Deviation from intention to treat analysis in randomised trials and treatment effect estimates: meta-epidemiological study

Iosief Abraha,[1] Antonio Cherubini,[2] Francesco Cozzolino,[1] Rita De Florio,[3] Maria Laura Luchetta,[4] Joseph M Rimland,[2] Ilenia Folletti,[6] Mauro Marchesi,[5] Antonella Germani,[5] Massimiliano Orso,[1] Paolo Eusebi,[1] Alessandro Montedori[1]

[1]Health Planning Service, Regional Health Authority of Umbria, Department of Epidemiology, 06124 Perugia, Italy

[2]Geriatrics and Geriatric Emergency Care, Italian National Research Centre on Aging Ancona, Italy

[3]General Medicine, Azienda USL Umbria 2, Terni, Italy

[4]General Medicine, Azienda USL Umbria 1, Perugia

[5]Transfusion Medicine Service, Azienda Ospedaliera di Perugia, Perugia

[6]Department of Clinical and Experimental Medicine, University of Perugia, Perugia

Correspondence to I Abraha iosief_a@yahoo.it

ABSTRACT

OBJECTIVE
To examine whether deviation from the standard intention to treat analysis has an influence on treatment effect estimates of randomised trials.

DESIGN
Meta-epidemiological study.

DATA SOURCES
Medline, via PubMed, searched between 2006 and 2010; 43 systematic reviews of interventions and 310 randomised trials were included.

ELIGIBILITY CRITERIA FOR SELECTING STUDIES
From each year searched, random selection of 5% of intervention reviews with a meta-analysis that included at least one trial that deviated from the standard intention to treat approach. Basic characteristics of the systematic reviews and randomised trials were extracted. Information on the reporting of intention to treat analysis, outcome data, risk of bias items, post-randomisation exclusions, and funding were extracted from each trial. Trials were classified as: ITT (reporting the standard intention to treat approach), mITT (reporting a deviation from the standard approach), and no ITT (reporting no approach). Within each meta-analysis, treatment effects were compared between mITT and ITT trials, and between mITT and no ITT trials. The ratio of odds ratios was calculated (value <1 indicated larger treatment effects in mITT trials than in other trial categories).

RESULTS
50 meta-analyses and 322 comparisons of randomised trials (from 84 ITT trials, 118 mITT trials, and 108 no ITT trials; 12 trials contributed twice to the analysis) were examined. Compared with ITT trials, mITT trials showed a larger intervention effect (pooled ratio of odds ratios 0.83 (95% confidence interval 0.71 to 0.96), P=0.01; between meta-analyses variance τ^2=0.13). Adjustments for sample size, type of centre, funding, items of risk of bias, post-randomisation exclusions, and variance of log odds ratio yielded consistent results (0.80 (0.69 to 0.94), P=0.005; τ^2=0.08). After exclusion of five influential studies, results remained consistent (0.85 (0.75 to 0.98); τ^2=0.08). The comparison between mITT trials and no ITT trials showed no statistical difference between the two groups (adjusted ratio of odds ratios 0.92 (0.70 to 1.23); τ^2=0.57).

CONCLUSIONS
Trials that deviated from the intention to treat analysis showed larger intervention effects than trials that reported the standard approach. Where an intention to treat analysis is impossible to perform, authors should clearly report who is included in the analysis and attempt to perform multiple imputations.

Introduction

Post-randomisation exclusions occur in most randomised trials and generally are ascribed to missing data or deviation from protocol. In the case of missing data, the outcome of interest is generally unknown and the application of a true intention to treat approach—that is, analysing patients according to their original allocation—becomes impractical.[1] Deviations from protocol can occur because of several reasons. Patients did not receive the intended treatment, were mistakenly enrolled, were randomised before information on eligibility was obtained and therefore became ineligible, or died or developed the outcome of interest before receiving treatment. In this case, strictly speaking, the outcome of interest should be known and the intention to treat approach can still be applied, although Fergusson and colleagues suggest that these exclusions are appropriate.[2]

When managing missing data or deviations from protocol, trial authors may face difficulties and deviate from the intention to treat approach, potentially disrupting the prognostic balance among treatment allocation groups.[2-4] These difficulties can explain the substantial increase of randomised trials that use a modified intention to treat (mITT) approach in the medical literature.[5] In an initial study, we found that of 475 trials using an mITT analysis, when the description was examined, 40% reported one type of deviation from the intention to treat approach and 55% reported two or more types. In addition, these trials were characterised by post-randomisation exclusions and

---

WHAT IS ALREADY KNOWN ON THIS TOPIC

The publication of randomised trials that report deviations from the intention to treat approach or report a modified approach has become more common in the medical literature

Reporting of deviations is significantly associated with post-randomisation exclusions, funding, and authors' conflicts of interest

An increase in significant results has been reported among trials that have deviated from the intention to treat approach, but no evidence exists to indicate whether such trials overestimate the treatment effect

WHAT THIS STUDY ADDS

Trials that deviate from the intention to treat approach overestimate the treatment effect of meta-analyses compared with those trials that report a standard approach

funding from private enterprises. The prevalence of the mITT approach has been reported as 0.6% of trials published in 2000-06 in PubMed, and 4.9% of trials published in 2000-06 in four highly cited general medical journals. These proportions are certainly underestimated, because trials that deviated from the intention to treat approach but did not use the word "modified" were not considered. As we have shown in a second study, trials that deviated from an intention to treat analysis, without specifically using the word "modified," generated a higher than expected number of trials (32 trials, in addition to 24 trials that used the word "modified").[6]

Given the preponderance of significant results among the trials that have deviated from the intention to treat approach,[5] we hypothesise that such trials overestimate the treatment effect compared with those trials reporting standard intention to treat analysis independent from the occurrence of exclusions.

Methods
Data sources and search strategy
To assess the impact of trials that deviate from a standard intention to treat analysis on the treatment effect of any clinical condition, we used a meta-epidemiological analysis, which is considered the most efficient approach to assess bias.[7] A meta-epidemiological analysis is a statistical technique used to assess whether a primary study characteristic can be a source of bias. It is a collection of meta-analyses, in each of which the trials are classified according to the presence of the characteristic under consideration. The influence of inadequate allocation concealment, of the absence of blinding, or of the presence of post-randomisation exclusion has been investigated in meta-epidemiological studies.[7-9]

To identify the reviews of interest, we launched Montori's highly specific search strategy in PubMed (Medline [title/abstract] OR (systematic[title/abstract] AND review[title/abstract] OR meta-analysis[publication type]).[10] From each year searched between 2006 and 2010, we randomly selected 5% of the records.

Study selection
Process of systematic review selection
We analysed only systematic reviews of therapeutic or preventive interventions, with at least one meta-analysis with categorical data, each containing at least two randomised trials of which at least one had to report a deviation from the intention to treat approach. Reviews with meta-analyses that included only trials using a deviation were excluded. We also excluded intervention reviews with head to head comparison trials, because experimental and control interventions could not be ascertained. Diagnostic, prognostic, epidemiological, and cost-effectiveness studies; animal studies; and reviews with meta-analyses of cluster or crossover trials were excluded. We excluded non-English language reviews. Pairs of investigators independently assessed relevant reviews for eligibility. Disagreement was resolved by discussion.

Data extraction and quality assessment
We developed two separate data collection forms for systematic reviews and randomised trials, by testing them on a random sample of three systematic reviews (with 20 randomised trials) that were not included in the final analyses.[11 12 13]

Systematic reviews
From each systematic review, the following information was retrieved: the journal in which the study was published, publication year, clinical area of interest, type of intervention in the experimental and control groups, type of primary outcome, funding, and the author's conflict of interest. The number of events, population, primary outcome measure, and results of the meta-analyses with its method of pooling (random or fixed effects model) were recorded. When necessary, authors of the systematic reviews were contacted to clarify uncertainties or obtain unpublished full texts of primary studies.

Randomised trials
From each randomised trial, the following general characteristics were collected: journal in which the trial was published; year of publication; clinical area investigated; number of patients allocated in each arm, together with the number of events; funding type; and conflict of interest of the author(s). Trials were subsequently classified according to the type of intervention (pharmacological or non-pharmacological), the type of centre (single-centre or multicentre) and the sample size.

For each primary outcome, the number of patients allocated to the experimental and control groups and the number of patients who experienced the outcome in each group were extracted. In the case of multiple allocations, we collected information on groups that were included in the meta-analysis only. For outcomes recorded at different times, the time point selected in the meta-analysis was used. In the case of number discrepancies between primary studies and meta-analyses, the primary study numbers were analysed. For each primary outcome, the number of post-randomisation exclusions was collected for experimental and control groups, wherever possible.

Pairs of reviewers independently extracted the data and disagreement was resolved by discussion. When necessary, authors of the randomised trials were contacted to clarify uncertainties.

Classification of trials based on intention to treat reporting
The classification based on intention to treat reporting formed the basis of our assessment. The issue of intention to treat is often interlinked with the post-randomisation exclusions. However, because exclusions are not always reported in trials and sometimes it is not possible to know their appropriateness, we classified trials using three categories of reporting, without using the occurrence of exclusions for the purpose of classification:[6]

First, ITT trials reported the phrase "intention to treat" with no apparent deviation in the description or

trials that correctly described the intention to treat principle. If a trial did not use the phrase but intended to analyse the patient data according to the original allocation of the patients, then it was classified in this category. A trial that reported analysis based on both standard and deviated approaches was classified in this category.

Second, mITT trials explicitly used the term "modified intention to treat" or reported a deviation from the intention to treat approach.[6] The number and type of deviations were retrieved and deviations were classified as treatment related deviation, baseline assessment related deviation, target condition related deviation, and post-baseline assessment related deviation. Third, "no ITT" trials did not refer any intention to treat approach and did not fall into the previous two categories.

To ensure uniform judgment of the reporting of intention to treat analyses, a pair of reviewers electronically copied the description from each trial, allowing for immediate availability and future reference. Both reviewers performed this task independently and classified the trials. Disagreement was resolved by consensus and, where necessary, by a third reviewer. This third reviewer also had the task of re-evaluating the trials that were classified as a no ITT trial.

Box 1 provides examples of real trials, classified according to the type of intention to treat reporting. Box 2 gives examples of descriptions of deviations from the intention to treat approach. A complete list of all the included trials and their classification is available in web appendix 1.

Post-randomisation exclusions

Post-randomisation exclusions in randomised trials were carefully assessed by checking the results and the tables, and taking into account the outcome of interest. If there was no apparent exclusion, the trial was categorised as "no exclusion." If exclusions occurred, but were not related to the outcome of interest, the trial was categorised as "no exclusion." Pairs of reviewers, who were not involved in classifying trials based on the intention to treat approach, independently extracted data and, where available, copied the related text for future reference. Discrepancies were resolved by discussion or by consulting a third reviewer.

Outcome of interest

We were interested in outcomes that were dichotomous. If a review reported results for both continuous and dichotomous outcomes, we selected the dichotomous set. When the outcome of interest was reported as a secondary outcome, the review was excluded. If results for more than one binary outcome were combined, we selected the objective outcome (such as all cause mortality or the data from a biological test, if reported).

Assessment of risk of bias of individual randomised trials

We used the approach recommended by the Cochrane Collaboration[28] to assess the risk of bias of individual primary studies for the following items: random sequence generation, allocation concealment, blinding of participants and personnel and blinding of the outcome assessor, and incomplete outcome data (web appendix 2). We were unable to evaluate selective outcome reporting bias, since we obtained the study protocol for only 14 trials (web appendix 3). The risk of bias of each item was rated as low, high, or unclear. Pairs of reviewers independently extracted data, and resolved any disagreements by discussion.

Funding sources

Information regarding the sources of funding for each trial was extracted from the randomised trials by reviewing the main body of the article, conflict of interest section, information on funding (if present), and acknowledgments section. Funding sources were abstracted and categorised as follows: (1) not-for-profit organisation; (2) for-profit agency; (3) co-financed, indicating funding from both not-for-profit organisation(s) and for-profit agency(s); (4) no funding; and (5) not reported. The presence or absence of author conflicts of interest was also determined, but since the results were similar to the funding data, they are not presented to avoid redundancy.

Data synthesis and analysis

Categorical variables were described by frequencies and percentages, and quantitative variables with means and standard deviations or medians and ranges. We used the κ coefficient to determine the degree of agreement between reviewers. Agreement between reviewers during the screening process and data abstraction was high (κ=0.96 for title and abstract screening, 0.87 for full text screening, 0.90 for intention to treat classification, and 0.86 for risk of bias assessment).

Meta-epidemiological analysis

Intervention effects were estimated using odds ratios. We recoded outcome events so that an odds ratio less than 1 indicated a positive or beneficial effect in favour of the experimental intervention. To estimate differences in effect estimates between the three types of trials, we performed a multivariable, multilevel meta-epidemiological analysis following the approach suggested by Siersma and colleagues.[29] Results are expressed as the average ratio of odds ratios (ROR)—that is, the ROR between mITT trials and ITT trials, and the ROR between mITT trials and no ITT trials. An ROR less than 1 indicated a larger effect estimate in mITT trials compared with ITT trials or no ITT trials. The heterogeneity across meta-analyses was measured with τ^2 (web appendix 4).

Heterogeneity assessment

A graphical method suggested by Baujat and colleagues[30] was proposed to identify studies that affect heterogeneity or the meta-analytic odds ratio estimate. The graphical approach also can be applied to identify meta-analyses that affect heterogeneity or the meta-analytical ROR estimate (web appendix 4).

Box 1: Examples of randomised trials classified according to intention to treat reporting

ITT trials

Sparano[14]

"TTP (time to progression), TTF (time to treatment failure), PFS (progression free survival), and OS (overall survival) were analyzed using the intent-to-treat population (all randomly assigned patients)";

"ORR (objective response rate) was analyzed using the assessable population (all randomly assigned patients who received at least one dose of study medication and who had at least one post-baseline tumor assessment)".

The primary outcome of interest of the systematic review[15] was progression free survival. The analysis described is based on an intention to treat approach. A modified intention to treat approach was used to assess the objective response rate, which was a secondary endpoint. Thus, this trial was classified as an ITT trial.

Hollander[16]

"Analyses were performed on an intent-to-treat basis with the last observation carried forward. All 73 patients enrolled were evaluated for efficacy, safety, and immunogenicity".

The trial declared an intention to treat based analysis and indicated an imputation method for missing data. No deviation was reported.

Thigpen[17]

"The primary treatment comparisons of benefit include all eligible patients, regardless of the amount of study treatment received. Both eligible and ineligible patients are included in the analyses and compared by treatment assignment when an intent-to-treat analysis is specifically indicated in this article. The summaries of toxicity include all patients who received any study treatment; those who did not receive study treatment are not included in these summaries".

The trial declared an intention to treat analysis for the efficacy analysis.

Cameron[18]

"To summarize the methods used in this report, the intent-to-treat (ITT) population, comprising all women who underwent randomization, was used for all efficacy analyses. The safety population comprised all randomized patients who received at least one treatment dose and was based on the actual treatment received".

The trial declared an intention to treat analysis for the efficacy analysis. The safety analysis was carried out based according to the treatment received.

mITT trials

Asakura[19]

"Efficacy and safety analyses were carried out on all patients who received at least one dose of the study medication and for whom at least one valid post-baseline efficacy evaluation was obtained".

A two type deviation (treatment related and post-baseline assessment related[5]) from the intention to treat approach was described. Consequently, the study was classified as an mITT trial.

Kärkkäinen[20]

"Data was analysed on a modified intention to treat (ITT) basis by retaining allocation to groups according to randomization and by including all subjects in whom we had endpoint information".

A modified intention to treat reporting trial, based on the presence of a post-baseline assessment related deviation.

No ITT trial

Pederzoli[21]

"Statistical analysis was based on the U test."

The authors provided no description of an intention to treat analysis or did not provide how to analyse data.

Box 2: Examples of types of deviations from the intention to treat approach

The following examples of mITT trials illustrate different types of deviations[5][6]:

- Treatment related deviation: any description that was related to the treatment, independently of the dose. For example: "an intent-to-treat population defined as all patients that took at least one dose of study medication"[22] or "analyses included all modified intent to treat patients: that is, all patients randomized to treatment who took their assigned medication for 4 weeks or more".[23]

- Baseline assessment related deviation: a participant was excluded from the analysis in the absence of a baseline characteristic. For example: "An intent-to-treat analysis included data from all patients who . . . had at least one baseline evaluation".[24]

- Target condition related deviation: investigators use a specific condition as a recruitment criterion for their cohort of patients. For example, patients are required to have a specific outcome at entry: "a surgically proven complicated intra-abdominal infection",[25] which is difficult to determine or is only suspected at enrolment. In this case, investigators randomise patients, but when they are subsequently found not to have the specific outcome, the investigators exclude these patients from their analysis.

- Post-baseline assessment related deviation: authors use the presence of a post-baseline assessment to establish their analysis. For example: "All primary analyses used the intent-to-treat (ITT) principle using data on each randomized subject with at least 1 post-randomization measure".[26]

- Trials may fall into multiple categories depending on the number of deviations. For example: "intention to treat population which is defined as all patients who received the trial drug, provided baseline efficacy data, and from whom at least one measurement after baseline was obtained".[27]

Adjusted analyses

To assess the robustness of our primary analysis, we performed additional adjusted analyses to take into account several potential confounding factors that can influence the estimate of the treatment effect. The confounders were identified from the medical literature for which a meta-epidemiological assessment was made: type of centre (single centre v multicentre) in which the trials were conducted,[31] sample size (quarters within each meta-analysis[8] and small v big sample size[32][33]), presence of post-randomisation exclusions,[9] and items of the risk of bias tool (random sequence generation and allocation concealment, blinding of participants and personnel, blinding of outcome assessor, and incomplete outcome data),[34] as well as the type of funding. For this purpose, we performed a multivariable multilevel meta-epidemiological analysis following the approach suggested by Siersma and colleagues[29] (web appendix 4). To control for publication bias, we adjusted models for the variance of the log odds ratio for each randomised trial, as suggested by Moreno and colleagues.[35]

Subgroup analysis

Subgroup analyses included the type of intervention (pharmacological v non-pharmacological), type of intervention in the control group (placebo v non-placebo) and type of outcome (objective v subjective). We did interaction tests to assess whether ROR estimates varied by the type of intervention, use of placebo, and type of outcome. Analyses were done by Stata/SE, version 13 for Windows (StataCorp) with metan, metareg, and mixed routines.

Sensitivity analysis

To assess the robustness of our results, using a sensitivity analysis, we re-estimated differences in effect estimates (1) between mITT trials versus ITT trials or no ITT trials; and (2) between mITT trials and the combined results of ITT trials and no ITT trials. We used a two step, meta-epidemiological approach described by Sterne and colleagues.[7] In the first step, we performed a metaregression for each meta-analysis, estimating the ROR. In the second step, we meta-analysed the resulting ROR and the 95% confidence interval among meta-analyses using a DerSimonian and Laird random effects model (web appendix 4).

Patient involvement

There was no patient involvement in this study.

Results

Figure 1 describes the screening process for both systematic reviews and randomised trials. We identified 43 reviews with categorical data with 310 trials for analysis. Web appendix 5 lists excluded reviews and excluded trials with reasons.

Characteristics of the included reviews

Web appendix 6 shows characteristics of the selected reviews with meta-analyses.[15][36-77] The median number of trials per meta-analysis was seven (range 2-22) and the median number of participants per meta-analysis was 1269 (range 209-365 368). None of the meta-analyses evaluated a non-pharmacological intervention. The outcomes were judged objective in 19 (38%) of the meta-analyses, whereas in 34 (68%), the experimental intervention was compared with a placebo control. Pooled odds ratios from individual random-effects meta-analyses ranged from 0.15 to 1.25. In 49 meta-analyses (98%), the experimental intervention had an effect, but only 30 (60%) of these were significant. Thirty four (68%) of the meta-analyses showed substantial heterogeneity ($I^2 \geq 50\%$).

Characteristics of ITT trials, mITT trials, and no ITT trials

Of 310 primary studies included in the final analyses, 84 (27%) were categorised as ITT trials, 118 (38%) as mITT trials, and 108 (35%) as no ITT trials. The studies were published between 1970 and 2009. The three categories of trials differed in terms of several variables. Compared with ITT trials, mITT trials were more likely to report post-randomisation exclusions (cluster weighted χ^2, P=0.03), have high or unclear incomplete outcome data (P=0.03), and receive funding from private enterprises (P=0.02).

Compared with no ITT trials, mITT trials were more likely to be multicentre studies (cluster weighted χ^2, P<0.001), report post-randomisation exclusions (P=0.03), report sample size calculation (P<0.001), and receive funding from private enterprises (P<0.001). Year of publication was more recent for mITT trials than for no ITT trials: 79 (67%) mITT trials were published between 2001 and 2009 compared
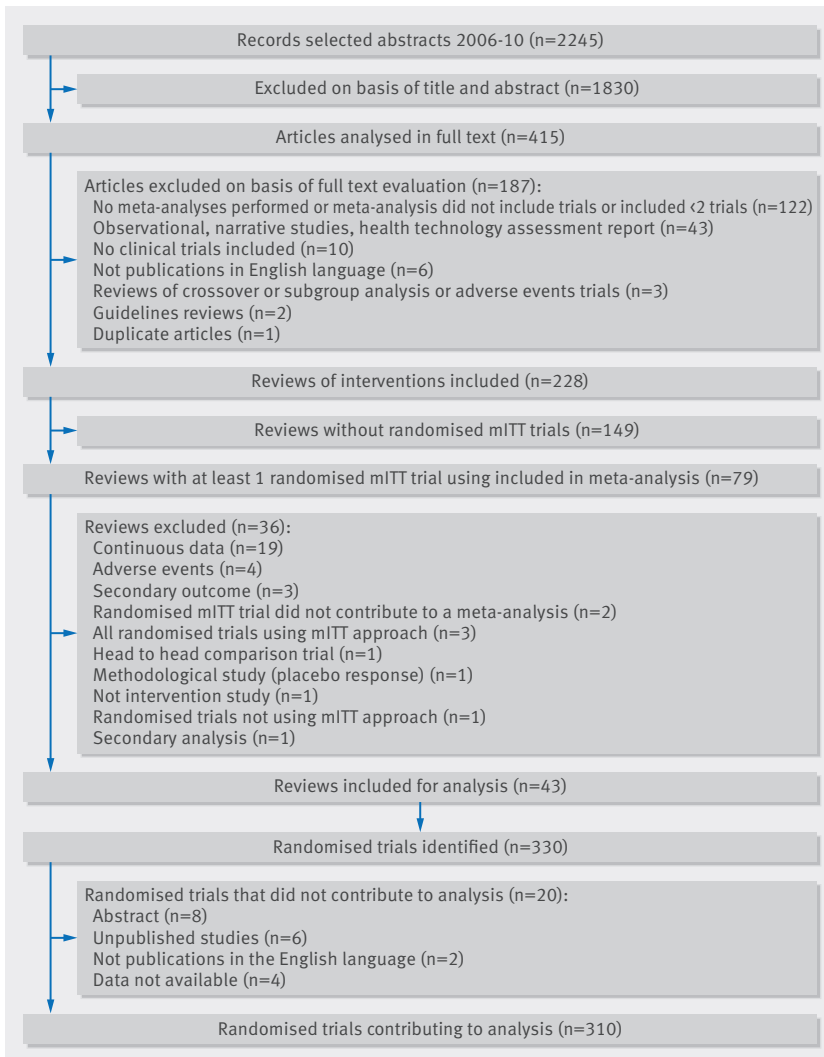
Fig 1 | Study screening process

with 34 (31%) no ITT trials (for trend, P<0.001). Studies classified as no ITT trials had a smaller sample size (median=83) than those classified as mITT trials (252) and ITT trials (272). The table shows the characteristics of included trials.

Types of deviation in mITT trials

Overall, of 118 trials classified as mITT, seven explicitly reported the intention to treat approach as "modified," and 111 described analyses of outcome data that deviated from a true intention to treat approach and thus were categorised as mITT trials. Sixty three (53%) trials reported one type of mITT deviation, and 55 (47%) reported more than one type of mITT deviation. In detail, 89 (75%) trials reported treatment related deviations (of which 47 had additional deviations); 56 (47%) reported a post-baseline assessment related deviation (of which 42 had at least one additional deviation); 22 (19%) reported a baseline assessment related deviation (of which 21 had at least one other type of deviation); and 15 (13%) used a target condition for the mITT deviation. Two (2%) trials with a protocol deviation did not fall into any of the previous categories and were classified as "others."

Estimate of the treatment effect among ITT trials, mITT trials, and no ITT trials

The 43 reviews provided 50 meta-analyses. Twelve trials contributed twice to the analyses; therefore, there were 322 comparisons in total. The treatment effect of mITT trials with respect to ITT trials was inflated by 17% (unadjusted ROR 0.83, 95% confidence interval 0.71 to 0.97; P=0.01; fig 2), with moderate variance between meta-analyses (τ^2=0.13). After adjusting the comparison for use of placebo, sample size, type of centre, items of risk of bias, post-randomisation exclusions, funding, and publication bias, the ROR was 0.80 (0.69 to 0.94; P=0.005; τ^2=0.08). The comparison between mITT trials and no ITT trials provided an ROR of 1.00 (0.75 to 1.33; P=0.99; τ^2=0.57; fig 2). This result changed after adjusting for potential confounders, but remained non-significant (0.92, 0.70 to 1.23).

Effect of heterogeneity

In a sensitivity analysis, we excluded five meta-analyses[15 47 62 63 76] that accounted for the most heterogeneity between the mITT and ITT trials comparison (fig 3). The results remained consistent, and we saw a reduction in the between meta-analyses variance (τ^2=0.08). The ROR between mITT and ITT trials, after excluding the meta-analyses, was 0.85 (95% confidence interval 0.75 to 0.98).

In the comparison between mITT trials and no ITT trials, we excluded seven studies[44 47 49 62 63 65] that contributed the most heterogeneity (fig 3). The between meta-analyses variance was 0.29, and the treatment effect between trial categories remained unchanged (ROR 1.21 (95% confidence interval 0.76 to 1.12)).

Subgroup analyses

For 19 meta-analyses assessing objective outcomes, the ROR between mITT trials and ITT trials was 0.80 (95% confidence interval 0.60 to 1.06; τ^2=0.18). For 31 meta-analyses assessing subjective outcomes, the corresponding result was 0.84 (0.70 to 1.01; τ^2=0.11). The interaction between the type of outcome and the difference between estimated treatment effect was not significant (P=0.96). For 34 meta-analyses of trials using a placebo, the ROR between mITT trials and ITT trials was 0.81 (95% confidence interval 0.82 to 1.00; τ^2=0.12). For 16 meta-analyses not using placebo, the corresponding results was 0.87 (0.69 to 1.08; τ^2=0.12). The interaction between the use of placebo (or not) and the difference between the estimated treatment effect was not significant (P=0.99).

Sensitivity analyses

Using the two step meta-epidemiological approach, we did sensitivity analyses on three comparisons. First, the estimate of the treatment effect was larger in mITT trials than in ITT trials (31 meta-analyses; ROR 0.84 (95% confidence interval 0.74 to 0.95, P=0.006); heterogeneity,

Characteristics of included randomised trials

Characteristic	Trials classified according to intention to treat category		
	ITT trials (n=84)	mITT trials (n=118)	No ITT trials (n=108)
Year of publication			
1970-94	7 (8)	7 (6)	53 (49)
1995-2000	30 (36)	32 (27)	21 (19)
2001-04	22 (26)	36 (31)	20 (19)
2005-09	25 (30)	43 (36)	14 (13)
Sample size			
Median (IQR)	272 (128-459)	252 (146-438)	83 (53-211)
No of allocated patients*			
>200	63 (75)	85 (72)	71 (66)
≤200	21 (25)	33 (28)	37 (34)
Sample size calculation			
Described	48 (57)	73 (62)	28 (26)
Not described	36 (43)	45 (38)	80 (74)
Use of placebo			
No	35 (42)	46 (39)	37 (34)
Yes	49 (58)	72 (61)	71 (66)
Type of centre			
Multicentre	69 (82)	111 (94)	60 (56)
Single centre	15 (18)	7 (6)	47 (44)
Use of flowchart			
No	55 (65)	67 (57)	99 (92)
Yes	29 (35)	51 (43)	9 (8)
Post-randomisation exclusion*			
No apparent exclusion	49 (56)	29 (24)	57 (51)
Exclusions	39 (44)	94 (76)	54 (49)
Proportion of post-randomisation exclusions†			
Median (IQR)	4.2 (0.9-9.4)	3.4 (1.4-6.4)	7.4 (3.3-16.1)
Risk of bias			
Sequence generation			
Low risk	28 (33)	36 (31)	24 (22)
High risk	0	0	0
Unclear	56 (67)	82 (69)	84 (78)
Allocation concealment			
Low risk	25 (30)	45 (38)	19 (18)
High risk	0	0	0
Unclear	59 (70)	73 (62)	89 (82)
Blinding of patients and personnel			
Low risk	39 (46)	62 (53)	41 (38)
High risk	21 (25)	18 (15)	20 (19)
Unclear	24 (29)	38 (32)	47 (44)
Blinding of outcome assessor			
Low risk	19 (23)	24 (20)	18 (17)
High risk	12 (14)	11 (9)	13 (12)
Unclear	53 (63)	83 (70)	77 (71)
Incomplete outcome data			
Low risk	39 (46)	32 (27)	32 (30)
High risk	5 (6)	30 (25)	24 (22)
Unclear	40 (48)	56 (47)	52 (48)
Funding source			
Not funded or public funding	11 (13)	2 (2)	12 (11)
Co-financed	9 (11)	9 (8)	8 (7)
For-profit funding	53 (63)	87 (74)	36 (33)
Not reported	11 (13)	20 (17)	52 (48)

Data are no (%) of trials, unless otherwise specified. IQR=interquartile range.
*Numbers and percentages are based on number of comparisons (n=322). Thus, exclusions and sample sizes might change owing to trials that contributed twice in the analyses.
†Calculation done within trials that performed exclusions.

$I^2=0\%$, P=0.51; $\tau^2=0.00$), corroborating results of the primary analysis (fig 4). The second comparison between mITT trials and no ITT trials also provided similar results to the primary analysis (28; 0.98 (0.79 to 1.23, P=0.9); $I^2=31\%$, P=0.10; $\tau^2=0.099$).

Figure 5 shows the third comparison, between the odds ratio of mITT trials and the combined odds ratios of ITT trials and no ITT trials (45 meta-analyses; ROR 0.90 (95% confidence interval 0.79 to 1.02, P=0.08); moderate heterogeneity ($I^2=27\%$; P=0.05); $\tau^2=0.05$). Using the approach by Baujat and colleagues,[30] we identified six influential studies[47 62 63 65 66] that accounted for most of the heterogeneity (fig 6). After excluding these trials, the ROR was 0.92 (0.82 to 1.02), P<0.08; $I^2=9.2\%$, P=0.3; $\tau^2=0.01$).

Discussion

Summary of findings

In this meta-epidemiological analysis, we compared the estimated treatment effect of 50 meta-analyses based on the description (reporting) of the intention to treat approach of 310 trials, comprising 322 comparisons. We found that in meta-analyses of trials that deviated from the intention to treat reporting, the treatment effect increased significantly. Our results remained consistent, after adjustment for risk of bias items, post-randomisation exclusions, and other potential sources of bias (such as sample size, type of centre, funding, and publication bias).

Strength and limitations

The current results are consistent with those of our previous meta-epidemiological assessment of the impact of modified intention to treat reporting, presented at the Cochrane Colloquium in 2009 in Singapore.[78] In that study, we assessed 355 trials and found that mITT trials overestimated the treatment effect by 18% compared with ITT trials (ROR 0.82 (95% confidence interval 0.72 to 0.96)).

The 2009 investigation had two principal limitations. First, the identified meta-analyses were limited to trials of specific interventions, such as treatment with antibiotics and cyclo-oxygenase-2 inhibitors. Second, we included trials that deviated from the intention to treat analysis without using the word "modified" as an inclusion criterion. To overcome these limitations and increase the generalisability of these findings, we decided not to limit the search to specific interventions and considered any type of deviation. Hence, the present study's main strength is that the results are based on the analysis of randomly selected reviews published over a five year period, representing a variety of medical conditions, and covering any deviation from the intention to treat reporting.

We acknowledge some limitations of our current study. First, our categorisation of trials was based on the authors' reporting. In addition, the classification of ITT trials was independent of the presence or absence of post-randomisation exclusions, hence we cannot claim that all the ITT trials actually followed a strict intention to treat analysis. Furthermore, mITT trials do
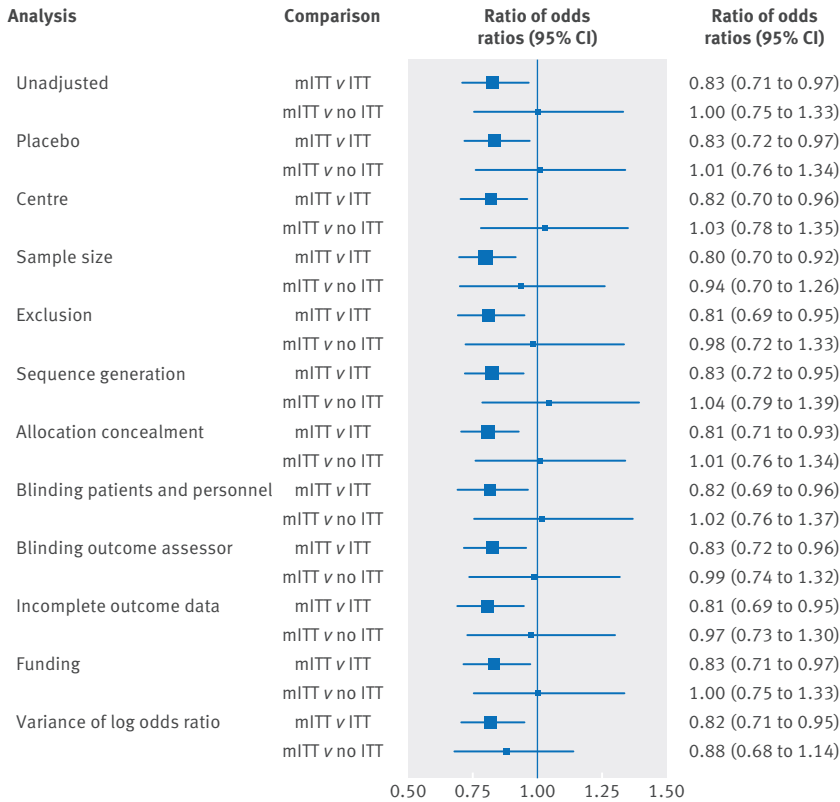
Analysis	Comparison	Ratio of odds ratios (95% CI)	Ratio of odds ratios (95% CI)
Unadjusted	mITT v ITT		0.83 (0.71 to 0.97)
	mITT v no ITT		1.00 (0.75 to 1.33)
Placebo	mITT v ITT		0.83 (0.72 to 0.97)
	mITT v no ITT		1.01 (0.76 to 1.34)
Centre	mITT v ITT		0.82 (0.70 to 0.96)
	mITT v no ITT		1.03 (0.78 to 1.35)
Sample size	mITT v ITT		0.80 (0.70 to 0.92)
	mITT v no ITT		0.94 (0.70 to 1.26)
Exclusion	mITT v ITT		0.81 (0.69 to 0.95)
	mITT v no ITT		0.98 (0.72 to 1.33)
Sequence generation	mITT v ITT		0.83 (0.72 to 0.95)
	mITT v no ITT		1.04 (0.79 to 1.39)
Allocation concealment	mITT v ITT		0.81 (0.71 to 0.93)
	mITT v no ITT		1.01 (0.76 to 1.34)
Blinding patients and personnel	mITT v ITT		0.82 (0.69 to 0.96)
	mITT v no ITT		1.02 (0.76 to 1.37)
Blinding outcome assessor	mITT v ITT		0.83 (0.72 to 0.96)
	mITT v no ITT		0.99 (0.74 to 1.32)
Incomplete outcome data	mITT v ITT		0.81 (0.69 to 0.95)
	mITT v no ITT		0.97 (0.73 to 1.30)
Funding	mITT v ITT		0.83 (0.71 to 0.97)
	mITT v no ITT		1.00 (0.75 to 1.33)
Variance of log odds ratio	mITT v ITT		0.82 (0.71 to 0.95)
	mITT v no ITT		0.88 (0.68 to 1.14)

0.50   0.75   1.00   1.25   1.50

Fig 2 | Differences in intervention effect estimates between mITT trials and ITT or no ITT trials. Unadjusted and adjusted analyses are shown

Fig 3 | Influence of meta-analyses contributing most heterogeneity in mITT v ITT and mITT v no ITT trial comparisons

exclusions,[9] although we cannot claim that it exhaustively deals with this issue.

Second, we only searched for articles in Medline, and therefore our results cannot be generalised to reviews that were published in other electronic databases. Third, we only analysed trials with binary outcomes, thus our results may not be applicable to trials with continuous data. Fourth, the number of prepublication protocols obtained was limited, and we cannot be sure whether reporting of the intention to treat analysis might have changed in the published study. Finally, our study was based on published information, which is a limitation common to other meta-epidemiological studies.

Deviation from the intention to treat analysis
Interpreting the use of deviation from intention to treat

The emerging phenomenon of trials with deviations can partly be attributed to the issue of missing data. Missing data is a common problem in randomised trials.[79] Experts report that participants with missing outcomes cannot be included in the analysis unless their outcomes are imputed.[80] In other words, intention to treat is not the best method to use; usually, the population analysed is known as "available cases" or "complete case"[80] and the analysis known as "efficacy analysis."[81] In our sample, 56 (47%) mITT trials, which had missing data, were those trials with a post-baseline assessment deviation. Authors reported different types of terminology to define their analysis such as "efficacy analysis"; or defined their evaluated population as "patients evaluable for efficacy,"[82] "intent-to-treat subset,"[83] or "population with observed cases,"[84] which may confuse readers.

Trial authors should be encouraged to explicitly state that the intention to treat analysis cannot be performed because of missing data and provide the analysis on "available cases," rather than use different nomenclatures to describe the intention to treat analysis or population. Researchers have proposed the use of multiple imputation as a general approach to deal with incomplete data.[85] Multiple imputation may conform to an intention to treat analysis trial,[80] although pitfalls and limitations must be accounted for and sensitivity analyses are strongly advocated.[80 86]

A separate issue related to deviation from the intention to treat approach is deviation from protocol, in which participants who did not adequately adhere to the protocol are excluded. This type of analysis is known as "per protocol," which is particularly useful for interpreting non-inferiority trials and, to some extent, for analysing the adverse effects of treatments. The reasons provided for deviation from protocol can be related to the amount of the treatment received, absence of a baseline assessment, or absence of a target specific condition at enrolment. After exclusion of the mITT trials with missing data in our study, 63 (53%) mITT trials reported a protocol deviation in which authors performed essentially a per protocol analysis.

A major issue arises if missing data and protocol deviation coexist in the mITT description. Authors sometimes do not provide a complete description of exclusions,

not have a single agreed definition, and their categorisation was again based on the reported description. Therefore, this classification was specific to our study. However, a strength of this study was that it provided an alternative approach to examine the bias related to
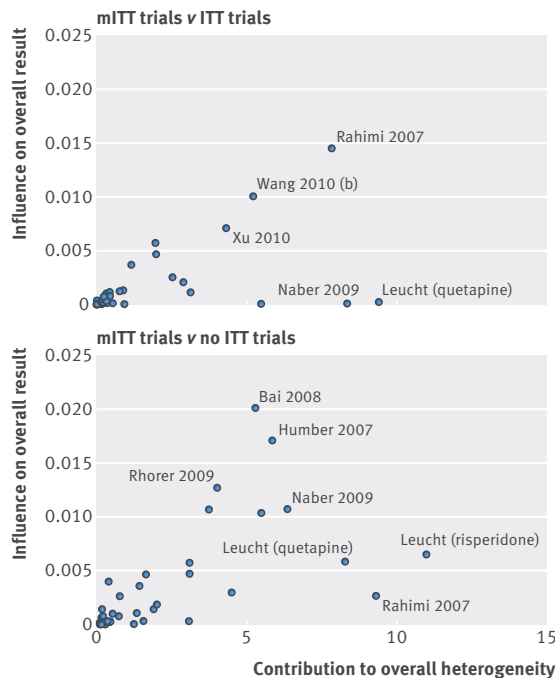
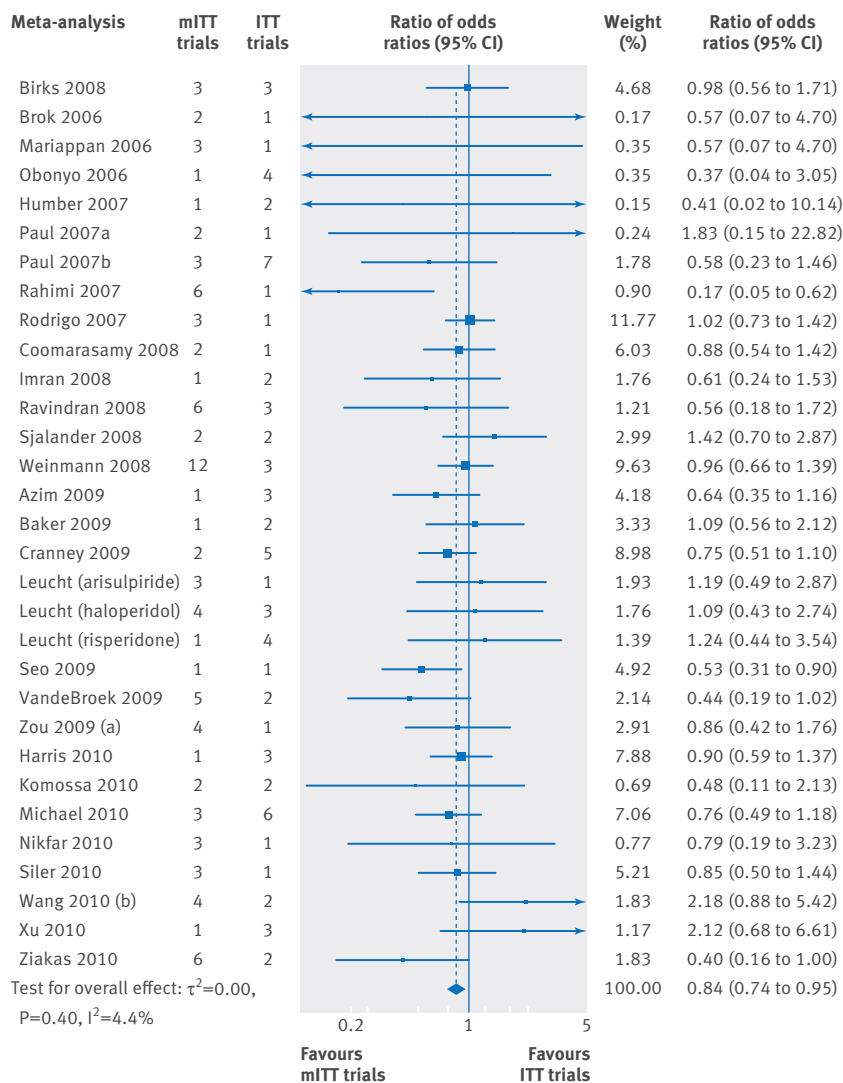| Meta-analysis	mITT trials	ITT trials	Ratio of odds ratios (95% CI)	Weight (%)	Ratio of odds ratios (95% CI)
Birks 2008	3	3		4.68	0.98 (0.56 to 1.71)
Brok 2006	2	1		0.17	0.57 (0.07 to 4.70)
Mariappan 2006	3	1		0.35	0.57 (0.07 to 4.70)
Obonyo 2006	1	4		0.35	0.37 (0.04 to 3.05)
Humber 2007	1	2		0.15	0.41 (0.02 to 10.14)
Paul 2007a	2	1		0.24	1.83 (0.15 to 22.82)
Paul 2007b	3	7		1.78	0.58 (0.23 to 1.46)
Rahimi 2007	6	1		0.90	0.17 (0.05 to 0.62)
Rodrigo 2007	3	1		11.77	1.02 (0.73 to 1.42)
Coomarasamy 2008	2	1		6.03	0.88 (0.54 to 1.42)
Imran 2008	1	2		1.76	0.61 (0.24 to 1.53)
Ravindran 2008	6	3		1.21	0.56 (0.18 to 1.72)
Sjalander 2008	2	2		2.99	1.42 (0.70 to 2.87)
Weinmann 2008	12	3		9.63	0.96 (0.66 to 1.39)
Azim 2009	1	3		4.18	0.64 (0.35 to 1.16)
Baker 2009	1	2		3.33	1.09 (0.56 to 2.12)
Cranney 2009	2	5		8.98	0.75 (0.51 to 1.10)
Leucht (arisulpiride)	3	1		1.93	1.19 (0.49 to 2.87)
Leucht (haloperidol)	4	3		1.76	1.09 (0.43 to 2.74)
Leucht (risperidone)	1	4		1.39	1.24 (0.44 to 3.54)
Seo 2009	1	1		4.92	0.53 (0.31 to 0.90)
VandeBroek 2009	5	2		2.14	0.44 (0.19 to 1.02)
Zou 2009 (a)	4	1		2.91	0.86 (0.42 to 1.76)
Harris 2010	1	3		7.88	0.90 (0.59 to 1.37)
Komossa 2010	2	2		0.69	0.48 (0.11 to 2.13)
Michael 2010	3	6		7.06	0.76 (0.49 to 1.18)
Nikfar 2010	3	1		0.77	0.79 (0.19 to 3.23)
Siler 2010	3	1		5.21	0.85 (0.50 to 1.44)
Wang 2010 (b)	4	2		1.83	2.18 (0.88 to 5.42)
Xu 2010	1	3		1.17	2.12 (0.68 to 6.61)
Ziakas 2010	6	2		1.83	0.40 (0.16 to 1.00)
Test for overall effect: τ^2=0.00, P=0.40, I^2=4.4%				100.00	0.84 (0.74 to 0.95)

Fig 4 | Differences in treatment effect between mITT trials and ITT trials using the two step meta-epidemiological approach

creating difficulties for readers to make assumptions. In 42 (36%) of mITT trials in our study, there were descriptions that could be related both to missing data and deviation from protocol. In some circumstances, authors reported clearly who deviated from protocol and were missing from follow-up (for example, "Among the 227 patients included in the study, one patient on revision was found not to fulfil the inclusion criteria and a further three patients were lost to follow up. This left 223 patients for intention to treat analysis"[87]). However, in other instances, no clear detail was provided (for example, "4 in the venlafaxine group, 1 in the fluoxetine group, and 1 in the placebo group could not be included in the efficacy analysis due to lack of efficacy assessments either at pretreatment or posttreatment"[88]).

Post-randomisation exclusions and deviation from the intention to treat approach
Based on descriptions of the deviations reported in mITT trials in our study, exclusions seem to be a natural consequence of a deviation. However, 34 (28%) mITT

trials in our study did not report exclusions, but at the same time did not report absence of exclusions.

In our analysis, we decided not to use the occurrence of exclusions to classify trials because a study may declare an intention to treat approach but exclude patients without transparently reporting those exclusions,[89] leading to an evident misclassification. ITT trials remained conservative in terms of the treatment effect despite half of them reporting exclusions. One possibility is that the exclusions performed by ITT trials and mITT trials have different weights (impacts), and exclusions in mITT trials exert a major influence on the exaggerated estimate of the treatment effect in meta-analyses. Further empirical evidence is needed to clarify this issue, for example, by explicitly defining the criteria to exclude patients and correlating these with the type of intention to treat analysis.[90]

Comparison with other studies
The bias related to trial analysis has been investigated in different ways and the magnitude and direction of the bias may vary depending on the clinical topic examined, definition of the exclusions, and methods used for analysis.[89-92] In an analysis of individual patient data of 14 meta-analyses on cancer, Tierney and Stewart found that the experimental intervention had a positive effect when exclusions were performed, compared with the effect measured when a true intention to treat analysis was done.[91] In a systematic review of randomised trials that used both intention to treat and per protocol analyses, Porta and colleagues found that, on average, per protocol estimates were higher than intention to treat estimates.[93] Similarly, Melander and colleagues assessed placebo controlled trials of serotonin reuptake inhibitors and found that drug treatment had a favourable effect when a per protocol analysis was used instead of an intention to treat analysis.[94] More recently, a meta-epidemiological study investigated the impact of exclusions on 14 meta-analyses of osteoarthritis randomised trials, and found that exclusion of patients from the analysis of trials could bias the estimates of treatment effects; however, the direction and magnitude of the bias were difficult to predict.[9]

Conclusion
Deviation from an intention to treat analysis in randomised trials is a potential source of biased estimates of treatment effects. Clinicians (and, by extension, patients) should be cautious when interpreting results from a modified intention to treat approach, or a deviation from the intention to treat approach, as the sole primary analysis. Trialists should also be aware that deviation from intention to treat analysis could lead to an overestimation of the treatment effect. Trialists should explicitly report who is included in the analysis as suggested by the CONSORT statement. Authors of systematic reviews should routinely assess the reporting of the intention to treat analysis. In the case of deviations and where possible, review authors should perform a sensitivity analysis—with and without the included mITT trials—when meta-analysing data.
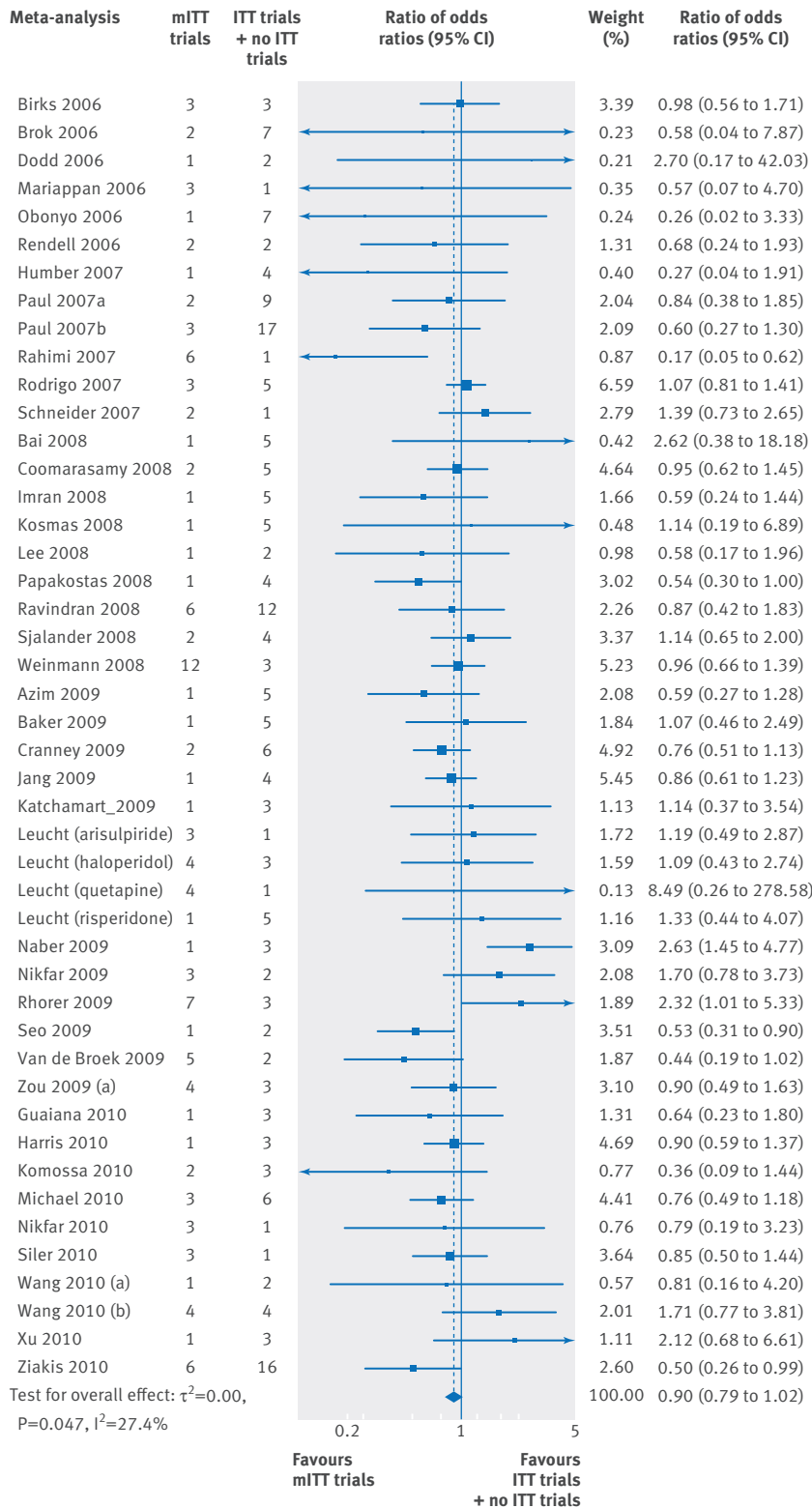
Meta-analysis	mITT trials	ITT trials + no ITT trials	Ratio of odds ratios (95% CI)	Weight (%)	Ratio of odds ratios (95% CI)
Birks 2006	3	3		3.39	0.98 (0.56 to 1.71)
Brok 2006	2	7		0.23	0.58 (0.04 to 7.87)
Dodd 2006	1	2		0.21	2.70 (0.17 to 42.03)
Mariappan 2006	3	1		0.35	0.57 (0.07 to 4.70)
Obonyo 2006	1	7		0.24	0.26 (0.02 to 3.33)
Rendell 2006	2	2		1.31	0.68 (0.24 to 1.93)
Humber 2007	1	4		0.40	0.27 (0.04 to 1.91)
Paul 2007a	2	9		2.04	0.84 (0.38 to 1.85)
Paul 2007b	3	17		2.09	0.60 (0.27 to 1.30)
Rahimi 2007	6	1		0.87	0.17 (0.05 to 0.62)
Rodrigo 2007	3	5		6.59	1.07 (0.81 to 1.41)
Schneider 2007	2	1		2.79	1.39 (0.73 to 2.65)
Bai 2008	1	5		0.42	2.62 (0.38 to 18.18)
Coomarasamy 2008	2	5		4.64	0.95 (0.62 to 1.45)
Imran 2008	1	5		1.66	0.59 (0.24 to 1.44)
Kosmas 2008	1	5		0.48	1.14 (0.19 to 6.89)
Lee 2008	1	2		0.98	0.58 (0.17 to 1.96)
Papakostas 2008	1	4		3.02	0.54 (0.30 to 1.00)
Ravindran 2008	6	12		2.26	0.87 (0.42 to 1.83)
Sjalander 2008	2	4		3.37	1.14 (0.65 to 2.00)
Weinmann 2008	12	3		5.23	0.96 (0.66 to 1.39)
Azim 2009	1	5		2.08	0.59 (0.27 to 1.28)
Baker 2009	1	5		1.84	1.07 (0.46 to 2.49)
Cranney 2009	2	6		4.92	0.76 (0.51 to 1.13)
Jang 2009	1	4		5.45	0.86 (0.61 to 1.23)
Katchamart_2009	1	3		1.13	1.14 (0.37 to 3.54)
Leucht (arisulpiride)	3	1		1.72	1.19 (0.49 to 2.87)
Leucht (haloperidol)	4	3		1.59	1.09 (0.43 to 2.74)
Leucht (quetapine)	4	1		0.13	8.49 (0.26 to 278.58)
Leucht (risperidone)	1	5		1.16	1.33 (0.44 to 4.07)
Naber 2009	1	3		3.09	2.63 (1.45 to 4.77)
Nikfar 2009	3	2		2.08	1.70 (0.78 to 3.73)
Rhorer 2009	7	3		1.89	2.32 (1.01 to 5.33)
Seo 2009	1	2		3.51	0.53 (0.31 to 0.90)
Van de Broek 2009	5	2		1.87	0.44 (0.19 to 1.02)
Zou 2009 (a)	4	3		3.10	0.90 (0.49 to 1.63)
Guaiana 2010	1	3		1.31	0.64 (0.23 to 1.80)
Harris 2010	1	3		4.69	0.90 (0.59 to 1.37)
Komossa 2010	2	3		0.77	0.36 (0.09 to 1.44)
Michael 2010	3	6		4.41	0.76 (0.49 to 1.18)
Nikfar 2010	3	1		0.76	0.79 (0.19 to 3.23)
Siler 2010	3	1		3.64	0.85 (0.50 to 1.44)
Wang 2010 (a)	1	2		0.57	0.81 (0.16 to 4.20)
Wang 2010 (b)	4	4		2.01	1.71 (0.77 to 3.81)
Xu 2010	1	3		1.11	2.12 (0.68 to 6.61)
Ziakis 2010	6	16		2.60	0.50 (0.26 to 0.99)
Test for overall effect: τ^2=0.00, P=0.047, I^2=27.4%				100.00	0.90 (0.79 to 1.02)

0.2   1   5

Favours mITT trials   Favours ITT trials + no ITT trials

Fig 5 | Differences in treatment effect between mITT trials and ITT/no ITT trials combined using the two step meta-epidemiological approach

Finally, ethics committee members have a unique opportunity, compared to other professionals, to detect potential methodological flaws in study designs, including the recognition of potential deviation from the intention to treat analysis.
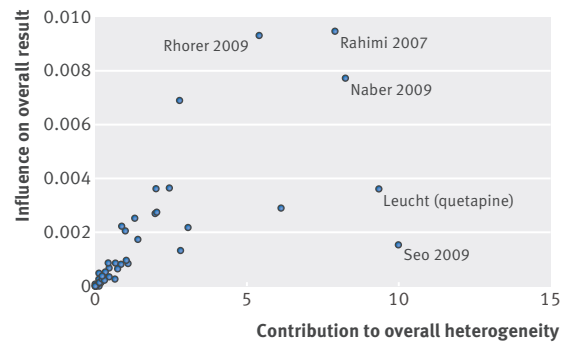
Fig 6 | Influence of meta-analyses contributing most heterogeneity in mITT v ITT/no ITT trial comparison

1 Wood AM, White IR, Thompson SG. Are missing outcome data adequately handled? A review of published randomized controlled trials in major medical journals. Clin Trials 2004;1:368-76.
2 Fergusson D, Aaron SD, Guyatt G, Hebert P. Post-randomisation exclusions: the intention to treat principle and excluding patients from analysis. BMJ 2002;325:652-4.
3 Newell DJ. Intention-to-treat analysis: implications for quantitative and qualitative research. Int J Epidemiol 1992;21:837-41.
4 Lewis JA, Machin D. Intention to treat—who should use ITT? Br J Cancer 1993;68:647-50.
5 Abraha I, Montedori A. Modified intention to treat reporting in randomised controlled trials: a systematic review. BMJ 2010:340;c2697.
6 Montedori A, Bonacini MI, Casazza G, et al. Modified versus standard intention-to-treat reporting: are there differences in methodological quality, sponsorship, and findings in randomized trials? A cross-sectional study. Trials 2011;12:58.
7 Sterne JA, Juni P, Schulz KF, Altman DG, Bartlett C, Egger M. Statistical methods for assessing the influence of study characteristics on treatment effects in 'meta-epidemiological' research. Stat Med 2002;21:1513-24.
8 Dechartres A, Trinquart L, Boutron I, Ravaud P. Influence of trial sample size on treatment effect estimates: meta-epidemiological study. BMJ 2013;346:f2304.

9   Nuesch E, Trelle S, Reichenbach S, et al. The effects of excluding patients from the analysis in randomised controlled trials: meta-epidemiological study. *BMJ* 2009;339:b3244.

10 Montori VM, Wilczynski NL, Morgan D, Haynes RB, Hedges Team. Optimal search strategies for retrieving systematic reviews from Medline: analytical survey. *BMJ* 2005;330:68.

11 Avouac J, Wipff J, Kahan A, Allanore Y. Effects of oral treatments on exercise capacity in systemic sclerosis related pulmonary arterial hypertension: a meta-analysis of randomised controlled trials. *Ann Rheum Dis* 2008;67:808-14.

12 Gao F, Liao C, Liu L, Tan A, Cao Y, Mo Z. The effect of aspirin in the recurrence of colorectal adenomas: a meta-analysis of randomized controlled trials. *Colorectal Dis* 2009;11:893-901.

13 Vardakas KZ, Samonis G, Michalopoulos A, Soteriades ES, Falagas ME. Antifungal prophylaxis with azoles in high-risk, surgical intensive care unit patients: a meta-analysis of randomized, placebo-controlled trials. *Crit Care Med* 2006;34:1216-24.

14 Sparano JA, Makhson AN, Semiglazov VF, et al. Pegylated liposomal doxorubicin plus docetaxel significantly improves time to progression without additive cardiotoxicity compared with docetaxel monotherapy in patients with advanced breast cancer previously treated with neoadjuvant-adjuvant anthracycline therapy: results from a randomized phase III study. *J Clin Oncol* 2009;27:4522-9.

15 Xu HB, Xu Q, Li L. A literature-based meta-analysis taxane-based doublet versus single-agent taxane chemotherapy in patients with advanced breast cancer. *J Cancer Res Clin Oncol* 2011;137:1005-13.

16 Hollander E, Baldini Rossi N, Sood E, Pallanti S. Risperidone augmentation in treatment-resistant obsessive-compulsive disorder: a double-blind, placebo-controlled study. *Int J Neuropsychopharmacol* 2003;6:397-401.

17 Thigpen JT, Brady MF, Homesley HD, et al. Phase III trial of doxorubicin with or without cisplatin in advanced endometrial carcinoma: a gynecologic oncology group study. *J Clin Oncol* 2004;22:3902-8.

18 Cameron D, Casey M, Press M, et al. A phase III randomized comparison of lapatinib plus capecitabine versus capecitabine alone in women with advanced breast cancer that has progressed on trastuzumab: updated efficacy and biomarker analyses. *Breast Cancer Res Treat* 2008;112:533-43.

19 Asakura S, Tajima O, Koyama T. Fluvoxamine treatment of generalized social anxiety disorder in Japan: a randomized double-blind, placebo-controlled study. *Int J Neuropsychopharmacol* 2007;10:263-74.

20 Kärkkäinen MK, Tuppurainen M, Salovaara K, et al. Does daily vitamin D 800IU and calcium 1000mg supplementation decrease the risk of falling in ambulatory women aged 65–71 years? A 3-year randomized population-based trial (OSTPRE-FPS). *Maturitas* 2010;65:359-65.

21 Pederzoli P, Bassi C, Vesentini S, Campedeli A. A randomized multicenter clinical trial of antibiotic prophylaxis of septic complications in acute necrotizing pancreatitis with imipenem. *Surg Gynecol Obstet* 1993;176:480-3.

22 McKenna F, Borenstein D, Wendt H, Wallemark C, Lefkowith JB, Geis GS. Celecoxib versus diclofenac in the management of osteoarthritis of the knee. *Scand J Rheumatol* 2001;30:11-8.

23 Navarro V, Gasto C, Torres X, Marcos T, Pintor L. Citalopram versus nortriptyline in late-life depression: a 12-week randomized single-blind study. *Acta Psychiatr Scand* 2001;103:435-40.

24 Benkert O, Grunder G, Wetzel H, Hackett D. A randomized, double-blind comparison of a rapidly escalating dose of venlafaxine and imipramine in inpatients with major depression and melancholia. *J Psychiatr Res* 1996;30:441-51.

25 Dupont H, Carbon C, Carlet J. Monotherapy with a broad-spectrum beta-lactam is as effective as its combination with an aminoglycoside in treatment of severe generalized peritonitis: a multicenter randomized controlled trial. The Severe Generalized Peritonitis Study Group. *Antimicrob Agents Chemother* 2000;44:2028-33.

26 Dmochowski RR, Miklos JR, Norton PA, Zinner NR, Yalcin I, Bump RC. Duloxetine versus placebo for the treatment of North American women with stress urinary incontinence. *J Urol* 2003;170:1259-63.

27 Fraser SM, Hopkins R, Hunter JA, Neumann V, Capell HA, Bird HA. Sulphasalazine in the management of psoriatic arthritis. *Br J Rheumatol* 1993;32:923-5.

28 Higgins JPT, Altman DG. Assessing risk of bias in included studies. Cochrane handbook for systematic reviews of interventions. John Wiley and Sons, 2008:187-241.

29 Siersma V, Als-Nielsen B, Chen W, Hilden J, Gluud LL, Gluud C. Multivariable modelling for meta-epidemiological assessment of the association between trial quality and treatment effects estimated in randomized clinical trials. *Stat Med* 2007;26:2745-58.

30 Baujat B, Mahe C, Pignon JP, Hill C. A graphical method for exploring heterogeneity in meta-analyses: application to a meta-analysis of 65 trials. *Stat Med* 2002;21:2641-52.

31 Dechartres A, Boutron I, Trinquart L, Charles P, Ravaud P. Single-center trials show larger treatment effects than multicenter trials: evidence from a meta-epidemiologic study. *Ann Intern Med* 2011;155:39-51.

32 Nuesch E, Trelle S, Reichenbach S, et al. Small study effects in meta-analyses of osteoarthritis trials: meta-epidemiological study. *BMJ* 2010;341:c3515.

33 Papageorgiou SN, Antonoglou GN, Tsiranidou E, Jepsen S, Jager A. Bias and small-study effects influence treatment effect estimates: a meta-epidemiological study in oral medicine. *J Clin Epidemiol* 2014;67:984-92.

34 Savovic J, Jones H, Altman D, et al. Influence of reported study design characteristics on intervention effect estimates from randomised controlled trials: combined analysis of meta-epidemiological studies. *Health Technol Assess* 2012;16:1-82.

35 Moreno SG, Sutton AJ, Ades AE, et al. Assessment of regression-based methods to adjust for publication bias through a comprehensive simulation study. *BMC Med Res Methodol* 2009;9:2.

36 Birks J. Cholinesterase inhibitors for Alzheimer's disease. *Cochrane Database Syst Rev* 2006;1:CD005593.

37 Brok J, Gluud LL, Gluud C. Ribavirin monotherapy for chronic hepatitis C infection: a Cochrane Hepato-Biliary Group systematic review and meta-analysis of randomized trials. *Am J Gastroenterol* 2006;101:842-7.

38 Dodd JM, Flenady VF, Cincotta R, Crowther CA. Prenatal administration of progesterone for preventing preterm birth. *Cochrane Database Syst Rev* 2006;1:CD004947.

39 Jefferson T, Demicheli V, Rivetti D, Jones M, Di Pietrantonj C, Rivetti A. Antivirals for influenza in healthy adults: systematic review. *Lancet* 2006;367:303-13.

40 Obonyo CO, Lau J. Efficacy of Haemophilus influenzae type b vaccination of children: a meta-analysis. *Eur J Clin Microbiol Infect Dis* 2006;25:90-7.

41 Paul M, Silbiger I, Grozinsky S, Soares-Weiser K, Leibovici L. Beta lactam antibiotic monotherapy versus beta lactam-aminoglycoside antibiotic combination therapy for sepsis. *Cochrane Database Syst Rev* 2006;1:CD003344.

42 Rendell JM, Gijsman HJ, Bauer MS, Goodwin GM, Geddes GR. Risperidone alone or in combination for acute mania. *Cochrane Database Syst Rev* 2006;1:CD004043.

43 Schneider LS, Dagerman K, Insel PS. Efficacy and adverse effects of atypical antipsychotics for dementia: meta-analysis of randomized, placebo-controlled trials. *Am J Geriatr Psychiatry* 2006;14:191-210.

44 Humber CE, Tierney JF, Symonds RP, et al. Chemotherapy for advanced, recurrent or metastatic endometrial cancer: a systematic review of Cochrane collaboration. *Ann Oncol* 2007;18:409-20.

45 Kosmas IP, Tatsioni A, Kolibianakis EM, et al. Effects and clinical significance of GnRH antagonist administration for IUI timing in FSH superovulated cycles: a meta-analysis. *Fertil Steril* 2008;90:367-72.

46 Mariappan P, Alhasso A, Ballantyne Z, Grant A, N'Dow J. Duloxetine, a serotonin and noradrenaline reuptake inhibitor (SNRI) for the treatment of stress urinary incontinence: a systematic review. *Eur Urol* 2007;51:67-74.

47 Rahimi R, Nikfar S, Abdollahi M. Do anti-tumor necrosis factors induce response and remission in patients with acute refractory Crohn's disease? A systematic meta-analysis of controlled clinical trials. *Biomed Pharmacother* 2007;61:75-80.

48 Rodrigo GJ, Nannini LJ. Tiotropium for the treatment of stable chronic obstructive pulmonary disease: a systematic review with meta-analysis. *Pulm Pharmacol Ther* 2007;20:495-502.

49 Bai Y, Gao J, Zou DW, Li ZS. Prophylactic antibiotics cannot reduce infected pancreatic necrosis and mortality in acute necrotizing pancreatitis: evidence from a meta-analysis of randomized controlled trials. *Am J Gastroenterol* 2008;103:104-10.

50 Coomarasamy A, Afnan M, Cheema D, van der Veen F, Bossuyt PM, van Wely M. Urinary hMG versus recombinant FSH for controlled ovarian hyperstimulation following an agonist long down-regulation protocol in IVF or ICSI treatment: a systematic review and meta-analysis. *Hum Reprod* 2008;23:310-5.

51 Imran H, Tleyjeh IM, Arndt CA, et al. Fluoroquinolone prophylaxis in patients with neutropenia: a meta-analysis of randomized placebo-controlled trials. *Eur J Clin Microbiol Infect Dis* 2008;27:53-63.

52 Lee YH, Woo JH, Rho YH, Choi SJ, Ji JD, Song GG. Meta-analysis of the combination of TNF inhibitors plus MTX compared to MTX monotherapy, and the adjusted indirect comparison of TNF inhibitors in patients suffering from active rheumatoid arthritis. *Rheumatol Int* 2008;28:553-9.

53 Papakostas GI, Nelson JC, Kasper S, Moller HJ. A meta-analysis of clinical trials comparing reboxetine, a norepinephrine reuptake inhibitor, with selective serotonin reuptake inhibitors for the treatment of major depressive disorder. *Eur Neuropsychopharmacol* 2008;18:122-7.

54 Ravindran V, Scott DL, Choy EH. A systematic review and meta-analysis of efficacy and toxicity of disease modifying anti-rheumatic drugs and biological agents for psoriatic arthritis. *Ann Rheum Dis* 2008;67:855-9.

55 Sjalander A, Jansson JH, Bergqvist D, Eriksson H, Carlberg B, Svensson P. Efficacy and safety of anticoagulant prophylaxis to prevent venous thromboembolism in acutely ill medical inpatients: a meta-analysis. *J Intern Med* 2008;263:52-60.

56 Weinmann S, Becker T, Koesters M. Re-evaluation of the efficacy and tolerability of venlafaxine vs SSRI: meta-analysis. *Psychopharmacology* 2008;196:511-20.

57 Azim HA Jr, Elattar I, Loberiza FR Jr, Azim H, Mok T, Ganti AK. Third generation triplet cytotoxic chemotherapy in advanced non-small cell lung cancer: a systematic overview. *Lung Cancer* 2009;64:194-8.

58 Baker WL, Silver D, White CM, et al. Dopamine agonists in the treatment of early Parkinson's disease: a meta-analysis. *Parkinsonism Relat Disord* 2009;15:287-94.

59 Cranney A, Wells GA, Yetisir E, et al. Ibandronate for the prevention of nonvertebral fractures: a pooled analysis of individual patient data. *Osteoporos Int* 2009;20:291-7.

60 Jang YG, Ilodigwe D, Macdonald RL. Metaanalysis of tirilazad mesylate in patients with aneurysmal subarachnoid hemorrhage. *Neurocrit Care* 2009;10:141-7.

61 Katchamart W, Trudeau J, Phumethum V, Bombardier C. Efficacy and toxicity of methotrexate (MTX) monotherapy versus MTX combination therapy with non-biological disease-modifying antirheumatic drugs in rheumatoid arthritis: a systematic review and meta-analysis. *Ann Rheum Dis* 2009;68:1105-12.

62 Leucht S, Arbter D, Engel RR, Kissling W, Davis JM. How effective are second-generation antipsychotic drugs? A meta-analysis of placebo-controlled trials. *Mol Psychiatry* 2009;14:429-47.

63 Naber KG, Cho YH, Matsumoto T, Schaeffer AJ. Immunoactive prophylaxis of recurrent urinary tract infections: a meta-analysis. *Int J Antimicrob Agents* 2009;33:111-9.

64 Nikfar S, Rahimi R, Rezaie A, Abdollahi M. A meta-analysis of the efficacy of sulfasalazine in comparison with 5-aminosalicylates in the induction of improvement and maintenance of remission in patients with ulcerative colitis. *Dig Dis Sci* 2009;54:1157-70.

65 Rhorer J, Ambrose CS, Dickinson S, et al. Efficacy of live attenuated influenza vaccine in children: a meta-analysis of nine randomized clinical trials. *Vaccine* 2009;27:1101-10.

66 Seo JH, Kim YH, Kim JS. Meta-analysis of pre-operative aromatase inhibitor versus tamoxifen in postmenopausal woman with hormone receptor-positive breast cancer. *Cancer Chemother Pharmacol* 2009;63:261-6.

67 van den Broek WW, Mulder PG, van Os E, Birkenhager TK, Pluijms E, Bruijn JA. Efficacy of venlafaxine compared with tricyclic antidepressants in depressive disorder: a meta-analysis. *J Psychopharmacol* 2009;23:708-13.

68 Zou Z, Xi GL, Yuan HB, Zhu QF, Shi XY. Telmisartan versus angiotension-converting enzyme inhibitors in the treatment of hypertension: a meta-analysis of randomized controlled trials. *J Hum Hypertens* 2009;23:339-49.

69 Guaiana G, Barbui C, Cipriani A. Hydroxyzine for generalised anxiety disorder. *Cochrane Database Syst Rev* 2010:CD006815.

70 Harris CA, Ward RL, Dobbins TA, Drew AK, Pearson S. The efficacy of HER2-targeted agents in metastatic breast cancer: a meta-analysis. *Ann Oncol* 2010;22:1308-17.

71 Komossa K, Depping AM, Meyer M, Kissling W, Leucht S. Second-generation antipsychotics for obsessive compulsive disorder. *Cochrane Database Syst Rev* 2010;12:CD008141.

72 Michael YL, Whitlock EP, Lin JS, Fu R, O'Connor EA, Gold R. Primary care-relevant interventions to prevent falling in older adults: a systematic evidence review for the U.S. Preventive Services Task Force. *Ann Intern Med* 2010;153:815-25.

73 Nikfar S, Mirfazaelian H, Abdollahi M. Efficacy and tolerability of immunoregulators and antibiotics in fistulizing Crohn's disease: a systematic review and meta-analysis of placebo-controlled trials. *Curr Pharm Des* 2010;16:3684-98.

74 Roskell NS, Beard SM, Zhao Y, Le TK. A meta-analysis of pain response in the treatment of fibromyalgia. Pain Pract 2011;11:516-27.

75 Siler AC, Gardner H, Yanit K, Cushman T, McDonagh M. Systematic review of the comparative effectiveness of antiepileptic drugs for fibromyalgia. *J Pain* 2011;12:407-15.

76 Wang XF, Li JD, Peng Y, Dai Y, Shi G, Xu W. Interleukin-2 receptor antagonists in liver transplantation: a meta-analysis of randomized trials. *Transplant Proc* 2010;42:4567-72.

77 Ziakas PD, Kourbeti IS, Voulgarelis M, Mylonakis E. Effectiveness of systemic antifungal prophylaxis in patients with neutropenia after chemotherapy: a meta-analysis of randomized controlled trials. *Clin Ther* 2010;32:2316-36.

78 Abraha I, Duca P, Montedori A. Impact of randomised controlled trials reporting modified intention to treat on meta-analysis: preliminary results. Oral presentation at the 17th Cochrane Colloquium; 2009 Oct 11-14, Singapore [abstr]. *Cochrane Database Syst Rev* 2009;(suppl CD000001):7.

79 Gravel J, Opatrny L, Shapiro S. The intention-to-treat approach in randomized controlled trials: are authors saying what they do and doing what they say? *Clin Trials* 2007;4:350-6.

80 Moher D, Hopewell S, Schulz KF, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010;340:c869.

81 Lachin JM. Statistical considerations in the intent-to-treat principle. *Control Clin Trials* 2000;21:167-89.

82 Rotstein C, Bow EJ, Laverdiere M, Ioannou S, Carr D, Moghaddam N. Randomized placebo-controlled trial of fluconazole prophylaxis for neutropenic cancer patients: benefit based on purpose and intensity of cytotoxic therapy. The Canadian Fluconazole Prophylaxis Study Group. *Clin Infect Dis* 1999;28:331-40.

83 Rudolph RL, Feiger AD. A double-blind, randomized, placebo-controlled trial of once-daily venlafaxine extended release (XR) and fluoxetine for the treatment of depression. *J Affect Disord* 1999;56:171-81.

84 Leizorovicz A, Cohen AT, Turpie AG, Olsson CG, Vaitkus PT, Goldhaber SZ. Randomized, placebo-controlled trial of dalteparin for the prevention of venous thromboembolism in acutely ill medical patients. *Circulation* 2004;110:874-9.

85 Sterne JA, White IR, Carlin JB, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ* 2009;338:b2393.

86 Molenberghs G, Thijs H, Jansen I, et al. Analyzing incomplete longitudinal clinical trial data. *Biostatistics* 2004;5:445-64.

87 Kiilerich S, Ladefoged K, Rannem T, Ranlov PJ. Prophylactic effects of olsalazine v sulphasalazine during 12 months maintenance treatment of ulcerative colitis. The Danish Olsalazine Study Group. *Gut* 1992;33:252-5.

88 Nemeroff CB, Thase ME. A double-blind, placebo-controlled comparison of venlafaxine and fluoxetine treatment in depressed outpatients. *J Psychiatr Res* 2007;41:351-9.

89 Schulz KF, Grimes DA, Altman DG, Hayes RJ. Blinding and exclusions after allocation in randomised controlled trials: survey of published parallel group trials in obstetrics and gynaecology. *BMJ* 1996;312:742-4.

90 Akl EA, Briel M, You JJ, et al. Potential impact on estimated treatment effects of information lost to follow-up in randomised controlled trials (LOST-IT): systematic review. *BMJ* 2012;344:e2809.

91 Tierney JF, Stewart LA. Investigating patient exclusion bias in meta-analysis. *Int J Epidemiol* 2005;34:79-87.

92 Balk EM, Bonis PA, Moskowitz H, et al. Correlation of quality measures with estimates of treatment effect in meta-analyses of randomized controlled trials. *JAMA* 2002;287:2973-82.

93 Porta N, Bonet C, Cobo E. Discordance between reported intention-to-treat and per protocol analyses. *J Clin Epidemiol* 2007;60:663-9.

94 Melander H, Ahlqvist-Rastad J, Meijer G, Beermann B. Evidence b(i)ased medicine—selective reporting from studies sponsored by pharmaceutical industry: review of studies in new drug applications. *BMJ* 2003;326:1171-3.

Web appendix 1: List of included randomised trials with related review and classification according to the type of intention to treat reporting

Web appendix 2: Definitions used for assessing risk of bias items in individual randomised trials

Web appendix 3: Protocol search and assessment

Web appendix 4: Statistical analyses methodology

Web appendix 5: List of excluded reviews with reasons

Web appendix 6: Characteristics of included meta-analyses