Evaluation of the use of the intention-to-treat approach in randomized controlled trials.

Do authors say what they do and do what they say?

Jocelyn Gravel

Department of Epidemiology and Biostatistics
McGill University, Montreal

Submitted in August 2004

A thesis submitted to McGill University in partial fulfilment of the requirements of the master degree in epidemiology

Your file Votre référence
ISBN: 0-494-12457-1
Our file Notre référence
ISBN: 0-494-12457-1

Table of contents

Acknowledgements

There is only one name on the title page but many people have contributed to the present manuscript. Without their contribution, this thesis would never have been completed.

The idea of doing a Masters thesis in epidemiology was initially suggested to me by Dr. Marie Gauthier, the former director of the Paediatric department of Sainte-Justine Hospital. I want to thank her for her enthusiasm and advice, which proved to be the spark that ignited this two-year process. I also want to thank the Paediatric department who, for the first year, provided financial support in association with the Université de Montréal through a fund called "Bourse en jumelage de fond" and the medical staff of the Emergency department of Sainte-Justine Hospital who always respected the fact that I was a student and therefore gave priority to my academic responsibilities over the clinical demands of the department.

I want to recognize Dr. Lucie Opatrny who reviewed some of the articles, participated in the construction of the evaluation forms and therefore helped to improve the reliability of the study. Dr Barbara Cummins-McManus also helped edit the manuscript.

Dr. Stan Shapiro was a great supervisor. I will be forever grateful for his precious advice. He was always generous with his time and his diplomatic approach in supervising my work was much appreciated. He taught me many important epidemiological concepts, but it was his attention to details that made him an excellent role model. Dr Michael Kramer was a member of my thesis committee. His revision of the manuscript was much appreciated.

Finally, I would like to thank my family. First, my parents Pierre Gravel and Yvette Michaud, who taught me the importance of hard work and education 28 years ago. They probably never expected that I would stay in school for such a long time. I want to thank them for their constant support and their keen interest. Also, I would like to thank my partner, Isabelle D'Amours, and my daughters, Clémentine and Madeleine, because they are the ones that illuminate my life and make every morning happier.

Abstract

Background: The intention-to-treat (ITT) approach is an analytic approach for the analysis of randomized controlled trials (RCT) in which patients are analysed as randomized regardless of the treatment received.

Objective: To evaluate the 1) proportion of articles describing a randomized trial in main medical journals in 2002 reporting the use of ITT, 2) proportion violating a major component of ITT, 3) distribution and management of missing data in the analysis of the studies reporting an ITT analysis.

Method: We conducted a cross-sectional literature review of RCTs reported in 10 medical journals in 2002. A single rater, using a standardized form, evaluated all articles. A second rater evaluated a 10% sample to assess reliability. The proportion of articles reporting the use of ITT was calculated. Among these, the proportion of articles that "analyzed patients as randomized" and the proportion and management of missing data was evaluated using standardized definitions.

Results: Of the 403 articles, 249 reported the use of ITT. Among these, available patients were analyzed as randomized in 192 articles (77±5%). However, more than 60% of the articles had missing data in their primary analysis. The main reason for missing data was loss to follow-up. Few articles reported a strategy for missing data.

Conclusion: This study emphasizes the fact that authors use the label "intention-to-treat" quite differently. Its most common use refers to the analysis of all AVAILABLE subjects as randomized.

Résumé

Introduction: On défini l'analyse « intention de traitement (IT)» comme étant l'analyse de tous les participants tels que randomisés quel que soit le traitement reçu.
Objectifs: Pour les essaies cliniques randomisés publiés en 2002, évaluer : 1. La proportion rapportant l'utilisation d'IT. 2. La proportion de ces derniers violant un principe fondamental de l'IT. 3. Les distribution et approche face aux données manquantes.
Méthodes: Une évaluation des essaies cliniques randomisés publiés dans 10 revues médicales influentes en 2002. Un évaluateur, utilisant une approche standardisée, a extrait les informations pertinentes de chaque article. Une deuxième personne a évalué la fiabilité en analysant un échantillon de 10% des articles. La proportion d'articles rapportant l'utilisation d'IT fut calculée. De ces articles, la proportion de ceux ayant vraiment *analysé les patients tels que randomisés* ainsi que les proportion et approche des données manquantes furent calculées.
Résultats : 249/403 articles rapportèrent l'utilisation d'IT. De ce groupe, les patients furent clairement *analysés tel que randomisés* dans 192 articles (77±5%). Plus de 60% avaient des données manquantes. La cause première était la perte de patients au suivi. Peu d'articles ne proposaient une méthode pour inférer les données manquantes.
Conclusion: Cette étude démontre que l'IT est utilisée de façon variable. Elle désigne habituellement une analyses de tout les patients DISPONIBLES tel que randomisés.

Background

The randomized controlled trial

Randomized controlled trials occupy a critical place in medical research. They are widely viewed as being at the highest level in the hierarchy of epidemiologic research designs[1;2]. A randomized controlled trial is the study design that most resembles the basic science experimental design.

The number of reports of randomized controlled trials in the medical literature is growing: Searching through PUBMED using the publication type "randomized controlled trial", we found 2,051 reports in 1980, 6,745 reports in 1990 and 11,011 reports in the year 2000. Using the Cochrane controlled trials register [3], we found reports of 5,801 randomized controlled trials published in 1980, 14,777 reports in 1990 and 18,467 reports in 2000.

A crucial aspect of the randomized controlled trial is the randomization process. The purposes of randomization are multiple[4;5]: First, randomization controls for bias that could be engendered when the investigator knows the treatment that will be allocated to the study subject before deciding if the subject is eligible for the study. Second, it distributes equitably, among the different treatment groups, the baseline characteristics that could influence the outcome and thereby confound the effect of treatment. In addition to minimizing bias engendered by known prognostic factors, randomization minimizes bias engendered by unknown factors. Indirectly, it facilitate blinding the identity of treatments to the investigators, participants, and evaluators[6]. This property permits reducing information bias introduced after assignment of treatments. In order to maintain the advantages of randomization, study subjects must be analyzed in the treatment arm to which they were initially randomized. Finally, randomization provides a basis for statistical inference.

Despite several decades of use, randomized controlled trials are still reported with variable adequacy [7-13]. For example, Dickinson et al. evaluated the quality of reporting for randomized controlled trials regarding head trauma[11]. They found that only 21% of these articles reported whether outcome assessors were blinded. Dersimonian et al.

evaluated the frequency of reporting on 11 important aspects of design of RCT in articles published in four high-impact journals[13]. They reported that: "*Of all 11 items in the 67 trials published in all four journals, 56 per cent were clearly reported, 10 per cent were ambiguously mentioned, and 34 per cent were not reported at all...Although information about whether patients were blind to treatment was given in 55 per cent, information about whether there was blind assessment of outcome was reported in only 30 per cent*". The authors conclude that: "*This variability can impair readers' ability to assess the quality of the methodology*". Moreover, many reports have shown that trials with improper methodology may be associated with biased results, most often in the direction toward a greater effect for the study intervention [14-16]. For example, Schulz et al. evaluated the relationship between randomized controlled trial design and treatment effect in a database of 33 meta-analyses gathered from the Cochrane Pregnancy and Childbirth Database [15]. The investigators, concluded that "*Compared with trials in which authors reported adequately concealed treatment allocation, trials in which concealment was either inadequate or unclear (did not report or incompletely reported a concealment approach) yielded larger estimates of treatment effects (P < .001). Odds ratios were larger by 41% for inadequately concealed trials and by 30% for unclearly concealed trials (adjusted for other aspects of quality)*". All these articles called for better reporting of the methodology used in randomized controlled trials.

The CONSORT statement

In order to standardize and improve the reports of randomized controlled trials, an international group of trialists, statisticians, epidemiologists, and biomedical editors joined their efforts and proposed the CONSORT statement in 1996[17-20]. It comprises a flow diagram and a checklist of items that should be addressed in a report of a randomized controlled trial. The flow diagram provides the reader with the flow of participants through the study from the evaluation for participation in the study to the final analysis. The checklist is composed of items that have been suggested necessary to be reported on in a RCT. The CONSORT statement can be used in writing, reviewing, or assessing reports of a parallel arms randomized controlled trial. According to their website (www.consort-statement.org), in December 2003, the CONSORT statement had been adopted by 54 medical journals. Key medical journals (e.g.: BMJ, JAMA, Lancet, Annals of Internal Medicine) adhere to the CONSORT

statement. The adhering journals usually stipulate in their "Instructions for Authors" section that they adhere to the CONSORT statement, that all randomized controlled trials submitted be identified as such, and that there should be a CONSORT checklist filled out associated with the paper submission. Many editorial groups such as the International Committee of Medical Journal Editors, the World Association of Medical Editors or the Council of Science Editors adhere to the CONSORT statement. Similarly, organizations as diverse as the American College of Physicians, the American Academy of Medical Acupuncture, the Cochrane Effective Practice & Organization of Care Group and some companies of the pharmaceutical industry endorse the CONSORT statement

CONSORT members meet regularly to discuss and monitor biomedical publications. They report in their website journals and organizations adhering to their statement, evidence of its beneficial effects and a complete and detailed version of the checklist. A revised version of the CONSORT statement was published in April 2001 [21-25]. It is now composed of a checklist of 22 items and a diagram describing the flow of subjects throughout the study. One of the items implemented in the 2001 revised CONSORT statement concerns the method used for the analysis of the primary outcome: Authors are asked to specify whether the analysis was by " intention-to-treat" or not (item #16 of the checklist)[22].

The intention-to-treat analysis

The first use of the phrase "intention to treat" appears to be attributable to Sir Austin Bradford Hill in his well-known book entitled "Principles of Medical Statistics" published in 1948[26]. He suggested that the exclusion of subjects after treatment assignment could affect the validity that randomization sought to provide. He gave the example of a pneumonectomy vs. radiotherapy in a trial evaluating treatment for cancer of the lung. He suggested that exclusion of subjects for whom pneumonectomy was impossible to perform at the time of operation could introduce a systematic difference between the two groups. Thus, patients with the worst prognosis would be excluded from one treatment group. He concluded that: "*Unless the losses are very few and therefore unimportant, we may inevitably have to keep such patients in the comparison group and thus measure the intention to treat in a given way rather than the actual treatment.*"

The intention-to-treat approach has been generally defined as: "*All patients allocated to each arm of the treatment regimen are analyzed together as representing that treatment arm, whether or not they received or completed the prescribed regimen*" [27;28]. Many authorities use a similar definition[3;29;30]. The intention-to-treat analysis has two main components. The first one regards allocation: subjects should be analyzed as randomized even if they did not receive the allocated treatment. This is in contrast to an "as treated" approach where subjects are analyzed according to the treatment they actually received. The second component regards subjects excluded from analysis: ALL randomized subjects should be analyzed. This is in contrast to a "per protocol" approach where only subjects who followed the study protocol and were available at the measurement of the primary outcome are analyzed. Failure to fulfill the second criterion may be due to loss of information (for example: subjects lost to follow-up) or to exclusion of information (for example: non compliant subjects excluded by the investigators).

Intention-to-treat has been advocated as the most appropriate approach for the analysis of outcome in difference seeking clinical trials [31-33]. It is viewed as a method that preserves the prognostic balance between treatment arms and minimizes the risk of finding a difference between two treatments when there is no difference [34]. It generally gives a conservative estimate of the treatment effect compared with what would be expected if there was complete compliance. This approach seeks to minimize overly optimistic estimates of the efficacy of an intervention resulting from the removal of non-compliant subjects. By accepting that non-compliance and protocol deviations are likely to occur in actual clinical practice, the intention-to-treat approach essentially tests a treatment policy or strategy; it reflects the effectiveness of a treatment in clinical practice.

Investigators sometimes use a different definition of intention-to-treat and analyze all *available* subjects as randomized, paying less attention to missing data. According to the CONSORT-statement website (WWW.consort-statement.org): "*The "intention-to-treat" strategy is not always straightforward to implement. It is common for some patients not to complete a study ... and thus not be assessed at the end. Although those participants cannot*

be included in the analysis, it is customary still to refer to analysis of all available participants as an intention-to-treat analysis".

Peto et al. have discussed the management of missing data[35], suggesting that subjects missing from the analysis of a randomized controlled trial could be divided into 3 categories: "*exclusions, withdrawals, and losses.*" They suggested that the *exclusions* are subjects that are excluded before randomization and do not bias the results, the *withdrawals* may lead to bias because they are subjects randomized but deliberately omitted from statistical analysis and, finally, the *losses* are subjects inadvertently lost during follow-up and whose experience can be included in the analysis only until last observation before being lost to follow-up. The losses may lead to bias if the rate and/or type of loss is different between the different treatment options.

The controversy regarding which subjects should be counted and which treatment should be incriminated in the analysis of a clinical trial was also discussed by Sackett and Gent in 1979[36]. According to the authors, the type of analysis used should depend on the type of question one wants to answer: an intention-to-treat approach should be used for a "management" or an "effectiveness" trial. It is a better reflection of clinical practice because it takes into account external factors that would influence treatment in a real-life situation. The inclusion of subjects who were not compliant and those who were treated but were later found not to fulfill the inclusion criteria would give a more realistic picture of the effect of an intervention in clinical practice; a subject wrongly included in the strict and rigid format of a randomized controlled trial would probably be treated in a real-life situation. Explanatory or efficacy trial would be more appropriate to evaluate the true biological efficacy of a treatment but it could compromise the benefit of randomization.

Many authors have demonstrated that exclusion of subjects from the analysis can bias the results. This is because there are generally differences between subjects who comply and those who do not. These differences can produce a difference in outcome [37;38]. A frequently discussed example comes from the Coronary Drug Project [39]: a randomized controlled trial that aimed to compare death rates in subjects assigned a cholesterol lowering drug vs a placebo. A total of 8,341 participants were randomized to receive a placebo (2,789 participants) or one of five lipid lowering drugs

(approximately 1,100 participants per drug). When analyzing the results of compliant subjects taking at least 80% of clofibrate vs all those who did not take the medication (the confirmed non-compliant subjects and those randomized to the placebo), the investigators found a difference in death rates (15% clofibrate vs 20% placebo or non-compliant). When they analyzed all subjects according to their randomized allocation, they found equivalent death rates (18% vs 19%). The death rate of the non-compliant subjects was similar among those randomized to placebo and those randomized to the study drug (approximately 25%). This study has been reported by many as an example of a bias that can be prevented by an intention-to-treat analysis [38;40;41].

Another example is the Joint Study of Extracranial Arterial Occlusion[42] which evaluated the prognosis following surgery or non-surgical treatment for transient cerebral ischemic attacks and cervical carotid artery lesions. The investigators first restricted their evaluation to subjects "available for follow-up" and reported a significant risk reduction in the primary outcome among subjects undergoing surgery. To be "available for follow-up", subjects had to be discharged alive and free of stroke after their initial hospitalization. This policy led to the exclusion of 19 surgically treated subjects and 2 medically treated subject. The inclusion of these 21 subjects in the analysis changed the estimated risk ratio from 0.74 (95%CI 0.59-0.83 with the subjects excluded) to 0.82 (95%CI 0.74-0.91 intention-to-treat).

Missing outcomes secondary to loss to follow-up may also induce bias in studies because of differential rate of losses for the different treatment groups. For example, Farwell et al. reported a randomized controlled trial to study the effects of phenobarbital treatment in febrile seizure in children on intelligence[43]. They showed that children who failed to report for the evaluation at two years of treatment were those expected to perform less well on testing based on their baseline characteristics. In addition, there were more losses to follow-up in the phenobarbital group than in the placebo group.

The risks of misuse with an evolving statistical method

In many fields, the introduction of a new technique is usually associated with a variable quality of utilization generally defined as the "learning curve". The concept of a learning curve can be applied to any advancement in fields as diverse as the

introduction of a new surgical procedure, a new informatics program for the management of references in the construction of a manuscript, or the introduction of a new way of throwing a curve ball for a baseball pitcher. For example, it has been shown for laparoscopic surgery that the operative time is usually longer until the surgeon gains experience with the technique[44;45]. It is now possible to quantify these learning curves. For example, Taipale et al. reported that the sensitivity in screening for major malformations by early ultrasonography for pregnant women can be achieved after a learning curve of 3-4 years[46].

In keeping with the introduction of other new technologies, new statistical methods should first be rigorously evaluated by biostatisticians and epidemiologists. When they are well studied and they seem accurate, they should slowly be integrated and used by other investigators (physicians, for example). The misuse of a statistical method has occasionally been reported with its democratization. For example, an article reporting a randomized controlled trial of the treatment for myocardial infarction inappropriately reported the use of random allocation[47]. A second analysis of the study showed that the randomization was in fact quasi-randomization based on an odd/even days of admission[48]. It led to very suspicious imbalance between treatment arms. Greenhouse et al. described the case of a previously reported two-phase maintenance therapy trial where subjects who responded to treatment during the acute phase were then randomized to maintenance therapy[5]. They demonstrated a confounding effect that biased the results toward a positive result in the trial. Another example comes from the use of cluster randomization. Simpson et al. demonstrated that many investigators (12 out of 21 reports) did not account for the effect of cluster randomization in their analysis for cluster studies [49]. It is now well known that omitting the cluster effect in the analysis of cluster studies may lead to exaggerated claims of statistical significance.

The best analytical method can still give biased results if it is not appropriately used. According to Baumgardner: *Historically, the number of scientific articles published in which inappropriate statistical analyses were performed is alarming. Only when the reader understands the problem and demands change will this situation improve. The consequence of inaction is to be mired with an array of poorly designed articles that, at the very least, do not advance the field of study and, at worse, may influence*

practitioners not well versed in statistics to expose patients to useless, unnecessary, or even harmful procedures [50].

Previous evaluation of the use of the intention-to-treat principle

It seemed important to evaluate whether or not investigators who claimed to perform an intention-to-treat analysis really did so. Upon review of the medical literature using the Pubmed database, only two articles evaluated the rate and adequacy of reports of an intention-to-treat analysis. Hollis et al. evaluated all randomized controlled trials reported in 4 main medical journals (BMJ, JAMA, Lancet, New England Journal of Medicine) during 1997 [51]. They reported that 48% of the randomized controlled trials reported the use of an intention-to-treat analysis. Of the 119 articles that reported the use of an intention-to-treat analysis, 12 excluded subjects who did not start the allocated intervention and 3 did not analyzed all randomized subjects as allocated. The authors concluded that the intention-to-treat analysis is often inadequately described and applied. The rate of intention-to-treat analysis utilization could have changed with the revision of the CONSORT statement because authors have to specify whether they used an intention-to-treat analysis in the revised CONSORT statement. In their study, Hollis et al. described thoroughly deviant reports on the "analyzed as randomized" component of the intention-to-treat principle, but the description of the management of missing data was limited.

In the second article, the authors evaluated a sample of 100 randomized controlled trials that used the word "intention-to-treat" or "intent-to-treat" in their abstract [52]. They reported that only 42% of their study population included all randomized subjects in their analysis. The main reasons for missing data were subjects lost to follow-up and those excluded because they received no treatment. This study was performed on articles published before the revision of the CONSORT statement and mainly concerned the exclusion of subjects from the analysis; little attention was paid to whether subjects were analyzed as randomized. Another important limitation of the study is selection of the articles; articles that mentioned "intention-to-treat" in their abstract may have used the term differently than others that mentioned it only in the Methods section. One could suppose that authors who mention the intention-to-treat analysis in their Abstract attach more importance to the concept than those who name

it only in the Methods section. The importance paid by the investigator to the concept could be related to its appropriateness of use.

Predictive factors of the appropriateness of use of intention-to-treat

Many studies have evaluated journal, investigator, or research subject characteristics associated with methodological quality in randomized controlled trials. Journal characteristics (high citation rates, impact factors, low acceptance rate) have been positively associated with higher methodological quality [53]. Other studies showed that the presence of an epidemiologist or a biostatistician as an author is associated with more rigorous method [54;55]. It has been reported that better methodologicval quality was associated with a lower rate of statistical significance, source of funding and larger size of the study [12;53;54;56]. Finally, the use of the CONSORT statement has been associated with better quality of reports[57]. Upon a systematic review of the literature, we found no report of the evaluation of factors associated with the appropriateness of use of intention-to-treat analysis.

The objectives and rationale of this study

This study was performed to answer four questions:

1. What proportion of the randomized controlled trials published in major medical journals during the year 2002 reported the use of the intention-to-treat approach?
2. What proportion of the trials reporting the use of the intention-to-treat approach deviated from any of its major aspects (e.g.: subject not analyzed as randomized)?
3. In the primary analysis of randomized controlled trials that report the use of an intention-to-treat approach, what is the rate of missing data, and at what point post randomization are subjects lost?
4. What are the strategies used for analyzing missing data in articles that report the use of an intention-to-treat approach?

Methods

Study design:

This study was a cross-sectional literature review of a sample of the randomized controlled trials articles published in 2002.

Source material:

To be included the articles had to fulfill three criteria:

- Published in one of 10 selected medical journals: *Annals of Emergency Medicine, Annals of Internal Medicine, British Medical Journal, Canadian Medical Association Journal, JAMA, Journal of Pediatrics, Lancet, The New England Journal of Medicine, Pediatric Emergency Care* or *Pediatrics*.
- Published from January 1st to December 31st 2002.
- Report a randomized controlled trial, as defined by the Cochrane Collaboration[3]: "*the participants (or other units) were definitely or possibly assigned prospectively to one or two (or more) alternative forms of health care using a process of random allocation*".

The 10 journals were a non-random sample chosen to be comparable with a previous study about the use of the intention-to-treat principle[51] and were based on the clinical interest of the principal investigator (pediatric emergency).These journals are described in Appendix I. Five journals had an impact factor higher than 5.0 (from the Institute for Scientific Information[58]): Four were the journals evaluated in a previous study of the intention-to-treat principle[51]and one was another major general medical journal. The 10 medical journals were selected from different strata of journal impact factor.

Articles with any one of the following criteria were excluded:

- Abstract or resumé of a full report previously published in another journal. This was done to decrease the risk of having the same study counted twice.
- Abstract of an oral presentation (because there is an important loss of information in an abstract that could bias the results of our study). Also, it has been reported that only 35-50% of those abstracts will eventually be published [59-61].

- Articles that did not report the results of a randomized controlled trial. This eliminated articles describing the methodology of a randomized controlled trial that and eliminated observational (non-experimental) analysis of a RCT study population.

Article retrieval:

Over the past years, investigators have questioned the sensitivity of a Medline/Pubmed search to retrieve randomized controlled trials[62-64]. Depending on the search strategy used, Medline searches had a sensitivity ranging from 50 to 98%. The diminished sensitivity was explained by the fact that not all journals are indexed in Medline, many randomized controlled trials are reported as a type of report that is not indexed (abstract, comments, conference proceeding, etc), and because of misclassification of reports. The high variability has led authors to suggests: "*a combination of MEDLINE and hand searching is required to identify adequately reports of randomized controlled trials*" [63]. The hand-search being performed on the non-indexed part of journals (supplements, letters to the editors, comments).

The Cochrane Collaboration [3] is an international non-profit organisation, dedicated to making up-to-date, accurate information about the effects of treatment readily available worldwide. It produces and disseminates systematic reviews of healthcare interventions and promotes the search for evidence from clinical trials and other studies of interventions. One of its objectives is to identify, through hand-searching of journals of all languages, as many randomized controlled trials as possible. It is now viewed as the most complete database of randomized controlled trials[65]. This database is constructed from hand-search and articles are usually registered earlier in the electronic databases (e.g., Pubmed) than in the Cochrane database. The delay in report completeness is approximately 2 years for high-impact journals and can be many years for lower-impact journals. The ideal search strategy used to retrieve randomized controlled trials would be highly sensitive (retrieve all articles that fulfill the inclusion/exclusion criteria), precise (would not retrieve articles that do not fulfill the inclusion/exclusion criteria), reliable, up-to-date and easy to perform. No strategy has thus far fulfilled all these criteria. When performing a literature search, one has to balance efficiency (the time needed to retrieve the reports) and completeness. The literature search strategy used may depend on the study purpose: While a systematic

review of the effectiveness of a specific treatment demands the evaluation of all relevant reports (abstract and full-reports), a methodological evaluation of randomized controlled trials reports may have to evaluate only full reports.

In the present work, the sensitivity of a Pubmed search in retrieving articles that report a randomized controlled trial published in the 10 selected journals in 2002 using two different strategies was compared to a hand-search strategy. For each journal, issues of one randomly selected month were evaluated. All these issues were examined using 3 search strategies:

1. Hand-search: A hand-search was performed by a single investigator (JG) for each of the selected issues. The selected issues were read from cover to cover looking for reports of randomized controlled trials, as defined by the Cochrane Collaboration[3]: "*the participants (or other units) were definitely or possibly assigned prospectively to one or two (or more) alternative forms of health care using a process of random allocation*". Each article were read until satisfied that the report was either 1. an article reporting a randomized controlled trial, 2. an abstract/resumé of a randomized controlled trial, or 3. not a randomized controlled trial.

2. Pubmed search (RCT type): In the Pubmed database (www.ncbi.nlm.nih.gov): a search was performed using the journal name and the limitations: publication date: "*2002-months-01 to 2002-months-30*" and "publication type: *randomized controlled trial*". The selected reports were read and classified as previously described.

3.Pubmed search (Cochrane strategy): The second Pubmed search used the journal name, the limitations: publication date: "*2002-months-01 to 2002-months-30*" and the search strategy reported by Robinson and Dickersin [66]:

(Randomized controlled trial [pt] OR controlled clinical trial [pt] OR randomized controlled trials [mh] OR random allocation [mh] OR double-blind method [mh] OR single-blind method [mh] OR clinical trial [pt] OR AND (mask*[tw] OR blind* [tw])) OR ("latin square"[tw]) OR placebos [mh] OR placebo* [tw] OR random* [tw] OR research design [mh:noexp] OR comparative study [mh] OR evaluation studies [mh] OR follow-up studies [mh] OR prospective studies [tw] OR cross-over studies [mh] OR control* [tw] OR prospectiv* [tw] OR volunteer* [tw]) NOT (animal [mh] NOT human [mh]).

This search strategy is an adaptation of the highly sensitive strategy developed by the Cochrane collaboration for OVID-Medline[67]. Abstracts of all selected articles were evaluated and classified according to the 3 possibilities previously described. An article selected by the search strategy that was not an article reporting a randomized controlled trial was considered a false positive[62-64].

A pool of all the articles that fulfilled our inclusion/exclusion criteria was constructed from the results of all 3 searches. It served as the gold standard. The sensitivity and positive predictive value (and their 95% confidence intervals) were calculated for the 3 search strategies. The positive predictive value for each search strategy was defined as the number of articles fulfilling the inclusion/exclusion criteria selected by the strategy divided by the total number of articles selected by that same strategy. We also calculated the "number-needed-to-read" as defined by Bachmann et al. [68]. This number is the number of selected articles that someone must read to find an article fulfilling the inclusion/exclusion criteria. It is equal to: 1/ (positive predictive value). An ideal search strategy would have 100% sensitivity and a number-needed-to-read of 1.0. It was initially decided that the Pubmed (RCT type) search strategy would be used for the completion of the article retrieval if its sensitivity was > 95%, because the gain in efficiency would then overcome the very small loss in completeness. It was also expected that there would be no bias engendered by the article selection strategy, because we did not expect to observe any relationship between the use of an intention-to-treat analysis and the risk of an article not being retrieved through the selection process.

Measurements

Outcome measures (see Appendix II):

The presence of the words: "intention-to-treat", "intent-to-treat" or "analyzed as randomized" for the description of the primary analysis in the article was the first outcome measure. The primary analysis was defined as the analysis that was performed to answer the question of the primary objective as described in the Methods section [1]. It was usually reported in the Results section of the Abstract [2] and

specified as the primary analysis in the Results section [3] of the article. In case of discrepancy among these three sources, the hierarchy of importance we used was [1] before [3] before [2], because we suspected that the primary objective defined in the Methods section was more likely to reflect the initially planned objective of the study. When more than one outcome fulfilled the defined criteria (for example, when the outcome was measured at two different time periods), the rater used his judgment based on what seemed to be the most clinically relevant outcome. This was usually the last measured outcome.

For the articles reporting the use of an intention-to-treat approach, four aspects of the analysis were evaluated (analysis actually used, type and extent of missing data, management of missing data, and management of improper inclusions):

1. Based on the information provided in the Methods and Results sections, the rater classified the method actually used for the primary analysis (ignoring missing data). The classifications were:
   - Per protocol (analysis of only the subjects who followed the assigned protocol, exclusion of non-compliant subjects as defined by the authors).
   - Per treatment (analysis of subjects according to the treatment they received regardless of their randomization)
   - Modified intention-to-treat (analysis of subjects as randomized, but with some restrictions specified by the investigators)
   - Intention-to-treat (analysis of subjects as randomized)
   - Unable to conclude from the reported information

2. The second aspect regarded missing data: We compared the number of subjects randomized and analyzed in the primary analysis. In the Result section of each article, we retrieved, when possible, for each intervention arm:
   - Number of randomized subjects
   - Number of subjects withdrawn before the beginning of treatment
   - Number of subjects lost to follow-up
   - Number of subjects excluded from the analysis because the subject had been improperly included
   - Number of subjects excluded from the analysis for non-compliance

- Number of subjects excluded from the analysis for other or unclear reasons
- Number of analyzed subjects

The evaluators specified when they were unable to identify the number of subjects randomized or analyzed. The number of randomized subjects should be equal to the number of analyzed subjects + all the missing data categories. When the type of missing data was unclear, it was entered in the "subjects excluded for other reasons" category.

3. The approach used for missing data was evaluated for articles that reported the use of an intention-to-treat approach but had missing data. The evaluators described the single method reported by the investigators for the management of missing data in their primary analysis and all methods reported for all primary and secondary analyses. The possible answers were:

-
- Subjects with missing data excluded
- Subjects with missing data kept in the denominator for the analysis (missing subjects were assumed not to have the outcome of interest)
- Use of worst/best case scenario
- Last observation carried forward
- Regression (description of any regression method used)
- Replacement by a mean
- Multiple imputation
- Other method
- Unclear management of missing data (the evaluator was unable to clearly state the analytic strategy because it was not stated in the text and it could not be infered from the results)

4. Management of improper inclusion: We described the method used by the trials authors to analyze subjects who were initially randomized but subsequently were found to meet exclusion criteria or failed to fulfill all the inclusion criteria. This was classified as:
   - No mention, or specifically stated that there were no improper inclusions
   - Improper inclusions included in the primary analysis

- Improper inclusions excluded from the primary analysis
- Performed the analysis with and without improper inclusions

Covariates

Several independent variables were measured in an attempt to understand variation in the intention-to-treat utilization:

1. Characteristics of the journal where the article was reported:
   a. Adherence to the CONSORT statement (based on the Instructions to Authors section of each journal in 2002)
   b. Impact factor [58] (Impact factor higher or lower than 5.0)
   c. Specialized or general medical journal.

2. Characteristics of the trials' investigators:
   a. Number of authors for the paper
   b. Presence of an affiliation with an epidemiology or biostatistics department in the authorship.

   These were retrieved from the author and affiliation section of the article. (When there was a writing committee, only the members in the writing committee were considered for assessing this covariate.)

3. Characteristics of the study:
   a. Type of intervention (drug/ nutritional supplement, surgical, behavioral, device, physical therapy, diagnosis/screening procedure, other)
   b. Source of funding (industry, public, no funding declared or combination)
   c. Single vs. multiple center study
   d. Use of placebo
   e. Nature of the primary outcome (categorical, continuous, survival)
   f. Type of trial design (parallel, cross-over/matched, factorial)
   g. Number of study participants
   h. Type of randomization (individual or group)
   i. Nature of hypothesis (equivalency or difference-seeking)

Evaluation of the reports

All articles were downloaded from the electronic versions of the journals. They were printed in a pdf format using Acrobat Reader 5.1 (Adobe, Acrobat Reader). All selected reports were evaluated using a standardized form (see Appendix II). During the inter-rater evaluation phase, two evaluators completely read the articles to retrieve the pertinent information. They did not discuss the articles with each other during the inter-rater evaluation phase. In order to increase the validity of their evaluation, however, the evaluators were allowed to discuss unclear aspects about articles or topics that were not part of the inter-rater reliability evaluation. Given the workload, all articles were evaluated by a single evaluator (JG) after successful demonstration of satisfactory inter-rater agreement.

Evaluators

The two evaluators, physicians (JG and LO) who were enrolled in the McGill University MSc program in epidemiology, trained themselves to evaluate articles. The training was composed of a period of description and clarification of the evaluation form and practice on five articles. A random sample of 40 articles (10% of the total number of articles) were then evaluated separately by both raters at the beginning of the study. Discrepancies were resolved by discussion. Inter-rater reliability was calculated using kappa statistics[69] for nominal variables and intra-class correlation coefficient (ICC) for continuous variables[70-72]. The ICC was calculated using SAS software version 8.2 (SAS institute, Inc., Cary, North Carolina). It was specified that the remaining articles would be evaluated by a single evaluator if there was "substantial agreement" between the evaluators using the criteria of Landis and Koch (kappa coefficient> 0.60) [73].

Data analysis

Database

Data were manually transferred from the evaluation charts to an electronic database by a single study investigator (JG). All data were entered on an Excel file and analyzed using SAS software version 8.2 (SAS institute, Inc., Cary, North Carolina).

Cleaning of the data was performed using two methods. First, for each variable the data were ranked using the "*sort by*" tool of the Excel software. The second method was through the use of the "*proc univariate*" and "*proc freq*" procedure of the SAS software. These procedures permitted identifying the missing data and extreme values. The outliers were then examined by using the study forms and sometimes by reviewing the article.

The number of articles that reported the use of the intention-to-treat approach divided by the number of reports evaluated provided the proportion of articles that claimed the use of intention-to-treat analysis.

The second objective was evaluated by the calculation of a simple proportion:

$$= \frac{\text{articles } \textbf{that actually used} \text{ an ITT approach according to the evaluator}}{\text{articles that } \textbf{reported} \text{ the use of an ITT approach}}$$

A table was used to describe the analytic approach actually used (according to the evaluators) for articles that reported the use of an intention-to-treat approach.

The third objective was to describe missing data in the analysis of articles that reported the use of an ITT analysis. There are several resulting tables:

- Table 7 provides missing data rates (all treatment arms combined) in the primary analysis. This missing data rate is calculated by:

$$\frac{\text{\# of subjects randomized - \# of subjects analyzed}}{\text{\# of subjects randomized}}$$

- Table 8 provides rates of losses to follow-up (all treatment arms combined) in the primary analysis. This rate is calculated by:

$$\frac{\text{\# of subjects lost to follow-up}}{\text{\# of subjects randomized}}$$

- Table 9 provides rates of exclusion (all treatment arms combined) in the primary analysis. This rate of exclusion is calculated by:

$$\frac{\text{\# of subjects excluded for non-compliance+ \# of subjects excluded for improper inclusion + \# of subjects excluded for other or unclear reasons}}{\text{\# of subjects randomized}}$$

Finally, the fourth objective concerned the **management** of missing data in articles that reported the use of an ITT analysis. Two tables describe the distribution of the analytic strategy for missing data.

- Table 10 reports the single strategy used for missing data in the primary analysis (as described above) in articles that reported the use of an intention-to-treat approach. In this table, each article had only one approach.

All proportions are reported with their corresponding 95% confidence intervals.

Sample size:

The 2nd and 3rd objectives of the study were the parts with the highest expected number of observations needed. Consequently, they served as a basis for the calculation of the sample size. Based on previous studies[51;52] it was estimated that approximately 10 % of articles that report an intention-to-treat approach would violate the "analyzed as randomized" aspect of the principle. A 95% confidence interval width of ± 5% for that proportion seemed reasonable. Previous studies also showed that missing data categories vary from 1 % to 50%. We decided that the confidence interval width for these proportions should be 6% for proportions lower than 5% (for example, 5% ±3%), 10% for proportions between 5 and 20% (for example, 15±5%) and 14% for higher proportions (for example, 23-37%). We have not taken into account the cluster effect engendered by the fact that these articles came from only 10 journals because we did not have the interclass and intraclass correlation needed for the calculation. The table below provides the number of articles needed for the different possible scenarios using the following formula:

$N = 4 * (1.96)^2 * P * (1-P) / (W)^2$

Where N= total number of articles required

P= expected proportion of outcome of interest

W= desired width of the confidence interval

Proportion	CI ± 0.03	CI ± 0.05	CI ± 0.06	CI ± 0.07
0.01	43			
0.03	124	45		
0.05	202	73	51	
0.10	384	139	96	71
0.20	683	245	171	126
0.30	897	323	225	165
0.50	1067	385	256	196

Based on this table, it was felt that a sample size of approximately 400 articles would be sufficient to provide the desired precision assuming that approximately 50% of randomized controlled trial articles would report the use of an intention-to-treat analysis [51].

We estimated that the evaluation of all articles reporting a randomized controlled trial published in 10 journals during a complete year would be sufficient to reach this number, because journals with high impact factors reported between 40 and 90 randomized controlled trials in the year 2002, while moderate impact journals reported between 5 and 40 randomized controlled trials. The evaluation of the selected journals was expected to provide approximately 450 articles.

Results

Evaluation of the performance of the 3 search strategies

A total of 22 journal issues and 1 supplement were retrieved for the evaluation of the retrieval strategies. These were hand-searched and evaluated using the 2 Pubmed search strategies. This resulted in a combined pool of 38 articles fulfilling the inclusion/exclusion criteria. Table 1 shows the sensitivity and number-needed-to-read to find articles fulfilling the inclusions/exclusions criteria for each search strategy. In brief, all 3 methods had a similarly high sensitivity (from 97% to 100%). However, the number-needed-to-read was quite variable. We had 1.1 abstracts to read to find a relevant article using a Pubmed search using "publication type randomized controlled trial" and 9.0 abstracts to read using hand-search. For 3 articles, the paper format of the journal reported only the Abstract, but it was fully reported in the electronic pages of the journal (*Pediatrics*). These reports were found by all methods. The hand-search and the Pubmed search using "publication type randomized controlled trial" missed the same single report, a randomized controlled trial reported in a letter to the editor. This trial was retrieved only by the Cochrane strategy.

The retrieval of articles for the remaining journal issues was performed using the Pubmed search using "publication type randomized controlled trial" because of the high sensitivity and efficiency of this method for retrieving articles fulfilling the inclusion/exclusion criteria.

Articles studied

A total of 299 issues from the 10 journals were published in 2002 and 403 articles reporting a randomized controlled trial were published in that year. All articles were all retrieved and evaluated. Depending on the length of the article, the evaluation took approximately 20 to 30 minutes per article. All the information previously described was obtained for each article, and there were no missing data.

Table 2 shows the number of articles retrieved from the different journals. There was a wide range of articles per journal (from 2 to 94).

Table 3 reports the basic characteristics of the articles. Approximately half of them were retrieved from journals that do not specify that they adhere to the CONSORT statement. More than three quarters of the reports came from journals with impact factors higher than 5.0. The affiliation of at least one author with an epidemiology or biostatistics department was difficult to evaluate in some journals, because the journals' editorial policy did not always provide the affiliated department of all authors; hence the estimate of 38.7±4.8% is conservative. Most of the studies were trying to show a difference between multiple treatment strategies (difference-seeking design), about half were industry-sponsored (totally or in combination with public funding) and multi-center studies. The number of participants per study ranged from 10 to 67 800 but was skewed to the right, as demonstrated by the large difference between the mean (1667) and the median (255).

Inter-rater evaluation

A total of 40 articles were randomly selected to be read independently by the two evaluators. Table 4 reports the kappa scores for inter-rater agreement. This agreement was very good, with kappa scores generally higher than 0.75. Only 2 of the 14 variables studied had a kappa score lower than 0.75 (management of missing data and management of improper inclusion). These two categorical variables had multiple possible answers. The low number of articles studied limited the precision for that high number of possible answers and the may explain the moderate inter-rater reliability. Moreover, these 2 variables were part of secondary objectives of our study.

Table 5 reports the intra-class correlations coefficient for the continuous measures. This correlation was very high (>0.98) for 6 of the 8 continuous measures. The 2 variables with lower intra-class correlation coefficient were "the number of subjects withdrawn before treatment" (ICC=0.01) and "the number of subjects excluded for non compliance" (ICC= 0.61). There were situations where both evaluators concluded that subjects were excluded from the analysis but disagreed on the sub-category of excluded subjects. In order to minimize the variability of the sub-categories of excluded subjects (improper inclusion, non-compliance, unclear), all excluded patients

were merged into a single category called "all excluded subjects". The intra-class correlation coefficient for this new variable was excellent (ICC= 0.9997).

Given that the pre-specified inter-rater correlation criteria were satisfied for the main variables of interest, I, alone evaluated the remaining articles.

Reporting of intention-to-treat approach

Of the 403 articles, 249 (61.8±4.7%) reported the use of an intention-to-treat analysis, 18 (4.2±2.0%) reported another approach, and the approach used was unclear or not mentioned in 134 (33.3±4.6%). Table 2 shows the high variability (from 27.3% to 83.6%) in the use of the intention-to-treat approach among the 10 journals. When reported, the method of analysis was mainly found in the Methods section. It was solely reported in Methods section in 132 articles, and was mentioned in the Methods and another section in 97 reports. It was reported only in the Abstract of 18 studies, only in the Results section of 15, and only in the Discussion of 1 article. The authors usually used the terms "intention-to-treat" or "intent-to-treat" and rarely used the term "analyzed as randomized". Few specifically stated that they did not use an intention-to-treat approach.

Use of intention-to-treat approach

Table 6 provides a description of the method of analysis used according to the evaluators in the 249 papers that reported the use of an intention-to-treat analysis. In 192 articles (77.1±5.2%), the subjects were clearly analyzed as randomized. In 23 reports (9.2±3.6%), the investigators described the utilization of a "*modified intention-to-treat*" analysis. In these studies, subjects were analyzed as randomized with some modifications in the intention-to-treat principle regarding exclusion or missing data. For example, in a study evaluating the effectiveness of caspofungin and amphotericin B for invasive candidiasis [74], the authors excluded all subjects who did not receive the first day of treatment and all those who failed to fulfill inclusion criteria after randomization: "*The modified intention-to-treat analysis (the primary analysis) included patients who had a documented diagnosis of invasive candidiasis and who received the study treatment for at least one day.*" The most commonly used definition of a "*modified intention-to-treat*" analysis was that all subjects that fulfilled the

inclusion criteria and received at least one dose of the study medication were analyzed in the treatment group to which they were assigned.

There were 17 articles (6.8±3.1%) that clearly violated the "analyzed as randomized" principle, even though they reported the use of an intention-to-treat approach. For example, in a study evaluating intranasal mupirocin to prevent postoperative Staphylococcus aureus infections[75] , the authors reported the exclusion after randomization of approximately 4% of the study population for improper inclusion or non-compliance. Of the 4030 subjects initially randomized, "*166 were excluded from the analysis, because they were not undergoing an eligible operation (49 in the mupirocin group and 48 in the placebo group), they received no study medication (22 and 26, respectively), or they met both exclusion criteria (8 and 13, respectively).*" The authors conclude that: "*3864 patients (95.9 percent) were included in the intention-to-treat analysis.*"

In another study regarding the clinical efficacy of homoeopathic of house dust mite (homoeopathic immunotherapy) in asthmatic people allergic to house dust mite[76], only 202 of the 242 randomized participants were analyzed. Seventeen were withdrawn because of a major protocol violation (oral steroids) that could have been secondary to clinical deterioration and 1 was excluded because of asthma exacerbation. One could say that they have been excluded from the study because of a worsening in their primary outcome that necessitated a more powerful treatment.

A last example concerns a study evaluating the efficacy of a structured half-day workshop format in improving subsequent review quality scores for reviewers of a medical journal[77]. It is reported in the Methods section that the authors performed 2 randomized trials and the analysis was based on intention-to-treat. In the first study, the investigators invited all 173 reviewers of a journal to attend the structured workshop. The intervention group was composed of the reviewers who attended the workshop while the control group was a sample of matched reviewers who did not attend. This is not true randomization and the analysis is more likely to be an "as treated" design. In the second study, 150 reviewers were randomized to either an intensive recruitment strategy to participate to the workshop or not to be recruited. Only the 11 subjects who participated to the workshop were analyzed in the

intervention group. They were compared to a random sample of 11 reviewers who were not invited to participate to the workshop. This design is more likely to be called a "per protocol" analysis.

The seventeen articles that violated the intention-to-treat principles are briefly described in Appendix II.

Finally, in 17 articles (6.8%±3.1%) the approach to analysis used was unclear. Most of these papers did not provide a flow diagram describing the subject pathway through the study (11 articles) or failed to state the number of analyzed subjects (9 articles). These articles probably violated the "analyzed as randomized" aspect of the principle, but it was difficult to conclude this with confidence based on the data reported.

Proportion and distribution of missing data

The proportion of missing data in articles that reported the use of an intention-to-treat analysis is presented in Table 7. The proportion could not be calculated for 9 articles because the number of randomized or analyzed subjects was unclear. Only 39.0% of articles had no missing data in their primary analysis. More than 20% of the articles had more than 10% of data missing. The three highest missing data rates encountered in the 249 articles were 58%, 56%, and 45%. One of these studies evaluated the effectiveness of inviting teenagers to general practice consultations to discuss health behaviour[78]. Of the 1488 teenagers initially randomized, only 659 (44%) provided answers for the primary outcome by responding to a mail questionnaire sent one year after the intervention of interest. All the non-responders were excluded from the analysis and there were few comments about the missing data in the discussion. This large missing data rate is an important limitation that could have biased the results.

In general, the main reason for missing data in the primary analysis was subjects lost to follow-up. Table 8 shows the distribution of articles according to their rate of subject loss to follow-up. Approximately 16% of the articles reported a rate of loss to follow-up higher than 10% of the number of randomized subjects. It was very difficult to differentiate loss to follow-up from withdrawals. Very few articles reported that

subjects who withdrew from the study treatment were still followed for outcome evaluation.

Table 9 shows articles according to their rate of missing data due to exclusion of subjects by investigators (for any reason). In more than 80% there were no exclusions reported. Table 9 also shows that in approximately 10%, the investigators excluded more than 1% of the study population from the analysis.

Strategies for missing data

A total of 152 articles claimed the use of an intention-to-treat analysis but had missing data. Table 10 shows that few of these articles reported a strategy for missing data in their primary analysis of the primary outcome. Subjects with missing data were clearly excluded from the primary analysis in 59% of the articles. The management of missing data was unclear in 17% of articles. In these, the authors did not mention the strategy used for missing data and the reviewer was unable to determine if the subjects were excluded because the number of subjects analyzed was not reported. In 18 of the 90 articles that used a categorical primary outcome, missing subjects were simply put in the denominator without taking into account the fact that the outcome was unknown. This represented 12% of the 152 articles. Finally, an imputation strategy was used in only 18 articles (12%). The most frequently used imputation strategy was a last observation carried forward (in 12 of the 62 articles that reported a continuous primary outcome). Only 5 studies performed a sensitivity analysis providing the spectrum of results according to multiple scenarios; all of these discussed the possible implications of their missing data.

Secondary analysis of the primary outcome was also examined regarding the analysis of missing data. Some articles reported multiple strategies. Although many articles used combinations of strategies, there were 103/152 (68%) articles that only used either the exclusion of the subject or did not state a strategy for missing data. The most frequently used imputation strategy was the last observation carried forward (23 articles).

Use of intention-to-treat without missing data

Figure 1 shows the distribution of articles accounting for the two components of the intention-to-treat principle together (missing data and "analyzed as randomized"). Of the 249 articles that reported the use of an intention-to-treat analysis, only 95 (38%) fulfilled both criteria and analyzed *all* subjects as randomized. In addition, 25 articles used an intention-to-treat analysis and provided an imputation strategy to account for missing data. More than 50% of the papers did not use an intention-to-treat approach or had missing data that were not taken into account in the analysis. Unsurprisingly, there was a strong positive correlation between the use of an intention-to-treat analysis and the absence of missing data in the analysis of our articles.

Management of improper inclusion

Tables 11 and 12 demonstrate the different management strategies for the analysis of subjects whose failure to fulfill the inclusion criteria was detected after randomization. For articles that reported the use of an intention-to-treat approach, we observed no mention of improper inclusion or reports of no improper inclusion in 75%. Subjects who failed to fulfill inclusion/exclusion criteria were excluded from the analysis in 16% and included in 8%. An analysis was performed with and without the improper inclusions in 3 articles (1%). The results for all articles (403 articles) were similar to the sub-group of articles that reported the use of intention-to-treat as shown in Table 12.

Discussion

Proportion of articles that reported the use of an intention-to-treat approach

We have shown that approximately 62% of the randomized controlled trials published in major medical journals during the year 2002 reported the use of an intention-to-treat approach for the analysis of their primary outcome.

This proportion may seem low, considering that the intention-to-treat approach is the generally recommended method for analysis in difference-seeking clinical trials[30]. A possible explanation for the low proportion is that most articles that did not report the use of an intention-to-treat analysis might have been equivalency studies. However, this was not the case: Of the 154 articles that did not report the use of an intention-to-treat analysis, 128 were for a difference-seeking study. Also, the proportions of articles that reported the use of an intention-to-treat analysis was similar for the group of all study articles (62%) and for the sub-group of difference-seeking studies (63%). Even though being a difference-seeking study was positively associated with the use of an intention-to-treat analysis (OR: 1.76; 95% CI: 1.03-3.02), the fact that more than 85% of the studies were difference-seeking dilutes the effect of study type on the proportion of articles that reported an intention-to-treat analysis.

Another possible explanation was that the articles that did not use an intention-to-treat analysis were cluster randomized trials. Again, this was not the case. Only 27 of the 403 articles reported cluster randomization. Of the 154 articles that did not report the use of an intention-to-treat analysis, only 13 used cluster randomization. The use of a cluster randomization was not a statistically significant predictor of the reporting of an intention-to-treat analysis (OR 0.63; 95% CI: 0.29-1.40).

To our knowledge, the proportion of articles reporting the use of an intention-to-treat analysis has been evaluated only once before. Hollis et al.[51] reported that 48 ±6% of the randomized controlled trials published in 4 main medical journals (BMJ, JAMA, Lancet, NEJM) in 1997 reported the use of an intention-to-treat analysis. Two differences between their study articles and ours might explain the difference in the

proportions (48% vs. 62%). The first difference is the year of publication. Papers in the Hollis et al. study were reported in 1997 compared to the year 2002 for our study. A revised version of the CONSORT statement was published in April 2001 [21-25]. One of the items implemented in the revised CONSORT statement concerns the method used for the analysis of the primary outcome; Authors are now asked to specify whether the analysis was by "intention-to-treat" or not (item #16 of the checklist)[22]. The rate of report of intention-to-treat may well have increased after revision of the CONSORT statement. This would correlate with previous studies that showed that the CONSORT statement was associated with an improvement in the reporting of randomized controlled trials[57]. However, this possibility should be evaluated in a longitudinal study.

The second difference between our articles and those of Hollis et al. is the type of journals evaluated. Their study was limited to four high-impact general medical journals, while ours examined 10 journals with a wider range of impact factors and included specialty journals. If we limit our evaluation to the 4 journals evaluated by Hollis et al., the reported use of an intention-to-treat analysis increases from 62% to 71% (201/283). This suggests that the differences between our results and those of Hollis et al. are not due to the different journal type.

One could hypothesize a difference in the reporting of an intention-to-treat analysis depending on the types of journals evaluated. Even though it was not the purpose of the present study, this hypothesis was evaluated by a sub-analysis. Table 13 shows that there is a statistically significant association between being a high-impact factor journal (OR 3.4; 95% CI: 2.1-5.4), adherence to CONSORT (OR 2.4; 95%CI: 1.6-3.6) or being a general journal (OR 3.6; 95% CI: 2.3-5.7) and the likelihood of reporting an intention-to-treat. A multivariate analysis including the variables being a general medicine journal, being a high-impact factor journal, and adherence to the CONSORT statement was conducted (Table 14). Being a general journal and adherence to CONSORT remained good predictors of the report of an intention-to-treat analysis (OR 2.8; 95% CI 1.2-6.9 and OR 1.8; 95%CI: 1.1-3.1) but there was no statistically significant association for the high-impact factor variable. However, the high correlation between these 3 variables may have confounded their relationship to the reporting of intention-to-treat.

Even though the use of intention-to-treat seems to be on the rise, only 62% of randomized controlled trials reported using it. Furthermore, 33% did not specify the method of analysis used. This is noteworthy considering the fact that approximately 60% of the articles were published in journals that adhere to the CONSORT statement. These journals are supposed to ask investigators to specify whether the analysis was by "intention-to-treat". The proportion of articles that reported an intention-to-treat approach was approximately 70±6% for articles published in journals adhering to the CONSORT statement.

The use of an intention-to-treat approach

Among the 249 articles that reported the use of an intention-to-treat analysis, 17 clearly violated the "analyzed as randomized" principle, 23 reported the use of a modified intention-to-treat approach, and in 17 the approach was unclear.

The articles that reported the use of a modified intention-to-treat analysis were analyzed separately. Some may view these articles as inappropriate uses of intention-to-treat analysis, and would say that they should have been counted as violating the "analyzed as randomized" principle because they excluded subjects after randomization. For example, Brown reported that the pharmaceutical industry commonly uses a different definition of intention-to-treat[79]. He notes, however, that:" *it is not a definition that is supported in regulatory guidelines. Section 5.2.1 of ICH E9 clearly advocates the inclusion of all randomized subjects in the analysis."* In our study, many articles that reported the use of a modified intention-to-treat analysis stated that they "*analyzed patients that received at least 1 dose of the allocated drug*". These articles would have been counted as violating the intention-to-treat principle had we used the same definitions as Hollis et al.[51] and would increase the number that violated the intention-to-treat principle from 17 to 32 (from 7 to 13% of articles that reported the use of intention-to-treat). These results can be compared to the results reported by Hollis et al.[51] who reported that of the 119 articles that reported the use of an intention-to-treat analysis, 12 (10%) excluded subjects who did not receive the treatment allocated and 3 were not analyzed as randomized.

Others would say that articles that reported a modified intention-to-treat analysis adhered to the intention-to-treat principle with regard to the "analyzed as randomized" aspect. Also, these articles provided a clearer definition of their analysis method and a better description of their exclusions. This greater clarity permitted the reader to evaluate the risk of bias engendered by the "modification". In some situations, a modified intention-to-treat design would decrease the risk of bias toward the null hypothesis for efficacy studies. For example, Friedlander et al. reported a study evaluating the efficacy of terbinafin in the treatment of tinea capitis[80]. There were 176 subjects enrolled and randomized on clinical suspicion of tinea capitis. The treatment was initially started based on clinical grounds. The final diagnosis was confirmed by culture 1 week after randomization in 159 subjects and excluded in 17 subjects. The investigators reported the cure rate on these 159 subjects in a modified intention-to-treat analysis. Including the 17 subjects that had a negative culture would have homogenized the results between the 2 treatment arms because it would have increased the rate of negative culture at week 12 in both treatment arms. This would have biased the results toward the null hypothesis and decreased the apparent efficacy of the intervention. Conversely, the inclusion of those 17 subjects would provide results that are more comparable to real-life situations, where physicians may prescribe the treatment on a clinical basis. This latter approach would have provided the result for an effectiveness evaluation.

The main objective of our study was to evaluate whether authors who reported an intention-to-treat analysis really did what they purported to do. The use of a modified intention-to-treat analysis violated some basic aspects of intention-to-treat but, at least, alerts the reader to that fact.

At least 17 articles that reported the use of an intention-to-treat analysis violated one of its basic principles. The main reason for violation was the exclusion of some subjects because they did not receive the completed allocated treatment. Some studies excluded only a few subjects for reasons that are not in accord with the generally accepted definition of an intention-to-treat analysis (all subjects are analyzed as randomized). For example, Eichenfield et al.[81] assessed the efficacy of topical anaesthesia during venipuncture procedures. Only 1 of the 120 randomized subjects was excluded because of a protocol deviation (the cream was applied for only 5

minutes instead of 60 minutes). It has previously been demonstrated that the exclusion of non-compliant subjects increases the risk of bias[37;38]. Other studies excluded as many as 10% of their randomized subjects for diverse reasons (withdrawal between randomization and intervention or refusal to consent after randomization). It is possible that investigators who claimed the use of an intention-to-treat analysis but excluded 10% of the subjects for other reasons used a different definition of intention-to-treat: AVAILABLE subjects are analyzed as randomized. This is not, however, the definition provided by most authorities in the field (CONSORT statement[17], Cochrane collaboration[3], Last's Dictionary of epidemiology[27], etc). A single study that claimed to be a randomized controlled trial using an intention-to-treat approach[77] (described in the Result section) clearly used an as-treated analysis. It should not have categorized this an intention-to-treat analysis because subjects were not even randomized.

The percentage of violations was lower in our study than in the one reported by Hollis et al.[51](7% vs. 13%). Several explanations could account for this difference. First, as already mentioned, we separated the articles reporting the use of a modified intention-to-treat analysis and the rate of violation would have been higher had we not done so. Second, the intention-to-treat approach may have been used more appropriately in our articles than in the papers of the previous study. When considering only the articles coming from the 4 journals in the study by Hollis et al.[51], the rate of violation was lower than for articles coming from our 10 journals (11/201 articles, 5.5%).

A univarate analysis of the data showed that being a general journal (OR 2.5; 95 CI 1.1-5.7) and being a high-impact factor journal (OR 2.6; 95 CI 1.1-5.9) were predictors of the appropriate use of intention-to-treat. However, being a journal that adhered to the CONSORT statement (OR 1.3; 95 CI 0.6-2.7) was less often associated with appropriate use of intention-to-treat (Table 15). In a multivariate analysis, none of these three factors was statistically significant (Table16).

Based on these results, we suggest that the lower rate of inadequately applied intention-to-treat analysis reported in our study compared to the results reported by Hollis et al.[51] is probably not related to journal characteristics. Another possible explanation may be the difference in the year of publication of the two study articles (1997 vs. 2002). Intention-to-treat is an evolving method that received much publicity in recent years, as demonstrated by its inclusion in the revised CONSORT statement in 2001. It is quite possible that investigators have improved their knowledge and utilization of intention-to-treat over the past few years. A longitudinal study would be the best approach to address this question.

In 17 articles, the analytic approach was unclear. Some, or many, of these articles may have violated a basic aspect of intention-to-treat. This could, at worst, double the percentage of articles that violated a basic aspect of intention-to-treat. Most of these articles reported a continuous primary outcome and failed to report the number of analyzed subjects or failed to describe the reasons for missing data. Also, the absence of a flow diagram in many of these articles limited the ability to describe the subjects excluded from the analysis. Some subjects may have been excluded from the primary analysis because they were not compliant, but we could not identify them because they were not described in a flow diagram or in the Results section. At best, all these articles appropriately used intention-to-treat and this would have no influence on our results.

When evaluating the results without taking into account the effect of missing data, we can conclude that between 7% to 23% of the articles violated a basic aspect of intention-to-treat. The difference depends on how one treats to the 17 unclear studies and to the 23 studies using a modified intention-to-treat approach.

Missing data

As one would expect, missing data were a major problem in the application of the intention-to-treat analysis in our articles. Only 39% of the articles that reported the use of intention-to-treat analysis had no missing data in their primary analysis and more than 20% had a missing data rate higher than 10%. These results are similar to those reported by Kruse et al.[52]. In their study, 42% of the articles had no missing subjects

and 11% excluded more than 10% of their randomized subjects. In the Hollis et al. study[51], 25% of their articles reported no missing data and 24% had a missing data rate greater than 10%. The strictest definition of the intention-to-treat analysis is that ALL subjects be included in the final analysis, regardless of the treatment they received. In more than 60% of our articles some subjects were not included. This was because of missing data (e.g., subjects lost to follow-up) or from exclusion of subjects (examples: improper inclusion, non-compliance). These two factors are associated with an increase risk of bias[82]. The high proportion (20%) of articles that had a rate of missing data greater than 10% is of concern, as exclusions of this magnitude have a greater potential to bias the findings[83].

The main reason for missing data was subjects lost to follow-up (95 articles). This occurred in 38% of articles that reported intention-to-treat or 67% of the articles with intention-to-treat that had missing data. It is usually not reasonable to assume that subjects lost to follow-up have the same risk of primary outcome than the others. These subjects may lead to bias if they were not lost randomly.

The high proportion of studies with exclusion of subjects is of concern. The investigators had information about these subjects that was not included in the analysis for different reasons (improper inclusion, non-compliant, etc). Exclusion of subjects should not occur in an intention-to-treat analysis. Such excluded subjects represent a sub-group of the study population that may have a different risk in achieving the primary outcome. Many articles usually excluded a small proportion of their study subjects. Some investigators argued that these exclusions are not of concern because their small numbers would probably have no impact on the conclusion of the study. Conversely, one could argue that if it has no effect, there is no problem of including them in the analysis and if it has an impact, it would be better to know it and to include them in the analysis.

Few studies reported a strategy for analysis of missing data in their primary analysis. Of the 152 articles with missing data, only 18 used an imputation strategy in their primary analysis and 49 in their secondary analysis. These imputation strategies can improve the quality of the analysis, but they all carry a risk for potential bias[84]. The low use of a management strategy is of concern. There are multiple possible

explanations for this low use. First, many authors reported that most imputation strategies may lead to bias[85-87]. For example, Engels and Diehr evaluated 14 imputation strategies for missing data. They identified situations where subjects had known data in a longitudinal study and treated them as unknown using different imputation strategies. They reported that "*Most imputation methods were biased toward estimating the "missing value" as too healthy, and most estimates had a variance that was too low*".

Even though no imputation strategy can be exempt from bias, a sensitivity analysis allows one to evaluate the spectrum of results while taking the effect of the missing data into account. However, this approach weakens the conclusion that authors can address, because it may increase the spectrum of final results for the study. In many studies, this could modify the conclusion. Another possible explanation for the lack of a management strategy for missing data may be the researcher's ignorance about the possible bias engendered by missing data or about the imputation strategies available.

Improper inclusions

The management of subjects that were found to have failed to satisfy inclusion/exclusion criteria after being randomized is controversial. Some authors suggest including all randomized subjects in the analysis while others suggest that the exclusion of these subjects has little potential for bias. Fergusson et al.[34] recommend including all subjects in the analysis with few exceptions. They suggested that it may be legitimate to exclude subjects already randomized when study personnel make a mistake in the implementation of the entry criteria, or when the subject never received the treatment because of an evolution of his clinical condition. The exclusions should be performed by an independent, blinded, adjudication committee.

Our results show the same diversity with regard to the management of subjects that failed to fulfill the inclusion/exclusion criteria after randomization. Nearly 25% of our articles randomized ineligible patients. The most common approach used for such subjects was their exclusion from analysis. Whether this approach was the most appropriate in order to limit bias and optimize power was not evaluated. Few studies performed an analysis with and without the subjects. Such studies could be considered

the most useful for readers because they provide the spectrum of results engendered by the improper inclusions. However, they also make explicit the increase the uncertainty resulting from the randomization of ineligible patients.

Limitations

Most of the articles were evaluated by a single evaluator and this could decrease the study's external validity and certainly affect its perceived external validity. In order to increase external validity, a detailed and clear standardized evaluation form was developed. Also, the first 40 articles were read by two evaluators and inter-rater reliability was assessed. The high concordance obtained between the two raters suggests that the assessment of the articles was consistent and standard. Good internal validity is the first step toward good external validity. The fact that our final results are comparable with the previous literature increases our confidence about the external validity of the study.

One of the primary outcomes, the use of intention-to-treat, was based on the subjective evaluation of the rater. Appendix III provides a brief description of all 17 articles that claimed to have used an intention-to-treat approach but were assessed as not doing so. It provides comments describing the basic aspect of intention-to-treat that was violated for each of these articles. This should help readers to draw their own conclusions regarding whether the articles really violated the intention-to-treat principles and limits the number of false-positives. However, there is the possibility that some articles that violated the intention-to-treat principle were missed (false negatives) and not included in the study.

Finally, there could be a difference between what was reported and the way the analysis was really performed. Other than discussing this with the investigators, it would be difficult to evaluate the concordance between the method reported in the article and Method actually used in the trial.

Conclusions

More than 60% of the 403 randomized controlled trials published in our sample of medical journals in 2002 reported the use of an intention-to-treat approach. Only 39% of these 249 articles actually analyzed all subjects as randomized. This low percentage is mainly due to the high proportion of articles with missing data in their primary analysis. There was 61% of articles that "analyzed as randomized" and had less than 5% of missing data

These results emphasize the fact that authors use the label "intention-to-treat" quite differently. While many authorities define intention-to-treat as the analysis of ALL subjects as randomized, the most common use of the term (in published reports of randomized controlled trials) refers to the analysis of all AVAILABLE subjects as randomized (i.e., without taking the missing data aspect into account). There are two solutions to this problem. The first is to change the definition of intention-to-treat to define a method of analysis in which available subjects are analyzed as randomized. This definition opens the door to major bias engendered by differential rate of exclusion for compliant and non-compliant subjects that are lost to follow-up. The second option is to preserve the classical definition and to name an analysis where there is missing data as a *modified intention-to-treat analysis*. Investigators using the latter option should define the modification used in their Method section. In either case, a greater transparency and completeness would permit a better analysis of the results of a randomized controlled trial.

Figure 1.Distribution of articles according to "analyzed as randomized" status and missing data status.

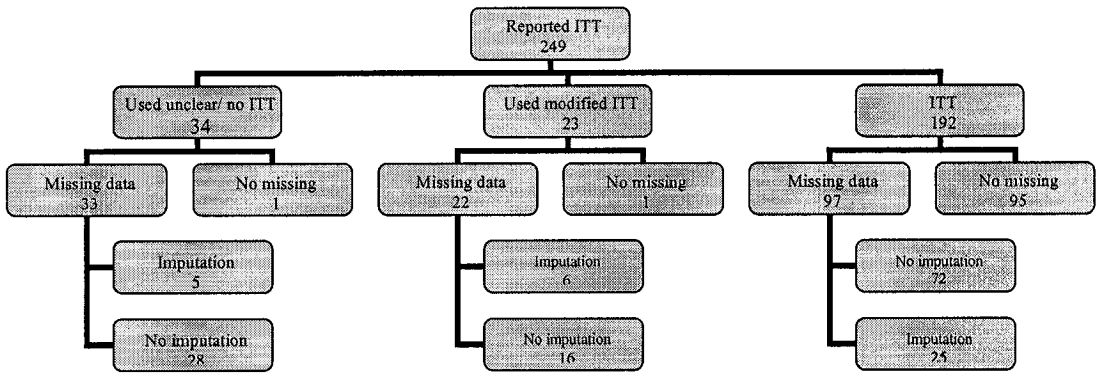

Table 1 Sensitivity and positive predictive value of each search strategy

Method	Number of articles evaluated	# of articles fulfilling the inclusion criteria	Sensitivity (95% CI)	Number-needed-to-read (95% CI)
Hand-search	334	37	0.97 (0.92-1.00)	9.0 (6.9 to 13.0)
Pubmed Publication Type: RCT	40	37	0.97 (0.92 to 1.00)	1.1 (0.99 to 1.2)
Pubmed Cochrane strategy	308	38	1.00 (0.97 to 1.00)	8.1 (6.2 to 11.5)

RCT: Randomized controlled trials, CI: confidence interval

Table 2 Number of RCTs per journal and reported method of analysis among the 10 journals (n=403).

		Proportion of articles that reported a method of analysis				
Journals	# of RCT	Intention-to-treat %±2sd (n)	Per protocol %±2sd (n)	As treated %±2sd (n)	Unclear %±2sd (n)	No mention %±2sd (n)
NEJM	74	60.8±11.1 (45)	2.7±3.7 (2)	0 (0)	0 (0)	36.4±10.1 (27)
BMJ	54	57.4 ±13.2 (31)	1.9±3.6 (1)	0	0	40.7±13.1 (22)
Lancet	94	78.8±8.3 (74)	11.7±6.5 (11)	0	0	9.6±5.9 (9)
JAMA	61	83.6±9.2 (51)	0	0	1.6±3.2 (1)	14.8±8.9 (9)
CMAJ	7	42.8±37.4 (3)	0	0	0	57.2±36.7 (4)
Annals Int Med	19	42.1±22.1 (8)	5.3±10.0 (1)	5.3±10.0 (1)	0	47.4±22.5 (9)
Ann Emerg Med	11	54.5±29.2 (6)	0	0	0	45.5±22.5 (5)
Ped Emerg Care	2	0	0	0	0	100 (2)
J Peds	33	27.3±15.0 (9)	3.3±5.8 (1)	0	3.3±5.8 (1)	66.7±16.1 (22)
Pediatrics	48	45.8±1.4 (22)	2.1±4.0 (1)	0	0	52.1±14.1 (25)
Total	403	61.8 ±4.7 (249)	4.2±2.0 (17)	0.3±0.5 (1)	0.5±0.7 (2)	33.3±4.6 (134)

Table 3. Characteristics of the articles (n=403).

Categorical variables	Proportion (%)
Published in a specialized journal	113/403 (28.0±4.4)
Published in a journal adherent to the CONSORT statement	246/403 (61.0±4.8)
Published in a journal with high impact factor	302/403 (74.9±4.2)
Affiliation of an author with an epidemiology/biostatistic department	156/403 (38.7±4.8)
Industry funded study	173/403 (42.9±4.8)
Multi-center study	263/403 (65.3±4.6)
Presence of a placebo	156/403 (38.7±4.8)
Group randomization	27/403 (6.7±2.4)
Presence of a flow diagram	239/403 (59.3±4.8)
Difference seeking design	346/403 (85.6±3.4)

Continuous variables	Mean (SD)	Median	1st quartile	3rd quartile
Number of randomized subjects	1667 (5658)	255	108	1004
Number of authors	8.7 (5.6)	7.0	5	11

Table 4 Inter-rater agreement (kappa score) for categorical measures

Variables	Simple kappa (95% CI)
Individual or group randomization	1.0 (1.00-1.00)
Individual or group analysis	1.0 (1.00-1.00)
Method of analysis reported	1.0 (1.00-1.00)
Method of analysis used (according to the rater)	0.91 (0.79-1.03)
Management of missing data	0.65 (0.46-0.83)
Management of improper inclusion	0.47 (0.29-0.64)
Presence of a flow diagram	1.00 (1.00-1.00)
Affiliation with an epidemiology/statistics department	0.79 (0.60-0.98)
Funding source	0.76 (0.58-0.93)
Multi or single center	0.95 (0.86-1.01)
Type of outcome	0.75 (0.57-0.93)
Use of a placebo	0.94 (0.85-1.05)
Study design	0.83 (0.60-1.06)
Difference-seeking study	0.94 (0.82-1.06)

Table 5 Inter-rater agreement (intra-class correlation coefficient) for continuous measures

Variable	Coefficient
Number of randomized subjects	0.9995
Number of withdrawals before treatment	0.0077
Number of losses to follow-up	0.9850
Number of exclusions for improper inclusion	1.0000
Number of exclusions for non compliance	0.6087
Number of exclusions for other or unclear reasons	0.9999
Number of "all excluded"	0.9997
Number of analyzed subjects	0.9999
Number of authors	1.0000

Table 6 Evaluator assessed method of analysis for articles reporting an intention-to-treat analysis (n=249)

Method	# of RCT (%±2 SD)
Intention-to-treat	192 (77.1±5.2)
Modified intention-to-treat	23 (9.2±3.6)
Per protocol	17 (6.8±3.1)
Unclear	17 (6.8±3.1)

Table 7 Proportion of missing data among articles that reported the use of an ITT approach (n=249)

Missing data proportion	Number of articles (%±2SD)
Unclear	9 (3.6±2.3)
No missing data	97 (39.0±5.9)
0-1%	27 (10.8±3.8)
1-5%	41 (16.5±5.6)
5-10%	23 (9.2±3.6)
>10%	52 (20.9±5.1)

Table 8. Proportion of losses to follow-up among articles that reported the use of an ITT approach (n=249)

Lost to follow-up proportion	Number of articles (%±2SD)
No loss to follow-up	154 (61.9±6.0)
0-1%	14 (5.7±2.9)
1-2%	9 (3.6±2.3)
2-5%	17 (6.9±3.1)
5-10%	16 (6.5±3.0)
> 10%	39 (15.7±4.5)

Table 9 Proportion of exclusions (all treatment arms combined) among articles that reported the use of an ITT approach (n=249)

Proportion of exclusion	Number of articles (%±2SD)
No exclusion	202 (81.1±4.8)
0-1%	15 (6.0±2.8)
1-2 %	7 (2.8±2.0)
2-5%	11 (4.4±2.5)
5-10%	8 (3.2±2.3)
>10%	6 (2.4±1.9)

Table 10 Strategy used for missing data in the PRIMARY analysis for articles that reported an intention-to-treat analysis (n=249).

Management strategy	Number of reports (%±2SD)
No missing data	97 (39.4±6.1)
Exclude subjects with missing data	89 (58.6±8.1)
Keep the subjects in the denominator	18 (11.8±5.2)
Use of worst case/ best case scenario	2 (1.3±1.8)
Last observation carried forward	12 (7.9±4.3)
Regression	0
Replace by a mean	0
Multiple imputation	1 (0.7±1.2)
Other	3 (2.0±2.2)
Unclear	27 (17.8±6.1)

Table 11 Management of subjects that failed inclusion/exclusion criteria after randomization in articles that reported the use of an ITT analysis (n=249)

Management	Number of article (%±2SD)
No report or no improper inclusion	186 (74.7±5.4)
Include subjects in the analysis	20 (8.0±3.4)
Exclude subjects from analysis	40 (16.1±4.5)
Perform analysis with and without the subjects	3 (1.2±1.4)

Table 12 Management of subjects who failed inclusion/exclusion criteria after randomization for all the articles (n=403)

Management	Number of article (%±2SD)
No report or no improper inclusion	307 (76.2±4.1)
Include subjects in the analysis	24 (6.0±2.3)
Exclude subjects from analysis	69 (17.1±3.7)
Perform analysis with and without the subjects	3 (0.7±0.9)

Table 13. Univariate analysis for predictors of reported use of intention-to-treat (n=403)

Factors	Odds ratio (95% CI)
Being a general journal	3.6 (2.3-5.7)
Being a high impact journal	3.4 (2.1-5.4)
Being a journal adhering to CONSORT	2.4 (1.6-3.6)
Use of group randomization	0.6 (0.3-1.4)
Affiliation of an author with an epidemiology/biostatatistics department	1.9 (1.3-3.0)
Use of a placebo	1.2 (0.8-1.9)
Difference-seeking study	1.8 (1.0-2.5)
Industry-sponsored study	1.9 (1.2-2.8)
Multi-center study	2.0 (1.3-3.1)
Study design Matched Parallel Factorial	 1.0 (ref) 4.9 (1.9-12.7) 10.8 (2.8-41.9)
Nature of primary outcome Continuous Categorical Survival	 1.0 1.8 (1.2-2.8) 4.1 (2.1-8.1)
Number of authors	1.12 (1.06-1.18)

Table 14.Multi-variate analysis for predictors of reported use of intention-to-treat (n=403)

Factors	Odds ratio (95% CI)
Being a general journal	2.8 (1.2-7.0)
Being a high-impact journal	0.7 (0.3-2.0)
Being a journal adhering to CONSORT	1.8 (1.1-3.1)
Affiliation of an author with an epidemiology/biostatistics department	1.6 (0.97-2.7)
Use of a placebo	1.2 (0.8-2.2)
Difference-seeking study	1.1 (0.6-2.1)
Industry-sponsored study	1.7 (1.0-2.7)
Multi-center study	1.0 (0.6-1.7)
Number of authors	1.1 (1.0-1.2)

Table 15.Univariate analysis for predictors of appropriate use of intention-to-treat (n=249)

Factors	Odds ratio (95% CI)
Being a general journal	2.5 (1.1-5.7)
Being a high-impact journal	2.6 (1.1-5.9)
Being a journal adhering to CONSORT	1.3 (0.6-2.7)
Use of group randomization	0.13 (0.04-0.40)
Affiliation of an author with an epidemiology/biostatistics department	1.2 (0.6-1.4)
Use of a placebo	1.5 (0.7-3.2)
Difference-seeking study	0.6 (0.2-2.3)
Industry-sponsored study	1.6 (0.6-3.8)
Multi-center study	1.4 (0.6-3.0)
Study design Matched Parallel Factorial	 1.0 (ref) 3.1 (0.6-17.8) 8.5 (0.6-11.8)
Nature of primary outcome Continuous Categorical Survival	 1.0 3.7 (1.7-8.1) 99.9 (0.0-999.9)
Number of authors	1.1 (1.0-1.3)

Table 16.Multi-variate analysis for predictors of appropriate use of intention-to-treat (n=249)

Factors	Odds ratio (95% CI)
Being a general journal	2.8 (0.4-19.6)
Being a high-impact journal	1.0 (0.1-7.8)
Being a journal adhering to CONSORT	1.3 (0.4-3.7)
Use of group randomization	0.1 (0.01-0.3)
Affiliation of an author with an epidemiology/biostatistics department	0.9 (0.4-2.1)
Difference-seeking study	0.4 (0.1-1.5)
Use of a placebo	1.8 (0.7-4.9)
Industry-sponsored study	0.7 (0.2-2.0)
Multi-center study	0.6 (0.2-1.8)
Number of authors	1.1 (1.0-1.2)

Reference List

1. Brighton B, Bhandari M, Tornetta P, III, Felson DT. Hierarchy of evidence: from case reports to randomized controlled trials. *Clin.Orthop.* 2003;19-24.

2. Devereaux PJ,.Yusuf S. The evolution of the randomized controlled trial and its role in evidence-based decision making. *J.Intern.Med.* 2003;**254**:105-13.

3. Cochrane Collaboration. Cochrane Controlled Trials Register. http://www.ovid.com/site/products/ovidguide/cctrdb.htm . 2002.
Ref Type: Electronic Citation

4. Schulz KF. Randomized controlled trials. *Clin.Obstet.Gynecol.* 1998;**41**:245-56.

5. Greenhouse JB,.Meyer MM. A note on randomization and selection bias in maintenance therapy clinical trials. *Psychopharmacol.Bull.* 1991;**27**:225-9.

6. Armitage P. The role of randomization in clinical trials. *Stat.Med.* 1982;**1**:345-52.

7. Junker CA. Adherence to published standards of reporting: a comparison of placebo-controlled trials published in English or German. *JAMA* 1998;**280**:247-9.

8. Adetugbo K,.Williams H. How well are randomized controlled trials reported in the dermatology literature? *Arch.Dermatol.* 2000;**136**:381-5.

9. Thornley B,.Adams C. Content and quality of 2000 controlled trials in schizophrenia over 50 years. *BMJ* 1998;**317**:1181-4.

10. Hotopf M, Lewis G, Normand C. Putting trials on trial--the costs and consequences of small trials in depression: a systematic review of methodology. *J.Epidemiol.Community Health* 1997;**51**:354-8.

11. Dickinson K, Bunn F, Wentz R, Edwards P, Roberts I. Size and quality of randomised controlled trials in head injury: review of published studies. *BMJ* 2000;**320**:1308-11.

12. Ruiz-Canela M, Martinez-Gonzalez MA, Irala-Estevez J. Intention to treat analysis is related to methodological quality. *BMJ* 2000;**320**:1007-8.

13. DerSimonian R, Charette LJ, McPeek B, Mosteller F. Reporting on methods in clinical trials. *N.Engl.J.Med.* 1982;**306**:1332-7.

14. Moher D, Pham B, Jones A, Cook DJ, Jadad AR, Moher M *et al.* Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet* 1998;**352**:609-13.

15. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995;**273**:408-12.

16. Juni P, Altman DG, Egger M. Systematic reviews in health care: Assessing the quality of controlled clinical trials. *BMJ* 2001;**323**:42-6.

17. Altman DG. Better reporting of randomised controlled trials: the CONSORT statement. *BMJ* 1996;**313**:570-1.

18. Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I *et al.* Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA* 1996;**276**:637-9.

19. Huston P,.Hoey J. CMAJ endorses the CONSORT statement. CONsolidation of Standards for Reporting Trials. *CMAJ.* 1996;**155**:1277-82.

20. Rennie D. How to report randomized controlled trials. The CONSORT statement. *JAMA* 1996;**276**:649.

21. Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D *et al.* The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann.Intern.Med.* 2001;**134**:663-94.

22. Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet* 2001;**357**:1191-4.

23. Moher D, Schulz KF, Altman D. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *JAMA* 2001;**285**:1987-91.

24. Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *Ann.Intern.Med.* 2001;**134**:657-62.

25. Shapiro S. The revised CONSORT statement: honing the cutting edge of the randomized controlled trial. *CMAJ.* 2001;**164**:1157-8.

26. Bradford Hill A. Principles of Medical Statistics. 1961.

27. Last J.M. A Dictionary of Epidemiology. Oxford University press, 2001.

28. Newell DJ. Intention-to-treat analysis: implications for quantitative and qualitative research. *Int.J.Epidemiol.* 1992;**21**:837-41.

29. Ware JH. Interpreting incomplete data in studies of diet and weight loss. *N.Engl.J.Med.* 2003;**348**:2136-7.

30. CONSORT-statement. www.consort-statement.org . 2004.
Ref Type: Electronic Citation

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

32. Lee YJ, Ellenberg JH, Hirtz DG, Nelson KB. Analysis of clinical trials by treatment actually received: is it really an option? *Stat.Med.* 1991;**10**:1595-605.

33. Montori VM,.Guyatt GH. Intention-to-treat principle. *CMAJ.* 2001;**165**:1339-41.

34. Fergusson D, Aaron SD, Guyatt G, Hebert P. Post-randomisation exclusions: the intention to treat principle and excluding patients from analysis. *BMJ* 2002;**325**:652-4.

35. Peto R, Pike MC, Armitage P, Breslow NE, Cox DR, Howard SV *et al.* Design and analysis of randomized clinical trials requiring prolonged observation of each patient. I. Introduction and design. *Br.J.Cancer* 1976;**34**:585-612.

36. Sackett DL,.Gent M. Controversy in counting and attributing events in clinical trials. *N.Engl.J.Med.* 1979;**301**:1410-2.

37. Peduzzi P, Detre K, Wittes J, Holford T. Intent-to-treat analysis and the problem of crossovers. An example from the Veterans Administration coronary bypass surgery study. *J.Thorac.Cardiovasc.Surg.* 1991;**101**:481-7.

38. Peduzzi P, Wittes J, Detre K, Holford T. Analysis as-randomized and the problem of non-adherence: an example from the Veterans Affairs Randomized Trial of Coronary Artery Bypass Surgery. *Stat.Med.* 1993;**12**:1185-95.

39. Influence of adherence to treatment and response of cholesterol on mortality in the coronary drug project. *N.Engl.J.Med.* 1980;**303**:1038-41.

40. May GS, DeMets DL, Friedman LM, Furberg C, Passamani E. The randomized clinical trial: bias in analysis. *Circulation* 1981;**64**:669-73.

41. Nagelkerke N, Fidler V, Bernsen R, Borgdorff M. Estimating treatment effects in randomized clinical trials in the presence of non-compliance. *Stat.Med.* 2000;**19**:1849-64.

42. Fields WS, Maslenikov V, Meyer JS, Hass WK, Remington RD, Macdonald M. Joint study of extracranial arterial occlusion. V. Progress report of prognosis following surgery or nonsurgical treatment for transient cerebral ischemic attacks and cervical carotid artery lesions. *JAMA* 1970;**211**:1993-2003.

43. Farwell JR, Lee YJ, Hirtz DG, Sulzbacher SI, Ellenberg JH, Nelson KB. Phenobarbital for febrile seizures--effects on intelligence and on seizure recurrence. *N.Engl.J.Med.* 1990;**322**:364-9.

44. Gaston KE, Moore DT, Pruthi RS. Hand-assisted laparoscopic nephrectomy: prospective evaluation of the learning curve. *J.Urol.* 2004;**171**:63-7.

45. Dagash H, Chowdhury M, Pierro A. When can I be proficient in laparoscopic surgery? A systematic review of the evidence. *J.Pediatr.Surg.* 2003;**38**:720-4.

46. Taipale P, Ammala M, Salonen R, Hiilesmaa V. Learning curve in ultrasonographic screening for selected fetal structural anomalies in early pregnancy. *Obstet.Gynecol.* 2003;**101**:273-8.

47. WRIGHT IS, BECK DF, MARPLE CD. Myocardial infarction and its treatment with anticoagulants; summary of findings in 1031 cases. *Lancet* 1954;**266**:92-5.

48. Stewart LA,.Parmar MK. Bias in the analysis and reporting of randomized controlled trials. *Int.J.Technol.Assess.Health Care* 1996;**12**:264-75.

49. Simpson JM, Klar N, Donnor A. Accounting for cluster randomization: a review of primary prevention trials, 1990 through 1993. *Am.J.Public Health* 1995;**85**:1378-83.

50. Baumgardner KR. A review of key research design and statistical analysis issues. *Oral Surg.Oral Med.Oral Pathol.Oral Radiol.Endod.* 1997;**84**:550-6.

51. Hollis S,.Campbell F. What is meant by intention to treat analysis? Survey of published randomised controlled trials. *BMJ* 1999;**319**:670-4.

52. Kruse RL, Alper BS, Reust C, Stevermer JJ, Shannon S, Williams RH. Intention-to-treat analysis: who is in? Who is out? *J.Fam.Pract.* 2002;**51**:969-71.

53. Lee KP, Schotland M, Bacchetti P, Bero LA. Association of journal quality indicators with methodological quality of clinical research articles. *JAMA* 2002;**287**:2805-8.

54. Delgado-Rodriguez M, Ruiz-Canela M, Irala-Estevez J, Llorca J, Martinez-Gonzalez A. Participation of epidemiologists and/or biostatisticians and methodological quality of published controlled clinical trials. *J.Epidemiol.Community Health* 2001;**55**:569-72.

55. Bhandari M, Richards RR, Sprague S, Schemitsch EH. The quality of reporting of randomized trials in the Journal of Bone and Joint Surgery from 1988 through 2000. *J Bone Joint Surg.Am.* 2002;**84-A**:388-96.

56. Clifford TJ, Barrowman NJ, Moher D. Funding source, trial outcome and reporting quality: are they related? Results of a pilot study. *BMC.Health Serv.Res.* 2002;**2**:18.

57. Moher D, Jones A, Lepage L. Use of the CONSORT statement and quality of reports of randomized trials: a comparative before-and-after evaluation. *JAMA* 2001;**285**:1992-5.

58. ISI web of knowledge. http://isiwebofknowledge.com/ . 2003.
Ref Type: Internet Communication

59. Scherer RW, Dickersin K, Langenberg P. Full publication of results initially presented in abstracts. A meta-analysis. *JAMA* 1994;**272**:158-62.

60. Bhandari M, Devereaux PJ, Guyatt GH, Cook DJ, Swiontkowski MF, Sprague S *et al.* An observational study of orthopaedic abstracts and subsequent full-text publications. *J Bone Joint Surg.Am.* 2002;**84-A**:615-21.

61. Callaham M, Wears RL, Weber E. Journal prestige, publication bias, and other characteristics associated with citation of published studies in peer-reviewed journals. *JAMA* 2002;**287**:2847-50.

62. Langham J, Thompson E, Rowan K. Identification of randomized controlled trials from the emergency medicine literature: comparison of hand searching versus MEDLINE searching. *Ann.Emerg.Med.* 1999;**34**:25-34.

63. Hopewell S, Clarke M, Lusher A, Lefebvre C, Westby M. A comparison of handsearching versus MEDLINE searching to identify reports of randomized controlled trials. *Stat.Med.* 2002;**21**:1625-34.

64. Suarez-Almazor ME, Belseck E, Homik J, Dorgan M, Ramos-Remus C. Identifying clinical trials in the medical literature with electronic databases: MEDLINE alone is not enough. *Control Clin.Trials* 2000;**21**:476-87.

65. Dickersin K, Manheimer E, Wieland S, Robinson KA, Lefebvre C, McDonald S. Development of the Cochrane Collaboration's CENTRAL Register of controlled clinical trials. *Eval.Health Prof.* 2002;**25**:38-64.

66. Robinson KA,.Dickersin K. Development of a highly sensitive search strategy for the retrieval of reports of controlled trials using PubMed. *Int.J Epidemiol.* 2002;**31**:150-3.

67. Dickersin K, Scherer R, Lefebvre C. Identifying relevant studies for systematic reviews. *BMJ* 1994;**309**:1286-91.

68. Bachmann LM, Coray R, Estermann P, Ter Riet G. Identifying diagnostic studies in MEDLINE: reducing the number needed to read. *J Am.Med.Inform.Assoc.* 2002;**9**:653-8.

69. Cohen, J. A coefficient of agreement for nominal scales. Educational and Psych.Meas. 20, 37-46. 1960.

Ref Type: Generic

70. Shrout PE,.Fleiss JL. Intraclass Correlations: Uses in Assessing Rater Reliabilit. *Psychological Bulletin* 1979;**2**:420-8.

71. Bravo G,.Potvin L. Estimating the reliability of continuous measures with Cronbach's alpha or the intraclass correlation coefficient: toward the integration of two traditions. *J Clin Epidemiol.* 1991;**44**:381-90.

72. Muller R,.Buttner P. A critical discussion of intraclass correlation coefficients. *Stat.Med.* 1994;**13**:2465-76.

73. Landis JR,.Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;**33**:159-74.

74. Mora-Duarte J, Betts R, Rotstein C, Colombo AL, Thompson-Moya L, Smietana J *et al.* Comparison of caspofungin and amphotericin B for invasive candidiasis. *N.Engl.J.Med.* 2002;**347**:2020-9.

75. Perl TM, Cullen JJ, Wenzel RP, Zimmerman MB, Pfaller MA, Sheppard D *et al.* Intranasal mupirocin to prevent postoperative Staphylococcus aureus infections. *N.Engl.J.Med.* 2002;**346**:1871-7.

76. Lewith GT, Watkins AD, Hyland ME, Shaw S, Broomfield JA, Dolan G *et al.* Use of ultramolecular potencies of allergen to treat asthmatic people allergic to house dust mite: double blind randomised controlled clinical trial. *BMJ* 2002;**324**:520.

77. Callaham ML,.Schriger DL. Effect of structured workshop training on subsequent performance of journal peer reviewers. *Ann.Emerg.Med.* 2002;**40**:323-8.

78. Walker Z, Townsend J, Oakley L, Donovan C, Smith H, Hurst Z *et al.* Health promotion for adolescents in primary care: randomised controlled trial. *BMJ* 2002;**325**:524.

79. Brown DJ. ICH E9 guideline 'Statistical principles for clinical trials':a case study Response to A. Phillips and V. Haudiquet. *Stat.Med.* 2003;**22**:13-7.

80. Friedlander SF, Aly R, Krafchik B, Blumer J, Honig P, Stewart D *et al.* Terbinafine in the treatment of Trichophyton tinea capitis: a randomized, double-blind, parallel-group, duration-finding study. *Pediatrics* 2002;**109**:602-7.

81. Eichenfield LF, Funk A, Fallon-Friedlander S, Cunningham BB. A clinical study to evaluate the efficacy of ELA-Max (4% liposomal lidocaine) as compared with eutectic mixture of local anesthetics cream for pain reduction of venipuncture in children. *Pediatrics* 2002;**109**:1093-9.

82. Choi SC,.Lu IL. Effect of non-random missing data mechanisms in clinical trials. *Stat.Med.* 1995;**14**:2675-84.

83. Unnebrink K,.Windeler J. Intention-to-treat: methods for dealing with missing values in clinical trials of progressively deteriorating diseases. *Stat.Med.* 2001;**20**:3931-46.

84. Lagakos SW, Lim LL, Robins JM. Adjusting for early treatment termination in comparative clinical trials. *Stat.Med.* 1990;**9**:1417-24.

85. Engels JM,.Diehr P. Imputation of missing longitudinal data: a comparison of methods. *J.Clin.Epidemiol.* 2003;**56**:968-76.

86. Revicki DA, Gold K, Buckman D, Chan K, Kallich JD, Woolley JM. Imputing physical health status scores missing owing to mortality: results of a simulation comparing multiple techniques. *Med.Care* 2001;**39**:61-71.

87. Streiner DL. The case of the missing data: methods of dealing with dropouts and other research vagaries. *Can.J.Psychiatry* 2002;**47**:68-75.

88. Reid IR, Brown JP, Burckhardt P, Horowitz Z, Richardson P, Trechsel U *et al.* Intravenous zoledronic acid in postmenopausal women with low bone mineral density. *N.Engl.J.Med.* 2002;**346**:653-61.

89. King M, Davidson O, Taylor F, Haines A, Sharp D, Turner R. Effectiveness of teaching general practitioners skills in brief cognitive behaviour therapy to treat patients with depression: randomised controlled trial. *BMJ* 2002;**324**:947-50.

90. Dey P, Bundred N, Gibbs A, Hopwood P, Baildam A, Boggis C *et al.* Costs and benefits of a one stop clinic compared with a dedicated breast clinic: randomised controlled trial. *BMJ* 2002;**324**:507.

91. Verhoef H, West CE, Nzyuko SM, de Vogel S, van d, V, Wanga MA *et al.* Intermittent administration of iron and sulfadoxine-pyrimethamine to control anaemia in Kenyan children: a randomised controlled trial. *Lancet* 2002;**360**:908-14.

92. Benator D, Bhattacharya M, Bozeman L, Burman W, Cantazaro A, Chaisson R *et al.* Rifapentine and isoniazid once a week versus rifampicin and isoniazid twice a week for treatment of drug-susceptible pulmonary tuberculosis in HIV-negative patients: a randomised clinical trial. *Lancet* 2002;**360**:528-34.

93. Molyneux EM, Walsh AL, Forsyth H, Tembo M, Mwenechanya J, Kayira K *et al.* Dexamethasone treatment in childhood bacterial meningitis in Malawi: a randomised controlled trial. *Lancet* 2002;**360**:211-8.

94. John JH, Ziebland S, Yudkin P, Roe LS, Neil HA. Effects of fruit and vegetable consumption on plasma antioxidant concentrations and blood pressure: a randomised controlled trial. *Lancet* 2002;**359**:1969-74.

95. Detmar SB, Muller MJ, Schornagel JH, Wever LD, Aaronson NK. Health-related quality-of-life assessments and patient-physician communication: a randomized controlled trial. *JAMA* 2002;**288**:3027-34.

96. Weeks JC, Nelson H, Gelber S, Sargent D, Schroeder G. Short-term quality-of-life outcomes following laparoscopic-assisted colectomy vs open colectomy for colon cancer: a randomized trial. *JAMA* 2002;**287**:321-8.

97. Johnson PC, Wheat LJ, Cloud GA, Goldman M, Lancaster D, Bamberger DM *et al.* Safety and efficacy of liposomal amphotericin B compared with conventional amphotericin B for induction therapy of histoplasmosis in patients with AIDS. *Ann.Intern.Med.* 2002;**137**:105-9.

98. Hanas R, Adolfsson P, Elfvin-Akesson K, Hammaren L, Ilvered R, Jansson I *et al.* Indwelling catheters used from the onset of diabetes decrease injection pain and pre-injection anxiety. *J.Pediatr.* 2002;**140**:315-20.

99. Hovell MF, Meltzer SB, Wahlgren DR, Matt GE, Hofstetter CR, Jones JA *et al.* Asthma management and environmental tobacco smoke exposure reduction in Latino children: a controlled trial. *Pediatrics* 2002;**110**:946-56.

100. Nager AL,.Wang VJ. Comparison of nasogastric and intravenous methods of rehydration in pediatric patients with acute dehydration. *Pediatrics* 2002;**109**:566-72.

Appendix I

Characteristics of the 10 journals

Journals	General Vs Specialized	Impact factor	Adhere to CONSORT
NEJM	General	31.7	No
BMJ	General	7.6	Yes
Lancet	General	15.4	Yes
JAMA	General	16.8	Yes
CMAJ	General	3.2	Yes
Annals Int Med	Specialized	11.4	Yes
Ann Emerg Med	Specialized	2.1	Yes
Ped Emerg Care	Specialized	0.6	No
J Peds	Specialized	3.2	No
Pediatrics	Specialized	3.4	No

Appendix II standardized article evaluation form

Specifications for the evaluation of the articles:

Item # 4 and 5: Unit of randomization and analysis

The unit of randomization/analysis is evaluated by comparing the relationship between randomization/analysis to the unit where the outcome was measured. For example: it is a group randomization if the randomization is performed at the level of a hospital, a cluster or a physician and the outcome is measured at the patient level.

Item #6

The primary analysis is defined as the analysis performed to answer the question of the primary objective described in the Methods section [1]. It is usually reported in the Results section of the Abstract [2] and specified as the primary analysis in the Results section [3]. In case of discrepancy between these three, the hierarchy of importance between the three is [1] before [3] before [2]. When more than one outcomes fulfill these criteria (for example when the outcome is measured at two different periods of time), the rater uses his judgment to decide the primary analysis based on what seems the most important outcome.

The authors used the following terms:

1. "Intention-to-treat" or "intent-to-treat" or "analyzed as randomized"
2. "per protocol" "analysis performed only on subject who followed the protocol"
3. "per treatment" "analyzed as treated" "analyzed according to the treatment received"

Item #7

There could be multiple answers. These are all the places where the authors used the words "intention-to-treat, per protocol..." in the text.

Items # 8

Use the same definitions as item #6. This item regards all analyses (primary and secondary) performed on the primary outcome.

There could be multiple answers.

Item #9

A modified ITT is a restriction suggested by the investigators with regards to their intention-to-treat principle. For example: « *we defined intention-to-treat as all patients that received at least one dose of treatment A or placebo* » In this example, if the investigators excluded all subjects who were not compliant, it is a per protocol analysis but if they exclude only those who did not received any treatment and there is a justification to exclude them, it is called modified intention-to-treat. Please describe the modification.

Item #10

Can have multiple answers

Item #11-18

Ideally, the rater should give the numbers for each treatment arms when it is possible. For each of these items provide the crude numbers. If impossible, give the percentage of the total number of randomized subjects. It is preferable to have all treatment arms merged together and have a complete table rather than having missing data for treatment arms evaluated separately.

If there is a control group (standard of care or placebo), write it in the first column. Identify column by treatment type.

Item 13-15 refer to missing data.

Item 14 refers to subjects for which there is no information. A subject who withdrew from the study intervention but who is still followed and can provide information for the analysis should not be included in this item. If this subject's information was

analyzed, it is not considered missing data. If the information was present but excluded from the analysis, it should count as part of item #16.

Item 16-18 refers to information that the investigators may have but did not use in the analysis.

Item #19 refers to the primary analysis. There should be 1 answer. Authors should usually name the imputation mode used for their analysis or the way they managed the missing information. The item "keep the subject in the denominator" refers to the cases where the investigators include all subjects in the denominator for the calculation of the proportions regardless of their missing data status. It is also usually used in the survival analysis.

Item #20
Multiple answers are possible. Circle all the methods named or described by the authors.

Item #21
Refers to the subjects who were randomized but were not supposed to be (e.g: they did not, in fact, meet the inclusion criteria or did meet certain exclusion criteria but had actually been randomized).

Item # 23
This is the number of authors who are cited at the bottom of the study title. If there is a writing committee, it includes only the writing committee members as defined in the authorship description

Item #24
Is there any affiliation with an epidemiology, biostatistics or public health department in the affiliation described in the authorship?

Item #25

"Industry" means funding by a private company. "Public" means funded by: government, public association, others. There is private funding when it is stipulated that a private company provided the study drugs or devices.

Item # 27

There can be multiple answers

- Drug/nutritional supplement are every medication, vitamins or supplement. It can be administrated IV, IM, oral, inhaled.
- Surgical: surgery vs non-surgery, different types of approach, different tools or techniques used
- Behavioral interventions are all interventions that regard lifestyle and knowledge (training programs, subjects or health professional education, psychotherapy, etc)
- Devices are patented devices (pacemaker, tracheostomy, feeding tube, etc).
- Physical therapy encompasses: massage, physiotherapy, ergotherapy, rehabilitation
- Diagnosis/screening is for all the trials that evaluate the effect of the use of a test on an outcome

Item #28

This is only for the primary analysis.

Item # 31

A difference-seeking trial is a trial in which it is stated in the Introduction or in the Methods that the hypothesis of the study is that there is a difference between the different interventions. If it is reported that the investigators aimed to show the equivalency or the non-inferiority of the interventions, then it is NOT a difference seeking trial.

Item # 32

Do the investigators report in the Abstract a difference between the two treatments for the primary analysis?

1. Article study number: _____ **2.Reviewer ID number**: _____

Article title: _____ Vol_____ First author: _____

3. 1NEJM 2BMJ 3Lancet 4JAMA 5CMAJ 6AnnIntMed 7AnnEmergMed 8PEC 9J. Peds 10Pediatrics

4. Randomization: 1 individual 2 group

5. Analysis: 1 individual 2 group 3 both 4 unclear

6. Method REPORTED by the authors for PRIMARY analysis (Page_____)
0. No mention 1.Intention-to-treat or intent-to-treat 2.Per protocol 3.As treated
4.Other_____

7. Places where the method for primary analysis is REPORTED (can be multiple)**:**
0. No reports 1.Abstract 2. Method 3. Results 4. Discussion

8. ALL methods REPORTED for analysis (including the primary analysis)
0. None 1. Intention to treat 2. "Per protocol" 3. "As treated"

9. Method USED for primary analysis (don't bother about missing data):
1. Subjects were analyzed as randomized regardless of the treatment they received (ITT)
2. Subjects were analyzed as randomized with some investigator's specific modification (modified ITT)_____
3. Only the subjects who followed the protocol were analyzed
4. Subjects were analyzed according to the treatment they received regardless of their randomization
5. Unclear
6. Other_____

10. Places where the description of the management of missing data for primary analysis is made:
0. No description 1. Abstract 2. Method 3. Results 4. Discussion

Rate of missing data in the PRIMARY ANALYSIS (subjects that were not analyzed)**:**

	Group _____	**Group_____**	**Group_____**	**Group_____**
11 Total number of randomized subjects				
12 Total number of analyzed subjects				
13 Withdrawal before beginning of treatment				
14 Loss to follow-up, withdrawal				
15 Missing data for other reason				
16 Improper inclusion				
17 Exclusion for non compliance				
18 Exclusion for other reason _____				

19. Strategy for missing data (PRIMARY analysis):
0. No mention
1. No missing data
2. Exclude subjects with missing data
3. Keep the subject in denominator
4. Use of worst case / best case scenario
5. Last observation carried forward (LOCF)
6. Regression
7. Replace by the mean
8. Multiple imputations
9. Other imputation_____

20. Other strategy for missing data (all analysis. May have multiple answers)**:**
0. Exclude subjects with missing data
1. Keep the subject in denominator
2. Use of worst / best case scenario
3. LOCF
4. Regression
5. Replace by the mean
6. Multiple imputations
7. Other imputation_____

21. Improper inclusion (subjects that were retrospectively declared ineligible)**:**
1. Say that there were none

2. No mention
3. Include them in analysis
4. Exclude them from analysis
5. Performed the analysis with and without them
6.Other _____

22. There is a diagram for the flow of participant: 0.No 1.Yes

23. Number of authors _____

24. Affiliation with an epidemiology/biostats/ public health department in the authorship: 0.No 1.Yes

25. Funding: 0. No funding declared 1. Industry 2. Public 3. Combination 4. Other

26. Number of centers: 1. One 2. Multi-center 3. No info

27. Type of intervention related to the primary analysis (can be multiple)
1. Drug/supplement 2. Surgical 3. Behavioral 4.Device 5. Physical therapy 6. Diagnosis/ screening procedure 7. Other _____

28. Nature of primary outcome
1. Categorical or dichotomous 2. Continuous 3. Survival 4. Other
_____

29. Use of a placebo
0. No 1. Yes

30. Type of design
1. Parallel 2. Cross-over/ matched 3. Factorial 4. Other _____

31. Difference-seeking trial
0. No 1. Yes 2. Unclear

32 Difference found
0. No 1. Yes 2. Unclear

Appendix III Summary of the 17 articles that violated a basic aspect of the intention-to-treat principle

Reference: Perl and al.[75]
Study resumé: A study evaluating intranasal mupirocin to prevent postoperative Staphylococcus aureus infections. The authors reported the exclusion after randomization of approximately 4% of the study population for improper inclusion or non-compliance. Of the 4030 subjects initially randomized, "*166 were excluded from the analysis, because they were not undergoing an eligible operation (49 in the mupirocin group and 48 in the placebo group), they received no study medication (22 and 26, respectively), or they met both exclusion criteria (8 and 13, respectively).*" The authors conclude that: "*3864 patients (95.9 percent) were included in the intention-to-treat analysis.*"

Reference: Lewith and al.[76]

Study resumé: A study regarding the clinical efficacy of homoeopathic potencies of house dust mite (homoeopathic immunotherapy) in asthmatic people allergic to house dust mite. Only 202 of the 242 randomized participants were analyzed. Seventeen of the participants were withdrawn because of a major protocol violation (oral steroids). There is also 1 patient that was excluded because of exacerbation of his asthma. One could conclude that they have been excluded from the study because of a worsening in their primary outcome that necessitated a more powerful treatment.

Reference: Reid and al.[88]

Study resumé: A study evaluating the effects of five regimens of zoledronic acid, the most potent bisphosphonate, on bone turnover and density in 351 postmenopausal women with low bone mineral density. The authors state in the Methods section that "*35 women were withdrew from the study, most commonly for personal reasons (15 women) or because of adverse events (14 women)*".Those non-compliant women were not evaluated in the final analysis.

Reference: King and al.[89]

Study resumé: This was a study to assess the effectiveness of teaching general practitioners skills in brief cognitive behaviour therapy. The intervention was a training package of 4 half days on brief cognitive behaviour therapy and the control

arm of the study had no intervention. A total of 116 practitioners were randomized but 32 withdrew before the intervention and only 51 were analyzed. An intention-to-treat analysis would mandate that all randomized practitioners should be analyzed. The effect of compliance to the intervention has to be taken into account in an intention-to-treat analysis.

Reference: Dey and al.[90]

Study resumé: This is a study comparing the cost of a 1-stop clinic vs a dedicated breast clinic for the assessment of women with suspected breast cancer. The women were randomized before consent. Women randomized in the 1-stop clinic could decide to be seen in a dedicated clinic. Of the 695 randomized subjects, 633 received the allocated intervention. All the women who did not receive the allocated intervention were not included in the analysis (per protocol analysis). Also, the rate of withdrawal was different between the two groups leaving room for bias.

Reference: Verhoef and al.[91]

Study resumé: This is a study evaluating the effect on hemoglobin concentrations of intermittent administration of iron supplements and sulfadoxine-pyrimethamine in symptom-free children in a region endemic for malaria. The primary outcome was hemoglobin concentration. Twenty-one of the 328 randomized subjects were not analyzed because of losses to follow-up, and 1 subject was withdrawn because of severe anaemia. These subjects should have been included in an intention-to-treat analysis.

Reference: Tuberculosis Trials Consortium[92]

Study resumé: This was a study evaluating the clinical and bacteriological efficacy of a once a week treatment of rifapentine and isoniazid vs a twice a week treatment for pulmonary tuberculosis. The primary outcome was the rate of relapse/failure. Of the 1004 subjects randomized, 9 died. They were included as not having the primary outcome. To be conservative, they should have been classified as having the primary outcome.

Reference: Molyneux and al.[93]

Study resumé: This study was performed in order to assess the effectiveness of dexamethasone in management of acute bacterial meningitis in a developing country. Of the 602 randomized subjects, 1 was excluded from the analysis because he did not receive the full treatment (he received only 1 dose).

Reference: John and al.[94]

Study resumé: This study was performed to evaluate the effect of an intervention to increase fruit and vegetable consumption on plasma concentrations of antioxidant vitamins, daily fruit and vegetable intake, and blood pressure. A total of 729 subjects were randomized. Of these, 22 did not attend the first appointment (intervention or control). These non-compliant subjects were excluded from the analysis.

Reference: Detmar and al.[95]

Study resumé: This study was performed to evaluate the efficacy of standardized Health Related Quality of Life (HRQL) assessments in facilitating patient-physician communication and increasing physicians' awareness of their subjects' HRQL-related problems. The intervention of interest was the use of a standardized HRQL questionnaire by the subjects before being seen by the physician. The primary outcome was the evaluation of the quantification of the HRQL topics discussed in the physician-patient visit. The study used a cross-over design and the unit of randomization was at the physician level. The analysis was at the subject level. A total of 273 subjects agreed to participate in the study and 109 declined. There was also the loss of 59 subjects during follow-up. The randomization of subjects was performed before they gave their consent but only the subjects who agreed to participate and receive the intervention were analyzed. All the subjects who were randomized but refused to participate were excluded. This is in conflict with the concept that all randomized subjects are analysed.

Reference: Weeks and al.[96]

Study resumé: This study evaluated quality of life outcomes after laparoscopic assisted colectomy vs open colectomy for colon cancer. A total of 576 subjects were randomized but 4 subjects were excluded from the analysis because of refusal to the allocated intervention.

Reference: Johnson and al.[97]

Study resumé: This is a study to compare amphotericin B vs. liposomal amphotericin B for induction therapy of moderate to severe disseminated histoplasmosis in subjects with AIDS. The investigators reported that "outcome analysis was performed on an intention-to-treat basis". A total of 81 subjects were randomly assigned to a treatment but 3 subjects were excluded from the "intention-to-treat safety analysis" because they withdrew consent before treatment but after randomization. Also another 4 subjects were excluded from the efficacy analysis because they had no histoplasmosis.

Reference: Callaham and Schriger[77]

Study resumé: Two studies were reported in this paper. In the first study, the investigators invited all 173 reviewers of a journal to attend the structured workshop. The intervention group was composed of the reviewers who attended the workshop while the control group was a sample of matched reviewers who did not attend. This is not true randomization and the analysis is more likely to be an "as treated" design. In the second study, 150 reviewers were randomized to either an intensive recruitment strategy to participate to the workshop or not to be recruited. Only the 11 subjects who participated to the workshop were analyzed in the intervention group. They were compared to a random sample of 11 reviewers who were not invited to participate to the workshop. This design is more likely to be called a "per protocol" analysis.

Reference: Hanas and al.[98]

Study resumé: This study evaluated the use of indwelling catheters as injection aids at diabetes onset to reduce injection pain and pre-injection anxiety. Forty-one of the 44 randomized subjects completed the study. One of them was excluded from the analysis because the parents decided to change him from the control group to the intervention group after the first day. This subject should have been analyzed in the control group to fulfill the requirements of an intention-to-treat analysis.

Reference: Hovell and al.[99]

Study resumé: This study tested the efficacy of coaching to reduce environmental tobacco smoke (ETS) exposure among asthmatic Latino children. The intervention consisted of 7 in-home sessions of 30-45 minutes over 3 months with a "coach" plus a booster phone call at the end. Five of the 193 randomized families dropped outs of the

study before the analysis. There is no information in the text whether these drop out were during or after the intervention. None of them was included in the analysis.

Reference: Eichenfield and al.[81]

Study resumé: This study assessed the efficacy of ELAMAX Vs EMLA cream for topical anesthesia during venipuncture procedures. It was a cross-over study and the primary outcome was a visual analog scale. One of the 120 randomized subjects was excluded because of a protocol deviation (the cream has been applied for only 5 minutes instead of 60 minutes). This subject should have been included in an intention-to-treat analysis.

Reference: Nager and Wang. [100]

Study resumé: This study assessed the safety, efficacy, and cost-effectiveness of rapid nasogastric hydration (RNG) and rapid intravenous hydration (RIV) administered in the emergency department (ED) to young children suffering with uncomplicated, acute moderate dehydration. The primary outcome was the failure rate. Of the 96 enrolled subjects, 3 subjects were excluded from the study because of treatment failure and another 2 subjects were excluded post-randomization because of severe dehydration. The conclusion of the authors is that RNG and RIV are safe and efficacious, but they excluded the subjects with treatment failure and those with severe dehydration.