ELSEVIER

Journal of Clinical Epidemiology 63 (2010) 142−153

Journal of Clinical Epidemiology

SYSTEMATIC REVIEWS

A substantial and confusing variation exists in handling of baseline covariates in randomized controlled trials: a review of trials published in leading medical journals

Peter C. Austin[a,b,c,*], Andrea Manca[d], Merrick Zwarenstein[a,c,e], David N. Juurlink[a,f], Matthew B. Stanbrook[a,f]

[a]*Institute for Clinical Evaluative Sciences, Toronto, Ontario, Canada*
[b]*Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada*
[c]*Department of Health Policy, Management and Evaluation, University of Toronto, Toronto, Ontario, Canada*
[d]*Centre for Health Economics, The University of York, York, United Kingdom*
[e]*Centre for Health Services Sciences, Sunnybrook Hospital, Toronto, Ontario, Canada*
[f]*Department of Medicine, University of Toronto, Toronto, Ontario, Canada*

Accepted 16 June 2009

Abstract

Objective: Statisticians have criticized the use of significance testing to compare the distribution of baseline covariates between treatment groups in randomized controlled trials (RCTs). Furthermore, some have advocated for the use of regression adjustment to estimate the effect of treatment after adjusting for potential imbalances in prognostically important baseline covariates between treatment groups.

Study Design and Setting: We examined 114 RCTs published in the *New England Journal of Medicine*, the *Journal of the American Medical Association*, *The Lancet*, and the *British Medical Journal* between January 1, 2007 and June 30, 2007.

Results: Significance testing was used to compare baseline characteristics between treatment arms in 38% of the studies. The practice was very rare in British journals and more common in the U.S. journals. In 29% of the studies, the primary outcome was continuous, whereas in 65% of the studies, the primary outcome was either dichotomous or time-to-event in nature. Adjustment for baseline covariates was reported when estimating the treatment effect in 34% of the studies.

Conclusions: Our findings suggest the need for greater editorial consistency across journals in the reporting of RCTs. Furthermore, there is a need for greater debate about the relative merits of unadjusted vs. adjusted estimates of treatment effect.

Keywords: Baseline covariates; Randomized controlled trial; Significance testing; Analysis of covariance; Regression adjustment; CONSORT statement; Clinical trial

1. Introduction

In randomized controlled trials (RCTs), randomization ensures that, on average, the distribution of baseline covariates will be similar between the different treatment arms. However, in any given RCT, there is likely to be residual differences in baseline characteristics between the treatment arms [1]. Adherence to the CONSORT guidelines on the reporting of RCTs requires that baseline demographic and clinical characteristics be described in each treatment group [2]. With few exceptions, the statistical literature is uniform in its agreement on the inappropriateness of using hypothesis testing to compare the distribution of baseline covariates between treated and untreated subjects in RCTs [1,3−6]. Senn writes that, in an RCT, ‘‘over all the randomizations the groups are balanced; and that for a particular randomization they are unbalanced’’ [1]. Thus, in an RCT, the only reason to use a significance test would be to examine the process of randomization itself. As Begg suggests, ‘‘a significance test of the association between the covariate and the treatment assignment is a test of the hypothesis that the treatments are randomly distributed. In other words, it is a test of a null hypothesis that is known to be true’’ [3]. Similarly, Altman writes that ‘‘performing a significance test to compare baseline variables is to assess the probability of

* Corresponding author. Institute for Clinical Evaluative Sciences, G1 06, 2075 Bayview Avenue, Toronto, Ontario M4N 3M5, Canada. Tel.: +416-480-6131; fax: +416-480-6048.
E-mail address: peter.austin@ices.on.ca (P.C. Austin).

0895-4356/10/$ − see front matter
doi: 10.1016/j.jclinepi.2009.06.002

What is new?

Key findings

- Substantial variation exists among leading medical journals in the use of statistical significance testing to compare baseline covariates between treatment groups in randomized controlled trials (RCTs).
- Unadjusted estimates of treatment effect were reported more frequently than adjusted analyses.
- Primary outcomes were more frequently either binary or time-to-event in nature than continuous in published RCTs.

What this adds to what was known?

- In the current era, there remain inconsistencies between the reporting of analyses in RCTs and what is considered the best statistical practice.

What is the implication and what should change now?

- There is a need for greater editorial consistency in enforcing the statistical proscription against the use of statistical significance testing for comparing baseline covariates between treatment groups in RCTs.
- There is a need for an informed debate about the relative interpretability and utility for clinical and policy decision making of unadjusted vs. adjusted measures of treatment effect for binary and time-to-event outcomes.

something having occurred by chance when we know that it did occur by chance. Such a procedure is clearly absurd'' [4]. Although randomization will, on average, balance covariates between treated and untreated subjects, it need not do so in any particular randomization.

A related issue when estimating the effect of treatment on outcomes in RCTs is adjusting for baseline covariates. Different authors have argued that regression analysis should be used when estimating the effect of the treatment on the outcome [1,7–9]. This allows one to adjust for chance imbalance in prognostically important baseline covariates. Furthermore, the use of regression analysis can result in increased precision when estimating the treatment effect [8,9]. Importantly, both Senn and Lavori et al. argue that the covariates should be selected a priori, and that the selected covariates should be strongly prognostic of the outcome [1,9]. The selection of covariates for inclusion in the regression model should not be based on significance tests comparing their distribution across arms of the trial.

It is not known to what extent the prohibition against using significance testing to compare baseline covariates between treatment arms and the prescription of the use of regression adjustment are adhered to in the current medical literature. The objective of the current study was to conduct a review of the treatment of baseline covariates in published RCTs in the general medical literature.

2. Methods for the systematic review

We conducted a systematic review of RCTs published in four leading general medical journals: the *New England Journal of Medicine* (*NEJM*), the *Journal of the American Medical Association* (*JAMA*), *The Lancet*, and the *British Medical Journal* (*BMJ*). We searched PubMed for articles published in these four journals between January 1, 2007 and June 30, 2007. We limited our search to reports of RCTs in humans. We excluded cluster randomization trials, reports of subgroup analyses of prior RCTs, secondary analyses of prior RCTs, studies in which no primary outcome was identified, and cost-effectiveness analyses conducted alongside RCTs.

The CONSORT statement requires that authors describe the baseline demographic and clinical characteristics of each group (item 15) [2]. In an explanation and elaboration of the CONSORT statement, it is suggested that baseline information is ''efficiently presented in a table'' [10]. We examined whether each published report of an RCT included a table describing baseline demographic and clinical characteristics of each arm. When such a table was not reported, we examined whether a comparison of baseline characteristics was described in the text.

Many RCTs use stratified randomization. Because it is commonly agreed that stratification factors should be accounted for in the analysis estimating treatment efficacy [11], an analysis was not considered an adjusted analysis if it adjusted only for stratification factors. Similarly, many studies adjust the effect of treatment on outcome for the baseline value of the response variable. An analysis that adjusted for the value of the response at baseline was not considered an adjusted analysis. Thus, an adjusted analysis had to adjust for at least one baseline covariate that was not a stratification factor and that was not the value of the response variable at baseline.

For each identified study, we abstracted the following information: (1) the primary outcome; (2) the nature of the primary outcome; (3) whether a table was published comparing baseline characteristics between treatment arms. When such a table was not published, we examined whether between-group comparisons of baseline characteristics were described in the text; (4) when such a table was published, we examined whether it was reported that significance testing had been used to compare the distribution of baseline covariates across treatment arms; (5) whether an unadjusted analysis of the effect of treatment on the primary outcome was reported. If an unadjusted analysis was reported for a continuous outcome, we examined whether the analysis had adjusted for the baseline value of the response variable; (6) whether an adjusted analysis of the effect of treatment on

the primary outcome was reported; and (7) how authors reported selecting the covariates for adjustment when an adjusted analysis was reported.

When multiple primary outcomes were listed, we selected the primary outcome that was described as the ''first'' primary outcome. If none of the primary outcomes was described as the ''first'' primary outcome, we selected the primary outcome that was used in the sample-size calculation. If multiple primary outcomes were listed and each was used in a sample-size calculation, then we used all identified primary outcomes.

When recording the method used for selecting baseline covariates for use in an adjusted analysis, we allowed for three possible options. First, we noted whether the authors explicitly reported that covariates for adjustment were selected before the analysis. Second, we noted whether the selection of covariates was reported to be based on a post hoc analysis (e.g., observed imbalance in baseline covariates, statistical significant differences in baseline covariates between arms of the study, or by the use of an automated variable-selection method, such as backward variable elimination). Third, in the absence of the reported use of one of the first two methods for covariate selection, we described the method of covariate selections as unclear.

Each identified article was read independently by two study authors. One study author (P.C.A.) read all the identified articles. Two of the authors (A.M. and M.B.S.) each read one-third of the identified articles. A fourth study author (M.Z.) abstracted some items from the remaining third, whereas the fifth study author (D.N.J.) abstracted additional items from these articles. Disagreements between abstractors were resolved by the two abstractors reviewing the disputed study and arriving at a consensus.

As a prespecified analysis, we used a chi-square test to test whether there were differences among the four journals in the proportion of articles that reported using statistical hypothesis testing to compare baseline characteristics between randomization arms.

3. Results of the systematic review

Results of the review are summarized in Table 1 and are described in the following subsections.

3.1. Included studies

The initial search strategy identified 135 published reports of RCTs [12−146]. Of these, 21 were excluded for the following reasons: ancillary study of prior RCT (one study) [75], cluster randomization trials (seven studies) [37,50,60,125,130,135,144], cost-effectiveness studies (one study) [133], no primary outcome identified (three studies) [26,116,142], pooled subgroup analysis of two RCTs (one study) [55], secondary analysis of an RCT (six studies) [25,31,71,88,112,128], subgroup analysis of an RCT (one study) [61], and substudy of an RCT (one study) [131]. This resulted in the inclusion of 114 published reports of RCTs. Of these, 42 (36.8%) were published in the *NEJM*, 25 (21.9%) in the *JAMA*, 28 (24.6%) in *The Lancet*, and 19 (16.7%) in the *BMJ*.

3.2. Comparison of baseline characteristics between treatment arms

Of the included 114 randomized trials, 110 (96.5%) reported a table in which baseline demographic and clinical characteristics were compared between treatment arms. Of the 110 randomized trials reporting such a table, significance testing was used in 42 (38.2%) to compare the distribution of baseline covariates across treatment arms. When we examined the prevalence of significance testing to compare baseline covariates in each journal separately, the results were as follows: 29 of 40 (72.5%) articles in the *NEJM*, 11 of 24 (45.8%) articles in the *JAMA*, 2 of 28 (7.1%) articles in *The Lancet*, and 0 of 18 (0%) articles in the *BMJ* reported using statistical hypothesis testing to compare baseline covariates between randomization groups. A chi-square test was used to compare the proportion of studies in which significance was used to compare baseline covariates between the four journals. The significance level of the chi-square test was less than 0.0001, indicating that there were differences in the probability of using significance testing to compare baseline covariates across these four journals. In inspecting the aforementioned results, one notes that only two of the 46 studies (4.4%) published in the two British journals (*The Lancet* and the *BMJ*) used significance testing to compare baseline covariates across treatment arms. In contrast, 40 of the 64 (62.5%) studies published in the two American journals (*NEJM* and *JAMA*) used significance testing ($P < 0.0001$ for the comparison between the two British journals and the two American journals). Of the four studies that did not report a table comparing baseline characteristics between treatment groups, three provided brief qualitative or quantitative comparisons of the baseline characteristics between study groups. Of these three studies, one study reported the result of a significance test comparing a baseline characteristic between the study groups. One study did not report any comparison of baseline characteristics between randomization arms.

3.3. Nature of the primary outcome

In the 114 included studies, the nature of primary outcome was as follows: continuous (29.0%), binary (28.1%), time-to-event (28.1%), binary/time-to-event (8.8%), count (2.6%), continuous/binary (0.9%), continuous/time-to-event (0.9%), and ordinal (1.8%). In several studies, the primary outcome was analyzed as both a binary and as a time-to-event outcome [14,36,66,79,90,93,106,110,134,137]. Journal-specific prevalences of the different types of outcomes are reported in Table 1.

Table 1
Overall and journal-specific results of review of handling of baseline covariates in reports of RCTs

Item	*BMJ*	*JAMA*	*Lancet*	*NEJM*	Overall
Comparison of baseline characteristics					
Table comparing baseline characteristics	18/19	24/25	28/28	40/42	110/114
Reported significance testing to compare characteristics in published table	0/18	11/24	2/28	29/40	42/110
Nature of primary outcome					
Continuous	8/19	10/25	3/28	12/42	33/114
Binary	4/19	5/25	10/28	13/42	32/114
Time-to-event	2/19	8/25	12/28	10/42	32/114
Binary/time-to-event	2/19	2/25	1/28	5/42	10/114
Count	2/19	0/25	0/28	1/42	3/114
Continuous/binary	1/19	0/25	0/28	0/42	1/114
Continuous/time-to-event	0/19	0/25	0/28	1/42	1/114
Ordinal	0/19	0/25	2/28	0/42	2/114
Estimates of treatment effect					
Unadjusted analysis reported	17/19	21/25	28/28	41/42	107/114
Adjusted analysis reported	7/19	10/25	12/28	10/42	39/114
Reported method for the selection of covariates for adjusted analyses					
A priori selection	2/7	1/10	2/12	1/10	6/39
Post hoc selection	2/7	2/10	3/12	0/10	7/39
Unclear	3/7	7/10	7/12	9/10	26/39

Abbreviations: RCTs, randomized controlled trials; *BMJ*, *British Medical Journal*; *JAMA*, *Journal of the American Medical Association*; *NEJM*, *New England Journal of Medicine*.

3.4. Unadjusted and adjusted estimates of treatment effect

Of the 114 included articles, 107 (93.9%) presented unadjusted analyses of treatment effect, six (5.3%) did not report unadjusted analysis, and for one study (0.9%), it was unclear whether an unadjusted analysis had been conducted. Similarly, 39 (34.2%) presented adjusted analyses of treatment effect, 72 (63.2%) did not report adjusted analysis, and for three studies (2.6%), it was unclear whether an adjusted analysis had been conducted. Journal-specific frequencies of unadjusted and adjusted analyses are reported in Table 1. Of the reviewed articles, 34 (31.8%) explicitly presented both unadjusted and adjusted estimates of treatment effect.

In 33 (29.0%) studies, the nature of the primary outcome was only continuous. We excluded one study in which it was unclear whether an unadjusted analysis had been conducted. In the remaining 32 studies, 27 (84.4%) reported using an unadjusted analysis, 11 (34.4%) reported an adjusted analysis, and six (18.8%) reported both unadjusted and adjusted analysis. In 74 (64.9%) studies, the nature of the primary outcome was either binary, time-to-event, or binary/time-to-event. We excluded the two studies in which it was unclear whether unadjusted or adjusted analyses had been conducted. In the 72 remaining studies, 71 (98.6%) reported using an unadjusted analysis, 25 (34.7%) reported an adjusted analysis, and 25 (34.7%) reported both unadjusted and adjusted analyses. In 32 of these 74 studies, the primary outcome was binary. Among these 32 studies, 31 (96.9%) reported an unadjusted analysis, 11 (34.4%) reported an adjusted analysis, and 11 (34.4%) reported both unadjusted and adjusted analyses. In 32 of the aforementioned 74 studies, the primary outcome was time-to-event in nature. Among these 32 studies, 32 (100%) reported an unadjusted analysis, 12 (37.5%) reported an adjusted analysis, and 12 (37.5%) reported both unadjusted and adjusted analyses. Finally, in eight of the aforementioned studies, the primary outcome was both binary and time-to-event in nature. In these eight studies, eight (100%) reported an unadjusted analysis, two (25.0%) reported an adjusted analysis, and two (25.0%) reported both unadjusted and adjusted analyses.

There were 28 studies that reported an unadjusted estimate for the effect of treatment on a primary outcome that was only a continuous variable. Of these, 15 (53.6%) reported adjusting for the baseline value of the response variable (in one of these 15 studies, the continuous outcome was also the change from baseline). Twelve (42.9%) studies did not adjust for the baseline value of the response variable. In these 12 studies, the outcome variable was the change from baseline in four studies and the percent change from baseline in one study. In one of the 28 studies, it was unclear whether the analysis had adjusted for the baseline value of the response variable.

3.5. Selection of covariates for use in adjusted analyses

Thirty-nine studies reported the results of an adjusted analysis. Of these, six (15.4%) clearly reported that the

variables used for adjustment had been selected a priori. In two (5.1%) studies, it was reported that stepwise regression was used to select the baseline characteristics for inclusion in the adjusted model. In three studies (7.7%), the selection of covariates was reported to be based on significance testing, whereas in two (5.1%) studies, the selection of covariates was based on a post hoc observation or decision. Finally, in 26 (66.7%) of the studies, it was unclear how the covariates for inclusion in the regression model were selected. Journal-specific prevalences of the different methods for selection covariates for adjustment are reported in Table 1. One notes that the proportion of studies in which it was reported that covariates were selected before the analysis varied across the four journals.

4. Discussion

The objective of the current study was to examine the treatment of baseline covariates in RCTs published in leading general medical journals. The CONSORT statement requires that authors describe the baseline demographic and clinical characteristics of each group (item 15) [2]. In an explanation and elaboration of the CONSORT statement, it is suggested that baseline information is ''efficiently presented in a table'' [10]. We found that all but four (3.5%) of the 114 trials reported a table comparing baseline characteristics between the treatment arms. Brief quantitative or qualitative comparisons of the similarity of baseline characteristics between treatment groups were reported in the text in three of the four articles in which a table was not published. Several authors have criticized the use of significance testing to compare the distribution of baseline covariates in RCTs [1,3−6]. We found that 38.2% of studies reported using significance testing to compare baseline covariates between treatment arms. The use of significance testing varied across journals. In general, the practice was rare in the two British journals (*BMJ* and *The Lancet*), whereas it occurred in most of the studies published in the two U.S. journals (*NEJM* and *JAMA*). We speculate that the observed differences between journals may reflect editorial preferences, with the two British journals having an editorial proscription against the use of significance testing, whereas the two U.S. journals may leave the decision to individual authors or associate editors.

Roberts and Torgerson suggest that carrying out statistical tests comparing baseline characteristics between study arms is the result of confusion between three different questions: (1) Has the randomization been properly conducted? (2) Could imbalance in baseline characteristics cause chance bias? (3) Should the analysis be adjusted for baseline variables? [147]. The European Agency for the Evaluation of Medical Products states: ''Statistical testing for baseline imbalance has no role in a trial where the handling of the randomisation and blinding has been fully satisfactory. Baseline summaries with respect to the main covariates should be presented and discussed from a clinical point of view, irrespective of whether a statistical test indicated a ''statistically significant'' difference between treatment groups. If the process of allocating patients to treatment has, in fact, not been random then any resulting bias cannot be corrected by any statistical adjustment'' [148]. Fayers and King provide a brief summary of methods for handling baseline characteristics in RCTs [149].

Given the strong condemnation in the statistical and methodological literature of the use of significance testing to compare baseline characteristics between treatment arms, our study suggests the need for greater consistency in editorial policies across journals concerning the reporting of RCTs. We suggest that journal editors adopt a policy against the use of significance testing to compare baseline covariates between treatment arms. As Roberts and Torgerson suggest, authors may be incorrectly selecting covariates for inclusion in a regression model to estimate an adjusted treatment effect based on the results of statistical tests comparing baseline characteristics between treatment groups [147]. Senn argues that the selection of covariates for inclusion in a regression model should be based on an a priori judgment of which variables are strongly prognostic of the outcome [1]. An inappropriate selection of covariates for adjustment may result in biased estimates of treatment effect. Pocock et al. suggest that there is often inadequate prior knowledge as to which baseline characteristics are related to prognosis, and that further research is needed in variable selection algorithms [150]. Furthermore, there is the danger that ad hoc variable-selection methods can lead to biased inferences [150]. In a specific setting, Pocock et al. suggest that if the correlation between a single covariate and a continuous outcome is weak, then even statistically significant covariate imbalance is unimportant, whereas if a covariate is strongly correlated with the outcome, then it is important to adjust for that covariate, regardless of the statistical significance of the imbalance [150].

Several authors have suggested that, rather than presenting unadjusted estimates, regression adjustment be used to estimate treatment effects [1,7−9,150]. Advantages to this approach include the ability to adjust for chance imbalance in prognostically important baseline covariates between treatment arms, and greater precision in estimating the treatment effect. We found that approximately one-third of studies reported an adjusted estimate of treatment effect. In most of the studies, only unadjusted estimates of treatment effect were reported.

The suggestion for reporting adjusted estimates has typically been made in the context of a linear treatment effect and a continuous outcome variable. In the current study, we found that 34.4% of the studies in which the outcome was continuous reported an adjusted estimate of treatment effect. We found that adjustment for baseline covariates was used in the medical literature for dichotomous outcomes and for time-to-event (survival) outcomes as well as for continuous outcomes. Of studies in which the

primary outcome was binary or time-to-event in nature, 34.7% reported an adjusted analysis. However, there are theoretical limitations to the use of regression adjustment when estimating the effect of treatment on dichotomous and time-to-event outcomes. Gail et al. found that the unadjusted treatment effect in a randomized clinical trial can lead to an attenuated estimate of treatment effect compared with the adjusted treatment effect when the outcome is dichotomous or time-to-event [151]. When estimating the effect of treatment on a binary or time-to-event outcome, the unadjusted estimate of the treatment effect will be systematically closer to the null value compared with the adjusted treatment effect, even in the presence of randomization. However, for linear treatment effects on continuous outcomes, the adjusted and unadjusted treatment effects will, on average, coincide. This phenomenon is related to the issue of the collapsibility of an estimator that has been discussed by Greenland et al. [152,153]. An estimator is collapsible when the population average or marginal effect is the same as the conditional, adjusted, or subject-specific effect. Differences in means and risk differences are collapsible. The unadjusted difference in means is equal to the subject-specific difference in means. For binary outcomes, regression adjustment would usually entail the use of a logistic regression model. However, the odds ratio is not a collapsible estimator. Thus, the marginal odds ratio will differ from the conditional or subject-specific odds ratio. For this and other reasons, the use of the odds ratio in prospective studies is discouraged [152,154]. For binary outcomes, risk differences and relative risks (assuming a uniform relative risk) are collapsible estimators [152]. However, their use precludes the use of regression adjustment (although an adjusted relative risk can be estimated using a log-binomial generalized linear model, its use in practice may be problematic owing to difficulties with estimation and convergence). When outcomes are binary or time-to-event, adjusted and unadjusted estimates can be expected to differ. There is disagreement in the literature as to which estimate (the marginal/unadjusted/crude estimate vs. the conditional/adjusted) is more informative to clinicians, policy makers, and health care funders. Hauck et al. argue that the conditional estimate is more meaningful from a clinical perspective [155]. In contrast, Martens et al. suggest that the marginal treatment effect is better defined and appears to be of greater interest [156]. This issue requires further attention, because in most of the studies examined, the primary outcome was either binary or time-to-event in nature: settings in which the conditional and marginal estimates do not coincide. In only a few studies (29.0%) was the primary outcome continuous in nature.

In two recent articles, methods for estimating absolute risk reductions and numbers needed to treat from adjusted logistic regression models and adjusted survival models have been proposed [157,158]. These methods are applicable both in the context of RCTs, in which one wants to adjust for potential imbalance in prognostically important baseline covariates, and in observational studies, in which there may be systematic differences between treated and untreated subjects. Increased use of these methods in RCTs with binary or time-to-event outcomes would allow for the reporting of risk differences, which are collapsible: the population-average risk difference is equivalent to the subject-specific risk difference. Furthermore, absolute measures of treatment effect may be more useful for clinical decision making than relative measures, such as the odds ratio.

Our study suggests the need for an informed debate about the relative merits of the interpretation of unadjusted vs. adjusted estimates of treatment effect in settings with binary or time-to-event outcomes. In particular, the relative merits of adjusted vs. unadjusted estimates of treatment effect for clinicians, policy makers, and health care funders require further examination. If marginal or population-average treatment effects are more meaningful from either a clinical or a policy perspective, then solely reporting adjusted estimates of treatment effect may result in a misunderstanding about the magnitude of the benefit of treatment at the population level [151].

A second criticism of the advocacy of regression adjustment in RCTs is that it only allows adjustment for measured prognostic variables and not for unmeasured variables. Some have argued that if measured prognostic variables are imbalanced, it is possible that prognostically important unmeasured variables are also imbalanced [159]. Regression adjustment using measured baseline variables may result in the unmeasured baseline variables remaining imbalanced between the different arms of the trial. Regression adjustment using measured baseline variables may provide false comfort to investigators and readers.

We found that six of the 114 (5.3%) studies did not report the results of an unadjusted analysis. There are meta-analytic implications of presenting only adjusted estimates of treatment effect. Meta-analyses pool estimates of treatment effects across RCTs examining the same exposure and outcome. The pooling of unadjusted treatment effects is relatively simple from a conceptual perspective, because the nature of the treatment effect is the same across all studies—an unadjusted or population-average treatment effect. However, interpreting the results of a meta-analysis that pooled adjusted treatment effects is more difficult. This is because each study may have adjusted for different baseline covariates. Thus, each study reports a different adjusted treated effect. Martens et al. argue that, in a given study, there exist multiple conditional treatment effects: adjusting for a different set of covariates results in a different conditional treatment effect [156]. From a conceptual perspective, it is unclear how to interpret the pooling of adjusted treatment effects, when different sets of covariates were controlled for in each study. Additionally, pooling adjusted and unadjusted treatment effects is problematic when outcomes are binary or time-to-event in nature. This is because one is pooling estimates of conditional and marginal treatment effects. Because of the noncollapsibility of the odds ratio and the

hazard ratio, these are estimating different quantities. For these reasons, from a meta-analytic perspective, the reporting of unadjusted estimates is strongly encouraged. These problems are exacerbated when the analysis of summary data involves the use of multiparameter evidence synthesis or mixed treatment-comparison models [160].

Lavori et al. examined the reports of 35 RCTs published in the *NEJM* in 1978 and 1979 [9]. They found that a ubiquitous error was the use of significance testing to compare baseline balance in covariates. Altman and Dore reviewed the reports of 80 RCTs published in 1987 and 1988 in four general medical journals (including three that were considered in our review) [6]. They found that hypothesis tests were used to compare baseline variables in 58% of the trials (Altman and Dore note that, of the 600 hypothesis tests conducted in these 46 trials, 4% were significant at the 5% level). However, they did not report journal-specific results for the use of significance testing to compare baseline characteristics. Pocock et al. examined the reports of 50 RCTs published in 1997 in the same four general medical journals that were examined in the current study [150]. They found that hypothesis tests were used to compare baseline variables in 48% of the trials. In comparison, we found that 38.2% of studies reported using significance testing to compare baseline characteristics. The study by Pocock et al. did not report journal-specific results. Thus, journal-specific comparisons cannot be made with our study. The use of significance testing to compare baseline covariates between treatment arms appears to have decreased substantially across the eras examined by Lavori et al. (1978 and 1979), by Altman and Dore (1987 and 1988), by Pocock et al. (1997), and by the current study (2007).

Altman and Dore found that when estimating treatment effects, 64% of studies reported either unadjusted estimates or estimates that accounted for change from baseline. Only 26% of the studies reported using statistical modeling that adjusted for baseline covariates when estimating the treatment effect. In the study by Altman and Dore, journal-specific rates of reporting the use of an adjusted analyses were 40%, 10%, 20%, and 35% for the *Annals of Internal Medicine*, the *BMJ*, *The Lancet*, and the *NEJM*, respectively. In the current review, it was found that 34.2% of studies reported an analysis that adjusted for baseline covariates. Thus, there was a modest increase in the use of regression adjustment for estimating treatment effects between 1987 and 2007. In the current review, journal-specific rates of reporting an adjusted analysis were 36.8%, 40.0%, 42.9%, and 23.8% for the *BMJ*, the *JAMA*, *The Lancet*, and the *NEJM*, respectively. Therefore, the reported use of an adjusted analysis increased in both the *BMJ*, and *The Lancet* between the era examined by Altman and Dore (1987/1988) and that considered in the current review (2007). However, the use of adjusted analyses decreased in the *NEJM* between these two periods.

There are several limitations to the current study. First, we only examined RCTs published in four leading general medical journals over a 6-month period in 2007. We did not examine RCTs published in other journals. Thus, we do not know the extent to which our findings are generalizable to other journals. However, despite restricting our review to leading general medical journals, we did observe substantial between-journal heterogeneity in some aspects of the reporting of the results of RCTs. Second, we were only able to examine what authors reported, and not what was actually done. A prior study found that in a substantial proportion of studies, there were unacknowledged discrepancies between the published reports and the study protocols [161]. It is conceivable that there were discrepancies between what was done and what was reported in the studies that we examined.

In summary, we found that the statistical treatment of baseline covariates in leading general medical journals differs from the guidance provided in the statistical literature. In the *NEJM* and in the *JAMA*, statistical hypothesis testing was routinely used to compare baseline covariates between treatment arms, whereas the practice was rare in the *BMJ* and *The Lancet*. Adjustment for potential imbalance in baseline covariates was only performed in a few studies. In most of the RCTs examined, the primary outcomes were either binary or time-to-event in nature. Further examination of the relative merits of adjusted and unadjusted estimates of treatment effects in this context is warranted.

Acknowledgments

The Institute for Clinical Evaluative Sciences (ICES) is supported in part by a grant from the Ontario Ministry of Health and Long Term Care. The opinions, results, and conclusions are those of the authors and no endorsement by the Ministry of Health and Long-Term Care or by the Institute for Clinical Evaluative Sciences is intended or should be inferred. Dr. Austin is supported in part by a Career Investigator Award from the Heart and Stroke Foundation of Ontario, and Dr. Juurlink was supported by a New Investigator Award from the Canadian Institutes for Health Research.

References

[1] Senn S. Testing for baseline balance in clinical trials. Stat Med 1994;13:1715−26.
[2] 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.
[3] Begg CB. Significance tests of covariate imbalance in clinical trials. Control Clin Trials 1990;11:223−5.
[4] Altman DG. Comparability of randomised groups. Statistician 1985;34:125−36.
[5] Rothman KJ. Epidemiologic methods in clinical trials. Cancer 1977;39:1771−5.
[6] Altman DT, Dore CJ. Randomisation and baseline comparisons in clinical trials. Lancet 1990;335:149−53.

[7] Senn SJ. Covariate imbalance and random allocation in clinical trials. Stat Med 1989;8:467−75.
[8] Altman DG, Dore CJ. Baseline comparisons in randomized clinical trials. Stat Med 1991;10:797−802.
[9] Lavori PW, Louis TA, Bailar JC III, Polansky M. Designs for experiments—parallel comparisons of treatment. N Engl J Med 1983;309:1291−8.
[10] Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, et al. for the CONSORT Group. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. Ann Intern Med 2001;134:663−94.
[11] Cook TD, DeMets DL, editors. Introduction to statistical methods for clinical trials. Boca Raton, FL: Chapman & Hall/CRC; 2008.
[12] Savoye M, Shaw M, Dziura J, Tamborlane WV, Rose P, Guandalini C, et al. Effects of a weight management program on body composition and metabolic parameters in overweight children: a randomized controlled trial. JAMA 2007;297:2697−704.
[13] Cole BF, Baron JA, Sandler RS, Haile RW, Ahnen DJ, Bresalier RS, et al. Polyp Prevention Study Group. Folic acid for the prevention of colorectal adenomas: a randomized clinical trial. JAMA 2007;297:2351−9.
[14] Dorsey G, Staedke S, Clark TD, Njama-Meya D, Nzarubara B, Maiteki-Sebuguzi C, et al. Combination therapy for uncomplicated falciparum malaria in Ugandan children: a randomized trial. JAMA 2007;297:2210−9.
[15] Leslie T, Mayan MI, Hasan MA, Safi MH, Klinkenberg E, Whitty CJ, et al. Sulfadoxine-pyrimethamine, chlorproguanil-dapsone, or chloroquine for the treatment of *Plasmodium vivax* malaria in Afghanistan and Pakistan: a randomized controlled trial. JAMA 2007;297:2201−9.
[16] Ebbeling CB, Leidig MM, Feldman HA, Lovesky MM, Ludwig DS. Effects of a low-glycemic load vs low-fat diet in obese young adults: a randomized trial. JAMA 2007;297:2092−102.
[17] Church TS, Earnest CP, Skinner JS, Blair SN. Effects of different doses of physical activity on cardiorespiratory fitness among sedentary, overweight or obese postmenopausal women with elevated blood pressure: a randomized controlled trial. JAMA 2007;297:2081−91.
[18] Erne P, Schoenenberger AW, Burckhardt D, Zuber M, Kiowski W, Buser PT, et al. Effects of percutaneous coronary interventions in silent ischemia after myocardial infarction: the SWISSI II randomized controlled trial. JAMA 2007;297:1985−91.
[19] Mebazaa A, Nieminen MS, Packer M, Cohen-Solal A, Kleber FX, Pocock SJ, et al. SURVIVE Investigators. Levosimendan vs dobutamine for patients with acute decompensated heart failure: the SURVIVE randomized trial. JAMA 2007;297:1883−91.
[20] Morrow DA, Scirica BM, Karwatowska-Prokopczuk E, Murphy SA, Budaj A, Varshavsky S, et al. MERLIN-TIMI 36 Trial Investigators. Effects of ranolazine on recurrent cardiovascular events in patients with non-ST-elevation acute coronary syndromes: the MERLIN-TIMI 36 randomized trial. JAMA 2007;297:1775−83.
[21] Tardif JC, Grégoire J, L'Allier PL, Ibrahim R, Lespérance J, Heinonen TM, et al. Effect of rHDL on Atherosclerosis-Safety and Efficacy (ERASE) Investigators. Effects of reconstituted high-density lipoprotein infusions on coronary atherosclerosis: a randomized controlled trial. JAMA 2007;297:1675−82.
[22] Alexander JH, Reynolds HR, Stebbins AL, Dzavik V, Harrington RA, Van de Werf F, et al. TRIUMPH Investigators. Effect of tilarginine acetate in patients with acute myocardial infarction and cardiogenic shock: the TRIUMPH randomized controlled trial. JAMA 2007;297:1657−66.
[23] Treanor JJ, Schiff GM, Hayden FG, Brady RC, Hay CM, Meyer AL, et al. Safety and immunogenicity of a baculovirus-expressed hemagglutinin influenza vaccine: a randomized controlled trial. JAMA 2007;297:1577−82.
[24] Halonen J, Halonen P, Järvinen O, Taskinen P, Auvinen T, Tarkka M, et al. Corticosteroids for the prevention of atrial fibrillation after cardiac surgery: a randomized controlled trial. JAMA 2007;297:1562−7.
[25] Rossouw JE, Prentice RL, Manson JE, Wu L, Barad D, Barnabei VM, et al. Postmenopausal hormone therapy and risk of cardiovascular disease by age and years since menopause. JAMA 2007;297:1465−77.
[26] Brandes JL, Kudrow D, Stark SR, O'Carroll CP, Adelman JU, O'Donnell FJ, et al. Sumatriptan-naproxen for acute treatment of migraine: a randomized trial. JAMA 2007;297:1443−54.
[27] Gheorghiade M, Konstam MA, Burnett JC Jr, Grinfeld L, Maggioni AP, Swedberg K, et al. Efficacy of Vasopressin Antagonism in Heart Failure Outcome Study with Tolvaptan (EVEREST) Investigators. Short-term clinical effects of tolvaptan, an oral vasopressin antagonist, in patients hospitalized for heart failure: the EVEREST Clinical Status Trials. JAMA 2007;297:1332−43.
[28] Konstam MA, Gheorghiade M, Burnett JC Jr, Grinfeld L, Maggioni AP, Swedberg K, et al. Efficacy of Vasopressin Antagonism in Heart Failure Outcome Study with Tolvaptan (EVEREST) Investigators. Effects of oral tolvaptan in patients hospitalized for worsening heart failure: the EVEREST Outcome Trial. JAMA 2007;297:1319−31.
[29] Nissen SE, Nicholls SJ, Wolski K, Howey DC, McErlean E, Wang MD, et al. Effects of a potent and selective PPAR-alpha agonist in patients with atherogenic dyslipidemia or hypercholesterolemia: two randomized controlled trials. JAMA 2007;297:1362−73.
[30] Crouse JR III, Raichlen JS, Riley WA, Evans GW, Palmer MK, O'Leary DH, et al. METEOR Study Group. Effect of rosuvastatin on progression of carotid intima-media thickness in low-risk individuals with subclinical atherosclerosis: the METEOR Trial. JAMA 2007;297:1344−53.
[31] Denes P, Larson JC, Lloyd-Jones DM, Prineas RJ, Greenland P. Major and minor ECG abnormalities in asymptomatic women and risk of cardiovascular events and mortality. JAMA 2007;297:978−85.
[32] Gardner CD, Kiazand A, Alhassan S, Kim S, Stafford RS, Balise RR, et al. Comparison of the Atkins, Zone, Ornish, and LEARN diets for change in weight and related risk factors among overweight premenopausal women: the A TO Z Weight Loss Study: a randomized trial. JAMA 2007;297:969−77.
[33] Schnurr PP, Friedman MJ, Engel CC, Foa EB, Shea MT, Chow BK, et al. Cognitive behavioral therapy for posttraumatic stress disorder in women: a randomized controlled trial. JAMA 2007;297:820−30.
[34] van Dijk D, Spoor M, Hijman R, Nathoe HM, Borst C, Jansen EW, et al. Octopus Study Group. Cognitive and cardiac outcomes 5 years after off-pump vs on-pump coronary artery bypass graft surgery. JAMA 2007;297:701−8.
[35] Zacharski LR, Chow BK, Howes PS, Shamayeva G, Baron JA, Dalman RL, et al. Reduction of iron stores and cardiovascular outcomes in patients with peripheral arterial disease: a randomized controlled trial. JAMA 2007;297:603−10.
[36] Stone GW, Bertrand ME, Moses JW, Ohman EM, Lincoff AM, Ware JH, et al. ACUITY Investigators. Routine upstream initiation vs deferred selective use of glycoprotein IIb/IIIa inhibitors in acute coronary syndromes: the ACUITY Timing trial. JAMA 2007;297:591−602.
[37] Thiam S, LeFevre AM, Hane F, Ndiaye A, Ba F, Fielding KL, et al. Effectiveness of a strategy to improve adherence to tuberculosis treatment in a resource-poor setting: a cluster randomized controlled trial. JAMA 2007;297:380−6.
[38] Lespérance F, Frasure-Smith N, Koszycki D, Laliberté MA, van Zyl LT, Baker B, et al. CREATE Investigators. Effects of citalopram and interpersonal psychotherapy on depression in patients with coronary artery disease: the Canadian Cardiac Randomized Evaluation of Antidepressant and Psychotherapy Efficacy (CREATE) trial. JAMA 2007;297:367−79.
[39] Oettle H, Post S, Neuhaus P, Gellert K, Langrehr J, Ridwelski K, et al. Adjuvant chemotherapy with gemcitabine vs observation in patients undergoing curative-intent resection of pancreatic cancer: a randomized controlled trial. JAMA 2007;297:267−77.

[40] Armstrong PW, Granger CB, Adams PX, Hamm C, Holmes D Jr, O'Neill WW, et al. APEX AMI Investigators. Pexelizumab for acute ST-elevation myocardial infarction in patients undergoing primary percutaneous coronary intervention: a randomized controlled trial. JAMA 2007;297:43−51.

[41] Strasser H, Marksteiner R, Margreiter E, Pinggera GM, Mitterberger M, Frauscher F, et al. Autologous myoblasts and fibroblasts versus collagen for treatment of stress urinary incontinence in women: a randomised controlled trial. Lancet 2007;369:2179−86.

[42] Paavonen J, Jenkins D, Bosch FX, Naud P, Salmerón J, Wheeler CM, et al. HPV PATRICIA Study Group. Efficacy of a prophylactic adjuvanted bivalent L1 virus-like-particle vaccine against infection with human papillomavirus types 16 and 18 in young women: an interim analysis of a phase III double-blind, randomised controlled trial. Lancet 2007;369:2161−70.

[43] Darboe MK, Thurnham DI, Morgan G, Adegbola RA, Secka O, Solon JA, et al. Effectiveness of an early supplementation scheme of high-dose vitamin A versus standard WHO protocol in Gambian mothers and infants: a randomised controlled trial. Lancet 2007;369:2088−96.

[44] Solomon SD, Janardhanan R, Verma A, Bourgoun M, Daley WL, Purkayastha D, et al. Valsartan In Diastolic Dysfunction (VALIDD) Investigators. Effect of angiotensin receptor blockade and antihypertensive drugs on diastolic function in patients with hypertension and diastolic dysfunction: a randomised trial. Lancet 2007;369: 2079−87.

[45] Khan MS, Dar O, Sismanidis C, Shah K, Godfrey-Faussett P. Improvement of tuberculosis case detection and reduction of discrepancies between men and women by simple sputum-submission instructions: a pragmatic randomised controlled trial. Lancet 2007;369:1955−60.

[46] von Hertzen H, Piaggio G, Huong NT, Arustamyan K, Cabezas E, Gomez M, et al. WHO Research Group on Postovulatory Methods of Fertility Regulation. Efficacy of two intervals and two routes of administration of misoprostol for termination of early pregnancy: a randomised controlled equivalence trial. Lancet 2007;369:1938−46.

[47] Gilligan D, Nicolson M, Smith I, Groen H, Dalesio O, Goldstraw P, et al. Preoperative chemotherapy in patients with resectable non-small cell lung cancer: results of the MRC LU22/NVALT 2/EORTC 08012 multicentre randomised trial and update of systematic review. Lancet 2007;369:1929−37.

[48] Andang'o PE, Osendarp SJ, Ayah R, West CE, Mwaniki DL, De Wolf CA, et al. Efficacy of iron-fortified whole maize flour on iron status of schoolchildren in Kenya: a randomised controlled trial. Lancet 2007;369:1799−806.

[49] Chan FK, Wong VW, Suen BY, Wu JC, Ching JY, Hung LC, et al. Combination of a cyclo-oxygenase-2 inhibitor and a proton-pump inhibitor for prevention of recurrent ulcer bleeding in patients at very high risk: a double-blind, randomised trial. Lancet 2007;369: 1621−6.

[50] Griffiths C, Sturdy P, Brewin P, Bothamley G, Eldridge S, Martineau A, et al. Educational outreach to promote screening for tuberculosis in primary care: a cluster randomised controlled trial. Lancet 2007;369:1528−34.

[51] Kuse ER, Chetchotisakd P, da Cunha CA, Ruhnke M, Barrios C, Raghunadharao D, et al. Micafungin Invasive Candidiasis Working Group. Micafungin versus liposomal amphotericin B for candidaemia and invasive candidosis: a phase III randomised double-blind trial. Lancet 2007;369:1519−27.

[52] Mochizuki S, Dahlöf B, Shimizu M, Ikewaki K, Yoshikawa M, Taniguchi I, et al. Jikei Heart Study Group. Valsartan in a Japanese population with hypertension and other cardiovascular disease (Jikei Heart Study): a randomised, open-label, blinded endpoint morbidity-mortality study. Lancet 2007;369:1431−9.

[53] Sherman DG, Albers GW, Bladin C, Fieschi C, Gabbai AA, Kase CS, et al. PREVAIL Investigators. The efficacy and safety of enoxaparin versus unfractionated heparin for the prevention of venous thromboembolism after acute ischaemic stroke (PREVAIL Study): an open-label randomised comparison. Lancet 2007;369: 1347−55.

[54] Grinsztejn B, Nguyen BY, Katlama C, Gatell JM, Lazzarin A, Vittecoq D, et al. Protocol 005 Team. Safety and efficacy of the HIV-1 integrase inhibitor raltegravir (MK-0518) in treatment-experienced patients with multidrug-resistant virus: a phase II randomised controlled trial. Lancet 2007;369:1261−9.

[55] Clotet B, Bellos N, Molina JM, Cooper D, Goffard JC, Lazzarin A, et al. POWER 1 and 2 Study Groups. Efficacy and safety of darunavir-ritonavir at week 48 in treatment-experienced patients with HIV-1 infection in POWER 1 and 2: a pooled subgroup analysis of data from two randomised trials. Lancet 2007;369:1169−78.

[56] Yokoyama M, Origasa H, Matsuzaki M, Matsuzawa Y, Saito Y, Ishikawa Y, et al. Japan EPA Lipid Intervention Study (JELIS) Investigators. Effects of eicosapentaenoic acid on major coronary events in hypercholesterolaemic patients (JELIS): a randomised open-label, blinded endpoint analysis. Lancet 2007;369:1090−8.

[57] François B, Bellissant E, Gissot V, Desachy A, Normand S, Boulain T, et al. Association des Réanimateurs du Centre-Ouest (ARCO). 12-h pretreatment with methylprednisolone versus placebo for prevention of postextubation laryngeal oedema: a randomised double-blind trial. Lancet 2007;369:1083−9.

[58] Marson AG, Al-Kharusi AM, Alwaidh M, Appleton R, Baker GA, Chadwick DW, et al. SANAD Study Group. The SANAD study of effectiveness of valproate, lamotrigine, or topiramate for generalised and unclassifiable epilepsy: an unblinded randomised controlled trial. Lancet 2007;369:1016−26.

[59] Marson AG, Al-Kharusi AM, Alwaidh M, Appleton R, Baker GA, Chadwick DW, et al. SANAD Study Group. The SANAD study of effectiveness of carbamazepine, gabapentin, lamotrigine, oxcarbazepine, or topiramate for treatment of partial epilepsy: an unblinded randomised controlled trial. Lancet 2007;369:1000−15.

[60] Sazawal S, Black RE, Ramsan M, Chwaya HM, Dutta A, Dhingra U, et al. Effect of zinc supplementation on mortality in children aged 1-48 months: a community-based randomised placebo-controlled trial. Lancet 2007;369:927−34.

[61] Stone GW, White HD, Ohman EM, Bertrand ME, Lincoff AM, McLaurin BT, et al. Acute Catheterization and Urgent Intervention Triage strategy (ACUITY) Trial Investigators. Bivalirudin in patients with acute coronary syndromes undergoing percutaneous coronary intervention: a subgroup analysis from the Acute Catheterization and Urgent Intervention Triage strategy (ACUITY) trial. Lancet 2007;369:907−19.

[62] Nadel S, Goldstein B, Williams MD, Dalton H, Peters M, Macias WL, et al. REsearching severe Sepsis and Organ dysfunction in children: a gLobal perspective (RESOLVE) Study Group. Drotrecogin alfa (activated) in children with severe sepsis: a multicentre phase III randomised controlled trial. Lancet 2007;369:836−43.

[63] Hirsch A, Windhausen F, Tijssen JG, Verheugt FW, Cornel JH, de Winter RJ Invasive versus Conservative Treatment in Unstable coronary Syndromes (ICTUS) Investigators. Long-term outcome after an early invasive versus selective invasive treatment strategy in patients with non-ST-elevation acute coronary syndrome and elevated cardiac troponin T (the ICTUS trial): a follow-up study. Lancet 2007;369:827−35.

[64] Ratcliff A, Siswantoro H, Kenangalem E, Maristela R, Wuwung RM, Laihad F, et al. Two fixed-dose artemisinin combinations for drug-resistant falciparum and vivax malaria in Papua, Indonesia: an open-label randomised comparison. Lancet 2007;369:757−65.

[65] Heijnen EM, Eijkemans MJ, De Klerk C, Polinder S, Beckers NG, Klinkert ER, et al. A mild treatment strategy for in-vitro fertilisation: a randomised non-inferiority trial. Lancet 2007;369:743−9.

[66] Gray RH, Kigozi G, Serwadda D, Makumbi F, Watya S, Nalugoda F, et al. Male circumcision for HIV prevention in men in Rakai, Uganda: a randomised trial. Lancet 2007;369:657−66.

[67] Bailey RC, Moses S, Parker CB, Agot K, Maclean I, Krieger JN, et al. Male circumcision for HIV prevention in young men in Kisumu, Kenya: a randomised controlled trial. Lancet 2007;369:643−56.
[68] Coombes RC, Kilburn LS, Snowdon CF, Paridaens R, Coleman RE, Jones SE, et al. Intergroup Exemestane Study. Survival and safety of exemestane versus tamoxifen after 2-3 years' tamoxifen treatment (Intergroup Exemestane Study): a randomised controlled trial. Lancet 2007;369:559−70.
[69] Zongo I, Dorsey G, Rouamba N, Tinto H, Dokomajilar C, Guiguemde RT, et al. Artemether-lumefantrine versus amodiaquine plus sulfadoxine-pyrimethamine for uncomplicated falciparum malaria in Burkina Faso: a randomised non-inferiority trial. Lancet 2007;369:491−8.
[70] Malhotra-Kumar S, Lammens C, Coenen S, Van Herck K, Goossens H. Effect of azithromycin and clarithromycin therapy on pharyngeal carriage of macrolide-resistant streptococci in healthy volunteers: a randomised, double-blind, placebo-controlled study. Lancet 2007;369:482−90.
[71] Durga J, van Boxtel MP, Schouten EG, Kok FJ, Jolles J, Katan MB, et al. Effect of 3-year folic acid supplementation on cognitive function in older adults in the FACIT trial: a randomised, double blind, controlled trial. Lancet 2007;369:208−16.
[72] Conter V, Valsecchi MG, Silvestri D, Campbell M, Dibar E, Magyarosy E, et al. Pulses of vincristine and dexamethasone in addition to intensive chemotherapy for children with intermediate-risk acute lymphoblastic leukaemia: a multicentre randomised trial. Lancet 2007;369:123−31.
[73] Smith I, Procter M, Gelber RD, Guillaume S, Feyereislova A, Dowsett M, et al. HERA Study Team. 2-year follow-up of trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer: a randomised controlled trial. Lancet 2007;369:29−36.
[74] Home PD, Pocock SJ, Beck-Nielsen H, Gomis R, Hanefeld M, Jones NP, et al. RECORD Study Group. Rosiglitazone evaluated for cardiovascular outcomes—an interim analysis. N Engl J Med 2007;357:28−38.
[75] Manson JE, Allison MA, Rossouw JE, Carr JJ, Langer RD, Hsia J, et al. WHI and WHI-CACS Investigators. Estrogen therapy and coronary-artery calcification. N Engl J Med 2007;356:2591−602.
[76] Sundar S, Jha TK, Thakur CP, Sinha PK, Bhattacharya SK. Injectable paromomycin for visceral leishmaniasis in India. N Engl J Med 2007;356:2571−81.
[77] Manzoni P, Stolfi I, Pugni L, Decembrino L, Magnani C, Vetrano G, et al. Italian Task Force for the Study and Prevention of Neonatal Fungal Infections. Italian Society of Neonatology. A multicenter, randomized trial of prophylactic fluconazole in preterm neonates. N Engl J Med 2007;356:2483−95.
[78] Reboli AC, Rotstein C, Pappas PG, Chapman SW, Kett DH, Kumar D, et al. Anidulafungin Study Group. Anidulafungin versus fluconazole for invasive candidiasis. N Engl J Med 2007;356:2472−82.
[79] Dember LM, Hawkins PN, Hazenberg BP, Gorevic PD, Merlini G, Butrimiene I, et al. Eprodisate for AA Amyloidosis Trial Group. Eprodisate for the treatment of renal disease in AA amyloidosis. N Engl J Med 2007;356:2349−60.
[80] Hudes G, Carducci M, Tomczak P, Dutcher J, Figlin R, Kapoor A, et al. Global ARCC Trial. Temsirolimus, interferon alfa, or both for advanced renal-cell carcinoma. N Engl J Med 2007;356:2271−81.
[81] Weinstein JN, Lurie JD, Tosteson TD, Hanscom B, Tosteson AN, Blood EA, et al. Surgical versus nonsurgical treatment for lumbar degenerative spondylolisthesis. N Engl J Med 2007;356:2257−70.
[82] Peul WC, van Houwelingen HC, van den Hout WB, Brand R, Eekhof JA, Tans JT, et al. Leiden-The Hague Spine Intervention Prognostic Study Group. Surgery versus prolonged conservative treatment for sciatica. N Engl J Med 2007;356:2245−56.
[83] Albo ME, Richter HE, Brubaker L, Norton P, Kraus SR, Zimmern PE, et al. Urinary Incontinence Treatment Network. Burch colposuspension versus fascial sling to reduce urinary stress incontinence. N Engl J Med 2007;356:2143−55.
[84] Papi A, Canonica GW, Maestrelli P, Paggiaro P, Olivieri D, Pozzi E, et al. BEST Study Group. Rescue use of beclomethasone and albuterol in a single inhaler for mild asthma. N Engl J Med 2007;356:2040−52.
[85] Peters SP, Anthonisen N, Castro M, Holbrook JT, Irvin CG, Smith LJ, et al. American Lung Association Asthma Clinical Research Centers. Randomized comparison of strategies for reducing treatment in mild persistent asthma. N Engl J Med 2007;356:2027−39.
[86] Garland SM, Hernandez-Avila M, Wheeler CM, Perez G, Harper DM, Leodolter S, et al. Females United to Unilaterally Reduce Endo/Ectocervical Disease (FUTURE) I Investigators. Quadrivalent vaccine against human papillomavirus to prevent anogenital diseases. N Engl J Med 2007;356:1928−43.
[87] FUTURE II Study Group. Quadrivalent vaccine against human papillomavirus to prevent high-grade cervical lesions. N Engl J Med 2007;356:1915−27.
[88] Jacobson AM, Musen G, Ryan CM, Silvers N, Cleary P, Waberski B, et al. Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications Study Research Group. Long-term effect of diabetes and its treatment on cognitive function. N Engl J Med 2007;356:1842−52.
[89] Sezer M, Oflaz H, Gören T, Okçular I, Umman B, Nişanci Y, et al. Intracoronary streptokinase after primary percutaneous coronary intervention. N Engl J Med 2007;356:1823−34.
[90] Black DM, Delmas PD, Eastell R, Reid IR, Boonen S, Cauley JA, et al. HORIZON Pivotal Fracture Trial. Once-yearly zoledronic acid for treatment of postmenopausal osteoporosis. N Engl J Med 2007;356:1809−22.
[91] Sachs GS, Nierenberg AA, Calabrese JR, Marangell LB, Wisniewski SR, Gyulai L, et al. Effectiveness of adjunctive antidepressant treatment for bipolar depression. N Engl J Med 2007;356:1711−22.
[92] Lau JY, Leung WK, Wu JC, Chan FK, Wong VW, Chiu PW, et al. Omeprazole before endoscopy in patients with gastrointestinal bleeding. N Engl J Med 2007;356:1631−40.
[93] Lacroix J, Hébert PC, Hutchison JS, Hume HA, Tucci M, Ducruet T, et al. TRIPICU Investigators; Canadian Critical Care Trials Group; Pediatric Acute Lung Injury and Sepsis Investigators Network. Transfusion strategies for patients in pediatric intensive care units. N Engl J Med 2007;356:1609−19.
[94] Kastelein JJ, van Leuven SI, Burgess L, Evans GW, Kuivenhoven JA, Barter PJ, et al. RADIANCE 1 Investigators. Effect of torcetrapib on carotid atherosclerosis in familial hypercholesterolemia. N Engl J Med 2007;356:1620−30.
[95] Cuba IPV Study Collaborative Group. Randomized, placebo-controlled trial of inactivated poliovirus vaccine in Cuba. N Engl J Med 2007;356:1536−44.
[96] Keime-Guibert F, Chinot O, Taillandier L, Cartalat-Carel S, Frenay M, Kantor G, et al. Association of French-Speaking Neuro--Oncologists. Radiotherapy for glioblastoma in the elderly. N Engl J Med 2007;356:1527−35.
[97] Larsen CM, Faulenbach M, Vaag A, Vølund A, Ehses JA, Seifert B, et al. Interleukin-1-receptor antagonist in type 2 diabetes mellitus. N Engl J Med 2007;356:1517−26.
[98] Boden WE, O'Rourke RA, Teo KK, Hartigan PM, Maron DJ, Kostuk WJ, et al. COURAGE Trial Research Group. Optimal medical therapy with or without PCI for stable coronary disease. N Engl J Med 2007;356:1503−16.
[99] Fawzi WW, Msamanga GI, Urassa W, Hertzmark E, Petraro P, Willett WC, et al. Vitamins and perinatal outcomes among HIV-negative women in Tanzania. N Engl J Med 2007;356:1423−31.
[100] Cox G, Thomson NC, Rubin AS, Niven RM, Corris PA, Siersted HC, et al. AIR Trial Study Group. Asthma control during the year after bronchial thermoplasty. N Engl J Med 2007;356:1327−37.

[101] Nissen SE, Tardif JC, Nicholls SJ, Revkin JH, Shear CL, Duggan WT, et al. ILLUSTRATE Investigators. Effect of torcetrapib on the progression of coronary atherosclerosis. N Engl J Med 2007;356:1304−16.

[102] Tonetti MS, D'Aiuto F, Nibali L, Donald A, Storry C, Parkar M, et al. Treatment of periodontitis and endothelial function. N Engl J Med 2007;356:911−20.

[103] Shrestha MP, Scott RM, Joshi DM, Mammen MP Jr, Thapa GB, Thapa N, et al. Safety and efficacy of a recombinant hepatitis E vaccine. N Engl J Med 2007;356:895−903.

[104] Nagot N, Ouédraogo A, Foulongne V, Konaté I, Weiss HA, Vergne L, et al. ANRS 1285 Study Group. Reduction of HIV-1 RNA levels with therapy to suppress herpes simplex virus. N Engl J Med 2007;356:790−9.

[105] Calverley PM, Anderson JA, Celli B, Ferguson GT, Jenkins C, Jones PW, et al. TORCH Investigators. Salmeterol and fluticasone propionate and survival in chronic obstructive pulmonary disease. N Engl J Med 2007;356:775−89.

[106] Belshe RB, Edwards KM, Vesikari T, Black SV, Walker RE, Hultquist M, et al. CAIV-T Comparative Efficacy Study Group. Live attenuated versus inactivated influenza vaccine in infants and young children. N Engl J Med 2007;356:685−96.

[107] Cahen DL, Gouma DJ, Nio Y, Rauws EA, Boermeester MA, Busch OR, et al. Endoscopic versus surgical drainage of the pancreatic duct in chronic pancreatitis. N Engl J Med 2007;356:676−84.

[108] Newburger JW, Sleeper LA, McCrindle BW, Minich LL, Gersony W, Vetter VL, et al. Pediatric Heart Network Investigators. Randomized trial of pulsed corticosteroid therapy for primary treatment of Kawasaki disease. N Engl J Med 2007;356:663−75.

[109] Krueger GG, Langley RG, Leonardi C, Yeilding N, Guzzo C, Wang Y, et al. CNTO 1275 Psoriasis Study Group. A human interleukin-12/23 monoclonal antibody for the treatment of psoriasis. N Engl J Med 2007;356:580−92.

[110] Legro RS, Barnhart HX, Schlaff WD, Carr BR, Diamond MP, Carson SA, et al. Cooperative Multicenter Reproductive Medicine Network. Clomiphene, metformin, or both for infertility in the polycystic ovary syndrome. N Engl J Med 2007;356:551−66.

[111] Lautrette A, Darmon M, Megarbane B, Joly LM, Chevret S, Adrie C, et al. A communication strategy and brochure for relatives of patients dying in the ICU. N Engl J Med 2007;356:469−78.

[112] Kuhle J, Pohl C, Mehling M, Edan G, Freedman MS, Hartung HP, et al. Lack of association between antimyelin antibodies and progression to multiple sclerosis. N Engl J Med 2007;356:371−8.

[113] Edwards RD, Moss JG, Lumsden MA, Wu O, Murray LS, Twaddle S, et al. Committee of the Randomized Trial of Embolization versus Surgical Treatment for Fibroids. Uterine-artery embolization versus surgery for symptomatic uterine fibroids. N Engl J Med 2007;356:360−70.

[114] Cornely OA, Maertens J, Winston DJ, Perfect J, Ullmann AJ, Walsh TJ, et al. Posaconazole vs. fluconazole or itraconazole prophylaxis in patients with neutropenia. N Engl J Med 2007;356:348−59.

[115] Ullmann AJ, Lipton JH, Vesole DH, Chandrasekar P, Langston A, Tarantolo SR, et al. Posaconazole or fluconazole for prophylaxis in severe graft-versus-host disease. N Engl J Med 2007;356:335−47.

[116] Paradise JL, Feldman HM, Campbell TF, Dollaghan CA, Rockette HE, Pitcairn DL, et al. Tympanostomy tubes and developmental outcomes at 9 to 11 years of age. N Engl J Med 2007;356:248−61.

[117] Lockman S, Shapiro RL, Smeaton LM, Wester C, Thior I, Stevens L, et al. Response to antiretroviral therapy after a single, peripartum dose of nevirapine. N Engl J Med 2007;356:135−47.

[118] Escudier B, Eisen T, Stadler WM, Szczylik C, Oudard S, Siebels M, et al. TARGET Study Group. Sorafenib in advanced clear-cell renal-cell carcinoma. N Engl J Med 2007;356:125−34.

[119] Motzer RJ, Hutson TE, Tomczak P, Michaelson MD, Bukowski RM, Rixe O, et al. Sunitinib versus interferon alfa in metastatic renal-cell carcinoma. N Engl J Med 2007;356:115−24.

[120] Hickson M, D'Souza AL, Muthu N, Rogers TR, Want S, Rajkumar C, et al. Use of probiotic Lactobacillus preparation to prevent diarrhoea associated with antibiotics: randomised double blind placebo controlled trial. BMJ 2007;335:80.

[121] Farmer A, Wade A, Goyder E, Yudkin P, French D, Craven A, et al. Impact of self monitoring of blood glucose in the management of patients with non-insulin treated diabetes: open parallel group randomised trial. BMJ 2007;335:132.

[122] Goodyer I, Dubicka B, Wilkinson P, Kelvin R, Roberts C, Byford S, et al. Selective serotonin reuptake inhibitors (SSRIs) and routine specialist care with and without cognitive behaviour therapy in adolescents with major depression: randomised controlled trial. BMJ 2007;335:142.

[123] Gohel MS, Barwell JR, Taylor M, Chant T, Foy C, Earnshaw JJ, et al. Long term results of compression therapy alone versus compression plus surgery in chronic venous ulceration (ESCHAR): randomised controlled trial. BMJ 2007;335:83.

[124] Montgomery AA, Emmett CL, Fahey T, Jones C, Ricketts I, Patel RR, et al. DiAMOND Study Group. Two decision aids for mode of delivery among women with previous caesarean section: randomised controlled trial. BMJ 2007;334:1305.

[125] Ranson MK, Sinha T, Chatterjee M, Gandhi F, Jayswal R, Patel F, et al. Equitable utilisation of Indian community based health insurance scheme among its rural membership: cluster randomised controlled trial. BMJ 2007;334:1309.

[126] Ronco G, Cuzick J, Pierotti P, Cariaggi MP, Dalla Palma P, Naldoni C, et al. Accuracy of liquid based versus conventional cytology: overall results of new technologies for cervical cancer screening: randomised controlled trial. BMJ 2007;335:28.

[127] Duffy M, Gillespie K, Clark DM. Post-traumatic stress disorder in the context of terrorism and other civil conflict in Northern Ireland: randomised controlled trial. BMJ 2007;334:1147.

[128] Kang JH, Cook N, Manson J, Buring JE, Grodstein F. Low dose aspirin and cognitive function in the women's health study cognitive cohort. BMJ 2007;334:987.

[129] Holland R, Brooksby I, Lenaghan E, Ashton K, Hay L, Smith R, et al. Effectiveness of visits from community pharmacists for patients with heart failure: HeartMed randomised controlled trial. BMJ 2007;334:1098.

[130] de Groot M, de Keijser J, Neeleman J, Kerkhof A, Nolen W, Burger H. Cognitive behaviour therapy to prevent complicated grief among relatives and spouses bereaved by suicide: cluster randomised controlled trial. BMJ 2007;334:994.

[131] Salter C, Holland R, Harvey I, Henwood K. ''I haven't even phoned my doctor yet.'' The advice giving role of the pharmacist during consultations for medication review with patients aged 80 or more: qualitative discourse analysis. BMJ 2007;334:1101.

[132] Hutchings J, Gardner F, Bywater T, Daley D, Whitaker C, Jones K, et al. Parenting intervention in Sure Start services for children at risk of developing conduct disorder: pragmatic randomised controlled trial. BMJ 2007;334:678.

[133] Edwards RT, Céilleachair A, Bywater T, Hughes DA, Hutchings J. Parenting programme for parents of children at risk of developing conduct disorder: cost effectiveness analysis. BMJ 2007;334:682.

[134] Alho OP, Koivunen P, Penna T, Teppo H, Koskela M, Luotonen J. Tonsillectomy versus watchful waiting in recurrent streptococcal pharyngitis in adults: randomised controlled trial. BMJ 2007;334:939.

[135] Howden-Chapman P, Matheson A, Crane J, Viggers H, Cunningham M, Blakely T, et al. Effect of insulating existing houses on health inequality: cluster randomised study in the community. BMJ 2007;334:460.

[136] Mutrie N, Campbell AM, Whyte F, McConnachie A, Emslie C, Lee L, et al. Benefits of supervised group exercise programme for women being treated for early stage breast cancer: pragmatic randomised controlled trial. BMJ 2007;334:517.

[137] Campbell IA, Bentley DP, Prescott RJ, Routledge PA, Shetty HG, Williamson IJ. Anticoagulation for three versus six months in

patients with deep vein thrombosis or pulmonary embolism, or both: randomised trial. BMJ 2007;334:674.
[138] Bech BH, Obel C, Henriksen TB, Olsen J. Effect of reducing caffeine intake on birth weight and length of gestation: randomised controlled trial. BMJ 2007;334:409.
[139] Reyburn H, Mbakilwa H, Mwangi R, Mwerinde O, Olomi R, Drakeley C, et al. Rapid diagnostic tests compared with malaria microscopy for guiding outpatient treatment of febrile illness in Tanzania: randomised trial. BMJ 2007;334:403.
[140] Petrie KJ, Müller JT, Schirmbeck F, Donkin L, Broadbent E, Ellis CJ, et al. Effect of providing information about normal test results on patients' reassurance: randomised controlled trial. BMJ 2007;334:352.
[141] Banu SH, Jahan M, Koli UK, Ferdousi S, Khan NZ, Neville B. Side effects of phenobarbital and carbamazepine in childhood epilepsy: randomised controlled trial. BMJ 2007;334:1207.
[142] Koh TH, Butow PN, Coory M, Budge D, Collie LA, Whitehall J, et al. Provision of taped conversations with neonatologists to mothers of babies in intensive care: randomised controlled trial. BMJ 2007;334:28.
[143] Sazawal S, Dhingra U, Dhingra P, Hiremath G, Kumar J, Sarkar A, et al. Effects of fortified milk on morbidity in young children in north India: community based, randomised, double masked placebo controlled trial. BMJ 2007;334:140.
[144] Henderson M, Wight D, Raab GM, Abraham C, Parkes A, Scott S, et al. Impact of a theoretically based sex education programme (SHARE) delivered by teachers on NHS registered conceptions and terminations: final results of cluster randomised trial. BMJ 2007;334:133.
[145] Brown CT, Yap T, Cromwell DA, Rixon L, Steed L, Mulligan K, et al. Self management for men with lower urinary tract symptoms: randomised controlled trial. BMJ 2007;334:25.
[146] Zar HJ, Cotton MF, Strauss S, Karpakis J, Hussey G, Schaaf HS, et al. Effect of isoniazid prophylaxis on mortality and incidence of tuberculosis in children with HIV: randomised controlled trial. BMJ 2007;334:136.
[147] Roberts C, Torgerson DJ. Understanding controlled trials: baseline imbalance in randomised controlled trials. BMJ 1999;319:185.
[148] European Agency for the Evaluation of Medicinal Products (EMEA)-Committee for Proprietary Medical Products (CPMP). Points to consider on adjustment for baseline covariates. Available at: http://www.emea.europa.eu/pdfs/human/ewp/286399en.pdf. 2003; Accessed October 3, 2008.
[149] Fayers P, King M. The baseline characteristics did not differ significantly. Qual Life Res 2008;17:1047−8.
[150] Pocock SJ, Assmann SE, Enos LE, Kasten LE. Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems. Stat Med 2002;21:2917−30.
[151] Gail MH, Wieand S, Piantadosi S. Biased estimates of treatment effect in randomized experiments with nonlinear regressions and omitted covariates. Biometrika 1984;7:431−44.
[152] Greenland S. Interpretation and choice of effect measures in epidemiologic analyses. Am J Epidemiol 1987;125:761−8.
[153] Greenland S, Robins JM, Pearl J. Confounding and collapsibility in causal inference. Stat Sci 1999;14:29−46.
[154] Newcombe RG. A deficiency of the odds ratio as a measure of effect size. Stat Med 2006;25:4235−40.
[155] Hauck WW, Anderson S, Marcus SM. Should we adjust for covariates in nonlinear regression analyses of randomized trials? Control Clin Trials 1998;19:249−56.
[156] Martens EP, Pestman WR, Klungel OH. Conditioning on the propensity score can result in biased estimation of common measures of treatment effect: a Monte Carlo study. (p n/a) by Peter C. Austin, Paul Grootendorst, Sharon-Lise T. Normand, Geoffrey M. Anderson, Stat Med, Published Online: June 16, 2006. DOI: 10.1002/sim. 2618. Stat Med 2007;26:3208−10.
[157] Austin PC. Absolute risk reductions, relative risks, and numbers needed to treat can be obtained from a logistic regression model. J Clin Epidemiol 2010;63:2−6.
[158] Austin PC. Absolute risk reductions and numbers needed to treat can be obtained from adjusted survival models for time-to-event outcomes. J Clin Epidemiol. 2010;63:46−55.
[159] Shrier I, Platt RW, Steele RJ. Re: ''Variable selection for propensity score models''. Am J Epidemiol 2007;166:238−89.
[160] Ades AE, Sutton AJ. Multiparameter evidence synthesis in epidemiology and medical decision-making: current approaches. J R Stat Soc Ser A 2006;169:5−35.
[161] Chan AW, Hrobjartsson A, Jorgensen KJ, Gotzsche PC, Altman DG. Discrepancies in sample size calculations and data analyses reported in randomised trials: comparison of publications with protocols. BMJ 2008;337:a2299. DOI: 10.1136/bmj.a2299.