Allowing for uncertainty due to LOCF-imputed and missing outcome data in meta-analysis

Dimitris Mavridis
Assistant Professor in Statistics
University of Ioannina

Acknowledgements: Andrea Cipriani, Anna Chaimani, Georgia Salanti, Julian Higgins, Ian White, Toshi Furukawa

Decision making in medicine

- Are atypical antipsychotics more effective than typical antipsychotics in reducing the symptoms of schizophrenia?

- To determine whether the administration of intravenous streptokinase early in the course of acute myocardial infarction would limit myocardial damage

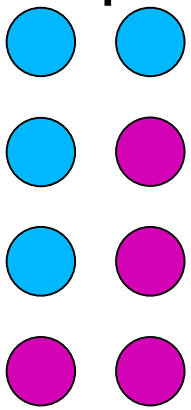
- To evaluate the efficacy of glucose-lowering drugs in patients with type 2 diabetes

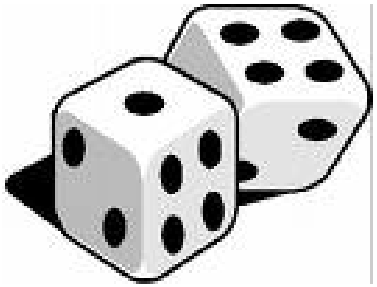
Randomized clinical trials (RCT)

Randomization distributes individual differences equally across groups and any difference in the outcome can be attributed to the intervention received

RCTs are the gold standard for clinical trials

participants





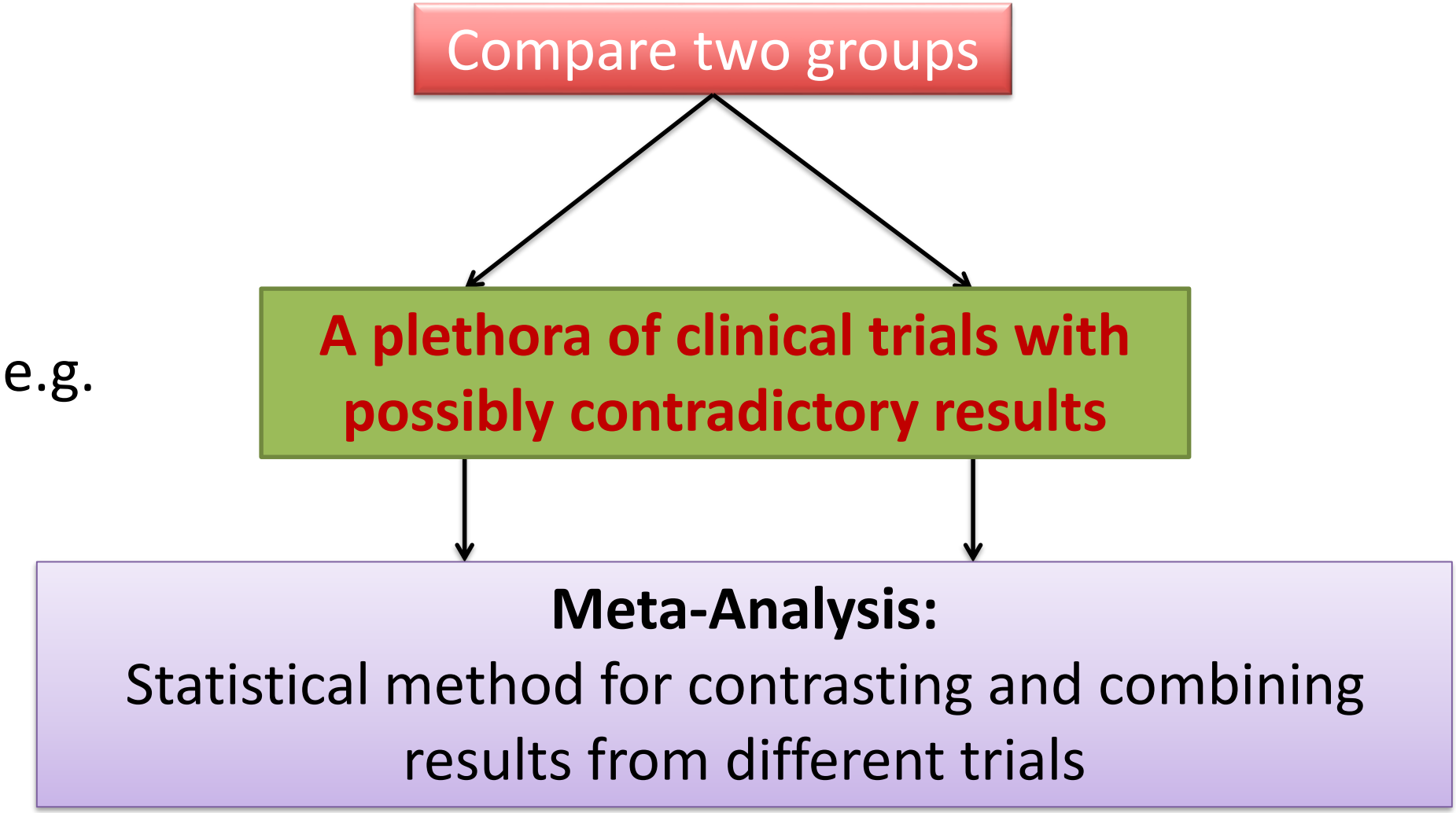
atypical antipsychotic

typical antipsychotic

Lots of studies with contradictory results

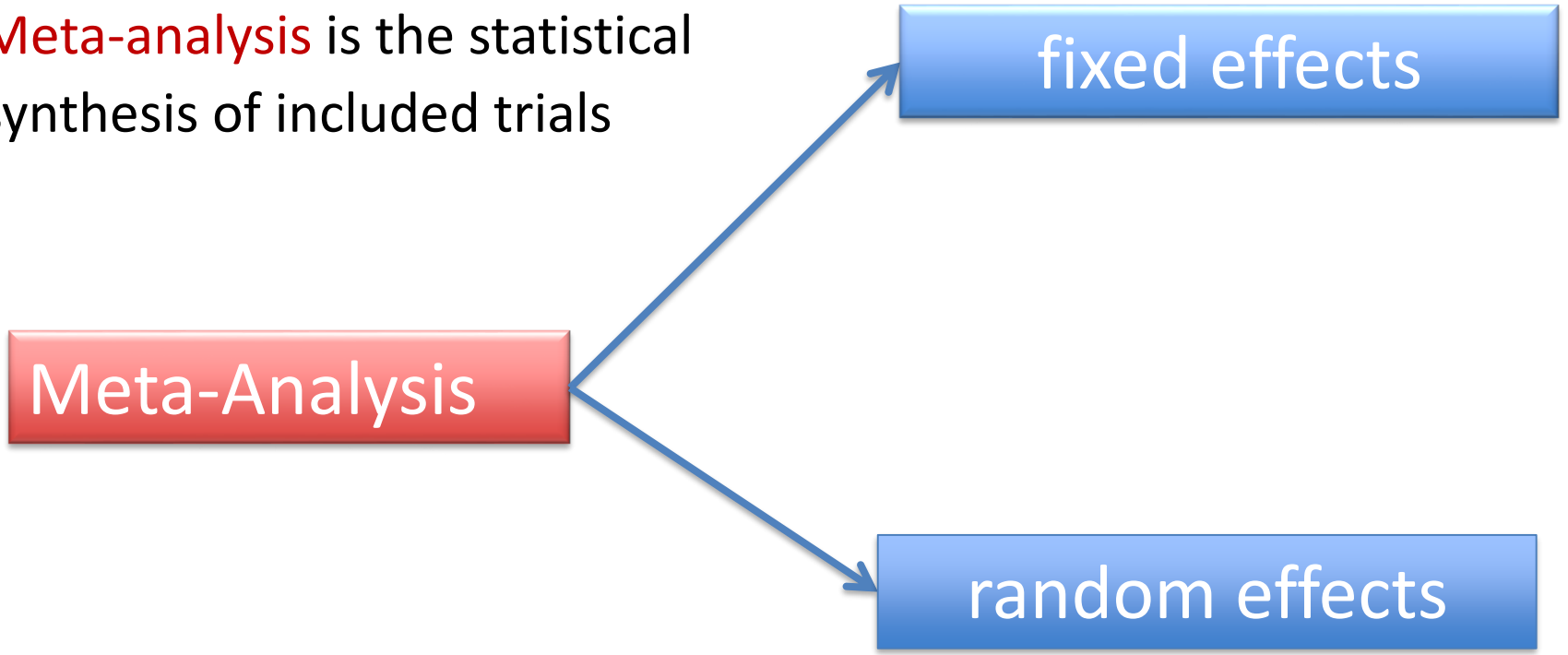
How to quantify all this information?

Meta-analysis



Meta-analysis

Meta-analysis is the statistical synthesis of included trials



- Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. A basic introduction to fixed-effect and random-effects models for meta-analysis. Research Synthesis Methods 2010;1:60-86
- Nikolakopoulou A, Mavridis D, Salanti G. Demystifying fixed and random effects meta-analysis. *Evidence-Based Mental Health* 2014; **17**(2): 53–57.

Meta-Analysis

- Meta-analysis is a two-stage procedure. The unit of analysis is the trial and not the individual (unless you have IPD)

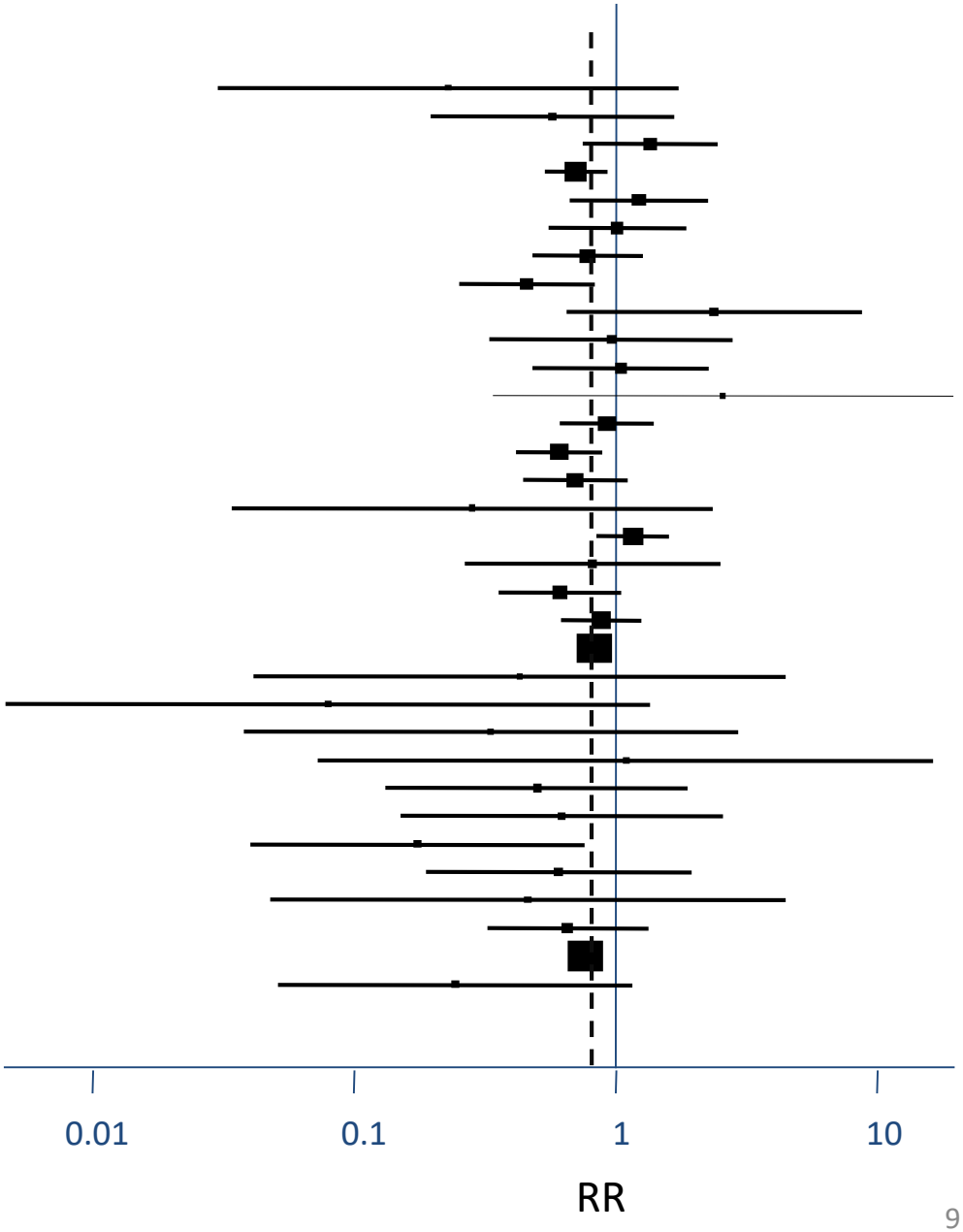
- 1st stage: extract data from the included trials. Compute a summary statistic (mean difference, odds/risk ratio etc) for each trial that describes the intervention effect (effect size) and quantify its uncertainty

- 2nd stage: Estimate a summary (pooled) intervention effect as a weighted average of the intervention effects estimated in individual studies

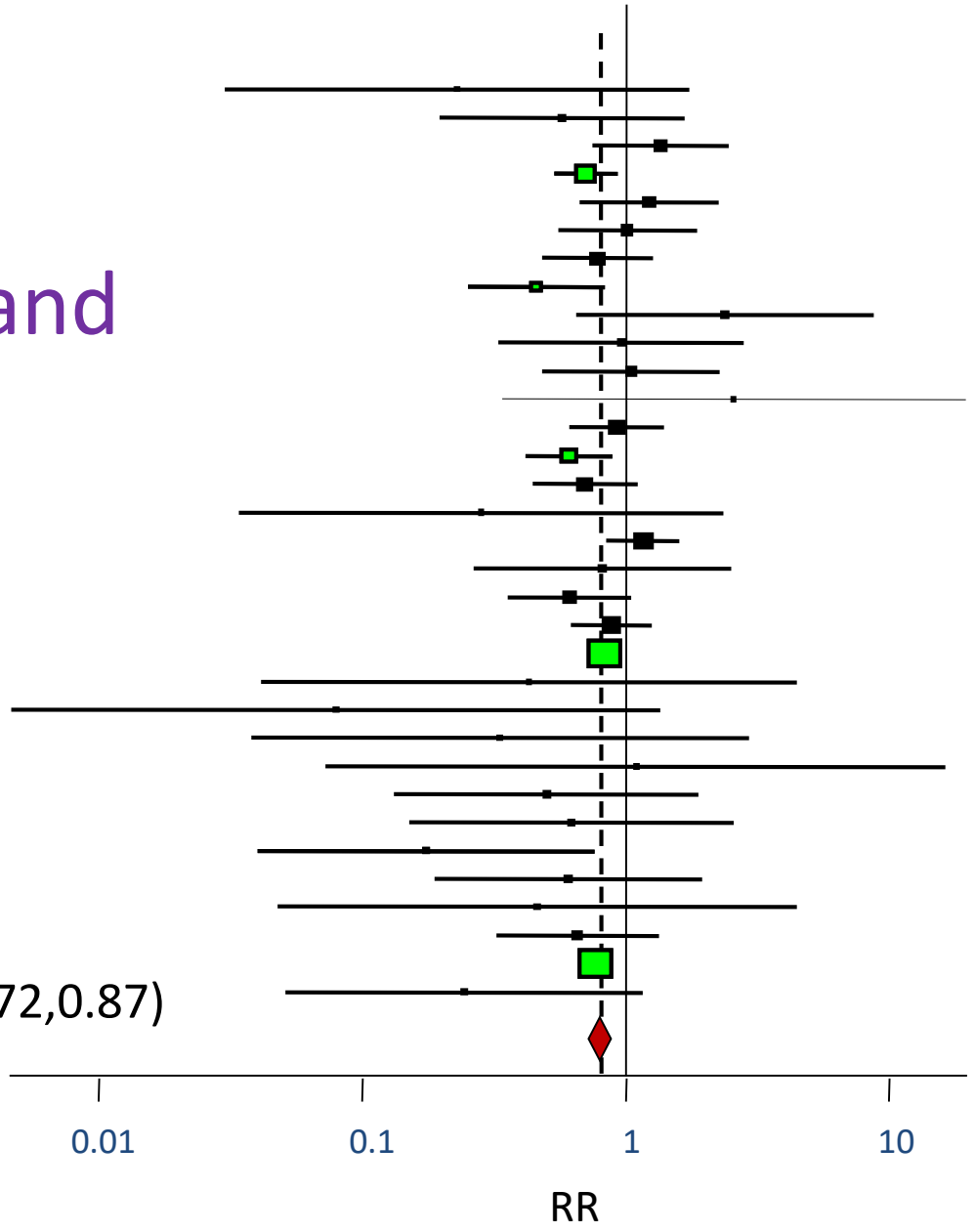
Advantages of meta-analysis

- To increase power and precision
  - detect effect as statistically significant; narrower Cis
- To quantify effect sizes and their uncertainty
  - reduce problems of interpretation due to sampling variation

Streptokinase and myocardial infarction



Streptokinase and myocardial infarction



RR = 0.79 (95% CI 0.72,0.87)

Medical decision making

- Administration of intravenous streptokinase for myocardial infarction
- Since 1977 there were lots of RCTs (5000 patients in total), for which a statistical synthesis clearly shows a significant reduction in mortality
- We waited for an extra 10 years (and randomized an extra 30 000 patients!) until streptokinase was adopted
- 15000 patients were randomized to a less effective treatment and ran a higher risk of death

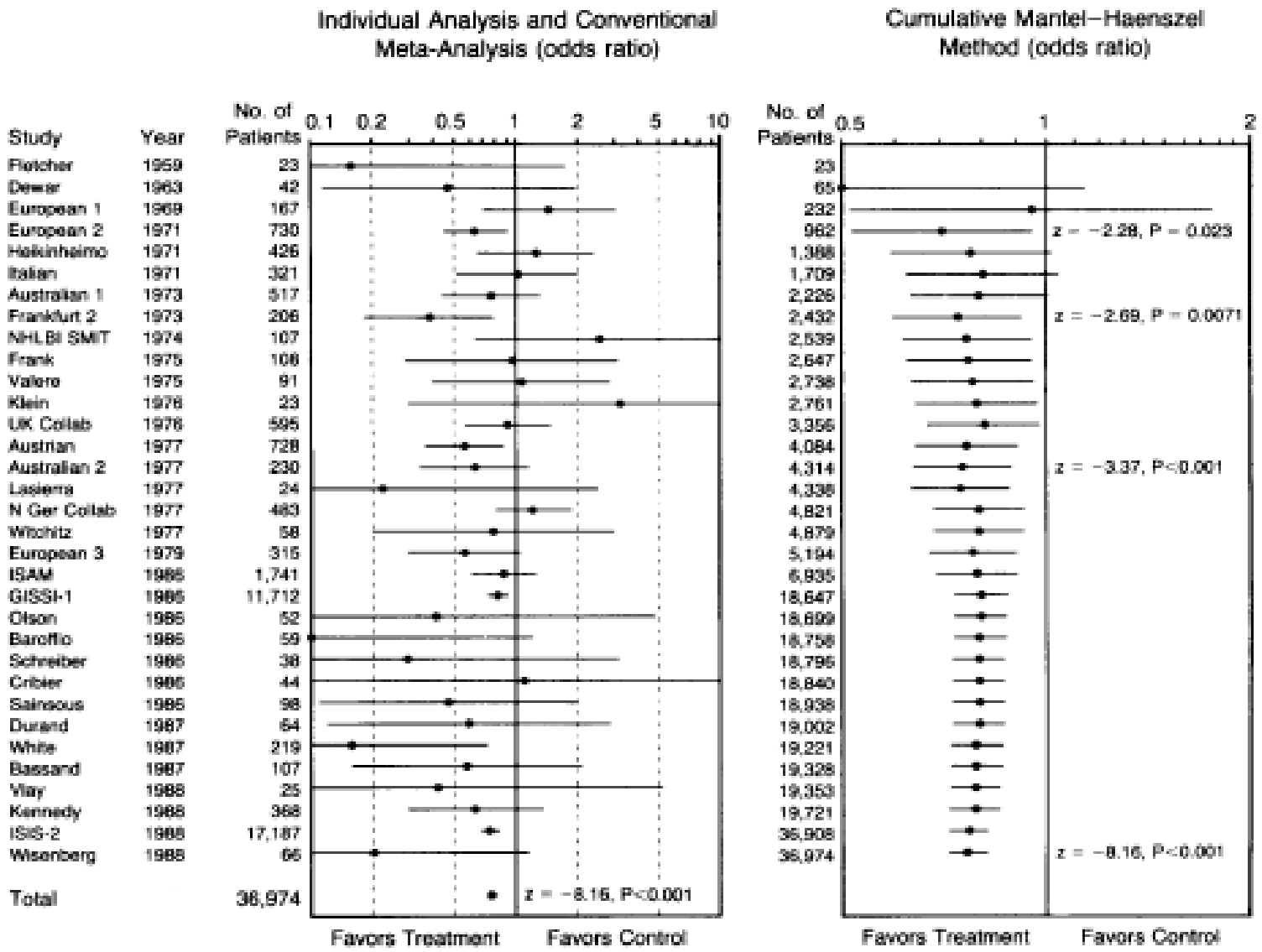


Figure 1. Conventional and Cumulative Meta-Analyses of 33 Trials of Intravenous Streptokinase for Acute Myocardial Infarction. The odds ratios and 95 percent confidence intervals for an effect of treatment on mortality are shown on a logarithmic scale. A bibliography of the published trial reports is available from the authors.

Lau J et al. 1992. Cumulative meta-analysis of therapeutic trials for myocardial infarction. New England Journal of Medicine 327(4): 248-254

Why missing outcome data matter ?

Why missing outcome data matter

- Missing outcome data are common in RCT's
  - In mental health, the dropout rate may exceed 50%

It creates two main problems at RCT level:

- **loss in power and precision**
  - Because the sample size decreases
- **Bias (maybe)**
  - **Any analysis must make an untestable assumption about missing data**
  - wrong assumptions ⟹ biased estimates
- There is **no remedy for missing data** - we can only do sensitivity analyses and see how much the results change under different assumptions
- **Any meta-analysis makes an untestable assumption about missing data – even if reviewers don't realize it!**

Assumptions about missing outcome data

Missing At Random (MAR)

The probability that data are missing does not depend on the outcome or unobserved factors that impact on the outcome

- In an RCT of antihypertensives that measures blood pressure (BP) data, older participants are more likely to have their BP recorded. Missing data are MAR if at any age, individuals with low and high BP are equally likely to have their BP recorded

Missing Not At Random (MNAR) or Informatively Missing (IM)

The probability that data are missing depends on the outcome

- In an RCT of antipsychotics individuals with relapse are more likely to leave the study early in the placebo group

Intention-to-treat (ITT) analysis

- Analyze all participants according to the randomization group
- An imputation method is needed
- Some imputation methods do not take uncertainty of imputation into account and consider imputed data as observed data, inflating sample size and producing spuriously narrow confidence intervals

RCT: Haloperidol vs. placebo in schizophrenia (Beasley 1998)

	Success	Failure	Missing
Haloperidol	29	18	22
Placebo	20	14	34

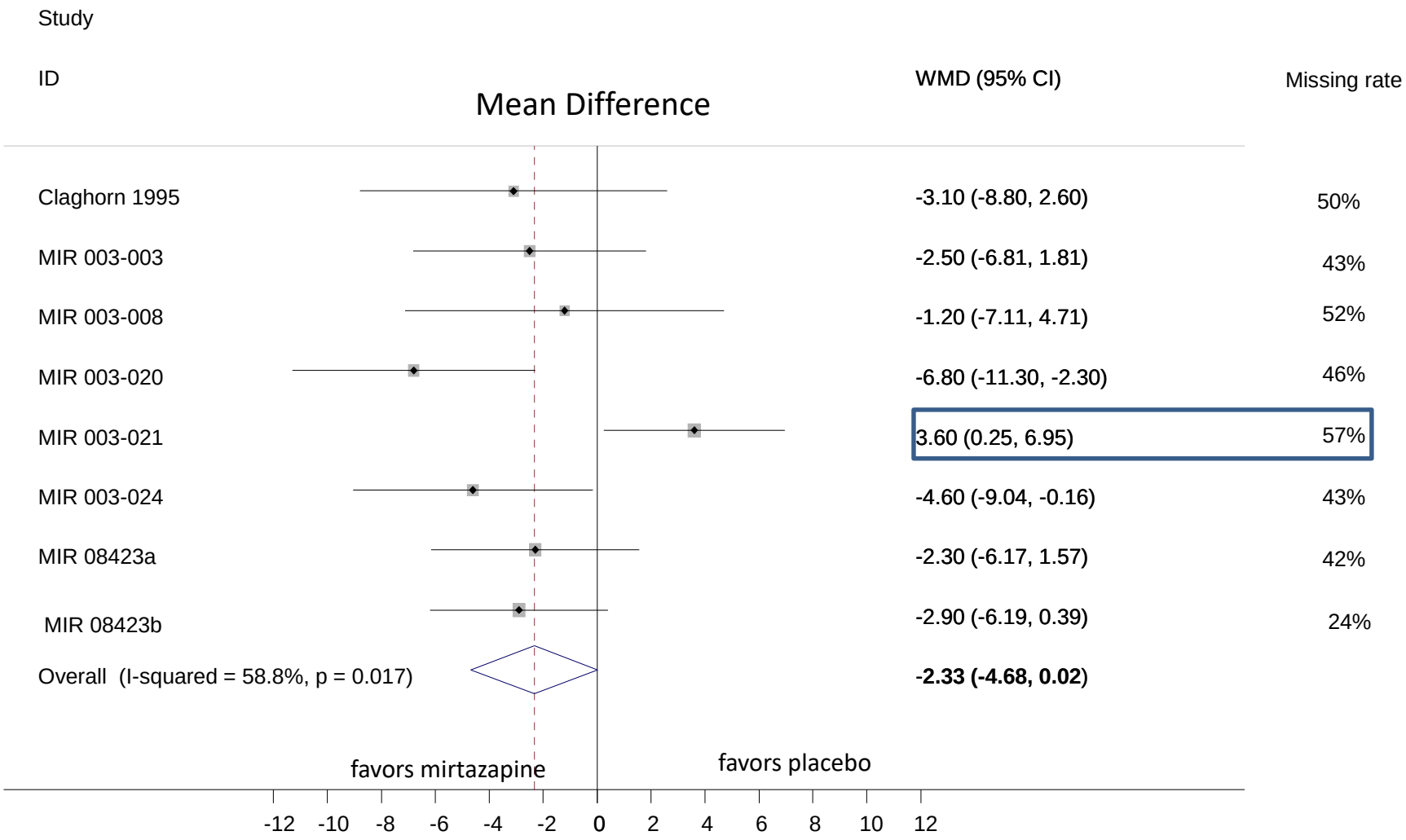
- Outcome: clinical global improvement (yes/no)
- RR=1.03 (0.66,1.61)
- Missing rates are 32% for haloperidol and 50% for placebo
- **How do systematic reviewers analyze these data?**

RCT: Haloperidol vs. placebo in schizophrenia (Beasley 1998)

	Success	Failure	Missing
Haloperidol	29	18	22
Placebo	20	14	34

- **Success rates: 29/47=0.62 vs 20/34=0.59 (Available Cases Analysis, ACA)**
- *Which is the assumption behind?*
- *MAR!*
- **Success rates: 29/69=0.42 vs 20/68=0.29**
- *Which is the assumption behind?*
- *We assume that successes have no chance to dropout!*
- **ANY analysis makes assumptions which, if wrong, produces biased results!**

Random effect meta-analysis of mean change in HAMD21 score. Mirtazapine vs placebo. Complete case analysis



Imputation methods

- Single imputation (Last Observation Carried Forward –LOCF, mean imputation, worst/best case scenario etc)
- Statistical models (inverse probability weighting-selection model, likelihood methods, Bayesian methods, multiple imputation, pattern-mixture models)

Many recently published papers in top medical journals suggest single imputation methods! Many recent RCTs employ single imputation schemes such as LOCF

Summary table of possible analyses (Cochrane Handbook)

Analysis	Outcome	Description of method/how it handles missing participants	Assumptions about missing outcome data	Adequacy for addressing missing data
Available cases	binary continuous	ignore them	a random sample of all participants	valid under missing at random (MAR)
worst (best)-case scenario	binary	imputes failures in the treatment group and successes in the control (or vice-versa)	worse in the experimental group (better in the experimental group)	*inflates sample size and erroneously increase precision/reduce standard errors*
mean imputation	continuous	imputes the mean value	the same as observed	
other simple imputation	binary continuous	imputes specific number of successes/mean value	explicit assumptions about them	
gamble-hollis	binary	downweight studies according to best/worst case scenarios	studies with large differences between best/worst case scenario are less reliable	too extreme downweighting.
The suggested model	binary continuous	downweight studies with high missing rates	the more the missing rate the less reliable is the estimate	**Accounts for uncertainty in the missing outcome data - Expert opinion can also be used.**

Pattern mixture models

$$Y = \left(Y^{obs}, Y^{miss}\right)'$$

i refers to study
j refers to arm
k refers to individual

$$R_{ijk} = \begin{cases} 1 & if \quad outcome \quad is \quad reported \\ 0 & otherwise \end{cases}$$

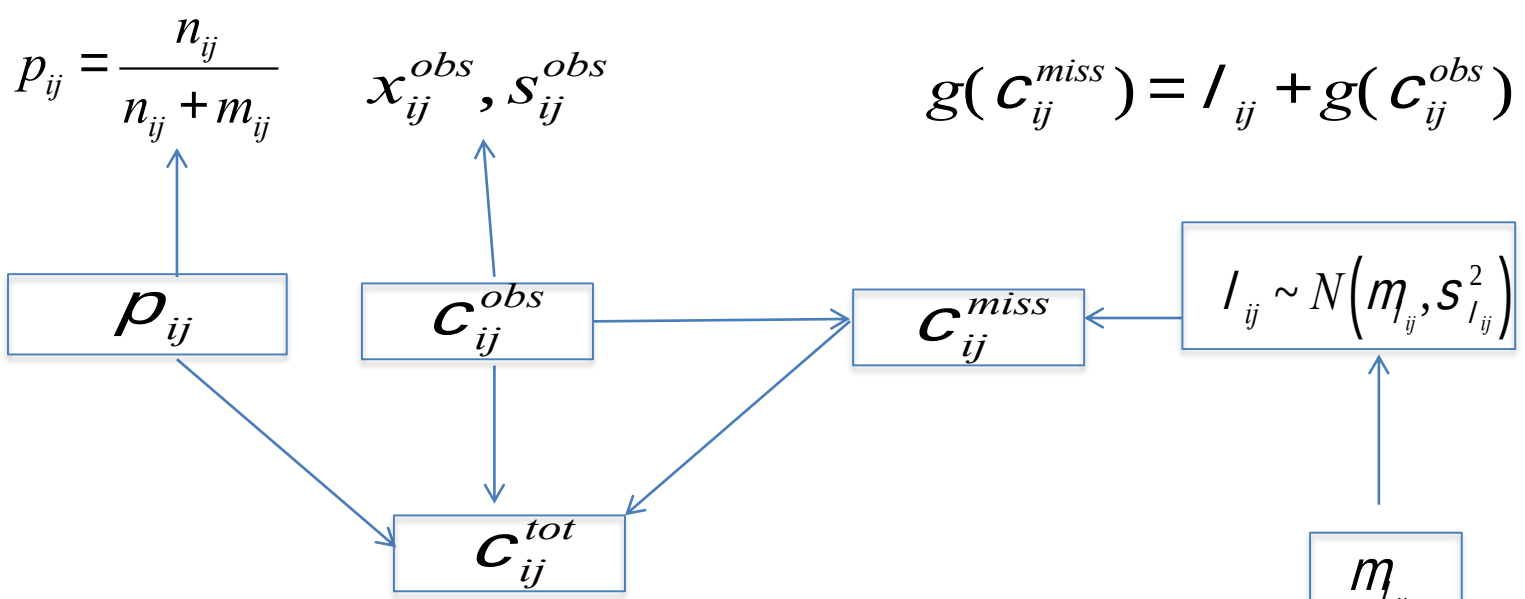
$$P\left(R_{ijk} = 1\right) = p_{ij}^{obs}$$

$$E\left(Y_{ijk} \mid R_{ijk} = 1\right) = c_{ij}^{obs}$$

$$E\left(Y_{ijk} \mid R_{ijk} = 0\right) = c_{ij}^{miss}$$

$$f\left(Y, R\right) = f(Y \mid R) f(R)$$

Model for arm *j* of study *i* pattern mixture model



$$x_{ij}^{tot} = p_{ij} x_{ij}^{obs} + (1 - p_{ij}) x_{ij}^{miss}$$

$$x_{ij}^{tot} = p_{ij} x_{ij}^{obs} + (1 - p_{ij}) g^{-1}\left(l_{ij} + g\left(x_{ij}^{obs}\right)\right)$$

studies i, arms j

Continuous outcome

Informative Missingness Difference in means

$$g(c_{ij}^{miss}) = I_{ij} + g(c_{ij}^{obs})$$

g is the identity function

$$I_{ij} = c_{ij}^{obs} - c_{ij}^{obs}$$

IMP =λ= mean in missing – mean in observed

λ=1 the mean in the missing participants exceed the mean in the observed participants by one unit

λ=-1 the mean in the missing participant is one unit less compared to the mean of the observed participants

λ=0 the data is missing at random

We work out the total means starting from IMP!

- Ask a clinician (or several!) with experience in clinical trials in the field

"Out of 100 patients randomized in drug X, 60 finished the study and had a mean score 3 whereas 40 patients did not finish. What do you guess would be the mean score in those who did not finish?"

-he answered "the mean score in those who did not finish is on average 4"

λ=IMP = mean in missing – mean in observed=4-3=1

Fictional example: Studies with same standard deviations and observed sample sizes per arm, but different missing rates

Study	Observed	Naïve SE (relative weight)	Randomized
1	100	0.07 (20%)	100
2	100	0.07 (20%)	120
3	100	0.07 (20%)	150
4	100	0.07 (20%)	200
5	100	0.07 (20%)	300

Would you give each study the same weight?

No, because uncertainty should be larger when you have more missing data!
The assumption (MAR or a specific form of IM) you will make to estimate IMP has more impact on study 5 rather than on study 2!
The observed sample size is not the only source of uncertainty!

First source of extra uncertainty: Proportion of missing data!

Fictional example: Studies with same mean, standard deviations and observed sample sizes per arm, but different missingness rates

Study	Observed	Naïve SE (relative weight)	Randomized
1	100	0.07 (20%)	100
2	100	0.07 (20%)	120
3	100	0.07 (20%)	150
4	100	0.07 (20%)	200
5	100	0.07 (20%)	300

We want to assume IMDOM=0

- We can NEVER be sure that the mean in the missing is exactly the same as in the observed
- We have some **uncertainty as to what exactly is the mean in the missing data**
- This can be represented by uncertainty in IMDOM!
- **We assume IMDOM=0 with uncertainty interval (-1, 1)**

Second source of extra uncertainty:

Uncertainty about the assumption and IM parameter

Assumptions about the informatively missingness parameter

- Missing at random (MAR) $\lambda_{ij} = 0$
- Free λ_{ij}; $\lambda_{ij} \sim N(\mu_{ij}, \sigma^2_{\lambda_{ij}})$
- Study specific λ: $\lambda_i \sim N(\mu_i, \sigma^2_{\lambda_i})$
- Correlated λ's

$$\begin{pmatrix} \lambda_{iC} \\ \lambda_{iT} \end{pmatrix} \sim N \left(\begin{pmatrix} \mu_{iC} \\ \mu_{iT} \end{pmatrix}, \begin{pmatrix} \sigma^2_{\lambda_{iC}} & \rho_\lambda \sigma_{\lambda_{iC}} \sigma_{\lambda_{iT}} \\ \rho_\lambda \sigma_{\lambda_{iC}} \sigma_{\lambda_{iT}} & \sigma^2_{\lambda_{iT}} \end{pmatrix} \right)$$

Adjusted effect sizes

$$b_i = f\left(c_{iT}^{tot}\right) - f\left(c_{iC}^{tot}\right)$$

- χ is the mean outcome for continuous outcomes and the risk for dichotomous outcomes
- If f is the identity function, then β is the mean/risk difference for continuous/dichotomous outomes
- If f is the logarithmic function, then β is the log mean/risk ratio for continuous/dichotomous outomes
- If If f is the identity function divided by the pooled standard deviation, then β is the standardized mean difference for continuous outomes
- If If f is the logit funtion, then β is the log odds ratio for dichotomous outomes

The key thing is the estimation of the SE of the effect size

To estimate SE(logRR), SE(logOR) and SE(SMD) you need mathematical manipulations or simulations (rather cumbersome!)

Likely, Stata will do the trick for you (metamiss2 command)!

Using Monte Carlo

Using a Taylor series approximation

For all mathematical details see:

Mavridis D., White I., Higgins J., Salanti G **Addressing continuous missing outcomes in pairwise and network meta-analysis** *Statistics in Medicine* 2015, **34**:721-41

White IR, Higgins JPT, Wood AM: **Allowing for uncertainty due to missing data in meta-analysis-Part 1 : Two-stage methods.** *Statistics in Medicine* 2008, **27**:711-727

Estimating E(β) and var(β) Taylor Series Approximation/Monte Carlo

$$b_i = f\left(x_{iT}^{tot}\right) - f\left(x_{iC}^{tot}\right)$$

$$c_{ij}^{tot} = p_{ij}\, c_{ij}^{obs} + \left(1 - p_{ij}\right) g^{-1}\left(g(c_{ij}^{obs}) + I_{ij}\right)$$

- E(β) and var(β) are straightforwardly calculated if *f* and *g* are identity functions

$$E(\beta_i) = (1 - p_{iT})\mu_{\lambda_{iT}} + \mathrm{x}_{iT}^{obs} - (1 - p_{iC})\mu_{\lambda_{iC}} - \mathrm{x}_{iC}^{obs}$$

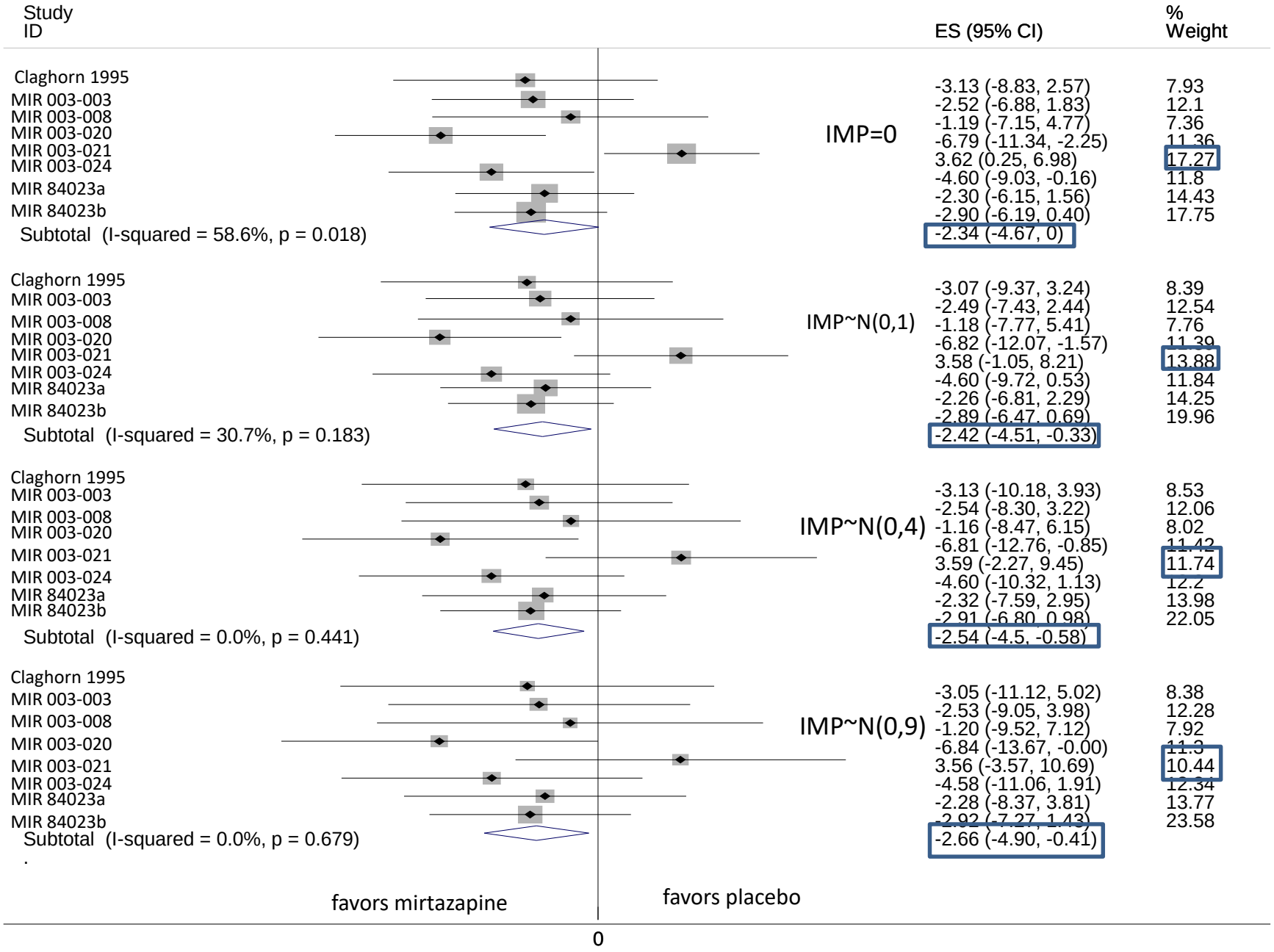
$$var(\beta_i) \approx \sum_{j=C,T} \left[\frac{p_{ij}(1 - p_{ij})}{n_{ij} + m_{ij}} \left(\mu_{\lambda_{ij}}^2 + \sigma_{\lambda_{ij}}^2\right) + \frac{s_{ij}^2}{n_{ij}} + \sigma_{\lambda_{ij}}^2 (1 - p_{ij})^2 \right] - 2\rho_{\lambda_i}\sigma_{\lambda_{iC}}\sigma_{\lambda_{iT}}(1 - p_{iT})(1 - p_{iC})$$

Fictional example: Studies with same mean, standard deviations and observed sample sizes per arm, but different missingness rates

Study	Observed	**Naïve SE (relative weight)**	Randomized	**Corrected SE (relative weight)**
1	100	0.07 (20%)	100	0.07 (57%)
2	100	0.07 (20%)	120	0.11 (25%)
3	100	0.07 (20%)	150	0.17 (10%)
4	100	0.07 (20%)	200	0.24 (5%)
5	100	0.07 (20%)	300	0.32 (3%)

We assume IMP=0 with uncertainty interval (-1, 1)

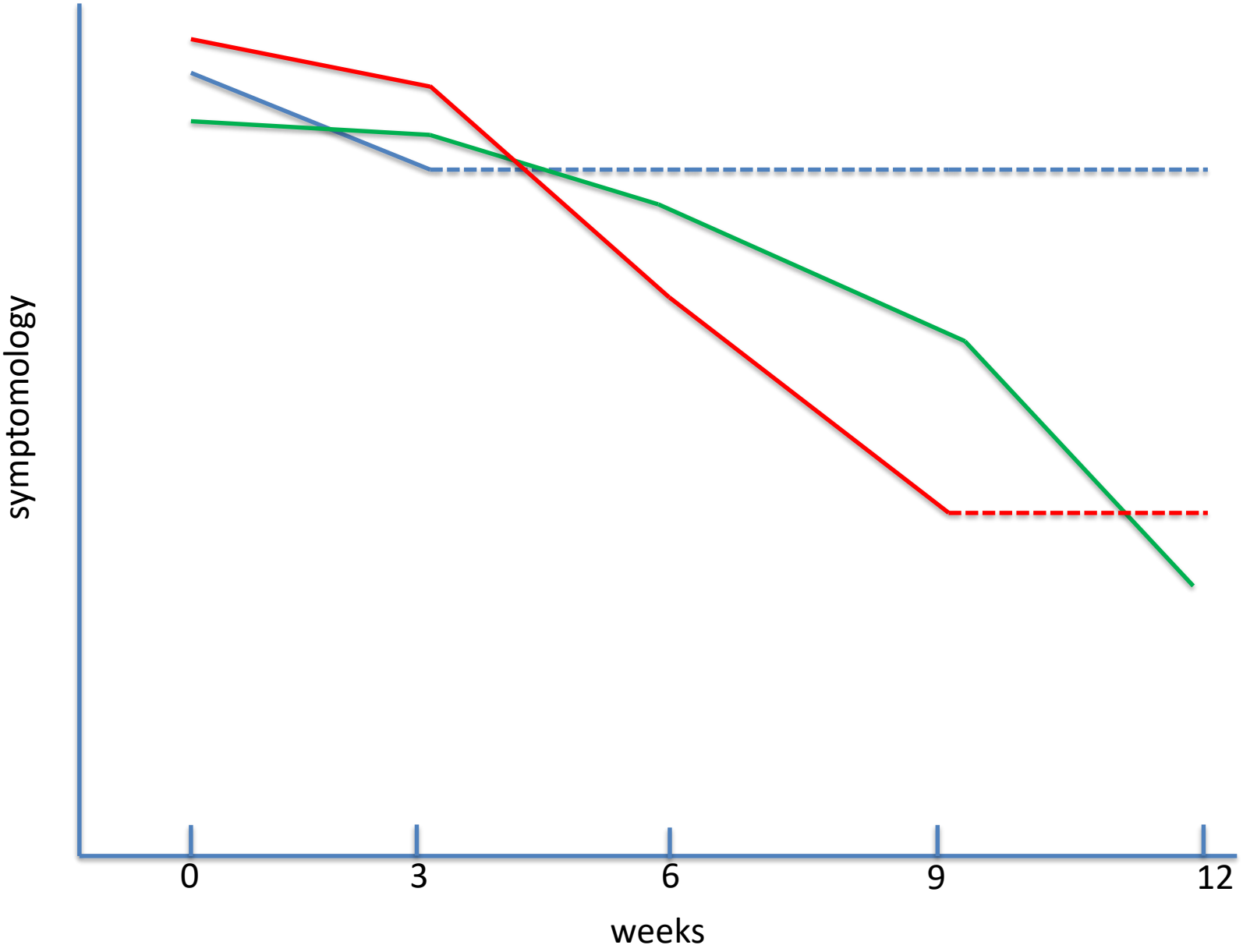
Studies with more missing data get less weight!


Study ID
ES (95% CI)
% Weight
Claghorn 1995
MIR 003-003
MIR 003-008
MIR 003-020
MIR 003-021
MIR 003-024
MIR 84023a
MIR 84023b
Subtotal (I-squared = 58.6%, p = 0.018)
IMP=0
-3.13 (-8.83, 2.57) 7.93
-2.52 (-6.88, 1.83) 12.1
-1.19 (-7.15, 4.77) 7.36
-6.79 (-11.34, -2.25) 11.36
3.62 (0.25, 6.98) 17.27
-4.60 (-9.03, -0.16) 11.8
-2.30 (-6.15, 1.56) 14.43
-2.90 (-6.19, 0.40) 17.75
-2.34 (-4.67, 0)
Claghorn 1995
MIR 003-003
MIR 003-008
MIR 003-020
MIR 003-021
MIR 003-024
MIR 84023a
MIR 84023b
Subtotal (I-squared = 30.7%, p = 0.183)
IMP~N(0,1)
-3.07 (-9.37, 3.24) 8.39
-2.49 (-7.43, 2.44) 12.54
-1.18 (-7.77, 5.41) 7.76
-6.82 (-12.07, -1.57) 11.39
3.58 (-1.05, 8.21) 13.88
-4.60 (-9.72, 0.53) 11.84
-2.26 (-6.81, 2.29) 14.25
-2.89 (-6.47, 0.69) 19.96
-2.42 (-4.51, -0.33)
Claghorn 1995
MIR 003-003
MIR 003-008
MIR 003-020
MIR 003-021
MIR 003-024
MIR 84023a
MIR 84023b
Subtotal (I-squared = 0.0%, p = 0.441)
IMP~N(0,4)
-3.13 (-10.18, 3.93) 8.53
-2.54 (-8.30, 3.22) 12.06
-1.16 (-8.47, 6.15) 8.02
-6.81 (-12.76, -0.85) 11.42
3.59 (-2.27, 9.45) 11.74
-4.60 (-10.32, 1.13) 12.2
-2.32 (-7.59, 2.95) 13.98
-2.91 (-6.80, 0.98) 22.05
-2.54 (-4.5, -0.58)
Claghorn 1995
MIR 003-003
MIR 003-008
MIR 003-020
MIR 003-021
MIR 003-024
MIR 84023a
MIR 84023b
Subtotal (I-squared = 0.0%, p = 0.679)
IMP~N(0,9)
-3.05 (-11.12, 5.02) 8.38
-2.53 (-9.05, 3.98) 12.28
-1.20 (-9.52, 7.12) 7.92
-6.84 (-13.67, -0.00) 11.3
3.56 (-3.57, 10.69) 10.44
-4.58 (-11.06, 1.91) 12.34
-2.28 (-8.37, 3.81) 13.77
-2.92 (-7.27, 1.43) 23.58
-2.66 (-4.90, -0.41)
favors mirtazapine
favors placebo
0

Why LOCF-imputed outcome data matter ?

Haloperidol vs. placebo in schizophrenia

r: success **f**: failures **m**:missing	**Haloperidol**			**Placebo**		
	rh	**fh**	**mh**	**rp**	**fp**	**mp**
Arvanitis	25	25	2	18	33	0
Beasley	29	18	22	20	14	34
Bechelli	12	17	1	2	28	1
Borison	3	9	0	0	12	0
Chouinard	10	11	0	3	19	0
Durost	11	8	0	1	14	0
Garry	7	18	1	4	21	1
Howard	8	9	0	3	10	0
Marder	19	45	2	14	50	2
Nishikawa 82	1	9	0	0	10	0
Nishikawa 84	11	23	3	0	13	0
Reschke	20	9	0	2	9	0
Selman	17	1	11	7	4	18
Serafetinides	4	10	0	0	13	1
Simpson	2	14	0	0	7	1
Spencer	11	1	0	1	11	0
Vichaiya	9	20	1	0	29	1


symptomology
0
3
6
9
12
weeks

The BILOCF parameter

- Bias in the LOCF outcome

BILOCF=γ = true mean outcome – LOCF imputed outcome

Assumptions about the BILOCF parameter

- Missing at random (MAR) $\gamma_{ij} = 0$
- Free γ_{ij}; $\gamma_{ij} \sim \mathrm{N}(\mu_{\gamma_{ij}}, \sigma^2_{\gamma_{ij}})$
- Study specific γ: $\gamma_i \sim \mathrm{N}(\mu_{\gamma_i}, \sigma^2_{\gamma_i})$
- Correlated γ's

$$\begin{pmatrix} \gamma_{iC} \\ \gamma_{iT} \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} \mu_{\gamma_{iC}} \\ \mu_{\gamma_{iT}} \end{pmatrix}, \begin{pmatrix} \sigma^2_{\gamma_{iC}} & \rho_\gamma \sigma_{\gamma_{iC}} \sigma_{\gamma_{i\mathrm{T}}} \\ \rho_\gamma \sigma_{\gamma_{iC}} \sigma_{\gamma_{i\mathrm{T}}} & \sigma^2_{\gamma_{i\mathrm{T}}} \end{pmatrix} \right)$$

Expert opinion

- *Participants randomized to fluoxetine were observed to have a mean score of 25 at the HAMD21 scale with 95% confidence interval [20-30] at 8 weeks after onset of the treatment. What is your prediction about their outcome at 12 weeks?*

IM and BILOCF parameters

parameter	Interpretation
Informative Missingness (IM)	Difference in the mean outcome between missing participants and completers
Bias in the LOCF (BILOCF)	Difference in the mean outcome between LOCF-imputed outcomes and their true value

When we adjust the weigh of a study, we need to take into account

- The observed data
- The missing rate
- Uncertainty in the IMP
- The imputation rate
- Uncertainty in the BILOCF parameter

These parameters are unknown. We can inform them through

- Expert opinion
- Sensitivity analysis
- External data (e.g. if trials report both results from completers and completers+imputed outcomes

Results from IMP+COM and COM analyses

	IMP+COM	COM
Pooled effect 95% CI	-0.22 (-0.41, -0.03)	-0.15 (-0.30,0.01)
Heterogeneity variance I^2	0.075 80%	0.028 53%

- The effect size in completers is not only smaller but also non-significant
- Heterogeneity is larger in completers+imputed. This makes sense conceptually, as analysis involves measuring the outcome at different time points.
- Although completers involve less participants, they have a more precise pooled effect due to the trade-off between within-study and between-study variance

Reboxetine vs placebo for depression

STUDY	TREATMENT	IMP	MEAN IMP+COM	SD IMP+COM	COM	MEAN COM	SD COM	MISSING
Study 1	reboxetine	4	12,60	10,30	22	10,10	8,20	0
	placebo	16	29,50	13,30	10	16,30	10,20	0
Study 2	reboxetine	7	17,18	4,75	17	16,59	4,73	2
	placebo	5	16,6	5,14	21	15,52	4,78	1

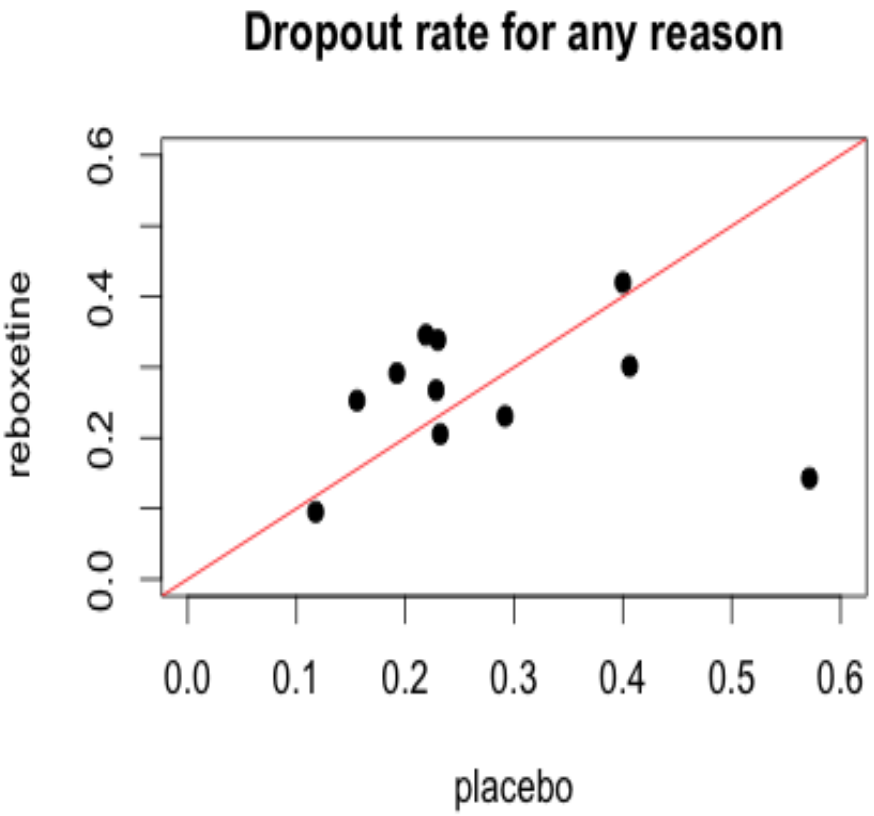
There are 11 studies, 10 report results from both completers and completers+imputed outcomes
We compute SMDs and its standard error for each study

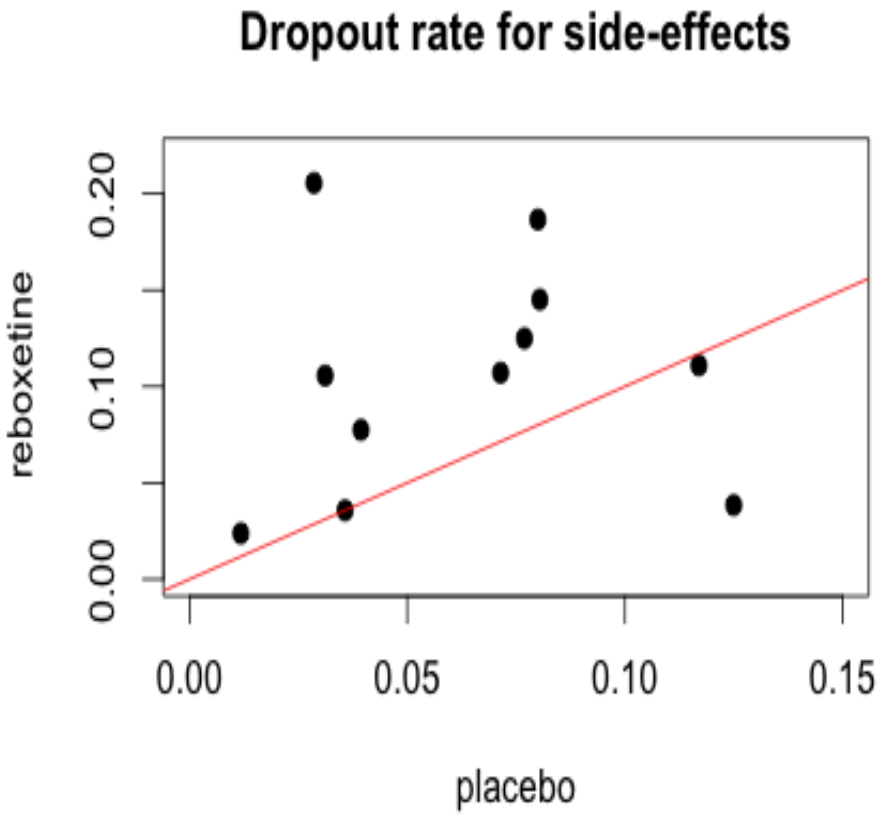
Reboxetine vs placebo for depressions

STUDY	TREATMENT	Imputation rate	SMD	SMD Completers only
Study 1	reboxetine	14%	-1.40	-0.68
	placebo	57%	(-1.99,-0.81)	(-1.45,0.09)
Study 2	reboxetine	19%	0.12	0.22
	placebo	29%	(-0.46, 0.69)	(-0.42,0.86)

- LOCF imputation will favour the drug that has lower imputation rate since participants randomized to that drug have more time under treatment
- The first study has a large effect size. MD=-16.9 or SMD=-1.40 suggesting reboxetine is very effective
- The placebo imputation rate is 57% while for reboxetine is 14%! A difference of 43% in absolute terms or 307% in relative terms!

Number of dropouts and subsequent LOCF imputations





The dropout rate for any reason is balanced across the two arms but is much bigger for reboxetine when it comes to dropout for side-effects. People leave placebo for lack of improvement and reboxetine for side-effects. If one of the groups have a faster dropout time, then the other drug benefits from their comparison using LOCF

IM and BILOCF parameters

parameter	placebo	reboxetine
Informative Missingness (IM)	[5,15]	[5,15]
Bias in the LOCF (BILOCF)	[-15,-5]	[-15,-5]

Summary estimate -0.23 (-0.39, -0.06)

This is a rational assumption if

-We believe missing participants are alike in the two groups

IM and BILOCF parameters

parameter	placebo	reboxetine
Informative Missingness (IM)	[5,15]	[0,10]
Bias in the LOCF (BILOCF)	[-15,-5]	[-10,0]

Summary estimate -0.09 (-0.25, 0.11)

This is a rational assumption if
-placebo dropouts leave the study earlier than reboxetine dropouts
-people on reboxetine leave because they have improved and there is no need to stay on therapy (side-effects)

conclusions

- Missing and LOCF-imputed outcome data are likely to bias treatment effects.
- The drug with the largest missing rate is favored
- The drug with the lowest (single) imputation rate is favored

conclusions

- we suggest models that can
  - account for the fact that the presence of missing and LOCF-imputed data introduce uncertainty in the study estimates
  - naturally downweight studies with lots of missing and imputed data
  - can model MAR or departures from MAR
- metamiss command in STATA (Ian White & Julian Higgins); metamiss2 command in STATA (Anna Chaimani and Ian White, forthcoming)

References

- Higgins JPT, White IR, Wood AM: Imputation methods for missing outcome data in meta-analysis of clinical trials. *Clinical Trials* 2008; **5**, pp. 225-239
- Mavridis D, Chaimani A, Efthimiou O, Leucht S, Salanti G.:Addressing missing outcome data in meta-analysis. Statistics in Practice. *Evidence Based Mental Health* 2014; **17**, pp 85-89.
- Mavridis D., White I., Higgins J., Salanti G.: Addressing continuous missing outcomes in pairwise and network meta-analysis *Statistics in Medicine* 2015, **34**:721-41
- White IR, Higgins JPT, Wood AM: Allowing for uncertainty due to missing data in meta-analysis-Part 1 : Two-stage methods. *Statistics in Medicine* 2008, **27**, pp. 711-727
- White IR, Higgins JPT. Meta-analysis with missing data. Stata J. 9(1):57–69.