NO EVIDENCE OF A TRIAL EFFECT OUTSIDE OF TREATMENT ACCESS

DO PATIENTS MANAGED WITHIN A TRIAL EXPERIENCE DIFFERENT OUTCOMES THAN THEIR COUNTERPARTS MANAGED OUTSIDE THE TRIAL? A SYSTEMATIC REVIEW AND META-ANALYSIS OF RANDOMIZED AND OBSERVATIONAL STUDIES.

By

NATASHA A FERNANDES, DIANNE BRYANT, LAUREN GRIFFITH, MOHAMED EL- RABBANY[1], NISHA FERNANDES[1], CRYSTAL KEAN[1], JACQUELYN MARSH[1], SIDDHI MATHUR[1], REBECCA MOYER[1], CLARE READE[1], JOHN RIVA[1], LYNDSAY SOMERVILLE[1], NEERA BHATNAGAR

[1] These authors contributed equally to this work.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfilment of the Requirements for the Degree

Masters of Science

MASTER OF SCIENCE (2012) **McMaster University**

Health Research Methodology **Hamilton, Ontario**

Title: Do patients managed within a trial experience different outcomes than their counterparts managed outside the trial? A systematic review and meta- analysis of randomized and observational studies.

Author: Natasha A. Fernandes, BHSc

Supervisor: Dr. Dianne M Bryant

Number of Pages: viii, 198

Abstract

Context: It is unclear whether the construct of a randomised controlled trial (RCT) itself could confer benefit or harm to trial participants beyond any effect of the experimental treatment under study (trial effect).

Objective: To determine whether there is a trial effect appreciated by RCT participants (insiders) compared to similar patients who do not participate (outsiders). Although we are most interested in the pragmatic comparison of insiders to outsiders, we will also conduct the explanatory comparison of insiders to outsiders when the intervention is the same.

Data Sources: We searched electronic health research databases, including CENTRAL (1960-2010), MEDLINE (1966-2010), EMBASE(1980-2010) and PsycINFO (1880- 2010).

Study Selection: Eligible studies included those that reported the outcomes of insiders and a group of parallel or consecutive outsiders and reported the same health outcome at the same endpoint.

Results: We included 147 articles out of the 42493 identified in our initial search. Five out of the 147 studies randomized patients to be insiders or outsiders, the remaining were observational designs. The heterogeneity of our overall result was reduced by grouping studies based on whether the intervention being investigated was effective and whether treatment inside and outside of the RCT was the same or different. There was no significant difference in outcomes between insiders and outsiders when the experimental intervention was ineffective (standard mean difference [95% confidence interval]: -0.03 [-0.1, 0.04]), or when it was effective and received by both insiders and outsiders (0.04 [-0.04,0.13]). If the experimental intervention was effective but was not administered to outsiders, they experienced worse health outcomes (-0.36 [-0.61, -0.12]).

Conclusions: There is no evidence to support any benefit or harm associated with trial participation. There is some evidence that better outcomes are experienced by insiders who had access to effective treatments not offered or available to outsiders.

Acknowledgments

I would like to thank all those individuals who dedicated time and effort to the development of this project. Foremost, Dr. Dianne Bryant has provided continual support both through the course of my thesis and throughout my Masters program. It was her guidance and tireless editing that has led to the refined end product.

My thesis committee members have also shared their respective expertise in the development of this project. Dr. Lauren Griffith helped inform my statistical plan for the meta-analyses. Neera Bhatnagar created our search strategy and provided our initial search yield. Her guidance in navigating RefWorks was very much appreciated. Dr. David Sackett has provided invaluable guidance from the very beginning of the project. His comments on my protocol and final paper ensured that my purpose was always focused and my thoughts were well developed.

Completion of this project would not have been possible without the help of my reviewer team (Nisha Fernandes, Crystal Kean, Jacquelyn Marsh, Siddhi Mathur, Rebecca Moyer, Clare Reade, John Riva, Lyndsay Somerville). Your endless hours screening, reviewing and data extracting provided me with the data to analyse and I appreciate the care and dedication with which you worked.

Last but certainly not least, I would like to thank my immensely supportive family, and most specifically my parents, Orlando and Patricia Fernandes. Enough appreciation cannot be expressed for your support. Thank you for making it so easy for me to pursue my goals.

Financial support

This project was partly funded by a CIHR Frederick Banting and Charles Best Canadian Graduate Scholarship, and an Ontario Graduate Scholarship.

Table of Contents

Abstract ii

Acknowledgements iv

Financial support v

Declaration of Academic Achievement viii

1. Purpose 1

2. Hypothesis 1

3. Background 1

4. Implications of Results 7

5. Literature Review 7

6. Methods 13

6. 1 Cochrane Registration 13

6. 2 Selection Criteria 13

6. 3 Search Strategy 13

6. 4 Titles and Abstract Screening 14

6. 5 Full Text Screening 14

6. 6 Pilot Testing of Data Extraction Form 14

6. 7 Data Extraction 15

6. 8 Plan for Statistical Analysis 17

6. 9 Pre-specified Causes of Heterogeneity 18

6. 10 Additional Analysis 23

7. Results 23

7. 1 Summary of Evidence 23

7. 2 Risk of Bias 25

7. 3 Subgroups 25

7. 4 Additional Analysis 29

8. Discussion **29**

9. Limitations **33**

10. References **34**

List of Tables and Figures

Figure 1: PRSIMA Flow Chart **60**

Table 1: List of imputations/assumptions **61**

Table 2: List of Excludes Articles **62**

Table 3: List of Articles Requiring Further Information **77**

Included Study Characteristics **80**

Forest Plots

1.1 Subgroups based on treatment effectiveness **143**

1.2 Subgroups based on baseline characteristics **150**

1.3 Subgroups based on outcome **157**

1.4 Subgroups based on methodological features **165**

1.5 Subgroups based on detection bias **170**

1.6 Subgroups based on exclusion bias **175**

1.7 Subgroups based on type of care received **181**

1.8 Multiple events outcomes (i.e. relapses) **189**

1.9 Subgroups based on treatment effectiveness (Vist analysis) **190**

Funnel Plots

Figure 2: Funnel plot for continuous outcomes **196**

Figure 3: Funnel plot for mortality outcomes **197**

Figure 4: Funnel plot for dichotomous non-mortality outcomes **198**

Declaration of Academic Achievement

As the primary author on this project I was responsible for the majority of the work presented. I drafted the protocol which involved designing the study, refining the research question and the methods. I also attended to the administrative needs of the study. It involved assigning the titles and abstracts to each of the reviewers, creating the data extraction form, securing copies of each of the studies selected for full text review and managing the data extraction database. I organized and trained the reviewers to help standardize the review process. I also played a main role in the title and abstract screening, reviewing full texts for inclusion, and extracting data.

I conducted all of the statistical analyzes- the descriptive statistics, reviewer agreement and imputation calculations. I also created all the forest and funnel plots included in this paper.

Finally, I drafted the results and discussion sections for this manuscript.

Purpose:

To determine whether receiving care within a randomised controlled trial affects patient outcomes. We will compare the outcomes of patients who participate in randomized controlled trials (insiders) to the outcomes of similar patients who were managed outside a trial (outsiders). Patients outside the trial will be considered similar to those inside the trial if they were eligible for trial participation.

Hypothesis:

We expect that patients treated inside a clinical trial will experience better outcomes compared to similar patients managed outside a trial.

Background:

Randomized controlled trials (RCTs) are at the top of the hierarchy of research studies when trying to prove a causal relationship between an intervention and a specific outcome[1]. Due to the nature of its design, RCTs are able to circumvent many of the biases (i.e. confounding factors, selection bias, etc.) that are associated with non-randomized designs. What remains unclear is whether just participating in an RCT has an effect on patient outcomes. It may be true that within an RCT, patients receive superior treatment and care than that received outside the trial. It may also be true that the patient may be exposed to additional risks due to the unknown consequences of the new intervention.

Peppercorn et al. proposed four explanations to account for differences in outcomes of insiders when compared to outsiders[2]. First, are the influences of any types of biases (i.e.: selection bias). Second, is the presence of confounding factors, where a confounder is defined as a variable that is correlated to two variables independently and due to its effect on both variables, misleads researchers to believe that a causal relationship between those two variables exist. The

final two effects, and the ones that will be the focus of this study, are the treatment and trial effects. A treatment effect is expressed by the finding of a difference in outcome(s) between the interventions being tested between two or more groups. A trial effect is expressed by the finding of a difference in outcome(s) between two or more groups despite any effect of the interventions being tested. Specifically, the participation effect captures differences in care received by patients inside the trial. Braunholtz et al. added four further subdivisions to the participation effect: the method in which the treatment is delivered (protocol effect), the quality of care received by the patient (care effect), any alterations in how the patient or doctor behave because they are aware of being monitored (Hawthorne effect) and finally, the psychological impact of being part of a trial (placebo effect)[3]. Unlike Peppercorn et al, we argue that the trial effect should include the treatment effect when the interventions within the RCT are not available or offered to those outside of the RCT.

We have further expanded on Braunholtz's subcategories. Braunholtz discussed as part of the protocol effect, the strict standardized treatment and follow-up in RCTs that could benefit the patients[3]. Furthermore, we believe that the additional monitoring of compliance demanded in RCTs would also encourage better outcomes.

The 'care effect' may be present because the physicians receive additional training within the RCT[3]. The clinician running the trial is probably an expert in the field, or at least has a lot of experience within the area[3]. Co-interventions are also another factor to consider, since within trials patients may receive supplemental care not received outside the trial[3]. We should also recognize that those centers more likely to participate in a trial tend to have more resources to facilitate better outcomes for its patients. For example, larger academic institutions are more likely to participate in trials than smaller centers[4]. Participation in a trial at an institution that has a lot of experience with RCTs also appears to be a further indicator of patient success. One study

in particular showed that hospitals conducting more RCTs than their counterparts had lower mortality rates, even after controlling for hospital characteristics (i.e. size, setting) [4]. Whether or not the institution conducting the RCT is privately or publicly funded may further impact the quality of care received. For instance, Devereaux et al. have shown that for- profit institutions have higher mortality rates than not- for- profit centers even after risk adjustment[5].

Only the positive nature of the psychological impact of RCT participation is considered by Braunholtz et al[3]. Indeed, believing you are receiving a treatment even if there is no active agent being administered, has shown to have a positive effect on patients. However, we would also like to consider the potentially negative psychological impact of trial enrolment. The informed consent process may serve to create further anxiety for patients. During informed consent, the degree of uncertainty is amplified since what is unknown is clearly stated to the patient. For patients facing life or death decisions, it can be unsettling when the physicians themselves are uncertain[6]. Similarly, current research is investigating the presence of a 'selection' or 'choice' effect[7]. It has been observed that when patients are allowed to make an informed choice about their treatment they experience better outcomes than their randomized counterparts[7]. Possible explanations for this phenomenon are that when patients are allowed to choose, they are taking ownership of their illness and will be more compliant and less frustrated with any side effects[7]. While there is much debate on whether such an effect exists [8-10], if it is present it would certainly be a part of the placebo effect.

In fact, policy makers in the field of oncology have assumed that one or more of these benefits will be appreciated by patients who participate in an RCT. Thus, the Federation of Clinical Oncologic Societies consensus statement states that to receive the best treatment a patient should be enrolled in a clinical trial[11]. Although they do not provide a rationale for their guideline, we could speculate that inside a trial patients have access to a theoretically more

promising therapy. There is also the chance that patients could receive better or more comprehensive ancillary care. Most trials are led by major health care institutions thus providing patients access to the best equipment and staff. Furthermore, the primary investigator overseeing their care is likely an expert in the field. The opportunity to receive such high level care may not present itself outside of an RCT. For these reasons, oncology clinical trials which used to be the last approach for patients who failed all other treatment options, are now being recommended to patients even if all other options are not yet exhausted[13].

On the contrary, some clinicians are concerned that RCTs may expose patients to additional risks not faced by patients receiving care outside of a randomized trial[14, 22-26]. In fact, in a study published in 1994, by Joseph, found that only 3% of new cancer patients were enrolled into RCTs each year.[15] In a review of thirty five articles from North America, Australia, Sweden and the UK, Castel et al. identified six articles where clinicians admitted to being especially reluctant to enrol their elderly patients into trials[16-21], and four of these articles further stated that patients with a worse prognosis were less likely to be enrolled even if these patients were eligible for the study[17, 18, 20, 21]. Some oncologists felt that their patients would be exposed to greater levels of toxicity from the treatment given in the trial when compared to levels expected outside the trial[27].

Ford et al.[28] conducted a systematic review and identified that patients who were consistently underrepresented in oncology research (i.e.: rural, ethnic minorities, low income) had specific negative views about clinical trials that made them reluctant to participate. The articles included in their review covered a diverse spectrum of trials (prevention as well as therapeutic) and settings (community and hospitals). The theme found in the majority of studies (20 studies) was that patients were reluctant to participate because they did not trust research studies and that they feared harm from study participation (18 studies). Some patients also felt

that the randomization process trivialized their condition; feeling that their life or death situation was being decided by a game of chance rather than through careful deliberation and thought[24].

The question we are investigating is pragmatic in nature. There is a continuum that follows RCTs from explanatory to pragmatic. Explanatory trials try to isolate the treatment effect by observing it under the best circumstances, whereas pragmatic trials investigate the treatment effect under usual care.

Explanatory RCTs differ from their more pragmatic counterparts in eight major areas. First, explanatory trials have more restrictive eligibility criteria, deliberately selecting for patients more likely to comply, at a higher risk of having the event, more responsive to treatment or a population that can be conveniently sampled. Pragmatic trials include all patients who would be treated in usual practice. Second, the administration of the intervention can be pragmatic or explanatory. An explanatory design includes standardized protocols for administration of the intervention and management of adverse events, and co-interventions are minimized. More pragmatic trials will allow treatments to be administered as they would in regular care with no restrictions on co-interventions. Third, the type of clinician providing care in explanatory trials tends to also be more highly qualified and may have had to provide evidence of their expertise. Conversely, pragmatic trials include all clinicians who would normally provide the treatment. Fourth, explanatory trials also follow a more structured follow-up schedule that usually includes more visits than normally occur in regular practice. Fifth, explanatory trials look for outcomes that are easily identified within a short period of time (usually surrogate), while pragmatic designs usually follow patients for a longer period of time to capture patient important outcomes that may take longer to develop. Sixth, patient compliance is more strictly monitored and regulated in explanatory trials whereas pragmatic trials do not monitor patient compliance and make no effort to change or improve it outside of efforts that would normally take place in

regular practice. Seventh, explanatory trials go to various lengths to ensure that practitioners adhere to the study protocol which is not practiced in more pragmatic designs. Finally, explanatory trials are more selective when including participants in the analysis. For example, patients who were noncompliant or lost-to-follow-up may be excluded. Pragmatic trials however follow a more conservative intention- to-treat analysis, including all patients in the analysis who were randomized regardless of their behaviour thereafter.

Thus, when we claim that our systematic review adopted a more pragmatic approach, we mean that we included all types of studies and did not restrict our eligibility criteria to certain clinical areas or insist that protocols (interventions, expertise of the clinician) were similar between insiders and outsiders. Since trials share similar design features in all areas of medicine, the benefit or harm they confer should be relatively similar regardless of the clinical area. Our intention is that the results of this review can be applied across medicine. Further, an explanatory approach to this systematic review might have insisted that the intervention between the insiders and outsiders were the same (making comparisons only between insiders and outsiders who received an identical intervention or control). However, in our pragmatic approach we combined the results of outsiders (treatment with control patients) and the results of insiders (treatment with control) and compared insiders to outsiders even if the outsiders received different interventions that the insiders. Finally, our main analysis did not differentiate by whether insiders and outsiders received treatment by a clinician with similar expertise, follow-up schedules or reasons for excluding patients (although these were addressed in our attempts to explain the between-study heterogeneity).

Implications of Results:

The results from our study could have many implications. Foremost, if patients within RCTs experience either better or worse outcomes, then the ethics of conducting RCTs must be revised. If the very construct of the RCT confers better outcomes it implies that future studies should either be an RCT or contain the elements usually found within RCTs that confer the superior results[29]. A positive trial effect would also imply that perhaps some of the features of RCTs, that infer superior outcomes, should be introduced into regular practice including strict monitoring of compliance to protocols, additional clinician training to a certain level of expertise, and closer follow-up. If there is evidence of a positive participation effect, then we would expect an increase in the recruitment rates into RCTs. Clinicians may feel less apprehensive when enrolling patients into RCTs and their patients may feel better about accepting the unknown risks associated with the new intervention or even the uncertainty within the clinical community around treating their disease. Increased enrolment would lead to larger studies that are better suited to make reliable conclusions.

Conversely, if RCTs are shown to expose enrolled patients to worse outcomes than those outside an RCT, then the process of informed consent will need to be revised. Patients would need to be informed at the onset not only about the benefits and risks associated with the intervention, but also about the benefits or risks associated with participation in a study.

Literature Review:

We were able to find seven systematic reviews that investigated differences in outcomes between RCT insiders and outsiders. Furthermore, all except one could not provide an overall effect because they did not conduct a meta-analysis. Interestingly, the results of the other

systematic reviews lack consistency. One review concluded that there was no evidence, free of bias, to indicate any benefit or harm from RCT participation[2]. Four other reviews concluded that being enrolled in an RCT had better outcomes [3,30,31,32] and one other stated that there were no negative effects from being enrolled[33]. Only one review concluded that there was no trial effect associated with participating in an RCT[34]. Most of these reviews were underpowered due to inclusion of a few studies, that even if an effect was present it was difficult to find a statistically significant difference. We intend to include more studies in our analysis by contacting authors and asking for unpublished data, instead of excluding the study altogether. This was a step that the other reviews missed, with the exception of Vist et al., and which negatively impacted the size of their reviews and ability to draw conclusions that are applicable to every practice.

The largest and most current review was done by Vist et al[34]. Their systematic review concluded that there was no trial effect from participating in an RCT when both patients inside and outside the trial received the same treatment. They reviewed studies published on or before March 2007 and included five RCTs and 80 cohort studies, which created a total of 136 comparisons. Vist et al. divided their comparisons based on outcomes measured. In the group of studies investigating dichotomous outcomes, there were 85 comparisons that were not statistically significantly different, eight comparisons found a significant benefit to being enrolled in an RCT and in five comparisons it was found to be harmful. In 38 comparisons where continuous outcomes were considered, 30 comparisons were not statistically significant and in the comparisons which were significant, three described benefits from being enrolled in an RCT and five found a negative outcome associated from enrolment. Although a summary estimate was not provided for any of these comparisons, a confidence interval (CI) around the summary estimate was stated. For the dichotomous outcomes the 95% CI ranged from 0.93 to 1.06, and for

the continuous outcomes the CI ranged from -0.05 to 0.11. Both indicate no significant difference between trial insiders and outsiders.

Due to the nature of the research question posed in Vist et al. the number of studies included was limited. The authors excluded studies which compared patients outside the RCT group who were given either a different intervention or control treatment than the patients inside the trial. After full text review, 140 articles were excluded. Of the 140 articles, 34 were excluded because treatment was different across study arms. Vist et al. excluded studies where treatment was not the same between the two groups because they thought it would control for the treatment effect and be able to better capture the trial effect. For our study however, we plan to include those excluded studies in order to more accurately capture any differences between the two groups. As was stated in the Background section of this proposal, some critics of RCTs argue that insiders are being exposed to potentially more harmful interventions than outsiders. By including all comparisons, even those where different treatments are being used in the RCT and non-RCT trials, we will be able to identify if this is really the case. Furthermore, Vist et al. were unable to explain the heterogeneity they observed in their meta-analysis, despite performing a sub group analysis. In our review we will be comparing more studies which may reduce the issues associated with performing a meta-analysis when there are a lot of clinical differences but fewer trials in each group.

The second largest review was done by Stiller et al. and included 53 studies published on or before 1994 which compared the outcomes of being in an RCT for patients with various forms of cancer[30]. They performed a comprehensive search of the oncology literature and compared differences in outcomes for stomach, colorectal, lung, cervical, ovarian, prostate, testicular, Hodgkin's disease, miscellaneous site and childhood cancers. They concluded that being treated

inside a clinical trial had beneficial effects because the RCT trials had lower mortality rates. However, they did not conduct a meta-analysis so this conclusion was based on a qualitative synthesis of the literature. No effect size was reported, nor could be calculated. No inclusion criteria were explicitly stated either; however, the research question suggests that they included any study that compared two groups where one group received special treatment either by being enrolled in an RCT or by being referred to a specialist centre. This question is much broader than our question. Only six of the 53 studies, compared patients treated inside an RCT with those outside. A further limitation of this study was that it was completed over 15 years ago and in the interim there would have been more studies published or change in guidelines, limiting the extent to which the results can be generalized to current practice.

The third largest systematic review by Gross et al. included 25 articles[33]. Although the authors did not provide a time period that the search covered, they did state that the latest publication included was from 2002. They had very specific inclusion criteria. The outsiders had to have been eligible for the trial, all patients had to have received concurrent care from the same facility, the outsiders had to have been given the choice of receiving the trial intervention, and all patients had to be recruited in the same way. They found that 21 out of the 25 articles showed no statistically significant difference in outcomes between those in the RCT and those outside. Their results were separated according to baseline characteristics between the two groups. In 15 out of the 17 studies where both groups had similar baseline characteristics there was no significant difference in clinical outcomes found. In the six out of seven studies where the patients enrolled in the RCT had poorer health than those outside the RCT, there was no significant difference in outcomes. Lastly, in the one study where patients enrolled in the RCT had better health than those outside the RCT, there was no significant difference between the two groups. Therefore, based on a qualitative synthesis of the data, they concluded that being treated inside an RCT has

no negative effect on patients. The limitation of this study is similar to Vist et al. because they excluded studies where the treatment received by the two groups was not similar. Furthermore, their inclusion criteria were overly restrictive such that the trial effect may have been controlled out.

The next largest systematic review was done by Peppercorn et al. and included 21 published studies and 26 total comparisons from studies up until 2002[2]. They included only cancer studies where the outsiders still met the eligibility criteria for the RCT. Of the 23 unadjusted comparisons they reported that 15 had a better outcome in the RCT and seven showed no difference between the two groups. They do not mention the outcome of the final unadjusted comparison. For the 17 adjusted comparisons, 12 had a better outcome in the RCT and five showed no difference between the two groups. They did not conduct a meta-analysis, nor state whether these differences were statistically significant. They concluded that there was not enough evidence to support that insiders had better outcomes than outsiders.

Braunholtz et al. authored the sixth review we found which included 14 articles that examined 21 trials up to 1996[3]. They included studies that compared insiders to outsiders. This was further restricted to only studies where outsiders were offered entry into the trial but refused, or patients who had the same medical condition as insiders but were not offered entry. The authors do not indicate whether the type of treatment the control group received factored into their inclusion decision. The study concluded that insiders had better outcomes than the outsiders. Braunholtz decided against pooling the results in a meta-analysis and instead used a qualitative synthesis of the articles to arrive at this conclusion. In eight articles the patients in the trial had statistically significantly better outcomes than those not included in the trial. In the remaining six articles the conclusions were not statistically significant. Three of these articles found a benefit associated with RCTs, one article found an improvement for both insiders and

outsiders, and in the final two articles there was no significant difference. A limitation on the extent that these results can be generalized to other practices is that the articles found were primarily from the cancer literature. We intend to include all articles across all types of illnesses in order to draw conclusions which can be applied across disease types.

The review on rheumatoid arthritis treatments searched for articles published before the end of 2005 and included only 11 articles[31]. They limited the studies they included to only those that compared rheumatoid arthritis medication (etanercept, infliximab, or adalimumab) to a placebo group. Only trials with comparable levels of dosage for each drug were compared. They found that the effect size favoured the RCT in comparison to what was observed in the clinical practice cohort. This was concluded from a qualitative synthesis instead of a meta-analysis. In five of the comparisons there was a difference between insiders and outsiders which was found to be statistically significant. The review only looked at studies that evaluated the efficacy of tumor necrosis factor alpha in treating rheumatoid arthritis. Such a narrowly focused research question limits the extent to which the results can be generalized to other fields.

The Emergency Care Research Institute performed the smallest review which included only 10 comparisons from nine articles[32]. The authors did not mention a search period; however, all of the articles found were published on or before 1999. They clearly stated their inclusion criteria. The studies were limited to only those they deemed had valid conclusions and were relevant to the patient population. The illness being treated had to be life- threatening, the patient population needed to be at least 18 years old, all patients needed to be eligible for the RCT trial and the study had to report patients' opinion on why they either chose or refused to participate. Of the nine articles, eight were investigating cancer, and the remaining two articles studied heart related conditions. They focused on mortality and patient- reported outcomes and excluded any outcomes that were surrogate since they were considered less important to patients. A qualitative

synthesis concluded that in eight articles a statistically significant difference was found, however, results in only five of the trials were designed such that the trial effect could be separated from selection bias. Of the five trials, four were found to have statistically significant better outcomes for patients in the RCTs compared to those outside the RCT. The authors themselves noted that due to the limited studies examined, the results may not be precise.

Methods:

Cochrane Registration

We registered our protocol as a Cochrane Review under the Cochrane Methodology Review Group.

Selection Criteria

Eligible studies included those that reported on the outcomes of patients who were treated inside an RCT (parallel group design only) and similar patients who were treated outside (any form of observational study). To be eligible, the study must have reported on outsiders who were followed in parallel or sequentially to the RCT insiders. If sequential, the follow-up could not be longer than two months after the RCT to minimize differences explained by changes in practice. Finally, to be eligible, the study had to report the same outcome at a similar time point for patients inside and outside the RCT.

Search Strategy

Neera Bhatnagar (Medical Librarian), at McMaster University, designed a highly sensitive search strategy to captured English articles relevant to our research question. The search included MEDLINE (1966 to November 2010), EMBASE (1980 to November 2010), Cochrane Central Register of Controlled Trials (CENTRAL; 1960 until the last quarter of 2010) and PsycINFO

(1880 to November 2010). After removing the duplicate articles, we compared our yield to the articles reviewed by Vist et al. that asked a similar research question; this step helped to validate our search.

Titles and Abstracts Screening

Using our previously developed eligibility criteria, pairs of reviewers independently screened titles and abstracts from the electronic search yield. There were eleven reviewers in total (Dr. Dianne Bryant, Mohamed El-Rabbany, Natasha Fernandes, Dr. Nisha Fernandes, Jacqueline Marsh, Dr. Clare Reade, John Riva, Lyndsay Somerville, Dr. Crystal Kean, Siddhi Mathur, Rebecca Moyer). After screening, the reviewers marked the study as either being included (i.e. eligible and uncertain) or excluded from full text review. The full text of any article marked for inclusion was reviewed, regardless of whether there was consensus among reviewing partners.

Full text Screening

A full text review was conducted of those articles identified in the earlier screening to confirm that these articles met our eligibility criteria. Each article was independently screened by two reviewers. If required, an independent third adjudicator (a reviewer from a different partnership) reviewed the paper and made a conclusion about its eligibility to resolve any disagreements. If after reading the full text, the reviewers agreed that the article was not eligible, they noted the specific reason. We calculated a weighted Kappa coefficient upon completion of this phase.

Pilot Testing of Data Extraction Form

We pilot tested our data extraction form using ten articles that were eligible after the full text review. All reviewing pairs independently completed the extraction form using the same articles

and their answers were compared. Based on feedback from reviewers, we revised those areas on the form with low inter-rater reliability or provided operational definitions of ambiguous terms.

Data Extraction

We provided each reviewer with a username and password to access our web-based data collection software. Upon accessing the system, each reviewer had access to a PDF of their articles for full-text review and its corresponding data extraction form. The reviewers did not have access to data input by any other reviewer. The administrator checked for disagreements and arranged for the adjudication of disagreements if the two reviewers could not form a consensus.

The data extraction form covered three major areas: fulfillment of eligibility criteria, methodological features, and outcomes. Methodological features included type of design, participant characteristics, interventions used in each study arm and the type of clinician that provided patient care. Reviewers were also asked to identify features of the study where the potential for bias or confounding was present and about the types of outcomes. Data extraction included recording the results for each group of insiders and outsiders.

We identified and recorded the presence of three main biases: detection bias, exclusion bias and selection bias. To evaluate the presence or absence of detection bias we focused on the frequency of visits during the follow-up period to determine whether those inside the RCT were evaluated to the same extent (frequency and type of test) as those outside.

To evaluate the presence or absence of exclusion bias we noted the number of patients in each arm that were excluded after inclusion into the study. We made the distinction between

those exclusions made appropriately (i.e. never eligible) and those inappropriately excluded (i.e. lost to follow-up, non-compliant, etc).

Finally, to evaluate the presence or absence of selection bias we reviewed the study's eligibility criteria and the description of and/or table that described patient characteristics. To be eligible for inclusion into our review, outsiders had to be eligible for the RCT. However, more flexibility was afforded to imbalances in patient characteristics. We classified differences between insiders and outsiders as either being balanced, having imbalances statistically controlled for in the analysis, having imbalances not controlled for at any stage, or that the study authors neglected to inform us whether there were any imbalances.

We also assessed the quality of the included studies. We noted the absence of specific methodological qualities (allocation concealment, blinding, and analysis of losses to follow-up) rather than give the study an overall quality rating. We contacted study authors for additional information or data not reported in their articles.

Our primary outcome was mortality. Secondary outcomes included patient reported or other clinically important outcomes. For dichotomous outcomes like mortality, we extracted the number of individuals and events per group and reported an odds ratio. In cases where both an odds ratio and an adjusted odds ratio were reported, we included the adjusted odds ratio[35]. In cases where an odds ratio was not reported and could not be calculated from the available data, we assumed that the relative risk approximated the odds ratio for low event-rate outcomes. For our dichotomous outcomes, where overall estimates of effect size and standard errors were provided (instead of events for each group), we used the generic inverse variance meta-analysis to input the data directly into Review Manager 5.1. We added a 0.5 correction to all cells to permit us to include the effect of studies where there was a zero event rate in one group. .

For continuous outcomes like self-reported functional ability, we reported the mean between-group difference and its standard deviation. We created rules for common scenarios where either the mean difference or the standard deviation was not provided but could be calculated using other statistical measures detailed in the study (Table 1). We imputed missing standard deviations from given standard errors or confidence intervals around the mean. We assumed a standard normal distribution was used to calculate the standard deviation for sample sizes greater than 100 in each group. For smaller sample sizes we based our calculations on a t-distribution. We could calculate the standard error of the difference in the means using the t- value corresponding to the p-value provided when no measure of dispersion was provided, to obtain a standard deviation. We also converted all change scores to final scores by using the baseline and change means and standard deviations, if provided. For those studies where the primary outcome only presented a range to describe their dispersion, we chose the secondary outcome instead. This was done because there is no robust way to convert range to standard deviation.

Plan for Statistical Analysis

All statistical calculations were performed on IBM SPSS Statistics Version 20. All of the forest plots and funnel plots were created on Review Manager (RevMan) Version 5.1. We used the more conservative random effects model when calculating the summary effect. We combined the means and standard deviations of the groups inside the RCT and performed the same calculation to combine the outcomes of groups outside the RCT, before entering the effect into RevMan. The standardized mean difference was used to combine the continuous outcomes. The associate standard error from the natural logarithm of the unadjusted relative risk was calculated to standardize non-mortality dichotomous outcomes across studies.

We initially separated the studies into two groups based on whether the overall study was a randomized trial (reported the results of patients who were randomized to be inside or outside the RCT) or observational design (reported the results of a cohort of insiders and outsiders). Next, we separated studies by their type of outcome; continuous, dichotomous non-mortality, and mortality outcomes.

Pre-specified Causes of Heterogeneity

Due to the diverse nature of the studies included in our systematic review we anticipated a high degree of variance in our meta-analysis. We used the I^2 statistic to measure the extent of inconsistency between studies. An I^2 of 25% indicated low, 50% indicated moderate and 75% indicated a high degree of heterogeneity.[36] We constructed several hypotheses to try to explain between-study heterogeneity should it be detected.

Types of Outcomes

We felt that if there was heterogeneity between studies that it might be explained by the type of outcome measure. Specifically, we divided outcomes into clinically important outcomes (subdivided into patient-reported outcomes and more objective clinical measures) versus surrogate outcomes. The patient reported outcomes group were further divided by the construct they measured including pain, quality of life, satisfaction and functional outcomes. Not all the studies presented their outcomes in the same direction. For example, one article may have presented the number of pregnancies, a 'good' outcome, while another article's primary outcome was the number of miscarriages, a 'bad' outcome. We accounted for this by converting all positive events into negative ones. Similarly, continuous outcomes were entered into RevMan

taking into account the direction of the scale (i.e. whether a higher score indicated improvement or decline). This was done by adding a negative sign to all values. The adjusted and unadjusted odds ratios were presented separately for mortality outcomes as well as all-cause and disease specific mortality.

Study Quality

We also felt that whether the patients treated outside the trial were followed in parallel with patients inside the RCT or whether they were treated and followed shortly before or after the patients inside the RCT could also potentially explain between-study heterogeneity. Finally, we felt that analyses that followed the intention-to-treat principle may report findings different from studies where intention-to-treat was not practiced.

Type of Care Provided

We felt that heterogeneity might be explained by whether or not the type of health care worker providing care for patients inside the RCT was different (in terms of expertise) than those outside the RCT. Thus, we proposed four subgroups; studies where the same individual with the same expertise provided care inside and outside the RCT, the studies where a different individual with the same expertise provided care for patients outside the RCT, studies where a different individual with less expertise provided care for patients treated outside the RCT and studies where it was unclear who provided care.

We hypothesized that heterogeneity within this unclear group would be further explained by separating studies with surgical interventions from those providing medical interventions, radiology interventions, counselling therapy and other areas of medicine. Since expertise bias is found predominantly within non-pharmaceutical specialties, we felt that it was important to separate them into their own subgroup. The learning curve may explain any differences between

the outcomes of patients inside and outside the RCT if physicians with less experience were systematically assigned to one particular arm of the study. For those groups where heterogeneity remained high we planned to create further subgroups based on similar or different care settings.

Detection Bias

The frequency of follow-up is another factor that may explain differences between studies. Due to the anticipated rigid protocols in RCTs compared to outside, we predicted that detection bias would be higher in patients treated inside the RCT than those treated outside the RCT. The RCT study should have greater resources to schedule a greater number of follow-up appointments and are often equipped with better tools with which to measure outcome or detect adverse events early. This may result in a potential bias that could go in either direction depending on whether the outcome being evaluated was positive (i.e. lower blood pressure) or negative (i.e. fetal distress). We separated studies based on whether follow-up was exactly identical, if patients inside the RCT had a different follow-up schedule than those outside the RCT, or if insufficient details were given on the follow-up process.

Exclusion Bias

We also felt that studies at greater risk of exclusion bias may have different outcomes than studies at lower risk. Studies at least risk formed the first group. This included studies that had no exclusions as well as those with deliberate but appropriate exclusions inside and outside the RCT. This group was compared to the studies that had any inappropriate exclusions in at least one group and to studies where it was unclear whether any exclusion had occurred. If heterogeneity was still high in the first group, we planned to further subdivide the groups into no exclusions and appropriate exclusions. We also thought to explain high heterogeneity within studies with inappropriate exclusions by separating those studies where the exclusions were

equal inside and outside the RCT from those that had an unequal number of exclusions inside and outside the RCT. Creation of this subgroup is based on the assumption that those studies where the proportion of inappropriate missing cases were unequal between arms also had a differential number of patients missing not at random.

Balance of Baseline Characteristics

Generally we expect the poorly balanced studies to demonstrate a treatment effect in favour of patients treated inside the RCT when compared to studies with stronger safeguards against selection bias[38]. This prediction is based on the fact that RCTs tend to enrol a healthier subset of a disease population[38]. Studies that we felt were at lesser risk of selection bias were those that demonstrated balance for known prognostic factors by chance or presented a statistically adjusted analysis. These two groups were combined and compared to studies where there was clearly an uncontrolled imbalance (as noted by the study authors) and to studies that were unclear about any imbalances. In the event that heterogeneity remained high within the group of studies where balance could not be ascertained, we planned to create a further subgroup that pooled studies with a sample size greater than or equal to 200 (RCT and observational groups combined) to those whose sample size was less than 200. We hypothesized that studies with a larger sample size were more likely balanced and therefore less vulnerable to selection bias than the smaller studies.

Treatment Provided

We hypothesized that the treatment provided in the trial and cohort groups could also explain heterogeneity. We proposed six subgroups. The first consisted of studies where one of the interventions within the RCT was effective (was statistically significantly different than the

comparator) and those outside the RCT received the same interventions. The second involved those studies where one of the interventions inside the RCT was effective and those treated outside the RCT received the same effective intervention only. The third, included those studies where one of the interventions inside the RCT was effective and those outside the RCT received the less effective or control intervention only. The fourth, included studies where one of the interventions inside the RCT was effective and patients outside the RCT received different interventions. The fifth consisted of studies where the RCT demonstrated no superior outcome between treatment groups. We did not sub-divide this final group because if there was no treatment effect inside the RCT then any differences between the outcomes of those treated inside and outside the RCT could be attributed to a trial effect. The final subgroup consisted of those studies where there was insufficient information provided about the effectiveness of the treatment in the trial, and/or insufficient details about the interventions received outside the trial.

We also analysed our data according to the analysis plan described in the published Vist et al review (see Forest plots 1.9).[34] Since their research question was explanatory in nature, they included comparisons only if insiders and outsiders were given identical treatments. To replicate their analysis, we isolated all studies that included a matching treatment/comparator/control for outsiders and insiders. Each of these comparisons was then separately entered into RevMan, annotated with an 'a' or 'b' or 'c' to indicate a separate treatment arm. For example, a study that compared a treatment and control in the trial and the same treatment and control outside the trial, would be analyzed by comparing the inside treatment to the outside treatment ('a') and the inside control to the outside control ('b'). In studies comparing three treatments, the additional arm was denoted with a 'c'.

Additional Analysis

We created a funnel plot (affect size versus standard error) to look for evidence of publication bias. Lastly, we ran a sensitivity analysis to determine the stability of our conclusions by removing studies that required us to make assumptions and impute data.

Results:

Summary of Evidence

Figure 1 summarizes the flow of studies through our screening process. We validated our search strategy by comparing our list of included studies to that of Vist et al[34]. There was only one study (Abraham 2004) that was not found by our search. Our initial search yielded 42493 articles. After removing the duplicate articles, 21045 articles remained and only 797 articles remained after reviewing the titles and abstracts. Following the full text review we could confirm that 147 articles met our eligibility criteria and provided sufficient information to be included in our analysis. **Table 2** lists the 554 excluded articles and the reason for their exclusion. The remaining 96 articles (**Table 3**) either met our eligibility criteria but did not provide sufficient information to be included in the analysis, or there was insufficient information to determine whether the study was eligible. In both cases the authors were contacted and asked to provide additional information but had not responded to our request at the time this manuscript was prepared. For a detailed description of each included study see **Included study characteristics.**

The calculated average of the weighted kappa was 0.68, which reflects a good level of agreement between reviewers at the full text screening stage. There was an 83% raw agreement between reviewers in the data extraction phase.

In 5 out of the 147 eligible studies, patients were randomly assigned to either treatment inside a trial or treatment outside the trial. In the remaining 142 studies there were a variety of reasons patients were not treated inside the trial; the most common reason being that either the patient or physician refused participation in the trial, the patients had a strong preference for a particular treatment or the study authors gave patients the option of choosing between being treated inside or outside the trial. There were a few studies that had unique reasons for treating patients outside the trial. In two studies non- trial patients lived too far away from the study site[54,55],. In a study by MacLennan al and Verdonck et al, patients were excluded primarily due to an administrative error (i.e. trial co-ordinators were not present that day)[113, 165] . Vind et al reported the outcomes of patients who refused to participate in the trial either because the patient felt they were too sick or too well to participate in the trial, had responsibilities to care for a partner, could not commit the time, or did not want to visit the hospital[166]. An outsider group was also created in the study by Woodhouse et al because the authors felt uncomfortable randomizing patients to a control group[179]. Patients were also treated outside the trial if the physicians wanted to test for any change in clinical practice during the period of the trial or if the physicians wanted an additional group to gain further training in the procedure[133]. Finally, West et al created an outsider group specifically to test whether a "trial effect" was present[171].

A diverse array of specialties were included in this review; radiology (n=1), mental health and addiction (n= 19), obstetrics/gynecology (n=25), anesthesiology (n=9), pediatrics (n=14), acupuncture (n=6), cardiology (n=19), oncology (n=19), weight loss and nutrition (n=4), surgery (n=5) and other subspecialties (n=19).

In total there were 49 continuous outcomes, 94 dichotomous outcomes of which 73 were non-mortality outcomes, 4 were recurring outcomes (such as relapse rates). There were 21 studies that reported mortality as an outcome.

Risk of bias

In terms of detection bias, the majority of the studies (n=100) had identical follow-up patterns between patients inside and outside the RCT. Only 23 studies had different follow-up and for 24 studies the details on follow-up were not provided. In terms of exclusion bias within the RCT, there were 67 studies that had no exclusions, only one study that had a deliberate but appropriate exclusion, 74 studies inappropriately excluded patients unequally between the two groups, and in 5 studies it was unclear whether there was any exclusion after randomization.

Subgroups

Our initial pooled analysis revealed a high degree of between-study heterogeneity and thus, we continued our analyses as per our a priori hypotheses. For our non-randomized mortality and dichotomous non-mortality outcomes the high degree of heterogeneity was not explained by any of our a priori hypotheses. We present the results of our non-randomized continuous outcomes and randomized comparisons according to the subgroups with the least amount of remaining heterogeneity.

Non- randomized comparisons

Dichotomous outcomes

Mortality outcomes

Heterogeneity was unacceptably high despite the creation of subgroups ($p<0.00001$, $I^2 = 83\%$) so the results were not pooled (Figure 2). There were 53714 patients inside the RCT and 25817 treated outside the RCT.

Non- mortality outcomes

None of the subgroups explained the heterogeneity found in this meta- analysis ($p<0.00001$, $I^2= 70\%$) (Figure 3). There were a total of 30253 patients treated inside and RCT and 30000 patients treated outside the RCT.

Continuous outcomes

We were able to pool all studies where the trial treatment was significantly better than the control and the exact same treatment and comparator were given to patients outside of the trial (Figure 4). This subgroup included seven studies that had 2905 patients in the trial and 6014 patients outside the trial. The heterogeneity was low to moderate ($p= 0.15$, $I^2= 37\%$), and the pooled result indicated no significant difference between the outcomes of patients inside and outside of the RCT (standardized mean difference [95% confidence interval]: 0.04 [-0.04,0.13]).

In three studies there was a positive treatment effect within the RCT and the patients treated outside the RCT were provided the same effective treatment. There were a total of 1205 patients in the RCT and 5258 treated outside the RCT. There was a high degree of heterogeneity among these studies ($p< 0.00001$, $I^2=95\%$). The same was true for four studies where the RCT demonstrated a positive treatment effect but those outside the RCT were only offered the same control intervention ($p=0.01$,$I^2= 74\%$). There were 5794 participants in the RCTs and 9035 patients outside of the RCTs. We did not pool either of these groups of studies.

Results could be pooled for the nine studies where there was a positive treatment effect inside the RCT but patients outside of the RCT received completely different treatments ($p=0.08$, $I^2= 43\%$). There were 604 patients inside of the RCT and 233 patients outside of the RCT. Within this subgroup the RCT patients had significantly better outcomes (-0.36 [-0.61, -0.12]).

The next subgroup was also the largest, consisting of 23 studies. In all of these studies there was no significant treatment effect inside the RCT. Patients treated outside the RCT were

either provided the same interventions, the same control only, the same treatment only, or completely different interventions. In total there were 4837 patients treated inside the RCT and 13030 patients treated outside the RCT. The heterogeneity among studies was low to moderate ($p=0.10$, $I^2= 29\%$). The pooled result showed that there was no difference in outcomes between those treated inside the trial and those treated outside the trial (-0.03 [-0.1, 0.04]).

The final subgroup consisted of only two studies. For both studies it was unclear whether there was a treatment effect or which interventions patients received outside of the RCT. We are awaiting clarification from the authors before including them in the analyses.

Randomized comparisons

Dichotomous outcomes

There was a moderate degree of heterogeneity between the four studies with dichotomous non-mortality outcomes ($p=0.06$, $I^2 = 60\%$) (Figure 6). One of the studies had an effective intervention that was also given to outsiders. One study had an ineffective intervention and in the remaining two studies it was not stated whether the intervention was effective. The overall pooled effect indicated no differences in outcomes when patients were treated inside a trial versus outside (relative risk [95% confidence interval]; 0.94[0.56, 1.57]).

Continuous outcomes

Only one of the studies where patients were randomized to be invited to participate in an RCT included a continuous outcome (Figure 5). There were 180 patients randomized to participate in an RCT and 97 patients randomized to the outside (i.e. patient or clinician preference) intervention arms.

Non- randomized comparisons according to matched treatments

Mortality

The high degree of heterogeneity ($p<0.00001$, $I^2 = 84\%$) did not allow for the results to be pooled (Figure 37).

Non- mortality

Within dichotomous non-mortality outcomes, the only analysis able to explain the heterogeneity between studies was when the trial treatment was ineffective ($p=0.18$, $I^2 = 19\%$) (Figure 38). In this circumstance, there was no statistically significant difference between insiders and outsiders (RR=0.95 [0.88, 1.04]). The overall effect could not be pooled due to inconsistency.

Continuous

Within the analysis of continuous outcomes, the only analysis able to explain the heterogeneity between studies was when the trial treatment was ineffective ($p=0.09$, $I^2 = 28\%$) (Figure 39). There was no statistically significant difference between the outcomes of insiders and outsiders (0.07 [-0.13, 0.27]).

Randomized comparisons

Continuous

We found only one randomized study reporting continuous outcomes. There was no statistically significant difference in outcomes between insiders and outsiders (-0.03 [-0.35, 0.28]).

Dichotomous

Studies reporting a dichotomous outcome had sufficiently low heterogeneity to warrant pooling the results ($p=0.07$, $I^2 = 53\%$). There was no statistically significant difference between insiders and outsiders (RR=0.92 [0.74, 1.15]).

Additional analysis

Our investigation into publication bias (see **Funnel Plots**) revealed that the studies with larger effect sizes (positive and negative) that had a smaller sample size and therefore with a greater standard error were missing. Because the included studies were symmetrical around the pooled estimate we were confident that our estimates were valid.

Our sensitivity analysis confirmed the robust nature of our imputations. Removing the studies that had imputed outcomes had no significant effect on our results.

Discussion

The purpose of this systematic review was to determine whether patients who are treated inside an RCT experience better outcomes than similar patients treated outside of an RCT (trial effect). We did not find evidence to support a trial effect (neither beneficial nor harmful) if the interventions were similar between patients inside and outside of an RCT or if the effect of the treatment being investigated within the RCT was similar to the control. We did find evidence of a treatment effect if the treatment being offered inside the RCT was different than what was available or offered to patients outside the RCT.

We found that there were two specific circumstances when patients managed outside of the RCT experienced similar outcomes as similar patients managed inside an RCT. The first circumstance was when the treatment being evaluated inside the RCT is shown to have a significant beneficial effect. In this situation, patients outside the trial who received identical interventions experienced similar outcomes. The second circumstance occurred when patients were given an ineffective treatment inside the RCT. In this case, patients outside the trial can expect the same outcomes as patients inside the RCT regardless of which interventions they receive.

Our findings do not support the theory of protocol and care effects proposed by Braunholtz et al[3]. Had there been better care because physicians were following strict study protocol, a difference would be detected between the groups where treatments were identical and amplified within the subgroup of studies where detection bias and expertise bias were most probable. Instead, our results indicate that evidence of a trial effect is only present when the interventions within the RCT are not available or offered to the patients outside the RCT. Differences in the health care worker providing care, the setting in which patients were treated, and the follow-up and attention the patients receive, do not impact their outcome.

The benefit observed in this scenario can be difficult to interpret. The effect size was -0.36 standard deviations, with a confidence interval ranging from -0.61 to -0.12. Cohen's suggestion for interpreting effect sizes can be applied here[184]. An absolute effect of 0.36 can be considered a small to medium change. Furthermore, a recent article by Norman observed that most minimally important differences (MID), the least difference that will be noticed by the patient, were a half standard deviation point[185]. Based on this assumption, the benefit we observed may be quite noticeably felt by participants. At the very least, although still statistically significant, the benefit may not be clinically relevant since the lower limit of the confidence interval falls below the MID of 0.50. Our results confirm those of the earlier systematic reviews done by Vist et al 2008 and Gross et al.[34,33] Both reviews concluded that there was no significant difference between patients treated within a trial and those treated outside the trial. As is expected, our analysis reflected the results of the Vist review once we controlled for the treatment effect. We felt that a shortcoming of those previous reviews was their insistence that the interventions for patients inside and outside the RCT must be identical. Our review posed a more pragmatic question - should patients enrol in a trial regardless of the treatment received outside the trial? Through our

more broad question we were also able to show a difference during those occasions when trial patients are offered an experimental treatment not accessible in standard practice.

Stiller et al [30] conducted an earlier review and found a beneficial effect associated with trial participation stating that trial mortality rates were lower. However, Stiller et al used a vote counting technique, whereby they added up the number of studies where patients inside the RCT faired significantly better than those outside the RCT but did not take into account the size of each study. Thus, larger studies would hold as much weight in the tally as smaller studies that are more prone to type II error. We, however, performed a random effects meta-analysis that takes into account the weight of each study and found no such benefit from trial participation.

Our findings do support encouraging clinicians and patients to participate in RCTs when pre-market novel interventions are being investigated. Thus, access to a novel therapy through trial participation carries with it the risks associated with the unknown side effects but also carries with it the unknown potential benefits of the novel treatment. Both are unknown. In order to fully inform the question, one would need to know several things. First, how many trials are conducted where novel treatments are being investigated? Second, how often are the results of these trials published and is there evidence of a publication bias? Third, what proportion of these trials favor the novel treatment? Without knowing the answers to these questions, the risk associated with participating in an RCT where a novel intervention is being evaluated is unknown. Our review only shows that when the novel treatment works, patients made a good choice in deciding to participate in the RCT.

Our findings and the information that is currently not available to answer the questions we raise, supports the opinions of Vickers[186] and Altman[187] who argue that not only should all clinical trials be reported in public access databases like clinicaltrials.gov but that the raw data

should also be made available. If this were the standard of reporting, our first question would be answered, the second would be moot and the third would be known or could be calculated.

Finally, it is possible that investigators who report the results of patients followed outside of the RCT have, by their follow-up of these patients affected their outcomes. By systematically collecting their outcomes (not wide spread routine practice), the clinician is inserting an intervention by making an effort to have the patient return for a follow-up visit (an effort that is usually initiated by the patient). Thus, one could infer that unless clinicians begin to routinely follow all patients in the same manner as they would for RCTs (i.e. proactively contacting patients and requesting a follow-up even when they fail to attend regular follow-up visits), then one could argue that the trial effect was neutralized by the investigators' curiosity as to whether the outcomes are different for patients inside versus outside a RCT. The majority of studies included in our review (68%) were at low risk of detection bias because both groups of patients were seen by their health care providers in an identical manner. There were differences however, in the number of post- randomization exclusions recorded within the trial and outside. This may affect the outcome if patients were excluded or lost for non-random reasons (either because patients were responding well or poorly) and the reason for exclusion was related to the treatment they received. If we assume that there are no additional differences between patients control outsiders and all outsiders, then the mortality outcome may be immune to this effect. In our review, we did not find evidence of a trail effect in the pooled results of studies that reported mortality.

Finally, it is also possible that patients who agree to attend follow-up visits as part of a surveillance routine (regardless of whether they are better or worse) may be different from those who do not agree to participate or whose participation depends on their outcome (continue to attend until something goes wrong and then consult a different clinician, only attend until they

are asymptomatic and then cannot justify the time/effort to return for a follow-up). In some studies, it is not explicitly stated whether the group of patients aggregated outside the RCT are inclusive of all eligible patients or restricted a subgroup of patients for whom data could be obtained (only those who agreed to participate, those who completed the final outcome, etc).

Limitations

Ideally our systematic review would include a greater number of randomized comparisons, where patients are randomly assigned to be inside or outside the trial. Randomized comparisons reduce the likelihood of selection bias and expertise bias. Unfortunately, our review only identified five such studies. The rest of our studies were observational designs, with only half of the studies having balanced prognostic characteristics. An additional concern with the observational design is that even if known prognostic factors are balanced, it does not mean that unknown characteristics are also balanced. The barrier faced by many systematic reviews is the lack of detail provided by study authors. Many of the studies we included in our review were ambiguous about the type of health care worker providing care, the setting and the presence of co-interventions. The limited reporting on co-interventions did not allow for a meta-analysis to be conducted based on that subgroup. Although we did create subgroups based on the setting and type of healthcare provider, the results were difficult to interpret. Therefore, it is possible that the factors that dictate whether a trial's treatment will be effective may also be related to those qualities on which we had insufficient details. Further, up until this point our review could not include the results of 96 studies because they contained insufficient information about their protocols or their results; 65% of the eligible literature. Inclusion of these studies could potentially influence our conclusions.

References

[1] Kunz, R., Vist, G., Oxman, AD. (2002) Randomisation to protect against selection bias in health- care trials (Cochrane methodology review). Issue 4. Oxford: Update Software.

[2] Peppercorn, J. M., Weeks, J. C., Cook, E. F. C., Joffe, S. (2004) Comparison of outcomes in cancer patients treated within and outside clinical trials: conceptual framework and structured review. *Lancet,* 363(9405), 263- 70.

[3] Braunholtz, D. A., Edwards, S. J. L., Lilford, R. J. (2001) Are randomised clinical trials good for us (in the short term)? Evidence for a "trial effect". *Journal of clinical epidemiology,* 54(3), 217-24.

[4] Majumdar, S. R., Roe, M. T., Peterson, E. D., Chen, A. Y., Gibler, W. B., Armstrong, P.W. (2008) Better outcomes for patients treated at hospitals that participate in clinical trials. *Archives of internal medicine,* 168(6), 657-662.

[5] Devereaux, P .J., Choi, P. T., Lacchetti, C., Weaver, B., Schünemann, H .J., Haines, T., Lavis, J. N., et al. (2002) A systematic review and meta-analysis of studies comparing mortality rates of private for-profit and private not-for-profit hospitals. *Canadian medical association journal,*166(11), 1399- 406.

[6] Tobias, J. S., Souhami, R. L.(1993) Fully informed consent can be needlessly cruel. *British medical journal,* 307(6913), 1199–1201.

[7] McCaffery, K. J., Turner, R., Macaskill, P., Walter, S. D., Chan, S. F., Irwig, L. (2011) Determining the impact of informed choice: separating treatment effects from the effects of choice and selection in randomized trials. *Medical decision making,* 31(2), 229- 36.

[8] Cooper, K. G., Parkin, D. E., Garratt, A. M., Grant, A. M. (1999) Two-year follow up of women randomised to medical management or transcervical resection of the endometrium for

heavy menstrual loss: clinical and quality of life outcomes. *British journal of obstetrics and gynaecology*,106(3), 258–65.

[9] Kitchener, H. C., Burns, S., Nelson, L., Myers, A. J., Fletcher, I., Desai, M., Dunn, G., et al. (2004) A randomised controlled trial of cytological surveillance versus patient choice between surveillance and colposcopy in managing mildly abnormal cervical smears. *British journal of obstetrics and gynecology*,111(1), 63–70.

[10] Noel, P. H., Larme, A. C., Meyer, J., Marsh, G., Correa, A., Pugh, J. A. (1998) Patient choice in diabetes education curriculum: nutritional versus standard content for type 2 diabetes. *Diabetes care*, 21(6), 896–901.

[11] American Federation of Clinical Oncologic Societies. (1998) Access to quality cancer care: consensus statement. *Journal of clinical oncology*,16(4), 1628- 30.

[12] King, N. M. (2000) Defining and describing benefit appropriately in clinical trials. *Journal of law, medicine, and ethics,* 28(4), 332-43.

[13] National Cancer Institute. (1998) Taking part in clinical trials: what cancer patients need to know. *NIH publication*. Available online at http://cancertrials.nci.nih.gov/understanding/bookshelf/treatment/index.html.

[14] Castel, P., Negrier, S., Boissel, J. P. (2006) Why don't cancer patients enter clinical trials? A review. *European journal of cancer*, 42(12),1744-8.

[15] Joseph, R. R. (1994) Viewpoints and concerns of a clinical trial participant. *Cancer*, 74(9 Suppl), 2692-3.

[16] Twelves, C. J., Thomson, C. S.,Young, J., Gould, A. (1998) Entry into clinical trials in breast cancer: the importance of specialist teams. *European journal of cancer*, 34(7), 1004–1007.

[17] Simon, M. S., Brown, D. R., Du, W., LoRusso, P., Kellogg, C. M. (1999) Accrual to breast cancer clinical trials at a university-affiliated hospital in metropolitan Detroit. *American journal of clinical oncology*, 22(1), 42–6.

[18] Siminoff, L. A., Zhang, A., Colabianchi, N., Sturm, C. M., Shen, Q. (2000) Factors that predict the referral of breast cancer patients onto clinical trials by their surgeons and medical oncologists. *Journal of clinical oncology*, 18(6), 1203–11.

[19] Simon, M. S., Du, W., Flaherty, L., Philip, P. A., Lorusso, P., Miree, C., Smith, D., et al. (2004) Factors associated with breast cancer clinical trials participation and enrolment at a large academic medical center. *Journal of clinical oncology*, 22(11), 2046–52.

[20] Kemeny, M. M., Peterson, B. L., Kornblith, A, B., Muss, H. B., Wheeler, J., Levine, E., Bartlett, N., et al. (2003) Barriers to clinical trial participation by older women with breast cancer. *Journal of clinical oncology*, 21(12), 2268–75.

[21] Townsley, C. A., Naidoo, K., Pond, G. R., Melnick W., Straus, S. E., Siu, L. L. (2003) Are older cancer patients being referred to oncologists? A mail questionnaire of Ontario primary care practitioners to evaluate their referral patterns. *Journal of clinical oncology*, 21(24), 4627–35.

[22] Fallowfield, L., Ratcliffe, D., Souhami, R. (1997) Clinicians' attitudes to clinical trials of cancer therapy. *European journal of cancer*, 33 (13), 2221–9.

[23] Lara, P. N., Higdon, R., Lim, N., Kwan, K., Tanaka, M., Lau, D. H., Wun, T., et al. (2001) Prospective evaluation of cancer clinical trial accrual patterns: identifying potential barriers to enrolment. *Journal of clinical oncology*, 19(6), 1728–33.

[24] Madsen, S. M., Mirza, M. R., Holm, S., Hilsted, K. L., Kampmann, K., Riis, P. (2002) Attitudes towards clinical research amongst participants and nonparticipants. *Journal of internal medicine.* 251 (2), 156–168.

[25] Grunfeld, E., Zitzelsberger, L., Coristine, M., Aspelund, F. (2002) Barriers and facilitators to enrollment in cancer clinical trials. *Cancer*, 95 (7), 1577– 83.

[26] Wright, J. R., Crooks, D., Ellis, P. M., Mings, D., Whelan, T. J. (2002) Factors that influence the recruitment of patients to phase III studies in oncology. *Cancer*, 95,1584–91.

[27] Aapro, M. S., Kohne, C., Cohen, H. J, Extermann, M. (2005) Never too old? Age should not be a barrier to enrolment in cancer clinical trials. *Geriatric Oncology,*10(3),198- 204.

[28] Ford, J. G., Howerton, M. W., Lai, G. Y., Gary, T. L., Bolen, S., Gibbons, M. C., Tilburt, J., et al. (2008) Barriers to recruiting underrepresented populations to cancer clinical trials: a systematic review. *Cancer,*112(2), 228-42.

[29] Segelov, E., Tattersall, M. H. N., Coates, A. S. (1992) Redressing the balance – the ethics of not entering an eligible patient on a randomised clinical trial. *Annals of oncology*, 3(2), 103-5

[30] Stiller, C. A. (1994) Centralised treatment, entry to trials and survival. *British journal of cancer*, 70(2) ,352- 62.

[31] Kievit, W., Fransen, J., Oerlemans, A. J., Kuper, H. H., van der Laar, M. A., de Rooij, D. J., De Gendt, C. M., et al. (2007) The efficacy of anti- TNF in rheumatoid arthritis, a comparison between randomised controlled trials and clinical practice. *Annals of rheumatic diseases,* 66(11), 1473- 8.

[32] The Emergency Care Research Institute (2002) Patients' reasons for participation in clinical trials and effect of trial participation on patient outcomes. *ECRI Health Technology Assessment Information Services.* Available online at www.ecri.org/Patient_Information/PatientReference_Guide/evidence.pdf (accessed July 2010)

[33] Gross, C. P., Krumholz, H. M., Van Wye, G., Emanuel, E. J., Wendler, D. (2006) Does random treatment assignment cause harm to research participants? *PLoS medicine*, 3(6), e188.

[34] Vist, G. E., Bryant, D., Somerville, L., Birminghem, T., Oxman, A. D. (2008) Outcomes of patients who participate in randomized controlled trials compared to similar patients receiving similar interventions who do not participate. *Cochrane Database of Systematic Reviews* Issue 3. Art. No.: MR000009. DOI: 10.1002/14651858.MR000009.pub4.

[35] DiCenso, A. (1995) Systematic overviews of the prevention and predictors of adolescent pregnancy. PhD thesis, University of Waterloo, Waterloo, Canada.

[36] Higgins, J. P. T., Thompson, S. G., Deeks, J. J., Altman, D. G. (2003) Measuring inconsistency in meta-analyses. *British medical journal*, 327(7414), 557-60.

[37] Reeves, B. C., van Binsbergen, J., van Weel, C. (2005) Systematic reviews incorporating evidence from nonrandomized study designs: reasons for caution when estimating health effects. *European journal of clinical nutrition*, 59(Suppl 1), S155–S161.

[38]Akaza, H. Hinotsu, S., Aso, Y., Kakizoe, T., Koiso, K. (1995) Bacillus Calmette-Guerin treatment of existing papillary bladder cancer and carcinoma in situ of the bladder. Four-year results. The Bladder Cancer BCG Study Group. *Cancer*, 75(2), 552-9.

[39]Amar, D., Roistacher, N., Burt, M. E, Rusch, V. W, Bains, M. S, Leung, D. H, Downey, R. J., et al. (1997) Effects of diltiazem versus digoxin on dysrhythmias and cardiac function after pneumonectomy. *The annals of thoracic surgery*, 63(5), 1374-81.

[40]Andersson, G., Lundstrom, P., Strom, L. (2003) Internet- based treatment of headache: does telephone contact add anything? *Headache*, 43(4), 353-61.

[41]Antman, A., Amato, D., Wood, W., Carson, J., Suit, H., Proppe, K., Carey, R., et al. (1985) Selection bias in clinical trials. *Journal of clinical oncology*, 3(8), 1142-7.

[42] Ashok, P. W., Kidd, A., Flett, G. M., Fitzmaurice, A., Graham,W., Templeton, A. (2002) A randomized comparison of medical abortion and surgical vacuum aspiration at 10-13 weeks gestation. *Human reproduction*, 17(1), 92-8.

[43] Bain, C., Cooper, K. G., Parkin, D. E. (2001) A partially randomized patient preference trial of microwave endometrial ablation using local anaesthesia and intravenous sedation or general anaesthesia: A pilot study. *Gynaecological endoscopy*, 10(4), 223-8.

[44] Bakker, A., Spinhoven, P., van Balkom, A. J., Vleugel, L., van Dyck, R. (2000) Cognitive therapy by allocation versus cognitive therapy by preference in the treatment of panic disorder. *Psychotherapy and psychosomatics*, 69(5), 240-3.

[45] Balmukhanov, S. B., Beisebaev, A. A., Aitkoolova, Z. I., Mustaphin, J. S., Philippenko, V. I., Rismuhamedova, R. S., Aisarova, A. M., et al. (1989) Intratumoral and parametrial infusion of metronidazole in the radiotherapy of uterine cervix cancer: preliminary report. *International journal of radiation oncology, biology, physics*, 16(4), 1061-3.

[46] Bannister, C. F., Brosius, K. K., Sigl, J. C., Meyer, B. J., Sebel, P.S. The effect of bispectral index monitoring on anesthetic use and recovery in children anesthetized with sevoflurane in nitrous oxide. *Anesthesia and analgesia*, 92(4), 877-81.

[47] Bedi, N., Chilvers, C., Churchill, R., Dewey, M., Duggan, C., Fielding, K., Gretton, V., et al. (2000) Assessing effectiveness of treatment of depression in primary care: Partially randomised preference trial. *British journal of psychiatry*, 177, 312-8.

[48] Bell, R., Palma, S. (2000) Antenatal exercise and birthweight. *The australian & new zealand journal of obstetrics & gynaecology*, 40(1), 70-3.

[49] Bhattacharya, S., Cameron, I. M., Mollison, J., Parkin, D. E., Abramovich, D. R., Kitchener, H. C. (1998) Admission-discharge policies for hysteroscopic surgery: a randomised comparison of day case with in-patient admission. *European journal of obstetrics, gynecology, and reproductive biology*, 76(1), 81-4.

[50] Biasoli, I., Franchi-Rezgui, P., Sibon, D., Briere, J., de Kerviler E., Thieblemont, C., Levy, V., et al. (2008) Analysis of factors influencing inclusion of 102 patients with stage III/IV Hodgkin's

lymphoma in a randomized trial for first-line chemotherapy. *Annals of oncology*, 19(11), 1915-20.

[51] Biederman, J., Herzog, D. B., Rivinus, T. M., Harper, G. P., Ferber, R. A., Rosenbaum, J. F., Harmatz, J. S., et al. (1985) Amitriptyline in the treatment of anorexia nervosa: a double-blind, placebo-controlled study. *Journal of clinical psychopharmacology*, 5(1), 10-6.

[52] Bijkerk, C. J., Muris, J. W., Knottnerus, J. A., Hoes, A. W., de Wit, N.J. (2008) Randomized patients in IBS research had different disease characteristics compared to eligible and recruited patients. *Journal of clinical epidemiology*, 61(11), 1176- 81.

[53] Blichert-Toft, M., Brincker, H., Andersen, J. A., Andersen, K. W., Axelsson, C. K., Mouridsen, H. T., Dombernowsky, P., et al. A Danish randomized trial comparing breast-preserving therapy with mastectomy in mammary carcinoma. Preliminary results. *Acta oncologica*, 27(6A), 671-7.

[54] Blumenthal, J. A., Jiang, W., Babyak, M. A., Krantz, D. S., Frid, D. J., Coleman, R. E., Waugh, R., et al. (1997) Stress management and exercise training in cardiac patients with myocardial ischemia. Effects on prognosis and evaluation of mechanisms. *Archives of internal medicine*, 157(19), 2213-23.

[55] Boesen, E. H., Boesen, S. H., Frederiksen, K., Ross, L., Dahlstrom, K., Schmidt, G., Naested, J., et al. (2007) Survival after a psychoeducational intervention for patients with cutaneous malignant melanoma: a replication study. *Journal of clinical oncology*, 25(36), 5698-703.

[56] Boezaart, A. P., Berry, R. A., Laubscher, J. J., Nell, M.L. (1998) Evaluation of anxiolysis and pain associated with combined peri- and retrobulbar eye block for cataract surgery. *Journal of clinical anesthesia*, 10(3), 204-10.

[57] Brinkhaus, B., Witt, C. M., Jena, S., Liecker, B., Wegscheider, K., Willich, S. N. (2008) Acupuncture in patients with allergic rhinitis: a pragmatic randomized trial. *Annals of allergy asthma and immunology*, 101, 535-43.

[58] Caplan, D. B., Buchanan, C. N. (1984) Treatment of lower respiratory tract infections due to Pseudomonas aeruginosa in patients with cystic fibrosis. *Reviews of infectious disease*, 6(Suppl 3), S705-10.

[59] Coronary artery surgery study (CASS) (1984) A randomized trial of coronary artery bypass surgery. Comparability of entry characteristics and survival in randomized patients and nonrandomized patients meeting randomization criteria. *Journal of the american college of cardiology*, 3(1), 114-28.

[60] Chauhan, S. P., Rutherford, S. E., Hess, L. W., Morrison, J.C. (1992) Prophylactic intrapartum amnioinfusion for patients with oligohydramnios. A prospective randomized study. *Journal of reproductive medicine*, 37(9), 817- 20.

[61] Chesebro, J. H., Fuster, V., Elveback, L. R., McGoon, D. C., Pluth, J. R., Puga, F. J., Wallace, R. B., et al. Trial of combined warfarin plus dipyridamole or aspirin therapy in prosthetic heart valve replacement: danger of aspirin compared with dipyridamole. *American journal of cardiology*, 51(9), 1537- 41.

[62] Bedi, N., Chilvers, C., Churchill, R., Dewey, M., Duggan, C., Fielding, K., Gretton, V., et al. (2000) Assessing effectiveness of treatment of depression in primary care. Partially randomised preference trial. *British journal of psychiatry*, 177, 312-8.

[63] Clagett, G. P., Youkey, J. R., Brigham, R. A., Orecchia, P. M, Salander, J. M., Collins, G. J., Rich, N.M. (1984) Asymptomatic cervical bruit and abnormal ocular pneumoplethysmography: a prospective study comparing two approaches to management. *Surgery*, 96(5), 823- 30.

[64] Clapp, D. W., Kliegman, R. M., Baley, J. E., Shenker, N., Kyllonen, K., Fanaroff, A. A., Berger, M. (1989) Use of intravenously administered immune globulin to prevent nosocomial sepsis in low birth weight infants: report of a pilot study. *Journal of pediatrics*, 115(6), 973-8.

[65] Clemens, J. D., van Loon, F. F., Rao, M., Sack, D. A., Ahmed, F., Chakraborty, J., Khan, M. R., Yunus, M., Harris, J. R., Svennerholm, A. M. (1992) Nonparticipation as a determinant of adverse health outcomes in a field trial of oral cholera vaccines. *American journal of epidemiology*, 135(8), 865-74.

[66] Cooper, K. G., Grant, A. M., Garratt, A.M. (1997) The impact of using a partially randomised patient preference design when evaluating alternative managements for heavy menstrual bleeding. *British journal of obstetrics and gynaecology*, 104(12), 1367- 73.

[67] Cowchock, F. S., Reece, E. A., Balaban, D., Branch, D. W., Plouffe, L. (1992) Repeated fetal losses associated with antiphospholipid antibodies: a collaborative randomized trial comparing prednisone with low-dose heparin treatment. *American journal of obstetrics and gynecology*, 166(5), 1318- 23.

[68] Creutzig, U., Ritter, J., Zimmermann, M., Schellong, G. (1993) Does cranial irradiation reduce the risk for bone marrow relapse in acute myelogenous leukemia? Unexpected results of the childhood acute myelogenous leukemia study BFM-87. *Journal of clinical oncology*, 11(2), 279-86.

[69] Dahan, R., Caulin, C., Figea, L., Kanis, J. A., Caulin, F., Segrestaa, J. M. (1986) Does informed consent influence therapeutic outcome? A clinical trial of the hypnotic activity of placebo in patients admitted to hospital. *British medical journal (clinical research ed.)*, 293(6543), 363-4.

[70] Dalal, H. M., Evans, P. H., Campbell, J. L., Taylor, R. S., Watt, A., Read, K. L., Mourant, A. J., Wingham, J., Thompson, D. R., Pereira Gray, D.J. (2997) Home-based versus hospital-based

rehabilitation after myocardial infarction: A randomized trial with preference arms--Cornwall Heart Attack Rehabilitation Management Study (CHARMS). *International journal of cardiology*, 119(2), 202- 11.

[71] Decensi, A., Robertson, C., Viale, G., Pigatto, F., Johansson, H., Kisanga, E. R., Veronesi, P., et al. (2003) A randomized trial of low-dose tamoxifen on breast cancer proliferation and blood estrogenic biomarkers. *Journal of the national cancer institute*, 95(11), 779- 90.

[72] Detre, K. M., Guo, P., Holubkov, R., Califf, R. M., Sopko, G., Bach, R., Brooks, M. M, et al. (1999) Coronary revascularization in diabetic patients: a comparison of the randomized and observational components of the Aypass Angioplasty Revascularization Investigation (BARI). *Circulation*, 99(5), 633- 40.

[73] Loeffler, M., Diehl, V., Pfreundschuh, M., Ruhl, U., Hasenclever, D., Nisters- Backes, H., Sieber, M., et al. (1997) Dose-response relationship of complementary radiotherapy following four cycles of combination chemotherapy in intermediate-stage Hodgkin's disease. *Journal of clinical oncology*, 15(6), 2275- 87.

[74]Eberhardt, K., Rydgren, L., Fex, E., Svensson, B., Wollheim, F. A. (1996) D-penicillamine in early rheumatoid arthritis: experience from a 2-year double blind placebo controlled study. *Clinical and experimental rheumatology*, 14(6), 625- 31.

[75] Edsmyr, F., Esposti, P. L., Johansson, B., Strindberg, B. (1978) Clinical experimental randomized study of 2.6-cis-diphenylhexamethylcyclotetrasiloxane and estramustine-17-phosphate in the treatment of prostatic carcinoma. *Journal of urology*, 120(6), 705-7.

[76] World Health Organization (WHO) Task Force on Oral Contraceptives. (1988) Effects of hormonal contraceptives on breast milk composition and infant growth. *Studies in family planning*, 19(6 Pt 1), 361-9.

[77] Ekstein, S., Elami, A., Merin, G., Gotsman, M.S., Lotan, C. (2002) Balloon angioplasty versus bypass grafting in the era of coronary stenting. *The Israel medical association journal*, 4(8), 583-9.

[78] Emery, M., Beran, M. D., Darwiche, J., Oppizzi, L., Joris, V., Capel, R., Guex, P., et al. (2003) Results from a prospective, randomized, controlled study evaluating the acceptability and effects of routine pre-IVF counselling. *Human reproduction*, 18(12), 2647- 53.

[79] Euler, A. R., Mitchell, D. K., Kline, R., Pickering, L.K. (2005) Prebiotic effect of fructo-oligosaccharide supplemented term infant formula at two concentrations compared with unsupplemented formula and human milk. *Journal of pediatric gastroenterology and nutrition*, 40(2), 157-64.

[80] Feit, F., Brooks, M.M., Sopko, G., Keller, N.M., Rosen, A., Krone, R., Berger, P.B., Shemin, R., et al. (2000) Long-term clinical outcome in the Bypass Angioplasty Revascularization Investigation Registry: comparison with the randomized trial. BARI Investigators. *Circulation*, 101(24), 2795- 802.

[81] Forbes, G.M., Collins, B.J. (2000) Nitrous oxide for colonoscopy: a randomized controlled study. *Gastrointestinal endoscopy*, 51(3), 271-7.

[82] Franz, M.J., Monk, A., Barry, B., McClain, K., Weaver, T., Cooper, N., Upham, P., Bergenstal, R., et al. (1995) Effectiveness of medical nutrition therapy provided by dietitians in the management of non-insulin-dependent diabetes mellitus: a randomized, controlled clinical trial. *Journal of the American dietetic association*, 95(9), 1009-17.

[83] Gall, C.A., Weller, D., Esterman, A., Pilotto, L., McGorm, K., Hammett, Z., Wattchow, D. (2007) Patient satisfaction and health-related quality of life after treatment for colon cancer. *Diseases of the colon and rectum*, 50(6), 801-9.

[84] Giron, M., Fernandez-Yanez, A., Mana-Alvarenga, S., Molina-Habas, A., Nolasco, A., Gomez-Beneyto, M. (2010) Efficacy and effectiveness of individual family intervention on social and clinical functioning and family burden in severe schizophrenia: A 2-year randomized controlled study. *Psychological medicine*, 40(1), 73- 84.

[85] Goodkin, D. E., Plencner, S., Palmer-Saxerud, J.,Teetzen, M., Hertsgaard, D. (1987) Cyclophosphamide in chronic progressive multiple sclerosis. Maintenance vs nonmaintenance therapy. *Archives of neurology*, 44(8), 823-7.

[86] Gossop, M., Johns, A., Green, L. (1986) Opiate withdrawal: inpatient versus outpatient programmes and preferred versus random assignment to treatment. *British medical journal*, 293(6539), 103-4.

[87] Grant, A.M., Wileman, S.M., Ramsay, C.R., Mowat, N.A., Krukowski, Z.H., Heading, R.C., Thursz, M. R., et al. (2008) Minimal access surgery compared with medical management for chronic gastro-oesophageal reflux disease: UK collaborative randomised trial. *British medical journal*, 15, 337.

[88] Gunn, T. R., Thompson, J. M., Jackson, H., McKnight, S., Buckthought, G., Gunn, A.J. (2000) Does early hospital discharge with home support of families with preterm infants affect breastfeeding success? A randomized trial. *Acta paediatrica*, 89(11), 1358- 63.

[89] Helsing, M., Bergman, B., Thaning, L., Hero, U. (1998) Quality of life and survival in patients with advanced non-small cell lung cancer receiving supportive care plus chemotherapy with carboplatin and etoposide or supportive care only. A multicentre randomised phase III trial. *European journal of cancer*, 34(7), 1036- 44.

[90] Henriksson, P., Edhaq, O. (1986) Orchidectomy versus oestrogen for prostatic cancer: cardiovascular effects. *British medical journal*, 293(6544), 413-5.

[91] Heuss, L.T., Drewe, J., Schnieper, P., Tapparelli, C.B., Pflimlin, E., Beglinger, C. (2004) Patient-controlled versus nurse-administered sedation with propofol during colonoscopy. A prospective randomized trial. *The American journal of gastroenterology*, 99(3), 511-8.

[92] Hoh, R., Pelfini, A., Neese, R.A., Chan, M., Cello, J.P., Cope, F.O., Abbruzese, B.C., Richards, E.W., et al. (1998) De novo lipogenesis predicts short-term body-composition response by bioelectrical impedance analysis to oral nutritional supplements in HIV-associated wasting. *The American journal of clinical nutrition*, 68(1), 154- 63.

[93] Howard, L.M., Flach, C., Leese, M., Byford, S., Killaspy, H., Cole, L., Lawlor, C., Johnson, S. (2009) Effectiveness and cost-effectiveness of admissions to women's crisis houses compared with traditional psychiatric wards: Pilot patient-preference randomised controlled trial. *The journal of nervous and mental disease*, 197(10), 722-7.

[94] Howie, F.L., Henshaw, R.C., Naji, S.A., Russell, I.T., Templeton, A.A. (1997) Medical abortion or vacuum aspiration? Two year follow up of a patient preference trial. *British journal of obstetrics and gynaecology*, 104(7), 829-33.

[95] Jena, S., Witt, C. M., Brinkhaus, B., Wegscheider, K., Willich, S.N. (2008) Acupuncture in patients with headache. *Cephalalgia*, 28(9), 969- 79.

[96] Jensen, L.B., Vestergaard, P., Hermann, A.P., Gram, J., Eiken, P., Abrahamsen, B., Brot, C., Kolthoff, N., et al. (2003) Hormone replacement therapy dissociates fat mass and bone mass, and tends to reduce weight gain in early postmenopausal women: a randomized controlled 5-year clinical trial of the Danish Osteoporosis Prevention Study. *Journal of bone and mineral research*, 18(2), 333-42.

[97] Kane, W.J. (1988) Direct current electrical bone growth stimulation for spinal fusion. *Spine*, 13(3), 363-5.

[98] Karande, V.C., Korn, A., Morris, R., Rao, R., Balin, M., Rinehart, J., Dohn, K., Gleicher, N. (1999) Prospective randomized trial comparing the outcome and cost of in vitro fertilization with that of a traditional treatment algorithm as first-line therapy for couples with infertility. *Fertility and sterility*, 71(3), 468-75.

[99] Kayser, B., Hulsebosch, R., Bosch, F. (2008) Low-dose acetylsalicylic acid analog and acetazolamide for prevention of acute mountain sickness. *High altitude medicine & biology*, 9(1), 15-23.

[100] Kendrick, D., Fielding, K., Bentley, E., Miller, P., Kerslake, R., Pringle, M. (2001) The role of radiography in primary care patients with low back pain of at least 6 weeks duration: a randomised (unblinded) controlled trial. *Health technology assessment*, 5(30), 1-69.

[101] Kieler, H., Hellberg, D., Nilsson, S., Waldenstrom ,U., Axelsson, O. (1998) Pregnancy outcome among non-participants in a trial on ultrasound screening. *Ultrasound in obstetrics & gynecology*, 11(2), 104-9.

[102] King, M., Sibbald, B., Ward, E., Bower, P., Lloyd, M., Gabbay, M., Byford, S. (2000) Randomised controlled trial of non-directive counselling, cognitive-behaviour therapy and usual general practitioner care in the management of depression as well as mixed anxiety and depression in primary care. *Health technology assessment*, 4(19), 1-83.

[103] Kirke, P.N., Daly, L.E., Elwood, J.H. (1992) A randomised trial of low dose folic acid to prevent neural tube defects. The Irish Vitamin Study Group. *Archives of disease in childhood*, 67(12), 1442-6.

[104] Koch-Henriksen, N., Sorensen, P.S., Christensen, T., Frederiksen, J., Ravnborg, M., Jensen, K., Heltberg, A., Kristensen, O., et al. (2006) A randomized study of two interferon-beta treatments in relapsing-remitting multiple sclerosis. *Neurology*, 66(7), 1056- 60.

[105] Lansky, D., Vance, M.A. (1983) School-based intervention for adolescent obesity: analysis of treatment, randomly selected control, and self-selected control subjects. *Journal of consulting and clinical psychology,* 51(1), 147-8.

[106] Lichtenberg, P,. Levinson, D., Sharshevsky, Y., Feldman, D., Lachman, M. (2008) Clinical case management of revolving door patients - a semi-randomized study. *Acta psychiatrica scandinavica,* 117(6), 449-54.

[107] Lidbrink, E., Frisell, J., Brandberg, Y., Rosendahl, I., Rutgvist, L.E. (1995) Nonattendance in the Stockholm mammography screening trial: relative mortality and reasons for nonattendance. *Breast cancer research and treatment,* 35(3), 267- 75.

[108] Link, M.P., Goorin, A.M., Horowitz, M., Meyer, W.H., Belasco, J., Baker, A., Ayala, A., Shuster, J. (1995) Adjuvant chemotherapy of high-grade osteosarcoma of the extremity. Updated results of the Multi-Institutional Osteosarcoma Study. *Breast cancer research and treatment*, 35(3), 267- 75.

[109] Liu, P.L., Shen, W.Z., Jing, T., Zhou, Z., Chen, Y., Li, X., Yuan, L. (2009) Effects of compound Red-rooted Salvia and aspirin on platelet aggregation and PKB activity in the elderly patients with ACS. *Chinese journal of new drugs,* 18(10), 900-2.

[110] Lock, C., Wilson, J., Steen, N., Eccles, M., Mason, H., Carrie, S., Clarke, R., Kubba, H., et al. (2010) North of England and Scotland Study of Tonsillectomy and Adeno-tonsillectomy in Children(NESSTAC): a pragmatic randomised controlled trial with a parallel non-randomised preference study. *Health technology assessment,* 14(13), 1-164.

[111] Luby, S., Agboatwalla, M., Schnell, B.M., Hoekstra, R.M., Rahbar, M.H., Keswick, B.H. (2002) The effect of antibacterial soap on impetigo incidence, Karachi, Pakistan. *The American journal of tropical medicine and hygiene,* 67(4), 430-5.

[112] Macdonald, J.H., Marcora, S.M., Jibani, M.M., Kumwenda, M.J., Ahmed, W., Lemmey, A.B. (2007) Nandrolone decanoate as anabolic therapy in chronic kidney disease: a randomized phase II dose-finding study. *Nephron clinical practice*, 106(3), c125-35.

[113] MacLennan, A.H., Kerin, J.F., Kirby, C., Grant, P., Warnes, G.M., Cox, L.W., Bryant-Greenwood, G., Greenwood, F. (2007) The effect of porcine relaxin vaginally applied at human embryo transfer in an in vitro fertilization programme. *The Australian & New Zealand journal of obstetrics & gynaecology*, 25(1), 68-71.

[114] MacMillan, J.F., Crow, T.J., Johnson, A.L., Johnstone, E.C. (1986) Short- term outcome in trial entrants and trial eligible patients. *British journal psychiatry*, 148, 128-33.

[115] Mahon, J., Laupacis, A., Donner, A., Wood, T. (1996) Randomised study of n of 1 trials versus standard practice. *British medical journal*, 312(7038), 1069- 74.

[116] Mahon, J., Laupacis, A., Hodder, R.V., McKim, D.A., Paterson, N.A., Wood, T.E., Donner, A. (1999) Theophylline for irreversible chronic airflow limitation : a randomized study comparing n of 1 trials to standard practice. *Chest*, 115(1), 38-48.

[117] Marcinczyk, M.J., Nicholas, G.G., Reed, J.F., Nastasee, S.A. (1997) Asymptomatic carotid endarterectomy: patient and surgeon selection. *Stroke*, 28(2), 291-6.

[118] Martin, L.F. (1994) Stress ulcers are common after aortic surgery. Endoscopic evaluation of prophylactic therapy. *The American surgeon*, 60(3), 169-74.

[119] Martinez-Amenos, A., Fernandez Ferre, M.L., Mota Vidal, C., Alsina Rocasalbas, J. (1990) Evaluation of two educative models in a primary care hypertension programme. *Journal of human hypertension*, 4(4), 362-4.

[120] Masood, J., Shah, N., Lane, T., Andrews, H., Simpson, P., Barua, J.M. (2002) Nitrous oxide (Entonox) inhalation and tolerance of transrectal ultrasound guided prostate biopsy: a double-blind randomized controlled study. *The journal of urology*, 168(1), 116-20.

[121] Mattila, P.S., Joki-Erkkila, V.P., Kilpi, T., Jokinen, J., Herva, E., Puhakka, H. (2003) Prevention of otitis media by adenoidectomy in children younger than 2 years. *Archives of otolaryngology*, 129(2), 163-8.

[122] Mayo Asymptomatic Carotid Endarterectomy Study Group. (1992) Results of a randomized controlled trial of carotid endarterectomy for asymptomatic carotid stenosis. *Mayo clinic proceedings*, 67(6), 513-8.

[123] McCaughey, E.S., Mulligan, J., Voss, L.D., Betts, P.R. (1998) Randomised trial of growth hormone in short normal girls. *Lancet*, 351(9107), 940-4.

[124] McKay, J.R., Alterman, A.I., McLellan, A.T., Boardman, C.R., Mulvaney, F.D., O'Brien, C.P. (1998) Random versus nonrandom assignment in the evaluation of treatment for cocaine abusers. *Journal of consulting and clinical psychology*, 66(4), 697-701.

[125] McKay, J.R., Alterman, A.I., McLellan, A.T., Snider, E.C., O'Brien, C.P. (1995) Effect of random versus nonrandom assignment in a comparison of inpatient and day hospital rehabilitation for male alcoholics. *Journal of consulting and clinical psychology*, 63(1), 70-8.

[126] Melchart, D., Steger, H.G., Linde, K., Makarian, K., Hatahet, Z., Brenke, R., Saller, R. (2002) Integrating patient preferences in clinical trials: a pilot study of acupuncture versus midazolam for gastroscopy. *Journal of alternative and complementary medicine*, 8(3), 265-74.

[127] Moertel, C., Childs, D.S., O'Fallon, J.R., Holbrook, M.A., Schutt, A.J., Reitemeier, R.J. (1984) Combined 5-Fluorouracil and radiation therapy as a surgical adjuvant for poor prognosis gastric carcinoma. *Journal of clinical oncology*, 2(11), 1249-54.

[128] Mori, A., Nobutoshi, F., Asano, T., Maruyama, T., Ohashi, N., Okumura, S., Inoue, H., Takekoshi, S., et al. (2006) Cardiovascular tolerance in unsedated upper gastrointestinal endoscopy: Prospective randomized comparison between transnasal and conventional oral procedures. *Digestive endoscopy*, 18(4), 282-87.

[129] Morrison, D.A., Sethi, G., Sacks, J., Henderson, W., Grover, F., Sedlis, S., Esposito, R., Ramanathan, K.B., et al. (2002) VA AWESOME (Angina With Extremely Serious Operative Mortality Evaluation) Multicenter Registry. Percutaneous coronary intervention versus coronary bypass graft surgery for patients with medically refractory myocardial ischemia and risk factors for adverse outcomes with bypass. *Journal of the American college of cardiology*, 39(2), 266-73.

[130] Nagel, H.T., Vandenbussche, F.P., Keirse, M.J., Oepkes, D., Oosterwijk, J.C., Beverstock, G., Kanhai, H.H. (1998) Amniocentesis before 14 completed weeks as an alternative to transabdominal chorionic villus sampling: a controlled trial with infant follow-up. *Prenatal diagnosis*, 18(5), 465-75.

[131] Neldam, S., Osler, M., Hansen, P.K., Nim, J., Smith, S.F., Hertel, J. (1986) Intrapartum fetal heart rate monitoring in a combined low- and high-risk population: a controlled clinical trial. *European journal of obstetrics, gynecology, and reproductive biology*, 23(1-2), 1-11.

[132] Nicolaides, K., Brizot, M.L., Patel, F., Snijders, R. (1994) Comparison of chorionic villus sampling and amniocentesis for fetal karyotyping at 10- 13 weeks' gestation. *Lancet*, 344(8920), 435-9.

[133] Ogden, J.A., Alvarez, R.G., Levitt, R.L., Johnson, J.E., Marlow, M.E. (2004) Electrohydraulic high-energy shock-wave treatment for chronic plantar fasciitis. *The journal of bone and joint surgery*, 86- A(10), 2216-28.

[134] Palmon, S.C., Liu, M., Moore, L.E., Kirsch, J.R. (1996) Capnography facilitates tight control of ventilation during transport. *Critical care medicine,* 24(4), 608-11.

[135] Panagopoulou, E., Montgomery, A., Tarlatzis, B. (2009) Experimental emotional disclosure in women undergoing infertility treatment: Are drop outs better off? *Social science & medicine*, 69(5), 678- 81.

[136] Paradise, J.L., Bluestone, C.D., Bachman, R.Z., Colborn, D.K., Bernard, B.S., Taylor, F.H., Rogers, K.D., Schwarzbach, R.H. et al. (1984) Efficacy of tonsillectomy for recurrent throat infection in severely affected children. Results of parallel randomized and nonrandomized clinical trials. *New England journal of medicine*, 310(11), 674-83.

[137] Petersen, M.K., Andersen, K.V., Andersen, N.T., Soballe, K. (2007) To whom do the results of this trial apply? External validity of a randomized controlled trial involving 130 patients scheduled for primary total hip replacement. *Acta orthopaedic*, 78(1), 12-8.

[138] Raistrick, D., West, D., Finnegan, O., Thistlethwaite, G., Brearley, R., Banbery, J. (2005) A comparison of buprenorphine and lofexidine for community opiate detoxification: results from a randomized controlled trial. *Addiction*, 100(12), 1860-7.

[139] Rigg, J.R., Jamrozik, K., Myles, P.S., Silbert, B., Peyton , P., Parsons, R.W., Collins, K. (2000) Design of the multicenter Australian study of epidural anesthesia and analgesia in major surgery: the MASTER trial. *Controlled clinical trials*, 21(3), 244- 56.

[140] Rorbye, C., Norgaard, M., Nilas, L. (2005) Medical versus surgical abortion: comparing satisfaction and potential confounders in a partly randomized study. *Human reproduction*, 20(3), 834- 8.

[141] Rosen, M.A., Roizen, M.F., Eger, E.I., Glass, R.H., Martin, M., Dandekar, P.V., Dailey, P.A., Litt, L. (1987) The effect of nitrous oxide on in vitro fertilization success rate. *Anesthesiology*, 67(1), 42-4.

[142] Salisbury, C., Francis, C., Rogers, C., Parry, K., Thomas, H., Chadwick, S., Turton, P. (2002) A randomised controlled trial of clinics in secondary schools for adolescents with asthma. *The British jounal of general practice*, 52(485), 988-96.

[143] Sesso, H.D., Gaziano, J.M., VanDenburgh, M., Hennekens, C.H., Glynn, R.J., Buring, J.E. (2002) Comparison of baseline characteristics and mortality experience of participants and

nonparticipants in a randomized clinical trial: the Physicians' Health Study. *Controlled clinical trials*, 23(6), 686-702.

[144] Shain, R.N., Ratsula, K., Toivonen, J., Lahteenmaki, P., Luukkainen, T., Holden, A.E., Rosenthal, M. (1989) Acceptability of an experimental intracervical device: results of a study controlling for selection bias. *Contraception*, 39(1), 73-84.

[145] Smith, P., Arnesen, H. (1990) Mortality in non-consenters in a post-myocardial infarction trial. *Journal of internal medicine*, 228(3), 253-6.

[146] Smuts, C.M., Borod, E., Peeples, J.M., Carlson, S.E. (2003) High-DHA eggs: feasibility as a means to enhance circulating DHA in mother and infant. *Lipids*, 38(4), 407-14.

[147] Stecksen-Blicks, C., Holgerson, P.L., Twetman, S. (2008) Effect of xylitol and xylitol-fluoride lozenges on approximal caries development in high-caries-risk children. *International journal of paediatric dentistry*, 18(3), 170-7.

[148] Stern, C., Chamley, L., Norris, H., Hale, L., Baker, H.W. (2003) A randomized, double-blind, placebo-controlled trial of heparin and aspirin for women with in vitro fertilization implantation failure and antiphospholipid or antinuclear antibodies. *Fertility and sterility*, 80(2), 376-83.

[149] Stith, S.M., Rosen, K.H., McCollum, E.E., Thomsen, C.J. (2004) Treating intimate partner violence within intact couple relationships: outcomes of multi-couple versus individual couple therapy. *Journal of marital and family therapy*, 30(3), 305-18.

[150] Stockton, K.A., Mengersen, K.A. (2009) Effect of multiple physiotherapy sessions on functional outcomes in the initial postoperative period after primary total hip replacement: a randomized controlled trial. *Archives of physical medicine and rehabilitation*, 90(10), 1652-7.

[151] Strandberg, T.E., Salomaa, V.V., Vanhanen, H.T., Naukkarinen, V.A., Sarna, S.J., Miettinen, T.A. (1995) Mortality in participants and non-participants of a multifactorial prevention study of

cardiovascular diseases: a 28 year follow up of the Helsinki Businessmen Study. *British heart journal*, 74(4), 449- 54.

[152] Suherman, S.K., Affandi, B., Korver, T. (1999)The effects of Implanon on lipid metabolism in comparison with Norplant. *Contraception*, 60(5), 281- 7.

[153] Sullivan, M.P., Fuller, L.M., Chen, T., Fisher, R., Fryer, C., Gehan, E., Gilchrist, G.S., Hays, D., et al. (1982) Intergroup Hodgkin's disease in children study of stages I and II: a preliminary report. *Cancer treatment reports*, 66(4), 937-47.

[154] Sundar, S., Rai, M., Chakravarty, J., Agarwal, D., Agrawal, N., Vaillant, M., Olliaro, P., Murray, H.W. (2008) New treatment approach in Indian visceral leishmaniasis: single-dose liposomal amphotericin B followed by short-course oral miltefosine. *Clinical infectious diseases*, 47(8), 1000-6.

[155] Taddio, A., Lee, C., Yip, A., Parvez, B., McNamara, P.J., Shah, V. (2006) Intravenous morphine and topical tetracaine for treatment of pain in neonates undergoing central line placement. *The journal of the American medical association*, 295(7), 793-800.

[156] Tanai, C., Nokihara, H., Yamamoto, S., Kunitoh, H., Yamamoto, N., Sekine, I., Ohe, Y., Tamura, T. (2009) Characteristics and outcomes of patients with advanced non-small-cell lung cancer who declined to participate in randomised clinical chemotherapy trials. *British journal of cancer*, 100(7), 1037- 42.

[157] Tanaka, S., Tsuchida, H., Namba, H., Namiki, A. (1994) Clonidine and lidocaine inhibition of isoflurane-induced tachycardia in humans. *Anesthesiology*, 81(6), 1341-9.

[158] Taplin, D., Meinking, T.L., Castillero, P.M., Sanchez, R. (1986) Permethrin 1% creme rinse for the treatment of Pediculus humanus var capitis infestation. *Pediatric dermatology*, 3(4), 344-8.

[159] Tenenbaum, A., Motro, M., Fisman, E.Z., Boyko,V., Mandelzweig, L., Shotan, A., Behar, S. (2002) Does participation in a long-term clinical trial lead to survival gain for patients with coronary artery disease? *The American journal of medicine*, 112(7), 545-8.

[160] Toprak, A., Erenus, M., Ilhan, A.H., Haklar, G., Fak, A.S., Oktay, A (2005) The effect of postmenopausal hormone therapy with or without folic acid supplementation on serum homocysteine level. *Climacteric*, 8(3), 279-86.

[161] Underwood, M., Ashby, D., Carnes, D., Catelnuovo, E., Cross, P., Harding, G., Hennessy, E., Letley, L., et al. (2008) Topical or oral ibuprofen for chronic knee pain in older people. The TOIB study. *Health technology assessment*, 12(22), iii-iv, ix-155.

[162] Urban, P., Stauffer, J.C., Bleed, D., Khatchatrian, N., Amann, W., Bertel, O., van den Brand, M., Danchin, N., et al. (1999) A randomized evaluation of early revascularization to treat shock complicating acute myocardial infarction. The (Swiss) Multicenter Trial of Angioplasty for Shock-(S)MASH. *European heart journal*, 20(14), 1030-8.

[163] van Bergen, P.F., Jonker, J.J., Molhoek, G.P., van der Burgh, P.H., van Domburg, R.T., Deckers, J.W., Hofman, A. (1995) Characteristics and prognosis of non-participants of a multi-centre trial of long-term anticoagulant treatment after myocardial infarction. *International journal of cardiology*, 49(2), 135-41.

[164] Van, H.L., Dekker, J., Koelen, J., Kool, S., van Aalst, G., Hendriksen, M., Peen, J., Schoevers, R. (2009) Patient preference compared with random allocation in short-term psychodynamic supportive psychotherapy with indicated addition of pharmacotherapy for depression. *Psychotherapy research*, 19(2), 205-12.

[165] Verdonck, L.F., van Putten, W.L., Hagenbeek, A., Schouten, H.C., Sonneveld, P., van Imhoff, G.W., Kluin-Nelemans, H.C., Raemaekers, J.M., et al. (1995) Comparison of CHOP

chemotherapy with autologous bone marrow transplantation for slowly responding patients with aggressive non-Hodgkin's lymphoma. *New England journal of medicine*, 332(16), 1045-51.

[166] Vind, A.B., Andersen, H.E., Pedersen, K.D., Jorgensen, T., Schwarz, P. (2009) Baseline and follow-up characteristics of participants and nonparticipants in a randomized clinical trial of multifactorial fall prevention in Denmark. *Journal of the American geriatrics society*, 57(10), 1844-9.

[167] Walker, W.S., Raychaudhury, T., Faichney, A., Prescott, R.J., Tonkin, R.W., Sang, C.T., Cameron, E.W., Reid, K.G., et al . (1986) Wound colonisation following cardiac surgery. *The journal of cardiovascular surgery*, 27(6), 662-6.

[168] Wallage, S., Cooper, K.G., Graham, W.J., Parkin, D.E. (2003) A randomised trial comparing local versus general anaesthesia for microwave endometrial ablation. *International journal of obstetrics and gynaecology*, 110(9), 799-807.

[169] Watzke, B., Ruddel, H., Jurgensen, R., Koch, U., Kriston, L., Grothgar, B., Schulz, H. (2010) Effectiveness of systematic treatment selection for psychodynamic and cognitive-behavioural therapy: Randomised controlled trial in routine mental healthcare. *The British journal of psychiatry*, 197(2), 96-105.

[170] Welt, S.I., Dorminy, J.H., Jelovsek, F.R., Crenshaw, M.C., Gall, S.A. (1981) The effect of prophylactic management and therapeutics on hypertensive disease in pregnancy: Preliminary studies. *Obstetrics and gynecology*, 57(5), 557- 65.

[171] West, J., Wright, J., Tuffnell, D., Jankowicz, D., West, R. (2005) Do clinical trials improve quality of care? A comparison of clinical processes and outcomes in patients in a clinical trial and similar patients outside a trial where both groups are managed according to a strict protocol. *Quality & safety in health care*, 14(3), 175- 8.

[172] Wetzner, S.M., Vincent, M.E., Robbins, A.H. (1979) Ceruletide-assisted cholecystography: a clinical assessment. *Radiology*, 131(1), 23-6.

[173] Wieringa-de Waard, M., Vos, J., Bonsel, G.J., Bindels, P.J., Ankum,W.M. (2002) Management of miscarriage: a randomized controlled trial of expectant management versus surgical evacuation. *Human reproduction*, 17(9), 2445-50.

[174] Williford, W.O., Krol, W.F., Buzby, G.P. (1993) Comparison of eligible randomized patients with two groups of ineligible patients: can the results of the VA Total Parenteral Nutrition clinical trial be generalized? *Journal of clinical epidemiology*, 46(9), 1025-34.

[175] Witt, C.M., Jena, S., Brinkhaus, B., Liecker, B.,Wegscheider, K., Willich, S.N. (2006) Acupuncture in patients with osteoarthritis of the knee or hip: a randomized, controlled trial with an additional nonrandomized arm. *Arthritis and rheumatism*, 54(11), 3485- 93.

[176] Witt, C.M., Jena, S., Brinkhaus, B., Liecker, B.,Wegscheider, K., Willich, S.N. (2006) Acupuncture for patients with chronic neck pain. *Pain*, 125(1-2), 98-106.

[177] Witt, C.M., Jena, S., Selim, D., Brinkhaus, B., Reinhold, T., Wruck, K., Liecker, B., Linde, K.,et al. (2006) Pragmatic randomized trial evaluating the clinical and economic effectiveness of acupuncture for chronic low back pain. *American journal of epidemiology,* 164(5), 487- 96.

[178] Witt, C.M., Reinhold, T., Brinkhaus, B., Roll, S., Jena, S., Willich, S.N. (2008) Acupuncture in patients with dysmenorrhea: a randomized study on clinical effectiveness and cost-effectiveness in usual care. *American journal of obstetrics and gynecology*, 198(2), 166 e1-8.

[179] Woodhouse, S.P., Cox, S., Boyd, P., Case, C., Weber, M. (1995) High dose and standard dose adrenaline do not alter survival, compared with placebo, in cardiac arrest. *Resuscitation*, 30(3), 243-9.

[180] Wyse, D.G., Hallstrom, A., McBride, R., Cohen, J.D., Steinberg, J.S., Mahmarian, J. (1991) Events in the Cardiac Arrhythmia Suppression Trial (CAST): mortality in patients surviving

open label titration but not randomized to double-blind therapy. *Journal of the American college of cardiology*, 18(1), 20-8.

[181] Yamamoto, H., Hughes, R.W., Schroeder, K.W., Viggiano, T.R., DiMagno, E.P. (1992) Treatment of benign esophageal stricture by Eder-Puestow or balloon dilators: a comparison between randomized and prospective nonrandomized trials. *Mayo clinic proceedings*, 67(3), 228-36.

[182] Yamani, M.H., Avery, R., Mawhorter, S.D., McNeill, A., Cook, D., Ratliff, N.B., Pelegrin, D., Colosimo, et al. (2005) The impact of CytoGam on cardiac transplant recipients with moderate hypogammaglobulinemia: a randomized single-center study. *The journal of heart and lung transplantation*, 24(11), 1766-9.

[183] Yersin, B., Besson, J., Duc-Mingot, S., Burnand, B. (1996) Screening and referral of alcoholic patients in a general hospital. *European addiction research*, 2(2), 94-101.

[184] Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). New York: Academic Press.

[185] Norman, G.R., Sloan, J.A., Wyrwich, K.W. (2003) Interpretation of changes in health-related quality of life. The remarkable universality of half a standard deviation. *Medical care*, 41(5), 582- 92.

[186] Vickers, A.J. (2006) Whose data set is it anyway? Sharing raw data from randomized trials. *Trials*, 16(7), 15.

[187] Altman, D.G., Cates, C. (2001) Authors should make their data available. *British medical journal*, 323 (7320), 1069- 70.

Tables and Figures

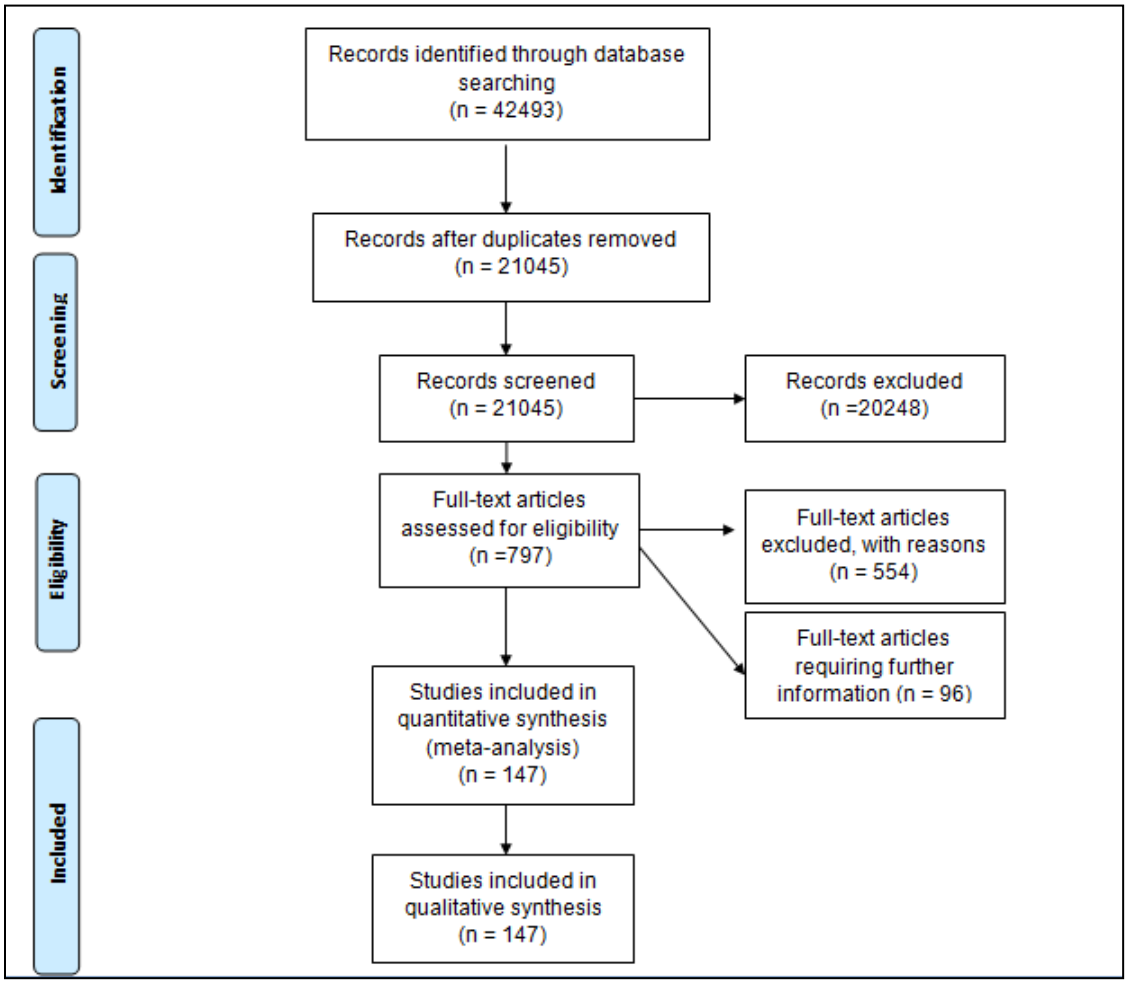


Figure 1: PRISMA flow chart of studies starting at our initial search yield, and ending with the number of studies included in our final meta- analysis.

Data available	**Data needed**	**Assumptions/imputations made**
Standard error (SE) of the difference	Standard deviation (SD) of the difference	Multiply the SE with the square root of the sample size
Confidence intervals around the difference	Standard deviation of the difference	(i) assumed a standard normal distribution for N> 100 (ii) assumed a t-distribution for N< 100
p- value, mean difference	Standard error of the difference	Converted the p-value to the t-value at that degree of freedom. Divide the mean difference by the t- value.
Baseline and change scores	Final score	Add/minus change score from baseline
Standard deviation of baseline and change scores	Standard deviation of final scores	Added the baseline and changes variances together
Range	Standard deviation	No appropriate conversion could be made

Table 1: List of assumptions and imputations used to calculate our missing data.

Author	Year	Reason for exclusion
"Abraham N"	2006	Eligible patients were not followed outside of RCT, within 2 months
"Abrams R"	1986	No RCT in this study
"Adenis A"	1995	Eligible patients were not followed outside of RCT, within 2 months
"Adis Data Information BV"	2006	No RCT in this study
"Adis Data Information"	2006	No RCT in this study
"Adriaensen M"	2004	No health outcome evaluated
"Agostoni C"	2007	Non-participants were almost eligible for participation..
"Ahluwalia J"	2002	Eligible patients were not followed outside of RCT, within 2 months
"Ai X"	2010	Eligible patients were not followed outside of RCT, within 2 months
"Al-Awadi K"	2008	Eligible patients were not followed outside of RCT, within 2 months
"Amundsen T"	2000	Eligible patients were not followed outside of RCT, within 2 months
"Anaya D"	2008	No RCT in this study
"Ardic F"	2007	Eligible patients were not followed outside of RCT, within 2 months
"Arndt C"	2009	Eligible patients were not followed outside of RCT, within 2 months
"Aronoff S"	1984	Eligible patients were not followed outside of RCT, within 2 months
"Arvanitakis"	2007	Only 56% of the cohort were eligible for the RCT.
"Åsenlöf P"	2009	Eligible patients were not followed outside of RCT, within 2 months
"Ashok P"	2005	No health outcome evaluated
"Ashok P"	2005	Same population as Ashok 2002.
"Auvinen A"	2004	Eligible patients were not followed outside of RCT, within 2 months
"Avenell A"	2004	Eligible patients were not followed outside of RCT, within 2 months
"Avrech O"	2004	Eligible patients were not followed outside of RCT, within 2 months
"Baer M"	2008	Eligible patients were not followed outside of RCT, within 2 months
"Bailey A"	2010	Eligible patients were not followed outside of RCT, within 2 months
"Bailie G"	1980	Eligible patients were not followed outside of RCT, within 2 months
"Balch C"	2001	Eligible patients were not followed outside of RCT, within 2 months
"Banno H"	2009	Eligible patients were not followed outside of RCT, within 2 months
"Bar F"	1992	Eligible patients were not followed outside of RCT, within 2 months
"Barrett B"	1992	Eligible patients were not followed outside of RCT, within 2 months
"Bastit L"	2008	Eligible patients were not followed outside of RCT, within 2 months
"Baum E"	1979	Eligible patients were not followed outside of RCT, within 2 months
"Behar J"	1975	Eligible patients were not followed outside of RCT, within 2 months
"Belkhadhir J"	1993	Eligible patients were not followed outside of RCT, within 2 months
"Bellandi F"	1995	Eligible patients were not followed outside of RCT, within 2 months
"Benasso M"	1997	Eligible patients were not followed outside of RCT, within 2 months
"Bergmann J"	1994	Patients considered to be outside the trial were also randomized, although they were not fully informed about trial participation
"Berry H"	1980	Eligible patients were not followed outside of RCT, within 2 months

"Bertelsen K"	1994	Eligible patients were not followed outside of RCT, within 2 months
"Bertrand O"	2008	The article was not in English
"Beutel M"	2005	Eligible patients were not followed outside of RCT, within 2 months
"Bezwoda W"	1995	Eligible patients were not followed outside of RCT, within 2 months
"Bhatia S"	2007	Eligible patients were not followed outside of RCT, within 2 months
"Bijkerk C"	2008	No health outcome evaluated
"Bilotta F"	2006	Eligible patients were not followed outside of RCT, within 2 months
"Bissada N"	1976	Cohort patients were not eligible for one of the trial arms
"Bissada N"	1977	Cohort patients were not eligible for one of the trial arms
"Bisschop M"	1985	Eligible patients were not followed outside of RCT, within 2 months
"Blackburn G"	2000	Eligible patients were not followed outside of RCT, within 2 months
"Blackshear J"	1996	Eligible patients were not followed outside of RCT, within 2 months
"Blanchon T"	2007	Eligible patients were not followed outside of RCT, within 2 months
"Blanco M"	2007	Eligible patients were not followed outside of RCT, within 2 months
"Blankenship J"	2009	Eligible patients were not followed outside of RCT, within 2 months
"Bonenkamp J"	1999	Eligible patients were not followed outside of RCT, within 2 months
"Botto G"	1997	Eligible patients were not followed outside of RCT, within 2 months
"Boulton D"	2010	Eligible patients were not followed outside of RCT, within 2 months
"Bousquet J"	2009	Eligible patients were not followed outside of RCT, within 2 months
"Bower P"	2000	No health outcome evaluated
"Boyle B"	2008	Eligible patients were not followed outside of RCT, within 2 months
"Boyle B"	2008	Eligible patients were not followed outside of RCT, within 2 months
"Brady III C"	1985	Eligible patients were not followed outside of RCT, within 2 months
"Brain E"	2005	Eligible patients were not followed outside of RCT, within 2 months
"Breslow N"	2004	Eligible patients were not followed outside of RCT, within 2 months
"Brewin T"	1996	No RCT in this study
"Brignole M"	1991	No RCT in this study
"Brinkhaus B"	2007	Eligible patients were not followed outside of RCT, within 2 months
"Brinkhaus B"	2006	Eligible patients were not followed outside of RCT, within 2 months
"Brocklehurst P"	1998	Eligible patients were not followed outside of RCT, within 2 months
"Brookes S"	2003	Eligible patients were not followed outside of RCT, within 2 months
"Brown B"	1997	Eligible patients were not followed outside of RCT, within 2 months
"Bryce R"	1991	Eligible patients were not followed outside of RCT, within 2 months
"Buist D"	2009	Eligible patients were not followed outside of RCT, within 2 months
"Buist D"	2004	Eligible patients were not followed outside of RCT, within 2 months
"Bulman A"	2009	No RCT in this study
"Burgers J"	2002	Only 50-25% of non-participants were eligible for the trial
"Busk M"	2008	Eligible patients were not followed outside of RCT, within 2 months
"Byrne C"	2005	Eligible patients were not followed outside of RCT, within 2 months

"Byrne W"	1996	No RCT in this study
"Califano L"	2003	Eligible patients were not followed outside of RCT, within 2 months
"Campo R"	2010	Eligible patients were not followed outside of RCT, within 2 months
"Canak V"	2006	Eligible patients were not followed outside of RCT, within 2 months
"Canfield R"	1977	Eligible patients were not followed outside of RCT, within 2 months
"Capell H"	2004	Eligible patients were not followed outside of RCT, within 2 months
"Cappellini M"	2007	Eligible patients were not followed outside of RCT, within 2 months
"Carroll R"	1989	No RCT in this study
"Caruzzo C"	1995	Eligible patients were not followed outside of RCT, within 2 months
"Cascinu S"	1999	Eligible patients were not followed outside of RCT, within 2 months
"Catalan J"	1991	Eligible patients were not followed outside of RCT, within 2 months
"Cervantes F"	2010	Eligible patients were not followed outside of RCT, within 2 months
"Chabot J"	2010	No RCT in this study
"Chalmers T"	1983	No RCT in this study
"Chambless D"	1990	Eligible patients were not followed outside of RCT, within 2 months
"Chan F"	2007	Eligible patients were not followed outside of RCT, within 2 months
"Charoenwat S"	1996	Eligible patients were not followed outside of RCT, within 2 months
"Chavannes N"	2009	Eligible patients were not followed outside of RCT, within 2 months
"Chazot C"	2009	Trial patients had to be given subcutaneous rHuEPO while cohort patients were given intravenous rHuEPO
"Chemtob C"	2002	Eligible patients were not followed outside of RCT, within 2 months
"Chen C"	2000	Only 43% of the cohort patients were eligible for the trial
"Chi B"	2007	Eligible patients were not followed outside of RCT, within 2 months
"Chi I"	2006	Eligible patients were not followed outside of RCT, within 2 months
"Chopra K"	1996	Eligible patients were not followed outside of RCT, within 2 months
"Choy E"	2008	Eligible patients were not followed outside of RCT, within 2 months
"Christianson J"	1992	Eligible patients were not followed outside of RCT, within 2 months
"Chutuape M"	2001	Eligible patients were not followed outside of RCT, within 2 months
"Clark N"	2008	Eligible patients were not followed outside of RCT, within 2 months
"Cloos F"	1977	Eligible patients were not followed outside of RCT, within 2 months
"Cohen C"	1983	Eligible patients were not followed outside of RCT, within 2 months
"Cok K"	2004	No RCT in this study
"Collaborative Ocular Melanoma Study Group"	2003	No health outcome evaluated
"Collaborative Ocular Melanoma Study Group"	1998	No health outcome evaluated
"Collinge J"	2009	No RCT in this study
"Condelli W"	2000	Eligible patients were not followed outside of RCT, within 2 months
"Cooper G"	2005	Eligible patients were not followed outside of RCT, within 2 months

"Cooper G"	2005	Eligible patients were not followed outside of RCT, within 2 months
"Coppin R"	2008	No health outcome evaluated
"Corey-Lisle P"	2004	Eligible patients were not followed outside of RCT, within 2 months
"Corwin P"	2005	Eligible patients were not followed outside of RCT, within 2 months
"Coward D"	2003	Exclude because the RCT had too few patients (4 after drop-outs)
"Crist W"	2001	Eligible patients were not followed outside of RCT, within 2 months
"Croxtall J"	2009	No RCT in this study
"Cunnigham A"	1989	Eligible patients were not followed outside of RCT, within 2 months
"Cutland C"	2009	Eligible patients were not followed outside of RCT, within 2 months
"D'Angelo R"	2005	Eligible patients were not followed outside of RCT, within 2 months
"Dahl- Jorgensen, K"	1992	Eligible patients were not followed outside of RCT, within 2 months
"Dalgard O"	2010	Eligible patients were not followed outside of RCT, within 2 months
"Danaher B"	2009	Eligible patients were not followed outside of RCT, within 2 months
"Dapp U"	2007	The article was not in English
"Davidson K"	2010	Eligible patients were not followed outside of RCT, within 2 months
"Davies L"	2000	No health outcome evaluated
"Davis S"	1985	Eligible patients were not followed outside of RCT, within 2 months
"Dawe R"	2002	No RCT in this study
"de Gara C"	1987	No RCT in this study
"de Jong Y"	2007	Eligible patients were not followed outside of RCT, within 2 months
"de Jong Z"	2004	No health outcome evaluated
"De Moerloose B"	2010	Eligible patients were not followed outside of RCT, within 2 months
"Dearnaley D"	2009	Eligible patients were not followed outside of RCT, within 2 months
"Dehghani S"	2009	Eligible patients were not followed outside of RCT, within 2 months
"Dellagrammaticas D"	2008	No RCT in this study
"Delmas P"	2008	Eligible patients were not followed outside of RCT, within 2 months
"Desbiens N"	2000	Eligible patients were not followed outside of RCT, within 2 months
"Detiege J"	1987	Eligible patients were not followed outside of RCT, within 2 months
"Di Mario C"	2008	Eligible patients were not followed outside of RCT, within 2 months
"DiMeglio L"	2004	Eligible patients were not followed outside of RCT, within 2 months
"Dobbin A"	2009	The endpoints were not similar inside and outside the trial
"Dodd J"	2007	No RCT in this study
"Dodd J"	2006	Eligible patients were not followed outside of RCT, within 2 months
"Dole V"	1969	No RCT in this study
"Dotzenrath C"	2005	Eligible patients were not followed outside of RCT, within 2 months
"Douglas H"	2005	No health outcome evaluated
"Dunn G"	2005	No RCT in this study
"Dwyer P"	2008	Eligible patients were not followed outside of RCT, within 2 months
"Ebers G"	2009	Eligible patients were not followed outside of RCT, within 2 months

"Edgar L"	2003	Eligible patients were not followed outside of RCT, within 2 months
"Edwards W"	1996	Eligible patients were not followed outside of RCT, within 2 months
"Ekman I"	2001	Eligible patients were not followed outside of RCT, within 2 months
"Elkjaer M"	2010	Eligible patients were not followed outside of RCT, within 2 months
"Ell K"	2008	Eligible patients were not followed outside of RCT, within 2 months
"Elliott"	1996	Trial patients were treated in a cluster randomized trial
"Elzi L"	2005	No RCT in this study
"Emkey R"	2005	Eligible patients were not followed outside of RCT, within 2 months
"Emslie G"	1998	No RCT in this study
"Enlund M"	2001	No RCT in this study
"Eriksson K"	1998	Eligible patients were not followed outside of RCT, within 2 months
"Erkan D"	2007	Eligible patients were not followed outside of RCT, within 2 months
"Espie C"	2007	Eligible patients were not followed outside of RCT, within 2 months
"Evers A"	2002	Eligible patients were not followed outside of RCT, within 2 months
"Facchinetti F"	2005	Eligible patients were not followed outside of RCT, within 2 months
"Facon T"	2006	Eligible patients were not followed outside of RCT, within 2 months
"Fair W"	1997	Eligible patients were not followed outside of RCT, within 2 months
"Falk R"	1983	No RCT in this study
"Fallowfield L"	1990	No RCT in this study
"Farrow J"	1999	Eligible patients were not followed outside of RCT, within 2 months
"Ferguson M"	2002	No RCT in this study
"Filardo G"	2008	No RCT in this study
"Fineberg N"	1992	No RCT in this study
"Fisher B"	1989	Eligible patients were not followed outside of RCT, within 2 months
"Fitzmaurice D"	2007	Eligible patients were not followed outside of RCT, within 2 months
"Floyd A"	2010	No health outcome evaluated
"Floyd A"	2010	No health outcome evaluated
"Forssel C"	1989	Only 36% of the cohort patients were eligible for the trial
"Fouladi M"	2003	Eligible patients were not followed outside of RCT, within 2 months
"Frasure-Smith N"	1987	Eligible patients were not followed outside of RCT, within 2 months
"Fries E"	1997	Eligible patients were not followed outside of RCT, within 2 months
"Frucht-Pery J"	2006	Only 63% of the cohort patients were eligible for the trial
"Galbrecht C"	1968	No RCT in this study
"Galeone M"	1992	Eligible patients were not followed outside of RCT, within 2 months
"Gallo C"	1995	No health outcome evaluated
"Gardin C"	2007	Eligible patients were not followed outside of RCT, within 2 months
"Garland W"	2007	No RCT in this study
"Geerts A"	2005	Eligible patients were not followed outside of RCT, within 2 months
"Geisler F"	2008	Eligible patients were not followed outside of RCT, within 2 months

"Giannetti A"	1984	No RCT in this study
"Giannini E"	2010	Eligible patients were not followed outside of RCT, within 2 months
"Gilsbach J"	1990	Eligible patients were not followed outside of RCT, within 2 months
"Giordano P"	2006	Eligible patients were not followed outside of RCT, within 2 months
"Giorlandino C"	2009	Eligible patients were not followed outside of RCT, within 2 months
"Glasgow R"	2009	No health outcome evaluated
"Glaspy J"	2009	Eligible patients were not followed outside of RCT, within 2 months
"Glaspy J"	2009	Eligible patients were not followed outside of RCT, within 2 months
"Glenn J"	1985	Eligible patients were not followed outside of RCT, within 2 months
"Goff J"	1986	Patients at the centers who were not stable enough to be moved to the University Hospital for laser treatment, formed the cohort group
"Gonwa T"	2002	Eligible patients were not followed outside of RCT, within 2 months
"Gordon P"	2008	Eligible patients were not followed outside of RCT, within 2 months
"Goss C"	2006	Patients not enrolled in the trial were taken from a population based registry
"Gottlieb A"	1984	Eligible patients were not followed outside of RCT, within 2 months
"Gregory R"	2000	Eligible patients were not followed outside of RCT, within 2 months
"Gridelli C"	2004	Eligible patients were not followed outside of RCT, within 2 months
"Griem K"	1987	No health outcome evaluated
"Groff A"	2004	No health outcome evaluated
"Grossarth-Maticek R"	2001	Eligible patients were not followed outside of RCT, within 2 months
"Grossarth-Maticek R"	2006	Eligible patients were not followed outside of RCT, within 2 months
"Grossarth-Maticek R"	2007	No RCT in this study
"Grunfeld E"	2006	Eligible patients were not followed outside of RCT, within 2 months
"Guan Z"	2010	Eligible patients were not followed outside of RCT, within 2 months
"Guilleminault C"	2002	Eligible patients were not followed outside of RCT, within 2 months
"Guyer R"	2008	Eligible patients were not followed outside of RCT, within 2 months
"Haberkern C"	1997	Not all patients in the cohort group were eligible (unclear the proportion)
"Hack T"	2007	Eligible patients were not followed outside of RCT, within 2 months
"Hadji P"	2008	No RCT in this study
"Haldeman S"	2010	No RCT in this study
"Halfvarson J"	2009	No RCT in this study
"Handelzalts J"	2010	No RCT in this study
"Handoll H"	2009	No RCT in this study
"Hare S"	1983	Eligible patients were not followed outside of RCT, within 2 months
"Harris O"	2009	Eligible patients were not followed outside of RCT, within 2 months
"Harrison J"	2003	Eligible patients were not followed outside of RCT, within 2 months
"Harvey I"	1989	No RCT in this study
"Hatlebakk J"	1998	Eligible patients were not followed outside of RCT, within 2 months

"Heaney R"	2009	Eligible patients were not followed outside of RCT, within 2 months
"Hegerl U"	2010	Eligible patients were not followed outside of RCT, within 2 months
"Helling T"	1992	Eligible patients were not followed outside of RCT, within 2 months
"Henshaw R"	1993	Same population as in Howie 1997
"Henshaw R"	1994	Same population as in Howie 1997
"Henshaw R"	1994	Same population as in Henshaw 1993
"Hernandez M"	2010	Eligible patients were not followed outside of RCT, within 2 months
"Hickson D"	1990	Eligible patients were not followed outside of RCT, within 2 months
"Hiday V"	2002	No health outcome evaluated
"Hlatky M"	1997	Eligible patients were not followed outside of RCT, within 2 months
"Hochman J"	2000	Eligible patients were not followed outside of RCT, within 2 months
"Hochman J"	1999	No RCT in this study
"Hofvind S"	2008	No health outcome evaluated
"Holliday M"	1999	Eligible patients were not followed outside of RCT, within 2 months
"Hollman G"	1998	Eligible patients were not followed outside of RCT, within 2 months
"Holm T"	2002	Eligible patients were not followed outside of RCT, within 2 months
"Honeycutt T"	2007	Eligible patients were not followed outside of RCT, within 2 months
"Hoogeboom T"	2010	Eligible patients were not followed outside of RCT, within 2 months
"Hoste E"	2005	Eligible patients were not followed outside of RCT, within 2 months
"Howard L"	2009	No health outcome evaluated
"Hreinsson J"	2003	Eligible patients were not followed outside of RCT, within 2 months
"Hsu D"	2009	No health outcome evaluated
"Hu C"	2006	Eligible patients were not followed outside of RCT, within 2 months
"Hudmon K"	1997	No RCT in this study
"Huf G"	2010	Conference proceeding/poster
"Hughes S"	2009	Eligible patients were not followed outside of RCT, within 2 months
"Hulse R"	2010	Eligible patients were not followed outside of RCT, within 2 months
"Hutton N"	1991	Eligible patients were not followed outside of RCT, within 2 months
"Hybbinette C"	1981	No RCT in this study
"Ihde D"	1994	Eligible patients were not followed outside of RCT, within 2 months
"Ikonomidis I"	2008	Eligible patients were not followed outside of RCT, within 2 months
"Ilankovan V"	1992	Eligible patients were not followed outside of RCT, within 2 months
"Innes G"	1997	Eligible patients were not followed outside of RCT, within 2 months
"Isler C"	2003	Eligible patients were not followed outside of RCT, within 2 months
"Ivancic M"	1996	Eligible patients were not followed outside of RCT, within 2 months
"Jackson H"	2005	Eligible patients were not followed outside of RCT, within 2 months
"Jacquillat C"	1980	Eligible patients were not followed outside of RCT, within 2 months
"Janson M"	2009	No health outcome evaluated
"Jantausch B"	2003	Eligible patients were not followed outside of RCT, within 2 months

"Jehn U"	1990	Eligible patients were not followed outside of RCT, within 2 months
"Jeremic B"	1999	Eligible patients were not followed outside of RCT, within 2 months
"Jerkeman M"	1999	The same construct was not measured inside and outside the trial
"Johnson C"	2010	No RCT in this study
"Johnson P"	2010	Eligible patients were not followed outside of RCT, within 2 months
"Johnson R"	2007	Eligible patients were not followed outside of RCT, within 2 months
"Kallmes D"	2009	The endpoints were not similar inside and outside the trial
"Kanlayanaphotporn R"	2009	Eligible patients were not followed outside of RCT, within 2 months
"Karlsson L"	2009	Eligible patients were not followed outside of RCT, within 2 months
"Karounis H"	2004	Eligible patients were not followed outside of RCT, within 2 months
"Karp D"	2009	Eligible patients were not followed outside of RCT, within 2 months
"Katsogridakis Y"	2008	Eligible patients were not followed outside of RCT, within 2 months
"Kaul N"	2007	Eligible patients were not followed outside of RCT, within 2 months
"Keilholz U"	2005	Eligible patients were not followed outside of RCT, within 2 months
"Kennedy T"	1969	Conference proceeding/poster
"Kerwin R"	2007	Eligible patients were not followed outside of RCT, within 2 months
"Keus F"	2008	Eligible patients were not followed outside of RCT, within 2 months
"Khan M"	2007	Eligible patients were not followed outside of RCT, within 2 months
"Khoo S"	2006	Eligible patients were not followed outside of RCT, within 2 months
"Kim G"	2009	Eligible patients were not followed outside of RCT, within 2 months
"Kim S"	1997	Eligible patients were not followed outside of RCT, within 2 months
"King III"	1997	Same patient population as King 2000
"Kitchener H"	2004	Eligible patients were not followed outside of RCT, within 2 months
"Klaber Moffett J"	2005	Eligible patients were not followed outside of RCT, within 2 months
"Klarlund M"	1997	Eligible patients were not followed outside of RCT, within 2 months
"Klosky J"	2009	No health outcome evaluated
"Koek M"	2009	Eligible patients were not followed outside of RCT, within 2 months
"Koek M"	2010	Eligible patients were not followed outside of RCT, within 2 months
"Koek M"	2009	The endpoints were not similar inside and outside the trial
"Konstantinidou E"	2002	Eligible patients were not followed outside of RCT, within 2 months
"Koo C"	2008	Eligible patients were not followed outside of RCT, within 2 months
"Koopmans C"	2009	Eligible patients were not followed outside of RCT, within 2 months
"Koopmans C"	2009	Eligible patients were not followed outside of RCT, within 2 months
"Korn E"	2010	No RCT in this study
"Korvick J"	1992	Eligible patients were not followed outside of RCT, within 2 months
"Kotsar A"	2006	No RCT in this study
"Krapf H"	2005	Eligible patients were not followed outside of RCT, within 2 months
"Kuhn L"	2008	Eligible patients were not followed outside of RCT, within 2 months
"Kurland A"	1966	Eligible patients were not followed outside of RCT, within 2 months

"Kushner S"	2009	Eligible patients were not followed outside of RCT, within 2 months
"Kushner S"	2009	Eligible patients were not followed outside of RCT, within 2 months
"Laatikainen L"	1987	Eligible patients were not followed outside of RCT, within 2 months
"Laigle-Donadey F"	2010	Eligible patients were not followed outside of RCT, within 2 months
"Lainez M"	2006	Eligible patients were not followed outside of RCT, within 2 months
"Lakerveld J"	2008	No RCT in this study
"Landis S"	1992	Eligible patients were not followed outside of RCT, within 2 months
"Lang P"	1997	Eligible patients were not followed outside of RCT, within 2 months
"Larsson P"	2006	Eligible patients were not followed outside of RCT, within 2 months
"Lasekan J"	1999	Eligible patients were not followed outside of RCT, within 2 months
"Lasser E"	1987	Eligible patients were not followed outside of RCT, within 2 months
"Laurila M"	2009	Patients outside the trial were chosen from a population based registry
"Lautenschlager N"	2008	Eligible patients were not followed outside of RCT, within 2 months
"Lazcano Ponce E"	2000	Eligible patients were not followed outside of RCT, within 2 months
"Lee K"	2005	Eligible patients were not followed outside of RCT, within 2 months
"Leese M"	2005	No RCT in this study
"Leeton J"	1985	Eligible patients were not followed outside of RCT, within 2 months
"Lefevre T"	1999	Eligible patients were not followed outside of RCT, within 2 months
"Leon A"	2005	Eligible patients were not followed outside of RCT, within 2 months
"Lerang F"	1997	Eligible patients were not followed outside of RCT, within 2 months
"Leroux-Roels I"	2010	Eligible patients were not followed outside of RCT, within 2 months
"Levine M"	1997	Eligible patients were not followed outside of RCT, within 2 months
"Levy H"	2005	Eligible patients were not followed outside of RCT, within 2 months
"Lewis B"	2005	Eligible patients were not followed outside of RCT, within 2 months
"Lichtiger S"	2009	No RCT in this study
"Lin C"	2009	Eligible patients were not followed outside of RCT, within 2 months
"Lin P"	2005	No RCT in this study
"Lipkovich I"	2008	Eligible patients were not followed outside of RCT, within 2 months
"Lloyd-Williams F"	2003	The endpoints were not similar inside and outside the trial
"Loeffler M"	1997	Eligible patients were not followed outside of RCT, within 2 months
"Logemann J"	2008	No RCT in this study
"Lubaki L"	2010	No RCT in this study
"Lurie J"	2008	No health outcome evaluated
"Lustig R"	1976	Eligible patients were not followed outside of RCT, within 2 months
"Lusuardi M"	2006	Eligible patients were not followed outside of RCT, within 2 months
"Mabeck C"	1979	Eligible patients were not followed outside of RCT, within 2 months
"Macrae D"	2010	No RCT in this study
"Madersbacher S"	2004	Eligible patients were not followed outside of RCT, within 2 months
"Magnussen L"	2009	Eligible patients were not followed outside of RCT, within 2 months

"Mahaffey K"	2005	Eligible patients were not followed outside of RCT, within 2 months
"Malmi H"	2010	No health outcome evaluated
"Manolis A"	1998	Eligible patients were not followed outside of RCT, within 2 months
"Mant J"	2008	Eligible patients were not followed outside of RCT, within 2 months
"Manuel De La Fuente J"	1994	Eligible patients were not followed outside of RCT, within 2 months
"Marcolongo R"	1985	Eligible patients were not followed outside of RCT, within 2 months
"Marks I"	1985	Eligible patients were not followed outside of RCT, within 2 months
"Marre M"	2000	The endpoints were not similar inside and outside the trial
"Marsa-Vila L"	1991	Eligible patients were not followed outside of RCT, within 2 months
"Marubini E"	1996	Eligible patients were not followed outside of RCT, within 2 months
"Marucci M"	2003	Eligible patients were not followed outside of RCT, within 2 months
"Mather H"	1976	No health outcome evaluated
"Mathew J"	2009	No RCT in this study
"Matilainen T"	1996	Eligible patients were not followed outside of RCT, within 2 months
"Mauer M"	2002	No health outcome evaluated
"McAfee P"	2006	Eligible patients were not followed outside of RCT, within 2 months
"McAvoy B"	1991	Eligible patients were not followed outside of RCT, within 2 months
"McCahon D"	2007	No RCT in this study
"McClung M"	2007	Eligible patients were not followed outside of RCT, within 2 months
"McElroy S"	2008	Eligible patients were not followed outside of RCT, within 2 months
"McFarlane A"	2001	Eligible patients were not followed outside of RCT, within 2 months
"McPherson K"	2008	No RCT in this study
"McPherson K"	1999	No RCT in this study
"Medical Research Council Multicentre Otitis Media Study Group"	2001	Eligible patients were not followed outside of RCT, within 2 months
"Medical Research Council's Working Party for Therapeutic Trials in Leukaemia"	1983	Eligible patients were not followed outside of RCT, within 2 months
"Meier P"	1985	Eligible patients were not followed outside of RCT, within 2 months
"Meier P"	1989	Eligible patients were not followed outside of RCT, within 2 months
"Melchionda N"	2006	No health outcome evaluated
"Mergl R"	2011	Eligible patients were not followed outside of RCT, within 2 months
"Michaels J"	2009	Eligible patients were not followed outside of RCT, within 2 months
"Michaux M"	1966	Eligible patients were not followed outside of RCT, within 2 months
"Miller L"	2008	No RCT in this study
"Miner M"	2008	No health outcome evaluated
"Miriam A"	2004	Eligible patients were not followed outside of RCT, within 2 months

"Mirone V"	2007	Eligible patients were not followed outside of RCT, within 2 months
"Moehring H"	2000	Eligible patients were not followed outside of RCT, within 2 months
"Moergel M"	2009	No RCT in this study
"Moergel M"	2009	No RCT in this study
"Molkenboer J"	2004	Eligible patients were not followed outside of RCT, within 2 months
"Mootsikapun P"	2005	Eligible patients were not followed outside of RCT, within 2 months
"Moran S"	2003	Eligible patients were not followed outside of RCT, within 2 months
"Moro E"	2005	Eligible patients were not followed outside of RCT, within 2 months
"Murphy D"	2009	Eligible patients were not followed outside of RCT, within 2 months
"Mwengee W"	2006	Eligible patients were not followed outside of RCT, within 2 months
"Myers S"	2007	Eligible patients were not followed outside of RCT, within 2 months
"Nadstawek"	2008	Eligible patients were not followed outside of RCT, within 2 months
"Nakache R"	2005	Eligible patients were not followed outside of RCT, within 2 months
"Narayan K"	1998	The endpoints were not similar inside and outside the trial
"Nashan B"	1996	Eligible patients were not followed outside of RCT, within 2 months
"Naslund G"	1994	No health outcome evaluated
"Naukkarinen V"	1989	Eligible patients were not followed outside of RCT, within 2 months
"Naylor P"	2008	Eligible patients were not followed outside of RCT, within 2 months
"Negrier S"	2007	Eligible patients were not followed outside of RCT, within 2 months
"Neutel J"	2006	Eligible patients were not followed outside of RCT, within 2 months
"Newman N"	2002	Eligible patients were not followed outside of RCT, within 2 months
"Niccols A"	2008	Eligible patients were not followed outside of RCT, within 2 months
"Nikkila E"	1984	Eligible patients were not followed outside of RCT, within 2 months
"Nio Y"	2004	Eligible patients were not followed outside of RCT, within 2 months
"Ohman E"	1996	Eligible patients were not followed outside of RCT, within 2 months
"Ohman J"	1989	Eligible patients were not followed outside of RCT, within 2 months
"Ojehagen A"	1992	Eligible patients were not followed outside of RCT, within 2 months
"Omata M"	2007	Eligible patients were not followed outside of RCT, within 2 months
"Owen J"	2009	Conference proceeding/poster
"Pajno G"	2003	Eligible patients were not followed outside of RCT, within 2 months
"Pak C"	1996	Eligible patients were not followed outside of RCT, within 2 months
"Pakkala S"	1992	Eligible patients were not followed outside of RCT, within 2 months
"Papageorgiou A"	2004	Eligible patients were not followed outside of RCT, within 2 months
"Papaldo P"	2005	Eligible patients were not followed outside of RCT, within 2 months
"Parker J"	2003	Eligible patients were not followed outside of RCT, within 2 months
"Parkinson's Disease Research Group in the United Kingdom"	1993	Eligible patients were not followed outside of RCT, within 2 months
"Parsons J"	1980	No RCT in this study

"Paterson C"	2003	Eligible patients were not followed outside of RCT, within 2 months
"Pendergast J"	1995	Eligible patients were not followed outside of RCT, within 2 months
"Petersen K"	1991	Eligible patients were not followed outside of RCT, within 2 months
"Peveler R"	2005	Eligible patients were not followed outside of RCT, within 2 months
"Phillips M"	1975	Eligible patients were not followed outside of RCT, within 2 months
"Phimda K"	2007	Eligible patients were not followed outside of RCT, within 2 months
"Pineda O"	2001	No RCT in this study
"Pizer B"	2006	Eligible patients were not followed outside of RCT, within 2 months
"Pocock S"	1997	Eligible patients were not followed outside of RCT, within 2 months
"Pollock J"	2000	Eligible patients were not followed outside of RCT, within 2 months
"Potter D"	1986	Only 28% of the cohort patients were eligible for the trial
"Price R"	1991	Eligible patients were not followed outside of RCT, within 2 months
"Quarmby L"	2007	Eligible patients were not followed outside of RCT, within 2 months
"Quilty L"	2008	Eligible patients were not followed outside of RCT, within 2 months
"Quinn C"	2009	No RCT in this study
"Ragab S"	2005	Eligible patients were not followed outside of RCT, within 2 months
"Rajchanuvong A"	1995	Eligible patients were not followed outside of RCT, within 2 months
"Rasmussen B"	2008	No RCT in this study
"Ravindranath Y"	1996	Eligible patients were not followed outside of RCT, within 2 months
"Raymond J"	1996	Eligible patients were not followed outside of RCT, within 2 months
"Reed K"	2008	Eligible patients were not followed outside of RCT, within 2 months
"Reed N"	2008	Eligible patients were not followed outside of RCT, within 2 months
"Reed S"	2009	No health outcome evaluated
"Regan J"	2006	Eligible patients were not followed outside of RCT, within 2 months
"Reinders M"	2008	Eligible patients were not followed outside of RCT, within 2 months
"Reinhart K"	2001	Eligible patients were not followed outside of RCT, within 2 months
"Resnick E"	2009	Eligible patients were not followed outside of RCT, within 2 months
"Rey E"	1999	No RCT in this study
"Reynolds K"	1997	No RCT in this study
"Rhomberg W"	2007	Eligible patients were not followed outside of RCT, within 2 months
"Ricco J"	2010	No RCT in this study
"Rokito S"	1995	The endpoints were not similar inside and outside the trial
"Roncucci L"	1993	Eligible patients were not followed outside of RCT, within 2 months
"Ros A"	2001	Eligible patients were not followed outside of RCT, within 2 months
"Rosendahl J"	2009	No health outcome evaluated
"Rosser W"	1992	No RCT in this study
"Roter D"	1977	No health outcome evaluated
"Röther J"	2002	No RCT in this study
"Rush J"	1990	Eligible patients were not followed outside of RCT, within 2 months

"Ryan M"	2005	No RCT in this study
"Rychtarik R"	1998	No health outcome evaluated
"Ryden L"	2008	No RCT in this study
"Rydhstrom H"	1991	Eligible patients were not followed outside of RCT, within 2 months
"S Slipp"	1978	Eligible patients were not followed outside of RCT, within 2 months
"Salisbury C"	2005	Eligible patients were not followed outside of RCT, within 2 months
"Salmon S"	1990	Eligible patients were not followed outside of RCT, within 2 months
"Savani N"	2001	Eligible patients were not followed outside of RCT, within 2 months
"Sbarbaro J"	1979	Eligible patients were not followed outside of RCT, within 2 months
"Schaar C"	2005	Eligible patients were not followed outside of RCT, within 2 months
"Scherer R"	2008	Eligible patients were not followed outside of RCT, within 2 months
"Schneider L"	2006	Eligible patients were not followed outside of RCT, within 2 months
"Schoot R"	2010	Eligible patients were not followed outside of RCT, within 2 months
"Schouten H"	2003	Eligible patients were not followed outside of RCT, within 2 months
"Schroer S"	2009	Eligible patients were not followed outside of RCT, within 2 months
"Segal R"	1999	Eligible patients were not followed outside of RCT, within 2 months
"Senoglu N"	2010	Eligible patients were not followed outside of RCT, within 2 months
"Senore C"	2010	Eligible patients were not followed outside of RCT, within 2 months
"Serruys P"	2010	Eligible patients were not followed outside of RCT, within 2 months
"Shah V"	2008	Eligible patients were not followed outside of RCT, within 2 months
"Shenfine J"	2005	Eligible patients were not followed outside of RCT, within 2 months
"Shiratori Y"	1999	Eligible patients were not followed outside of RCT, within 2 months
"Sika M"	2007	Eligible patients were not followed outside of RCT, within 2 months
"Siminoff L"	2006	Eligible patients were not followed outside of RCT, within 2 months
"Singh B"	1995	Eligible patients were not followed outside of RCT, within 2 months
"Singhal A"	2003	Eligible patients were not followed outside of RCT, within 2 months
"Smidt-Jensen S"	1991	Eligible patients were not followed outside of RCT, within 2 months
"Smith M"	2009	Eligible patients were not followed outside of RCT, within 2 months
"Soghikian K"	1992	Eligible patients were not followed outside of RCT, within 2 months
"Sorensen J"	1992	Eligible patients were not followed outside of RCT, within 2 months
"South East Asian Quinine Artesunate Malaria Trial (SEAQUAMAT) group"	2005	Eligible patients were not followed outside of RCT, within 2 months
"Stabile G"	2001	Eligible patients were not followed outside of RCT, within 2 months
"Stansfield S"	1984	No health outcome evaluated
"Stockle M"	1995	Eligible patients were not followed outside of RCT, within 2 months
"Strudler Wallston B"	1987	Eligible patients were not followed outside of RCT, within 2 months
"Sturmer T"	1998	Eligible patients were not followed outside of RCT, within 2 months
"Sudilovsky A"	1981	Eligible patients were not followed outside of RCT, within 2 months

"Sweetenham J"	2001	Eligible patients were not followed outside of RCT, within 2 months
"Szanto E"	1986	Eligible patients were not followed outside of RCT, within 2 months
"Taylor D"	2008	Eligible patients were not followed outside of RCT, within 2 months
"Tew J"	2007	Eligible patients were not followed outside of RCT, within 2 months
"The principal investigators of CASS and their associates "	1981	Eligible patients were not followed outside of RCT, within 2 months
"The Support Principal Investigators"	1995	Eligible patients were not followed outside of RCT, within 2 months
"Thiboutot D"	2009	Eligible patients were not followed outside of RCT, within 2 months
"Thompson J"	1982	Eligible patients were not followed outside of RCT, within 2 months
"Thorburn Bird S"	2001	Eligible patients were not followed outside of RCT, within 2 months
"Tonstad S"	2006	Eligible patients were not followed outside of RCT, within 2 months
"Toplak H"	2005	Eligible patients were not followed outside of RCT, within 2 months
"Torgerson D"	1996	Eligible patients were not followed outside of RCT, within 2 months
"Torti C"	2005	Eligible patients were not followed outside of RCT, within 2 months
"Tschaikowsky K"	2000	Eligible patients were not followed outside of RCT, within 2 months
"Ungerleider J"	1982	Eligible patients were not followed outside of RCT, within 2 months
"van Balen F"	1996	No health outcome evaluated
"Van de Wiel N"	2003	Eligible patients were not followed outside of RCT, within 2 months
"van den Berg-Wolf, M"	2008	Eligible patients were not followed outside of RCT, within 2 months
"van Meerbeeck J"	2007	Eligible patients were not followed outside of RCT, within 2 months
"van Weert E"	2005	Eligible patients were not followed outside of RCT, within 2 months
"Van Zanten S"	2007	Eligible patients were not followed outside of RCT, within 2 months
"Vass M"	2005	Eligible patients were not followed outside of RCT, within 2 months
"Vassilopoulou-Sellin R"	1999	No health outcome evaluated
"Veenhof C"	2006	Eligible patients were not followed outside of RCT, within 2 months
"Vela L"	2006	No RCT in this study
"Velasquez M"	2000	No health outcome evaluated
"Vickers A"	2009	Eligible patients were not followed outside of RCT, within 2 months
"Vuorma S"	2003	Eligible patients were not followed outside of RCT, within 2 months
"Wagner H"	1994	Eligible patients were not followed outside of RCT, within 2 months
"Wagner K"	2003	No RCT in this study
"Wallace P"	2002	Eligible patients were not followed outside of RCT, within 2 months
"Wallberg B"	2009	No health outcome evaluated
"Walsh D"	1991	Eligible patients were not followed outside of RCT, within 2 months
"Wang F"	2003	Eligible patients were not followed outside of RCT, within 2 months
"Ward E"	2000	Same patient population as King 2000
"Warshaw E"	2005	Eligible patients were not followed outside of RCT, within 2 months

"Webster-Stratton C"	1990	Eligible patients were not followed outside of RCT, within 2 months
"Weijmar Schultz W"	1996	Eligible patients were not followed outside of RCT, within 2 months
"Weisdorf D"	1997	Eligible patients were not followed outside of RCT, within 2 months
"Wells K"	2000	Eligible patients were not followed outside of RCT, within 2 months
"Wermeling P"	2010	No RCT in this study
"Wetzig N"	2005	Eligible patients were not followed outside of RCT, within 2 months
"Wharton T"	2004	Eligible patients were not followed outside of RCT, within 2 months
"Whegang S"	2010	Eligible patients were not followed outside of RCT, within 2 months
"Whitehurst D"	2007	Eligible patients were not followed outside of RCT, within 2 months
"Wiart L"	1997	Eligible patients were not followed outside of RCT, within 2 months
"Wilfred Germino F"	2009	Eligible patients were not followed outside of RCT, within 2 months
"Wilhelmsen L"	1986	Eligible patients were not followed outside of RCT, within 2 months
"Wilson S"	2010	Eligible patients were not followed outside of RCT, within 2 months
"Winters W"	1981	Eligible patients were not followed outside of RCT, within 2 months
"Witt C"	2005	Eligible patients were not followed outside of RCT, within 2 months
"Wolden S"	2001	No RCT in this study
"Wolter J"	2004	Eligible patients were not followed outside of RCT, within 2 months
"Wolters T"	2010	Eligible patients were not followed outside of RCT, within 2 months
"Woodcock N"	2001	Eligible patients were not followed outside of RCT, within 2 months
"Woods W"	2001	Eligible patients were not followed outside of RCT, within 2 months
"Wright P"	1988	Eligible patients were not followed outside of RCT, within 2 months
"Yangco B"	1987	Eligible patients were not followed outside of RCT, within 2 months
"Yealy D"	2005	Eligible patients were not followed outside of RCT, within 2 months
"Yoon J"	2009	Eligible patients were not followed outside of RCT, within 2 months
"Yuasa H"	2003	Eligible patients were not followed outside of RCT, within 2 months
"Yuen Loke A"	2005	Eligible patients were not followed outside of RCT, within 2 months
"Zlatnik F"	1993	Eligible patients were not followed outside of RCT, within 2 months

Table 2: List of the 554 excluded articles as well as the reason for their exclusions.

Author	Year
"Abraham N"	2004
"Alberto P"	1976
"Alvarez R"	2003
"Assmus B"	2002
"Atkins M"	1993
"Baker M"	1982
"Banach M"	2000
"Battin M"	2007
"Bausewein C"	2010
"Berglund G"	1997
"Berkeley A"	1985
"Birch E"	1992
"Black S"	1993
"Brown A"	2001
"Buijk S"	2002
"Burette A"	1992
"Chadwick D"	1991
"Comerota A"	2004
"Cooper J"	1999
"Corvo R"	2000
"Cross P"	2005
"Deuschle M"	2004
"Diaz E"	2004
"Dobbin A"	2009
"Ejlertsen B"	2008
"Englund J"	2005
"Exner D"	1999
"Fisher J"	2000
"Francis B"	2004
"Gowers S"	2010
"Grant A"	2008
"Grunfeld E"	1995
"Haan J"	1991
"Halbert C"	2010
"Herman R"	1992
"Hertegård S"	2002
"Holubkov R"	1999
"Jha P"	1996
"Jing-hong Z"	1990

"Jirmar R"	2008
"Kahan B"	2008
"Kamal S"	2006
"Kaufmann C"	1994
"Kleinschmidt S"	1999
"Krysztopik R"	2002
"Lawson P"	1984
"Link M"	1986
"Lundell L"	1998
"Majumdar A"	2010
"Mansergh G"	2010
"Masood J"	2002
"Mayberg M"	1991
"McGhee S"	1994
"Menon V"	2008
"Mori A"	2008
"Mosekilde L"	2000
"Moynihan C"	1998
"National Emphysema Treatment Trial Research Group"	2004
"Neill M"	1991
"Neudorf S"	2004
"O'Brien C"	1989
"Olbers T"	2003
"Oude Elberink J"	2006
"Oude Elberink J"	2009
"Papadopoulos E"	2006
"Paradise L"	1990
"Peterson A"	2006
"Playforth M"	1988
"Porter M"	2005
"Prescott R"	2007
"Quigley R"	1995
"Reeves B"	2004
"Rogers W"	1995
"Rovers M"	2001
"Schmoor C"	1996
"Spanos W"	1994
"Sperling L"	1993
"Stacey M"	1990

"Sterling R"	1997
"Stone P"	1990
"Straatsma B"	2003
"Tincello D"	2009
"Tofteng C"	2002
"Treanor J"	2010
"Underwood M"	2008
"Vass M"	2007
"Vetrhus M"	2002
"Vitiello B"	2009
"Walther B"	2003
"Weinstein J"	2006
"Whitehouse P"	2006
"Wieringa-de Waard M"	2004
"Wieringa-de Waard M"	2002
"Williams A"	1999
"Williams G"	1999
"Young J"	1996

Table 3: List of studies that require further information from authors, before they can be included in our meta- analysis.

Included Study Characteristics

Author	**Akaza 1995**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients refused to be randomized. An intention-to-treat analysis was used in the trial.
Selection Bias	Most likely there are differences in baseline characteristics between groups. N=120
Detection Bias	Unclear whether follow-up was the same between groups.
Exclusion Bias	No post-randomization exclusions.
Intervention and Co-interventions	***Trial*** Treatment: BCG prophylactic (maintenance) instillation group Control: Untreated observation group ***Cohort*** Control: Same Unclear if trial treatment was significantly different.
Care provider and setting	Unclear who is providing care, both groups treated in hospital.
Outcomes	Incidence of recurrence. Within 3 years of follow-up.

Author	**Amar 1997**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients were not entered into the trial either because they were originally scheduled for pulmonary lobectomy or because informed consent could not be obtained. An intention-to-treat analysis was used in the trial.
Selection Bias	There were no differences in baseline characteristics between groups.
Detection Bias	More frequent visits for those in RCT.
Exclusion Bias	No post-randomization exclusions.
Intervention and Co-interventions	***Trial*** Treatment 1: Prophylaxis treatment of dilitiazem Treatment 2: Prophylaxis treatment of digoxin ***Cohort*** Control: No prophylaxis given There was no significant difference between trial treatment groups.
Care provider and setting	Unclear who is providing care, both groups treated in the operating room and post-anesthesia care unit of the hospital.
Outcomes	Incidence of supraventricular dysrhythmias. Mean follow-up was 203 hours from surgery.

Author	**Andersson 2003**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort was recruited to act as a control sample. No intention-to-treat analysis used in the trial.
Selection Bias	Most likely there are differences in baseline characteristics between groups. N=52
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	There were 9 inappropriate exclusions in the treatment 1 trial, 11 in the

	treatment 2 trial, and none in the non- trial group.
Intervention and Co-interventions	***Trial*** Treatment 1: Web-based self-help program and telephone calls Treatment 2: Web-based self-help program only ***Cohort*** Treatment 2: Same There was no significant difference between trial treatment groups.
Care provider and setting	Therapist provided care when needed in the trial, unclear if there was access to a therapist in the non-trial arm. Both groups were treated in the same setting.
Outcomes	Average headache index over the duration of the trial (means of noted intensity for each day summed, divided by the total number of registration days).

Author	**Antman 1985**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients refused to be randomized. Intention-to-treat analysis was used in the trial.
Selection Bias	There were differences in baseline characteristics between the two groups.
Detection Bias	Unclear if follow-up was the same between groups.
Exclusion Bias	No post-randomization exclusions.
Intervention and Co-interventions	***Trial*** Treatment: Five cycles of adjuvant doxorubicin 90 mg/m^2 intravenously, every 3 weeks. Control: Observation alone. ***Cohort*** Treatment: Same Control: Same There was no significant difference between trial treatment groups.
Care provider and setting	Medical oncologist, surgeon, radiotherapist and pathologist all provided care in the trial, unclear who provided care outside trial. All patients were treated at the Women's Hospital and Cancer Center.
Outcomes	Number of patients who were not disease free. Within 40 months of follow-up.

Author	**Ashok 2002**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients had a strong preference for a particular treatment. Intention-to-treat analysis was used in the trial.
Selection Bias	There were no differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	There were 14 appropriate and 70 inappropriate exclusions in the trial treatment 1 group, 18 appropriate and 69 inappropriate exclusions in the trial treatment 2 group. There were 6 inappropriate exclusions in the cohort treatment 1 group, 9 appropriate and 26 inappropriate exclusions in the cohort treatment 2 group.
Intervention and Co-interventions	***Trial*** Treatment 1: Medical abortion

	Treatment 2: Surgically induced abortion ***Cohort*** Treatment 1: Same Treatment 2: Same There was a significant beneficial effect associated with the trial treatment 2.
Care provider and setting	Nurses provided care in both treatment 1 groups and surgeons in both treatment 2 groups. All patients were treated in the hospital gynecology ward.
Outcomes	Mean number of days bleeding. At 2-3 weeks from termination.

Author	**Bain 2001**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients had a strong preference for a particular treatment. Intention-to-treat analysis was used in the trial.
Selection Bias	There were no differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	No post-randomization exclusions.
Intervention and Co-interventions	***Trial*** Treatment 1: Local anaesthesia Treatment 2: General anaesthesia ***Cohort*** Treatment 1: Same Treatment 2: Same There was no significant difference between trial treatment groups.
Care provider and setting	Unclear who provided care inside and outside of the trial. All patients were treated in the same theatre suite.
Outcomes	Perceived pain post-operatively measured by the McGill Pain questionnaire. At discharge.

Author	**Bakker 2000**
Design and Methods	A randomized trial compared to a parallel-treated cohort of eligible patients. Cohort patients refused trial medication. Intention-to-treat analysis was used in the trial.
Selection Bias	There were no differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	There were 9 inappropriate exclusions in the trial treatment 1 group, 4 inappropriate exclusions in the trial treatment 2 group, 3 inappropriate exclusions in the trial treatment 3 group, 2 inappropriate exclusions in the trial control group. There were 7 inappropriate exclusions in the cohort treatment 1 group.
Intervention and Co-interventions	***Trial*** Treatment 1 : Cognitive therapy Treatment 2: Paroxetine (20-60 mg/day) Treatment 3: Clomipramine (50-150 mg/day) Control: Placebo ***Cohort*** Treatment 1: Same Unclear if trial treatment was significantly different.

Care provider and setting	A CT- trained psychologist and psychiatrist provided care in the trial treatment 1 group and psychiatrists provided care in the trial treatment 2,3, control groups. Unclear who provided care for the cohort patients but appears to be similarly trained professionals. All patients were treated at an outpatient clinic for anxiety disorders.
Outcomes	Mean panic attack frequency. Within 12 weeks follow-up.
Note	RCT information was extracted from: Bakker A et al. (1999) "Paroxetine, clomipramine, and cognitive therapy in the treatment of panic disorder." J Clin Psychiatry 60(12):831-8

Author	**Balmukhanov 1989**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Do not state why cohort patients did not participate in the trial. Intention-to-treat analysis was not used in the trial.
Selection Bias	Most likely no differences in baseline characteristics between groups. N=395
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	No post- randomization exclusions.
Intervention and Co-interventions	***Trial*** Treatment 1: Radiotherapy in combination with metronidazole Treatment 2: Radiotherapy alone ***Cohort*** Treatment 1: Same Treatment 2: Same Trail treatment 1 was significantly better than trial treatment 2.
Care provider and setting	Unclear who provided care both in the trial and cohort groups. All patients were treated at the Institute of Oncology.
Outcomes	Stage IIb tumors that did not clear. At 2 weeks.

Author	**Bannister 2001**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients were not enrolled because they were used to identify the presence of learning bias or a change in clinical practice. Intention-to-treat analysis was not used in the trial.
Selection Bias	There were no differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	No post- randomization exclusions.
Intervention and Co-interventions	***Trial*** Treatment: Bispectral index monitoring, co-intervention were a combination of medications (i.e. oral midacolam, fentanyl, opiods). Control: Standard practice, co-intervention were a combination of medications (i.e. oral midacolam, fentanyl, opiods). ***Cohort*** Control: Same Trial treatment was significantly better than the control.
Care provider and setting	The same anesthesiologists provided care in the trial and cohort groups. All

	patients were treated in the pediatric surgery operating room.
Outcomes	Mean time to discharge in minutes.

Author	**Bedi 2000**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients preferred choosing their own treatment. Intention-to-treat analysis was not used in the trial.
Selection Bias	There were no differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	There were 8 inappropriate exclusions in the trial treatment 1 group. 6 inappropriate exclusions in the trial treatment 2 group. 32 inappropriate exclusions in the cohort treatment 1 group. 24 inappropriate exclusions in the cohort treatment 2 group.
Intervention and Co-interventions	***Trial*** Treatment 1: Counseling Treatment 2: Antidepressant medication ***Cohort*** Treatment 1: Same Treatment 2: Same There were no significant difference between the trial treatments.
Care provider and setting	Experienced counselors provided care in the trial and cohort treatment 1 groups. General practitioner provided care in the trial and cohort treatment 2 groups. All patients were treated at their general practitioner's clinic.
Outcomes	Mean BDI score. At 8 weeks.

Author	**Bell 2000**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort refused to be randomized, most because they did not want to risk having to reduce their exercise levels. Intention-to-treat analysis was used in the trial.
Selection Bias	There were differences in baseline characteristics between groups.
Detection Bias	Unclear if follow-up was the same between groups.
Exclusion Bias	There were 2 inappropriate exclusions in each of the trial control and cohort control groups.
Intervention and Co-interventions	***Trial*** Treatment: Reducing their exercise program to less than or equal to three sessions weekly Control: Continuing their intended exercise program ***Cohort*** Control: Same There was no significant difference between the trial treatment and trial control.
Care provider and setting	Unclear who provided care in the trial and cohort groups. Unclear where the patients were treated.
Outcomes	Number of women who delivered prematurely (before 37 weeks).

Author	**Bhattacharya 1998**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients.

	Cohort patients lived more than 20 miles away from the hospital to participate. Intention-to-treat analysis was used in the trial.
Selection Bias	There were no differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	It was unclear the reason for any of the following exclusions: 12 in the trial treatment 1 group, 6 in the trial treatment 2 group, and 16 in the cohort treatment 2 group.
Intervention and Co-interventions	***Trial*** Treatment 1: Outpatients (discharged from the hospital on the same day as the procedure). Treatment 2: Admitted to 48 hour inpatient care. ***Cohort*** Treatment 2: Same There was no significant difference between the trial treatments. Note: Significantly more patients in the trial asked for and received concurrent sterilization with their regular procedure.
Care provider and setting	All groups were operated on by three experienced consultants or by trainees observed by these consultants. All patients were treated at the hospital.
Outcomes	Number of patients who were not "Very satisfied" with treatment. At 12 months.

Author	**Biasoli 2008**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort participants refused enrollment or their physician decided not to enroll the patient. Intention-to-treat analysis was not used in the trial.
Selection Bias	There were differences in baseline characteristics between groups.
Detection Bias	More frequent follow-up visits for those in the trial.
Exclusion Bias	No exclusions in the trial. 2 inappropriate exclusions in the cohort.
Intervention and Co-interventions	***Trial*** Treatment 1: Chemotherapy with ABVD (doxorubicin, bleomycin, vinblastine, dacarbazine) Treatment 2: Chemotherapy with BEACOPP regimen (bleomycin, etoposide, doxorubicin, cyclophosphamide, vincristine, procarbazine, prednisone) ***Cohort*** Treatment 1: Same Treatment 2: In one group four patients received four cycles of increased-dose BEACOPP and four cycles of standard BEACOPP and another four patients received four cycles of increased-dose BEACOPP and three cycles of ABVD Unclear if there was a significant difference between the trial treatments.
Care provider and setting	The same physicians provided care in the trial and cohort groups. All patients were treated at the same hospital.
Outcomes	Number who did not have complete remission at 1 month.

Author	**Biederman 1985**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort participants refused randomization. Intention-to-treat analysis was not used in the trial.

Selection Bias	There were no differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	There was only one inappropriate exclusion in the trial control group.
Intervention and Co-interventions	***Trial*** Treatment: Amitriptyline (up to 3mg/kg/day) Control: Placebo (Same doseage as above) ***Cohort*** Control: No drug treatment only psychosocial treatment There was no significant difference between the trial treatments.
Care provider and setting	Psychiatrist, medical and nursing staff provided care in the trial, unclear who provided care outside trial. All patients were treated at the Eating Disorder Unit at Child Psychiatry Services.
Outcomes	Number who had a response of greater than 30% on the Schedule for Affective Disorders and Schizophrenia-Change Version (SADS-C) Scale. At 5 weeks.

Author	**Bijker 2002**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Either the cohort patient or their physician chose their treatment. Intention-to-treat analysis was not used in the trial.
Selection Bias	Most likely no differences in baseline characteristics between groups (N= 433).
Detection Bias	More frequent follow-up in the RCT.
Exclusion Bias	There were 10 patients excluded for an unknown reason in the trial group.
Intervention and Co-interventions	***Trial*** Treatment 1: Local excision + radiotherapy Treatment 2: Local excision ***Cohort*** Treatment 1: Same Treatment 2: Same Treatment 3: Mastectomy There was a statistically significant benefit associated with the trial treatment 1.
Care provider and setting	Unclear who provided care in the trial. All patients were treated at the same setting.
Outcomes	Local recurrence. At 4 years.

Author	**Blichert- Toft 1988**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients declined randomization. Intention-to-treat analysis was used in the trial.
Selection Bias	There were no differences in baseline characteristics between groups.
Detection Bias	Unclear if follow-up was the same between groups.
Exclusion Bias	Protocol violation was the reason for all the exclusions- "the most important were erroneous allocation of systemic adjuvant therapy, surgical divergences from protocol, deficient pathoanatomical examination of the surgical specimen, disseminated disease demonstrated post-operatively". 21 patients were excluded in the trial treatment 1 group, 30 patients in the

	cohort treatment 1 group. 22 patients were excluded in the trial treatment 2 group, 19 in the cohort treatment 2 group.
Intervention and Co-interventions	***Trial*** Treatment 1: Breast preserving therapy Treatment 2: Mastectomy ***Cohort*** Treatment 1: Same Treatment 2: Same Unclear if there is a significant difference between the trial treatments.
Care provider and setting	Unclear who provided care in both the trial and cohort groups. All patients were treated in similar surgical departments.
Outcomes	Number of patients who had a recurrence of disease. At 3 years.

Author	**Blumenthal 1997**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients were unable to attend the 3 sessions/week requirement of the trial (mostly due to distance). Intention-to-treat analysis was used in the trial.
Selection Bias	There were no differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	1 and 2 patients were lost to follow-up in the trial and cohort control groups, respectively.
Intervention and Co-interventions	***Trial*** Treatment: Exercise training group Control: Stress Management Program ***Cohort*** Control: Usual Care Unclear whether the trial treatment was significantly different than the trial control.
Care provider and setting	Hospital affiliated cardiologists provided care in the trial, patients' usual local cardiologists provided care outside trial. Trial patients were treated at the Duke University Medical Centre. Cohort patients were treated at their local medical centre.
Outcomes	Number of cardiac related events. At 2 years.

Author	**Boesen 2007**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients declined trial participation either due to distance, the time commitment involved or because they felt they did not need support. Intention-to-treat analysis was used in the trial.
Selection Bias	There were differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	There were two appropriate exclusions and one exclusion because the patient received immunotherapy in the trial treatment group. In the trial control group one patient was inappropriately excluded.
Intervention and Co-interventions	***Trial*** Treatment: Psychoeducational intervention Control: Surgery alone ***Cohort***

	Control: Unclear what intervention, if any, this group received There was no significant difference between the trial treatment and control groups.
Care provider and setting	Plastic surgeons provided care in both trial and cohort groups. All patients were treated at the outpatient clinic.
Outcomes	Mortality at 5 years.

Author	**Boezaart 1998**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients requested no treatment (medication). Intention-to-treat analysis was not used in the trial.
Selection Bias	There was a difference in a baseline characteristic between the two groups.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	No post-randomization exclusions.
Intervention and Co-interventions	***Trial*** Treatment 1: 3mg bromazepam Treatment 2: 6mg bromazepam Treatment 3: 0.5mg alprazolam Treatment 4: 1mg alprazolam Treatment 5: 5mg diazepam Control: Placebo - multivitamin pill ***Cohort*** Control: No treatment There was no significant difference between the trial treatments and control.
Care provider and setting	The same surgeons/anesthesiologists provided care in the trial and cohort groups. Unclear who provided care outside trial. All patients were treated at a private hospital.
Outcomes	Mean anxiety scores during surgery (measured on the Visual Analog Scale).

Author	**Brinkhaus 2008**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients declined trial participation. Intention-to-treat analysis was not used in the trial.
Selection Bias	There were differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	There were 68 inappropriate exclusions in the trial treatment group, 475 inappropriate exclusion in the cohort treatment group, 52 inappropriate exclusion in the trial control group.
Intervention and Co-interventions	***Trial*** Treatment: Immediate acupuncture with routine care Control: No treatment (delayed acupuncture given after 3 months) ***Cohort*** Treatment: Same There was a significant beneficial effect in favor of the trial treatment.
Care provider and setting	Physicians with a certification in acupuncture provided care in the trial and cohort groups. Unclear where patients were treated, but it was the same

	setting for both groups.
Outcomes	Mean sum score on the Rhinitis Quality of Life Questionnaire. At 3 months.

Author	**Caplan 1984**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Unclear the reason cohort patients were not enrolled in the trial. Intention-to-treat analysis was not used in the trial.
Selection Bias	Most likely there were differences in baseline characteristics between groups. (N=75)
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	No post-randomization exclusions.
Intervention and Co-interventions	***Trial*** Treatment 1: Cefsulodin Treatment 2: Ticarcillin Treatment 3: Tobramycin ***Cohort*** Treatment 1: Same No clinically important differences between treatment groups.
Care provider and setting	Unclear who provided care in the trial and cohort groups. All patients were treated at the Emory University Cystic Fibrosis Center or similar setting.
Outcomes	Number of patients without complete resolution of infection.

Author	**CASS Principal Investigators and their associates 1984**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients either declined randomization (28%), their physician declined (69%) or for other reasons (3%). Intention-to-treat analysis was not used in the trial.
Selection Bias	There were differences in baseline characteristics between groups.
Detection Bias	More frequent visits for those in RCT.
Exclusion Bias	There was 1 inappropriate exclusion in the trial and 10 inappropriate exclusions in the cohort.
Intervention and Co-interventions	***Trial*** Treatment 1: Coronary artery bypass surgery Treatment 2: Medically treated ***Cohort*** Treatment 1: Same Treatment 2: Same No significant difference between trial treatments.
Care provider and setting	Unclear who provided care in the trial or cohort groups. Unclear where the patients were treated.
Outcomes	Mortality at 5 years

Author	**Chauhan 1992**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients did not give consent for randomization. Intention-to-treat analysis was not used in the trial.
Selection Bias	There were no differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.

Exclusion Bias	No post-randomization exclusions.
Intervention and Co-interventions	***Trial*** Treatment 1: Prophylactic amnio-infusion Treatment 2: No amnio-infusion ***Cohort*** Control: No amnio-infusion Unclear if there were any significant difference between trial treatments.
Care provider and setting	Physicians provided care in the trial, unclear who provided care outside trial. All patients were treated at the Portsmouth Naval Hospital, Labor suite.
Outcomes	Cesarean section due to fetal distress. At birth.

Author	**Chesebro 1983**
Design and Methods	Randomized trial and a separate cohort of eligible patients treated within two months of the trial follow-up period. Cohort was assembled because the authors decided after trial commencement that they needed a control group. Intention-to-treat analysis was not used in the trial.
Selection Bias	There were no differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	No post-randomization exclusions.
Intervention and Co-interventions	***Trial*** Treatment 1: Warfarin plus dipyridamole 100 mg orally 4 times a day Treatment 2: Warfarin plus aspirin 250 mg orally twice a day ***Cohort*** Control: Warfarin alone No significant difference between trial treatments.
Care provider and setting	Unclear who provided care in the trial or cohort groups. Unclear where the patients were treated.
Outcomes	Number not free of thromboembolism. At 3 years.

Author	**Chilvers 2001**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients did not give consent for randomization. Intention-to-treat analysis was not used in the trial.
Selection Bias	There were no differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	There were 4 inappropriate exclusions in the trial treatment 1 group and 11 inappropriate exclusions in the cohort treatment 1 group. There was 1 inappropriate exclusion in the trial treatment 2 group and 2 inappropriate exclusions in the cohort treatment 2 group.
Intervention and Co-interventions	***Trial*** Treatment 1: Counseling Treatment 2: Antidepressant drugs ***Cohort*** Treatment 1: Same Treatment 2: Same No significant difference between trial treatments.

Care provider and setting	Experienced counselors and general practitioners (GPs) provided care in the trial treatment 1 and 2 groups, respectively. In the cohort treatment 1 group different counselors were used who adopted the "most suitable counseling approach". In the cohort treatment 2 group different GPs were used who were given written guidelines on routine drug treatment of depression. All patients were treated at their counselor or GPs' practice.
Outcomes	Number who were not in remission from depression. At 12 months.

Author	**Clagett 1984**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients did not consent to randomization. Instead assignment was based on preference of the individual or their physician. Intention-to-treat analysis was not used in the trial.
Selection Bias	There were no differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	There were 2 inappropriate exclusions in the cohort treatment 1 group.
Intervention and Co-interventions	***Trial*** Treatment 1: 650 mg of aspirin twice a day Treatment 2: Arteriography and prophylactic carotid endarterectomy ***Cohort*** Treatment 1: Same Treatment 2: Same No significant difference between trial treatments.
Care provider and setting	Unclear who provided care in the trial and cohort groups. Unclear where the patients were treated.
Outcomes	Incidence of stroke caused by intervention. At 3 years.

Author	**Clapp 1989**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients' parents refused consent. Intention-to-treat analysis was used in the trial.
Selection Bias	There were differences in baseline characteristics between groups.
Detection Bias	Unclear if follow-up was the same between groups.
Exclusion Bias	No post-randomization exclusions.
Intervention and Co-interventions	***Trial*** Treatment: Intravenously administered immune globulin Control: Placebo ***Cohort*** Control: No treatment No significant difference between trial treatment and control groups.
Care provider and setting	Unclear who provided care in the trial and cohort groups. All patients were treated at the Rainbow Babies and Children's Hospital.
Outcomes	Mortality at hospital discharge.

Author	**Clemens 1992**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients refused to participate. Intention-to-treat analysis was used in the trial.

Selection Bias	There were no differences in baseline characteristics between groups.
Detection Bias	Unclear if follow-up was the same between groups.
Exclusion Bias	Unclear if there were any exclusions.
Intervention and Co-interventions	***Trial*** Treatment 1: B- subunit, killed whole cell vaccine Treatment 2: Killed whole cell vaccine without B subunits Control: Placebo of E.coli K12 strain ***Cohort*** Control: No treatment Unclear if there were any significant differences between trial treatments and control.
Care provider and setting	Unclear who provided care in the trial and cohort groups. All patients were treated at the International Centre for Diarrhoeal Disease Research.
Outcomes	Number of cholera episode. At 3 years.

Author	**Cooper 1997**
Design and Methods	Patients were randomized either to a randomized trial or a partially randomized preference trial (where patients were allowed to choose between a trial and preference arm). Intention-to-treat analysis was not used in the trial.
Selection Bias	There were differences in baseline characteristics between groups.
Detection Bias	Unclear if follow-up was the same between groups.
Exclusion Bias	No post-randomization exclusions.
Intervention and Co-interventions	***Trial*** Treatment 1: Transcervical resection of the endometrium Treatment 2: Medical management ***Cohort*** Treatment 1: Same Treatment 2: Same There was a statistically significant benefit associated with the trial treatment 1.
Care provider and setting	The same gynecologists provided care in all groups. All patients were treated at the same hospital.
Outcomes	Number of patients who were not satisfied with treatment. At 4 months.

Author	**Cowchock 1992**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients refused to participate. Intention-to-treat analysis was not used in the trial.
Selection Bias	There were differences in baseline characteristics between groups.
Detection Bias	Unclear if follow-up was the same between groups.
Exclusion Bias	No post-randomization exclusions.
Intervention and Co-interventions	***Trial*** Treatment 1: Low dose heparin Treatment 2: 40 mg prednisone daily ***Cohort*** Treatment 1: Same Treatment 2: Same

	No significant difference between trial treatments.
Care provider and setting	Unclear who provided care in the trial and cohort groups. Unclear where the patients were treated.
Outcomes	Number who did not have a live birth.

Author	**Creutzig 1993**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients refused randomization, and chose their treatment. Intention-to-treat analysis was used in the trial.
Selection Bias	There were no differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	No post-randomization exclusions.
Intervention and Co-interventions	***Trial*** Treatment: Irradiation Control: No irradiation ***Cohort*** Treatment: Same Control: Same No significant difference between trial treatment and control.
Care provider and setting	Unclear who provided care in the trial and cohort groups. Unclear where the patients were treated.
Outcomes	Relapse rate.

Author	**Dahan 1986**
Design and Methods	Patients were randomized either to a randomized trial or a cohort study. Intention-to-treat analysis was not used in the trial.
Selection Bias	Most likely there were differences in baseline characteristics between groups. (N=60)
Detection Bias	Unclear if follow-up was the same between groups.
Exclusion Bias	No post-randomization exclusions.
Intervention and Co-interventions	***Trial*** Treatment: Received informed consent ***Cohort*** Control: Received no informed consent (and therefore no knowledge of trial participation) All patients received placebo pills (so the trial treatment was not effective).
Care provider and setting	Unclear who provided care in the trial and cohort groups. Unclear where the patients were treated.
Outcomes	Number of side effects reported. After 1 day.

Author	**Dalal 2007**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients were allowed to choose their treatment in this preference-trial design. Intention-to-treat analysis was used in the trial.
Selection Bias	There were no differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	There were 10 inappropriate exclusions in the trial treatment 1 group, and 17 inappropriate exclusions in the cohort treatment 1 group. There were 10

	inappropriate exclusions in the trial treatment 2 group, and 9 inappropriate exclusions in the cohort treatment 2 group.
Intervention and Co-interventions	***Trial*** Treatment 1: Home based rehabilitation Treatment 2: Hospital based rehabilitation ***Cohort*** Treatment 1: Same Treatment 2: Same No significant difference between trial treatments.
Care provider and setting	Cardiac rehabilitation nurse provided care in the trial treatment 1, multidisciplinary team (cardiac rehabilitation nurse, physiotherapist or exercise therapist, with input from a psychologist or occupational therapist, pharmacist and dietician) provided care in the trial treatment 2 group. Patients in the cohort received care from similar individuals. Trial patients received care at the local hospital or community centres (treatment 1) or home (treatment 2). Unclear who provided care in the cohort group.
Outcomes	Mean Global MacNew score. At 9 months.

Author	**Decensi 2003**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients either underwent surgery in a different hospital after an initial screening biopsy (n=6) or were enrolled after the trial women assigned to receive tamoxifen had been randomly assigned (n=23). Intention-to-treat analysis was not used in the trial.
Selection Bias	There were no differences in baseline characteristics between groups.
Detection Bias	Unclear if follow-up was the same between groups.
Exclusion Bias	There was one inappropriate exclusion in the trial treatment 1 group and four in the trial treatment 3 group. There was 18 inappropriate exclusions in the cohort group.
Intervention and Co-interventions	***Trial*** Treatment 1: 1 mg/day of tamoxifen Treatment 2: 5 mg/day of tamoxifen Treatment 3: 20 mg/day of tamoxifen All patients received an initial dose of 20 mg of tamoxifen ***Cohort*** Control: No tamoxifen at all Significant difference between trial treatments.
Care provider and setting	The same experienced pathologist and reference physician provided care in the trial and cohort groups. All patients were treated at the European Institute of Oncology.
Outcomes	Mean level of plasma IGF-I. At 4 weeks.

Author	**Detre 1999**
Design and Methods	Randomized trial and a separate cohort of eligible patients treated within two and a half months from trial follow-up. Cohort patients refused randomization. Intention-to-treat analysis was not used in the trial.
Selection Bias	There were differences in baseline characteristics between groups.
Detection Bias	More frequent visits for those in RCT.

Exclusion Bias	No post-randomization exclusions.
Intervention and Co-interventions	***Trial*** Treatment 1: PTCA Treatment 2: CABG ***Cohort*** Treatment 1: Same Treatment 2: Same Significant beneficial effect of trial treatment 2 compared to trial treatment 1.
Care provider and setting	Unclear who provided care in the trial and cohort groups. All patients were treated at the same hospital.
Outcomes	Kaplan Meier cardiac mortality rates. At 5 years.

Author	**Diehl 1995**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients refused randomization. Intention-to-treat analysis was not used in the trial.
Selection Bias	There were no differences in baseline characteristics between groups.
Detection Bias	Unclear if follow-up was the same between groups.
Exclusion Bias	No post- randomization exclusions.
Intervention and Co-interventions	***Trial*** Treatment 1: Low-dose radiotherapy, co-intervention: persisting residual tumors received additional radiation up to a total dose of 40 Gy. Treatment 2: Chemotherapy ***Cohort*** Control: Chose no consolidation therapy No significant difference between trial treatments.
Care provider and setting	Unclear who provided care in the trial and cohort groups. All patients were treated at one of the study centers.
Outcomes	Relapse rate. At 6 years

Author	**Eberhardt 1996**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients did not participate in the trial due to fear of side effects, fear of not receiving an active drug or reasons not related to the trial medication. Intention-to-treat analysis was not used in the trial.
Selection Bias	There were no differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	There were 12 inappropriate exclusions in the treatment trial group. 19 inappropriate exclusions in the control trial group. It is unclear whether there were any exclusions in the cohort.
Intervention and Co-interventions	***Trial*** Treatment: D-Penicillamine (DPA) Control: Placebo ***Cohort*** Control: Slow-acting anti-rheumatic drugs (SAARDS) such as chloroquine No significant difference between trial treatments.
Care provider and setting	Unclear who provided care in the trial and cohort groups. All patients were

	treated at the Rheumatology Unit.
Outcomes	Number not in remission. At 2 years.

Author	**Edsmyr 1978**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Unclear why cohort patients did not participate in the trial. Intention-to-treat analysis was not used in the trial.
Selection Bias	Most likely there were differences in baseline characteristics between groups. (N=27)
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	No post- randomization exclusions.
Intervention and Co-interventions	***Trial*** Treatment 1: 100 mg of 2.6- cis given orally twice daily Treatment 2: 300 mg estramustine given orally twice daily ***Cohort*** Treatment 1: Same Unclear if there was a significant difference between trial treatments.
Care provider and setting	Unclear who provided care in the trial and cohort groups. Unclear where the patients were treated in both groups.
Outcomes	Number experiencing some degree of pain (on a 0-3 scale). At 3 months.

Author	**Ekstein 2002**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients refused randomization. Intention-to-treat analysis was used in the trial.
Selection Bias	Most likely there are no differences in baseline characteristics between groups. (N=1255)
Detection Bias	Less frequent visits for those in RCT.
Exclusion Bias	There were 3 inappropriate exclusions in the cohort treatment 1 group. There were 3 inappropriate exclusions in the trial treatment 2 group and 2 inappropriate exclusions in the cohort treatment 2 group.
Intervention and Co-interventions	***Trial*** Treatment 1: PTCA Treatment 2: CABG ***Cohort*** Treatment 1: Same Treatment 2: Same No significant difference between trial treatments.
Care provider and setting	An experienced interventional cardiologist and surgeon provided care in the treatment 1 and 2 trial groups, respectively. Similar professionals were providing care in the cohort groups. All patients were treated at the Hadassah University Hospital or similar setting.
Outcomes	Mean score on the mobility domain of the EuroQoL questionnaire (0-100 scored questionnaire). At 6 months.
Note	Additional information on the trial was extracted from: Serruys PW, Unger F, Eduardo Sousa J, et al. Comparison of coronary- artery bypass surgery and stenting for the treatment of multivessel disease. N Engl J Med 2001;344:1117±24.

Author	**Emery 2003**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients had a preference for the treatment. Intention-to-treat analysis was not used in the trial.
Selection Bias	There were no differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	Excluded the treatment cohort group because six drop-out patients from the trial were included in the cohort. There were 14 inappropriate exclusions in the treatment trial group, 18 inappropriate exclusions in the control trial group, 9 inappropriate exclusions in the cohort control group.
Intervention and Co-interventions	***Trial*** Treatment: IVF Counseling Control: No IVF counseling ***Cohort*** Control: Same No significant difference between trial treatment and control.
Care provider and setting	The counselors provided care in the treatment trial group. Unclear who provided care in the control trial and cohort groups. All patients were treated at the same IVF programme at the same hospital.
Outcomes	Mean score on the State Trait Anxiety Inventory (Trait). At 6 weeks after embryo transfer. *Assumed that the number of men and women were equal when calculating the combined mean scores (since the cohort group was an odd number, we arbitrarily divided the total such that the female group had an extra patient).*

Author	**Euler 2005**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients were being fed with human milk. Intention-to-treat analysis was used in the trial.
Selection Bias	There were no differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	No post-randomization exclusions.
Intervention and Co-interventions	***Trial*** Treatment 1: S-26Ò Gold with 3.0 g/L FOS (inulin) Treatment 2: S-26Ò Gold with 1.5 g/L FOS (inulin) ***Cohort*** Control: Human milk No significant difference between trial treatments.
Care provider and setting	Unclear who provided care in the trial and cohort groups. All patients were treated at the same centre.
Outcomes	Bifidobacterium organism count (log base 10). At 7 days.

Author	**Feit 2000**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients refused randomization. Intention-to-treat analysis was used in

	the trial.
Selection Bias	There were differences in baseline characteristics between groups.
Detection Bias	More frequent visits for those in RCT.
Exclusion Bias	There were 660 inappropriate exclusions in the trial and 674 in the cohort group.
Intervention and Co-interventions	***Trial*** Treatment 1: CABG Treatment 2: PTCA ***Cohort*** Treatment 1: Same Treatment 2: Same Control: Medical/No treatment Trial treatment 2 was significantly better than treatment 1.
Care provider and setting	Unclear who provided care in the trial and cohort groups. Trial patients were treated at the study hospital, cohort patients were treated in the community.
Outcomes	Mortality at 7 years.

Author	**Forbes 2000**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients refused randomization. Intention-to-treat analysis was used in the trial.
Selection Bias	There were differences in baseline characteristics between groups, but were adjusted for in the analysis.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	No post-randomization exclusions.
Intervention and Co-interventions	***Trial*** Treatment: Inhaled Entonox for 60 seconds prior to colonoscopy Control: Intravenous midazolam (0.06mg/kg) and meperidine (0.76mg/kg) ***Cohort*** Control: Same No significant difference between trial treatments.
Care provider and setting	The same gastroenterologist and colonoscopist provided care to all patients. All patients were treated at the Royal Perth Hospital.
Outcomes	Number who experienced an adverse events during colonoscopy.

Author	**Franz 1995**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients were not offered trial enrollment. Intention-to-treat analysis was not used in the trial.
Selection Bias	There were no differences in baseline characteristics between groups.
Detection Bias	Less frequent visits for those in the RCT.
Exclusion Bias	There were 68 inappropriate exclusions in the trial group.
Intervention and Co-interventions	***Trial*** Treatment 1: Practice guideline care Treatment 2: Basic nutrition care ***Cohort*** Control: No treatment

	Trial treatment 1 was significantly better than treatment 2.
Care provider and setting	Dieticians provided care in the trial, unclear who provided care outside trial. All patients were treated at the same centre.
Outcomes	Mean HbA1c. At 6 months.

Author	**Gall 2007**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Patient preference design; cohort patients refused randomization. Intention-to-treat analysis was not used in the trial.
Selection Bias	There were differences in baseline characteristics between groups, but were adjusted for in the analysis.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	There were 21 inappropriate exclusions in the treatment trial group, 25 inappropriate exclusions in the control trial group, 23 inappropriate exclusions in the cohort control group and 18 inappropriate exclusions in the cohort control group.
Intervention and Co-interventions	***Trial*** Treatment 1: Follow-up with a general practitioner (GP) Treatment 2: Follow-up with a surgeon ***Cohort*** Treatment 1: Same Treatment 2: Same No significant difference between trial treatment and control.
Care provider and setting	A GP and surgeon provided care in the trial treatment 1 and 2 groups, respectively. A different GP and surgeon provided care in the cohort groups. All patients were treated at the same hospital setting.
Outcomes	Number of patients with probable anxiety (HADS$\geq$ 11). At 24 months.

Author	**Giron 2010**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients refused randomization. Intention-to-treat analysis was used in the trial.
Selection Bias	There were no differences in baseline characteristics between groups.
Detection Bias	More frequent visits for those in RCT.
Exclusion Bias	The trial treatment 2 group was excluded from the analysis because their outcome was not provided. There was 1 inappropriate exclusion in the trial group and 2 inappropriate exclusions in the cohort control group.
Intervention and Co-interventions	***Trial*** Treatment: Family intervention + counseling + standard treatment ***Cohort*** Control: Unclear Significant beneficial effect of the trial treatment.
Care provider and setting	Psychiatrists, psychologists, social workers and nurses provided care in the trial treatment group. Different person provided care in the cohort control group. All patients were treated at the mental health centre.
Outcomes	Mean absolute change in number of psychiatric hospitalizations. During 2 years.

Author	**Goodkin 1987**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients did not agree to cyclophosphamide treatment. Intention-to-treat analysis was used in the trial.
Selection Bias	There were no differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	Do not provide the number of post-randomization exclusions.
Intervention and Co-interventions	***Trial*** Treatment 1: IV Cyclophosphamide induction treatment + alternate month maintence Treatment 2: IV Cyclophosphamide induction treatment ***Cohort*** Control: No treatment No significant difference between trial treatments.
Care provider and setting	Unclear who provided care in the trial, patients outside the trial received no care. Patients outside the trial did not receive care at the same setting as trial patients.
Outcomes	Number of patients who did not have their neurological status stabilized. At 24 months.

Author	**Gossop 1986**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients had a strong preference for one treatment. Intention-to-treat analysis was not used in the trial.
Selection Bias	There were no differences in baseline characteristics between groups.
Detection Bias	Unclear whether follow-up was the same between groups.
Exclusion Bias	No post-randomization exclusions.
Intervention and Co-interventions	***Trial*** Treatment 1: Inpatient treatment Treatment 2: Outpatient treatment ***Cohort*** Treatment 1: Same Treatment 2: Same Unclear whether there was a significant difference between trial treatments.
Care provider and setting	Unclear who provided care in any of these groups. All patients were treated at the same setting.
Outcomes	Number not withdrawn from opiates by the end of the supervised period.

Author	**Grant 2008**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients had a strong preference of treatment. Intention-to-treat analysis was used in the trial.
Selection Bias	There were differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	There were 33 inappropriate exclusions in the treatment trial group, 25 inappropriate exclusions in the control trial group, 49 inappropriate exclusions in the cohort treatment group and 29 inappropriate exclusions in the cohort

	control group.
Intervention and Co-interventions	***Trial*** Treatment 1: Fundoplication surgery Treatment 2: Medical management (GERD drug) ***Cohort*** Treatment 1: Same Treatment 2: Same Significant beneficial effect in favor of the trial treatment 1 group.
Care provider and setting	Same surgeon provided care in the trial and cohort treatment 1 groups, same gastroenterologist provided care in the trial and cohort treatment 2 groups. All patients were treated at the same hospital.
Outcomes	Mean reflux questionnaire score (questionnaire ranges from 0-100). At 12 months.

Author	**Gunn 2000**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients refused randomization. Intention-to-treat analysis was not used in the trial.
Selection Bias	There were differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	No post-randomization exclusions
Intervention and Co-interventions	***Trial*** Treatment 1: Early discharge Treatment 2: Routine discharge (a pattern of weight gain needed before discharge) ***Cohort*** Treatment 2: Same Trial treatment 2 was significantly better than trial treatment 1.
Care provider and setting	Nursing staff provided care in the trial and in the cohort. All patients were treated at the same unit in the hospital.
Outcomes	Mean weight (gms). At 6 weeks after discharge.

Author	**Helsing 1998**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients refused randomization. Intention-to-treat analysis was not used in the trial.
Selection Bias	There were no differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	There control group of the cohort was not included in the analysis because no outcomes were provided for this arm. 1 deliberate appropriate post-randomization exclusion in the trial control group.
Intervention and Co-interventions	***Trial*** Treatment 1: Supportive care + Palliative platinum based chemotherapy Treatment 2: Supportive care ***Cohort*** Treatment 1: Same

	Significant beneficial effect of trial treatment 1.
Care provider and setting	Unclear who provided care, but was the same professional in both trial and cohort groups. All patients were treated in the same setting.
Outcomes	Probability of mortality. At 1 year.

Author	**Henriksson 1986**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients refused randomization. Intention-to-treat analysis was not used in the trial.
Selection Bias	Most likely there were differences in baseline characteristics between groups. (N= 100)
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	No post-randomization exclusions.
Intervention and Co-interventions	***Trial*** Treatment 1: Oestrogen and ethinyloestradiol Treatment 2: Orchidectomy ***Cohort*** Treatment 1: Same Treatment 2: Same Significant beneficial effect in favor of trial treatment 2.
Care provider and setting	Unclear who provided care in the trial and cohort group. All patients were treated in the same setting.
Outcomes	Number of major cardiovascular events. At 1 year.

Author	**Heuss 2004**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients refused randomization (wanted to be asleep during the whole procedure, refused to be responsible for sedation, unclear the reason). Intention-to-treat analysis was not used in the trial.
Selection Bias	There were no differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	No post-randomization exclusions.
Intervention and Co-interventions	***Trial*** Treatment 1: Patients were connected to a PCA pump with one-way valve. Treatment 2: An intermittent bolus technique was used. ***Cohort*** Treatment 2: Same No significant difference between trial treatments.
Care provider and setting	All patients were treated by the same trained nurse and endoscopist. All patients were treated at the same department.
Outcomes	Mean pain (VAS scale) during colonoscopy.

Author	**Hoh 1998**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Unclear why cohort patients were not enrolled in the trial. Intention-to-treat analysis was not used in the trial.
Selection Bias	There were no differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.

Exclusion Bias	There were 10 inappropriate exclusions in the trial group.
Intervention and Co-interventions	***Trial*** Treatment 1: Standard whole-casein-protein based oral supplement Treatment 2: Digest of a soy, peptide-based supplement ***Cohort*** Control: Non-supplemented study arm No significant difference between trial treatments.
Care provider and setting	Unclear who provided care in the trial. Only trial patients were treated at the hospital.
Outcomes	Mean proportion of energy needs met at 6 weeks.

Author	**Howard 2010**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Preference trial where cohort patients refused randomization. Intention-to-treat analysis was not used in the trial.
Selection Bias	There were no differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	There were 5 inappropriate exclusions in the trial treatment 1 group, 8 in the trial treatment 2 group, 7 in the cohort treatment 1 group and 10 in the cohort treatment 2 group.
Intervention and Co-interventions	***Trial*** Treatment 1: Women's crisis house Treatment 2: Hospital admission (ward) ***Cohort*** Treatment 1: Same Treatment 2: Same There was no clinically important difference between trial treatments.
Care provider and setting	The same nurses and health care workers with a background in mental health provided care in the trial and cohort treatment 1 groups. Unclear who provided care in the treatment 2 group, but can assume it was the same professional. All patients were either treated at the center or ward.
Outcomes	Mean Global Assessment of Functioning (GAF) score. At 12 weeks.

Author	**Howie 1997**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Preference trial where cohort patients refused randomization. Intention-to-treat analysis was used in the trial.
Selection Bias	There were no differences in baseline characteristics between groups.
Detection Bias	Unclear if follow-up was the same between groups.
Exclusion Bias	There were 60 inappropriate exclusions in the trial treatment 1 group, 58 in the trial treatment 2 group, 46 in the cohort treatment 1 group and 59 in the cohort treatment 2 group.
Intervention and Co-interventions	***Trial*** Treatment 1: Medical abortion Treatment 2: Vacuum aspiration ***Cohort*** Treatment 1: Same Treatment 2: Same

	There was a significant beneficial effect in favor of the trial treatment 2.
Care provider and setting	Nurses provided care for all groups. All patients were treated at the gynecology ward.
Outcomes	Numbers of patients who did not find the procedure acceptable. At 2 years.

Author	**Jena 2008**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients refused randomization. Intention-to-treat analysis was used in the trial.
Selection Bias	There were differences in baseline characteristics between groups, but were controlled for in the analysis.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	There were 99 appropriate exclusions and 72 inappropriate exclusions in the trial treatment group, 95 appropriate and 124 inappropriate exclusions in the trial control group, and 944 appropriate and 520 inappropriate exclusions in the cohort treatment group.
Intervention and Co-interventions	***Trial*** Treatment: Acupuncture with routine care Control: Routine care ***Cohort*** Treatment: Same There was no significant difference between trial treatment and control.
Care provider and setting	Physician with certification in acupuncture provided care in both treatment groups. Unclear where patients were treated, but it was a similar setting across groups.
Outcomes	Mean number of days with headaches. At 6 months.

Author	**Jensen 2003**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients chose their own treatment. Intention-to-treat analysis was used in the trial.
Selection Bias	There were differences in baseline characteristics between groups, but were controlled for in the analysis.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	There were 54 inappropriate exclusions in the trial treatment group, 55 in the trial control group, 16 in the cohort treatment group and 89 in the cohort control group.
Intervention and Co-interventions	***Trial*** Treatment: First- line hormone replacement therapy Control: No hormone replacement therapy ***Cohort*** Treatment: Same Control: Same There was a significant beneficial effect in favor of the trial treatment.
Care provider and setting	Unclear who provided care in either the trial or cohort groups. All patients were treated at the same hospital.
Outcomes	Mean change in weight (kg). At 5 years.

Author	**Kane 1988**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients were selected because study authors wanted to collect a broader base of safety and effectiveness data. Intention-to-treat analysis was not used in the trial.
Selection Bias	Most likely there were differences in baseline characteristics between groups. (N= 175)
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	No post-randomization exclusions.
Intervention and Co-interventions	***Trial*** Treatment: Spinal fusion with the addition of the direct current bone growth stimulator Control: Spinal fusion without stimulation ***Cohort*** Control: Unclear what treatment, if any, they received There was a significant beneficial effect in favor of the trial treatment.
Care provider and setting	Unclear who provided care in either the trial or cohort groups. Unclear where patients were treated.
Outcomes	Number of unsuccessful radiographic fusions.

Author	**Karande 1999**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients refused randomization. Intention-to-treat analysis was not used in the trial.
Selection Bias	There were no differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	There were 19 inappropriate exclusions in the trial treatment group, 14 in the trial control group and 31 in the cohort control group.
Intervention and Co-interventions	***Trial*** Treatment: In vitro fertilization Control: Standard infertility treatment algorithm ***Cohort*** Control: Same There was a significant beneficial effect in favor of the trial control.
Care provider and setting	Unclear who provided care in either the trial or cohort groups. All patients were treated at the infertility clinic.
Outcomes	Number without pregnancy. At 22 months.

Author	**Kayser 2008**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients opted out and chose acetazolamide. Intention-to-treat analysis was not used in the trial.
Selection Bias	There were no differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	No post-randomization exclusions.
Intervention and Co-interventions	***Trial*** Treatment: Low dose of calcium carbasalate

	Control: Placebo ***Cohort*** Control: Low dose of acetazolamide No significant difference between trial treatment and control groups.
Care provider and setting	Same expert provided care to those in the trial and cohort groups. All patients were treated in the same setting.
Outcomes	Incidence of acute mountain sickness. At 6 days.

Author	**Kendrick 2001**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients had a preference for treatment. Intention-to-treat analysis was used in the trial.
Selection Bias	There were differences in baseline characteristics between groups.
Detection Bias	Unclear whether follow-up was the same between groups.
Exclusion Bias	There were 195 inappropriate exclusions in the trial treatment group, 12 in the trial control group, 3 in the cohort treatment group and 2 in the cohort control group.
Intervention and Co-interventions	***Trial*** Treatment: Lumbar spine radiography with usual care Control: Usual care ***Cohort*** Treatment: Same Control: Same There was a significant beneficial effect in favor of the trial control.
Care provider and setting	Unclear who provided care in the trial and cohort groups. All patients were treated at either a general practice or hospital.
Outcomes	Number of patients who still have back pain. At 9 months.

Author	**Kieler 1998**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients refused randomization. Intention-to-treat analysis was not used in the trial.
Selection Bias	There were differences in baseline characteristics, but these were controlled for in the analysis.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	There were 93 inappropriate exclusions in the trial treatment group and 99 in the trial control group.
Intervention and Co-interventions	***Trial*** Treatment: Screening ultrasound scan at 15 weeks Control: Non-screened control group, no ultrasound scan before at least 19 weeks ***Cohort*** Control: Same There was a significant beneficial effect in favor of the trial treatment.
Care provider and setting	Experienced midwives, clinical assistants and obstetricians provided care in the trial treatment group. Specialist obstetricians or general practitioners provided care in the trial and cohort control groups. All patients were treated

	at the antenatal care clinic.
Outcomes	Mean weight at birth.

Author	**King 2000**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients were given the option to choose their own treatment. Intention-to-treat analysis was used in the trial.
Selection Bias	There were differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	There were 10 inappropriate exclusions in the trial treatment group, 13 in the trial control group, and 9 in the trial treatment 2 group. There were 15 inappropriate exclusions in the cohort control group and 14 in the cohort treatment 2 group.
Intervention and Co-interventions	***Trial*** Treatment 1: Usual general practitioner (GP) care Treatment 2: Non-directive counseling (NDC) Control: Cognitive-behaviour therapy (CBT) ***Cohort*** Treatment 2: Same Control: Same No significant difference between trial treatments.
Care provider and setting	GP provided care in the trial and cohort treatment group, clinical psychologists in the trial and cohort treatment 2 group and in the trial and cohort treatment 2 group. All patients were treated in the same setting.
Outcomes	Mean score on the Beck Depression Inventory. At 12 months.

Author	**Kirke 1992**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients were initially pregnant and therefore could not be randomized. Intention-to-treat analysis was not used in the trial.
Selection Bias	There were no differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	There were a total of 3 inappropriate exclusions in the trial groups.
Intervention and Co-interventions	***Trial*** Treatment 1: Folic acid only Treatment 2: Multivitamins excluding folic acid Treatment 3: Folic acid plus multivitamins ***Cohort*** Control: No additional supplements No significant difference between trial treatments.
Care provider and setting	Unclear who provided care in any of these groups. Unclear where the cohort patients were treated.
Outcomes	Number of babies born with neural tube defects.

Author	**Koch-Henriksen 2006**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients refused randomization. Intention-to-treat analysis was not used in the trial.

Selection Bias	There were no differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	There were 44 inappropriate exclusions in the trial treatment 1 group, 33 in the trial treatment 2 group and 46 in the cohort group.
Intervention and Co-interventions	***Trial*** Treatment 1: IFNβ-1b (Betaferon) Treatment 2: IFNβ-1a (Rebif) ***Cohort*** Treatment 1: Same No significant difference between trial treatments.
Care provider and setting	Unclear who provided care in the trial and cohort groups. All patients were treated at the same setting.
Outcomes	Number of relapses within one year of treatment.

Author	**Lansky 1983**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients' parents refused randomization. Intention-to-treat analysis was not used in the trial.
Selection Bias	There were no differences in baseline characteristics between groups.
Detection Bias	More frequent visits for those in the trial.
Exclusion Bias	No post-randomization exclusions.
Intervention and Co-interventions	***Trial*** Treatment: 12 weekly 45-minute session. Children were asked to self-monitor food intake and exercise, a food-exchange plan was taught and children practiced aerobic activities. Control: No treatment ***Cohort*** Control: Same There was a significant beneficial effect in favor of the trial treatment.
Care provider and setting	A physical education instructor provided care in the trial treatment group. All other groups received no treatment. All patients were treated at the same setting.
Outcomes	Mean weight lost. At 12 weeks.

Author	**Lichtenberg 2008**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients could not be located or refused to participate. Intention-to-treat analysis was used in the trial.
Selection Bias	There were no differences in baseline characteristics between groups.
Detection Bias	More frequent visits for those in the trial.
Exclusion Bias	No post-randomization exclusions.
Intervention and Co-interventions	***Trial*** Treatment 1: Case management Treatment 2: Standard care ***Cohort*** Control: No treatment No significant difference between trial treatments.

Care provider and setting	A case manager provided care in the trial treatment 1 group and psychiatrist, nurse and social worker in the trial treatment 2 group. Unclear who provided care outside trial. Unclear where patients outside the trial were treated.
Outcomes	Number of rehospitalizations within 1 year.

Author	**Lidbrink 1995**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients did not attend the screening. Intention-to-treat analysis was not used in the trial.
Selection Bias	There were no differences in baseline characteristics between groups.
Detection Bias	Unclear whether follow-up was the same between groups.
Exclusion Bias	No post-randomization exclusions.
Intervention and Co-interventions	***Trial*** Treatment: Breast cancer screening Control: No screening during the period of the trial ***Cohort*** Control: No screening at all No significant difference between trial treatment and control groups.
Care provider and setting	Unclear who provided care in the trial and cohort group. Unclear where patients were treated.
Outcomes	Number of breast cancer- related deaths. Followed for an average of 7.4 years.

Author	**Link 1991**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients refused randomization. Intention-to-treat analysis was used in the trial.
Selection Bias	There were no differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	No post-randomization exclusions.
Intervention and Co-interventions	***Trial*** Treatment: Immediate, intensive adjuvant chemotherapy Control: Observation alone with no adjuvant therapy ***Cohort*** Treatment: Same Control: Same There was a significant beneficial effect in favor of the trial treatment.
Care provider and setting	Unclear who provided care in the trial and cohort groups. All patients were treated at the same setting.
Outcomes	Number of patients experiencing a recurrence. At 6 years.

Author	**Liu 1998**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients either did not give consent, the team arrived late, or the obstetrician requested not to intubate. Intention-to-treat analysis was used in the trial.
Selection Bias	There were no differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	No post-randomization exclusions.

Intervention and Co-interventions	***Trial*** Treatment: Routine meconium management (intubation) Control: No intubation ***Cohort*** Treatment: Same Control: Same No significant difference between trial treatment and control groups.
Care provider and setting	A respiratory therapist and nurse provided care in the trial treatment and control groups. Unclear who provided care outside trial. All patients were treated in a similar setting (delivery room).
Outcomes	Number of newborns experiencing respiratory symptoms requiring supplemental oxygen. Immediately after delivery.

Author	**Lock 2010**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients had a strong preference of treatment. Intention-to-treat analysis was used in the trial.
Selection Bias	There were differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	There were 12 inappropriate exclusions in the trial treatment 1 group, 25 in the trial treatment 2 group, 139 inappropriate exclusions in the cohort treatment 1 group and 19 in the cohort treatment 2 group.
Intervention and Co-interventions	***Trial*** Treatment 1: Surgery Treatment 2: Medical treatment ***Cohort*** Treatment 1: Same Treatment 2: Same There was a significant beneficial effect in favor of the trial treatment.
Care provider and setting	Surgeons provided care in the trial and cohort treatment 1 groups. General practitioners provided care in the trial and cohort treatment 2 groups. All patients were treated at the otolaryngology department.
Outcomes	Mean episodes of sore throat per month. At 2 years.

Author	**Luby 2002**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients lived in another colony, so were used as a control group. Intention-to-treat analysis was used in the trial.
Selection Bias	There were differences in baseline characteristics between groups, but they were controlled for in the analysis.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	No post-randomization exclusions.
Intervention and Co-interventions	***Trial*** Treatment: Medicated bar soap with 1.2 % triclocarban Control: Placebo soap ***Cohort*** Control: Standard habits and practices - provided with books, pens, pencils

	No significant difference between trial treatment and control groups.
Care provider and setting	Field workers and clinicians provided care in all groups. All patients were treated in their own households.
Outcomes	Total impetigo episodes. Maximum follow-up of 7161 person-weeks of observation.

Author	**Macdonald 2007**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients did not want active treatment. Intention-to-treat analysis was used in the trial.
Selection Bias	There were no differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	There were 2 inappropriate exclusions in the trial treatment 1 group and 1 in the trial treatment 2 group. In the trial treatment 3 group there was 1 appropriate exclusion and 2 inappropriate exclusions. In the cohort group there were 2 inappropriate exclusions.
Intervention and Co-interventions	***Trial*** Treatment 1: High dose of nandrolone decanoate (ND) Treatment 2: Low dose of ND Treatment 3: Medium dose of ND ***Cohort*** Control: No treatment There was a significant beneficial effect in favor of the trial treatment 1.
Care provider and setting	Dialysis providers provided care to all groups. All patients were treated at the same setting.
Outcomes	Mean appendicular lean mass post-treatment. At 24 weeks.

Author	**MacLennan 1985**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients were excluded from the trial due to administrative reasons (i.e. temporary absence of one of the researchers). Intention-to-treat analysis was not used in the trial.
Selection Bias	Most likely there were differences in baseline characteristics between groups. (N=169)
Detection Bias	Unclear whether follow-up was the same between groups.
Exclusion Bias	No post-randomization exclusions.
Intervention and Co-interventions	***Trial*** Treatment 1: Purified porcine relaxin Treatment 2: Placebo gel ***Cohort*** Control: No gel applied No significant difference between trial treatment and control groups.
Care provider and setting	IVF clinic staff provided care to all groups. All patients were treated at the IVF clinic.
Outcomes	Number of patients not pregnant. At 15 months.

Author	**MacMillan 1986**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients.

	Cohort patients were not enrolled in the trial either because of patient/relative/physician refusal, precipitous discharge, unsuitable medications, departure from the area or a combination of reasons. Intention-to-treat analysis was not used in the trial.
Selection Bias	There were differences in baseline characteristics between groups.
Detection Bias	More frequent visits for those in RCT.
Exclusion Bias	There were 6 inappropriate exclusions in the trial treatment group, 7 in the trial control group, and 67 in the cohort control group.
Intervention and Co-interventions	***Trial*** Treatment: Active study medication Control: Placebo ***Cohort*** Control: Standard care medications There was a significant beneficial effect in favor of the trial treatment.
Care provider and setting	Unclear who provided treatment in any of the groups. Unclear where patients were treated.
Outcomes	Number of patients who relapsed.

Author	**Mahon 1996**
Design and Methods	Patients were randomized either to a Nof1 randomized trial or a cohort study. Cohort patients were randomized to standard practice. Intention-to-treat analysis was not used in the trial.
Selection Bias	Most likely there was a difference in baseline characteristics between groups. (N= 31)
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	There were no post- randomization exclusions.
Intervention and Co-interventions	***Trial*** Treatment 1: Theophylline Treatment 2: Placebo ***Cohort*** Control: Standard practice Unclear whether there was a significant difference between trial treatments.
Care provider and setting	Unclear who provided care in the trial and cohort groups. All patients were treated at the tertiary care centre.
Outcomes	Number still taking theophylline. At 6 months.

Author	**Mahon 1999**
Design and Methods	Patients were randomized either to a Nof1 randomized trial or a cohort study. Cohort patients were randomized to standard practice. Intention-to-treat analysis was used in the trial.
Selection Bias	There were no differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	There were 6 inappropriate exclusions in the trial. There was 8 inappropriate exclusions in the cohort group.
Intervention and Co-interventions	***Trial*** Treatment 1: Theophylline Treatment 2: Placebo ***Cohort***

	Control: Standard practice Unclear whether there was a significant difference between trial treatments.
Care provider and setting	A primary care physician provided care in the trial and cohort group. All patients were treated at a primary care setting.
Outcomes	Number still taking theophylline. At 12 months.

Author	**Marcinczyk 1997**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients either refused participation or physician did not refer patient to the study. Intention-to-treat analysis was not used in the trial.
Selection Bias	There were no differences in baseline characteristics between groups.
Detection Bias	Unclear whether there was a difference in follow-up.
Exclusion Bias	No post-randomization exclusions.
Intervention and Co-interventions	***Trial*** Treatment: Carotid endarterectomy Control: Unclear ***Cohort*** Treatment: Same Unclear whether there was a significant difference between trial treatment and control.
Care provider and setting	Trial trained surgeons provided care in the trial, non-trial participating surgeons provided care outside trial. All patients were treated at the same hospital.
Outcomes	Mortality during hospital stay.

Author	**Martin 1994**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients refused participation. Intention-to-treat analysis was used in the trial.
Selection Bias	There were no differences in baseline characteristics between groups.
Detection Bias	More frequent visits for those in trial.
Exclusion Bias	No post-randomization exclusions.
Intervention and Co-interventions	***Trial*** Treatment 1: 200 mcg misoprostol and a placebo liquid antacid Treatment 2: Magnesium- aluminum hydroxide liquid antacid (Maalox) ***Cohort*** Control: Cimetidine 300 mg intravenously No significant difference between trial treatments.
Care provider and setting	GI specialists and ICU staff provided care in all groups. All patients were treated at the same tertiary care centre.
Outcomes	Hemorrhage from gastric lesions. At 3 days.

Author	**Martinez- Amenos 1990**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients refused participation. Intention-to-treat analysis was not used in the trial.
Selection Bias	There were differences in baseline characteristics between groups.

Detection Bias	Follow-up was the same between groups.
Exclusion Bias	No post-randomization exclusions.
Intervention and Co-interventions	***Trial*** Treatment 1: Team education Treatment 2: Individual education Control: No further education after initial assessment ***Cohort*** Treatment 2: Same Control: Same There was a significant beneficial effect in favor of the trial treatment 1.
Care provider and setting	The same primary care providers provided care in all groups. All patients were treated at the same primary care centre.
Outcomes	Number with blood pressure > 160/95. At 2 months.

Author	**Masood 2002**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients refused participation. Intention-to-treat analysis was used in the trial.
Selection Bias	Most likely there were differences in baseline characteristics between groups. (N= 110)
Detection Bias	More frequent visits for those in RCT.
Exclusion Bias	No post-randomization exclusions.
Intervention and Co-interventions	***Trial*** Treatment: Entonox via a breath activated device Control: Air ***Cohort*** Control: No treatment There was a significant beneficial effect in favor of the trial treatment.
Care provider and setting	An anesthesiologist provided care in all groups. All patients were treated at the same urology department.
Outcomes	Mean pain score (visual pain analog scale). At 30 minutes post-operatively.

Author	**Matilla 2003**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients refused randomization. Intention-to-treat analysis was not used in the trial.
Selection Bias	There were no differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	There were 3 inappropriate exclusions in the cohort treatment group.
Intervention and Co-interventions	***Trial*** Treatment: Tympanostomy with adenoidectomy Control: Tympanostomy without adenoidectomy ***Cohort*** Treatment: Same Control: Same No significant difference between trial treatment and control groups.
Care provider and setting	Surgeons and trained study physicians provided care in all groups. All

	patients were treated at the same study clinic.
Outcomes	Rates of otitis media episodes. At 7 months.

Author	**Mayo Asymptomatic Carotid Endarterectomy Study Group 1992**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Unclear the reason cohort patients did not enroll. Intention-to-treat analysis was used in the trial.
Selection Bias	There were no differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	No post-randomization exclusions.
Intervention and Co-interventions	***Trial*** Treatment 1: Aspirin 80mg/day orally Treatment 2: Carotid arteriography and endarterectomy ***Cohort*** Treatment 1: Same Treatment 2: Same Unclear whether there was a significant difference between trial treatments.
Care provider and setting	A surgeon provided care in the trial treatment 2 group. Unclear who provided care in all other groups. Unclear where patients were treated.
Outcomes	Incidence of transient ischemic attacks. Patients were followed-up for a mean of 23.6 months.
Note	Additional information gathered from: Mayo Asymptomatic Carotid Endarterectomy Study Group. Effectiveness of carotid endarterectomy for asymptomatic carotid stenosis: design of a clinical trial. Mayo Clin Proc. 64:897-904, 1989

Author	**McCaughey 1998**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients refused randomization. Intention-to-treat analysis was not used in the trial.
Selection Bias	There were no differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	There were 3 inappropriate exclusions in the trial treatment group, 2 in the trial control group and 3 in the non-trial group.
Intervention and Co-interventions	***Trial*** Treatment: Growth hormone Control: Intensive monitoring without treatment ***Cohort*** Control: No treatment There was a significant beneficial effect in favor of the trial treatment.
Care provider and setting	Unclear who provided care in any of the groups. All patients were treated at the same hospital.
Outcomes	Mean near- final- height. At 7 years.

Author	**McKay 1998**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients refused randomization. Intention-to-treat analysis was not used in the trial.

Selection Bias	There were differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	There were 6 inappropriate exclusions in the trial treatment group, 8 in the trial control group, 4 in the cohort treatment group and 1 in the cohort control group.
Intervention and Co-interventions	***Trial*** Treatment 1: Day hospital Treatment 2: In-patient treatment ***Cohort*** Treatment 1: Same Treatment 2: Same No significant difference between trial treatments.
Care provider and setting	Senior psychiatrist, social workers, nurses and counselors provided care in all groups. All patients were treated at the neuropsychiatric hospital and addiction recovery unit.
Outcomes	Mean number of days of cocaine use. Within 1 year of treatment.

Author	**McKay 1995**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients refused randomization due to strong preference for one treatment. Intention-to-treat analysis was not used in the trial.
Selection Bias	There were differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	There were 4 inappropriate exclusions in the trial treatment 1 group, 4 in the trial treatment 2 group, 8 in the cohort treatment 1 group and 8 in the cohort treatment 2 group.
Intervention and Co-interventions	***Trial*** Treatment 1: Day hospital Treatment 2: In-patient treatment ***Cohort*** Treatment 1: Same Treatment 2: Same No significant difference between trial treatments.
Care provider and setting	The same individual provided care in all groups. All patients were treated at the addiction recovery unit.
Outcomes	Mean number of drinking days. Within 1 year of treatment.

Author	**Melchart 2002**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients were given the option of choosing their preferred treatment. Intention-to-treat analysis was not used in the trial.
Selection Bias	There were no differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	There were 6 inappropriate exclusions in the trial treatment group, 8 in the trial control group, 4 in the cohort treatment group and 1 in the cohort control group. No post-randomization exclusions.

Intervention and Co-interventions	***Trial*** Treatment 1: Sedation with intravenous midazolam Treatment 2: Acupuncture ***Cohort*** Treatment 1: Same Treatment 2: Same Control: No treatment No significant difference between trial treatments.
Care provider and setting	Consultant neurologist and an assistant physician provided care in the trial treatment 2 group. Unclear who provided care in the other groups. All patients were treated at the same hospital.
Outcomes	Number of patients who would not undergo the same treatment again. At 2 hours.

Author	**Moertel 1984**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients refused randomization. Intention-to-treat analysis was not used in the trial.
Selection Bias	There were no differences in baseline characteristics between groups.
Detection Bias	Unclear whether follow-up was the same between groups.
Exclusion Bias	No post-randomization exclusions.
Intervention and Co-interventions	***Trial*** Treatment: Combined 5-Fluorouracil and radiation therapy Control: No further treatment ***Cohort*** Control: Same There was a significant beneficial effect in favor of the trial treatment.
Care provider and setting	Unclear who provided care in any of the groups. Patients in the trial treatment group were treated at the hospital, unclear where the rest of the patients were treated.
Outcomes	Mortality at 5 years.

Author	**Mori 2006**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients had a preference for treatment. Intention-to-treat analysis was not used in the trial.
Selection Bias	Most likely there were no differences in baseline characteristics between groups. (N= 927)
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	There were 9 inappropriate exclusions in the trial treatment 1 group, 48 in the cohort treatment 1 group.
Intervention and Co-interventions	***Trial*** Treatment 1: Nasal esophagogastroduodenoscopy Treatment 2: Oral esophagogastroduodenoscopy ***Cohort*** Treatment 1: Same Treatment 2: Same

	There was a significant beneficial effect in favor of the trial treatment 1.
Care provider and setting	Qualified specialists provided care in all groups. All patients were treated at the same hospital.
Outcomes	Mean satisfaction score. Immediately after endoscopy.

Author	**Morrison 2002**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients refused randomization. Intention-to-treat analysis was not used in the trial.
Selection Bias	There were differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	No post-randomization exclusions.
Intervention and Co-interventions	***Trial*** Treatment 1: Coronary artery bypass graft surgery Treatment 2: Percutaneous coronary intervention ***Cohort*** Treatment 1: Same Treatment 2: Same No significant difference between trial treatments.
Care provider and setting	Unclear who provided care in any of the groups. All patients were treated at the same hospital sites.
Outcomes	Kaplan-Meier estimates of the numbers who are not free of unstable angina or repeat revascularization. At 3 years.

Author	**Nagel 1998**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients refused randomization and were given the choice of either treatment. Intention-to-treat analysis was used in the trial.
Selection Bias	There were no differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	No post-randomization exclusions.
Intervention and Co-interventions	***Trial*** Treatment 1: Transabdominal chorionic villus sampling Treatment 2: Early amniocentesis ***Cohort*** Treatment 1: Same Treatment 2: Same There was a significant beneficial effect in favor of the trial treatment 1.
Care provider and setting	Experienced operators provided care in all groups. All patients were treated at the same hospital.
Outcomes	Number of viable fetus losses during pregnancy.

Author	**Neldam 1986**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients refused participation. Intention-to-treat analysis was not used in the trial.
Selection Bias	There were no differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.

Exclusion Bias	No post-randomization exclusions.
Intervention and Co-interventions	***Trial*** Treatment 1: Electronic fetal monitoring cardiotocography (EFM) Treatment 2: Stethoscope (AUS) ***Cohort*** Treatment 1: Routine care- EFM for high risk cases, EFM/AUS in all others No significant difference between trial treatments.
Care provider and setting	The same obstetricians provided care in all of the groups. All patients were treated at the same hospital ward.
Outcomes	Number with an Apgar score of 0-3. At 5 minutes post birth.

Author	**Nicolaides 1994**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients had a preference for one treatment. Intention-to-treat analysis was not used in the trial.
Selection Bias	There were no differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	There were 1 inappropriate exclusion in the cohort treatment group.
Intervention and Co-interventions	***Trial*** Treatment: Early amniocentesis Control: Chorionic villus sampling ***Cohort*** Treatment: Same Control: Same There was a significant beneficial effect in favor of the control treatment.
Care provider and setting	A specialist in fetal medicine, or a research registrar under his supervision provided care in all groups. All patients were treated at the same research centre.
Outcomes	Number of spontaneous fetal deaths.

Author	**Ogden 2004**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients were treated first and used to train the study physicians. Intention-to-treat analysis was not used in the trial.
Selection Bias	There were no differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	There were 4 inappropriate exclusions in all groups.
Intervention and Co-interventions	***Trial*** Treatment: High energy electrohydraulic shock wave treatment Control: Placebo "shock" wave treatment ***Cohort*** Treatment: Same There was a significant beneficial effect in favor of the trial treatment.
Care provider and setting	Study physician provided care in the trial groups, but a different professional provided care for the non-trial group. Unclear where patients were treated.
Outcomes	Treatment not considered a success. At 3 months.

Author	**Palmon 1996**
Design and Methods	Randomized trial and a separate cohort of eligible patients treated within two months of the follow-up of the trial. The cohort group was added to determine if end-tidal CO_2 could be more tightly controlled in the presence of the monitor. Intention-to-treat analysis was not used in the trial.
Selection Bias	There were differences in baseline characteristics between groups.
Detection Bias	Unclear whether follow-up was the same between groups.
Exclusion Bias	No post-randomization exclusions.
Intervention and Co-interventions	***Trial*** Treatment 1: No monitor Treatment 2: Monitor-blind ***Cohort*** Treatment 1: Monitor was given No significant difference between trial treatments.
Care provider and setting	Same junior and senior anesthesia residents provided care in all groups. All patients were treated at the same neuro-radiology centre.
Outcomes	Mean $PaCO_2$.

Author	**Panagopoulou 2009**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients refused to participate. Intention-to-treat analysis was not used in the trial.
Selection Bias	There were no differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	No post-randomization exclusions.
Intervention and Co-interventions	***Trial*** Treatment 1: Emotional-writing condition (in their diaries) Treatment 2: Fact-writing condition (in their diaries) Control: No treatment ***Cohort*** Control: Same No significant difference between trial treatments.
Care provider and setting	Unclear who provided care in any of the groups. All patients were treated at the same clinic.
Outcomes	Number not pregnant.

Author	**Paradise 1984**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients declined randomization, treatment was based on preference. Intention-to-treat analysis was not used in the trial.
Selection Bias	There were no differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	There were 21 inappropriate exclusions in the trial treatment group, 28 in the trial control group, 37 in the cohort treatment group and 31 in the cohort control group.
Intervention and Co-interventions	***Trial*** Treatment: Surgery (tonsillectomy with or without adenoidectomy) Control: No surgery

	Cohort Treatment: Same Control: Unclear if it was the exact same as the trial control There was a significant beneficial effect in favor of the trial treatment.
Care provider and setting	Pediatrician and pediatric nurse practitioner provided care in trial groups. Different professional provided care in the cohort group. All patients were treated at the same hospital.
Outcomes	Total episodes of throat infections. At 3 years.

Author	**Peteren 2007**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients refused randomization. Intention-to-treat analysis was used in the trial.
Selection Bias	There were differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	There were 3 inappropriate exclusions in the trial treatment group and 9 in the trial control group.
Intervention and Co-interventions	***Trial*** Treatment: Hip replacement fast track Control: Usual care ***Cohort*** Control: Same Unclear whether there was a significant difference between trial treatment and control.
Care provider and setting	Unclear who provided care in any of the groups. All patients were treated at the same hospital.
Outcomes	Number transferred to rehabilitation ward.

Author	**Raistrick 2005**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients wanted to choose their treatment. Intention-to-treat analysis was not used in the trial.
Selection Bias	There were no differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups..
Exclusion Bias	There were 15 inappropriate exclusions in the trial treatment group, 21 in the trial control group, 28 in the cohort treatment group and 18 in the cohort control group.
Intervention and Co-interventions	***Trial*** Treatment 1: Buprenorphine Treatment 2: Lofexidine ***Cohort*** Treatment 1: Same Treatment 2: Same No significant difference between trial treatments.
Care provider and setting	Experienced doctor initiated care, follow-up with either a nurse or doctor provided care in trial groups. Unclear who provided care in the cohort groups. All patients were treated at same addiction recovery unit.

Outcomes	Number of patients not abstinent at 1 month of follow-up.

Author	**Reddihough 1998**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients refused randomization. Intention-to-treat analysis was not used in the trial.
Selection Bias	There were no differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	There were 12 inappropriate exclusions in the trial group, and 13 in the cohort group.
Intervention and Co-interventions	***Trial*** Treatment 1: Conductive education (CE) based programme (average 75.6 hours of therapy) Treatment 2: Control (average 79.8 hours of therapy) ***Cohort*** Treatment 1: CE therapy (average 86.0 hours of therapy) Treatment 2: Control (average 59 hours of therapy) No significant difference between trial treatments.
Care provider and setting	Physiotherapist provided care in the trial. Unclear who provided care in the cohort. All patients were treated in the same setting.
Outcomes	Mean Gross Motor Function Measure (GMFM) score.

Author	**Rigg 2000**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Either cohort patients refused randomization or their physician did not enroll them in the study. Intention-to-treat analysis was not used in the trial.
Selection Bias	There were no differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	No post-randomization exclusions.
Intervention and Co-interventions	***Trial*** Treatment: Epidural block Control: No epidural ***Cohort*** Control: Unclear what treatment, if any, they received No significant difference between trial treatments.
Care provider and setting	Anesthesiologists provided care in all groups. All patients were treated at the same hospital.
Outcomes	Mortality at 30 days.

Author	**Rorbye 2005**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients refused randomization and were given the treatment of their preference. Intention-to-treat analysis was not used in the trial.
Selection Bias	There were differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	There were 29 inappropriate exclusions in the cohort treatment group and 166 in the cohort control group. Unclear if there were any losses in the trial group.
Intervention and Co-	***Trial***

interventions	Treatment: Medical abortion Control: Surgical abortion ***Cohort*** Treatment: Same Control: Same There was a significant beneficial effect in favor of the trial control.
Care provider and setting	Unclear who provided care in any of the groups. All patients were treated at the same hospital.
Outcomes	Number who were not "satisfied or very satisfied with the procedure". At 2 weeks.

Author	**Rosen 1987**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients refused randomization. Intention-to-treat analysis was not used in the trial.
Selection Bias	Most likely there were differences in baseline characteristics between groups. (N= 142)
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	No post-randomization exclusions.
Intervention and Co-interventions	***Trial*** Treatment 1: General anesthesia with nitrous oxide Treatment 2: General anesthesia without nitrous oxide ***Cohort*** Treatment 1: Same Treatment 2: Same No significant difference between trial treatments.
Care provider and setting	Unclear who provided care in any of the groups. All patients were treated at the same hospital.
Outcomes	Number of patients not pregnant after 1 cycle of in- vitro fertilization.

Author	**Salisbury 2002**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients were at schools not selected for the trial. Intention-to-treat analysis was used in the trial.
Selection Bias	There were no differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	There were 37 inappropriate exclusions in the trial treatment group, 15 in the trial control group and 12 in the cohort group.
Intervention and Co-interventions	***Trial*** Treatment: Nurse- run asthma clinic Control: Normal care in general practice ***Cohort*** Control: Same No significant difference between trial treatments.
Care provider and setting	Either a nurse or a doctor provided care in all groups. Patients in the trial control and cohort groups were treated in the same setting.
Outcomes	Mean "Paediatric Quality of Life Questionnaire Standard UK version" score.

	At 6 months.

Author	**Sesso 2002**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Unclear why cohort patients did not participate in the trial. Intention-to-treat analysis was not used in the trial.
Selection Bias	There were differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	No post-randomization exclusions.
Intervention and Co-interventions	***Trial*** Treatment: Aspirin Control: Beta-carotene placebo ***Cohort*** Unclear what treatment(s), if any, was given No significant difference between trial treatment and control.
Care provider and setting	Unclear who provided care in any of the groups. Unclear where patients were treated.
Outcomes	Mortality due to cardiovascular complications. Followed-up for an average of 5.39 years.

Author	**Shain 1989**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients wanted the experimental device. Intention-to-treat analysis was not used in the trial.
Selection Bias	There were differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	There were 16 inappropriate exclusions in the trial treatment group, 14 in the trial treatment 2 group and 12 in the cohort group.
Intervention and Co-interventions	***Trial*** Treatment 1: Nova-T intra-uterine device (IUD) Treatment 2: LNG-IUD ***Cohort*** Treatment 1: Intra-cervical device No significant difference between trial treatments.
Care provider and setting	Unclear who provided care in any of the groups. All patients were treated at the same clinic.
Outcomes	Number of patients who discontinued treatment. At 12 months.

Author	**Smith 1990**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients refused participation. Intention-to-treat analysis was used in the trial.
Selection Bias	There were differences in baseline characteristics between groups.
Detection Bias	Unclear whether follow-up was the same between groups.
Exclusion Bias	No post-randomization exclusions.
Intervention and Co-interventions	***Trial*** Treatment: Warfarin Control: Placebo

	Cohort Control: Unclear what was given There was a significant beneficial effect in favor of the trial treatment.
Care provider and setting	Cardiologists provided care in the treatment groups. Unclear who provided care in the cohort group. All patients were treated at the cardiology centre.
Outcomes	Mortality at a minimum follow-up of 2 years.

Author	**Smuts 2003**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients did not normally eat eggs. Intention-to-treat analysis was not used in the trial.
Selection Bias	There were no differences in baseline characteristics between groups.
Detection Bias	More frequent visits for those in RCT.
Exclusion Bias	No post-randomization exclusions.
Intervention and Co-interventions	***Trial*** Treatment 1: Omega Tech Eggs Treatment 2: Regular Eggs ***Cohort*** Treatment 1: Low egg intake No significant difference between trial treatments.
Care provider and setting	Clinical nurses provided care in all trial groups. Unclear who treated the cohort group. All patients were treated at the same setting.
Outcomes	Mean docosahexaenoic acid (DHA) level in plasma TAG. Measured at the third trimester.

Author	**Stecksen-blicks 2008**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients refused randomization. Intention-to-treat analysis was not used in the trial.
Selection Bias	There were no differences in baseline characteristics between groups.
Detection Bias	Unclear whether follow-up was the same between groups.
Exclusion Bias	There were 24 inappropriate exclusions in the trial treatment 1 group, 21 in the trial treatment 2 group, and 6 in the cohort control group.
Intervention and Co-interventions	***Trial*** Treatment 1: Xylitol lozenges Treatment 2: Xylitol/fluoride-containing lozenges ***Cohort*** Control: Conventional care was offered No significant difference between trial treatments.
Care provider and setting	Dentists provided care in all groups. All patients were treated at the same clinic.
Outcomes	Mean total proximal caries prevalence. Within 2 years.

Author	**Stern 2003**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Unclear why cohort patients were not enrolled in the trial. Intention-to-treat analysis was not used in the trial.

Selection Bias	There were no differences in baseline characteristics between groups.
Detection Bias	Unclear whether follow-up was the same between groups.
Exclusion Bias	No post-randomization exclusions.
Intervention and Co-interventions	***Trial*** Treatment: Heparin and aspirin Control: Placebo ***Cohort*** Control: Unclear what treatment, if any, they received No significant difference between trial treatment and control.
Care provider and setting	Unclear who provided care in any of the groups. All patients were treated at the same hospital.
Outcomes	Fetal heart implantation considered unsuccessful.

Author	**Stith 2004**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients were unable to attend treatment due to scheduling problems, or they no longer wanted to participate. Intention-to-treat analysis was not used in the trial.
Selection Bias	There were differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	There were 12 inappropriate exclusions in the trial treatment group, 11 in the trial control group, and 5 in the cohort group.
Intervention and Co-interventions	***Trial*** Treatment 1: Met therapist alone for co-therapy Treatment 2: Multi-couple group co-therapy ***Cohort*** Control: No treatment There was a significant beneficial effect in favor of the trial treatment 1.
Care provider and setting	Therapist provided care in the trial groups. Different professional provided care in the cohort. Patients in trial and cohort groups were treated in different settings.
Outcomes	Recidivism rates at 2 years.

Author	**Stockton 2009**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients received treatment based on their preference. Intention-to-treat analysis was used in the trial.
Selection Bias	Most likely there were differences in baseline characteristics between groups. (N= 78)
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	There were 4 inappropriate exclusions in the trial treatment 1 group and 5 in the trial treatment 2 group.
Intervention and Co-interventions	***Trial*** Treatment 1: Twice daily physiotherapy Treatment 2: Once daily physiotherapy ***Cohort*** Treatment 1: Participants chose hydrotherapy

	No significant difference between trial treatments.
Care provider and setting	Physiotherapists provided care in trial groups. Different physiotherapists provided care in the cohort group. All patients were treated at the same hospital.
Outcomes	Mean Iowa level of assistance score. At 6 days.

Author	**Strandberg 1995**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients refused randomization. Intention-to-treat analysis was not used in the trial.
Selection Bias	There were no differences in baseline characteristics between groups.
Detection Bias	More frequent visits for those in RCT.
Exclusion Bias	There were 158 inappropriate exclusions in the trial treatment group, 154 in the trial control group and 104 in the cohort group.
Intervention and Co-interventions	***Trial*** Treatment: Health checks Control: No health checks ***Cohort*** Treatment: Unclear There was a significant beneficial effect in favor of the trial control.
Care provider and setting	Unclear who provided care in any of the groups. Only trial patients were treated at the institute.
Outcomes	Mortality at 18 years.

Author	**Suherman 1999**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients volunteered to form a non-randomized control group. Intention-to-treat analysis was not used in the trial.
Selection Bias	There were no differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	There were 3 inappropriate exclusions in the trial treatment 1 group, 4 in the trial treatment 2 group and 16 in the cohort group.
Intervention and Co-interventions	***Trial*** Treatment 1: Implanon (single-rod contraceptive implant) Treatment 2: Norplant (six-rod implant) ***Cohort*** Treatment 1: Intra-uterine device No significant difference between trial treatments.
Care provider and setting	Unclear who provided care in any of the groups. Unclear where patients were treated.
Outcomes	Mean apolipoprotein AI concentration. At 2 years.

Author	**Sullivan 1982**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Unclear the reason cohort patients were not enrolled in the trial. Intention-to-treat analysis was not used in the trial.
Selection Bias	Most likely there were no differences in baseline characteristics between groups. (N= 269)

Detection Bias	Follow-up was the same between groups.
Exclusion Bias	There were 15 inappropriate exclusions in the trial treatment 1 group, 43 in the trial treatment 2 group and 21 in the trial treatment 3 group. There were 8 in the cohort treatment 1 group, 3 in the cohort treatment 2 group and 10 in the cohort treatment 3 group.
Intervention and Co-interventions	***Trial*** Treatment 1: Involved- field (IF) radiotherapy Treatment 2: IF radiotherapy and MOPP (Mechlorethamine, Oncovin, Procarbazine, Prednisone) chemotherapy Treatment 3: Extended field (EF) radiotherapy ***Cohort*** Treatment 1: Same Treatment 2: Same Treatment 3: Same Unclear whether there was a significant difference between trial treatments.
Care provider and setting	Unclear who provided care in any of the groups. Unclear where patients were treated.
Outcomes	Relapse after complete or partial remission.

Author	**Sundar 2008**
Design and Methods	Randomized trial and a separate cohort of eligible patients treated within two months of the trial follow-up. Cohort patients were treated non-randomly with the trial treatment because the trial patients were responding favorably to the treatment. Intention-to-treat analysis was used in the trial.
Selection Bias	There were differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	No post-randomization exclusions.
Intervention and Co-interventions	***Trial*** Treatment 1: 5 mg/kg active parenteral agent liposomal amphotericin B (L-AmB) once Treatment 2: 5 mg/kg L-AmB once + miltefosine 10 days Treatment 3: 5mg/kg L-AmB once + miltefosine 14 days Treatment 4: 3.75mg/kg L-Amb once + miltefosine 14 days ***Cohort*** Treatment 1: 5 mg/kg L-AmB once + miltefosine 7 days No significant difference between trial treatments.
Care provider and setting	Unclear who provided care in any of the groups. All patients were treated at the same inpatient unit.
Outcomes	Number of patients not cured. At 16 days after treatment.

Author	**Taddio 2006**
Design and Methods	Randomized trial and a separate cohort of eligible patients treated within two months of the trial follow-up. Cohort patients' parents refused to give consent for study drugs. Intention-to-treat analysis was used in the trial.
Selection Bias	There were no differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	There were 4 inappropriate exclusions in the trial treatment 1 group, 3 in the trial treatment 2 group, 6 in the trial treatment 3 group. There was 1

	inappropriate exclusion in the cohort control group.
Intervention and Co-interventions	***Trial*** Treatment 1: 0.5 g of tetracaine 4% gel applied to the insertion site Treatment 2: 0.1 mg/kg of intravenous morphine Treatment 3: 0.5 g of tetracaine 4% gel and 0.1 mg/kg of intravenous morphine ***Cohort*** Control: No treatment There was a significant beneficial effect in favor of the trial treatment 3.
Care provider and setting	A bedside nurse provided care in all trial groups. Cohort patients did not receive treatment. All patients were treated at the same hospital.
Outcomes	Mean proportion of time brow bulge observed during procedure.

Author	**Tanai 2009**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients refused to participate. Intention-to-treat analysis was not used in the trial.
Selection Bias	There were no differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	No post-randomization exclusions.
Intervention and Co-interventions	***Only included Trial 1 from this study because Trial 2 included a patient who was initially randomized.*** ***Trial*** Treatment 1: Cisplatin-irinotecan Treatment 2: Carboplatin-paclitaxel Treatment 3: Cisplatin-gemcitabine Treatment 4:Cisplatin-vinorelbine ***Cohort*** Treatment 1-4: Same No significant difference between trial treatments.
Care provider and setting	Physicians provided care in all trial groups. Unclear who provided care in the cohort groups. All patients were treated at the same hospital.
Outcomes	Mortality at 2 years.

Author	**Tanaka 1994**
Design and Methods	Randomized trial and a separate cohort of eligible patients treated within two months of the trial follow-up. Cohort patients were not enrolled in the trial to test whether systemic lidocaine was affecting the isoflurane level. Intention-to-treat analysis was not used in the trial.
Selection Bias	There were no differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	No post-randomization exclusions.
Intervention and Co-interventions	***Trial*** Treatment 1: Nasal lidocaine Treatment 2: Clonidine premedication Control: No treatment ***Cohort*** Treatment 1: Intravenous lidocaine

	There was a significant harmful effect associated with the trial control.
Care provider and setting	Unclear whether similar professionals provided care in the trial and cohort groups. Unclear where patients were treated.
Outcomes	Heart rate during the procedure.

Author	**Taplin 1986**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients were designated as a non-randomized positive control group. Intention-to-treat analysis was not used in the trial.
Selection Bias	There were no differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	No post-randomization exclusions.
Intervention and Co-interventions	***Trial*** Treatment: Permethrin 1% creme rinse Control: Placebo ***Cohort*** Treatment: 1% lindane shampoo (Kwell) There was a significant beneficial effect in favor of the trial treatment.
Care provider and setting	Designated member of the research team provided care in all groups. All patients were treated at the same setting.
Outcomes	Number of patients that are not free of lice and viable nits. At 14 days.

Author	**Tenenbaum 2002**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients refused randomization. Intention-to-treat analysis was not used in the trial.
Selection Bias	There were differences in baseline characteristics between groups.
Detection Bias	More frequent visits for those in RCT.
Exclusion Bias	No post-randomization exclusions.
Intervention and Co-interventions	***Trial*** Treatment: Bezafibrate 400 mg/day Control: Placebo ***Cohort*** Control: Community based treatment Unclear whether there was a significant difference between trial treatment and control.
Care provider and setting	Unclear who provided care in any of the groups. Patients in the study were treated in the clinic, and cohort patients were treated in the community.
Outcomes	Mortality within 9 years of study.

Author	**Toprak 2005**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients did not enroll in the trial because they either had benign breast or ovarian cancer, or were not willing to use hormone therapy. Intention-to-treat analysis was not used in the trial.
Selection Bias	There were no differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.

Exclusion Bias	No post-randomization exclusions.
Intervention and Co-interventions	***Trial*** Treatment: 5mg of folic acid + daily with 0.625 mg conjugated equine estrogen (CEE), continuously combined with 2.5 mg medroxyprogesterone acetate daily for 12 weeks Control: Placebo daily + 0.625 mg CEE, continuously combined with 2.5 mg medroxyprogesterone acetate daily for 12 weeks ***Cohort*** Control: No hormone therapy (or any other study medication) There was a significant beneficial effect in favor of the trial treatment.
Care provider and setting	Unclear who provided care in any of the groups. All patients were treated at the same clinic.
Outcomes	Mean serum homocysteine levels at 12 weeks.

Author	**Underwood 2008**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients had a preference for one treatment. Intention-to-treat analysis was used in the trial.
Selection Bias	There were differences in baseline characteristics between groups, but these were controlled for statistically in the analysis.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	There were 45 inappropriate exclusions in the trial treatment 1 group, 50 in the trial treatment 2 group, 26 in the cohort treatment 1 group and 6 in the cohort treatment 2 group.
Intervention and Co-interventions	***Trial*** Treatment 1: Topical ibuprofen Treatment 2: Oral ibuprofen ***Cohort*** Treatment 1: Same Treatment 2: Same No significant difference between trial treatments.
Care provider and setting	Either a GP or nurse provided care in all groups. All patients were treated at the similar primary care setting.
Outcomes	Mean WOMAC (Global) score. At 24 months.

Author	**Urban 1999**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients either refused to participate or their physician refused to enroll them in the trial. Intention-to-treat analysis was used in the trial.
Selection Bias	Most likely there were differences in baseline characteristics between groups. (N= 103)
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	No post-randomization exclusions.
Intervention and Co-interventions	***Trial*** Treatment 1: Early invasive group- PTCA or CABG were attempted if considered feasible Treatment 2: Early conservative group - did not undergo immediate coronary angiography

	Cohort Treatment 1: Same No significant difference between trial treatments.
Care provider and setting	Unclear who provided care in any of the groups. All patients were treated at the same centre.
Outcomes	Mortality at 30 days.

Author	**van Bergen 1995**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Unclear the reason cohort patients did not enroll in the trial. Intention-to-treat analysis was not used in the trial.
Selection Bias	There were differences in baseline characteristics between groups.
Detection Bias	More frequent visits for those in RCT.
Exclusion Bias	No post-randomization exclusions.
Intervention and Co-interventions	***Trial*** Treatment 1: Coumarin derivative Control: Placebo ***Cohort*** Control: Standard of care No significant difference between trial treatment and control.
Care provider and setting	Unclear who provided care in any of the groups. All patients treated at the same centre.
Outcomes	Mortality at 5 years.
Notes	Additional information extracted from: ASPECT Research Group. "Effect of long-term oral anticoagulant treatment on mortality and cardiovascular morbidity after myocardial infarction. Anticoagulants in the Secondary Prevention of Events in Coronary Thrombosis" Lancet: 1994 vol.343 iss.8896 pg.499 -503

Author	**Van 2009**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients had a preference for treatment. Intention-to-treat analysis was not used in the trial.
Selection Bias	There were no differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	There were 19 inappropriate exclusions in the trial groups and 15 in the cohort groups.
Intervention and Co-interventions	***The second treatment in the trial was not included in the analysis because the study did not provide outcomes for this arm due to a low enrollment rate.*** ***Trial*** Treatment 1: Supportive psychotherapy ***Cohort*** Treatment 1: Same Unclear whether there was a significant difference between trial treatments.
Care provider and setting	Either a trained psychiatrists or psychotherapists provided care in the trial groups. Unclear who provided care in the cohort groups. All patients were

	treated in the same setting.
Outcomes	Number who did not have at least a 50% reduction in the Hamilton Depression Rating Scale Score. At 24 weeks.
Note	Additional information extracted from: Dekker, J., Koelen, J. A., Van, H. L., Schoevers, R. A., Peen, J., Hendriksen, M., et al. (2008). Speed of action: The efficacy of short-term psychodynamic supportive psychotherapy versus pharmacotherapy in the treatment of depression.

Author	**Verdonck 1995**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients either refused to participate, did not participate due to psychological reasons, medical reasons or administrative errors. Intention-to-treat analysis was used in the trial.
Selection Bias	Most likely there were differences in baseline characteristics between groups. (N= 106)
Detection Bias	Unclear whether follow-up was the same between groups.
Exclusion Bias	No post-randomization exclusions.
Intervention and Co-interventions	***Trial*** Treatment 1: Additional five courses of CHOP (cyclophosphamide, doxorubicin, vincristine, prednisone) Treatment 2: High-dose chemoradiotherapy and autologous bone marrow transplantation ***Cohort*** Control: Unclear what treatment, if any, they received No significant difference between trial treatments.
Care provider and setting	Unclear who provided care in any of the groups. Unclear where patients were treated.
Outcomes	Number who are not in remission.

Author	**Vind 2009**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients refused participation either because they considered themselves too healthy, felt they were too ill/frail, have to care for a sick spouse, are too busy, or were not interested in visiting the hospital. Intention-to-treat analysis was used in the trial.
Selection Bias	There were differences in baseline characteristics between groups.
Detection Bias	More frequent visits for those in RCT.
Exclusion Bias	There were 136 inappropriate exclusions in the trial groups, and 150 in the cohort group.
Intervention and Co-interventions	***Trial*** Treatment: Multi-factorial fall prevention Control: Usual care ***Cohort*** Control: Unclear what treatment, if any, they received. Unclear whether there was a significant difference between trial treatment and control.
Care provider and setting	Either a doctor, nurse, or physical therapist provided care in the trial treatment group. Unclear who provided care in any of the other groups. Unclear where

	cohort patients were treated.
Outcomes	Mortality at 6 months.
Note	Additional information extracted from: "Vind AB, Andersen HE, Pedersen KD et al. An outpatient multifactorial falls prevention intervention does not reduce falls in high-risk elderly Danes. J Am Geriatr Soc 2009;57:971–977."

Author	**Walker 1986**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients were not randomized in order to assess the rate of wound colonization (if any) resulting from the control saline infiltration. Intention-to-treat analysis was not used in the trial.
Selection Bias	Most likely there were differences in baseline characteristics between groups. (N= 137)
Detection Bias	Unclear whether follow-up was the same between groups.
Exclusion Bias	There were 2 inappropriate exclusions in the trial treatment group.
Intervention and Co-interventions	***Trial*** Treatment: Cefuroxime sodium (750mg) with 20mL of normal saline Control: Normal saline ***Cohort*** Control: Conventional prophylactic regimen There was a significant beneficial effect in favor of the trial treatment.
Care provider and setting	Surgeons provided care in all treatment groups. Unclear who provided care in the cohort group. All patients were treated in the same setting.
Outcomes	Frequency of wound colonization.

Author	**Wallage 2003**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients had a preference for one type of treatment. Intention-to-treat analysis was used in the trial.
Selection Bias	There were no differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	There were 3 inappropriate exclusions in the trial treatment 1 group and 10 inappropriate exclusions in the trial treatment 2 group. There were 4 inappropriate exclusions in the cohort treatment 1 group and 4 in the cohort treatment 2 group.
Intervention and Co-interventions	***Trial*** Treatment 1: General anesthesia Treatment 2: Local anesthesia ***Cohort*** Treatment 1: Same Treatment 2: Same No clinically important difference between trial treatments.
Care provider and setting	An anesthesiologist provided care in the trial groups. A different anesthesiologist cared for the cohort patients. All patients were treated at the same hospital.
Outcomes	Number who felt the anesthesia was not acceptable after surgery.

Author	**Watzke 2010**
Design and Methods	Patients were randomized either to a randomized trial or a cohort study. Intention-to-treat analysis was used in the trial.
Selection Bias	There were differences in baseline characteristics between groups, but these were statistically controlled for in the analysis.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	There were 24 inappropriate exclusions in the trial treatment 1 group and 18 inappropriate exclusions in the trial treatment 2 group. There were 14 inappropriate exclusions in the cohort treatment 1 group and 9 in the cohort treatment 2 group.
Intervention and Co-interventions	***Trial*** Treatment 1: Cognitive-behavioural therapy Treatment 2: Psychodynamic therapy ***Cohort*** Treatment 1: Same Treatment 2: Same Unclear whether there was a significant difference between trial treatments.
Care provider and setting	Therapists provided care in the trial groups. Probably different person providing care in the cohort groups. All patients were treated at the same in-patient unit.
Outcomes	Marginal mean of the General Severity Index score (short version of the Symptom Checklist 90 revised). At 6 months.

Author	**Welt 1981**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients refused randomization because of a preference. Intention-to-treat analysis was not used in the trial.
Selection Bias	There were differences in baseline characteristics between groups.
Detection Bias	Unclear whether follow-up was the same between groups.
Exclusion Bias	No post-randomization exclusions.
Intervention and Co-interventions	***Trial*** Treatment: Antihypertensive drug therapy Control: Placebo ***Cohort*** Treatment: Same Control: Usual care No significant difference between trial treatment and control.
Care provider and setting	Neonatologist provided care in the trial groups. Different person providing care in the cohort groups. Cohort patients were treated at a different setting than the trial patients.
Outcomes	Number of preeclampsia events.

Author	**West 2005**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients were not enrolled in the trial in order to identify, if present, a "trial effect". Intention-to-treat analysis was not used in the trial.
Selection Bias	Most likely there were no differences in baseline characteristics between groups. (N= 408)

Detection Bias	More frequent visits for those in RCT.
Exclusion Bias	No post-randomization exclusions.
Intervention and Co-interventions	***Trial*** Treatment: Magnesium sulphate Control: Placebo (saline) ***Cohort*** Control: Usual care Unclear whether there was a significant difference between trial treatment and control.
Care provider and setting	Unclear who provided care in any of the groups or in which setting. However, it was the same person and same setting across groups.
Outcomes	Number of admissions to the ICU.

Author	**Wetzner 1979**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Unclear why cohort patients were not enrolled in the trial. Intention-to-treat analysis was not used in the trial.
Selection Bias	There were differences in baseline characteristics between groups.
Detection Bias	Less frequent visits for those in the RCT.
Exclusion Bias	No post-randomization exclusions.
Intervention and Co-interventions	***Trial*** Treatment: Intramuscular dose of ceruletide as an adjunct to oral cholecystography Control: Fatty meal assisted cholecystography ***Cohort*** Treatment: Same There was a significant beneficial effect in favor of the trial treatment.
Care provider and setting	Unclear who provided care in any of the groups. Unclear where patients were treated.
Outcomes	Number who demonstrated contraction with greater than 20%-40% reduction. At 20 minutes.

Author	**Wieringa- de Waard 2002**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients refused randomization and received treatment of preference. Intention-to-treat analysis was used in the trial.
Selection Bias	There were differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	There were 12 inappropriate exclusions in each of the cohort treatment 1 and treatment 2 groups.
Intervention and Co-interventions	***Trial*** Treatment 1: Expectant management Treatment 2: Surgical evacuation (curettage) ***Cohort*** Treatment 1: Same Treatment 2: Same There was a significant beneficial effect in favor of the trial treatment 2.

Care provider and setting	An attending physician provided care in all groups. All patients were treated at the same clinic.
Outcomes	Unsuccessful treatment at 6 weeks.

Author	**Williford 1993**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients refused randomization. Intention-to-treat analysis was not used in the trial.
Selection Bias	There were differences in baseline characteristics between groups.
Detection Bias	More frequent visits for those in RCT.
Exclusion Bias	There was only 1 exclusion in the trial group for an unknown reason.
Intervention and Co-interventions	***Trial*** Treatment: Total parenteral nutrition Control: No treatment ***Cohort*** Treatment: Unclear what treatment, if any, they received No significant difference between trial treatment and control.
Care provider and setting	Unclear who provided care in any of the groups. Unclear where patients were treated.
Outcomes	Septic complications. At 90 days.

Author	**Witt 2006a**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients refused randomization. Intention-to-treat analysis was used in the trial.
Selection Bias	There were differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	There were 83 inappropriate exclusions in the trial treatment group, 86 in the trial control group and 440 in the cohort treatment group.
Intervention and Co-interventions	***Trial*** Treatment: Acupunture Control: Usual medical care ***Cohort*** Treatment: Same There was a significant beneficial effect in favor of the trial treatment.
Care provider and setting	A different physician, but who was similarly trained, provided care in all groups. Unclear where patients were treated.
Outcomes	Mean WOMAC All Index Score. At 3 months.

Author	**Witt 2006b**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients refused randomization. Intention-to-treat analysis was used in the trial.
Selection Bias	There were differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	There were 190 inappropriate exclusions and 127 appropriate exclusions in the trial treatment group. There were 225 inappropriate exclusions and 188 appropriate exclusions in the trial control group. There were 5709

	inappropriate exclusions in the cohort control group.
Intervention and Co-interventions	***Trial*** Treatment: Immediate acupuncture with routine care Control: Delayed acupuncture after 3 months ***Cohort*** Control: Same There was a significant beneficial effect in favor of the trial treatment.
Care provider and setting	A physician with a certification in acupuncture provided care in all groups. All patients were treated in the same setting.
Outcomes	Mean neck pain and disability score. At 6 months.

Author	**Witt 2006c**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients refused randomization. Intention-to-treat analysis was not used in the trial.
Selection Bias	There were differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	There were 130 inappropriate exclusions and 98 appropriate exclusions in the trial treatment group. There were 193 inappropriate exclusions and 154 appropriate exclusions in the trial control group. There were 4636 inappropriate exclusions in the cohort control group.
Intervention and Co-interventions	***Trial*** Treatment: Immediate acupuncture with routine care Control: Delayed acupuncture after 3 months ***Cohort*** Control: Same There was a significant beneficial effect in favor of the trial treatment.
Care provider and setting	A physician with a certification in acupuncture provided care in all groups. All patients were treated in the same setting.
Outcomes	Mean back pain. At 6 months.

Author	**Witt 2008**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients refused randomization. Intention-to-treat analysis was not used in the trial.
Selection Bias	There were differences in baseline characteristics between groups, but these were controlled for statistically during the analysis.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	There were 8 inappropriate exclusions and 3 appropriate exclusions in the trial treatment group. There were 8 inappropriate exclusions and 4 appropriate exclusions in the trial control group. There were 59 inappropriate exclusions in the cohort control group.
Intervention and Co-interventions	***Trial*** Treatment: Immediate acupuncture with routine care Control: Routine care, delayed acupuncture by 3 months ***Cohort*** Control: Same

	There was a significant beneficial effect in favor of the trial treatment.
Care provider and setting	A physician with a certification in acupuncture provided care in all groups. All patients were treated in the same setting.
Outcomes	Mean pain intensity during the last menstruation before assessment. At 3 months.

Author	**Woodhouse 1995**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients were not enrolled in the trial because the authors wanted a non-randomized control group since they were uncomfortable with giving patients placebo treatment. Intention-to-treat analysis was used in the trial.
Selection Bias	There were differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	No post-randomization exclusions.
Intervention and Co-interventions	***Trial*** Treatment: 10 mg adrenaline Control: Placebo (saline) ***Cohort*** Treatment: Open 1 mg adrenaline No significant difference between trial treatments.
Care provider and setting	The same medical staff provided care in all groups. All patients were treated at the same hospital.
Outcomes	Number of deaths following conversion to sinus rhythm or ventricular tachycardia.

Author	**World Health Organization Task Force on Oral Contraceptives 1988**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients had a preference for treatment. Intention-to-treat analysis was not used in the trial.
Selection Bias	There were differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	There were 5 inappropriate exclusions in the trial treatment 1 group, 5 in the trial treatment 2 group, 9 in the cohort treatment 1 group and 11 in the cohort treatment 2 group.
Intervention and Co-interventions	***Included only the Bangkok centre because differences in protocol in the Szeged and Khon Kaen centres made it inappropriate to pool the centres together.*** ***Trial*** Treatment 1: Combined pill Treatment 2: Progesterone only pill ***Cohort*** Treatment 1: Intra-uterine device, sterilization or no contraception Treatment 2: Injectable progesterone There was a significant harmful effect associated with the trial treatment 1.
Care provider and setting	Unclear who provided care in all groups. All patients were treated at the same setting.
Outcomes	Mean change in milk volume. At 24 weeks.

Author	**Wyse 1991**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients refused randomization. Intention-to-treat analysis was not used in the trial.
Selection Bias	There were differences in baseline characteristics between groups.
Detection Bias	More frequent visits for those in RCT.
Exclusion Bias	No post-randomization exclusions.
Intervention and Co-interventions	***Trial*** Treatment: Anti-arrhythmic drugs (encainide and flecainide) Control: Placebo ***Cohort*** Control: Unclear what treatment, if any, they received Unclear whether there was a significant difference between trial treatment and control.
Care provider and setting	Unclear who provided care in any of the groups. Unclear where patients were treated.
Outcomes	Deaths due to arrhythmia.

Author	**Yamamoto 1992**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients were not randomized because the endoscopists were uncomfortable with randomizing all patients since they had considerable more experience with one method. Intention-to-treat analysis was not used in the trial.
Selection Bias	There were differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	No post-randomization exclusions.
Intervention and Co-interventions	***Trial*** Treatment 1: Eder-Puestow dilator Treatment 2: Medi-Tech balloon dilator ***Cohort*** Treatment 1: Same Treatment 2: Same No significant difference between trial treatments.
Care provider and setting	Endoscopists provided care in the trial groups. Unclear who provided care in the cohort groups. All patients were treated in the same setting.
Outcomes	Recurrent dysphagia during 4 years of follow-up.

Author	**Yamani 2005**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients refused randomization. Intention-to-treat analysis was used in the trial.
Selection Bias	Most likely there were differences in baseline characteristics between groups. (N= 56)
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	No post-randomization exclusions.
Intervention and Co-interventions	***Trial*** Treatment: Cytomegalovirus immunoglobulin (CytoGam) replacement

	Control: Placebo- 5% dextrose in water ***Cohort*** Control: Unclear what treatment, if any, they receive There was a significant beneficial effect in favor of the trial treatment.
Care provider and setting	Unclear who provided care in any of the groups. Unclear where patients were treated.
Outcomes	Incidence of cytomegalovirus infection.

Author	**Yersin 1996**
Design and Methods	Randomized trial and a separate parallel-treated cohort of eligible patients. Cohort patients refused randomization. Intention-to-treat analysis was not used in the trial.
Selection Bias	There were differences in baseline characteristics between groups.
Detection Bias	Follow-up was the same between groups.
Exclusion Bias	There were 7 inappropriate exclusions in the trial treatment 1 group, 9 in the trial treatment 2 group and 6 in the cohort treatment 2 group.
Intervention and Co-interventions	***Trial*** Treatment 1: Multi-axial individualized proposals Treatment 2: Abstinence counseling ***Cohort*** Treatment 2: Same There was no significant difference between trial treatments.
Care provider and setting	Complete medical team (resident and fellow in psychiatry, psychiatrist, social workers) provided care in the trial treatment 1 group. Same person treated both treatment 2 groups. All patients were treated at the same setting.
Outcomes	Number of patients not abstinent at 1 year.

Forest plots

1.1 Subgroups based on treatment effectiveness

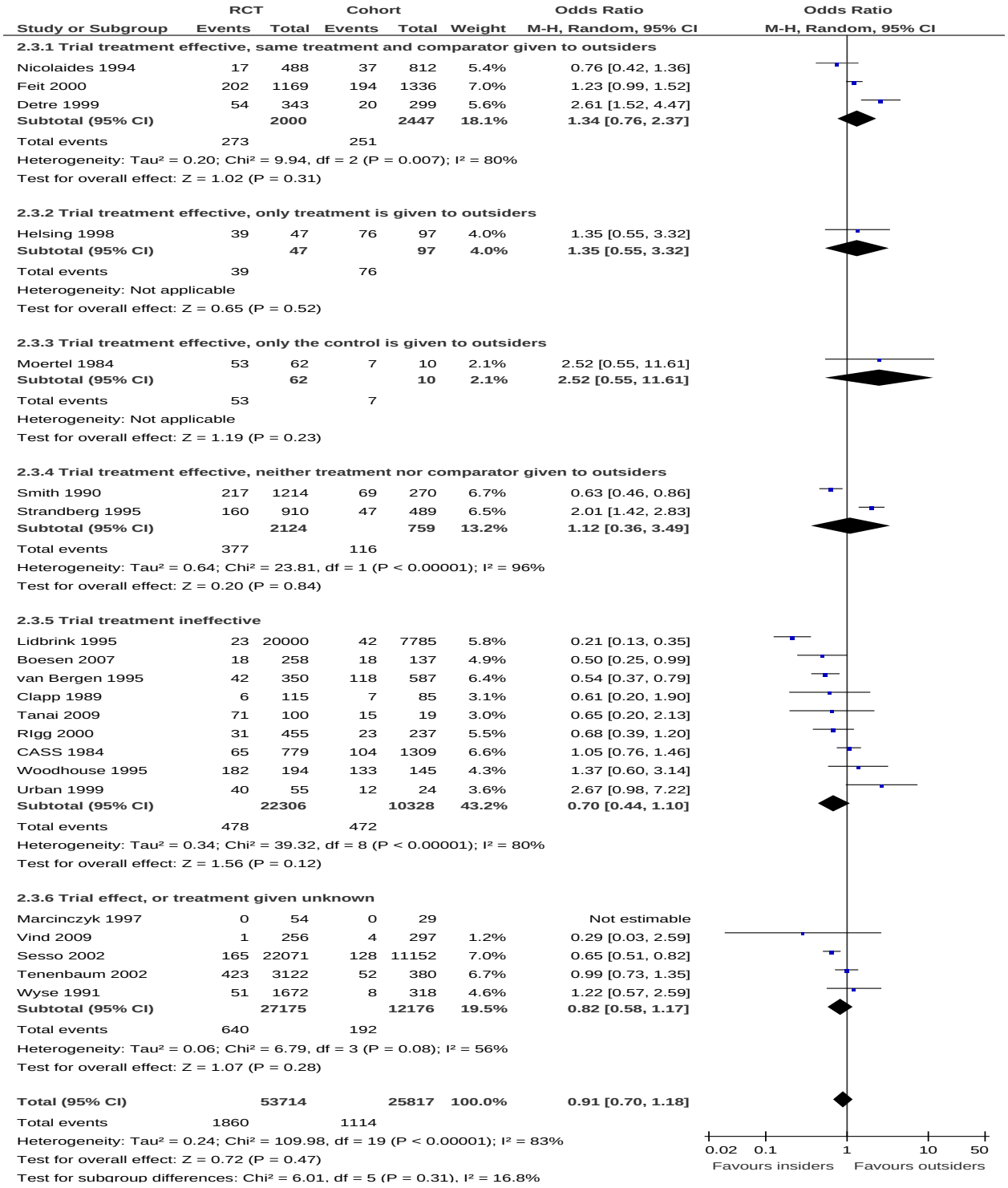
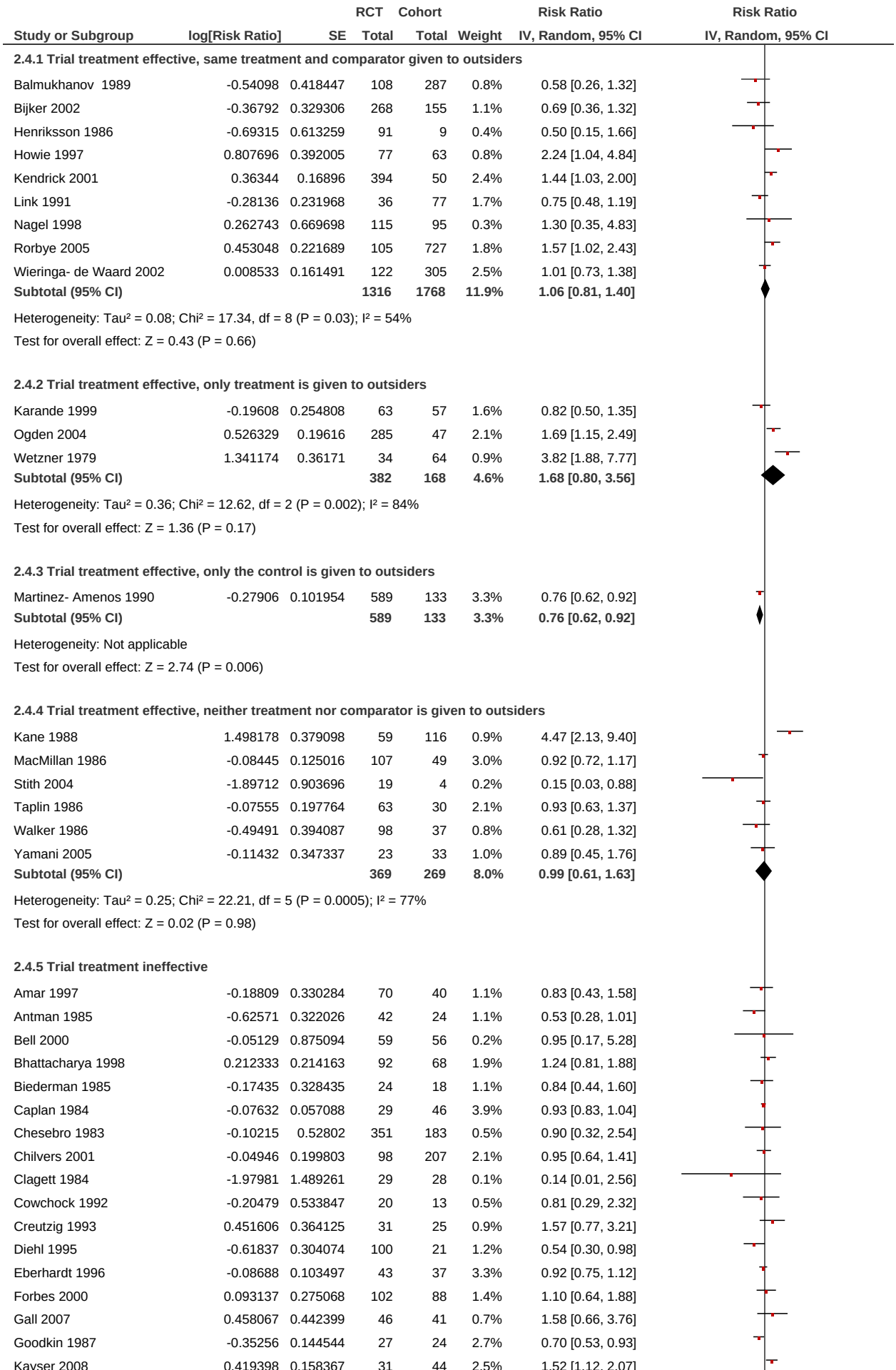


Figure 2: All studies in this meta-analysis feature non- randomized participants either to the trial or cohort group and had a mortality outcome. Studies in this forest plot are divided into subgroups based on trial treatment effectiveness to explain heterogeneity.


RCT Cohort Risk Ratio Risk Ratio
Study or Subgroup log[Risk Ratio] SE Total Total Weight IV, Random, 95% CI IV, Random, 95% CI
2.4.1 Trial treatment effective, same treatment and comparator given to outsiders
Balmukhanov 1989 -0.54098 0.418447 108 287 0.8% 0.58 [0.26, 1.32]
Bijker 2002 -0.36792 0.329306 268 155 1.1% 0.69 [0.36, 1.32]
Henriksson 1986 -0.69315 0.613259 91 9 0.4% 0.50 [0.15, 1.66]
Howie 1997 0.807696 0.392005 77 63 0.8% 2.24 [1.04, 4.84]
Kendrick 2001 0.36344 0.16896 394 50 2.4% 1.44 [1.03, 2.00]
Link 1991 -0.28136 0.231968 36 77 1.7% 0.75 [0.48, 1.19]
Nagel 1998 0.262743 0.669698 115 95 0.3% 1.30 [0.35, 4.83]
Rorbye 2005 0.453048 0.221689 105 727 1.8% 1.57 [1.02, 2.43]
Wieringa- de Waard 2002 0.008533 0.161491 122 305 2.5% 1.01 [0.73, 1.38]
Subtotal (95% CI) 1316 1768 11.9% 1.06 [0.81, 1.40]
Heterogeneity: Tau² = 0.08; Chi² = 17.34, df = 8 (P = 0.03); I² = 54%
Test for overall effect: Z = 0.43 (P = 0.66)
2.4.2 Trial treatment effective, only treatment is given to outsiders
Karande 1999 -0.19608 0.254808 63 57 1.6% 0.82 [0.50, 1.35]
Ogden 2004 0.526329 0.19616 285 47 2.1% 1.69 [1.15, 2.49]
Wetzner 1979 1.341174 0.36171 34 64 0.9% 3.82 [1.88, 7.77]
Subtotal (95% CI) 382 168 4.6% 1.68 [0.80, 3.56]
Heterogeneity: Tau² = 0.36; Chi² = 12.62, df = 2 (P = 0.002); I² = 84%
Test for overall effect: Z = 1.36 (P = 0.17)
2.4.3 Trial treatment effective, only the control is given to outsiders
Martinez- Amenos 1990 -0.27906 0.101954 589 133 3.3% 0.76 [0.62, 0.92]
Subtotal (95% CI) 589 133 3.3% 0.76 [0.62, 0.92]
Heterogeneity: Not applicable
Test for overall effect: Z = 2.74 (P = 0.006)
2.4.4 Trial treatment effective, neither treatment nor comparator is given to outsiders
Kane 1988 1.498178 0.379098 59 116 0.9% 4.47 [2.13, 9.40]
MacMillan 1986 -0.08445 0.125016 107 49 3.0% 0.92 [0.72, 1.17]
Stith 2004 -1.89712 0.903696 19 4 0.2% 0.15 [0.03, 0.88]
Taplin 1986 -0.07555 0.197764 63 30 2.1% 0.93 [0.63, 1.37]
Walker 1986 -0.49491 0.394087 98 37 0.8% 0.61 [0.28, 1.32]
Yamani 2005 -0.11432 0.347337 23 33 1.0% 0.89 [0.45, 1.76]
Subtotal (95% CI) 369 269 8.0% 0.99 [0.61, 1.63]
Heterogeneity: Tau² = 0.25; Chi² = 22.21, df = 5 (P = 0.0005); I² = 77%
Test for overall effect: Z = 0.02 (P = 0.98)
2.4.5 Trial treatment ineffective
Amar 1997 -0.18809 0.330284 70 40 1.1% 0.83 [0.43, 1.58]
Antman 1985 -0.62571 0.322026 42 24 1.1% 0.53 [0.28, 1.01]
Bell 2000 -0.05129 0.875094 59 56 0.2% 0.95 [0.17, 5.28]
Bhattacharya 1998 0.212333 0.214163 92 68 1.9% 1.24 [0.81, 1.88]
Biederman 1985 -0.17435 0.328435 24 18 1.1% 0.84 [0.44, 1.60]
Caplan 1984 -0.07632 0.057088 29 46 3.9% 0.93 [0.83, 1.04]
Chesebro 1983 -0.10215 0.52802 351 183 0.5% 0.90 [0.32, 2.54]
Chilvers 2001 -0.04946 0.199803 98 207 2.1% 0.95 [0.64, 1.41]
Clagett 1984 -1.97981 1.489261 29 28 0.1% 0.14 [0.01, 2.56]
Cowchock 1992 -0.20479 0.533847 20 13 0.5% 0.81 [0.29, 2.32]
Creutzig 1993 0.451606 0.364125 31 25 0.9% 1.57 [0.77, 3.21]
Diehl 1995 -0.61837 0.304074 100 21 1.2% 0.54 [0.30, 0.98]
Eberhardt 1996 -0.08688 0.103497 43 37 3.3% 0.92 [0.75, 1.12]
Forbes 2000 0.093137 0.275068 102 88 1.4% 1.10 [0.64, 1.88]
Gall 2007 0.458067 0.442399 46 41 0.7% 1.58 [0.66, 3.76]
Goodkin 1987 -0.35256 0.144544 27 24 2.7% 0.70 [0.53, 0.93]
Kayser 2008 0.419398 0.158367 31 44 2.5% 1.52 [1.12, 2.07]

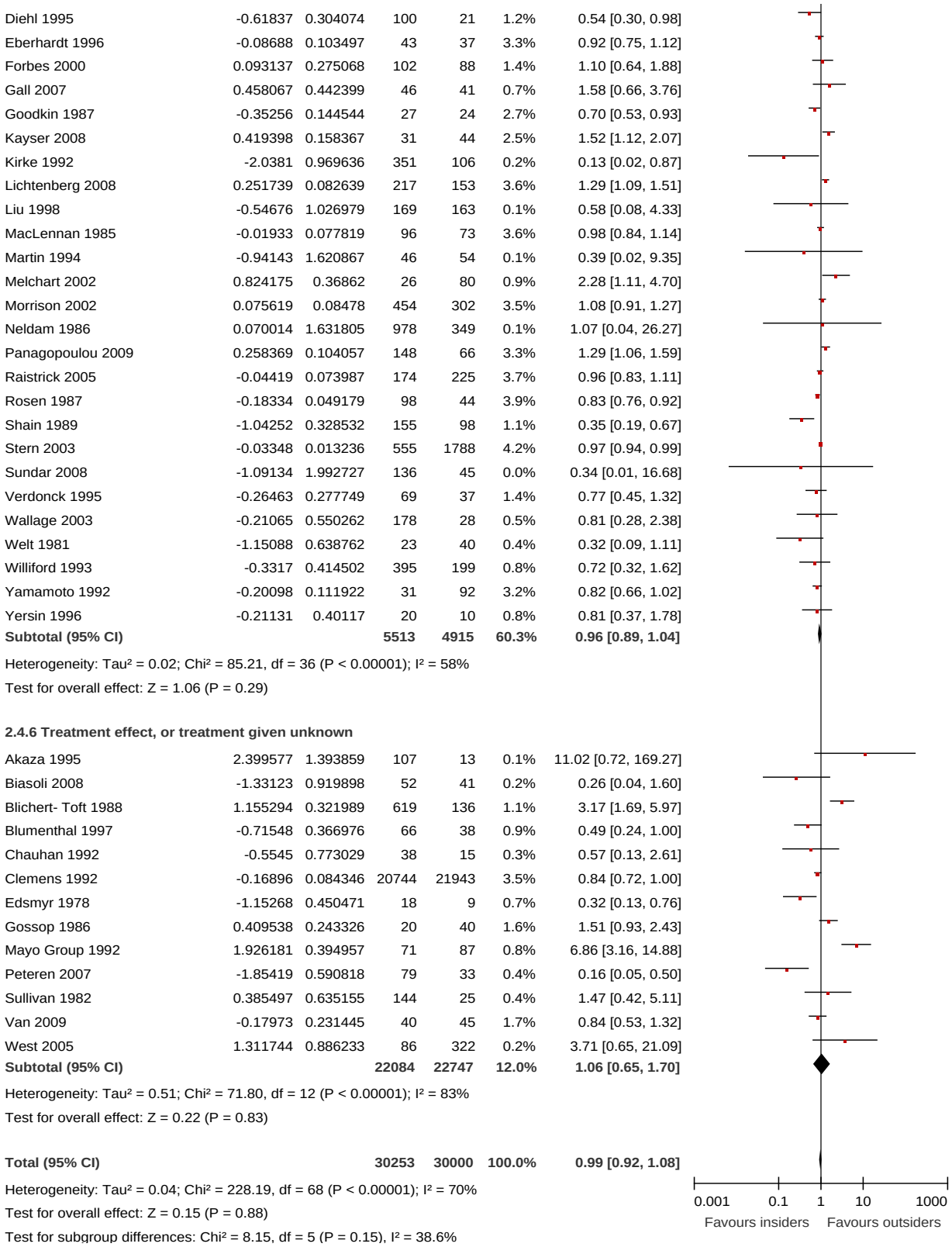
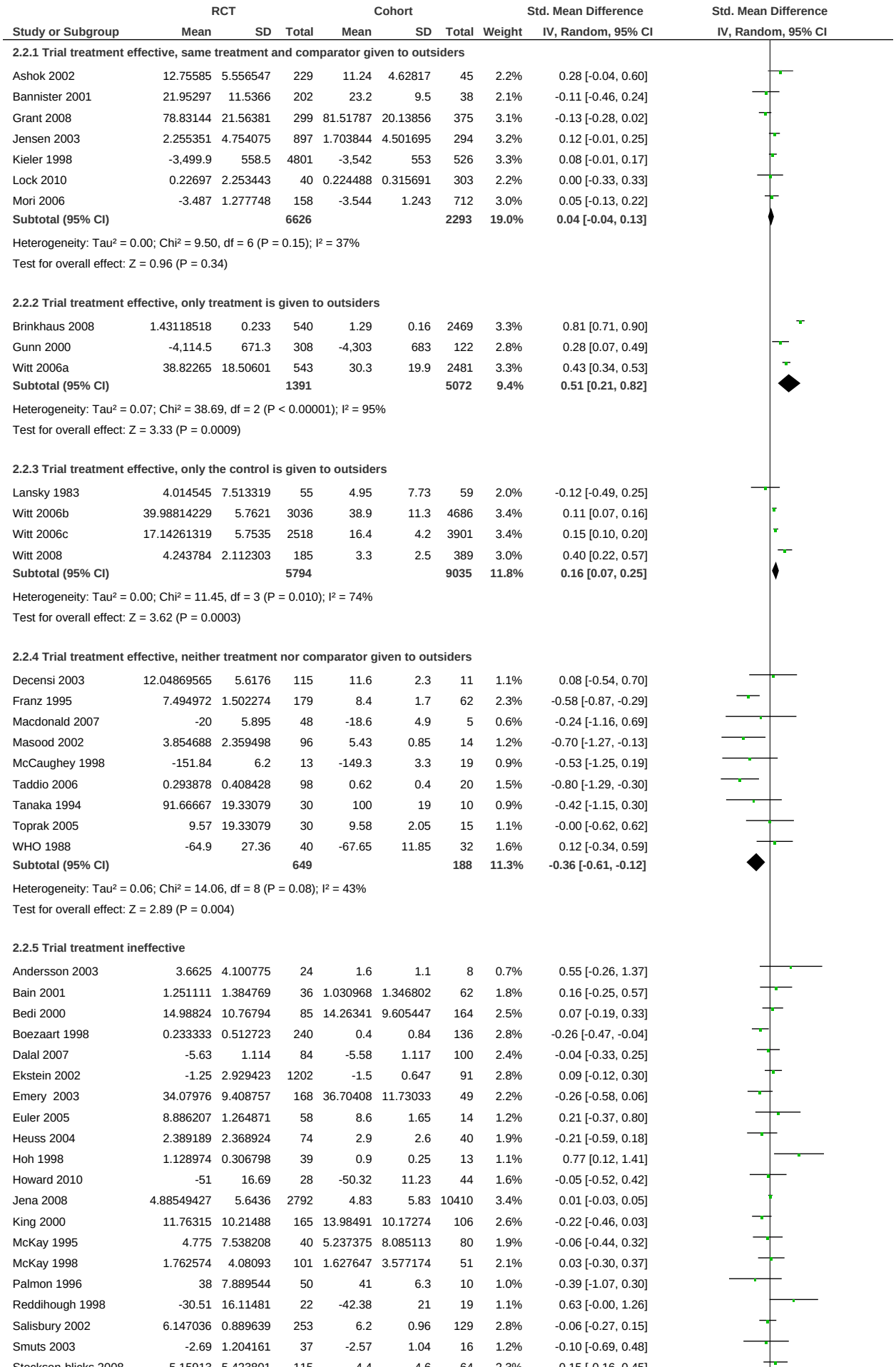


Figure 3: All studies in this meta-analysis were non- randomly assigned to either the trial or cohort group and had a dichotomous non-mortality outcome. Studies in this forest plot are divided into subgroups based on trial treatment effectiveness to explain heterogeneity.


RCT
Cohort
Std. Mean Difference
Std. Mean Difference
Study or Subgroup Mean SD Total Mean SD Total Weight IV, Random, 95% CI
IV, Random, 95% CI
2.2.1 Trial treatment effective, same treatment and comparator given to outsiders
Ashok 2002 12.75585 5.556547 229 11.24 4.62817 45 2.2% 0.28 [-0.04, 0.60]
Bannister 2001 21.95297 11.5366 202 23.2 9.5 38 2.1% -0.11 [-0.46, 0.24]
Grant 2008 78.83144 21.56381 299 81.51787 20.13856 375 3.1% -0.13 [-0.28, 0.02]
Jensen 2003 2.255351 4.754075 897 1.703844 4.501695 294 3.2% 0.12 [-0.01, 0.25]
Kieler 1998 -3,499.9 558.5 4801 -3,542 553 526 3.3% 0.08 [-0.01, 0.17]
Lock 2010 0.22697 2.253443 40 0.224488 0.315691 303 2.2% 0.00 [-0.33, 0.33]
Mori 2006 -3.487 1.277748 158 -3.544 1.243 712 3.0% 0.05 [-0.13, 0.22]
Subtotal (95% CI) 6626 2293 19.0% 0.04 [-0.04, 0.13]
Heterogeneity: Tau² = 0.00; Chi² = 9.50, df = 6 (P = 0.15); I² = 37%
Test for overall effect: Z = 0.96 (P = 0.34)
2.2.2 Trial treatment effective, only treatment is given to outsiders
Brinkhaus 2008 1.43118518 0.233 540 1.29 0.16 2469 3.3% 0.81 [0.71, 0.90]
Gunn 2000 -4,114.5 671.3 308 -4,303 683 122 2.8% 0.28 [0.07, 0.49]
Witt 2006a 38.82265 18.50601 543 30.3 19.9 2481 3.3% 0.43 [0.34, 0.53]
Subtotal (95% CI) 1391 5072 9.4% 0.51 [0.21, 0.82]
Heterogeneity: Tau² = 0.07; Chi² = 38.69, df = 2 (P < 0.00001); I² = 95%
Test for overall effect: Z = 3.33 (P = 0.0009)
2.2.3 Trial treatment effective, only the control is given to outsiders
Lansky 1983 4.014545 7.513319 55 4.95 7.73 59 2.0% -0.12 [-0.49, 0.25]
Witt 2006b 39.98814229 5.7621 3036 38.9 11.3 4686 3.4% 0.11 [0.07, 0.16]
Witt 2006c 17.14261319 5.7535 2518 16.4 4.2 3901 3.4% 0.15 [0.10, 0.20]
Witt 2008 4.243784 2.112303 185 3.3 2.5 389 3.0% 0.40 [0.22, 0.57]
Subtotal (95% CI) 5794 9035 11.8% 0.16 [0.07, 0.25]
Heterogeneity: Tau² = 0.00; Chi² = 11.45, df = 3 (P = 0.010); I² = 74%
Test for overall effect: Z = 3.62 (P = 0.0003)
2.2.4 Trial treatment effective, neither treatment nor comparator given to outsiders
Decensi 2003 12.04869565 5.6176 115 11.6 2.3 11 1.1% 0.08 [-0.54, 0.70]
Franz 1995 7.494972 1.502274 179 8.4 1.7 62 2.3% -0.58 [-0.87, -0.29]
Macdonald 2007 -20 5.895 48 -18.6 4.9 5 0.6% -0.24 [-1.16, 0.69]
Masood 2002 3.854688 2.359498 96 5.43 0.85 14 1.2% -0.70 [-1.27, -0.13]
McCaughey 1998 -151.84 6.2 13 -149.3 3.3 19 0.9% -0.53 [-1.25, 0.19]
Taddio 2006 0.293878 0.408428 98 0.62 0.4 20 1.5% -0.80 [-1.29, -0.30]
Tanaka 1994 91.66667 19.33079 30 100 19 10 0.9% -0.42 [-1.15, 0.30]
Toprak 2005 9.57 19.33079 30 9.58 2.05 15 1.1% -0.00 [-0.62, 0.62]
WHO 1988 -64.9 27.36 40 -67.65 11.85 32 1.6% 0.12 [-0.34, 0.59]
Subtotal (95% CI) 649 188 11.3% -0.36 [-0.61, -0.12]
Heterogeneity: Tau² = 0.06; Chi² = 14.06, df = 8 (P = 0.08); I² = 43%
Test for overall effect: Z = 2.89 (P = 0.004)
2.2.5 Trial treatment ineffective
Andersson 2003 3.6625 4.100775 24 1.6 1.1 8 0.7% 0.55 [-0.26, 1.37]
Bain 2001 1.251111 1.384769 36 1.030968 1.346802 62 1.8% 0.16 [-0.25, 0.57]
Bedi 2000 14.98824 10.76794 85 14.26341 9.605447 164 2.5% 0.07 [-0.19, 0.33]
Boezaart 1998 0.233333 0.512723 240 0.4 0.84 136 2.8% -0.26 [-0.47, -0.04]
Dalal 2007 -5.63 1.114 84 -5.58 1.117 100 2.4% -0.04 [-0.33, 0.25]
Ekstein 2002 -1.25 2.929423 1202 -1.5 0.647 91 2.8% 0.09 [-0.12, 0.30]
Emery 2003 34.07976 9.408757 168 36.70408 11.73033 49 2.2% -0.26 [-0.58, 0.06]
Euler 2005 8.886207 1.264871 58 8.6 1.65 14 1.2% 0.21 [-0.37, 0.80]
Heuss 2004 2.389189 2.368924 74 2.9 2.6 40 1.9% -0.21 [-0.59, 0.18]
Hoh 1998 1.128974 0.306798 39 0.9 0.25 13 1.1% 0.77 [0.12, 1.41]
Howard 2010 -51 16.69 28 -50.32 11.23 44 1.6% -0.05 [-0.52, 0.42]
Jena 2008 4.88549427 5.6436 2792 4.83 5.83 10410 3.4% 0.01 [-0.03, 0.05]
King 2000 11.76315 10.21488 165 13.98491 10.17274 106 2.6% -0.22 [-0.46, 0.03]
McKay 1995 4.775 7.538208 40 5.237375 8.085113 80 1.9% -0.06 [-0.44, 0.32]
McKay 1998 1.762574 4.08093 101 1.627647 3.577174 51 2.1% 0.03 [-0.30, 0.37]
Palmon 1996 38 7.889544 50 41 6.3 10 1.0% -0.39 [-1.07, 0.30]
Reddihough 1998 -30.51 16.11481 22 -42.38 21 19 1.1% 0.63 [-0.00, 1.26]
Salisbury 2002 6.147036 0.889639 253 6.2 0.96 129 2.8% -0.06 [-0.27, 0.15]
Smuts 2003 -2.69 1.204161 37 -2.57 1.04 16 1.2% -0.10 [-0.69, 0.48]

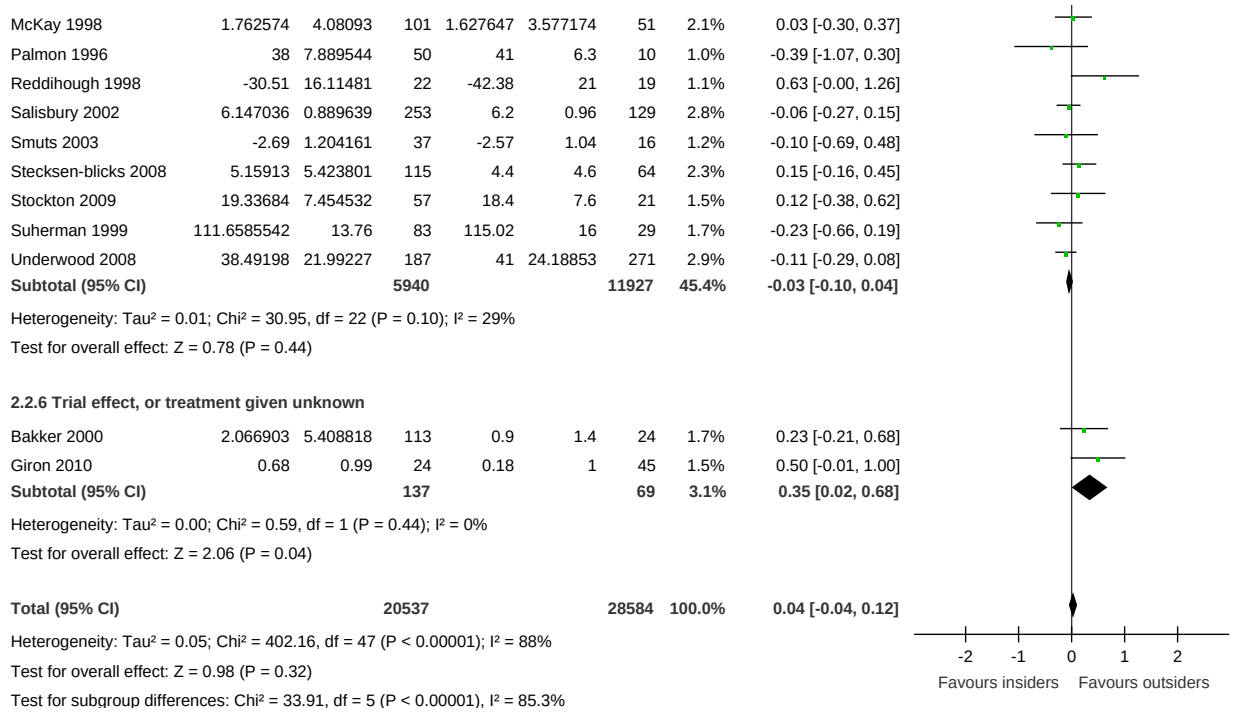


Figure 4: All studies in this meta-analysis non-randomly assigned participants either to the trial or cohort group and had a continuous endpoint as an outcome. Studies in this forest plot are divided into subgroups based on trial treatment effectiveness to explain heterogeneity. Note that negative signs were added to those scores that were based on a scale where lower scores indicated higher disease severity.

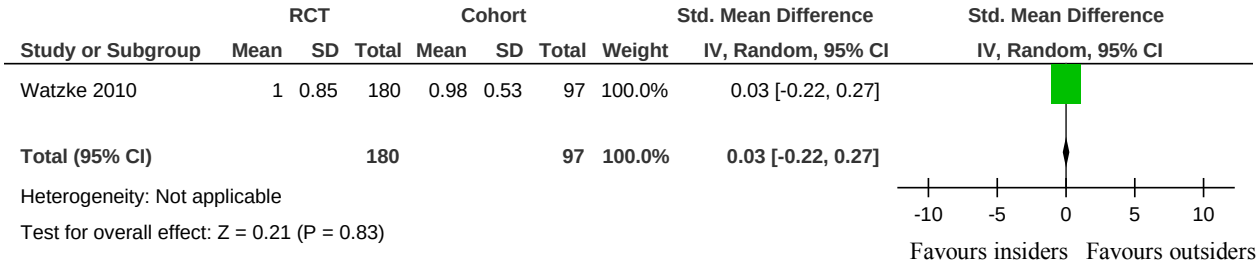


Figure 5: All studies in this meta-analysis randomly assigned participants either to the trial or cohort group and had a continuous endpoint as an outcome. Studies in this forest plot are divided into subgroups based on trial treatment effectiveness to explain heterogeneity.

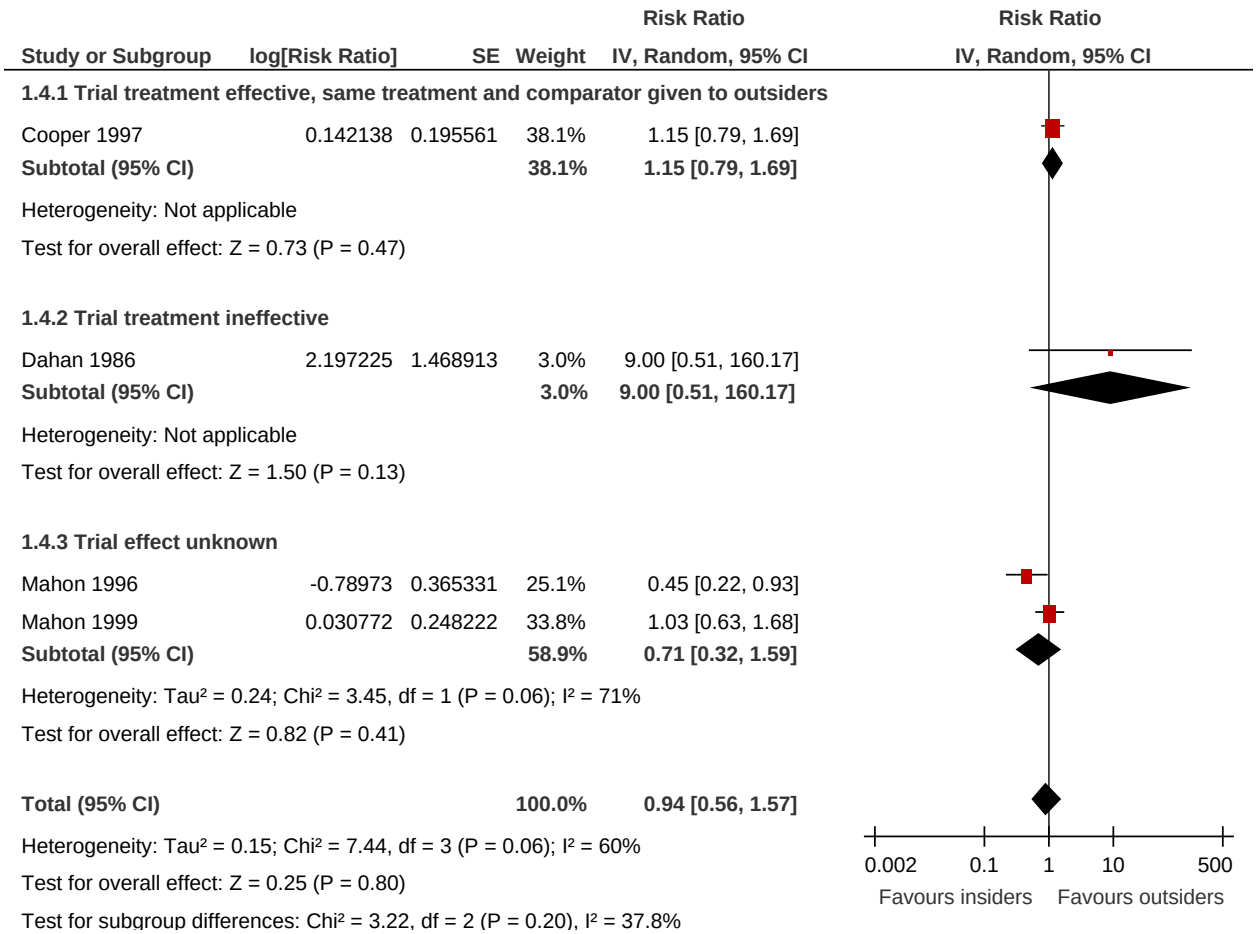


Figure 6: All studies in this meta-analysis randomly assigned participants either to the trial or cohort group and had a dichotomous non-mortality outcome. Studies in this forest plot are divided into subgroups based on trial treatment effectiveness to explain heterogeneity.

1.2 Subgroups based on baseline characteristics

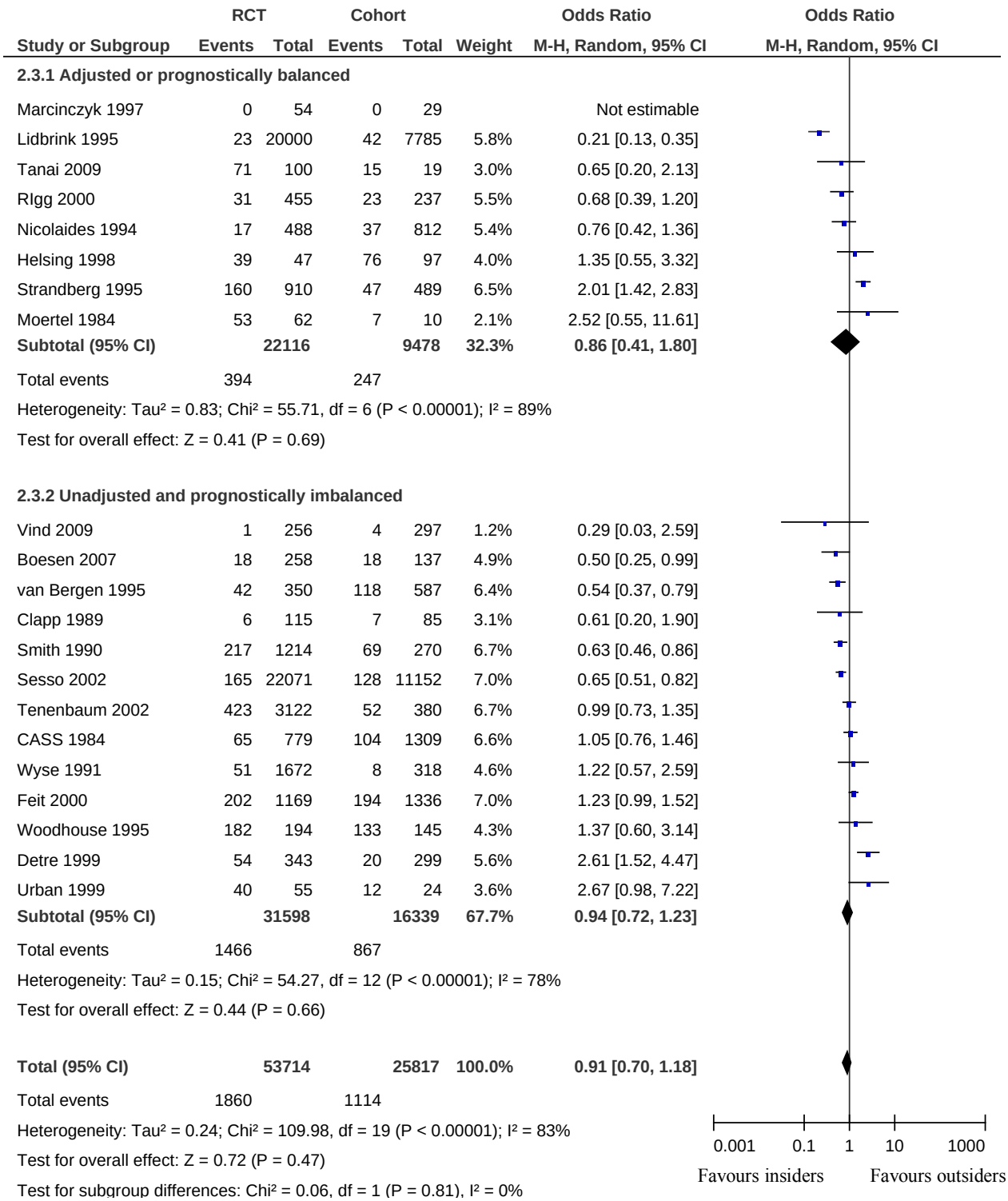
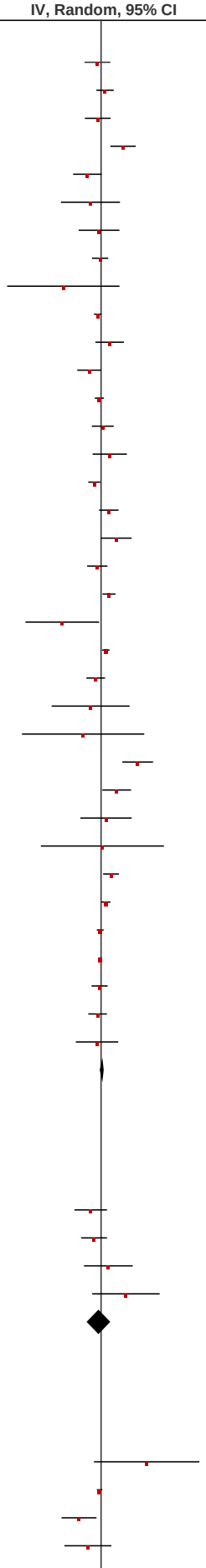


Figure 7: All studies in this meta-analysis non- randomly assigned participants either to the trial or cohort group and had a mortality outcome. Studies in this forest plot are divided into subgroups based on baseline characteristics to explain heterogeneity.

Study or Subgroup	log[Risk Ratio]	SE	RCT Total	Cohort Total	Weight	Risk Ratio IV, Random, 95% CI
2.4.1 Prognostic balance						
Amar 1997	-0.18809	0.330284	70	40	1.1%	0.83 [0.43, 1.58]
Bhattacharya 1998	0.212333	0.214163	92	68	1.9%	1.24 [0.81, 1.88]
Biederman 1985	-0.17435	0.328435	24	18	1.1%	0.84 [0.44, 1.60]
Blichert- Toft 1988	1.155294	0.321989	619	136	1.1%	3.17 [1.69, 5.97]
Blumenthal 1997	-0.71548	0.366976	66	38	0.9%	0.49 [0.24, 1.00]
Chauhan 1992	-0.5545	0.773029	38	15	0.3%	0.57 [0.13, 2.61]
Chesebro 1983	-0.10215	0.52802	351	183	0.5%	0.90 [0.32, 2.54]
Chilvers 2001	-0.04946	0.199803	98	207	2.1%	0.95 [0.64, 1.41]
Clagett 1984	-1.97981	1.489261	29	28	0.1%	0.14 [0.01, 2.56]
Clemens 1992	-0.16896	0.084346	20744	21943	3.5%	0.84 [0.72, 1.00]
Creutzig 1993	0.451606	0.364125	31	25	0.9%	1.57 [0.77, 3.21]
Diehl 1995	-0.61837	0.304074	100	21	1.2%	0.54 [0.30, 0.98]
Eberhardt 1996	-0.08688	0.103497	43	37	3.3%	0.92 [0.75, 1.12]
Forbes 2000	0.093137	0.275068	102	88	1.4%	1.10 [0.64, 1.88]
Gall 2007	0.458067	0.442399	46	41	0.7%	1.58 [0.66, 3.76]
Goodkin 1987	-0.35256	0.144544	27	24	2.7%	0.70 [0.53, 0.93]
Gossop 1986	0.409538	0.243326	20	40	1.6%	1.51 [0.93, 2.43]
Howie 1997	0.807696	0.392005	77	63	0.8%	2.24 [1.04, 4.84]
Karande 1999	-0.19608	0.254808	63	57	1.6%	0.82 [0.50, 1.35]
Kayser 2008	0.419398	0.158367	31	44	2.5%	1.52 [1.12, 2.07]
Kirke 1992	-2.0381	0.969636	351	106	0.2%	0.13 [0.02, 0.87]
Lichtenberg 2008	0.251739	0.082639	217	153	3.6%	1.29 [1.09, 1.51]
Link 1991	-0.28136	0.231968	36	77	1.7%	0.75 [0.48, 1.19]
Liu 1998	-0.54676	1.026979	169	163	0.1%	0.58 [0.08, 4.33]
Martin 1994	-0.94143	1.620867	46	54	0.1%	0.39 [0.02, 9.35]
Mayo Group 1992	1.926181	0.394957	71	87	0.8%	6.86 [3.16, 14.88]
Melchart 2002	0.824175	0.36862	26	80	0.9%	2.28 [1.11, 4.70]
Nagel 1998	0.262743	0.669698	115	95	0.3%	1.30 [0.35, 4.83]
Neldam 1986	0.070014	1.631805	978	349	0.1%	1.07 [0.04, 26.27]
Ogden 2004	0.526329	0.19616	285	47	2.1%	1.69 [1.15, 2.49]
Panagopoulou 2009	0.258369	0.104057	148	66	3.3%	1.29 [1.06, 1.59]
Raistrick 2005	-0.04419	0.073987	174	225	3.7%	0.96 [0.83, 1.11]
Stern 2003	-0.03348	0.013236	555	1788	4.2%	0.97 [0.94, 0.99]
Taplin 1986	-0.07555	0.197764	63	30	2.1%	0.93 [0.63, 1.37]
Van 2009	-0.17973	0.231445	40	45	1.7%	0.84 [0.53, 1.32]
Wallage 2003	-0.21065	0.550262	178	28	0.5%	0.81 [0.28, 2.38]
Subtotal (95% CI)			**26123**	**26509**	**54.6%**	**1.08 [0.97, 1.20]**
Heterogeneity: Tau² = 0.04; Chi² = 115.73, df = 35 (P < 0.00001); I² = 70%						
Test for overall effect: Z = 1.35 (P = 0.18)						
2.4.2 Most likely balanced						
Balmukhanov 1989	-0.54098	0.418447	108	287	0.8%	0.58 [0.26, 1.32]
Bijker 2002	-0.36792	0.329306	268	155	1.1%	0.69 [0.36, 1.32]
Sullivan 1982	0.385497	0.635155	144	25	0.4%	1.47 [0.42, 5.11]
West 2005	1.311744	0.886233	86	322	0.2%	3.71 [0.65, 21.09]
Subtotal (95% CI)			**606**	**789**	**2.4%**	**0.90 [0.49, 1.67]**
Heterogeneity: Tau² = 0.14; Chi² = 4.68, df = 3 (P = 0.20); I² = 36%						
Test for overall effect: Z = 0.34 (P = 0.74)						
2.4.3 Most likely imbalanced						
Akaza 1995	2.399577	1.393859	107	13	0.1%	11.02 [0.72, 169.27]
Caplan 1984	-0.07632	0.057088	29	46	3.9%	0.93 [0.83, 1.04]
Edsmyr 1978	-1.15268	0.450471	18	9	0.7%	0.32 [0.13, 0.76]
Henriksson 1986	-0.69315	0.613259	91	9	0.4%	0.50 [0.15, 1.66]



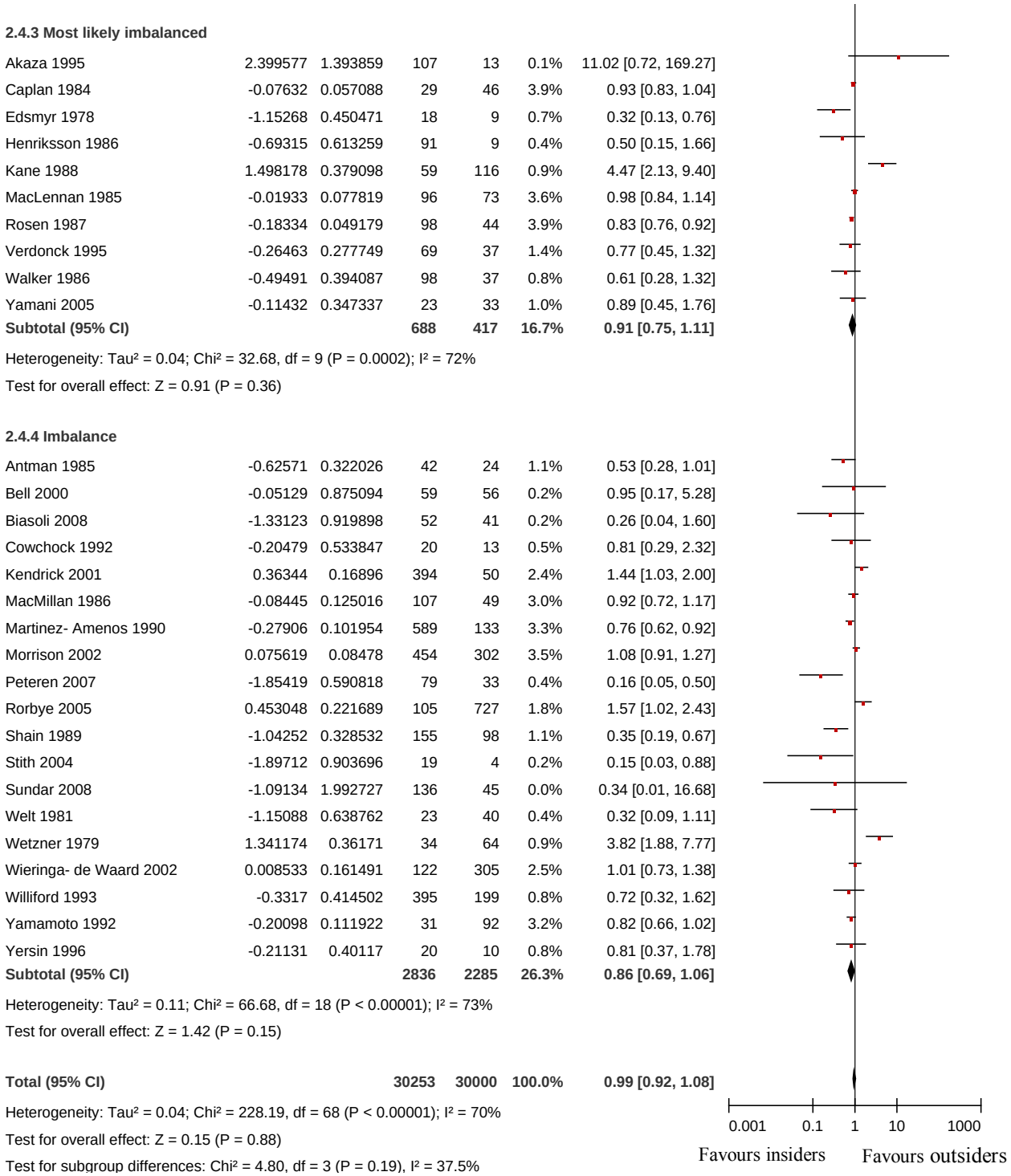
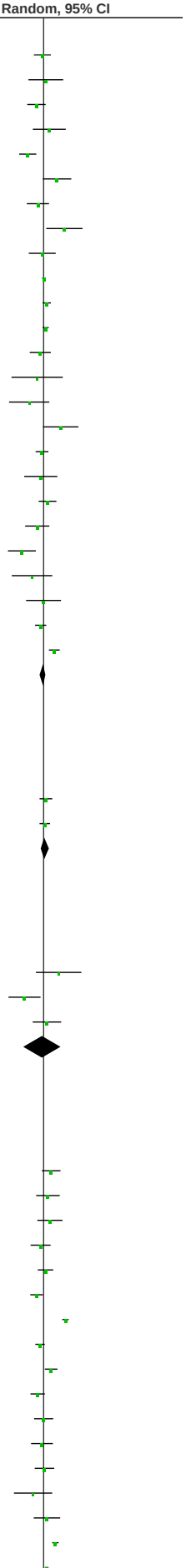


Figure 8: All studies in this meta-analysis non- randomly assigned participants either to the trial or cohort group and had a dichotomous non-mortality outcome. Studies in this forest plot are divided into subgroups based on baseline characteristics to explain heterogeneity.

Study or Subgroup	RCT Mean	RCT SD	RCT Total	Cohort Mean	Cohort SD	Cohort Total	Weight	Std. Mean Difference IV, Random, 95% CI
2.2.1 Prognostic balance								
Dalal 2007	-5.628	1.114	84	-5.579	1.117	100	2.4%	-0.04 [-0.33, 0.25]
Decensi 2003	12.04869565	5.6176	115	11.6	2.3	11	1.1%	0.08 [-0.54, 0.70]
Emery 2003	34.07976	9.408757	168	36.70408	11.73033	49	2.2%	-0.26 [-0.58, 0.06]
Euler 2005	8.886207	1.264871	58	8.6	1.65	14	1.2%	0.21 [-0.37, 0.80]
Franz 1995	7.494972	1.502274	179	8.4	1.7	62	2.3%	-0.58 [-0.87, -0.29]
Giron 2010	0.68	0.99	24	0.18	1	45	1.5%	0.50 [-0.01, 1.00]
Heuss 2004	2.389189	2.368924	74	2.9	2.6	40	1.9%	-0.21 [-0.59, 0.18]
Hoh 1998	1.128974	0.306798	39	0.9	0.25	13	1.1%	0.77 [0.12, 1.41]
Howard 2010	-51	16.69	28	-50.32	11.23	44	1.6%	-0.05 [-0.52, 0.42]
Jena 2008	4.885494	5.6436	2792	4.83	5.83	10410	3.4%	0.01 [-0.03, 0.05]
Jensen 2003	2.255351	4.754075	897	1.703844	4.501695	294	3.2%	0.12 [-0.01, 0.25]
Kieler 1998	-3,499.9	558.5	4801	-3,542	553	526	3.3%	0.08 [-0.01, 0.17]
Lansky 1983	4.014545	7.513319	55	4.95	7.73	59	2.0%	-0.12 [-0.49, 0.25]
Macdonald 2007	-20	5.895	48	-18.6	4.9	5	0.6%	-0.24 [-1.16, 0.69]
McCaughey 1998	-151.84	6.2	13	-149.3	3.3	19	0.9%	-0.53 [-1.25, 0.19]
Reddihough 1998	-30.51	16.11481	22	-42.38	21	19	1.1%	0.63 [-0.00, 1.26]
Salisbury 2002	6.147036	0.889639	253	6.2	0.96	129	2.8%	-0.06 [-0.27, 0.15]
Smuts 2003	-2.69	1.204161	37	-2.57	1.04	16	1.2%	-0.10 [-0.69, 0.48]
Stecksen-blicks 2008	5.15913	5.423801	115	4.4	4.6	64	2.3%	0.15 [-0.16, 0.45]
Suherman 1999	111.6585542	13.76	83	115.02	16	29	1.7%	-0.23 [-0.66, 0.19]
Taddio 2006	0.293878	0.408428	98	0.62	0.4	20	1.5%	-0.80 [-1.29, -0.30]
Tanaka 1994	91.66667	19.33079	30	100	19	10	0.9%	-0.42 [-1.15, 0.30]
Toprak 2005	9.57	19.33079	30	9.58	2.05	15	1.1%	-0.00 [-0.62, 0.62]
Underwood 2008	38.49198	21.99227	187	41	24.18853	271	2.9%	-0.11 [-0.29, 0.08]
Witt 2008	4.243784	2.112303	185	3.3	2.5	389	3.0%	0.40 [0.22, 0.57]
Subtotal (95% CI)			**10415**			**12653**	**47.1%**	**-0.02 [-0.11, 0.07]**
Heterogeneity: Tau² = 0.02; Chi² = 74.09, df = 24 (P < 0.00001); I² = 68%								
Test for overall effect: Z = 0.39 (P = 0.70)								
2.2.3 Most likely balanced								
Ekstein 2002	-1.25	2.929423	1202	-1.5	0.647	91	2.8%	0.09 [-0.12, 0.30]
Mori 2006	-3.487	1.277748	158	-3.544	1.243	712	3.0%	0.05 [-0.13, 0.22]
Subtotal (95% CI)			**1360**			**803**	**5.7%**	**0.06 [-0.07, 0.20]**
Heterogeneity: Tau² = 0.00; Chi² = 0.09, df = 1 (P = 0.76); I² = 0%								
Test for overall effect: Z = 0.91 (P = 0.36)								
2.2.4 Most likely imbalanced								
Andersson 2003	3.6625	4.100775	24	1.6	1.1	8	0.7%	0.55 [-0.26, 1.37]
Masood 2002	3.854688	2.359498	96	5.43	0.85	14	1.2%	-0.70 [-1.27, -0.13]
Stockton 2009	19.33684	7.454532	57	18.4	7.6	21	1.5%	0.12 [-0.38, 0.62]
Subtotal (95% CI)			**177**			**43**	**3.5%**	**-0.04 [-0.73, 0.64]**
Heterogeneity: Tau² = 0.27; Chi² = 7.54, df = 2 (P = 0.02); I² = 73%								
Test for overall effect: Z = 0.13 (P = 0.90)								
2.2.5 Imbalance								
Ashok 2002	12.75585	5.556547	229	11.24	4.62817	45	2.2%	0.28 [-0.04, 0.60]
Bain 2001	1.251111	1.384769	36	1.030968	1.346802	62	1.8%	0.16 [-0.25, 0.57]
Bakker 2000	2.066903	5.408818	113	0.9	1.4	24	1.7%	0.23 [-0.21, 0.68]
Bannister 2001	21.95297	11.5366	202	23.2	9.5	38	2.1%	-0.11 [-0.46, 0.24]
Bedi 2000	14.98824	10.76794	85	14.26341	9.605447	164	2.5%	0.07 [-0.19, 0.33]
Boezaart 1998	0.233333	0.512723	240	0.4	0.84	136	2.8%	-0.26 [-0.47, -0.04]
Brinkhaus 2008	1.431185	0.233	540	1.29	0.16	2469	3.3%	0.81 [0.71, 0.90]
Grant 2008	78.83144	21.56381	299	81.51787	20.13856	375	3.1%	-0.13 [-0.28, 0.02]
Gunn 2000	-4,114.52	671.3	308	-4,303	683	122	2.8%	0.28 [0.07, 0.49]
King 2000	11.76315	10.21488	165	13.98491	10.17274	106	2.6%	-0.22 [-0.46, 0.03]
Lock 2010	0.22697	2.253443	40	0.224488	0.315691	303	2.2%	0.00 [-0.33, 0.33]
McKay 1995	4.775	7.538208	40	5.237375	8.085113	80	1.9%	-0.06 [-0.44, 0.32]
McKay 1998	1.762574	4.08093	101	1.627647	3.577174	51	2.1%	0.03 [-0.30, 0.37]
Palmon 1996	38	7.889544	50	41	6.3	10	1.0%	-0.39 [-1.07, 0.30]
WHO 1988	-64.9	27.36	40	-67.65	11.85	32	1.6%	0.12 [-0.34, 0.59]
Witt 2006a	38.82265	18.50601	543	30.3	19.9	2481	3.3%	0.43 [0.34, 0.53]



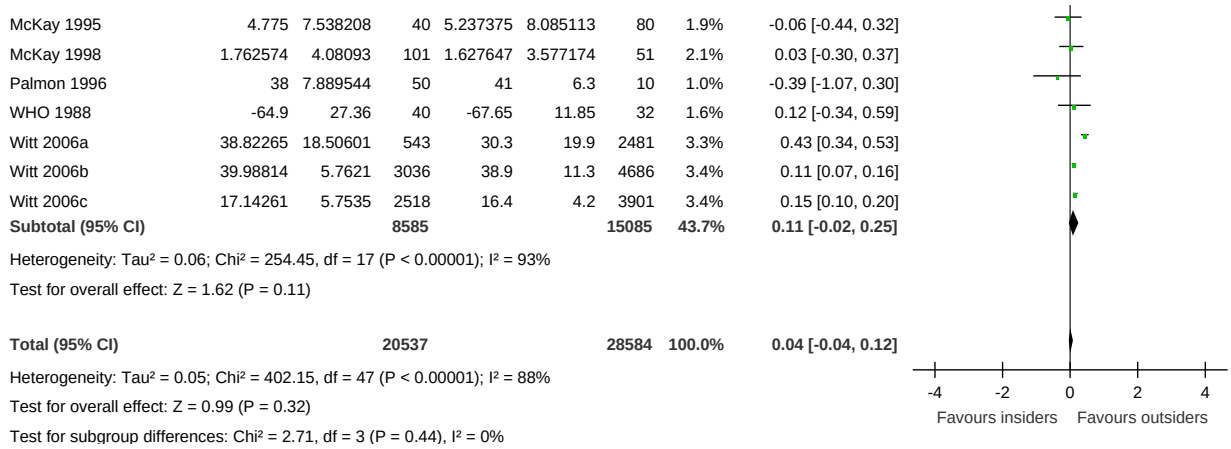


Figure 9: All studies in this meta-analysis non- randomly assigned participants either to the trial or cohort group and had a continuous endpoint as an outcome. Studies in this forest plot are divided into subgroups based on baseline characteristics to explain heterogeneity. Note that negative signs were added to those scores that were based on a scale where lower scores indicated higher disease severity.

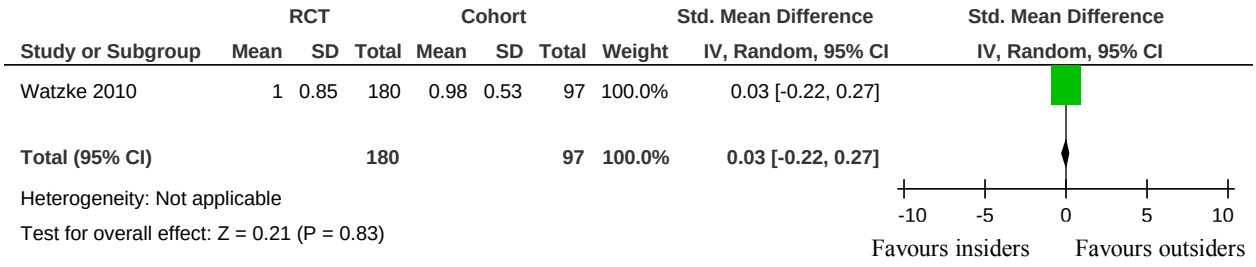


Figure 10: All studies in this meta-analysis randomly assigned participants either to the trial or cohort group and had a continuous endpoint as an outcome. Studies in this forest plot are divided into subgroups based on baseline characteristics to explain heterogeneity.

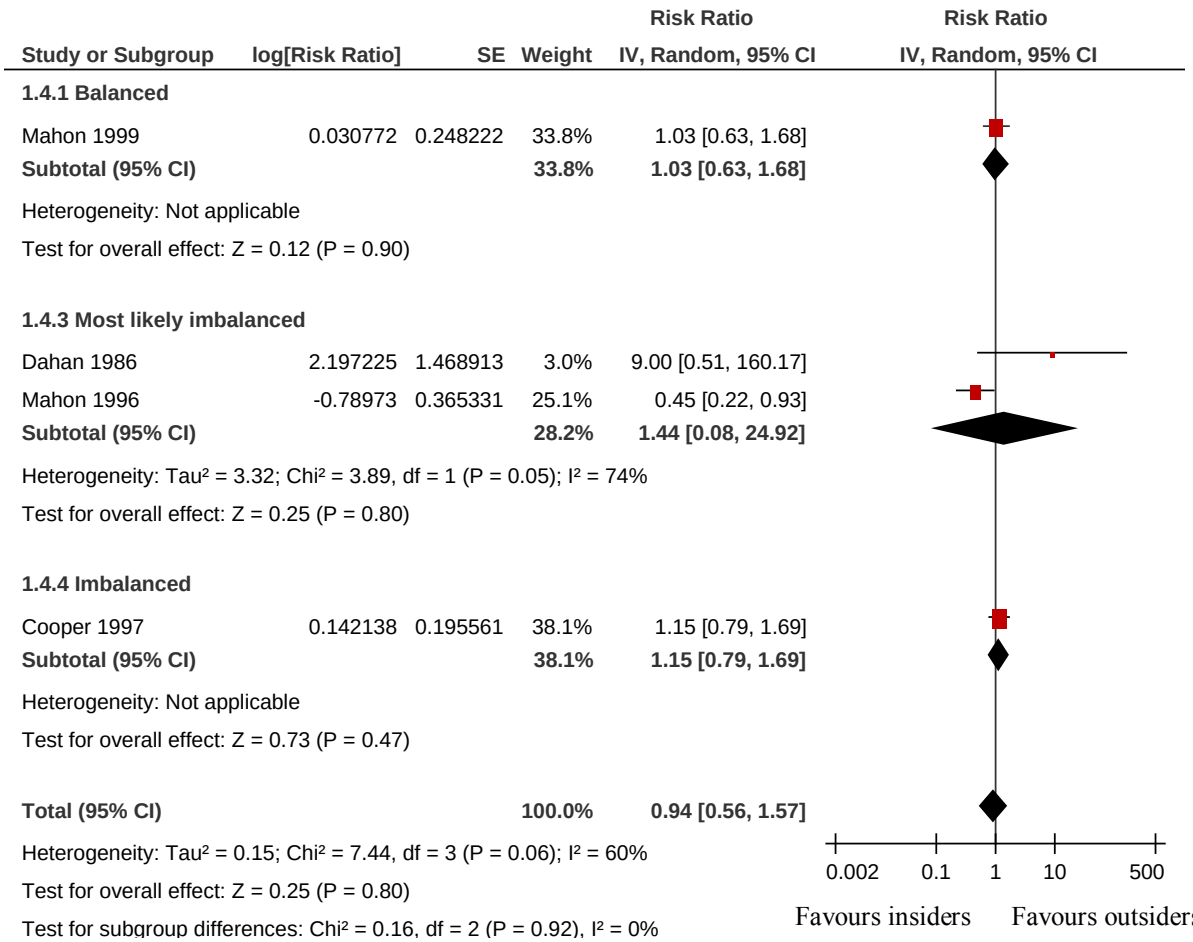
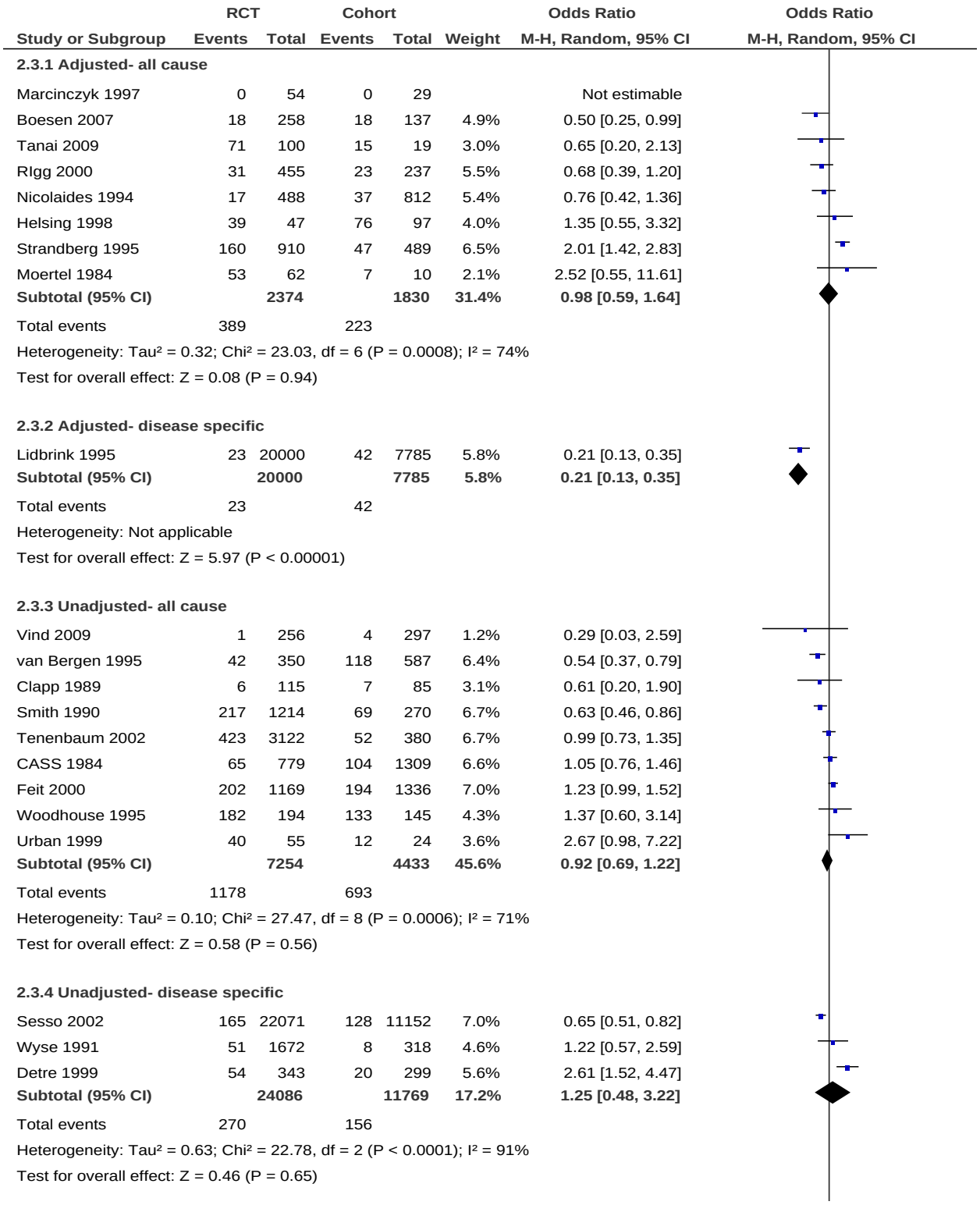
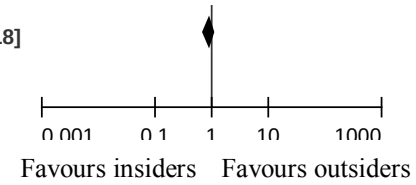


Figure 11: All studies in this meta-analysis randomly assigned participants either to the trial or cohort group and had a dichotomous non-mortality outcome. Studies in this forest plot are divided into subgroups based on baseline characteristics to explain heterogeneity.

1.3 Subgroups based on outcome





Total (95% CI)	**53714**	**25817**	**100.0%**	**0.91 [0.70, 1.18]**	
Total events	1860	1114			

Heterogeneity: $Tau^2 = 0.24$; $Chi^2 = 109.98$, df = 19 ($P < 0.00001$); $I^2 = 83\%$

Test for overall effect: $Z = 0.72$ ($P = 0.47$)

Test for subgroup differences: $Chi^2 = 27.69$, df = 3 ($P < 0.00001$), $I^2 = 89.2\%$

Figure 12: All studies in this meta-analysis non- randomly assigned participants either to the trial or cohort group and had mortality as an outcome. Studies in this forest plot are divided into subgroups based on study outcome to explain heterogeneity.

Study or Subgroup	log[Risk Ratio]	SE	RCT Total	Cohort Total	Weight	Risk Ratio IV, Random, 95% CI
2.4.1 Non- patient reported						
Akaza 1995	2.399577	1.393859	107	13	0.1%	11.02 [0.72, 169.27]
Amar 1997	-0.18809	0.330284	70	40	1.1%	0.83 [0.43, 1.58]
Antman 1985	-0.62571	0.322026	42	24	1.1%	0.53 [0.28, 1.01]
Balmukhanov 1989	-0.54098	0.418447	108	287	0.8%	0.58 [0.26, 1.32]
Bell 2000	-0.05129	0.875094	59	56	0.2%	0.95 [0.17, 5.28]
Biasoli 2008	-1.33123	0.919898	52	41	0.2%	0.26 [0.04, 1.60]
Biederman 1985	-0.17435	0.328435	24	18	1.1%	0.84 [0.44, 1.60]
Bijker 2002	-0.36792	0.329306	268	155	1.1%	0.69 [0.36, 1.32]
Blichert- Toft 1988	1.155294	0.321989	619	136	1.1%	3.17 [1.69, 5.97]
Blumenthal 1997	-0.71548	0.366976	66	38	0.9%	0.49 [0.24, 1.00]
Caplan 1984	-0.07632	0.057088	29	46	3.9%	0.93 [0.83, 1.04]
Chauhan 1992	-0.5545	0.773029	38	15	0.3%	0.57 [0.13, 2.61]
Chesebro 1983	-0.10215	0.52802	351	183	0.5%	0.90 [0.32, 2.54]
Chilvers 2001	-0.04946	0.199803	98	207	2.1%	0.95 [0.64, 1.41]
Clagett 1984	-1.97981	1.489261	29	28	0.1%	0.14 [0.01, 2.56]
Clemens 1992	-0.16896	0.084346	20744	21943	3.5%	0.84 [0.72, 1.00]
Cowchock 1992	-0.20479	0.533847	20	13	0.5%	0.81 [0.29, 2.32]
Creutzig 1993	0.451606	0.364125	31	25	0.9%	1.57 [0.77, 3.21]
Diehl 1995	-0.61837	0.304074	100	21	1.2%	0.54 [0.30, 0.98]
Eberhardt 1996	-0.08688	0.103497	43	37	3.3%	0.92 [0.75, 1.12]
Forbes 2000	0.093137	0.275068	102	88	1.4%	1.10 [0.64, 1.88]
Gall 2007	0.458067	0.442399	46	41	0.7%	1.58 [0.66, 3.76]
Goodkin 1987	-0.35256	0.144544	27	24	2.7%	0.70 [0.53, 0.93]
Gossop 1986	0.409538	0.243326	20	40	1.6%	1.51 [0.93, 2.43]
Henriksson 1986	-0.69315	0.613259	91	9	0.4%	0.50 [0.15, 1.66]
Kane 1988	1.498178	0.379098	59	116	0.9%	4.47 [2.13, 9.40]
Karande 1999	-0.19608	0.254808	63	57	1.6%	0.82 [0.50, 1.35]
Kayser 2008	0.419398	0.158367	31	44	2.5%	1.52 [1.12, 2.07]
Kirke 1992	-2.0381	0.969636	351	106	0.2%	0.13 [0.02, 0.87]
Lichtenberg 2008	0.251739	0.082639	217	153	3.6%	1.29 [1.09, 1.51]
Link 1991	-0.28136	0.231968	36	77	1.7%	0.75 [0.48, 1.19]
Liu 1998	-0.54676	1.026979	169	163	0.1%	0.58 [0.08, 4.33]
MacLennan 1985	-0.01933	0.077819	96	73	3.6%	0.98 [0.84, 1.14]
MacMillan 1986	-0.08445	0.125016	107	49	3.0%	0.92 [0.72, 1.17]
Martin 1994	-0.94143	1.620867	46	54	0.1%	0.39 [0.02, 9.35]
Mayo Group 1992	1.926181	0.394957	71	87	0.8%	6.86 [3.16, 14.88]
Morrison 2002	0.075619	0.08478	454	302	3.5%	1.08 [0.91, 1.27]
Nagel 1998	0.262743	0.669698	115	95	0.3%	1.30 [0.35, 4.83]
Ogden 2004	0.526329	0.19616	285	47	2.1%	1.69 [1.15, 2.49]
Panagopoulou 2009	0.258369	0.104057	148	66	3.3%	1.29 [1.06, 1.59]
Peteren 2007	-1.85419	0.590818	79	33	0.4%	0.16 [0.05, 0.50]
Raistrick 2005	-0.04419	0.073987	174	225	3.7%	0.96 [0.83, 1.11]
Rosen 1987	-0.18334	0.049179	98	44	3.9%	0.83 [0.76, 0.92]
Shain 1989	-1.04252	0.328532	155	98	1.1%	0.35 [0.19, 0.67]
Stern 2003	-0.03348	0.013236	555	1788	4.2%	0.97 [0.94, 0.99]
Stith 2004	-1.89712	0.903696	19	4	0.2%	0.15 [0.03, 0.88]
Sullivan 1982	0.385497	0.635155	144	25	0.4%	1.47 [0.42, 5.11]
Sundar 2008	-1.09134	1.992727	136	45	0.0%	0.34 [0.01, 16.68]
Taplin 1986	-0.07555	0.197764	63	30	2.1%	0.93 [0.63, 1.37]

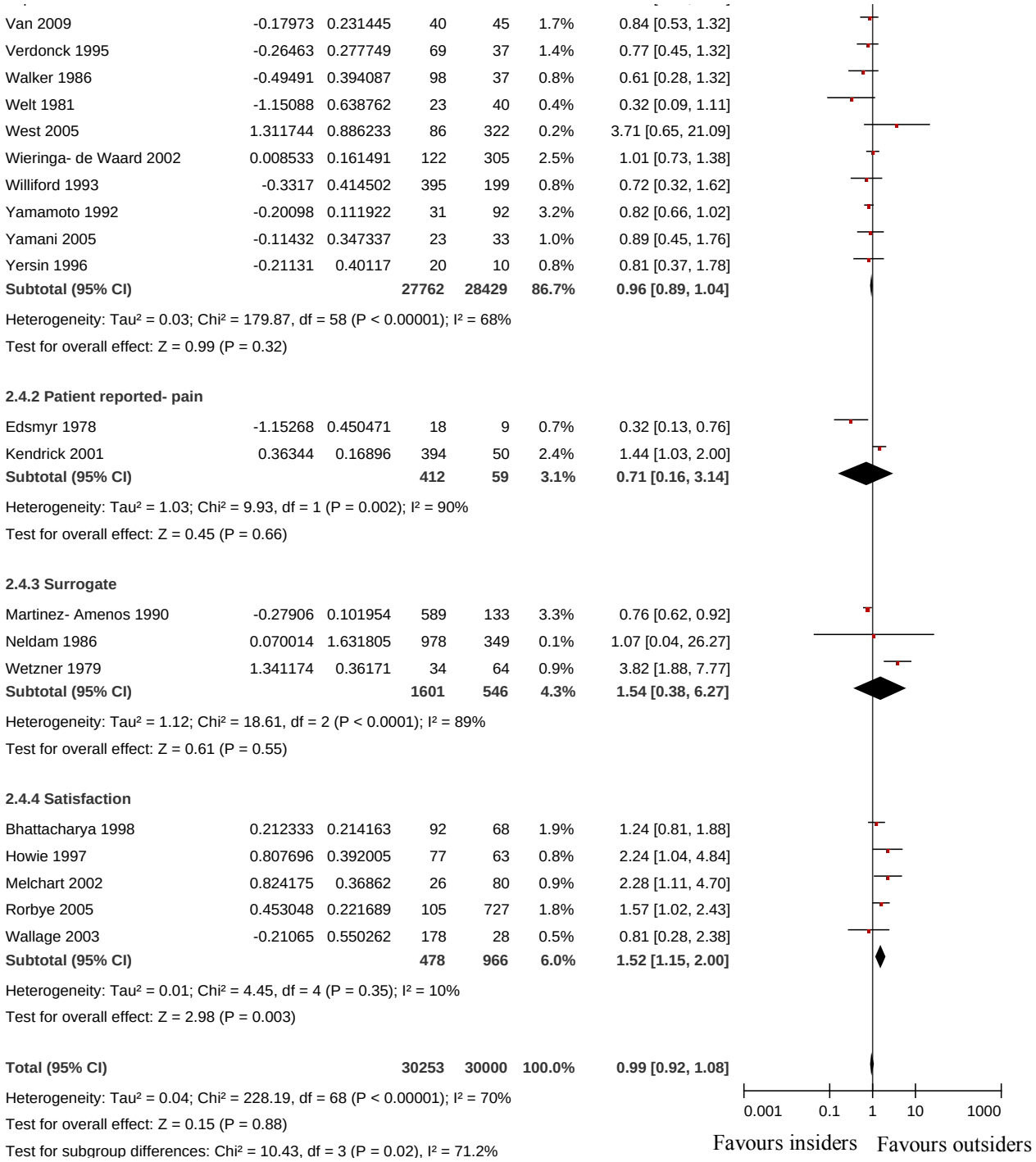
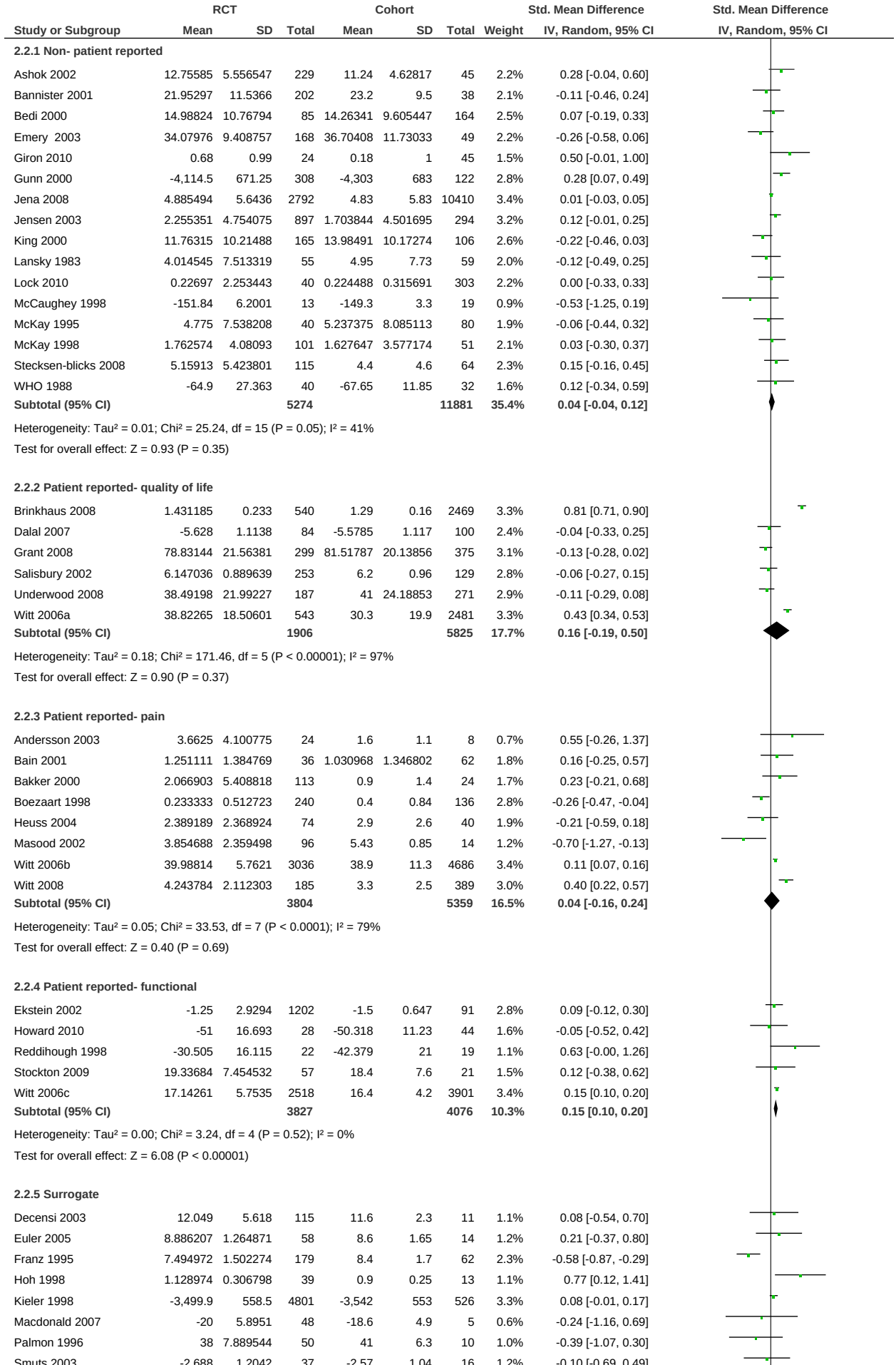


Figure 13: All studies in this meta-analysis non-randomly assigned participants either to the trial or cohort group and had a dichotomous non-mortality outcome. Studies in this forest plot are divided into subgroups based on study outcome to explain heterogeneity.


RCT
Cohort
Std. Mean Difference
Std. Mean Difference
Study or Subgroup Mean SD Total Mean SD Total Weight IV, Random, 95% CI
IV, Random, 95% CI
2.2.1 Non- patient reported
Ashok 2002 12.75585 5.556547 229 11.24 4.62817 45 2.2% 0.28 [-0.04, 0.60]
Bannister 2001 21.95297 11.5366 202 23.2 9.5 38 2.1% -0.11 [-0.46, 0.24]
Bedi 2000 14.98824 10.76794 85 14.26341 9.605447 164 2.5% 0.07 [-0.19, 0.33]
Emery 2003 34.07976 9.408757 168 36.70408 11.73033 49 2.2% -0.26 [-0.58, 0.06]
Giron 2010 0.68 0.99 24 0.18 1 45 1.5% 0.50 [-0.01, 1.00]
Gunn 2000 -4,114.5 671.25 308 -4,303 683 122 2.8% 0.28 [0.07, 0.49]
Jena 2008 4.885494 5.6436 2792 4.83 5.83 10410 3.4% 0.01 [-0.03, 0.05]
Jensen 2003 2.255351 4.754075 897 1.703844 4.501695 294 3.2% 0.12 [-0.01, 0.25]
King 2000 11.76315 10.21488 165 13.98491 10.17274 106 2.6% -0.22 [-0.46, 0.03]
Lansky 1983 4.014545 7.513319 55 4.95 7.73 59 2.0% -0.12 [-0.49, 0.25]
Lock 2010 0.22697 2.253443 40 0.224488 0.315691 303 2.2% 0.00 [-0.33, 0.33]
McCaughey 1998 -151.84 6.2001 13 -149.3 3.3 19 0.9% -0.53 [-1.25, 0.19]
McKay 1995 4.775 7.538208 40 5.237375 8.085113 80 1.9% -0.06 [-0.44, 0.32]
McKay 1998 1.762574 4.08093 101 1.627647 3.577174 51 2.1% 0.03 [-0.30, 0.37]
Stecksen-blicks 2008 5.15913 5.423801 115 4.4 4.6 64 2.3% 0.15 [-0.16, 0.45]
WHO 1988 -64.9 27.363 40 -67.65 11.85 32 1.6% 0.12 [-0.34, 0.59]
Subtotal (95% CI) 5274 11881 35.4% 0.04 [-0.04, 0.12]
Heterogeneity: Tau² = 0.01; Chi² = 25.24, df = 15 (P = 0.05); I² = 41%
Test for overall effect: Z = 0.93 (P = 0.35)
2.2.2 Patient reported- quality of life
Brinkhaus 2008 1.431185 0.233 540 1.29 0.16 2469 3.3% 0.81 [0.71, 0.90]
Dalal 2007 -5.628 1.1138 84 -5.5785 1.117 100 2.4% -0.04 [-0.33, 0.25]
Grant 2008 78.83144 21.56381 299 81.51787 20.13856 375 3.1% -0.13 [-0.28, 0.02]
Salisbury 2002 6.147036 0.889639 253 6.2 0.96 129 2.8% -0.06 [-0.27, 0.15]
Underwood 2008 38.49198 21.99227 187 41 24.18853 271 2.9% -0.11 [-0.29, 0.08]
Witt 2006a 38.82265 18.50601 543 30.3 19.9 2481 3.3% 0.43 [0.34, 0.53]
Subtotal (95% CI) 1906 5825 17.7% 0.16 [-0.19, 0.50]
Heterogeneity: Tau² = 0.18; Chi² = 171.46, df = 5 (P < 0.00001); I² = 97%
Test for overall effect: Z = 0.90 (P = 0.37)
2.2.3 Patient reported- pain
Andersson 2003 3.6625 4.100775 24 1.6 1.1 8 0.7% 0.55 [-0.26, 1.37]
Bain 2001 1.251111 1.384769 36 1.030968 1.346802 62 1.8% 0.16 [-0.25, 0.57]
Bakker 2000 2.066903 5.408818 113 0.9 1.4 24 1.7% 0.23 [-0.21, 0.68]
Boezaart 1998 0.233333 0.512723 240 0.4 0.84 136 2.8% -0.26 [-0.47, -0.04]
Heuss 2004 2.389189 2.368924 74 2.9 2.6 40 1.9% -0.21 [-0.59, 0.18]
Masood 2002 3.854688 2.359498 96 5.43 0.85 14 1.2% -0.70 [-1.27, -0.13]
Witt 2006b 39.98814 5.7621 3036 38.9 11.3 4686 3.4% 0.11 [0.07, 0.16]
Witt 2008 4.243784 2.112303 185 3.3 2.5 389 3.0% 0.40 [0.22, 0.57]
Subtotal (95% CI) 3804 5359 16.5% 0.04 [-0.16, 0.24]
Heterogeneity: Tau² = 0.05; Chi² = 33.53, df = 7 (P < 0.0001); I² = 79%
Test for overall effect: Z = 0.40 (P = 0.69)
2.2.4 Patient reported- functional
Ekstein 2002 -1.25 2.9294 1202 -1.5 0.647 91 2.8% 0.09 [-0.12, 0.30]
Howard 2010 -51 16.693 28 -50.318 11.23 44 1.6% -0.05 [-0.52, 0.42]
Reddihough 1998 -30.505 16.115 22 -42.379 21 19 1.1% 0.63 [-0.00, 1.26]
Stockton 2009 19.33684 7.454532 57 18.4 7.6 21 1.5% 0.12 [-0.38, 0.62]
Witt 2006c 17.14261 5.7535 2518 16.4 4.2 3901 3.4% 0.15 [0.10, 0.20]
Subtotal (95% CI) 3827 4076 10.3% 0.15 [0.10, 0.20]
Heterogeneity: Tau² = 0.00; Chi² = 3.24, df = 4 (P = 0.52); I² = 0%
Test for overall effect: Z = 6.08 (P < 0.00001)
2.2.5 Surrogate
Decensi 2003 12.049 5.618 115 11.6 2.3 11 1.1% 0.08 [-0.54, 0.70]
Euler 2005 8.886207 1.264871 58 8.6 1.65 14 1.2% 0.21 [-0.37, 0.80]
Franz 1995 7.494972 1.502274 179 8.4 1.7 62 2.3% -0.58 [-0.87, -0.29]
Hoh 1998 1.128974 0.306798 39 0.9 0.25 13 1.1% 0.77 [0.12, 1.41]
Kieler 1998 -3,499.9 558.5 4801 -3,542 553 526 3.3% 0.08 [-0.01, 0.17]
Macdonald 2007 -20 5.8951 48 -18.6 4.9 5 0.6% -0.24 [-1.16, 0.69]
Palmon 1996 38 7.889544 50 41 6.3 10 1.0% -0.39 [-1.07, 0.30]
Smuts 2003 -2.688 1.2042 37 -2.57 1.04 16 1.2% -0.10 [-0.69, 0.49]

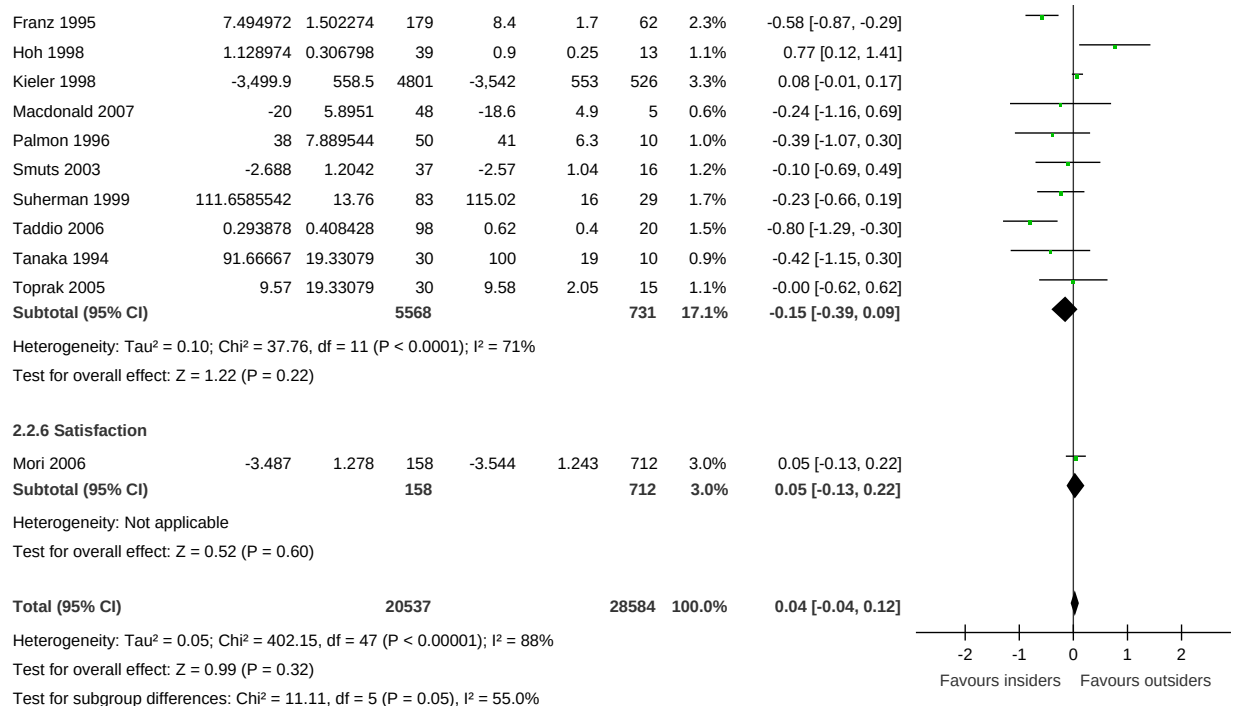


Figure 14: All studies in this meta-analysis non- randomly assigned participants either to the trial or cohort group and had a continuous mean score as an outcome. Studies in this forest plot are divided into subgroups based on study outcome to explain heterogeneity. Clinically important outcomes were defined as any outcome that is important from the patient's perspective. Surrogate outcomes are those outcomes that are not directly important to patients but that may predict future clinically important outcomes. Patient reported quality of life outcomes measure a combination of physical, mental and social constructs. The satisfaction outcome reflects how well tolerated the procedure was for the patient. Note that negative signs were added to those scores that were based on a scale where lower scores indicated higher disease severity.

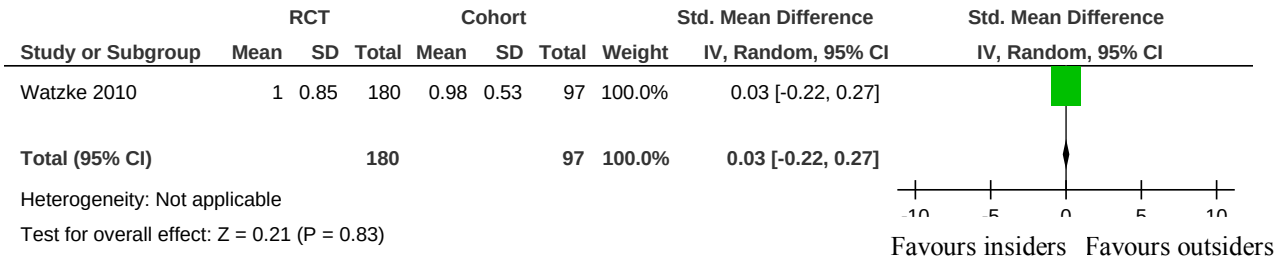


Figure 15: All studies in this meta-analysis randomized participants either to the trial or cohort group and had a continuous mean score as an outcome. Studies in this forest plot are divided into subgroups based on study outcome to explain heterogeneity.

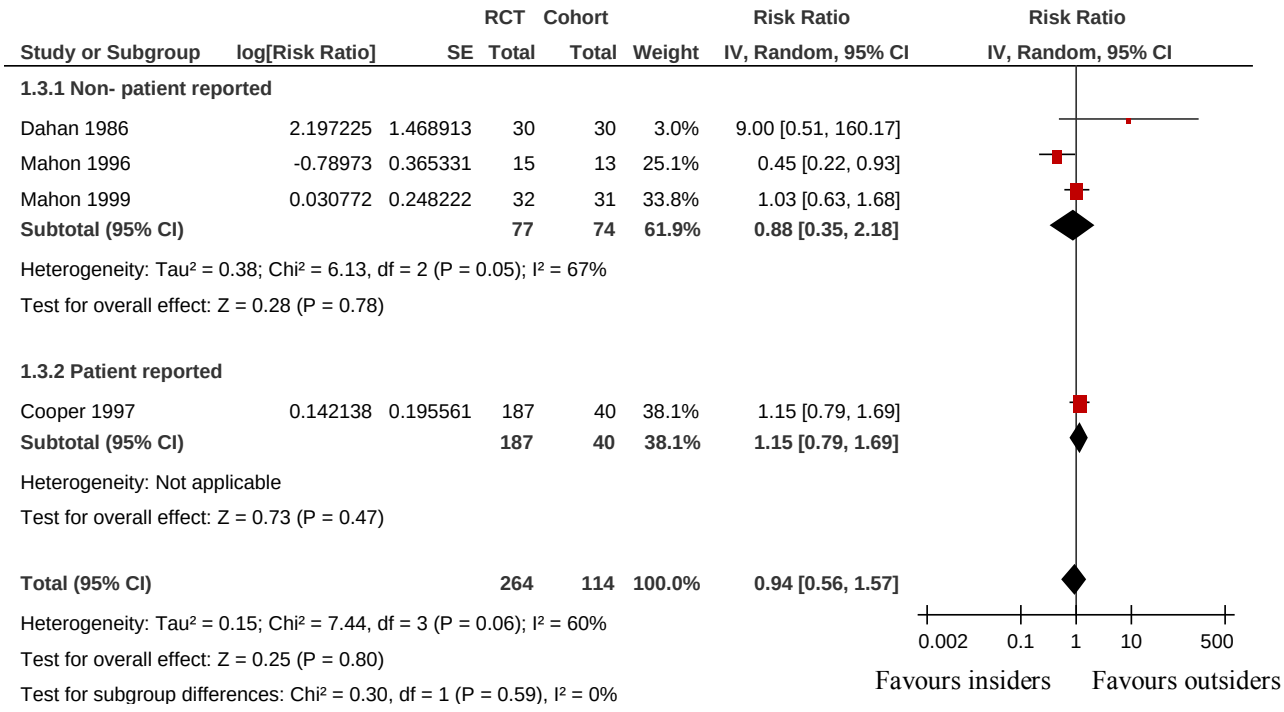


Figure 16: All studies in this meta-analysis randomized participants either to the trial or cohort group and had a dichotomous non-mortality outcome. Studies in this forest plot are divided into subgroups based on study outcome to explain heterogeneity. Clinically important outcomes were defined as any outcome that is important from the patient's perspective. Patient reported quality of life outcomes measure a combination of physical, mental and social constructs.

1.4 Subgroups based on methodological features

Study or Subgroup	RCT Events	RCT Total	Cohort Events	Cohort Total	Weight	Odds Ratio M-H, Random, 95% CI
2.3.1 In parallel + ITT						
Vind 2009	1	256	4	297	1.2%	0.29 [0.03, 2.59]
Boesen 2007	18	258	18	137	4.9%	0.50 [0.25, 0.99]
Clapp 1989	6	115	7	85	3.1%	0.61 [0.20, 1.90]
Smith 1990	217	1214	69	270	6.7%	0.63 [0.46, 0.86]
Feit 2000	202	1169	194	1336	7.0%	1.23 [0.99, 1.52]
Woodhouse 1995	182	194	133	145	4.3%	1.37 [0.60, 3.14]
Urban 1999	40	55	12	24	3.6%	2.67 [0.98, 7.22]
Subtotal (95% CI)		**3261**		**2294**	**30.8%**	**0.90 [0.58, 1.39]**
Total events	666		437			
Heterogeneity: Tau² = 0.19; Chi² = 21.96, df = 6 (P = 0.001); I² = 73%						
Test for overall effect: Z = 0.47 (P = 0.64)						
2.3.2 In parallel + No ITT						
Marcinczyk 1997	0	54	0	29		Not estimable
Lidbrink 1995	23	20000	42	7785	5.8%	0.21 [0.13, 0.35]
van Bergen 1995	42	350	118	587	6.4%	0.54 [0.37, 0.79]
Sesso 2002	165	22071	128	11152	7.0%	0.65 [0.51, 0.82]
Tanai 2009	71	100	15	19	3.0%	0.65 [0.20, 2.13]
RIgg 2000	31	455	23	237	5.5%	0.68 [0.39, 1.20]
Nicolaides 1994	17	488	37	812	5.4%	0.76 [0.42, 1.36]
Tenenbaum 2002	423	3122	52	380	6.7%	0.99 [0.73, 1.35]
CASS 1984	65	779	104	1309	6.6%	1.05 [0.76, 1.46]
Wyse 1991	51	1672	8	318	4.6%	1.22 [0.57, 2.59]
Helsing 1998	39	47	76	97	4.0%	1.35 [0.55, 3.32]
Strandberg 1995	160	910	47	489	6.5%	2.01 [1.42, 2.83]
Moertel 1984	53	62	7	10	2.1%	2.52 [0.55, 11.61]
Subtotal (95% CI)		**50110**		**23224**	**63.6%**	**0.83 [0.59, 1.17]**
Total events	1140		657			
Heterogeneity: Tau² = 0.27; Chi² = 69.99, df = 11 (P < 0.00001); I² = 84%						
Test for overall effect: Z = 1.07 (P = 0.29)						
2.3.3 Within 2 months						
Detre 1999	54	343	20	299	5.6%	2.61 [1.52, 4.47]
Subtotal (95% CI)		**343**		**299**	**5.6%**	**2.61 [1.52, 4.47]**
Total events	54		20			
Heterogeneity: Not applicable						
Test for overall effect: Z = 3.49 (P = 0.0005)						
Total (95% CI)		**53714**		**25817**	**100.0%**	**0.91 [0.70, 1.18]**
Total events	1860		1114			
Heterogeneity: Tau² = 0.24; Chi² = 109.98, df = 19 (P < 0.00001); I² = 83%						
Test for overall effect: Z = 0.72 (P = 0.47)						
Test for subgroup differences: Chi² = 13.20, df = 2 (P = 0.001), I² = 84.8%						

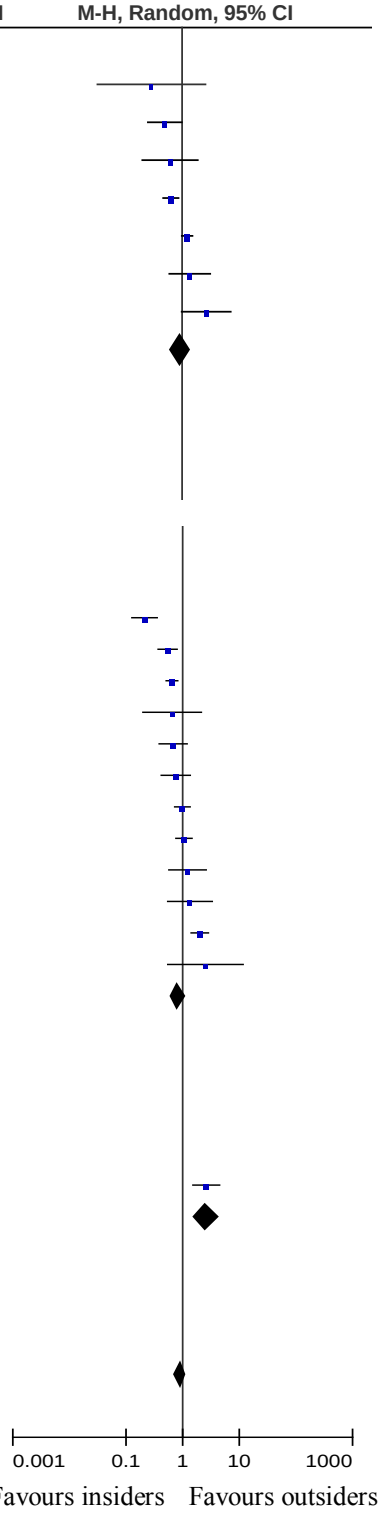
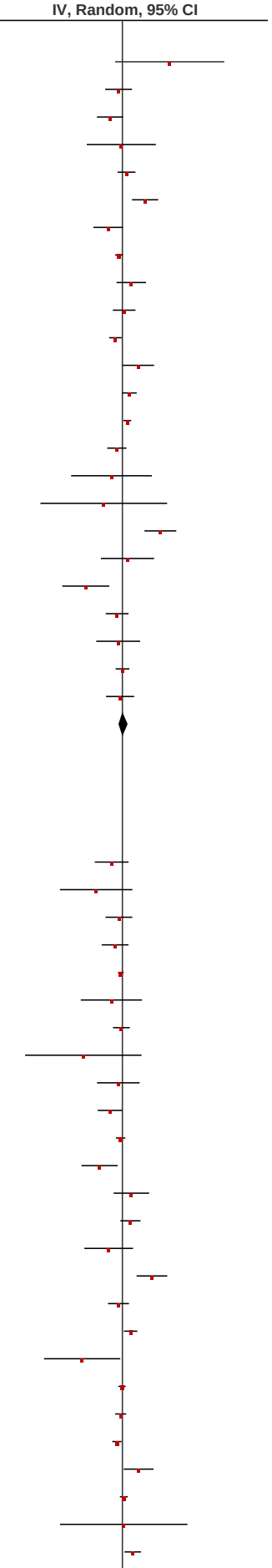


Figure17: All studies in this meta-analysis non- randomly assigned participants to either a trial or cohort group and had a mortality outcome. Studies in this forest plot are divided into subgroups based on methodological features to explain heterogeneity.

Study or Subgroup	log[Risk Ratio]	SE	RCT Total	Cohort Total	Weight	Risk Ratio IV, Random, 95% CI
2.4.1 In parallel + ITT						
Akaza 1995	2.399577	1.393859	107	13	0.1%	11.02 [0.72, 169.27]
Amar 1997	-0.18809	0.330284	70	40	1.1%	0.83 [0.43, 1.58]
Antman 1985	-0.62571	0.322026	42	24	1.1%	0.53 [0.28, 1.01]
Bell 2000	-0.05129	0.875094	59	56	0.2%	0.95 [0.17, 5.28]
Bhattacharya 1998	0.212333	0.214163	92	68	1.9%	1.24 [0.81, 1.88]
Blichert- Toft 1988	1.155294	0.321989	619	136	1.1%	3.17 [1.69, 5.97]
Blumenthal 1997	-0.71548	0.366976	66	38	0.9%	0.49 [0.24, 1.00]
Clemens 1992	-0.16896	0.084346	20744	21943	3.5%	0.84 [0.72, 1.00]
Creutzig 1993	0.451606	0.364125	31	25	0.9%	1.57 [0.77, 3.21]
Forbes 2000	0.093137	0.275068	102	88	1.4%	1.10 [0.64, 1.88]
Goodkin 1987	-0.35256	0.144544	27	24	2.7%	0.70 [0.53, 0.93]
Howie 1997	0.807696	0.392005	77	63	0.8%	2.24 [1.04, 4.84]
Kendrick 2001	0.36344	0.16896	394	50	2.4%	1.44 [1.03, 2.00]
Lichtenberg 2008	0.251739	0.082639	217	153	3.6%	1.29 [1.09, 1.51]
Link 1991	-0.28136	0.231968	36	77	1.7%	0.75 [0.48, 1.19]
Liu 1998	-0.54676	1.026979	169	163	0.1%	0.58 [0.08, 4.33]
Martin 1994	-0.94143	1.620867	46	54	0.1%	0.39 [0.02, 9.35]
Mayo Group 1992	1.926181	0.394957	71	87	0.8%	6.86 [3.16, 14.88]
Nagel 1998	0.262743	0.669698	115	95	0.3%	1.30 [0.35, 4.83]
Peteren 2007	-1.85419	0.590818	79	33	0.4%	0.16 [0.05, 0.50]
Verdonck 1995	-0.26463	0.277749	69	37	1.4%	0.77 [0.45, 1.32]
Wallage 2003	-0.21065	0.550262	178	28	0.5%	0.81 [0.28, 2.38]
Wieringa- de Waard 2002	0.008533	0.161491	122	305	2.5%	1.01 [0.73, 1.38]
Yamani 2005	-0.11432	0.347337	23	33	1.0%	0.89 [0.45, 1.76]
Subtotal (95% CI)			**23555**	**23633**	**30.7%**	**1.06 [0.86, 1.31]**
Heterogeneity: Tau² = 0.15; Chi² = 90.85, df = 23 (P < 0.00001); I² = 75%						
Test for overall effect: Z = 0.55 (P = 0.59)						
2.4.2 In parallel + No ITT						
Balmukhanov 1989	-0.54098	0.418447	108	287	0.8%	0.58 [0.26, 1.32]
Biasoli 2008	-1.33123	0.919898	52	41	0.2%	0.26 [0.04, 1.60]
Biederman 1985	-0.17435	0.328435	24	18	1.1%	0.84 [0.44, 1.60]
Bijker 2002	-0.36792	0.329306	268	155	1.1%	0.69 [0.36, 1.32]
Caplan 1984	-0.07632	0.057088	29	46	3.9%	0.93 [0.83, 1.04]
Chauhan 1992	-0.5545	0.773029	38	15	0.3%	0.57 [0.13, 2.61]
Chilvers 2001	-0.04946	0.199803	98	207	2.1%	0.95 [0.64, 1.41]
Clagett 1984	-1.97981	1.489261	29	28	0.1%	0.14 [0.01, 2.56]
Cowchock 1992	-0.20479	0.533847	20	13	0.5%	0.81 [0.29, 2.32]
Diehl 1995	-0.61837	0.304074	100	21	1.2%	0.54 [0.30, 0.98]
Eberhardt 1996	-0.08688	0.103497	43	37	3.3%	0.92 [0.75, 1.12]
Edsmyr 1978	-1.15268	0.450471	18	9	0.7%	0.32 [0.13, 0.76]
Gall 2007	0.458067	0.442399	46	41	0.7%	1.58 [0.66, 3.76]
Gossop 1986	0.409538	0.243326	20	40	1.6%	1.51 [0.93, 2.43]
Henriksson 1986	-0.69315	0.613259	91	9	0.4%	0.50 [0.15, 1.66]
Kane 1988	1.498178	0.379098	59	116	0.9%	4.47 [2.13, 9.40]
Karande 1999	-0.19608	0.254808	63	57	1.6%	0.82 [0.50, 1.35]
Kayser 2008	0.419398	0.158367	31	44	2.5%	1.52 [1.12, 2.07]
Kirke 1992	-2.0381	0.969636	351	106	0.2%	0.13 [0.02, 0.87]
MacLennan 1985	-0.01933	0.077819	96	73	3.6%	0.98 [0.84, 1.14]
MacMillan 1986	-0.08445	0.125016	107	49	3.0%	0.92 [0.72, 1.17]
Martinez- Amenos 1990	-0.27906	0.101954	589	133	3.3%	0.76 [0.62, 0.92]
Melchart 2002	0.824175	0.36862	26	80	0.9%	2.28 [1.11, 4.70]
Morrison 2002	0.075619	0.08478	454	302	3.5%	1.08 [0.91, 1.27]
Neldam 1986	0.070014	1.631805	978	349	0.1%	1.07 [0.04, 26.27]
Ogden 2004	0.526329	0.19616	285	47	2.1%	1.69 [1.15, 2.49]



Martinez- Amenos 1990	-0.27906	0.101954	589	133	3.3%	0.76 [0.62, 0.92]
Melchart 2002	0.824175	0.36862	26	80	0.9%	2.28 [1.11, 4.70]
Morrison 2002	0.075619	0.08478	454	302	3.5%	1.08 [0.91, 1.27]
Neldam 1986	0.070014	1.631805	978	349	0.1%	1.07 [0.04, 26.27]
Ogden 2004	0.526329	0.19616	285	47	2.1%	1.69 [1.15, 2.49]
Panagopoulou 2009	0.258369	0.104057	148	66	3.3%	1.29 [1.06, 1.59]
Raistrick 2005	-0.04419	0.073987	174	225	3.7%	0.96 [0.83, 1.11]
Rorbye 2005	0.453048	0.221689	105	727	1.8%	1.57 [1.02, 2.43]
Rosen 1987	-0.18334	0.049179	98	44	3.9%	0.83 [0.76, 0.92]
Shain 1989	-1.04252	0.328532	155	98	1.1%	0.35 [0.19, 0.67]
Stern 2003	-0.03348	0.013236	555	1788	4.2%	0.97 [0.94, 0.99]
Stith 2004	-1.89712	0.903696	19	4	0.2%	0.15 [0.03, 0.88]
Sullivan 1982	0.385497	0.635155	144	25	0.4%	1.47 [0.42, 5.11]
Taplin 1986	-0.07555	0.197764	63	30	2.1%	0.93 [0.63, 1.37]
Van 2009	-0.17973	0.231445	40	45	1.7%	0.84 [0.53, 1.32]
Walker 1986	-0.49491	0.394087	98	37	0.8%	0.61 [0.28, 1.32]
Welt 1981	-1.15088	0.638762	23	40	0.4%	0.32 [0.09, 1.11]
West 2005	1.311744	0.886233	86	322	0.2%	3.71 [0.65, 21.09]
Wetzner 1979	1.341174	0.36171	34	64	0.9%	3.82 [1.88, 7.77]
Williford 1993	-0.3317	0.414502	395	199	0.8%	0.72 [0.32, 1.62]
Yamamoto 1992	-0.20098	0.111922	31	92	3.2%	0.82 [0.66, 1.02]
Yersin 1996	-0.21131	0.40117	20	10	0.8%	0.81 [0.37, 1.78]
Subtotal (95% CI)			**6211**	**6139**	**68.7%**	**0.97 [0.89, 1.06]**
Heterogeneity: Tau² = 0.03; Chi² = 134.13, df = 42 (P < 0.00001); I² = 69%						
Test for overall effect: Z = 0.65 (P = 0.52)						
2.4.3 Within 2 months						
Chesebro 1983	-0.10215	0.52802	351	183	0.5%	0.90 [0.32, 2.54]
Sundar 2008	-1.09134	1.992727	136	45	0.0%	0.34 [0.01, 16.68]
Subtotal (95% CI)			**487**	**228**	**0.5%**	**0.85 [0.31, 2.30]**
Heterogeneity: Tau² = 0.00; Chi² = 0.23, df = 1 (P = 0.63); I² = 0%						
Test for overall effect: Z = 0.33 (P = 0.74)						
Total (95% CI)			**30253**	**30000**	**100.0%**	**0.99 [0.92, 1.08]**
Heterogeneity: Tau² = 0.04; Chi² = 228.19, df = 68 (P < 0.00001); I² = 70%						
Test for overall effect: Z = 0.15 (P = 0.88)						
Test for subgroup differences: Chi² = 0.65, df = 2 (P = 0.72), I² = 0%						

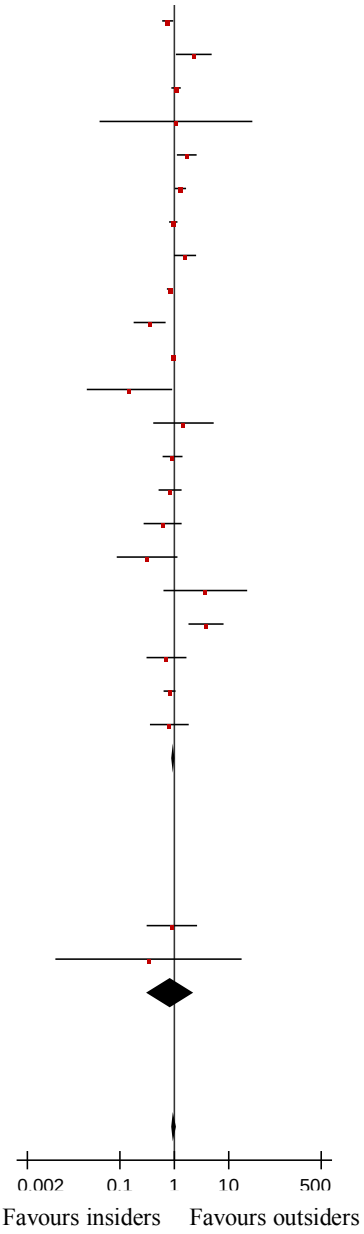
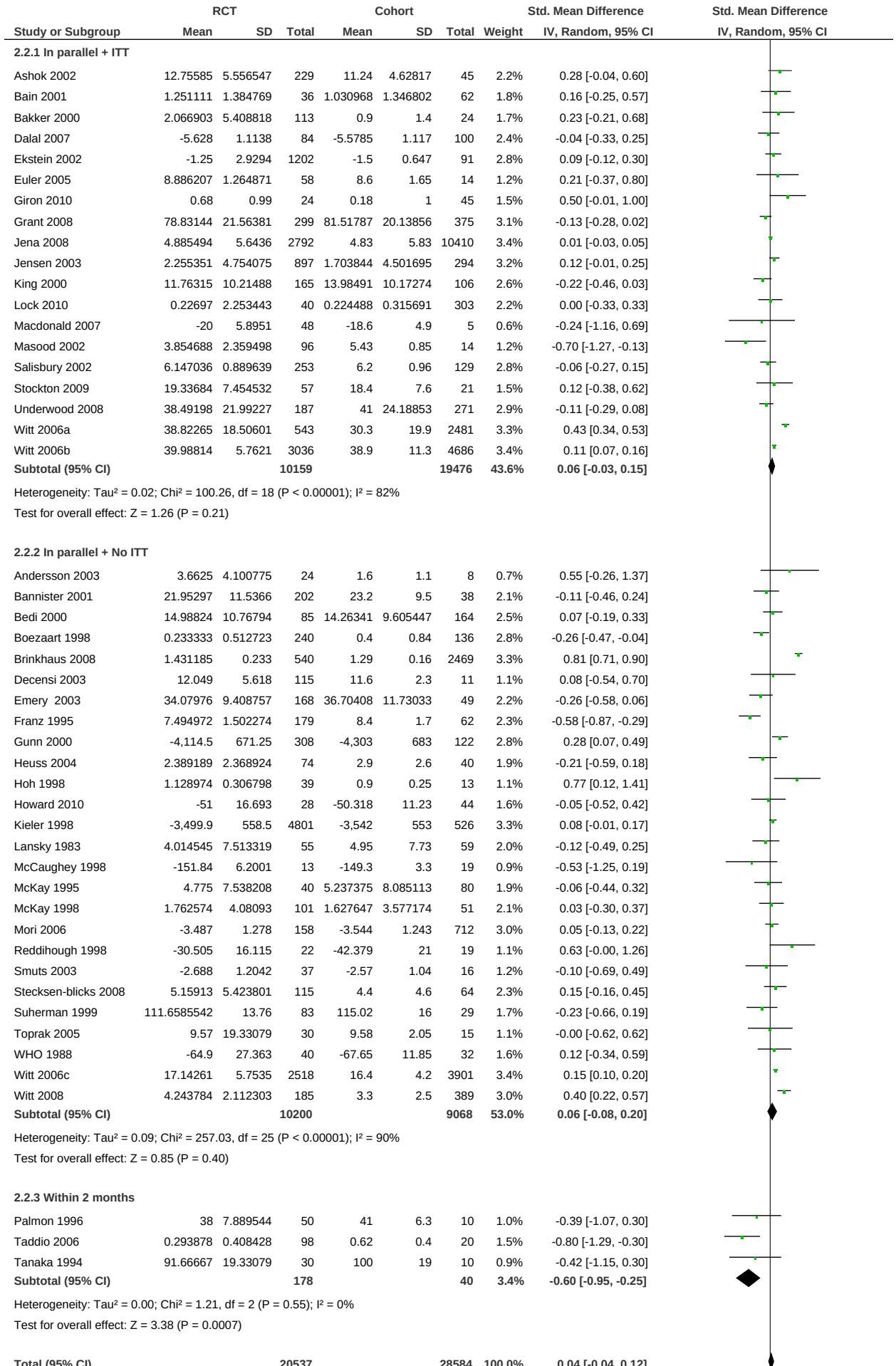


Figure 18: All studies in this meta-analysis randomized participants either to the trial or cohort group and had a dichotomous non-mortality outcome. Studies in this forest plot are divided into subgroups based on methodological features to explain heterogeneity.



Study or Subgroup	RCT Mean	RCT SD	RCT Total	Cohort Mean	Cohort SD	Cohort Total	Weight	Std. Mean Difference IV, Random, 95% CI
2.2.1 In parallel + ITT								
Ashok 2002	12.75585	5.556547	229	11.24	4.62817	45	2.2%	0.28 [-0.04, 0.60]
Bain 2001	1.251111	1.384769	36	1.030968	1.346802	62	1.8%	0.16 [-0.25, 0.57]
Bakker 2000	2.066903	5.408818	113	0.9	1.4	24	1.7%	0.23 [-0.21, 0.68]
Dalal 2007	-5.628	1.1138	84	-5.5785	1.117	100	2.4%	-0.04 [-0.33, 0.25]
Ekstein 2002	-1.25	2.9294	1202	-1.5	0.647	91	2.8%	0.09 [-0.12, 0.30]
Euler 2005	8.886207	1.264871	58	8.6	1.65	14	1.2%	0.21 [-0.37, 0.80]
Giron 2010	0.68	0.99	24	0.18	1	45	1.5%	0.50 [-0.01, 1.00]
Grant 2008	78.83144	21.56381	299	81.51787	20.13856	375	3.1%	-0.13 [-0.28, 0.02]
Jena 2008	4.885494	5.6436	2792	4.83	5.83	10410	3.4%	0.01 [-0.03, 0.05]
Jensen 2003	2.255351	4.754075	897	1.703844	4.501695	294	3.2%	0.12 [-0.01, 0.25]
King 2000	11.76315	10.21488	165	13.98491	10.17274	106	2.6%	-0.22 [-0.46, 0.03]
Lock 2010	0.22697	2.253443	40	0.224488	0.315691	303	2.2%	0.00 [-0.33, 0.33]
Macdonald 2007	-20	5.8951	48	-18.6	4.9	5	0.6%	-0.24 [-1.16, 0.69]
Masood 2002	3.854688	2.359498	96	5.43	0.85	14	1.2%	-0.70 [-1.27, -0.13]
Salisbury 2002	6.147036	0.889639	253	6.2	0.96	129	2.8%	-0.06 [-0.27, 0.15]
Stockton 2009	19.33684	7.454532	57	18.4	7.6	21	1.5%	0.12 [-0.38, 0.62]
Underwood 2008	38.49198	21.99227	187	41	24.18853	271	2.9%	-0.11 [-0.29, 0.08]
Witt 2006a	38.82265	18.50601	543	30.3	19.9	2481	3.3%	0.43 [0.34, 0.53]
Witt 2006b	39.98814	5.7621	3036	38.9	11.3	4686	3.4%	0.11 [0.07, 0.16]
Subtotal (95% CI)			**10159**			**19476**	**43.6%**	**0.06 [-0.03, 0.15]**
Heterogeneity: Tau² = 0.02; Chi² = 100.26, df = 18 (P < 0.00001); I² = 82%								
Test for overall effect: Z = 1.26 (P = 0.21)								
2.2.2 In parallel + No ITT								
Andersson 2003	3.6625	4.100775	24	1.6	1.1	8	0.7%	0.55 [-0.26, 1.37]
Bannister 2001	21.95297	11.5366	202	23.2	9.5	38	2.1%	-0.11 [-0.46, 0.24]
Bedi 2000	14.98824	10.76794	85	14.26341	9.605447	164	2.5%	0.07 [-0.19, 0.33]
Boezaart 1998	0.233333	0.512723	240	0.4	0.84	136	2.8%	-0.26 [-0.47, -0.04]
Brinkhaus 2008	1.431185	0.233	540	1.29	0.16	2469	3.3%	0.81 [0.71, 0.90]
Decensi 2003	12.049	5.618	115	11.6	2.3	11	1.1%	0.08 [-0.54, 0.70]
Emery 2003	34.07976	9.408757	168	36.70408	11.73033	49	2.2%	-0.26 [-0.58, 0.06]
Franz 1995	7.494972	1.502274	179	8.4	1.7	62	2.3%	-0.58 [-0.87, -0.29]
Gunn 2000	-4,114.5	671.25	308	-4,303	683	122	2.8%	0.28 [0.07, 0.49]
Heuss 2004	2.389189	2.368924	74	2.9	2.6	40	1.9%	-0.21 [-0.59, 0.18]
Hoh 1998	1.128974	0.306798	39	0.9	0.25	13	1.1%	0.77 [0.12, 1.41]
Howard 2010	-51	16.693	28	-50.318	11.23	44	1.6%	-0.05 [-0.52, 0.42]
Kieler 1998	-3,499.9	558.5	4801	-3,542	553	526	3.3%	0.08 [-0.01, 0.17]
Lansky 1983	4.014545	7.513319	55	4.95	7.73	59	2.0%	-0.12 [-0.49, 0.25]
McCaughey 1998	-151.84	6.2001	13	-149.3	3.3	19	0.9%	-0.53 [-1.25, 0.19]
McKay 1995	4.775	7.538208	40	5.237375	8.085113	80	1.9%	-0.06 [-0.44, 0.32]
McKay 1998	1.762574	4.08093	101	1.627647	3.577174	51	2.1%	0.03 [-0.30, 0.37]
Mori 2006	-3.487	1.278	158	-3.544	1.243	712	3.0%	0.05 [-0.13, 0.22]
Reddihough 1998	-30.505	16.115	22	-42.379	21	19	1.1%	0.63 [-0.00, 1.26]
Smuts 2003	-2.688	1.2042	37	-2.57	1.04	16	1.2%	-0.10 [-0.69, 0.49]
Stecksen-blicks 2008	5.15913	5.423801	115	4.4	4.6	64	2.3%	0.15 [-0.16, 0.45]
Suherman 1999	111.6585542	13.76	83	115.02	16	29	1.7%	-0.23 [-0.66, 0.19]
Toprak 2005	9.57	19.33079	30	9.58	2.05	15	1.1%	-0.00 [-0.62, 0.62]
WHO 1988	-64.9	27.363	40	-67.65	11.85	32	1.6%	0.12 [-0.34, 0.59]
Witt 2006c	17.14261	5.7535	2518	16.4	4.2	3901	3.4%	0.15 [0.10, 0.20]
Witt 2008	4.243784	2.112303	185	3.3	2.5	389	3.0%	0.40 [0.22, 0.57]
Subtotal (95% CI)			**10200**			**9068**	**53.0%**	**0.06 [-0.08, 0.20]**
Heterogeneity: Tau² = 0.09; Chi² = 257.03, df = 25 (P < 0.00001); I² = 90%								
Test for overall effect: Z = 0.85 (P = 0.40)								
2.2.3 Within 2 months								
Palmon 1996	38	7.889544	50	41	6.3	10	1.0%	-0.39 [-1.07, 0.30]
Taddio 2006	0.293878	0.408428	98	0.62	0.4	20	1.5%	-0.80 [-1.29, -0.30]
Tanaka 1994	91.66667	19.33079	30	100	19	10	0.9%	-0.42 [-1.15, 0.30]
Subtotal (95% CI)			**178**			**40**	**3.4%**	**-0.60 [-0.95, -0.25]**
Heterogeneity: Tau² = 0.00; Chi² = 1.21, df = 2 (P = 0.55); I² = 0%								
Test for overall effect: Z = 3.38 (P = 0.0007)								
Total (95% CI)			**20537**			**28584**	**100.0%**	**0.04 [-0.04, 0.12]**

Std. Mean Difference
IV, Random, 95% CI

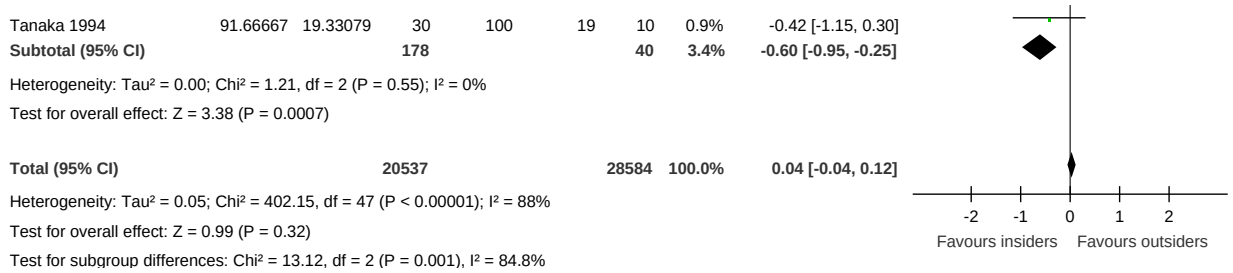


Figure 19: All studies in this meta-analysis randomized participants either to the trial or cohort group and had a continuous endpoint as an outcome. Studies in this forest plot are divided into subgroups based on methodological features to explain heterogeneity. Note that negative signs were added to those scores that were based on a scale where lower scores indicated higher disease severity.

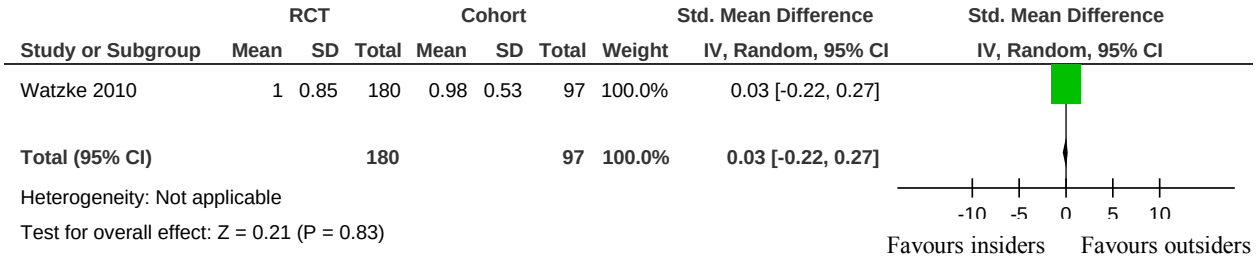


Figure 20: All studies in this meta-analysis non- randomly assigned participants to either a trial or cohort group and had a continuous endpoint as an outcome. Studies in this forest plot are divided into subgroups based on methodological features to explain heterogeneity.

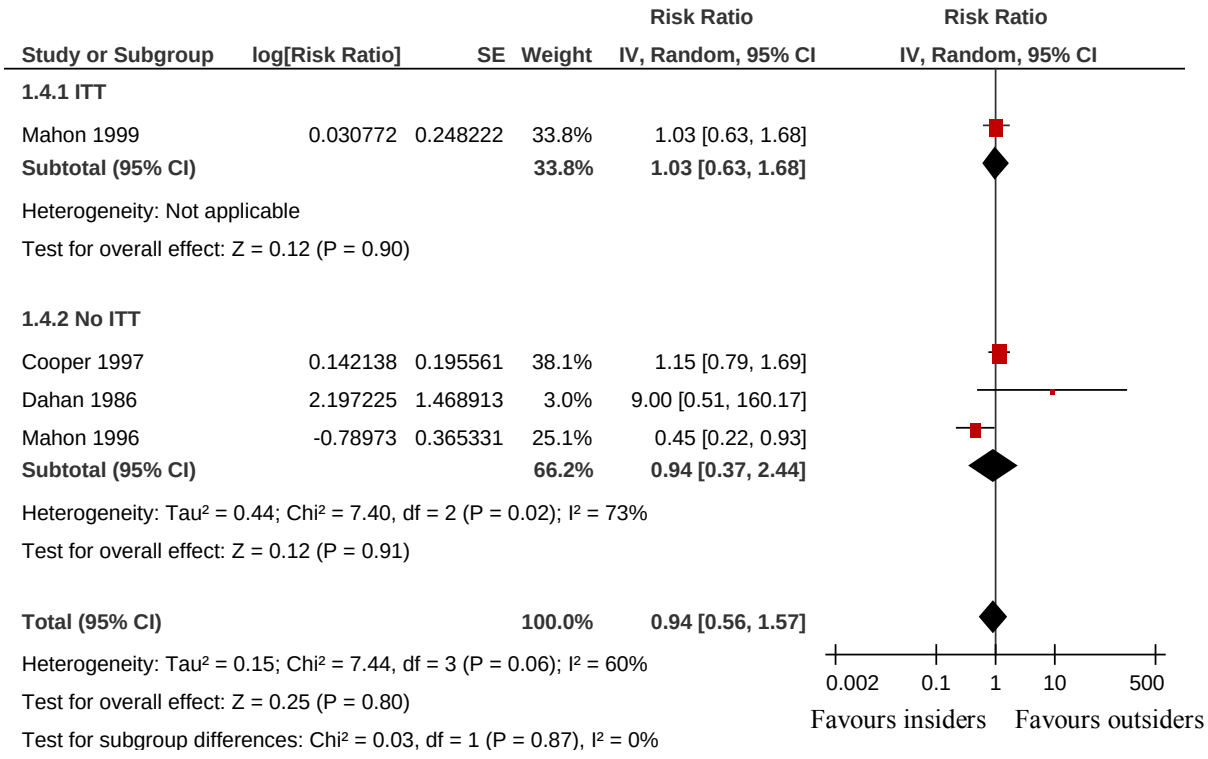


Figure 21: All studies in this meta-analysis non- randomly assigned participants to either a trial or cohort group and had a dichotomous non- mortality outcome. Studies in this forest plot are divided into subgroups based on methodological features to explain heterogeneity.

1.5 Subgroups based on potential for detection bias

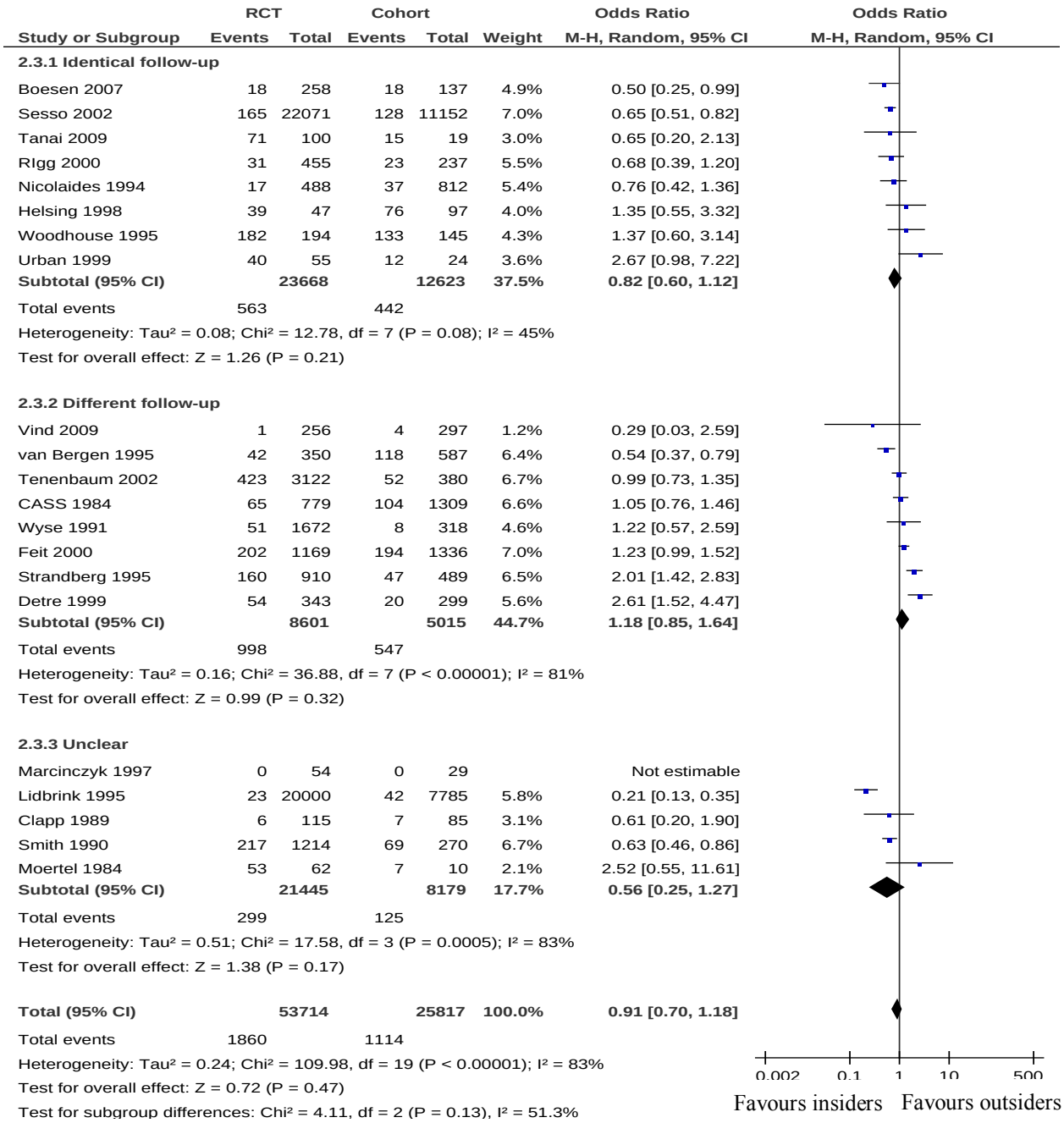


Figure 22: All studies in this meta-analysis non- randomly assigned participants to either a trial or cohort group and had mortality as an outcome. Studies in this forest plot are divided into subgroups based on potential for detection bias to explain heterogeneity.

Study or Subgroup	log[Risk Ratio]	SE	RCT Total	Cohort Total	Weight	Risk Ratio IV, Random, 95% CI
2.4.1 Identical follow-up						
Balmukhanov 1989	-0.54098	0.418447	108	287	0.8%	0.58 [0.26, 1.32]
Bhattacharya 1998	0.212333	0.214163	92	68	1.9%	1.24 [0.81, 1.88]
Biederman 1985	-0.17435	0.328435	24	18	1.1%	0.84 [0.44, 1.60]
Blumenthal 1997	-0.71548	0.366976	66	38	0.9%	0.49 [0.24, 1.00]
Caplan 1984	-0.07632	0.057088	29	46	3.9%	0.93 [0.83, 1.04]
Chauhan 1992	-0.5545	0.773029	38	15	0.3%	0.57 [0.13, 2.61]
Chesebro 1983	-0.10215	0.52802	351	183	0.5%	0.90 [0.32, 2.54]
Chilvers 2001	-0.04946	0.199803	98	207	2.1%	0.95 [0.64, 1.41]
Clagett 1984	-1.97981	1.489261	29	28	0.1%	0.14 [0.01, 2.56]
Creutzig 1993	0.451606	0.364125	31	25	0.9%	1.57 [0.77, 3.21]
Eberhardt 1996	-0.08688	0.103497	43	37	3.3%	0.92 [0.75, 1.12]
Edsmyr 1978	-1.15268	0.450471	18	9	0.7%	0.32 [0.13, 0.76]
Forbes 2000	0.093137	0.275068	102	88	1.4%	1.10 [0.64, 1.88]
Gall 2007	0.458067	0.442399	46	41	0.7%	1.58 [0.66, 3.76]
Goodkin 1987	-0.35256	0.144544	27	24	2.7%	0.70 [0.53, 0.93]
Henriksson 1986	-0.69315	0.613259	91	9	0.4%	0.50 [0.15, 1.66]
Kane 1988	1.498178	0.379098	59	116	0.9%	4.47 [2.13, 9.40]
Karande 1999	-0.19608	0.254808	63	57	1.6%	0.82 [0.50, 1.35]
Kayser 2008	0.419398	0.158367	31	44	2.5%	1.52 [1.12, 2.07]
Kirke 1992	-2.0381	0.969636	351	106	0.2%	0.13 [0.02, 0.87]
Link 1991	-0.28136	0.231968	36	77	1.7%	0.75 [0.48, 1.19]
Liu 1998	-0.54676	1.026979	169	163	0.1%	0.58 [0.08, 4.33]
Martinez- Amenos 1990	-0.27906	0.101954	589	133	3.3%	0.76 [0.62, 0.92]
Mayo Group 1992	1.926181	0.394957	71	87	0.8%	6.86 [3.16, 14.88]
Melchart 2002	0.824175	0.36862	26	80	0.9%	2.28 [1.11, 4.70]
Morrison 2002	0.075619	0.08478	454	302	3.5%	1.08 [0.91, 1.27]
Nagel 1998	0.262743	0.669698	115	95	0.3%	1.30 [0.35, 4.83]
Neldam 1986	0.070014	1.631805	978	349	0.1%	1.07 [0.04, 26.27]
Ogden 2004	0.526329	0.19616	285	47	2.1%	1.69 [1.15, 2.49]
Panagopoulou 2009	0.258369	0.104057	148	66	3.3%	1.29 [1.06, 1.59]
Peteren 2007	-1.85419	0.590818	79	33	0.4%	0.16 [0.05, 0.50]
Raistrick 2005	-0.04419	0.073987	174	225	3.7%	0.96 [0.83, 1.11]
Rorbye 2005	0.453048	0.221689	105	727	1.8%	1.57 [1.02, 2.43]
Rosen 1987	-0.18334	0.049179	98	44	3.9%	0.83 [0.76, 0.92]
Shain 1989	-1.04252	0.328532	155	98	1.1%	0.35 [0.19, 0.67]
Stith 2004	-1.89712	0.903696	19	4	0.2%	0.15 [0.03, 0.88]
Sullivan 1982	0.385497	0.635155	144	25	0.4%	1.47 [0.42, 5.11]
Sundar 2008	-1.09134	1.992727	136	45	0.0%	0.34 [0.01, 16.68]
Taplin 1986	-0.07555	0.197764	63	30	2.1%	0.93 [0.63, 1.37]
Van 2009	-0.17973	0.231445	40	45	1.7%	0.84 [0.53, 1.32]
Wallage 2003	-0.21065	0.550262	178	28	0.5%	0.81 [0.28, 2.38]
Welt 1981	-1.15088	0.638762	23	40	0.4%	0.32 [0.09, 1.11]
Wieringa- de Waard 2002	0.008533	0.161491	122	305	2.5%	1.01 [0.73, 1.38]
Yamamoto 1992	-0.20098	0.111922	31	92	3.2%	0.82 [0.66, 1.02]
Yamani 2005	-0.11432	0.347337	23	33	1.0%	0.89 [0.45, 1.76]
Yersin 1996	-0.21131	0.40117	20	10	0.8%	0.81 [0.37, 1.78]
Subtotal (95% CI)			**5978**	**4629**	**66.5%**	**0.96 [0.86, 1.08]**
Heterogeneity: Tau² = 0.07; Chi² = 151.47, df = 45 (P < 0.00001); I² = 70%						
Test for overall effect: Z = 0.64 (P = 0.52)						
2.4.2 Different follow-up						
Amar 1997	-0.18809	0.330284	70	40	1.1%	0.83 [0.43, 1.58]
Biasoli 2008	-1.33123	0.919898	52	41	0.2%	0.26 [0.04, 1.60]
Bijker 2002	-0.36792	0.329306	268	155	1.1%	0.69 [0.36, 1.32]
Lichtenberg 2008	0.251739	0.082639	217	153	3.6%	1.29 [1.09, 1.51]

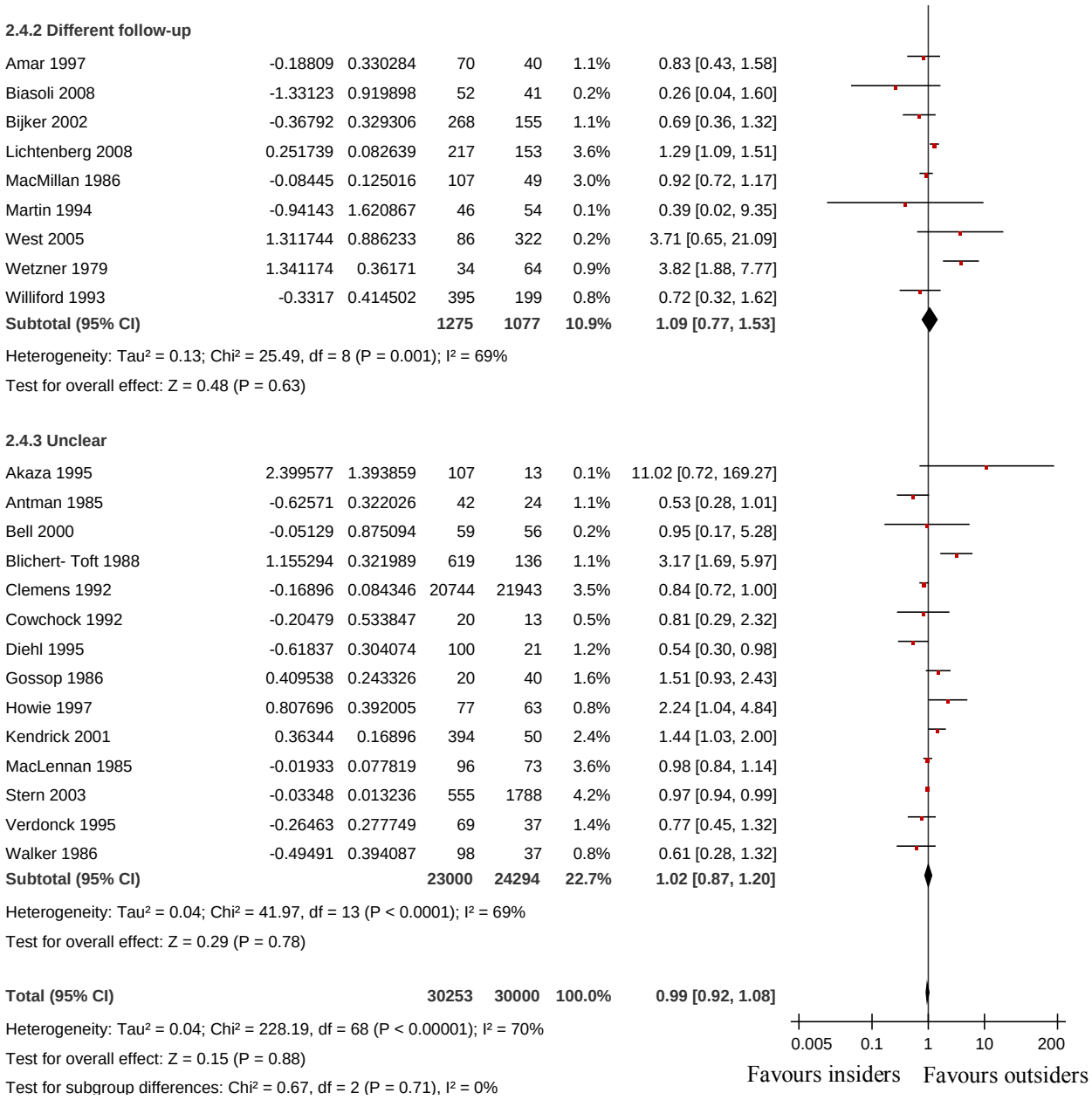
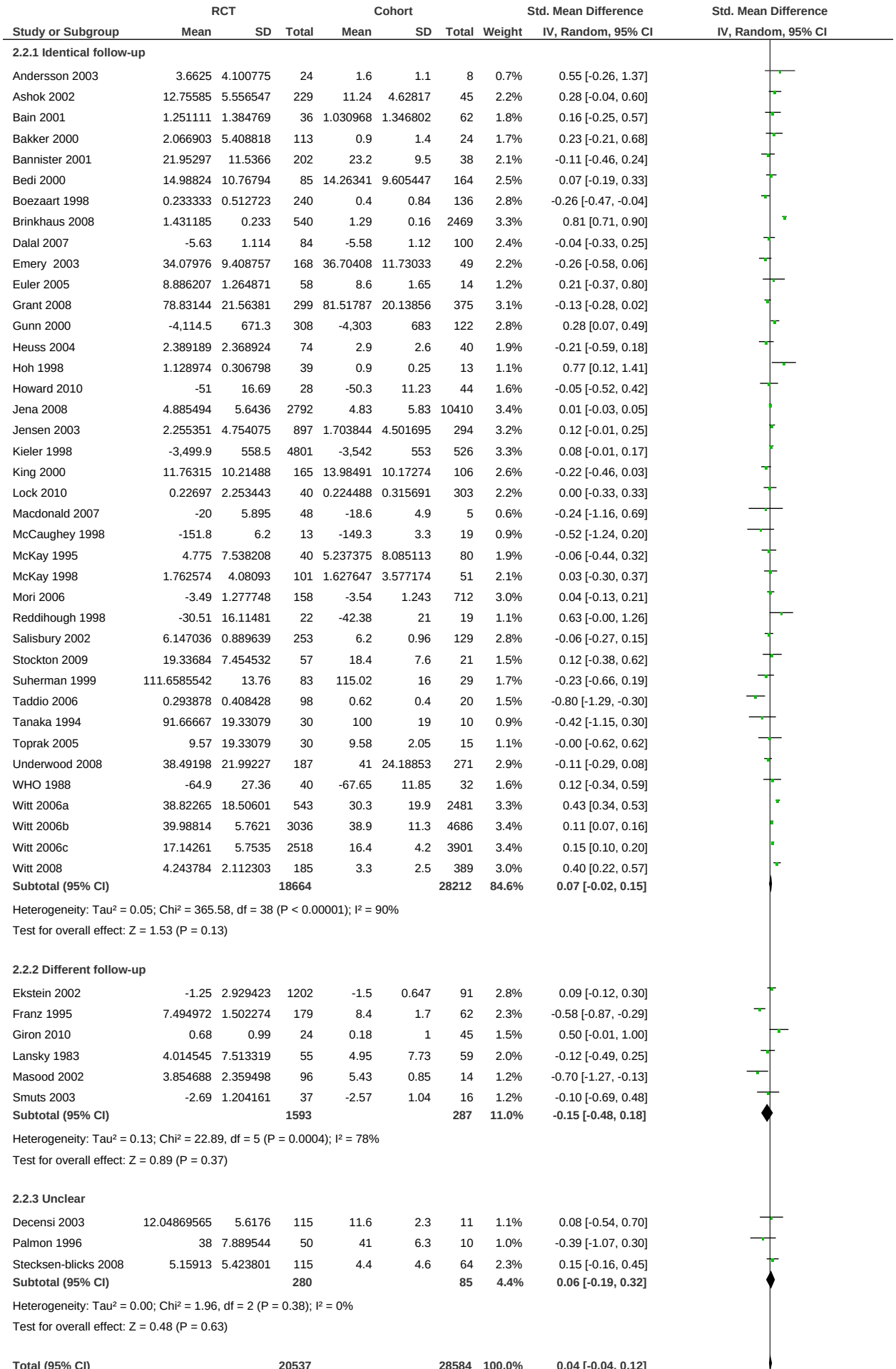


Figure 23: All studies in this meta-analysis non- randomly assigned participants to either a trial or cohort group and had a dichotomous non-mortality outcome. Studies in this forest plot are divided into subgroups based on potential for detection bias to explain heterogeneity.



Study or Subgroup	RCT Mean	RCT SD	RCT Total	Cohort Mean	Cohort SD	Cohort Total	Weight	Std. Mean Difference IV, Random, 95% CI
2.2.1 Identical follow-up								
Andersson 2003	3.6625	4.100775	24	1.6	1.1	8	0.7%	0.55 [-0.26, 1.37]
Ashok 2002	12.75585	5.556547	229	11.24	4.62817	45	2.2%	0.28 [-0.04, 0.60]
Bain 2001	1.251111	1.384769	36	1.030968	1.346802	62	1.8%	0.16 [-0.25, 0.57]
Bakker 2000	2.066903	5.408818	113	0.9	1.4	24	1.7%	0.23 [-0.21, 0.68]
Bannister 2001	21.95297	11.5366	202	23.2	9.5	38	2.1%	-0.11 [-0.46, 0.24]
Bedi 2000	14.98824	10.76794	85	14.26341	9.605447	164	2.5%	0.07 [-0.19, 0.33]
Boezaart 1998	0.233333	0.512723	240	0.4	0.84	136	2.8%	-0.26 [-0.47, -0.04]
Brinkhaus 2008	1.431185	0.233	540	1.29	0.16	2469	3.3%	0.81 [0.71, 0.90]
Dalal 2007	-5.63	1.114	84	-5.58	1.12	100	2.4%	-0.04 [-0.33, 0.25]
Emery 2003	34.07976	9.408757	168	36.70408	11.73033	49	2.2%	-0.26 [-0.58, 0.06]
Euler 2005	8.886207	1.264871	58	8.6	1.65	14	1.2%	0.21 [-0.37, 0.80]
Grant 2008	78.83144	21.56381	299	81.51787	20.13856	375	3.1%	-0.13 [-0.28, 0.02]
Gunn 2000	-4,114.5	671.3	308	-4,303	683	122	2.8%	0.28 [0.07, 0.49]
Heuss 2004	2.389189	2.368924	74	2.9	2.6	40	1.9%	-0.21 [-0.59, 0.18]
Hoh 1998	1.128974	0.306798	39	0.9	0.25	13	1.1%	0.77 [0.12, 1.41]
Howard 2010	-51	16.69	28	-50.3	11.23	44	1.6%	-0.05 [-0.52, 0.42]
Jena 2008	4.885494	5.6436	2792	4.83	5.83	10410	3.4%	0.01 [-0.03, 0.05]
Jensen 2003	2.255351	4.754075	897	1.703844	4.501695	294	3.2%	0.12 [-0.01, 0.25]
Kieler 1998	-3,499.9	558.5	4801	-3,542	553	526	3.3%	0.08 [-0.01, 0.17]
King 2000	11.76315	10.21488	165	13.98491	10.17274	106	2.6%	-0.22 [-0.46, 0.03]
Lock 2010	0.22697	2.253443	40	0.224488	0.315691	303	2.2%	0.00 [-0.33, 0.33]
Macdonald 2007	-20	5.895	48	-18.6	4.9	5	0.6%	-0.24 [-1.16, 0.69]
McCaughey 1998	-151.8	6.2	13	-149.3	3.3	19	0.9%	-0.52 [-1.24, 0.20]
McKay 1995	4.775	7.538208	40	5.237375	8.085113	80	1.9%	-0.06 [-0.44, 0.32]
McKay 1998	1.762574	4.08093	101	1.627647	3.577174	51	2.1%	0.03 [-0.30, 0.37]
Mori 2006	-3.49	1.277748	158	-3.54	1.243	712	3.0%	0.04 [-0.13, 0.21]
Reddihough 1998	-30.51	16.11481	22	-42.38	21	19	1.1%	0.63 [-0.00, 1.26]
Salisbury 2002	6.147036	0.889639	253	6.2	0.96	129	2.8%	-0.06 [-0.27, 0.15]
Stockton 2009	19.33684	7.454532	57	18.4	7.6	21	1.5%	0.12 [-0.38, 0.62]
Suherman 1999	111.6585542	13.76	83	115.02	16	29	1.7%	-0.23 [-0.66, 0.19]
Taddio 2006	0.293878	0.408428	98	0.62	0.4	20	1.5%	-0.80 [-1.29, -0.30]
Tanaka 1994	91.66667	19.33079	30	100	19	10	0.9%	-0.42 [-1.15, 0.30]
Toprak 2005	9.57	19.33079	30	9.58	2.05	15	1.1%	-0.00 [-0.62, 0.62]
Underwood 2008	38.49198	21.99227	187	41	24.18853	271	2.9%	-0.11 [-0.29, 0.08]
WHO 1988	-64.9	27.36	40	-67.65	11.85	32	1.6%	0.12 [-0.34, 0.59]
Witt 2006a	38.82265	18.50601	543	30.3	19.9	2481	3.3%	0.43 [0.34, 0.53]
Witt 2006b	39.98814	5.7621	3036	38.9	11.3	4686	3.4%	0.11 [0.07, 0.16]
Witt 2006c	17.14261	5.7535	2518	16.4	4.2	3901	3.4%	0.15 [0.10, 0.20]
Witt 2008	4.243784	2.112303	185	3.3	2.5	389	3.0%	0.40 [0.22, 0.57]
Subtotal (95% CI)			**18664**			**28212**	**84.6%**	**0.07 [-0.02, 0.15]**
Heterogeneity: Tau² = 0.05; Chi² = 365.58, df = 38 (P < 0.00001); I² = 90%								
Test for overall effect: Z = 1.53 (P = 0.13)								
2.2.2 Different follow-up								
Ekstein 2002	-1.25	2.929423	1202	-1.5	0.647	91	2.8%	0.09 [-0.12, 0.30]
Franz 1995	7.494972	1.502274	179	8.4	1.7	62	2.3%	-0.58 [-0.87, -0.29]
Giron 2010	0.68	0.99	24	0.18	1	45	1.5%	0.50 [-0.01, 1.00]
Lansky 1983	4.014545	7.513319	55	4.95	7.73	59	2.0%	-0.12 [-0.49, 0.25]
Masood 2002	3.854688	2.359498	96	5.43	0.85	14	1.2%	-0.70 [-1.27, -0.13]
Smuts 2003	-2.69	1.204161	37	-2.57	1.04	16	1.2%	-0.10 [-0.69, 0.48]
Subtotal (95% CI)			**1593**			**287**	**11.0%**	**-0.15 [-0.48, 0.18]**
Heterogeneity: Tau² = 0.13; Chi² = 22.89, df = 5 (P = 0.0004); I² = 78%								
Test for overall effect: Z = 0.89 (P = 0.37)								
2.2.3 Unclear								
Decensi 2003	12.04869565	5.6176	115	11.6	2.3	11	1.1%	0.08 [-0.54, 0.70]
Palmon 1996	38	7.889544	50	41	6.3	10	1.0%	-0.39 [-1.07, 0.30]
Stecksen-blicks 2008	5.15913	5.423801	115	4.4	4.6	64	2.3%	0.15 [-0.16, 0.45]
Subtotal (95% CI)			**280**			**85**	**4.4%**	**0.06 [-0.19, 0.32]**
Heterogeneity: Tau² = 0.00; Chi² = 1.96, df = 2 (P = 0.38); I² = 0%								
Test for overall effect: Z = 0.48 (P = 0.63)								
Total (95% CI)			**20537**			**28584**	**100.0%**	**0.04 [-0.04, 0.12]**

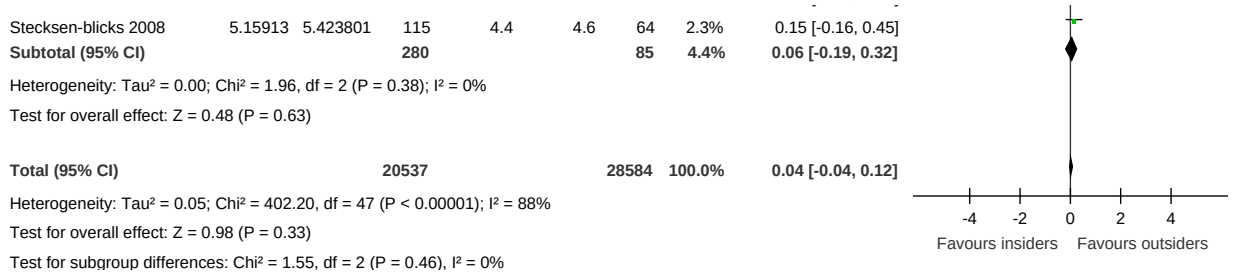


Figure 24: All studies in this meta-analysis non- randomly assigned participants to either a trial or cohort group and had a continuous mean score as an outcome. Studies in this forest plot are divided into subgroups based on potential for detection bias to explain heterogeneity. Note that negative signs were added to those scores that were based on a scale where lower scores indicated higher disease severity.

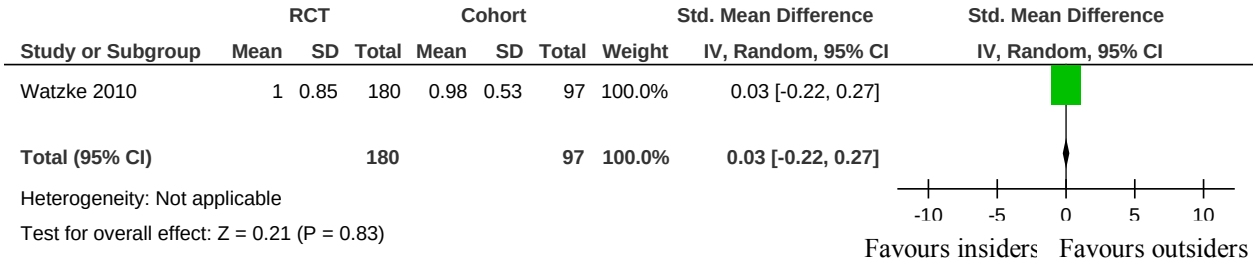


Figure 25: All studies in this meta-analysis randomly assigned participants to either a trial or cohort group and had a continuous mean score as an outcome. Studies in this forest plot are divided into subgroups based on potential for detection bias to explain heterogeneity.

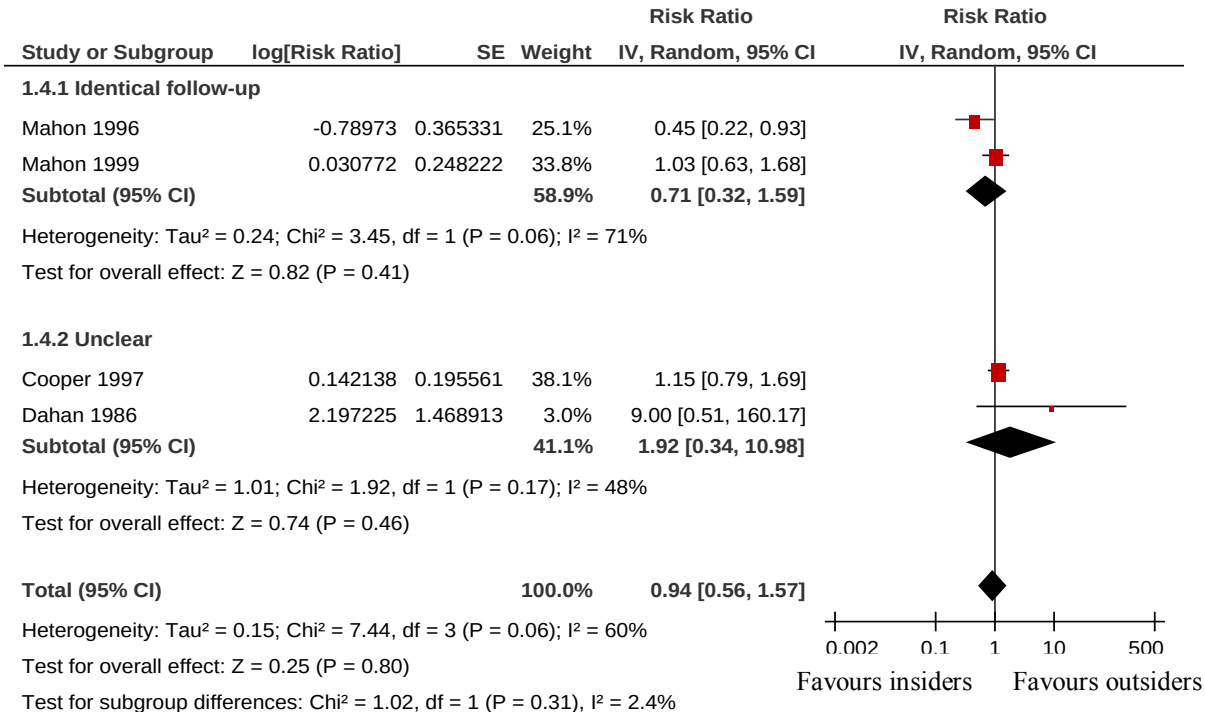


Figure 26: All studies in this meta-analysis randomly assigned participants to either a trial or cohort group and had a dichotomous non-mortality outcome. Studies in this forest plot are divided into subgroups based on potential for detection bias to explain heterogeneity.

1.6 Subgroups based on potential for exclusion bias

	RCT		Cohort			Odds Ratio	Odds Ratio
Study or Subgroup	**Events**	**Total**	**Events**	**Total**	**Weight**	**M-H, Random, 95% CI**	**M-H, Random, 95% CI**
2.3.1 No exclusions							
Marcinczyk 1997	0	54	0	29		Not estimable	
Lidbrink 1995	23	20000	42	7785	5.8%	0.21 [0.13, 0.35]	
van Bergen 1995	42	350	118	587	6.4%	0.54 [0.37, 0.79]	
Clapp 1989	6	115	7	85	3.1%	0.61 [0.20, 1.90]	
Smith 1990	217	1214	69	270	6.7%	0.63 [0.46, 0.86]	
Sesso 2002	165	22071	128	11152	7.0%	0.65 [0.51, 0.82]	
Tanai 2009	71	100	15	19	3.0%	0.65 [0.20, 2.13]	
Rigg 2000	31	455	23	237	5.5%	0.68 [0.39, 1.20]	
Tenenbaum 2002	423	3122	52	380	6.7%	0.99 [0.73, 1.35]	
Wyse 1991	51	1672	8	318	4.6%	1.22 [0.57, 2.59]	
Woodhouse 1995	182	194	133	145	4.3%	1.37 [0.60, 3.14]	
Moertel 1984	53	62	7	10	2.1%	2.52 [0.55, 11.61]	
Detre 1999	54	343	20	299	5.6%	2.61 [1.52, 4.47]	
Urban 1999	40	55	12	24	3.6%	2.67 [0.98, 7.22]	
Subtotal (95% CI)		**49807**		**21340**	**64.3%**	**0.85 [0.61, 1.19]**	
Total events	1358		634				
Heterogeneity: Tau² = 0.26; Chi² = 65.10, df = 12 (P < 0.00001); I² = 82%							
Test for overall effect: Z = 0.96 (P = 0.34)							
2.3.2 Appropriate exclusions							
Helsing 1998	39	47	76	97	4.0%	1.35 [0.55, 3.32]	
Subtotal (95% CI)		**47**		**97**	**4.0%**	**1.35 [0.55, 3.32]**	
Total events	39		76				
Heterogeneity: Not applicable							
Test for overall effect: Z = 0.65 (P = 0.52)							
2.3.3 Inappropriate (unequal)							
Vind 2009	1	256	4	297	1.2%	0.29 [0.03, 2.59]	
Boesen 2007	18	258	18	137	4.9%	0.50 [0.25, 0.99]	
Nicolaides 1994	17	488	37	812	5.4%	0.76 [0.42, 1.36]	
CASS 1984	65	779	104	1309	6.6%	1.05 [0.76, 1.46]	
Feit 2000	202	1169	194	1336	7.0%	1.23 [0.99, 1.52]	
Strandberg 1995	160	910	47	489	6.5%	2.01 [1.42, 2.83]	
Subtotal (95% CI)		**3860**		**4380**	**31.7%**	**1.05 [0.73, 1.50]**	
Total events	463		404				
Heterogeneity: Tau² = 0.12; Chi² = 19.49, df = 5 (P = 0.002); I² = 74%							
Test for overall effect: Z = 0.24 (P = 0.81)							
Total (95% CI)		**53714**		**25817**	**100.0%**	**0.91 [0.70, 1.18]**	
Total events	1860		1114				
Heterogeneity: Tau² = 0.24; Chi² = 109.98, df = 19 (P < 0.00001); I² = 83%							
Test for overall effect: Z = 0.72 (P = 0.47)							
Test for subgroup differences: Chi² = 1.27, df = 2 (P = 0.53), I² = 0%							

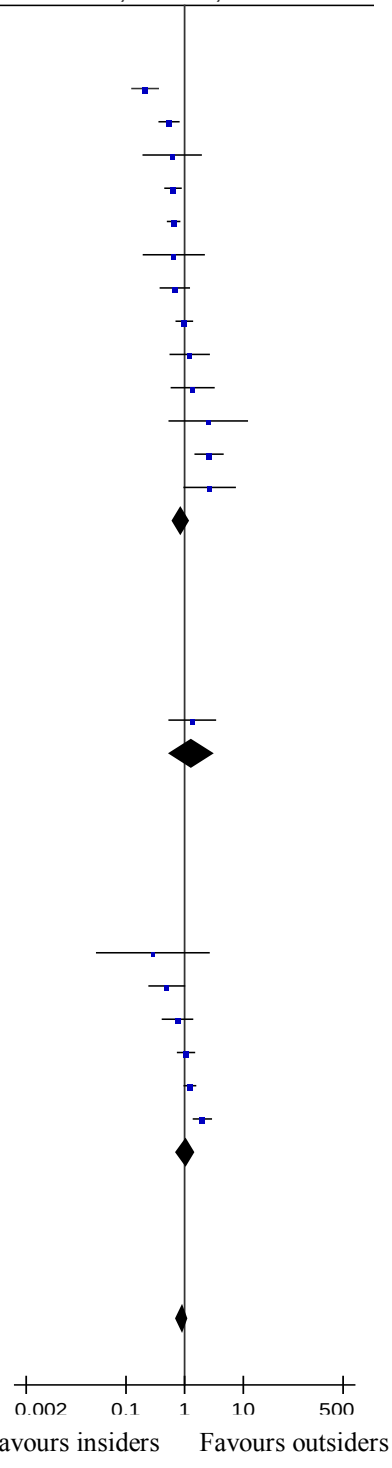


Figure 27: All studies in this meta-analysis non- randomly assigned participants to either a trial or cohort group and had a mortality outcome. Studies in this forest plot are divided into subgroups based on the potential for exclusion bias to explain heterogeneity.

Study or Subgroup	log[Risk Ratio]	SE	RCT Total	Cohort Total	Weight	Risk Ratio IV, Random, 95% CI
2.4.1 No exclusions						
Akaza 1995	2.399577	1.393859	107	13	0.1%	11.02 [0.72, 169.27]
Amar 1997	-0.18809	0.330284	70	40	1.1%	0.83 [0.43, 1.58]
Antman 1985	-0.62571	0.322026	42	24	1.1%	0.53 [0.28, 1.01]
Balmukhanov 1989	-0.54098	0.418447	108	287	0.8%	0.58 [0.26, 1.32]
Caplan 1984	-0.07632	0.057088	29	46	3.9%	0.93 [0.83, 1.04]
Chauhan 1992	-0.5545	0.773029	38	15	0.3%	0.57 [0.13, 2.61]
Chesebro 1983	-0.10215	0.52802	351	183	0.5%	0.90 [0.32, 2.54]
Cowchock 1992	-0.20479	0.533847	20	13	0.5%	0.81 [0.29, 2.32]
Creutzig 1993	0.451606	0.364125	31	25	0.9%	1.57 [0.77, 3.21]
Diehl 1995	-0.61837	0.304074	100	21	1.2%	0.54 [0.30, 0.98]
Edsmyr 1978	-1.15268	0.450471	18	9	0.7%	0.32 [0.13, 0.76]
Forbes 2000	0.093137	0.275068	102	88	1.4%	1.10 [0.64, 1.88]
Gossop 1986	0.409538	0.243326	20	40	1.6%	1.51 [0.93, 2.43]
Henriksson 1986	-0.69315	0.613259	91	9	0.4%	0.50 [0.15, 1.66]
Kane 1988	1.498178	0.379098	59	116	0.9%	4.47 [2.13, 9.40]
Kayser 2008	0.419398	0.158367	31	44	2.5%	1.52 [1.12, 2.07]
Lichtenberg 2008	0.251739	0.082639	217	153	3.6%	1.29 [1.09, 1.51]
Link 1991	-0.28136	0.231968	36	77	1.7%	0.75 [0.48, 1.19]
Liu 1998	-0.54676	1.026979	169	163	0.1%	0.58 [0.08, 4.33]
MacLennan 1985	-0.01933	0.077819	96	73	3.6%	0.98 [0.84, 1.14]
Martin 1994	-0.94143	1.620867	46	54	0.1%	0.39 [0.02, 9.35]
Martinez- Amenos 1990	-0.27906	0.101954	589	133	3.3%	0.76 [0.62, 0.92]
Mayo Group 1992	1.926181	0.394957	71	87	0.8%	6.86 [3.16, 14.88]
Morrison 2002	0.075619	0.08478	454	302	3.5%	1.08 [0.91, 1.27]
Nagel 1998	0.262743	0.669698	115	95	0.3%	1.30 [0.35, 4.83]
Neldam 1986	0.070014	1.631805	978	349	0.1%	1.07 [0.04, 26.27]
Panagopoulou 2009	0.258369	0.104057	148	66	3.3%	1.29 [1.06, 1.59]
Rosen 1987	-0.18334	0.049179	98	44	3.9%	0.83 [0.76, 0.92]
Stern 2003	-0.03348	0.013236	555	1788	4.2%	0.97 [0.94, 0.99]
Sundar 2008	-1.09134	1.992727	136	45	0.0%	0.34 [0.01, 16.68]
Taplin 1986	-0.07555	0.197764	63	30	2.1%	0.93 [0.63, 1.37]
Verdonck 1995	-0.26463	0.277749	69	37	1.4%	0.77 [0.45, 1.32]
Welt 1981	-1.15088	0.638762	23	40	0.4%	0.32 [0.09, 1.11]
West 2005	1.311744	0.886233	86	322	0.2%	3.71 [0.65, 21.09]
Wetzner 1979	1.341174	0.36171	34	64	0.9%	3.82 [1.88, 7.77]
Yamamoto 1992	-0.20098	0.111922	31	92	3.2%	0.82 [0.66, 1.02]
Yamani 2005	-0.11432	0.347337	23	33	1.0%	0.89 [0.45, 1.76]
Subtotal (95% CI)			**5254**	**5020**	**55.5%**	**1.02 [0.92, 1.13]**
Heterogeneity: Tau² = 0.03; Chi² = 136.29, df = 36 (P < 0.00001); I² = 74%						
Test for overall effect: Z = 0.36 (P = 0.72)						
2.4.3 Inappropriate (unequal)						
Bell 2000	-0.05129	0.875094	59	56	0.2%	0.95 [0.17, 5.28]
Biasoli 2008	-1.33123	0.919898	52	41	0.2%	0.26 [0.04, 1.60]
Biederman 1985	-0.17435	0.328435	24	18	1.1%	0.84 [0.44, 1.60]
Blichert- Toft 1988	1.155294	0.321989	619	136	1.1%	3.17 [1.69, 5.97]
Blumenthal 1997	-0.71548	0.366976	66	38	0.9%	0.49 [0.24, 1.00]
Chilvers 2001	-0.04946	0.199803	98	207	2.1%	0.95 [0.64, 1.41]
Clagett 1984	-1.97981	1.489261	29	28	0.1%	0.14 [0.01, 2.56]
Eberhardt 1996	-0.08688	0.103497	43	37	3.3%	0.92 [0.75, 1.12]
Gall 2007	0.458067	0.442399	46	41	0.7%	1.58 [0.66, 3.76]
Howie 1997	0.807696	0.392005	77	63	0.8%	2.24 [1.04, 4.84]
Karande 1999	-0.19608	0.254808	63	57	1.6%	0.82 [0.50, 1.35]
Kendrick 2001	0.36344	0.16896	394	50	2.4%	1.44 [1.03, 2.00]
Kirke 1992	-2.0381	0.969636	351	106	0.2%	0.13 [0.02, 0.87]

Eberhardt 1996	-0.08688	0.103497	43	37	3.3%	0.92 [0.75, 1.12]
Gall 2007	0.458067	0.442399	46	41	0.7%	1.58 [0.66, 3.76]
Howie 1997	0.807696	0.392005	77	63	0.8%	2.24 [1.04, 4.84]
Karande 1999	-0.19608	0.254808	63	57	1.6%	0.82 [0.50, 1.35]
Kendrick 2001	0.36344	0.16896	394	50	2.4%	1.44 [1.03, 2.00]
Kirke 1992	-2.0381	0.969636	351	106	0.2%	0.13 [0.02, 0.87]
MacMillan 1986	-0.08445	0.125016	107	49	3.0%	0.92 [0.72, 1.17]
Melchart 2002	0.824175	0.36862	26	80	0.9%	2.28 [1.11, 4.70]
Ogden 2004	0.526329	0.19616	285	47	2.1%	1.69 [1.15, 2.49]
Peteren 2007	-1.85419	0.590818	79	33	0.4%	0.16 [0.05, 0.50]
Raistrick 2005	-0.04419	0.073987	174	225	3.7%	0.96 [0.83, 1.11]
Rorbye 2005	0.453048	0.221689	105	727	1.8%	1.57 [1.02, 2.43]
Shain 1989	-1.04252	0.328532	155	98	1.1%	0.35 [0.19, 0.67]
Stith 2004	-1.89712	0.903696	19	4	0.2%	0.15 [0.03, 0.88]
Sullivan 1982	0.385497	0.635155	144	25	0.4%	1.47 [0.42, 5.11]
Van 2009	-0.17973	0.231445	40	45	1.7%	0.84 [0.53, 1.32]
Walker 1986	-0.49491	0.394087	98	37	0.8%	0.61 [0.28, 1.32]
Wallage 2003	-0.21065	0.550262	178	28	0.5%	0.81 [0.28, 2.38]
Wieringa- de Waard 2002	0.008533	0.161491	122	305	2.5%	1.01 [0.73, 1.38]
Yersin 1996	-0.21131	0.40117	20	10	0.8%	0.81 [0.37, 1.78]
Subtotal (95% CI)			**3473**	**2591**	**34.5%**	**0.97 [0.82, 1.17]**
Heterogeneity: Tau² = 0.11; Chi² = 81.03, df = 26 (P < 0.00001); I² = 68%						
Test for overall effect: Z = 0.28 (P = 0.78)						
2.4.4 Unclear						
Bhattacharya 1998	0.212333	0.214163	92	68	1.9%	1.24 [0.81, 1.88]
Bijker 2002	-0.36792	0.329306	268	155	1.1%	0.69 [0.36, 1.32]
Clemens 1992	-0.16896	0.084346	20744	21943	3.5%	0.84 [0.72, 1.00]
Goodkin 1987	-0.35256	0.144544	27	24	2.7%	0.70 [0.53, 0.93]
Williford 1993	-0.3317	0.414502	395	199	0.8%	0.72 [0.32, 1.62]
Subtotal (95% CI)			**21526**	**22389**	**10.0%**	**0.83 [0.70, 0.99]**
Heterogeneity: Tau² = 0.01; Chi² = 5.26, df = 4 (P = 0.26); I² = 24%						
Test for overall effect: Z = 2.02 (P = 0.04)						
Total (95% CI)			**30253**	**30000**	**100.0%**	**0.99 [0.92, 1.08]**
Heterogeneity: Tau² = 0.04; Chi² = 228.19, df = 68 (P < 0.00001); I² = 70%						
Test for overall effect: Z = 0.15 (P = 0.88)						
Test for subgroup differences: Chi² = 3.74, df = 2 (P = 0.15), I² = 46.5%						

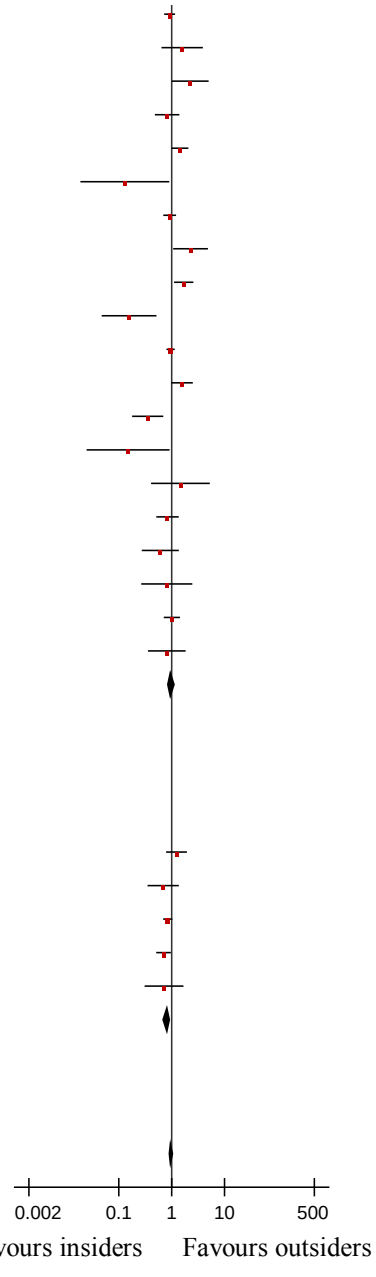
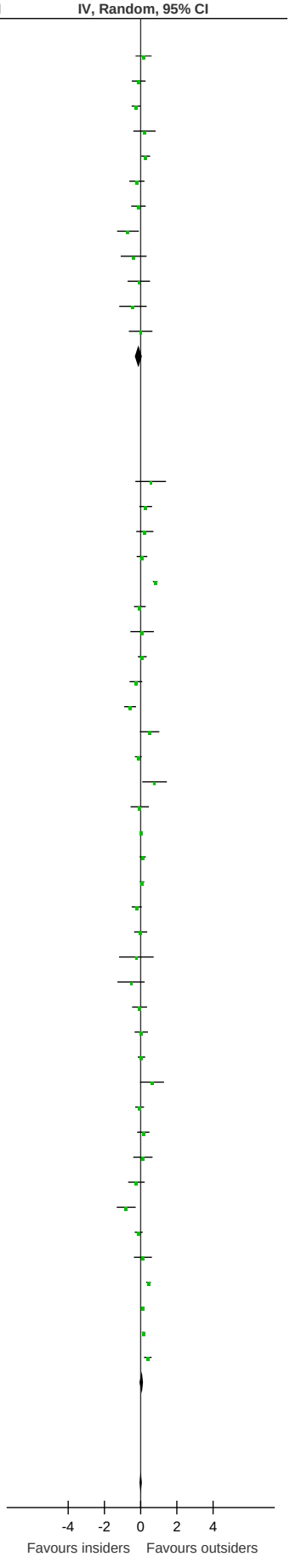


Figure 28: All studies in this meta-analysis non- randomly assigned participants to either a trial or cohort group and had a dichotomous non-mortality outcome. Studies in this forest plot are divided into subgroups based on the potential for exclusion bias to explain heterogeneity.

Study or Subgroup	RCT Mean	RCT SD	RCT Total	Cohort Mean	Cohort SD	Cohort Total	Weight	Std. Mean Difference IV, Random, 95% CI
2.2.1 No exclusions								
Bain 2001	1.251111	1.384769	36	1.030968	1.346802	62	1.8%	0.16 [-0.25, 0.57]
Bannister 2001	21.95297	11.5366	202	23.2	9.5	38	2.1%	-0.11 [-0.46, 0.24]
Boezaart 1998	0.233333	0.512723	240	0.4	0.84	136	2.8%	-0.26 [-0.47, -0.04]
Euler 2005	8.886207	1.264871	58	8.6	1.65	14	1.2%	0.21 [-0.37, 0.80]
Gunn 2000	-4,114.5	671.25	308	-4,303	683	122	2.8%	0.28 [0.07, 0.49]
Heuss 2004	2.389189	2.368924	74	2.9	2.6	40	1.9%	-0.21 [-0.59, 0.18]
Lansky 1983	4.014545	7.513319	55	4.95	7.73	59	2.0%	-0.12 [-0.49, 0.25]
Masood 2002	3.854688	2.359498	96	5.43	0.85	14	1.2%	-0.70 [-1.27, -0.13]
Palmon 1996	38	7.889544	50	41	6.3	10	1.0%	-0.39 [-1.07, 0.30]
Smuts 2003	-2.688	1.2042	37	-2.57	1.04	16	1.2%	-0.10 [-0.69, 0.49]
Tanaka 1994	91.66667	19.33079	30	100	19	10	0.9%	-0.42 [-1.15, 0.30]
Toprak 2005	9.57	19.33079	30	9.58	2.05	15	1.1%	-0.00 [-0.62, 0.62]
Subtotal (95% CI)			**1216**			**536**	**20.0%**	**-0.10 [-0.27, 0.07]**
Heterogeneity: Tau² = 0.04; Chi² = 22.72, df = 11 (P = 0.02); I² = 52%								
Test for overall effect: Z = 1.13 (P = 0.26)								
2.2.2 Inappropriate (unequal)								
Andersson 2003	3.6625	4.100775	24	1.6	1.1	8	0.7%	0.55 [-0.26, 1.37]
Ashok 2002	12.75585	5.556547	229	11.24	4.62817	45	2.2%	0.28 [-0.04, 0.60]
Bakker 2000	2.066903	5.408818	113	0.9	1.4	24	1.7%	0.23 [-0.21, 0.68]
Bedi 2000	14.98824	10.76794	85	14.26341	9.605447	164	2.5%	0.07 [-0.19, 0.33]
Brinkhaus 2008	1.431185	0.233	540	1.29	0.16	2469	3.3%	0.81 [0.71, 0.90]
Dalal 2007	-5.628	1.1138	84	-5.5785	1.117	100	2.4%	-0.04 [-0.33, 0.25]
Decensi 2003	12.049	5.618	115	11.6	2.3	11	1.1%	0.08 [-0.54, 0.70]
Ekstein 2002	-1.25	2.9294	1202	-1.5	0.647	91	2.8%	0.09 [-0.12, 0.30]
Emery 2003	34.07976	9.408757	168	36.70408	11.73033	49	2.2%	-0.26 [-0.58, 0.06]
Franz 1995	7.494972	1.502274	179	8.4	1.7	62	2.3%	-0.58 [-0.87, -0.29]
Giron 2010	0.68	0.99	24	0.18	1	45	1.5%	0.50 [-0.01, 1.00]
Grant 2008	78.83144	21.56381	299	81.51787	20.13856	375	3.1%	-0.13 [-0.28, 0.02]
Hoh 1998	1.128974	0.306798	39	0.9	0.25	13	1.1%	0.77 [0.12, 1.41]
Howard 2010	-51	16.693	28	-50.318	11.23	44	1.6%	-0.05 [-0.52, 0.42]
Jena 2008	4.885494	5.6436	2792	4.83	5.83	10410	3.4%	0.01 [-0.03, 0.05]
Jensen 2003	2.255351	4.754075	897	1.703844	4.501695	294	3.2%	0.12 [-0.01, 0.25]
Kieler 1998	-3,499.9	558.5	4801	-3,542	553	526	3.3%	0.08 [-0.01, 0.17]
King 2000	11.76315	10.21488	165	13.98491	10.17274	106	2.6%	-0.22 [-0.46, 0.03]
Lock 2010	0.22697	2.253443	40	0.224488	0.315691	303	2.2%	0.00 [-0.33, 0.33]
Macdonald 2007	-20	5.8951	48	-18.6	4.9	5	0.6%	-0.24 [-1.16, 0.69]
McCaughey 1998	-151.84	6.2001	13	-149.3	3.3	19	0.9%	-0.53 [-1.25, 0.19]
McKay 1995	4.775	7.538208	40	5.237375	8.085113	80	1.9%	-0.06 [-0.44, 0.32]
McKay 1998	1.762574	4.08093	101	1.627647	3.577174	51	2.1%	0.03 [-0.30, 0.37]
Mori 2006	-3.487	1.278	158	-3.544	1.243	712	3.0%	0.05 [-0.13, 0.22]
Reddihough 1998	-30.505	16.115	22	-42.379	21	19	1.1%	0.63 [-0.00, 1.26]
Salisbury 2002	6.147036	0.889639	253	6.2	0.96	129	2.8%	-0.06 [-0.27, 0.15]
Stecksen-blicks 2008	5.15913	5.423801	115	4.4	4.6	64	2.3%	0.15 [-0.16, 0.45]
Stockton 2009	19.33684	7.454532	57	18.4	7.6	21	1.5%	0.12 [-0.38, 0.62]
Suherman 1999	111.6585542	13.76	83	115.02	16	29	1.7%	-0.23 [-0.66, 0.19]
Taddio 2006	0.293878	0.408428	98	0.62	0.4	20	1.5%	-0.80 [-1.29, -0.30]
Underwood 2008	38.49198	21.99227	187	41	24.18853	271	2.9%	-0.11 [-0.29, 0.08]
WHO 1988	-64.9	27.363	40	-67.65	11.85	32	1.6%	0.12 [-0.34, 0.59]
Witt 2006a	38.82265	18.50601	543	30.3	19.9	2481	3.3%	0.43 [0.34, 0.53]
Witt 2006b	39.98814	5.7621	3036	38.9	11.3	4686	3.4%	0.11 [0.07, 0.16]
Witt 2006c	17.14261	5.7535	2518	16.4	4.2	3901	3.4%	0.15 [0.10, 0.20]
Witt 2008	4.243784	2.112303	185	3.3	2.5	389	3.0%	0.40 [0.22, 0.57]
Subtotal (95% CI)			**19321**			**28048**	**80.0%**	**0.08 [-0.01, 0.16]**
Heterogeneity: Tau² = 0.05; Chi² = 368.28, df = 35 (P < 0.00001); I² = 90%								
Test for overall effect: Z = 1.66 (P = 0.10)								
Total (95% CI)			**20537**			**28584**	**100.0%**	**0.04 [-0.04, 0.12]**
Heterogeneity: Tau² = 0.05; Chi² = 402.15, df = 47 (P < 0.00001); I² = 88%								
Test for overall effect: Z = 0.99 (P = 0.32)								
Test for subgroup differences: Chi² = 3.12, df = 1 (P = 0.08), I² = 68.0%								



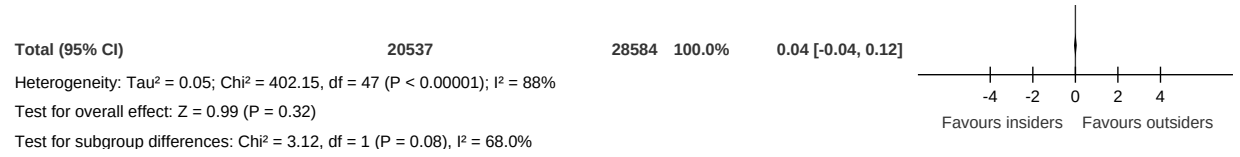


Figure 29: All studies in this meta-analysis non-randomly assigned participants to either a trial or cohort group and had a continuous endpoint as an outcome. Studies in this forest plot are divided into subgroups based on the potential for exclusion bias to explain heterogeneity. Note that negative signs were added to those scores that were based on a scale where lower scores indicated higher disease severity.

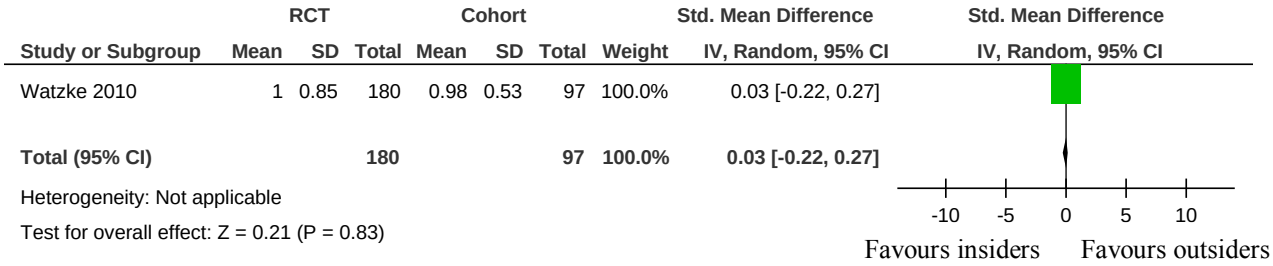


Figure 30: All studies in this meta-analysis randomly assigned participants to either a trial or cohort group and had a continuous mean score as an outcome. Studies in this forest plot are divided into subgroups based on the potential for exclusion bias to explain heterogeneity.

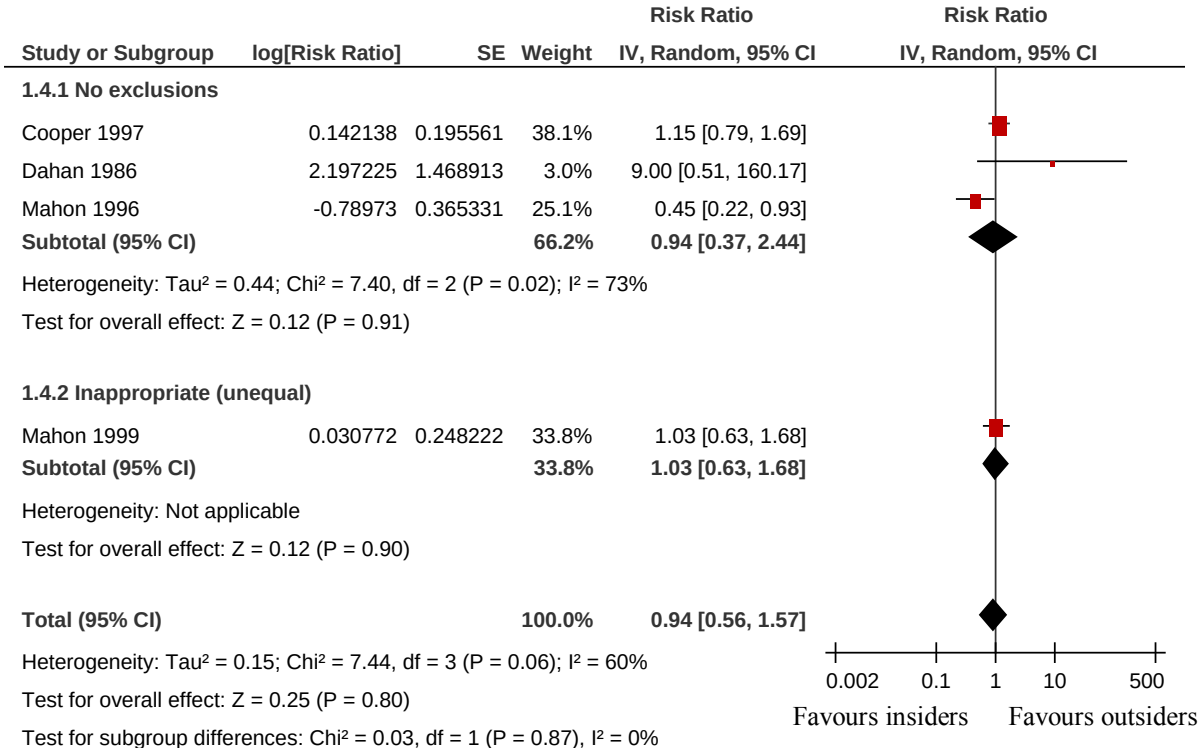
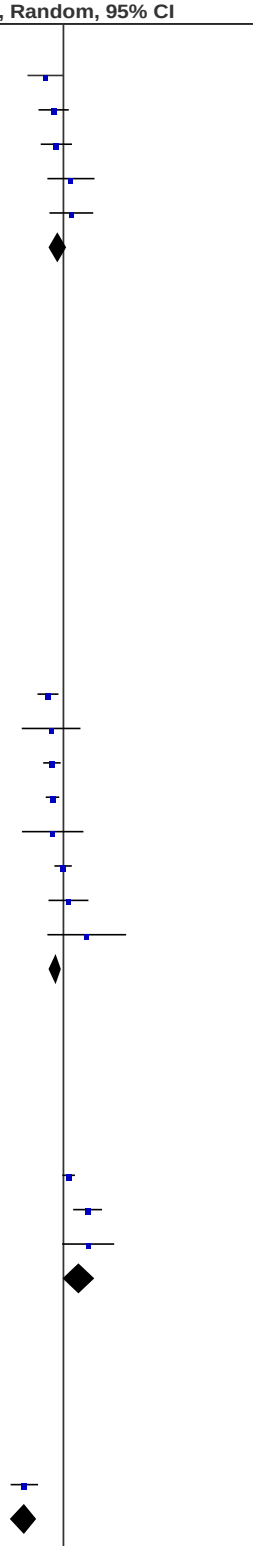


Figure 31: All studies in this meta-analysis randomly assigned participants to either a trial or cohort group and had a dichotomous non-mortality outcome. Studies in this forest plot are divided into subgroups based on the potential for exclusion bias to explain heterogeneity.

1.7 Subgroups based on type of care received

	RCT		Cohort			Odds Ratio	Odds Ratio
Study or Subgroup	**Events**	**Total**	**Events**	**Total**	**Weight**	**M-H, Random, 95% CI**	**M-H, Random, 95% CI**
2.3.1 Same individual, same expertise							
Boesen 2007	18	258	18	137	4.9%	0.50 [0.25, 0.99]	
RIgg 2000	31	455	23	237	5.5%	0.68 [0.39, 1.20]	
Nicolaides 1994	17	488	37	812	5.4%	0.76 [0.42, 1.36]	
Helsing 1998	39	47	76	97	4.0%	1.35 [0.55, 3.32]	
Woodhouse 1995	182	194	133	145	4.3%	1.37 [0.60, 3.14]	
Subtotal (95% CI)		**1442**		**1428**	**24.0%**	**0.79 [0.56, 1.12]**	
Total events	287		287				
Heterogeneity: Tau² = 0.03; Chi² = 5.07, df = 4 (P = 0.28); I² = 21%							
Test for overall effect: Z = 1.31 (P = 0.19)							
2.3.2 Different individual, same expertise							
Marcinczyk 1997	0	54	0	29		Not estimable	
Subtotal (95% CI)		**54**		**29**		**Not estimable**	
Total events	0		0				
Heterogeneity: Not applicable							
Test for overall effect: Not applicable							
2.3.3 Unclear- non-surgical							
van Bergen 1995	42	350	118	587	6.4%	0.54 [0.37, 0.79]	
Clapp 1989	6	115	7	85	3.1%	0.61 [0.20, 1.90]	
Smith 1990	217	1214	69	270	6.7%	0.63 [0.46, 0.86]	
Sesso 2002	165	22071	128	11152	7.0%	0.65 [0.51, 0.82]	
Tanai 2009	71	100	15	19	3.0%	0.65 [0.20, 2.13]	
Tenenbaum 2002	423	3122	52	380	6.7%	0.99 [0.73, 1.35]	
Wyse 1991	51	1672	8	318	4.6%	1.22 [0.57, 2.59]	
Moertel 1984	53	62	7	10	2.1%	2.52 [0.55, 11.61]	
Subtotal (95% CI)		**28706**		**12821**	**39.6%**	**0.73 [0.59, 0.91]**	
Total events	1028		404				
Heterogeneity: Tau² = 0.04; Chi² = 12.11, df = 7 (P = 0.10); I² = 42%							
Test for overall effect: Z = 2.78 (P = 0.005)							
2.3.4 Unclear- surgical							
Feit 2000	202	1169	194	1336	7.0%	1.23 [0.99, 1.52]	
Detre 1999	54	343	20	299	5.6%	2.61 [1.52, 4.47]	
Urban 1999	40	55	12	24	3.6%	2.67 [0.98, 7.22]	
Subtotal (95% CI)		**1567**		**1659**	**16.3%**	**1.88 [1.02, 3.46]**	
Total events	296		226				
Heterogeneity: Tau² = 0.21; Chi² = 8.13, df = 2 (P = 0.02); I² = 75%							
Test for overall effect: Z = 2.01 (P = 0.04)							
2.3.6 Unclear- radiology							
Lidbrink 1995	23	20000	42	7785	5.8%	0.21 [0.13, 0.35]	
Subtotal (95% CI)		**20000**		**7785**	**5.8%**	**0.21 [0.13, 0.35]**	



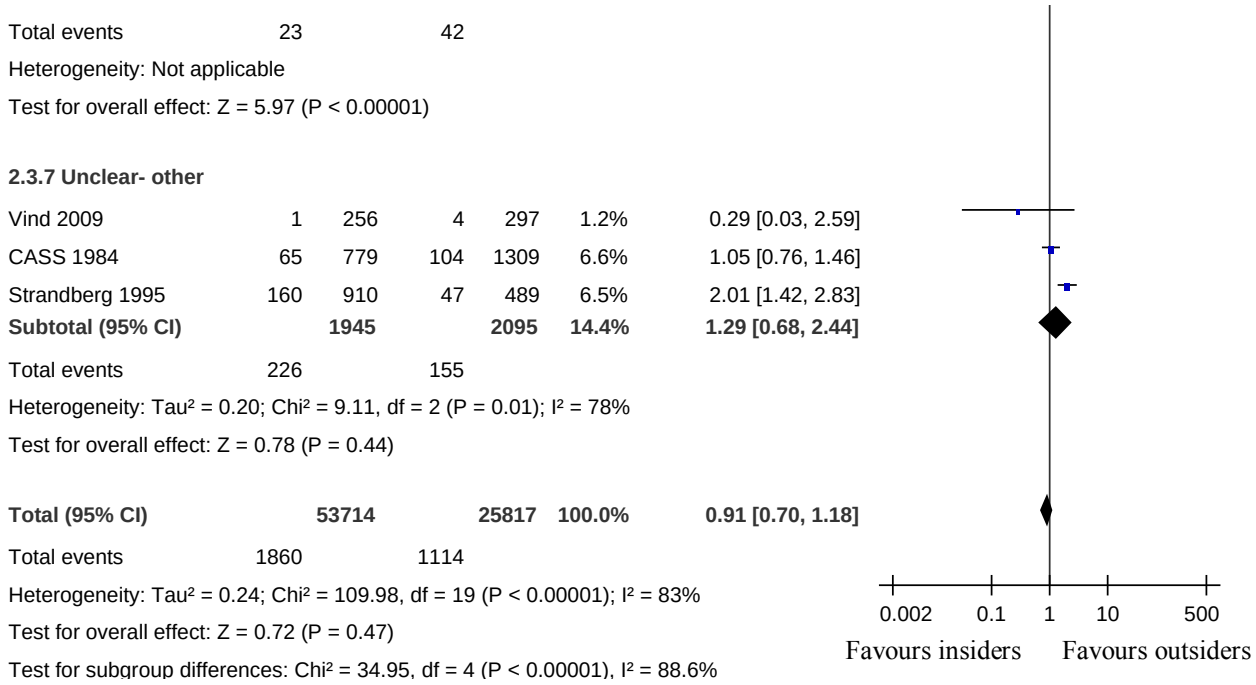
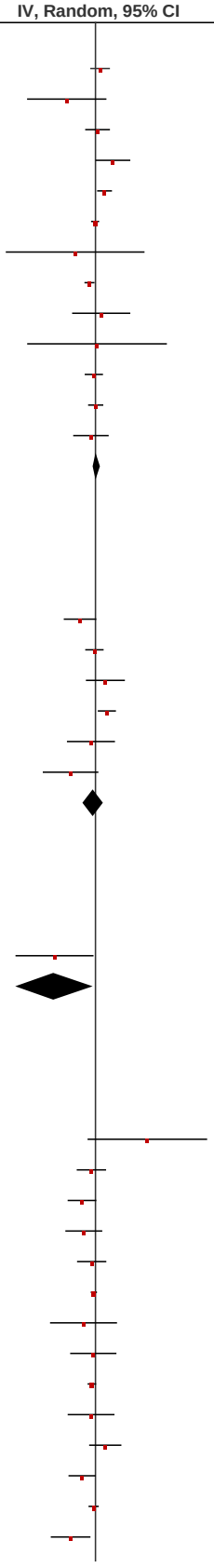
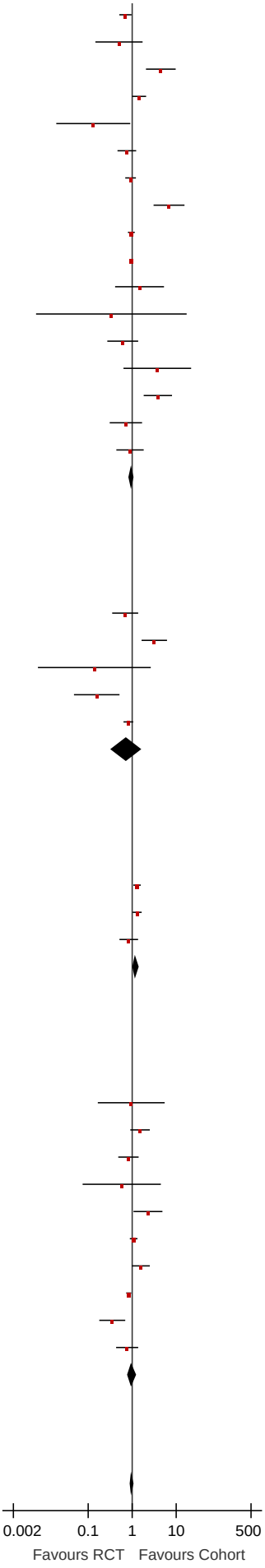


Figure 32: All studies in this meta-analysis non- randomly assigned participants to either a trial or cohort group and had mortality as an outcome. Studies in this forest plot are divided into subgroups based on the type of care provided to explain heterogeneity.

Study or Subgroup	log[Risk Ratio]	SE	RCT Total	Cohort Total	Weight	Risk Ratio IV, Random, 95% CI
2.4.1 Same individual, same expertise						
Bhattacharya 1998	0.212333	0.214163	92	68	1.9%	1.24 [0.81, 1.88]
Biasoli 2008	-1.33123	0.919898	52	41	0.2%	0.26 [0.04, 1.60]
Forbes 2000	0.093137	0.275068	102	88	1.4%	1.10 [0.64, 1.88]
Howie 1997	0.807696	0.392005	77	63	0.8%	2.24 [1.04, 4.84]
Kayser 2008	0.419398	0.158367	31	44	2.5%	1.52 [1.12, 2.07]
MacLennan 1985	-0.01933	0.077819	96	73	3.6%	0.98 [0.84, 1.14]
Martin 1994	-0.94143	1.620867	46	54	0.1%	0.39 [0.02, 9.35]
Martinez- Amenos 1990	-0.27906	0.101954	589	133	3.3%	0.76 [0.62, 0.92]
Nagel 1998	0.262743	0.669698	115	95	0.3%	1.30 [0.35, 4.83]
Neldam 1986	0.070014	1.631805	978	349	0.1%	1.07 [0.04, 26.27]
Taplin 1986	-0.07555	0.197764	63	30	2.1%	0.93 [0.63, 1.37]
Wieringa- de Waard 2002	0.008533	0.161491	122	305	2.5%	1.01 [0.73, 1.38]
Yersin 1996	-0.21131	0.40117	20	10	0.8%	0.81 [0.37, 1.78]
Subtotal (95% CI)			**2383**	**1353**	**19.6%**	**1.04 [0.87, 1.23]**
Heterogeneity: Tau² = 0.03; Chi² = 22.82, df = 12 (P = 0.03); I² = 47%						
Test for overall effect: Z = 0.43 (P = 0.67)						
2.4.2 Different individual, same expertise						
Blumenthal 1997	-0.71548	0.366976	66	38	0.9%	0.49 [0.24, 1.00]
Chilvers 2001	-0.04946	0.199803	98	207	2.1%	0.95 [0.64, 1.41]
Gall 2007	0.458067	0.442399	46	41	0.7%	1.58 [0.66, 3.76]
Ogden 2004	0.526329	0.19616	285	47	2.1%	1.69 [1.15, 2.49]
Wallage 2003	-0.21065	0.550262	178	28	0.5%	0.81 [0.28, 2.38]
Welt 1981	-1.15088	0.638762	23	40	0.4%	0.32 [0.09, 1.11]
Subtotal (95% CI)			**696**	**401**	**6.6%**	**0.93 [0.58, 1.49]**
Heterogeneity: Tau² = 0.20; Chi² = 15.02, df = 5 (P = 0.01); I² = 67%						
Test for overall effect: Z = 0.31 (P = 0.76)						
2.4.3 Different individual, lesser expertise						
Stith 2004	-1.89712	0.903696	19	4	0.2%	0.15 [0.03, 0.88]
Subtotal (95% CI)			**19**	**4**	**0.2%**	**0.15 [0.03, 0.88]**
Heterogeneity: Not applicable						
Test for overall effect: Z = 2.10 (P = 0.04)						
2.4.4 Unclear- non- surgical						
Akaza 1995	2.399577	1.393859	107	13	0.1%	11.02 [0.72, 169.27]
Amar 1997	-0.18809	0.330284	70	40	1.1%	0.83 [0.43, 1.58]
Antman 1985	-0.62571	0.322026	42	24	1.1%	0.53 [0.28, 1.01]
Balmukhanov 1989	-0.54098	0.418447	108	287	0.8%	0.58 [0.26, 1.32]
Biederman 1985	-0.17435	0.328435	24	18	1.1%	0.84 [0.44, 1.60]
Caplan 1984	-0.07632	0.057088	29	46	3.9%	0.93 [0.83, 1.04]
Chauhan 1992	-0.5545	0.773029	38	15	0.3%	0.57 [0.13, 2.61]
Chesebro 1983	-0.10215	0.52802	351	183	0.5%	0.90 [0.32, 2.54]
Clemens 1992	-0.16896	0.084346	20744	21943	3.5%	0.84 [0.72, 1.00]
Cowchock 1992	-0.20479	0.533847	20	13	0.5%	0.81 [0.29, 2.32]
Creutzig 1993	0.451606	0.364125	31	25	0.9%	1.57 [0.77, 3.21]
Diehl 1995	-0.61837	0.304074	100	21	1.2%	0.54 [0.30, 0.98]
Eberhardt 1996	-0.08688	0.103497	43	37	3.3%	0.92 [0.75, 1.12]
Edsmyr 1978	-1.15268	0.450471	18	9	0.7%	0.32 [0.13, 0.76]



Goodkin 1987	-0.35256	0.144544	27	24	2.7%	0.70 [0.53, 0.93]
Henriksson 1986	-0.69315	0.613259	91	9	0.4%	0.50 [0.15, 1.66]
Kane 1988	1.498178	0.379098	59	116	0.9%	4.47 [2.13, 9.40]
Kendrick 2001	0.36344	0.16896	394	50	2.4%	1.44 [1.03, 2.00]
Kirke 1992	-2.0381	0.969636	351	106	0.2%	0.13 [0.02, 0.87]
Link 1991	-0.28136	0.231968	36	77	1.7%	0.75 [0.48, 1.19]
MacMillan 1986	-0.08445	0.125016	107	49	3.0%	0.92 [0.72, 1.17]
Mayo Group 1992	1.926181	0.394957	71	87	0.8%	6.86 [3.16, 14.88]
Raistrick 2005	-0.04419	0.073987	174	225	3.7%	0.96 [0.83, 1.11]
Stern 2003	-0.03348	0.013236	555	1788	4.2%	0.97 [0.94, 0.99]
Sullivan 1982	0.385497	0.635155	144	25	0.4%	1.47 [0.42, 5.11]
Sundar 2008	-1.09134	1.992727	136	45	0.0%	0.34 [0.01, 16.68]
Walker 1986	-0.49491	0.394087	98	37	0.8%	0.61 [0.28, 1.32]
West 2005	1.311744	0.886233	86	322	0.2%	3.71 [0.65, 21.09]
Wetzner 1979	1.341174	0.36171	34	64	0.9%	3.82 [1.88, 7.77]
Williford 1993	-0.3317	0.414502	395	199	0.8%	0.72 [0.32, 1.62]
Yamani 2005	-0.11432	0.347337	23	33	1.0%	0.89 [0.45, 1.76]
Subtotal (95% CI)			**24506**	**25930**	**43.0%**	**0.96 [0.85, 1.08]**
Heterogeneity: Tau² = 0.04; Chi² = 101.18, df = 30 (P < 0.00001); I² = 70%						
Test for overall effect: Z = 0.70 (P = 0.48)						
2.4.5 Unclear- surgical						
Bijker 2002	-0.36792	0.329306	268	155	1.1%	0.69 [0.36, 1.32]
Blichert- Toft 1988	1.155294	0.321989	619	136	1.1%	3.17 [1.69, 5.97]
Clagett 1984	-1.97981	1.489261	29	28	0.1%	0.14 [0.01, 2.56]
Peteren 2007	-1.85419	0.590818	79	33	0.4%	0.16 [0.05, 0.50]
Yamamoto 1992	-0.20098	0.111922	31	92	3.2%	0.82 [0.66, 1.02]
Subtotal (95% CI)			**1026**	**444**	**5.9%**	**0.73 [0.32, 1.64]**
Heterogeneity: Tau² = 0.61; Chi² = 26.90, df = 4 (P < 0.0001); I² = 85%						
Test for overall effect: Z = 0.77 (P = 0.44)						
2.4.7 Unclear- counseling						
Lichtenberg 2008	0.251739	0.082639	217	153	3.6%	1.29 [1.09, 1.51]
Panagopoulou 2009	0.258369	0.104057	148	66	3.3%	1.29 [1.06, 1.59]
Van 2009	-0.17973	0.231445	40	45	1.7%	0.84 [0.53, 1.32]
Subtotal (95% CI)			**405**	**264**	**8.6%**	**1.22 [1.03, 1.45]**
Heterogeneity: Tau² = 0.01; Chi² = 3.26, df = 2 (P = 0.20); I² = 39%						
Test for overall effect: Z = 2.29 (P = 0.02)						
2.4.8 Unclear- other						
Bell 2000	-0.05129	0.875094	59	56	0.2%	0.95 [0.17, 5.28]
Gossop 1986	0.409538	0.243326	20	40	1.6%	1.51 [0.93, 2.43]
Karande 1999	-0.19608	0.254808	63	57	1.6%	0.82 [0.50, 1.35]
Liu 1998	-0.54676	1.026979	169	163	0.1%	0.58 [0.08, 4.33]
Melchart 2002	0.824175	0.36862	26	80	0.9%	2.28 [1.11, 4.70]
Morrison 2002	0.075619	0.08478	454	302	3.5%	1.08 [0.91, 1.27]
Rorbye 2005	0.453048	0.221689	105	727	1.8%	1.57 [1.02, 2.43]
Rosen 1987	-0.18334	0.049179	98	44	3.9%	0.83 [0.76, 0.92]
Shain 1989	-1.04252	0.328532	155	98	1.1%	0.35 [0.19, 0.67]
Verdonck 1995	-0.26463	0.277749	69	37	1.4%	0.77 [0.45, 1.32]
Subtotal (95% CI)			**1218**	**1604**	**16.3%**	**1.00 [0.79, 1.26]**
Heterogeneity: Tau² = 0.07; Chi² = 32.95, df = 9 (P = 0.0001); I² = 73%						
Test for overall effect: Z = 0.03 (P = 0.98)						
Total (95% CI)			**30253**	**30000**	**100.0%**	**0.99 [0.92, 1.08]**
Heterogeneity: Tau² = 0.04; Chi² = 228.19, df = 68 (P < 0.00001); I² = 70%						
Test for overall effect: Z = 0.15 (P = 0.88)						
Test for subgroup differences: Chi² = 10.69, df = 6 (P = 0.10), I² = 43.9%						



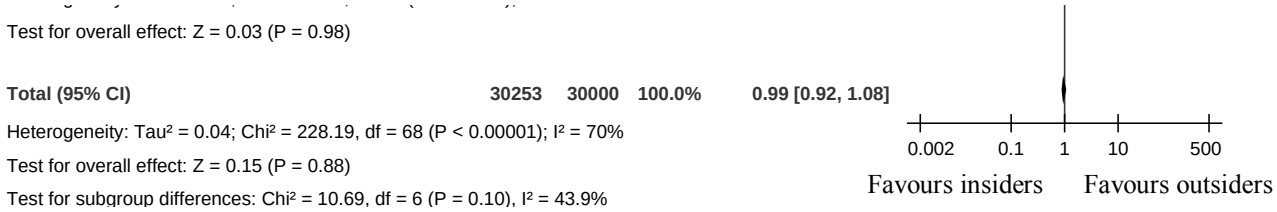
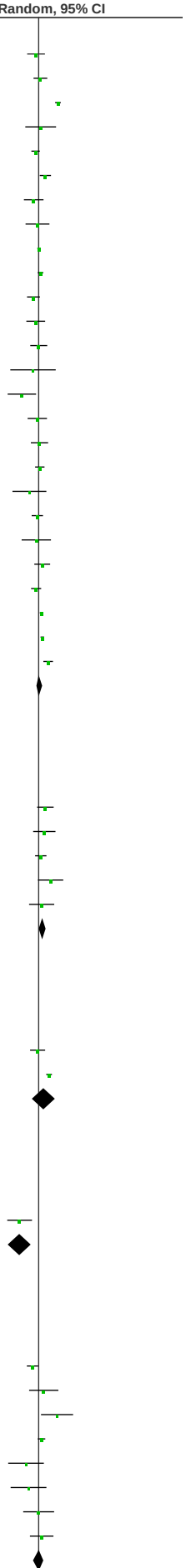


Figure 33: All studies in this meta-analysis non- randomly assigned participants to either a trial or cohort group and had a dichotomous non-mortality outcome. Studies in this forest plot are divided into subgroups based on the type of care provided to explain heterogeneity.

Study or Subgroup	RCT Mean	RCT SD	RCT Total	Cohort Mean	Cohort SD	Cohort Total	Weight	Std. Mean Difference IV, Random, 95% CI
2.2.1 Same individual, same expertise, same setting								
Bannister 2001	21.95297	11.5366	202	23.2	9.5	38	2.1%	-0.11 [-0.46, 0.24]
Bedi 2000	14.98824	10.76794	85	14.26341	9.605447	164	2.5%	0.07 [-0.19, 0.33]
Brinkhaus 2008	1.431185	0.233	540	1.29	0.16	2469	3.3%	0.81 [0.71, 0.90]
Decensi 2003	12.049	5.618	115	11.6	2.3	11	1.1%	0.08 [-0.54, 0.70]
Grant 2008	78.83144	21.56381	299	81.51787	20.13856	375	3.1%	-0.13 [-0.28, 0.02]
Gunn 2000	-4,114.5	671.25	308	-4,303	683	122	2.8%	0.28 [0.07, 0.49]
Heuss 2004	2.389189	2.368924	74	2.9	2.6	40	1.9%	-0.21 [-0.59, 0.18]
Howard 2010	-51	16.693	28	-50.318	11.23	44	1.6%	-0.05 [-0.52, 0.42]
Jena 2008	4.885494	5.6436	2792	4.83	5.83	10410	3.4%	0.01 [-0.03, 0.05]
Kieler 1998	-3,499.9	558.5	4801	-3,542	553	526	3.3%	0.08 [-0.01, 0.17]
King 2000	11.76315	10.21488	165	13.98491	10.17274	106	2.6%	-0.22 [-0.46, 0.03]
Lansky 1983	4.014545	7.513319	55	4.95	7.73	59	2.0%	-0.12 [-0.49, 0.25]
Lock 2010	0.22697	2.253443	40	0.224488	0.315691	303	2.2%	0.00 [-0.33, 0.33]
Macdonald 2007	-20	5.8951	48	-18.6	4.9	5	0.6%	-0.24 [-1.16, 0.69]
Masood 2002	3.854688	2.359498	96	5.43	0.85	14	1.2%	-0.70 [-1.27, -0.13]
McKay 1995	4.775	7.538208	40	5.237375	8.085113	80	1.9%	-0.06 [-0.44, 0.32]
McKay 1998	1.762574	4.08093	101	1.627647	3.577174	51	2.1%	0.03 [-0.30, 0.37]
Mori 2006	-3.487	1.278	158	-3.544	1.243	712	3.0%	0.05 [-0.13, 0.22]
Palmon 1996	38	7.889544	50	41	6.3	10	1.0%	-0.39 [-1.07, 0.30]
Salisbury 2002	6.147036	0.889639	253	6.2	0.96	129	2.8%	-0.06 [-0.27, 0.15]
Smuts 2003	-2.688	1.2042	37	-2.57	1.04	16	1.2%	-0.10 [-0.69, 0.49]
Stecksen-blicks 2008	5.15913	5.423801	115	4.4	4.6	64	2.3%	0.15 [-0.16, 0.45]
Underwood 2008	38.49198	21.99227	187	41	24.18853	271	2.9%	-0.11 [-0.29, 0.08]
Witt 2006b	39.98814	5.7621	3036	38.9	11.3	4686	3.4%	0.11 [0.07, 0.16]
Witt 2006c	17.14261	5.7535	2518	16.4	4.2	3901	3.4%	0.15 [0.10, 0.20]
Witt 2008	4.243784	2.112303	185	3.3	2.5	389	3.0%	0.40 [0.22, 0.57]
Subtotal (95% CI)			**16328**			**24995**	**60.6%**	**0.05 [-0.05, 0.14]**
Heterogeneity: Tau² = 0.04; Chi² = 285.34, df = 25 (P < 0.00001); I² = 91%								
Test for overall effect: Z = 0.92 (P = 0.36)								
2.2.2 Different individual, same expertise, same setting								
Ashok 2002	12.75585	5.556547	229	11.24	4.62817	45	2.2%	0.28 [-0.04, 0.60]
Bakker 2000	2.066903	5.408818	113	0.9	1.4	24	1.7%	0.23 [-0.21, 0.68]
Ekstein 2002	-1.25	2.9294	1202	-1.5	0.647	91	2.8%	0.09 [-0.12, 0.30]
Giron 2010	0.68	0.99	24	0.18	1	45	1.5%	0.50 [-0.01, 1.00]
Stockton 2009	19.33684	7.454532	57	18.4	7.6	21	1.5%	0.12 [-0.38, 0.62]
Subtotal (95% CI)			**1625**			**226**	**9.6%**	**0.19 [0.04, 0.33]**
Heterogeneity: Tau² = 0.00; Chi² = 2.70, df = 4 (P = 0.61); I² = 0%								
Test for overall effect: Z = 2.43 (P = 0.01)								
2.2.3 Different individual, same expertise, unclear setting								
Dalal 2007	-5.628	1.1138	84	-5.5785	1.117	100	2.4%	-0.04 [-0.33, 0.25]
Witt 2006a	38.82265	18.50601	543	30.3	19.9	2481	3.3%	0.43 [0.34, 0.53]
Subtotal (95% CI)			**627**			**2581**	**5.7%**	**0.22 [-0.25, 0.68]**
Heterogeneity: Tau² = 0.10; Chi² = 9.43, df = 1 (P = 0.002); I² = 89%								
Test for overall effect: Z = 0.90 (P = 0.37)								
2.2.4 Different individual, lesser expertise								
Taddio 2006	0.293878	0.408428	98	0.62	0.4	20	1.5%	-0.80 [-1.29, -0.30]
Subtotal (95% CI)			**98**			**20**	**1.5%**	**-0.80 [-1.29, -0.30]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 3.17 (P = 0.002)								
2.2.5 Unclear- non-surgical								
Boezaart 1998	0.233333	0.512723	240	0.4	0.84	136	2.8%	-0.26 [-0.47, -0.04]
Euler 2005	8.886207	1.264871	58	8.6	1.65	14	1.2%	0.21 [-0.37, 0.80]
Hoh 1998	1.128974	0.306798	39	0.9	0.25	13	1.1%	0.77 [0.12, 1.41]
Jensen 2003	2.255351	4.754075	897	1.703844	4.501695	294	3.2%	0.12 [-0.01, 0.25]
McCaughey 1998	-151.84	6.2001	13	-149.3	3.3	19	0.9%	-0.53 [-1.25, 0.19]
Tanaka 1994	91.66667	19.33079	30	100	19	10	0.9%	-0.42 [-1.15, 0.30]
Toprak 2005	9.57	19.33079	30	9.58	2.05	15	1.1%	-0.00 [-0.62, 0.62]
WHO 1988	-64.9	27.363	40	-67.65	11.85	32	1.6%	0.12 [-0.34, 0.59]
Subtotal (95% CI)			**1347**			**533**	**12.7%**	**0.00 [-0.22, 0.23]**



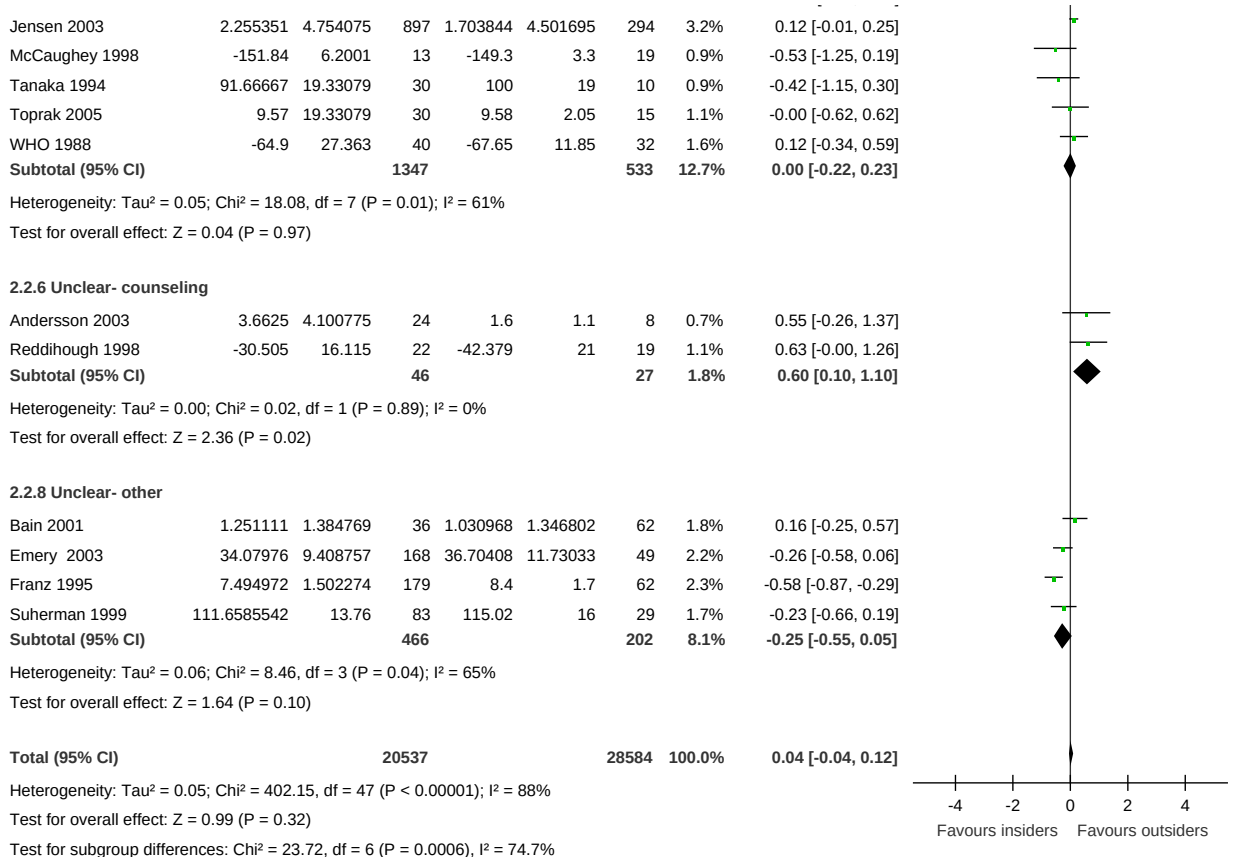


Figure 34: All studies in this meta-analysis non- randomly assigned participants to either a trial or cohort group and had a continuous score as an outcome. Studies in this forest plot are divided into subgroups based on the type of care provided to explain heterogeneity. Note that negative signs were added to those scores that were based on a scale where lower scores indicated higher disease severity.

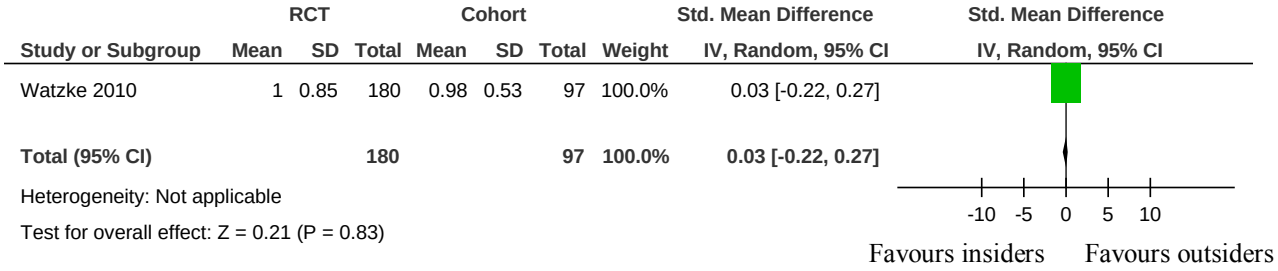


Figure 35: All studies in this meta-analysis randomly assigned participants to either a trial or cohort group and had a continuous endpoint as an outcome. Studies in this forest plot are divided into subgroups based on the type of care provided to explain heterogeneity.

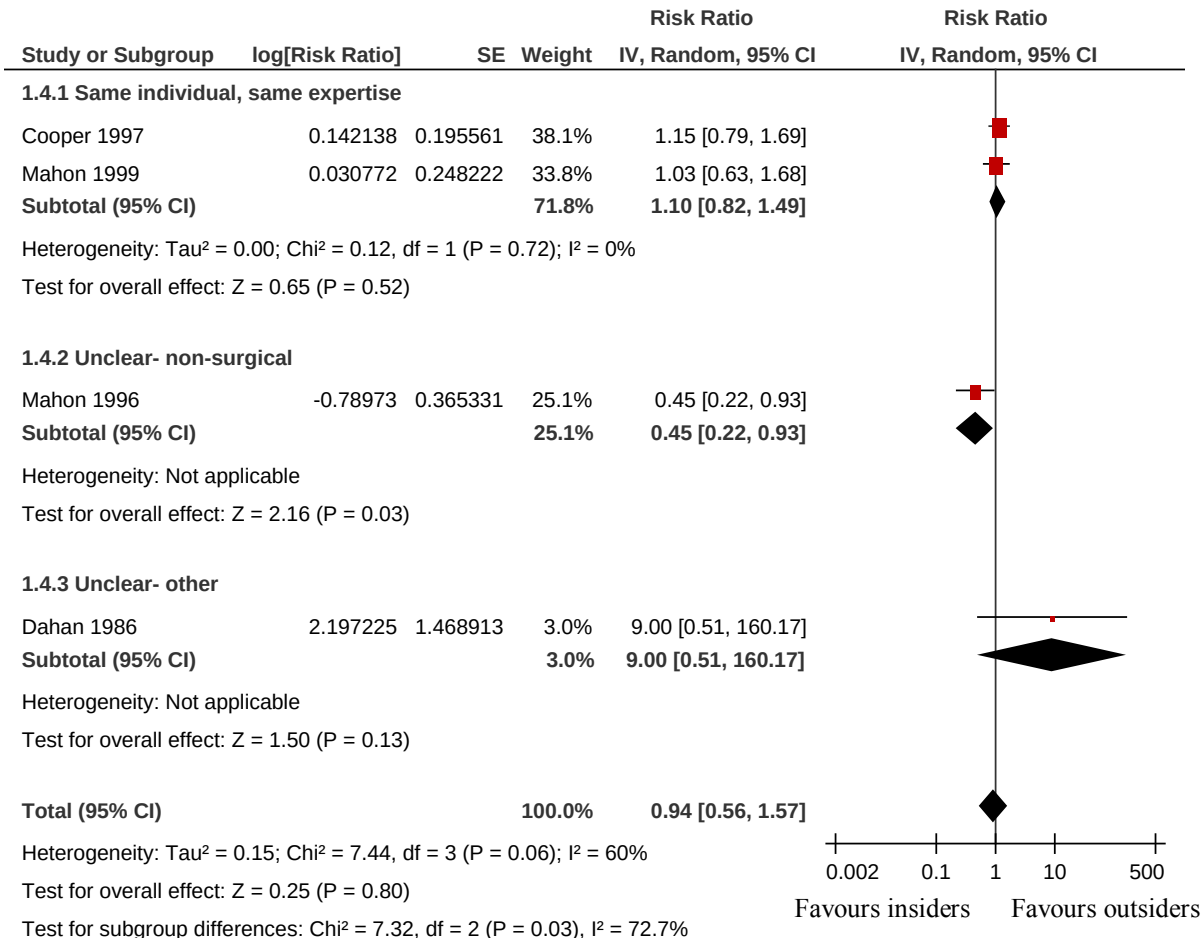


Figure 36: All studies in this meta-analysis randomly assigned participants to either a trial or cohort group and had a dichotomous non-mortality outcome. Studies in this forest plot are divided into subgroups based on the type of care provided to explain heterogeneity.

1.8 Multiple event outcomes (i.e. relapses)

	RCT		Cohort			Odds Ratio	Odds Ratio
Study or Subgroup	**Events**	**Total**	**Events**	**Total**	**Weight**	**M-H, Fixed, 95% CI**	**M-H, Fixed, 95% CI**
Koch-Henriksen 2006	209	224	123	74		Not estimable	
Luby 2002	186	162	133	79		Not estimable	
Matilla 2003	302	137	254	166		Not estimable	
Paradise 1984	83	42	63	28		Not estimable	
Total (95% CI)		**565**		**347**		**Not estimable**	
Total events	780		573				
Heterogeneity: Not applicable							
Test for overall effect: Not applicable							

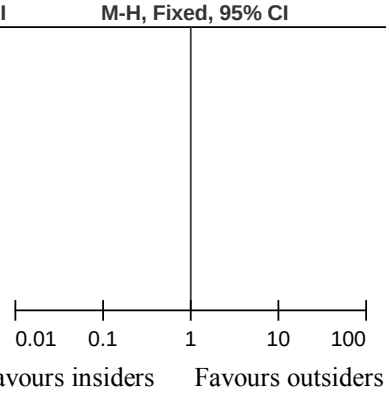


Figure 36: All studies in this meta-analysis non- randomly assigned participants to either a trial or cohort group and had a multiple event dichotomous outcome.

1.9 Subgroups based on treatment effectiveness (Vist analysis)

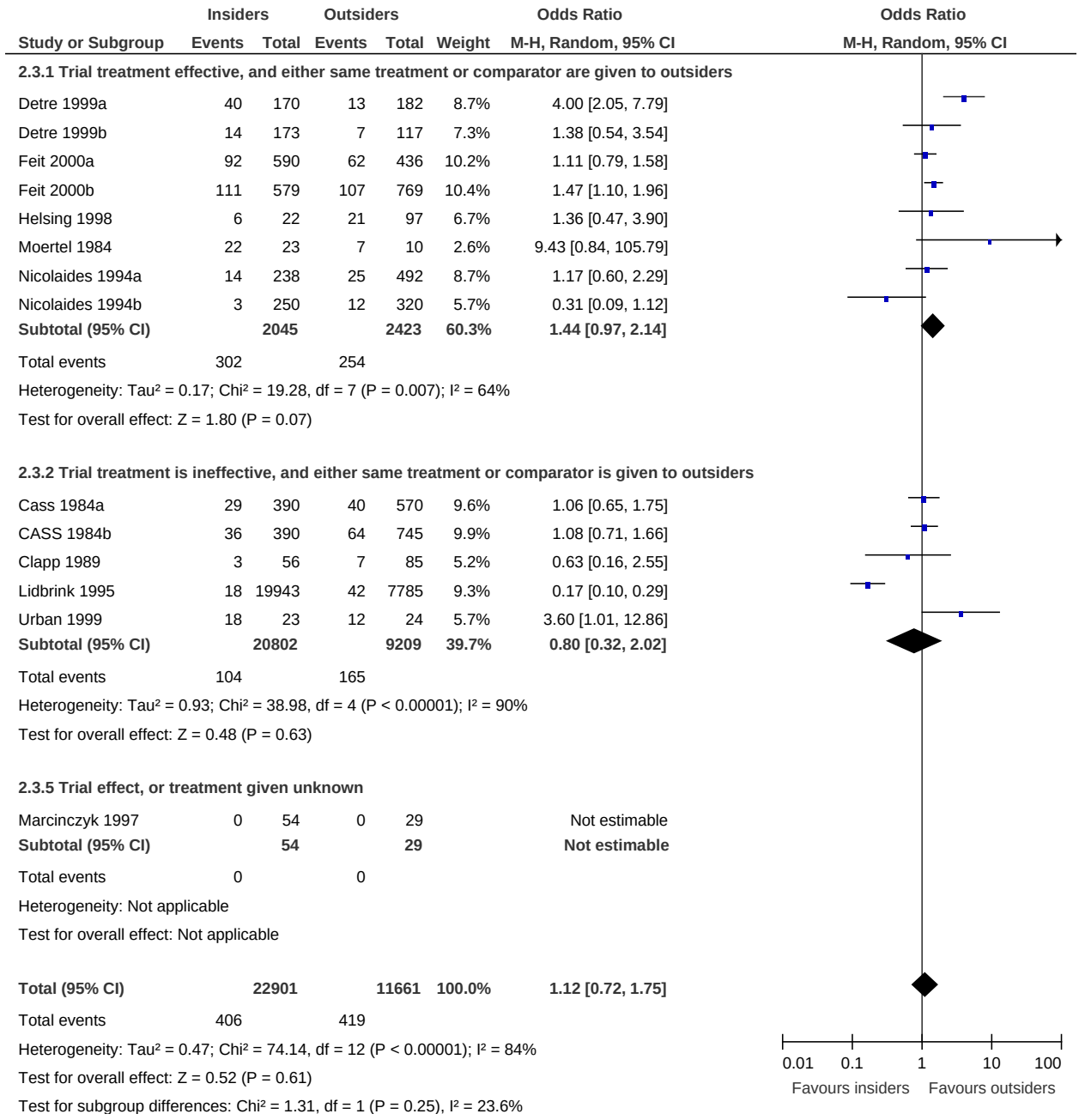
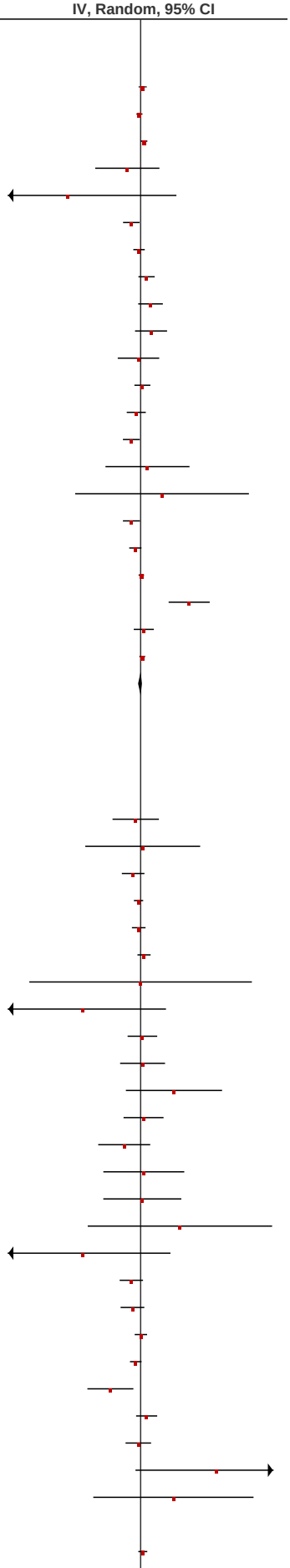


Figure 37: All studies in this meta-analysis feature non-randomized participants either to the trial or cohort group that had a mortality outcome and the same treatment given to the insiders and outsiders. Studies in this forest plot are divided into subgroups based on trial treatment effectiveness to explain heterogeneity. Each study arm is separated in this analysis (i.e. insider treatment compared to outsider treatment, insider control compared to outsider control).

Study or Subgroup	log[Risk Ratio]	SE	Weight	Risk Ratio IV, Random, 95% CI
2.1.1 Trial treatment effective, and either same treatment or comparator is given to the outsiders				
Balmukhanov 1989a	0	0		Not estimable
Balmukhanov 1989b	0.073888	0.061453	4.2%	1.08 [0.95, 1.21]
Bijker 2002a	-0.0395	0.04224	4.5%	0.96 [0.88, 1.04]
Bijker 2002b	0.114238	0.050208	4.4%	1.12 [1.02, 1.24]
Henriksson 1986a	-0.46287	0.558621	0.3%	0.63 [0.21, 1.88]
Henriksson 1986b	-2.54273	1.921409	0.0%	0.08 [0.00, 3.40]
Howie 1997a	-0.32211	0.135469	2.7%	0.72 [0.56, 0.94]
Howie 1997b	-0.05335	0.090372	3.6%	0.95 [0.79, 1.13]
Karande 1999	0.207065	0.133028	2.7%	1.23 [0.95, 1.60]
Kendrick 2001a	0.346383	0.207282	1.7%	1.41 [0.94, 2.12]
Kendrick 2001b	0.363975	0.27374	1.1%	1.44 [0.84, 2.46]
Link 1991a	-0.07351	0.356973	0.7%	0.93 [0.46, 1.87]
Link 1991b	0.06252	0.130438	2.8%	1.06 [0.82, 1.37]
Martinez- Amenos 1990b	-0.1507	0.157915	2.3%	0.86 [0.63, 1.17]
Martinez-Amenos 1990a	-0.32083	0.138295	2.6%	0.73 [0.55, 0.95]
Nagel 1998a	0.23923	0.73432	0.2%	1.27 [0.30, 5.36]
Nagel 1998b	0.745473	1.531096	0.0%	2.11 [0.10, 42.37]
Ogden 2004	-0.31815	0.140102	2.6%	0.73 [0.55, 0.96]
Rorbye 2005a	-0.18243	0.097134	3.4%	0.83 [0.69, 1.01]
Rorbye 2005b	0.024644	0.036872	4.6%	1.02 [0.95, 1.10]
Wetzner 1979	1.690127	0.353554	0.8%	5.42 [2.71, 10.84]
Wieringa- de Waard 2002b	0.111622	0.167539	2.2%	1.12 [0.81, 1.55]
Wieringa-de Waard 2002a	0.059238	0.039793	4.5%	1.06 [0.98, 1.15]
Subtotal (95% CI)			**51.9%**	**1.00 [0.93, 1.09]**
Heterogeneity: Tau² = 0.02; Chi² = 63.37, df = 21 (P < 0.00001); I² = 67%				
Test for overall effect: Z = 0.10 (P = 0.92)				
2.1.2 Trial treatment is ineffective, and either same treatment or comparator is given to the outsiders				
Antman 1985	-0.17063	0.40003	0.6%	0.84 [0.38, 1.85]
Bell 2000	0.073427	1.008574	0.1%	1.08 [0.15, 7.77]
Bhattacharya 1998	-0.25925	0.189121	1.9%	0.77 [0.53, 1.12]
Caplan 1984	-0.07146	0.071474	4.0%	0.93 [0.81, 1.07]
Chilvers 2001a	-0.06546	0.10946	3.2%	0.94 [0.76, 1.16]
Chilvers 2001b	0.119225	0.105472	3.3%	1.13 [0.92, 1.39]
Clagett 1984a	0	1.965215	0.0%	1.00 [0.02, 47.08]
Clagett 1984b	-2.0126	1.467049	0.1%	0.13 [0.01, 2.37]
Cowchock 1992a	0.062035	0.251711	1.3%	1.06 [0.65, 1.74]
Cowchock 1992b	0.068208	0.387924	0.6%	1.07 [0.50, 2.29]
Creutzig 1993a	1.161133	0.842552	0.2%	3.19 [0.61, 16.65]
Creutzig 1993b	0.101858	0.344243	0.8%	1.11 [0.56, 2.17]
Forbes 2000	-0.56947	0.45063	0.5%	0.57 [0.23, 1.37]
Gall 2007a	0.111918	0.707508	0.2%	1.12 [0.28, 4.48]
Gall 2007b	0.059898	0.681468	0.2%	1.06 [0.28, 4.04]
Liu 1998a	1.368393	1.626011	0.0%	3.93 [0.16, 95.14]
Liu 1998b	-2.00148	1.540516	0.0%	0.14 [0.01, 2.77]
Melchart 2002a	-0.32073	0.195922	1.8%	0.73 [0.49, 1.07]
Melchart 2002b	-0.28353	0.199336	1.8%	0.75 [0.51, 1.11]
Morrison 2002a	0.010751	0.098418	3.4%	1.01 [0.83, 1.23]
Morrison 2002b	-0.16722	0.091801	3.5%	0.85 [0.71, 1.01]
Panagopoulou 2009	-1.05124	0.396744	0.6%	0.35 [0.16, 0.76]
Raistrick 2005a	0.210615	0.17592	2.1%	1.23 [0.87, 1.74]
Raistrick 2005b	-0.08029	0.2152	1.6%	0.92 [0.61, 1.41]
Rosen 1987a	2.653242	1.43389	0.1%	14.20 [0.85, 235.95]
Rosen 1987b	1.142923	1.411725	0.1%	3.14 [0.20, 49.89]
Sundar 2008	0	0		Not estimable
Wallage 2003	0.076338	0.067333	4.0%	1.08 [0.95, 1.23]



Raistrick 2005a	0.210615	0.17592	2.1%	1.23 [0.87, 1.74]
Raistrick 2005b	-0.08029	0.2152	1.6%	0.92 [0.61, 1.41]
Rosen 1987a	2.653242	1.43389	0.1%	14.20 [0.85, 235.95]
Rosen 1987b	1.142923	1.411725	0.1%	3.14 [0.20, 49.89]
Sundar 2008	0	0		Not estimable
Wallage 2003	0.076338	0.067333	4.0%	1.08 [0.95, 1.23]
Yersin 1996	0	0.416125	0.6%	1.00 [0.44, 2.26]
Subtotal (95% CI)			**36.5%**	**0.95 [0.88, 1.04]**
Heterogeneity: Tau² = 0.01; Chi² = 33.54, df = 27 (P = 0.18); I² = 19%				
Test for overall effect: Z = 1.12 (P = 0.26)				
2.1.3 Trial effect, or treatment given unknown				
Akaza 1995	2.448539	1.396898	0.1%	11.57 [0.75, 178.83]
Blichert- Toft 1988a	-0.13808	0.047868	4.4%	0.87 [0.79, 0.96]
Blichert- Toft 1988b	-0.21833	0.041728	4.5%	0.80 [0.74, 0.87]
Chauhan 1992	-0.63219	0.972115	0.1%	0.53 [0.08, 3.57]
Edsmyr 1978	0.998529	0.576916	0.3%	2.71 [0.88, 8.41]
Mayo 1992a	3.100092	0.618047	0.3%	22.20 [6.61, 74.55]
Mayo 1992b	-1.21468	0.91144	0.1%	0.30 [0.05, 1.77]
Sullivan 1982a	1.91005	1.387055	0.1%	6.75 [0.45, 102.37]
Sullivan 1982b	0.099318	0.862779	0.1%	1.10 [0.20, 5.99]
Sullivan 1982c	-2.50738	1.066161	0.1%	0.08 [0.01, 0.66]
Van 2009a	0.435318	0.591834	0.3%	1.55 [0.48, 4.93]
Van 2009b	-0.24273	0.263108	1.2%	0.78 [0.47, 1.31]
Subtotal (95% CI)			**11.6%**	**1.04 [0.78, 1.37]**
Heterogeneity: Tau² = 0.07; Chi² = 47.21, df = 11 (P < 0.00001); I² = 77%				
Test for overall effect: Z = 0.25 (P = 0.80)				
Total (95% CI)			**100.0%**	**0.97 [0.91, 1.04]**
Heterogeneity: Tau² = 0.02; Chi² = 172.67, df = 61 (P < 0.00001); I² = 65%				
Test for overall effect: Z = 0.78 (P = 0.43)				
Test for subgroup differences: Chi² = 0.92, df = 2 (P = 0.63), I² = 0%				

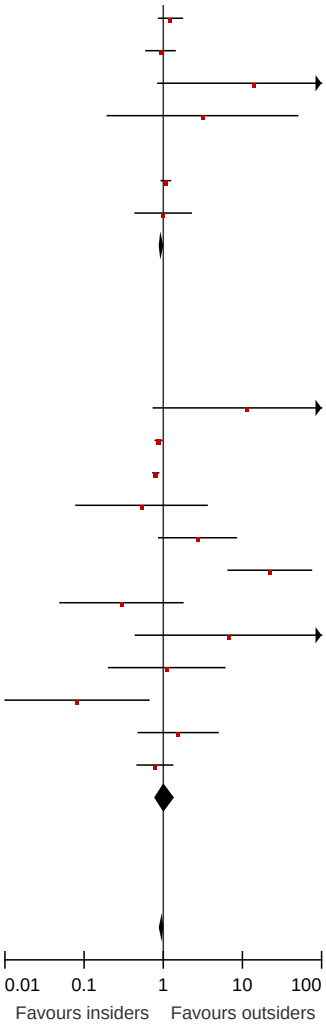
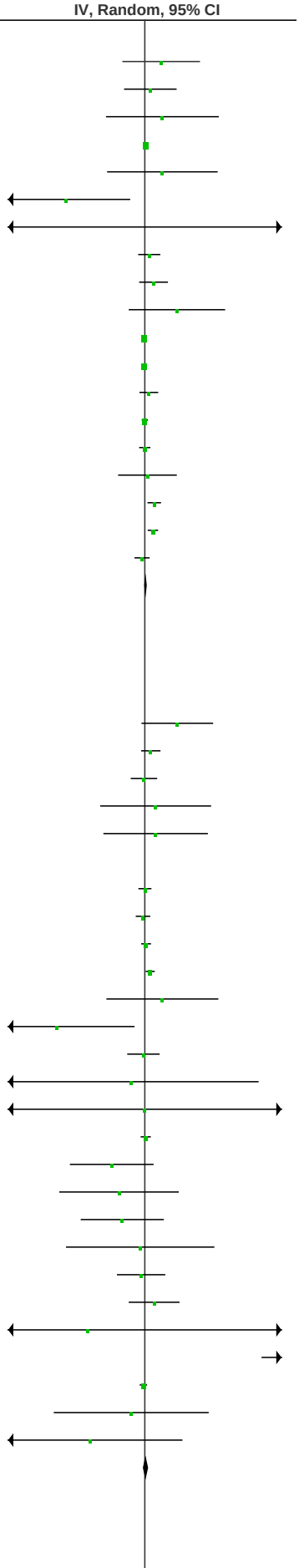


Figure 38: All studies in this meta-analysis were non-randomly assigned to either the trial or cohort group that had a dichotomous non-mortality outcome, and the same treatment was given to the insiders and outsiders. Studies in this forest plot are divided into subgroups based on trial treatment effectiveness to explain heterogeneity. Each study arm is separated in this analysis (i.e. insider treatment compared to outsider treatment, insider control compared to outsider control).

	Insiders			Outsiders				Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Random, 95% CI
2.2.1 Trial treatment effective, and either same treatment or comparator is given to the outsiders								
Ashok 2002a	14.21	4.8	118	13	4.1	9	0.1%	1.21 [-1.61, 4.03]
Ashok 2002b	11.21	5.9	111	10.8	4.7	36	0.2%	0.41 [-1.48, 2.30]
Bannister 2001	24.5	14	97	23.2	9.5	38	0.0%	1.30 [-2.81, 5.41]
Brinkhaus 2008	1.35	0.26	266	1.29	0.16	2469	14.3%	0.06 [0.03, 0.09]
Grant 2008a	84.6	17.9	145	83.3	20.7	212	0.0%	1.30 [-2.73, 5.33]
Grant 2008b	73.4	23	154	79.2	19.2	163	0.0%	-5.80 [-10.48, -1.12]
Gunn 2000	4,303	683	122	4,189	731	160	0.0%	114.00 [-51.89, 279.89]
Jensen 2003a	1.94	4.86	448	1.61	4.54	205	1.2%	0.33 [-0.44, 1.10]
Jensen 2003b	2.57	4.6	449	1.92	4.43	89	0.7%	0.65 [-0.36, 1.66]
Lansky 1983	7.32	7.4	25	4.95	7.73	59	0.1%	2.37 [-1.14, 5.88]
Lock 2010a	0.13	0.21	119	0.19	0.3	248	13.8%	-0.06 [-0.11, -0.01]
Lock 2010b	0.33	0.4	112	0.38	0.34	55	11.6%	-0.05 [-0.17, 0.07]
Masood 2002	5.73	1.6	45	5.43	0.85	14	1.6%	0.30 [-0.35, 0.95]
Mori 2006a	4	0.74	387	4	0.74	77	9.0%	0.00 [-0.18, 0.18]
Mori 2006b	3	1.48	325	3	1.5	81	4.1%	0.00 [-0.36, 0.36]
Witt 2006a	30.5	16.6	274	30.3	19.9	2481	0.2%	0.20 [-1.92, 2.32]
Witt 2006b	39.6	6.3	1563	38.9	11.3	4686	3.0%	0.70 [0.25, 1.15]
Witt 2006c	17	5.8	1321	16.4	4.2	3901	4.6%	0.60 [0.26, 0.94]
Witt 2008	3.1	2.2	93	3.3	2.5	389	2.4%	-0.20 [-0.71, 0.31]
Subtotal (95% CI)			**6174**			**15372**	**66.9%**	**0.07 [-0.03, 0.18]**
Heterogeneity: Tau² = 0.01; Chi² = 50.32, df = 18 (P < 0.0001); I² = 64%								
Test for overall effect: Z = 1.45 (P = 0.15)								
2.2.3 Trial is ineffective, and either same treatment or comparator is given to outsiders								
Andersson 2003	4	4.9	15	1.6	1.1	8	0.1%	2.40 [-0.19, 4.99]
Bain 2001a	1.5	1.24	20	1.06	1.11	32	1.5%	0.44 [-0.23, 1.11]
Bain 2001b	0.94	1.5	16	1	1.58	30	0.8%	-0.06 [-0.99, 0.87]
Bedi 2000a	15.2	11.6	40	14.4	9.8	108	0.0%	0.80 [-3.24, 4.84]
Bedi 2000b	14.8	10	45	14	9.3	56	0.1%	0.80 [-3.00, 4.60]
Boezaart 1998	0.15	0	40	0.4	0.84	136		Not estimable
Dalal 2007a	5.61	1.14	55	5.6	1.12	50	3.2%	0.01 [-0.42, 0.44]
Dalal 2007b	5.54	1.1	45	5.67	1.1	34	2.6%	-0.13 [-0.62, 0.36]
Ekstein 2002a	1.5	0.6	47	1.4	3.3	600	5.0%	0.10 [-0.21, 0.41]
Ekstein 2002b	1.5	0.7	44	1.1	2.5	602	5.7%	0.40 [0.11, 0.69]
Emery 2003a	35.3	9.45	86	34	8.9	24	0.0%	1.30 [-2.78, 5.38]
Emery 2003b	32.8	9.3	82	39.3	13.6	25	0.0%	-6.50 [-12.20, -0.80]
Heuss 2004	2.8	2.5	36	2.9	2.6	40	0.6%	-0.10 [-1.25, 1.05]
Howard 2010a	50	10.3	30	51	16.4	14	0.0%	-1.00 [-10.35, 8.35]
Howard 2010b	51	13.4	14	51	18	14	0.0%	0.00 [-11.75, 11.75]
Jena 2008	4.9	5.89	1442	4.83	5.83	10410	4.9%	0.07 [-0.25, 0.39]
King 2000a	11.96	9.93	107	14.4	9.9	66	0.1%	-2.44 [-5.48, 0.60]
King 2000b	11.4	11	58	13.3	10.7	40	0.0%	-1.90 [-6.26, 2.46]
McKay 1995a	2.85	5.19	20	4.51	7.65	57	0.1%	-1.66 [-4.68, 1.36]
McKay 1995b	6.7	9.1	20	7.04	9	23	0.0%	-0.34 [-5.77, 5.09]
McKay 1998a	1.52	4.18	52	1.79	3.84	33	0.2%	-0.27 [-2.00, 1.46]
McKay 1998b	2.02	4	49	1.33	3.12	18	0.2%	0.69 [-1.14, 2.52]
Reddihough 1998a	29	16.66	8	33.2	13.82	9	0.0%	-4.20 [-18.86, 10.46]
Reddihough 1998b	52.11	18.8	11	28.64	18	13	0.0%	23.47 [8.67, 38.27]
Salisbury 2002	6.1	0.9	134	6.2	0.96	129	7.4%	-0.10 [-0.33, 0.13]
Underwood 2008a	40	22	93	41	25	198	0.0%	-1.00 [-6.67, 4.67]
Underwood 2008b	37	22	94	41	22	73	0.0%	-4.00 [-10.73, 2.73]
Subtotal (95% CI)			**2703**			**12842**	**32.7%**	**0.07 [-0.13, 0.27]**
Heterogeneity: Tau² = 0.05; Chi² = 34.96, df = 25 (P = 0.09); I² = 28%								
Test for overall effect: Z = 0.72 (P = 0.47)								



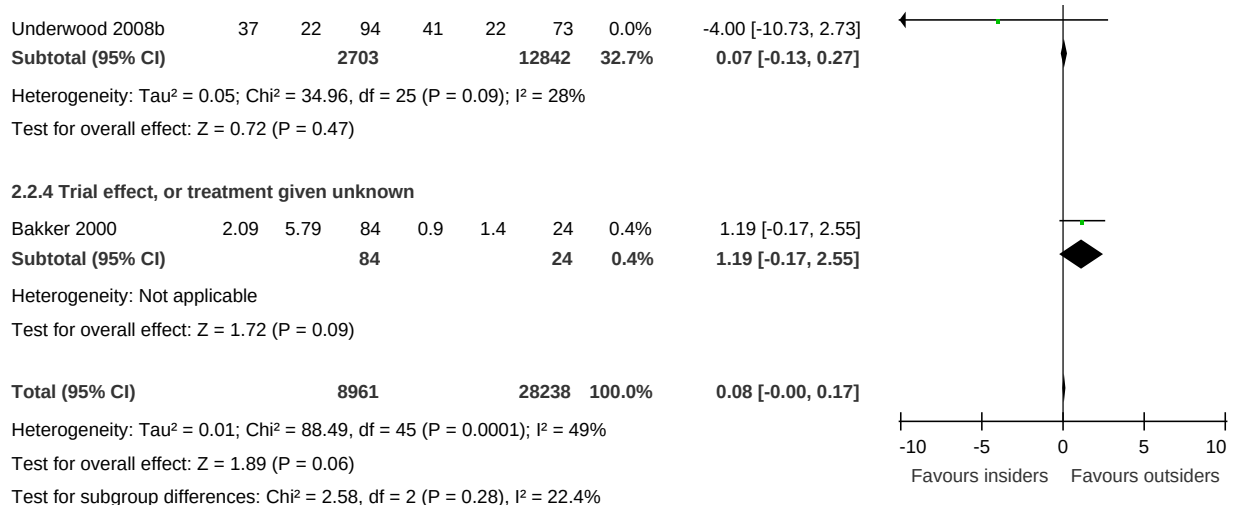


Figure 39: All studies in this meta-analysis non- randomly assigned participants either to the trial or cohort group that had a continuous endpoint as an outcome, and the same treatment was given to the insiders and outsiders. Studies in this forest plot are divided into subgroups based on trial treatment effectiveness to explain heterogeneity. Each study arm is separated in this analysis (i.e. insider treatment compared to outsider treatment, insider control compared to outsider control).

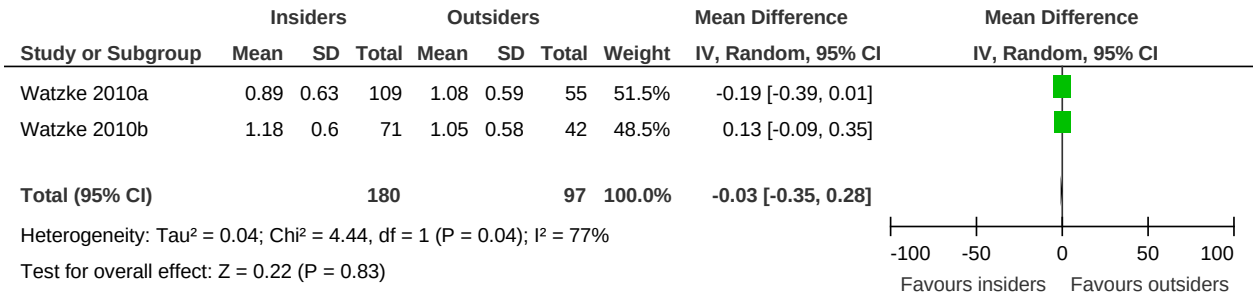


Figure 40: All studies in this meta-analysis randomly assigned participants either to the trial or cohort group that had a continuous endpoint as an outcome, and the same treatment was given to the insiders and outsiders. Studies in this forest plot are divided into subgroups based on trial treatment effectiveness to explain heterogeneity. Each study arm is separated in this analysis (i.e. insider treatment compared to outsider treatment, insider control compared to outsider control).

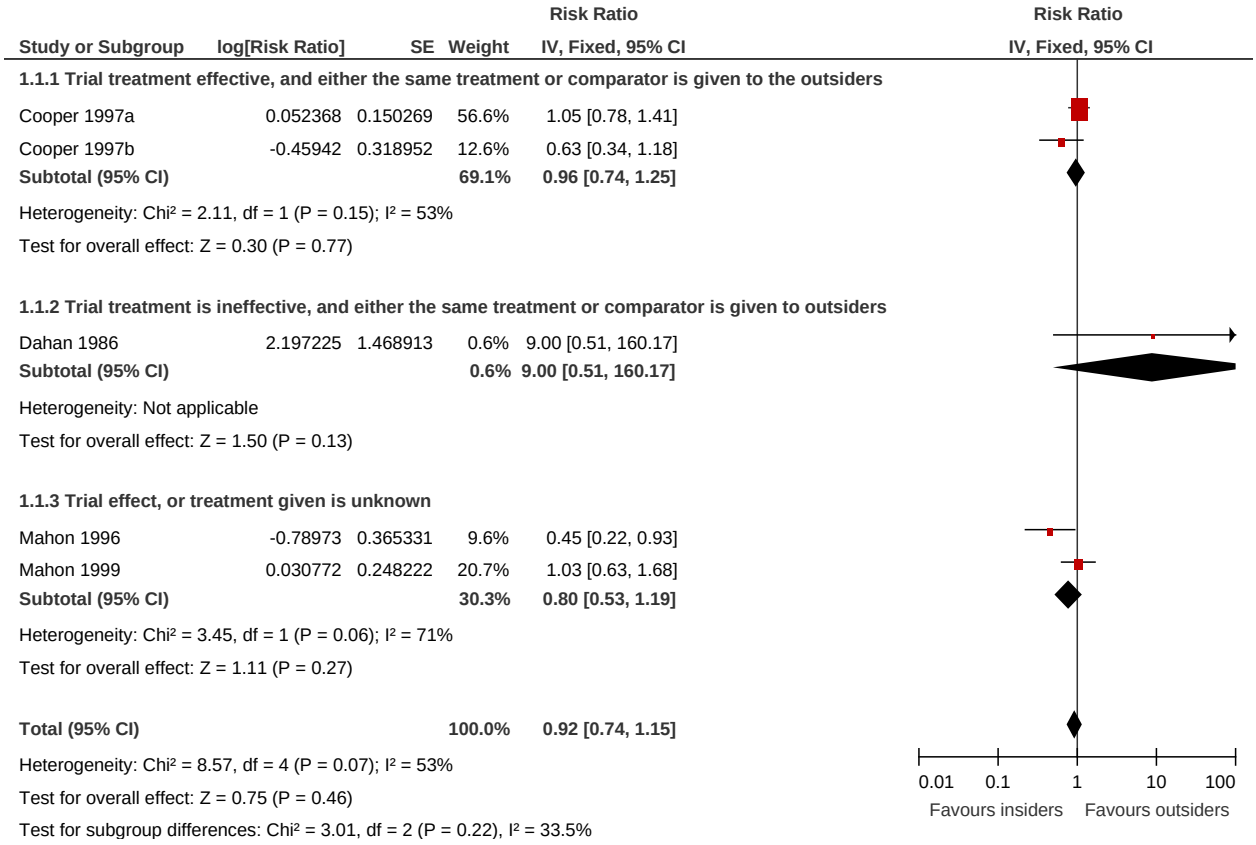


Figure 41: All studies in this meta-analysis randomly assigned participants either to the trial or cohort group that had a continuous endpoint as an outcome, and the same treatment was given to the insiders and outsiders. Studies in this forest plot are divided into subgroups based on trial treatment effectiveness to explain heterogeneity. Each study arm is separated in this analysis (i.e. insider treatment compared to outsider treatment, insider control compared to outsider control).

Funnel Plots

1.1 Subgroups based on trial treatment effectiveness

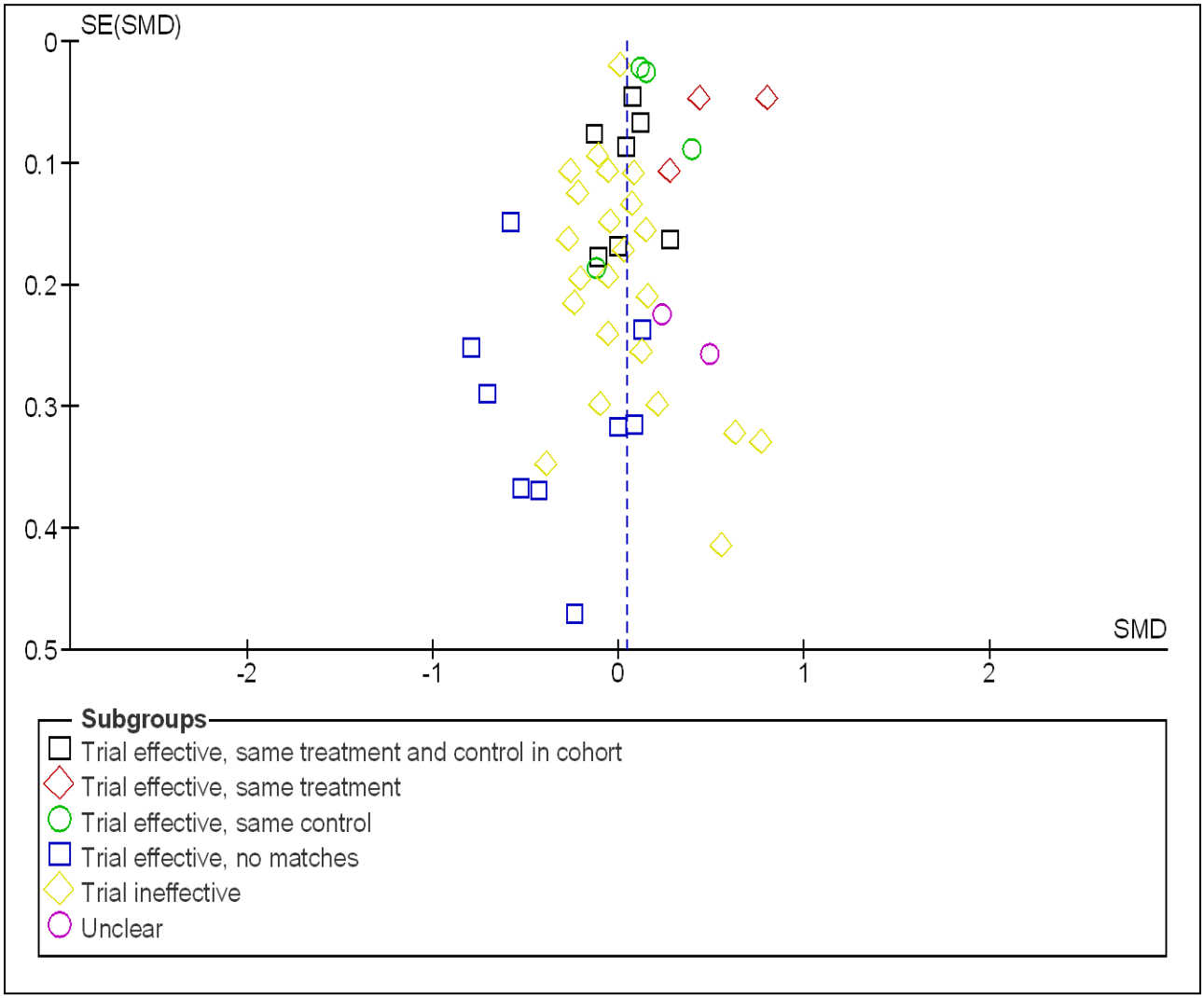


Figure 2: Funnel plot of the continuous outcomes subdivided based on trial treatment effectiveness. Subgroups are denoted according to the colors and shapes provided in the legend. The x- axis is the standardized mean difference (SMD), and the y-axis plots the standard error around the SMD.

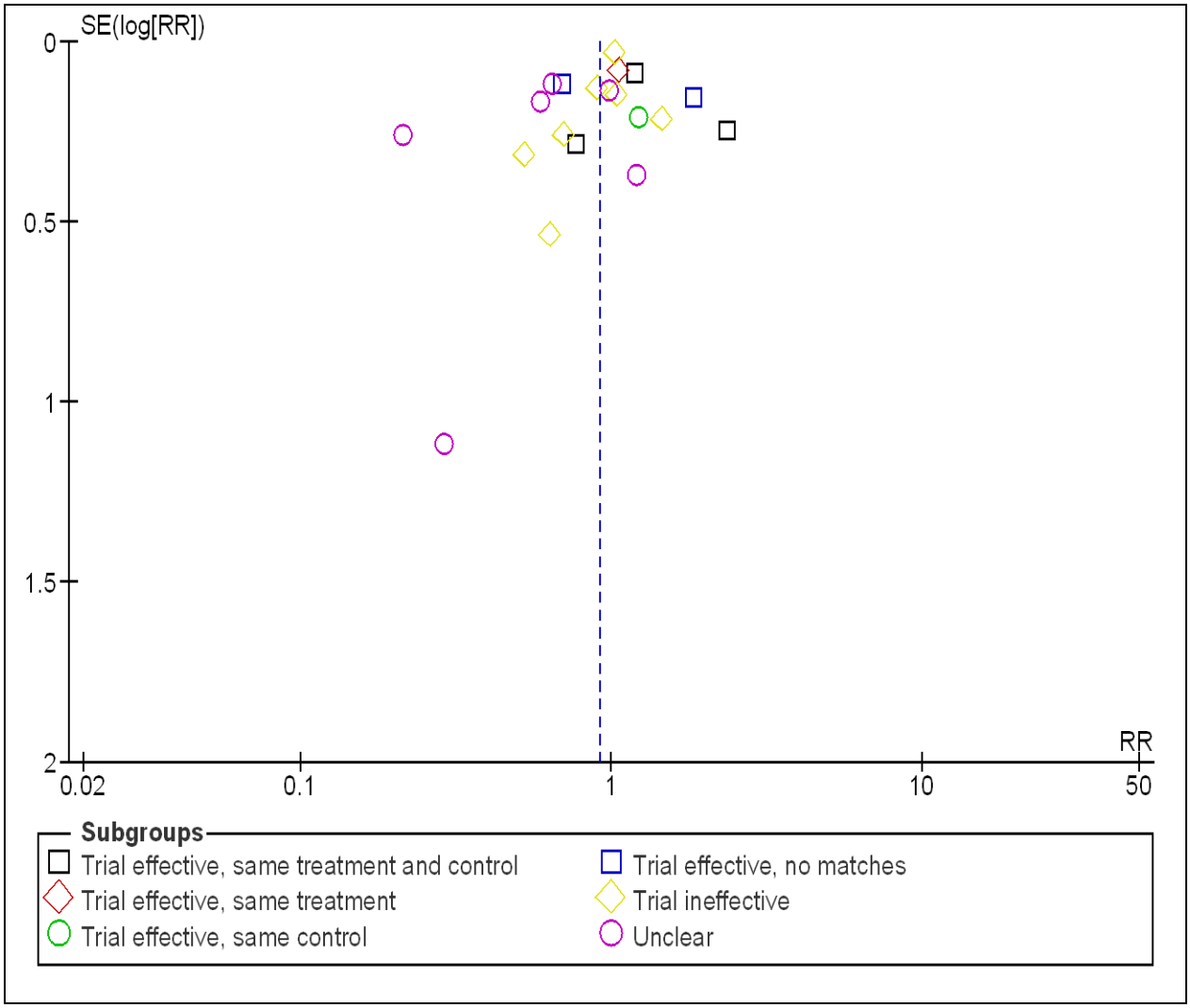


Figure 3: Funnel plot of the mortality outcomes subdivided based on trial treatment effectiveness. Subgroups are denoted according to the colors and shapes provided in the legend. The x- axis is the standardized mean difference (SMD), and the y-axis plots the standard error around the SMD.

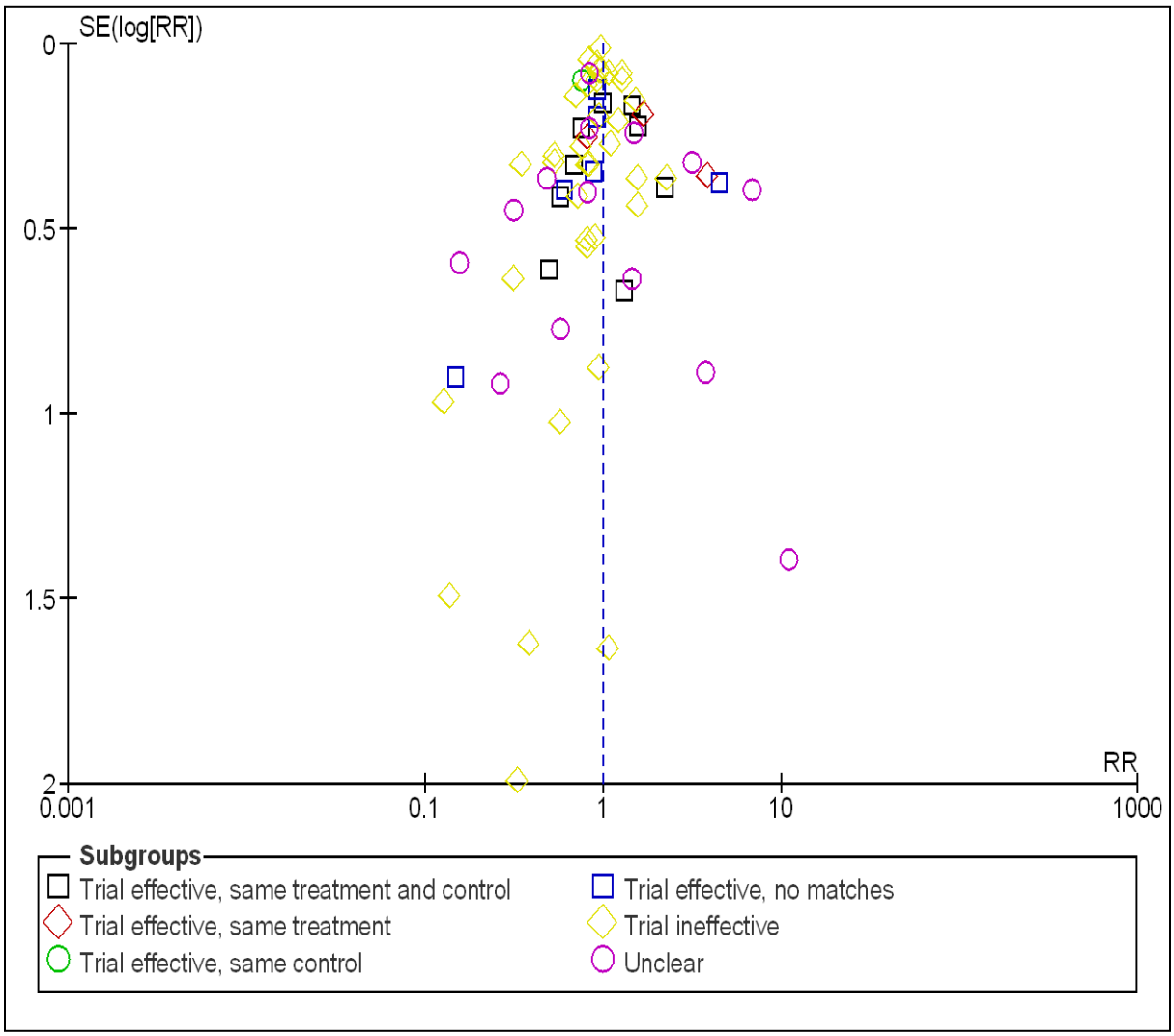


Figure 4: Funnel plot of the dichotomous non- mortality outcomes subdivided based on trial treatment effectiveness. Subgroups are denoted according to the colors and shapes provided in the legend. The x- axis is the standardized mean difference (SMD), and the y- axis plots the standard error around the SMD.