THEORY AND METHODS

Investigating patient exclusion bias in meta-analysis

Jayne F Tierney and Lesley A Stewart

Accepted 6 July 2004

Background Trial investigators frequently exclude patients from trial analyses which may bias estimates of the effect of treatment. Combining these estimates in a meta-analysis could aggregate any such biases.

Methods To investigate how excluding patients from trials can affect the results of both trials and meta-analyses, we used 14 meta-analyses of individual patient data (IPD) that addressed therapeutic questions in cancer. These included 133 randomized controlled trials (RCT) and 21,905 patients. We explored whether exclusions were related to trial characteristics and categorized the reasons for exclusions. For each RCT and meta-analysis, we compared results of an intention-to-treat analysis of all randomized patients with an analysis based on those patients included in the investigators’ analysis.

Results In all, 92 trials (69%) excluded between 0.3 and 38% of patients randomized. Trials excluding patients tended to be older and larger than those that did not. Most patients were excluded because of ineligibility or protocol violations. Exclusions varied substantially by meta-analysis, more patients tending to be excluded from the treatment arm. Comparing trial analyses there was no clear indication that exclusion of patients altered the results more in favour of either treatment or control. However, comparing meta-analysis results, there was a tendency for those based on ‘included’ patients to favour the research treatment ($P = 0.03$). Inconsistency of trial results was often increased as a result of the investigators’ exclusions.

Conclusions Trials, systematic reviews, and meta-analyses may be prone to bias associated with post-randomization exclusion of patients. Wherever possible, the level of such exclusions should be taken into account when assessing the potential for bias in trials, systematic reviews, and meta-analyses. Ideally, trials, systematic reviews, and meta-analyses should be based on all randomized patients.

Keywords Meta-analysis, systematic review, randomized controlled trial, exclusion bias, intention to treat, individual patient data

Random allocation of patients in a clinical trial should ensure the unbiased assignment and comparison of therapeutic groups. However, this can only be guaranteed if all randomized patients are analysed according to the treatments initially assigned; an intention-to-treat approach. Yet trial investigators frequently exclude patients from trial analyses sometimes for subjective, if well-intentioned reasons, which may threaten the validity of the treatment comparison (Box 1). Some argue that ineligible patients and sometimes those who never received the intervention may be excluded without introducing bias into the analysis of a trial, provided they are excluded in a blinded fashion. However, even if blinding is feasible, the subsequent analysis may not accurately reflect the impact of an intervention in practice. If patients are excluded from analyses for reasons that are related to treatment and outcome, this may bias estimates of the effect of treatment. For example, if patients died from toxic effects of an experimental treatment and were subsequently excluded from the trial analysis, perhaps as ‘early deaths’, the estimate of effect would be biased in favour of that experimental treatment. Moreover, combining trial estimates in a meta-analysis could aggregate any such biases. Indeed there is some evidence to suggest that exclusion of patients contributes to differences between results of meta-analyses of individual...
patient data (IPD) and meta-analyses of published aggregate data. However, we have found no other empirical studies comparing results of meta-analyses including all randomized patients with just those included in investigators’ analyses.

In our protocols for systematic reviews and meta-analyses of IPD, we explicitly request information on all randomized patients,7,8 even those excluded from the investigators’ own analyses and, where possible, brief reasons for these exclusions. Obviously, this relies on data from all randomized patients being available, sometimes many years after completion of a trial and on investigators being willing to provide us with such data. If we are able to obtain these excluded patients, and to date we have been very successful in doing so, we can then carry out an intention-to-treat analysis of all randomized patients, as standard. This is the most robust and least biased approach. No value judgements are made as to whether individual patient exclusions are appropriate. Indeed, this would only be possible with very detailed knowledge of the individual trials. IPD also provides us with an opportunity to investigate the potential influence of excluding patients in individual trials and in meta-analyses. Furthermore, because we collect all trials (published and unpublished), update follow-up and collect the same outcomes for each trial, other potential sources of bias such as publication bias, follow-up bias, and selective (outcome) reporting bias are limited (Box 1). We have examined the effect of excluding patients after randomization on the results of our group’s completed and published IPD meta-analyses in cancer and the trials they included.

Methods

We used 14 systematic reviews and meta-analyses of IPD from randomized controlled trials (RCT) that addressed therapeutic questions in cancer.9–16 These meta-analyses varied by cancer site (bladder, brain, lung, oesophagus, ovary, lung, and soft tissue sarcoma); start of recruitment (1966–1989); end of recruitment (1972–1995); the type and timing of the experimental treatment; the number of trials (4–19); the number of patients (479–3146); the endpoints examined (survival and recurrence), and the direction of the overall treatment effect (hazard ratios [HR] = 0.68–1.21, Table 1).

We investigated whether exclusion of patients was related to the year a trial started, the year a trial ended, the year of publication and the number of patients recruited. Furthermore, for those trials with exclusions, we investigated whether the proportion of patients excluded was related to these same trial characteristics. As these variables were somewhat skewed even if transformed, we used Mann Whitney U tests and Spearman Rank Correlations respectively to test these associations.

Where investigators provided reasons for patient exclusions, we classified them into six broad categories. These were: (1) ineligibility fixed at randomization e.g. age out of range; (2) ineligibility re-assessed after randomization e.g. pathological confirmation of tumour stage; (3) early outcome; (4) protocol violation; (5) missing data; and (6) exclusion for other reasons. As the reasons given were sometimes insufficiently detailed and/or ambiguous, then sometimes we had to make assumptions when we classified them. Therefore, it is appropriate to present these results only in a semi-quantitative fashion.

For the primary outcome of survival and for each RCT, as far as possible, we carried out (1) an intention-to-treat analysis of all randomized patients and (2) an analysis based on just those patients included in the investigators’ own analysis. The latter generally corresponded to the most recently published or

Box 1: Glossary of terms

**Intention-to-treat analysis**: Including and analysing all randomised patients according to their original treatment allocation, irrespective of whether they actually received that treatment. This preserves the unbiased comparison of treatment groups afforded by randomization.

**Exclusion bias**: Collective term covering the various potential biases that can result from the post-randomization exclusion of patients from a trial and subsequent analyses. This may also be referred to as attrition bias. Some common reasons for excluding patients are:

- **Ineligibility**: Where it is found that patients do not meet the trial eligibility criteria some time after randomization.
- **Protocol violation**: Where patients fail to receive their allocated treatment according to protocol. This can happen if patients are knowingly or mistakenly given a treatment other than their allocated treatment. Protocol violations can also occur because of a clinician’s withdrawal of patients from treatment, because of their poor condition, poor response to treatment, toxicity etc. Similarly, a patient’s lack of compliance or adherence to treatment because of poor condition, toxicity or other personal reasons can be another source of protocol violations.
- **Early outcome**: Where patients experience an outcome prior to, during, or shortly after a course of treatment. The rationale for this is that such outcomes are apparently not attributable to treatment.
- **Loss to follow-up**: Where patients stop contributing outcome data. This may be because they can no longer be contacted, for example, having moved away or because they actively want to drop-out of further participation in the trial. The latter may be related to clinician withdrawal or patient compliance.

Clearly, some of these reasons are related to treatment and/or outcome and others may be applied differentially according to treatment allocation. Thus, exclusion of patients for such reasons can lead to subjectively selected treatment groups and so a biased treatment comparison.

**Publication bias**: Trials with positive or null results are more likely to be published than those with negative results.

**Follow-up bias**: More or better follow-up is sought or available for one treatment group compared to another.

**Selective reporting bias**: Results reported are a biased representation of the full results. For example only those outcomes or subgroups of patients that show positive results are reported. Exclusion bias can be regarded as one type of selective reporting bias.
conducted analysis. Where there were no exclusions only one type of analysis was possible. The log rank expected number of deaths and variance were used to calculate individual trial HRs. Thus, the times to death for individual patients were used within trials to calculate the HR. This represents the risk of an event for those patients allocated to the research treatment, compared with those allocated to control.

We included all trials in the meta-analyses irrespective of whether they excluded patients or not. This allowed us to investigate the impact of exclusions in real situations. Trial HRs were pooled in the meta-analyses using the fixed effect model and the I² statistic was used to assess the degree of inconsistency in results across trials. We compared both trial and meta-analysis results based on all patients, with trial and meta-analysis results based on patients included in the investigators’ analyses. The log scale was used to enable use of parametric, paired t-tests. Also, I² statistics were informally compared. All P-values quoted are two-sided.

Results

Exclusions and trial characteristics

There were 138 trials included in the various meta-analyses, but for five of these the same arms had been used in two separate treatment questions. Of the 133 unique trials involving 21,905 patients, the results of 112 were published in papers or abstracts and the results of 21 were unpublished. There were 41 trials (31%) where investigators did not exclude any patients from their analyses and 7 (17%) of these were unpublished. Of the remaining 92 trials (69%) with patient exclusions, 14 (15%) were unpublished. Between 0.3 and 38% of patients randomized were excluded across these trials. There was also substantial variation in the level of patient exclusions across meta-analyses (Figure 1). Between 4 and 24% of patients randomized were excluded from the research arm and between 3 and 20% of patients randomized were excluded from the control arm. Although the proportions of patients excluded on treatment and control were broadly similar, there was a slight tendency for more patients to be excluded from the treatment arm (Figure 1). For example in the soft tissue sarcoma meta-analysis, 24% of patients were excluded in the treatment arm compared with 20% in the control arm and in the lung PORT meta-analysis, double the proportion of patients were excluded from the treatment (8%) compared with the control arm (4%).

Trials with patient exclusions had a slight tendency to be older, i.e. starting (P = 0.01) or finishing (P = 0.02) recruitment earlier, were published earlier (P = 0.02) and accrued more patients than those without exclusions (P = 0.03, Table 2). However, for the 92 trials that excluded at least one patient, there seemed to be no correlation between the proportion excluded and these same trial characteristics (year trial started, r_s = 0.163, P = 0.12; year trial ended, r_s = 0.072, P = 0.50; publication year, r_s = 0.036, P = 0.75 number of patients recruited, r_s = 0.061, P = 0.57).

Reasons for exclusions

Just fewer than 1800 patients were excluded from the trials in the meta-analyses and we had brief reasons for exclusion for 55% of them. Although sometimes we were forced to make inferences
when categorizing these reasons, some broad trends were apparent (Table 3). More than a third of patients were excluded because eligibility was classified or re-defined some time after randomization. Frequently, this related to the type or extent of disease such as tumour stage, tumour grade, or histological type. Around a quarter of patients were excluded because of protocol violations and sometimes they were simply described as such, while at other times the reasons given were more specific to the research or control treatment. In some cases the underlying cause of the protocol violation was provided e.g. the clinician stopped treatment because of poor patient condition or the patient refused further treatment. Approximately another quarter of patients were excluded for reasons of ineligibility that should have been known at randomization, such as advanced age. The remaining patients were excluded either because data were missing, because they had experienced an outcome prior to treatment, or for other reasons.

### Impact of exclusions

Comparing the trial analyses based on ‘included’ patients with (intention-to-treat) analyses of all randomized patients produced 12 equivalent results, 44 instances of results more in favour of the research treatment, and 36 cases where the results were more in favour of control. There was no clear indication that the exclusion of patients altered the results more in one direction than another ($t = 1.537$, $P = 0.13$). Mostly the differences between the HR were small, with 70% changing by 1 to 10%, but in 17% of trials the differences ranged from 11 to as much as 35% (Figure 2).

In contrast, comparing the 14 pooled meta-analysis results, there was a tendency for the HR for ‘included’ patients to be.
more in favour of the research treatment than the HR based on all patients \( (t = 2.401, P = 0.03, \text{Figure 3}) \). This was irrespective of whether the overall effect was in favour of this treatment or not. These differences tended to be small, between 1 and 5%, and generally might not have altered the interpretation of the results. However, in the soft tissue sarcoma meta-analysis, the HR changed from 0.90 (95% CI: 0.77, 1.04, \( P = 0.157 \))\(^2\) to 0.85, the CI narrowed (95% CI: 0.72, 1.00) and this result approached conventional levels of significance \( (P = 0.06) \), when patients were excluded (Table 4). Such changes could have had an impact on the clinical perception of the results of this particular meta-analysis.

There were eight treatment comparisons where the level of inconsistency was increased as a result of the investigators’ exclusions, two where it was reduced and four where it did not differ (Table 4).
Exclusion of patients from published analyses of randomized trials is widespread, although it is known that it can potentially introduce bias to these analyses and subsequent conclusions. Moreover, in meta-analysis, where enhanced statistical power increases the chance of detecting not only moderate treatment effects, but also moderate biases, greater caution is required. We examined why patients were excluded from trials and explored the potential impact of such exclusions on results at both the trial level and meta-analysis level. There were some instances where we could not obtain patients excluded from the investigators’ analyses or where patients were inadequately followed or lost to follow-up and strictly speaking, this prevented a proper intention-to-treat analysis of all patients for a few trials. Although, this could have introduced bias into our comparison, it is likely to be minimal.

Reasons for exclusions

One of the commonest reasons given for excluding patients was that they were re-assessed after randomization and found to be ineligible. This probably illustrates some of the practical difficulties in classifying tumour type and spread accurately in advance of randomization, particularly when this is done clinically rather than pathologically. In principle, it might be legitimate to conduct an analysis of just those patients who satisfy the eligibility criteria to provide an estimate of effect in the strictly eligible patients. However, such an analysis could only be guaranteed to be free from bias if the data pertaining to patient eligibility were reviewed by an external source, blinded to the treatment allocation. Apart from independent reviews of histology and grade in some of the sarcoma trials, blinded assessment of patient eligibility did not appear to have been standard practice in the trials. Thus, there is a strong possibility that eligibility criteria were knowingly or unknowingly applied differently for each treatment group, making subsequent exclusions imbalanced by group and potentially biased.

Discussion

Exclusion of patients from published analyses of randomized trials is widespread, although it is known that it can potentially introduce bias to these analyses and subsequent conclusions. Moreover, in meta-analysis, where enhanced statistical power increases the chance of detecting not only moderate treatment effects, but also moderate biases, greater caution is required. We examined why patients were excluded from trials and explored the potential impact of such exclusions on results at both the trial level and meta-analysis level. There were some instances where we could not obtain patients excluded from the investigators’ analyses or where patients were inadequately followed or lost to follow-up and strictly speaking, this prevented a proper intention-to-treat analysis of all patients for a few trials. Although, this could have introduced bias into our comparison, it is likely to be minimal.

Exclusions and trial characteristics

Our results, which are based on 133 randomized trials in cancer, show that trials with patient exclusions patients tended to be older than those without, albeit to a small degree, suggesting perhaps that trial methodology improved over the time period that the trials were conducted. Trials with exclusions also recruited more patients, which may reflect practical problems in the conduct of larger trials, or simply that excluding patients from larger trials seems more acceptable. Most of the trials in these meta-analyses started recruiting before 1990. There is evidence that the characteristics of cancer trials, at least those conducted in the UK, have changed markedly since then. They have been designed to recruit more patients, are more frequently multi-centre, and because of advances in trial methodology are perhaps of higher quality. Thus, in the future it would be interesting to re-appraise the impact of exclusion bias over an extended time period, by incorporating more recently conducted cancer trials. There was no clear evidence that the proportion of patients excluded varied by characteristics of the trial, but this too may be influenced by the relatively narrow time frame we were able to observe.
A substantial number of patients were excluded from the trials because they were ineligible for reasons that could not change after randomization. Examples include being too old, having received prior treatment, having inadequate blood counts, or having another chronic disease. Presumably these patients were mistakenly included, because information was incomplete at the time of randomization. At first sight it may seem that excluding these patients would not introduce bias, but it is possible that post hoc assessment of these patients’ eligibility was influenced by treatment allocation and events subsequent to randomization, such that bias cannot be ruled out. Ideally, any patients found to be ineligible after randomization should remain in a trial to provide not only an unbiased estimate of effect, but also to reflect the realities of assessing patients, diseases, and conditions in day-to-day clinical practice.

Not receiving the research or control treatment according to protocol was also a common reason for excluding patients from analyses. All of the trials compared a standard control treatment (e.g. surgery) with the same treatment plus another (e.g. surgery plus chemotherapy) or compared different chemotherapy regimens. In either situation it is very likely that protocol deviations in one treatment group would differ from those in the other, because generally more treatment is being compared with less, with probable repercussions for the duration, toxicity, and so, perhaps, tolerability of treatment. Certainly, we have evidence from some trials that protocol violations were predominantly on one arm. Moreover, we know that the underlying cause of the protocol violations were often poor patient condition or patient refusal of treatment, both of which could relate to treatment and final outcome. Thus, exclusion of patients for this reason cannot be justified and could seriously bias the analysis of the remaining patients.

A small proportion of patients were excluded because they had an early outcome, such as death prior to or during treatment, presumably on the grounds that the outcome could not be attributed to treatment, which is difficult to establish. More importantly such outcomes are defined with knowledge of the treatments being compared and the results, making objectivity difficult to maintain. In fact, occurrence of such early outcomes should be balanced by randomization, so there is no need to exclude patients who experience them. Another small group of patients were excluded because of missing data. While such data could potentially be missing at random, patients that are lost to follow-up maybe lost for reasons related to both treatment and outcome.

**Impact of exclusions**

The extent of patient exclusions was not necessarily trivial and was often greater in the research arm. This lends support to a previous study, which suggests that trials published in general medical journals may have fewer patients in the experimental arm because of post-randomization exclusion of patients. Based on the reasons we have been given, it was likely that some biases were introduced into the treatment comparisons as a result of these exclusions. However, the results for individual trials were altered in both directions, suggesting that exclusions can bias individual trial results both in favour of the treatment and control.

Situations did arise where exclusions altered most trials in a particular direction, with a greater potential for a biased meta-analysis result. In our examples, post-randomization exclusions commonly made the experimental treatment appear more beneficial. Although changes to the HR and associated statistics were often modest, it must be borne in mind that by using IPD we have already corrected for other potential biases associated with extracting data from published reports of trials, in particular publication bias and follow-up bias. Furthermore, without knowledge of patient exclusions, it is not possible to assess whether bias exists and to what degree. There may be good clinical reasons for actively excluding certain types of patients from a trial or meta-analysis and comparing this secondary analysis to the primary intention-to-treat analysis. If there is agreement between the two, this could strengthen the conclusions drawn. However to limit bias in the IPD meta-analysis setting, any deliberate exclusions should be pre-specified in the meta-analysis protocol and applied objectively and uniformly across all trials. Ideally, their impact should then be assessed by sensitivity analyses (including and excluding patients to determine whether it impacts on the treatment effect). In the soft tissue sarcoma meta-analysis described here, exclusion of patients changed the HR for survival from 0.90 \((P = 0.16)\) to 0.85 \((P = 0.06)\) (Table 3). In contrast, sensitivity analyses of pre-specified patient exclusions made little impact on the estimated effect of chemotherapy \(^{12}\) (Table 3), despite these same reasons being used in different combinations by investigators to exclude patients from their own trial analyses. Of course, this approach would not have been possible without IPD.

A reason that is often given by trial investigators to justify excluding patients after randomization is that it reduces clinical heterogeneity. This may be true within individual trials, but in the context of a meta-analysis, where the reasons for exclusion are likely to differ from trial to trial, there is a risk that differences between trials will increase. This mechanism may have been operating in some of the meta-analyses described here, where inconsistency was greater when patients were excluded.

**Generalizability of findings and implications**

Given the long history of RCTs in cancer, the trials described here are probably of relatively high quality compared with those in some other diseases or conditions. Furthermore, investigators were often able to update follow-up and provided missing data and so improve the quality of information available to us. We suspect therefore, that our results represent an optimistic view of the impact of exclusion bias on trial and meta-analysis results. By contrast, in a trial of a toxic drug, in a non-life threatening, but complex condition, eligibility of patients may be more difficult to establish, compliance may be poorer and outcome data more difficult obtain, than in the trials we have described. Thus, we imagine that in trials conducted in other diseases, examining different types of intervention and outcome, the extent of patient exclusion and consequent bias could be far greater.

Where substantial numbers of patients are excluded from reports of randomized trials, the trial results could be biased. However, to date there has been a lack of empirical evidence to support this notion.\(^{22}\) Unfortunately, this may be because many trial reports do not provide information about which patients were excluded or the reasons for their exclusions.\(^{23}\)
Paradoxically, such reports may represent trials that have actually excluded patients, are methodologically weaker and potentially more biased than reports that do provide details of exclusions. While the release of the CONSORT statement appears to have improved the quality of reporting of RCT in recent years, the specific description and application of intention-to-treat principles has continued to be inadequate. Certainly changes to the checklist and flow diagram in the revised CONSORT statement will make it much easier to track which patients are excluded from trials and whether an intention-to-treat analysis was carried out. These changes cannot however influence the quality of individual trial reports published before 2001 and therefore, cannot guarantee that published analyses will include all randomized patients.

It may not always be possible or desirable to obtain IPD for meta-analysis. However, systematic reviews and meta-analysis that rely on published or other summary data are known to be associated with a number of biases and patient exclusion bias cannot be ruled out. Thus, when appraising published trials, the level and types of patient exclusion should be judged alongside other potential sources of bias, such as the adequacy of allocation concealment. Sensitivity analyses could be used to assess indirectly the impact of a trial (or trials) with a high risk of bias, by including and excluding the trial (or trials). Ideally, if information on excluded patients is absent, reviewers should contact trial investigators to try to obtain further information. They may then carry out an appropriate intention-to-treat analysis of all randomized patients or at least assess the extent to which patient exclusion bias may be operating.

Conclusions

Exclusion of patients after randomization can affect trial, systematic review, and meta-analysis results. Wherever possible, the level of such exclusions should be taken into account when assessing the potential for bias in such studies. Ideally, though, trials, systematic reviews, and meta-analyses should be based on all randomized patients.

Acknowledgements

We would like to thank all the patients who took part in the trials and contributed to this research. We also would like to thank the collaborating investigators and institutions involved in the Advanced Bladder Overview Collaboration, the Advanced Ovarian Cancer Trialists’ Group, the Non-Small Cell Lung Collaborative Group, the Sarcoma Meta-analysis Collaboration, the Oesophageal Cancer Collaborative Group, the PORT Meta-analysis Trialists Group and Glioma Meta-analysis Trialists’ Group, who kindly supplied trial data for the various meta-analyses. We would also like to thank two anonymous referees for their suggestions to expand and improve the manuscript.

KEY MESSAGES

• Trial investigators frequently exclude patients from trial analyses and this can lead to estimates of the effect within trials that are biased in favour of either the research treatment or control. Combining these estimates in a systematic review or meta-analysis may aggregate these biases.

• Using individual patient data (IPD), we investigated whether patient exclusions were related to trial characteristics, the reasons why patients were excluded, and the impact of these exclusions on trial and meta-analysis results.

• Trials that excluded patients tended to be slightly older and larger than those that did not and, most commonly, patients were excluded for reasons of ineligibility or protocol violations.

• Exclusion biased trial results both in favour of the research treatment and control, but more often biased meta-analysis results in favour of the research treatment.

References

6 Stewart LA, Parmar MKB. Meta-analysis of the literature or of individual patient data: is there a difference? Lancet 1993;341:418–22.
8 Stewart LA, Tierney JF. To IPD or Not to IPD? Eval Health Prof 2002;25:76–97.
Commentary: Empirical evidence of attrition bias in clinical trials

Peter Juni1,2 and Matthias Egger1,2*

Excluding the 5 patients who died and the 10 who had permanent postoperative deficits, there remained 79 patients available for follow-up and at risk of subsequent persistent stroke and death.

This is how, in 1970, Fields and colleagues analysed the data from the Joint Study of Extracranial Arterial Occlusion, which had randomly allocated patients with bilateral carotid stenosis to carotid endarterectomy or medical treatment.1 Among the patients who had survived surgery and were ‘available for follow-up’, a 26% reduction in the risk of recurrent transient ischaemic attacks, stroke, or death was observed, compared with patients who had received conventional treatment (P = 0.02). Around the same time Bradford Hill, in the ninth edition of his Principles of Medical Statistics, pointed out that excluding patients after ‘admission to the treated or control group’ may affect the validity of clinical trials and that ‘unless the losses are very few and therefore unimportant, we may inevitably have to keep such patients in the comparison and thus measure the ‘intention to treat’ in a given way, rather than the actual treatment’.2 Indeed, when several years later Sackett and Gent3 re-analysed the study according to this intention to treat principle the results were less convincing: the reduction in the risk was 17% (P = 0.09) (Figure 1).

In more recent years, the debate has shifted from anecdotal evidence of bias in single trials to more sophisticated ‘meta-epidemiological’ research, based on many trials and meta-analyses.4 Schulz and colleagues5 pioneered this approach when they assessed the methodological quality of 250