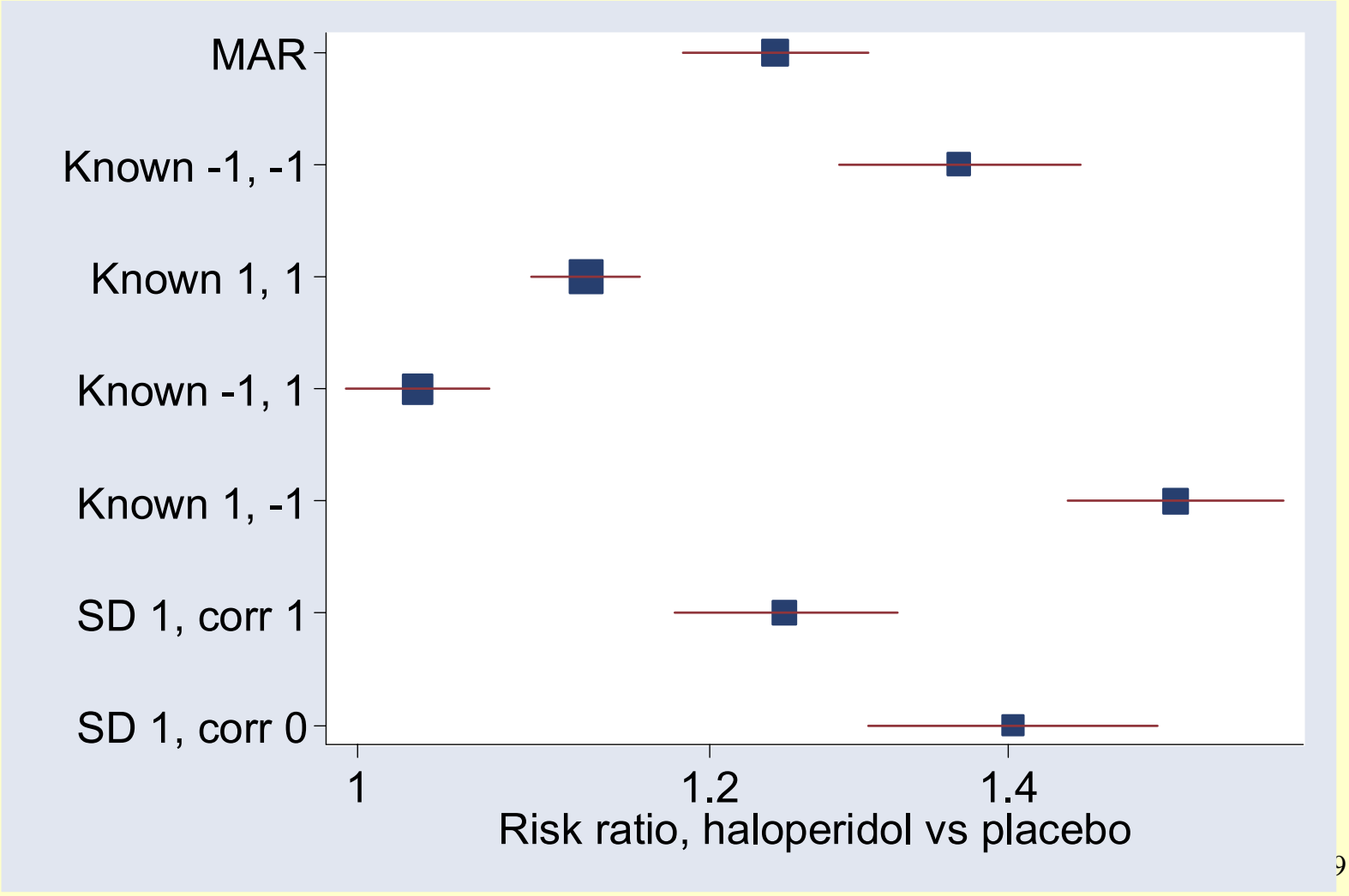
Meta-analysis

- The Beasley trial discussed above was part of a meta-analysis of 17 trials
- Two trials had substantial missingness
- Start with MAR meta-analysis
- Do sensitivity analyses to IM

4 sensitivity analyses

1. Fixed IMOR (same in all trials)
   a. same IMOR in both arms
   b. opposite IMORs
   → changes point estimates
2. Random IMOR (varies between trials)
   a. same IMOR in both arms
   b. IMORs uncorrelated between arms
   → standard error ↑, trial weight ↓

Haloperidol meta: sensitivity analysis



Hierarchical model for IM in meta-analysis

With

Julian Higgins and Angela Wood (BSU)

Nicky Welton and Tony Ades (Bristol)

1 or 2 stages?

- We have used a 2-stage method:
  - estimate effect & standard error for each trial, allowing for IM within trials
  - pool across trials
- Can we use a 1-stage method?
  - hierarchical model

Model

Outcome model:

π_{ir} = true success fraction in arm r of trial i

logit $\pi_{ir} = \mu_i + \beta_i r$

Treatment effect $\beta_i = \beta$ or $\mathrm{N}(\beta, \tau^2)$

Missingness model:

$\alpha_{ir0}, \alpha_{ir1}$ = probability of missing in successes, failures

$$IMOR_{ir} = \frac{\alpha_{ir1}}{1-\alpha_{ir1}} \Big/ \frac{\alpha_{ir0}}{1-\alpha_{ir0}}, \quad \delta_{ir} = \log(IMOR_{ir})$$

Need a model for δ_{ir}

Possible models for IMORs

- $(\delta_{i0},\delta_{i1})$ independent between trials with specified prior e.g.
  - $\delta_{i0}=1$, $\delta_{i1}=-1$ in all trials
  - $\delta_{ir}=N(0,1)$, $corr(\delta_{i0},\delta_{i1})=1$
- Allow correlation between trials, e.g.
  - $\delta_{ir}=\alpha+\beta_i+\gamma_r+\delta_{ir}$, each with specified variance
- Common IMORs e.g. $\delta_{ir}=\delta_r$ and vague prior on δ_r
- Exchangeable IMORs
  - $(\delta_{i0},\delta_{i1})=N(\mu,\Sigma)$ and vague prior on μ,Σ

Learning about δ

- Hierarchical models can in principle learn about δ
- e.g. if missingness is associated with effect size
- Seems dangerous! e.g. other aspects of trial quality might be associated with missingness and influence effect size
- I would prefer *not* to learn about δ

Hierarchical models: estimated log IMORs

Model for IMORs			Estimate	95% CI	
Common			-1.33	-2.74	+0.05
Arm-specific	Haloperidol		-0.69	-2.40	+1.03
	Placebo		-2.88	-5.54	-0.35
Exchange-able	Halo-peridol	Mean	-0.35	-29	+28
		SD	37	24	67
	Placebo (corr=0.01)	Mean	-16	-65	+20
		SD	46	28	100

- Looks as if we don't learn much about δ
- May be a safe framework to express our views about δ

7. Practicalities & discussion

IM analysis

- Need to go beyond MAR analysis, especially when outcome is measured only once
- Proposed approximate method is realistic and simple to apply
- Must consider different degrees of IM in different arms
  - Prior correlation is important

Alternative approach

- A non-Bayesian alternative is to use the elicited results to inform sensitivity analyses, assuming different fixed δ's.
- This is fine, but I prefer the Bayesian approach because it changes the "headline figure"

Eliciting priors

- Who provides the prior?
  - investigator?
  - independent expert?
  - meta-analyst?
  - you, the online reader?
- How many “experts”?
- Elicit before or after data collection?
- Need more expertise in eliciting priors
- Need a “library” of IM differences

Conservative analysis

- LOCF is sometimes claimed to be conservative
- The proposed IM analysis has a much better claim to be conservative
  - corrects point estimate if this is reasonable
  - inflates standard error to allow for uncertainty about missing data

I would like to see …

- … a policy (by journals and regulators) that any trial must
  - *either* find evidence about the degree of IM
  - *or* allow for a plausible degree of IM in the primary analysis

References

I. R. White, S. J. Thompson. Adjusting for partially missing baseline measurements in randomised trials. *Statistics in Medicine*, in press.

R. Henderson, P. Diggle, A. Dobson. Joint modelling of longitudinal measurements and event time data. *Biostatistics* 2000;1:465–480.

J. P. T. Higgins, I. R. White, A. Wood. Missing outcome data in meta-analysis of clinical trials: development and comparison of methods, with recommendations for practice. *Clinical Trials*, submitted.

J. J. Forster, P. W. F. Smith. Model-based inference for categorical survey data subject to non-ignorable non-response. *Journal of the Royal Statistical Society (B)* 1998;60:57–70.

S. Schroter, N. Black, S. Evans, J. Carpenter, F. Godlee, R. Smith. Effects of training on the quality of peer review: A randomised controlled trial. *British Medical Journal* 2004;328:673–5.

C-M. J. Beasley, G. Tollefson, P. Tran, W. Satterlee, T. Sanger, S. Hamilton. Olanzapine versus placebo and haloperidol: acute phase results of the North American double-blind olanzapine trial. *Neuropsychopharmacology* 1996;14:111–123.

D. B. Rubin. Formalizing subjective notions about the effect of nonrespondents in sample surveys. *Journal of the American Statistical Association* 1977;72:538–543.

R. J. A. Little. Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association* 1995;90:1112–1121.

A. Wood, I. R. White, M. Hotopf. Using number of failed contact attempts to adjust for non-ignorable non-response. *JRSSA*, submitted.