Justification of exclusion criteria was underreported in a review of cardiovascular trials

Amand F. Schmidt\textsuperscript{a,b,c,*}, Rolf H.H. Groenwold\textsuperscript{a,b}, Johannes J.M. van Delden\textsuperscript{a}, Yuri van der Does\textsuperscript{a,d}, Olaf H. Klunge\textsuperscript{a,b}, Kit C.B. Roes\textsuperscript{a}, Arno W. Hoes\textsuperscript{a}, Rieke van der Graaf\textsuperscript{a}

\textsuperscript{a}Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Stratenum 6.131, P.O. Box 85500, 3508 GA Utrecht, The Netherlands

\textsuperscript{b}Division of Pharmacoepidemiology & Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences, P.O. Box 80082, 3508 TB Utrecht, The Netherlands

\textsuperscript{c}Department of Farm Animal Health, Faculty of Veterinary Medicine, Utrecht University, Yalelaan 107, Utrecht 3584 CL, The Netherlands

\textsuperscript{d}Department of Emergency Medicine, Erasmus University Medical Center, ’s Gravendijkwal 230, 3015 CE, Rotterdam, The Netherlands

Accepted 16 December 2013; Published online 5 March 2014

Abstract

**Objectives:** Ethical guidelines for human subject research require that the burdens and benefits of participation be equally distributed. This study aimed to provide empirical data on exclusion of trial participants and reasons for this exclusion. As a secondary objective, we assessed to what extent exclusion affects generalizability of study results.

**Study Design and Setting:** Review of trials on secondary prevention of cardiovascular events.

**Results:** One hundred thirteen trials were identified, of which 112 reported exclusion criteria. One study justified the exclusion criteria applied. Ambiguous exclusion criteria due to the opinion of the physician (28 of 112 = 25\%) or physical disability (12 of 112 = 11\%) were reported. Within groups of trials that studied similar treatments (ie, beta-blocker, clopidogrel, or statin therapy), baseline characteristics differed among trials. For example, the proportion of women ranged between 23.1\%–47.4\%, 2.1\%–38.9\%, and 10.6\%–50.6\% for the clopidogrel, beta-blocker, and statin trials, respectively. Nevertheless, no evidence was found for heterogeneity of treatment effects.

**Conclusion:** Almost none of the articles justified the applied exclusion criteria. No evidence was found that inclusion of dissimilar participants affected generalizability. To allow for a normative discussion on equitable selection of study populations, researchers should not only report exclusion criteria but also the reasons for using these criteria. © 2014 Elsevier Inc. All rights reserved.

Keywords: Randomized controlled trials; Ethics; Medical; Exclusion criteria; Generalizability; Research; Human experimentation

1. Introduction

International ethical guidelines for medical research involving humans widely acknowledge that inclusion of human beings for research purposes has to be justified [1,2]. Inclusion of human participants in medical research such as randomized clinical trials (RCTs) is ethically difficult because we “use” humans primarily for the purposes of science and society [3,4]. Moreover, there have been serious wrongdoings and highly controversial cases in the past [5]. Because of the ethical and historical complexity, many have felt, and still feel, that specific groups should not be included in clinical trials such as (pregnant) women, children, and people from low- and middle-income countries [2,5,6]. However, such exclusion practices have resulted in underrepresentation in research of certain groups [7]. Therefore, current versions of ethical guidelines require not only justification of inclusion but also of exclusion [2,8,9]. For instance, the Council for International Organizations of Medical Sciences guideline for biomedical research involving human beings requires that “Groups or communities to be invited to be subjects of research should
What is new?

Key findings
- In our review of cardiovascular trials, virtually all trials reported the exclusion criteria that were applied. However, in only 1 out of the 113 trials that were reviewed, justification of the applied exclusion criteria was provided.

What this adds to what was known?
- Thus far, most research has focused on “unnecessary” use of exclusion criteria without considering the justifications of applied exclusion criteria. This study shows that authors do not feel obliged to report justification for exclusion of study populations in articles on randomized clinical trials in cardiovascular diseases. Moreover, we showed that it was possible to derive generalizable treatment effect estimates in a heterogeneous patient population.

What is the implication and what should change now?
- Given the importance of justification of potential trial populations, we recommend that editors not only make it mandatory to report on exclusion criteria used but also require that the authors give justifications for the applied exclusion criteria. Explicitly reporting both exclusion criteria and rationales for these criteria may decrease the use of ambiguous exclusion criteria, or at the very least, readers can more easily judge whether exclusion of groups of patients was justified.

be selected in such a way that the burdens and benefits of the research will be equitably distributed. The exclusion of groups or communities that might benefit from study participation must be justified” [2]. Likewise, the Canadian Tri-Council Policy Statement 2 [9] stresses that “taking into account the scope and objectives of their research, researchers should be inclusive in selecting participants. Researchers shall not exclude individuals […] unless there is a valid reason for the exclusion.”

Although ethical concerns of inappropriate exclusion of trial populations are expressed in guidelines [9], it is currently unknown to what extent benefits and burdens of research are equally distributed. It is not straightforward to evaluate the current selection of study location and population because trial databases do not require reporting which potential study populations have been excluded and why. In addition, considerations on equitable distribution of burdens and benefits may be part of the evaluation of study protocols in research ethics committees, but the notes of these meetings are usually not publicly available. Therefore, a logical first step to assess in what proportion of studies unbalanced selection of patient groups was applied is literature reviews on reporting exclusion criteria and the grounds for using these criteria. Information on exclusion criteria is likely to be available in articles because both the CONsolidated Standards of Reporting Trials (CONSORT) and the Standard Protocol Items: Recommendations for Interventional Trials (SPIRIT) statements require the reporting of exclusion criteria [10,11]. Although these data not necessarily reveal whether study populations have been deliberately excluded, they may nevertheless show whether and to what extent the reasons for exclusion of trial participants have been transparent on a more general level and hence whether there are concerns of unjustifiable exclusion of study populations. This article is the first to study the use of exclusion criteria in this way. There have been previous studies on the use of exclusion criteria, but they have focused on their unnecessary use [12–14]. Thus far, no study explored whether researchers themselves justified the exclusion criteria that were applied. In this article, we will report both the current status of the application of exclusion criteria and the justification of exclusion criteria using reported data from RCTs on secondary prevention of cardiovascular events. We have chosen studies on this topic because we expected a large number of trials from a large number of research groups, thus increasing representativeness of the sample.

Apart from ethical reasons for justifying exclusion, there are also methodological reasons: if certain patient populations are not represented in RCTs, this may reduce generalizability of trial results. The previous studies that explored the application of exclusion criteria assumed that any exclusion of potential subjects hampers generalizability [12–14]. However, studies comparing RCTs and non-randomized studies (typically with less stringent inclusion and exclusion criteria) found little differences in the treatment effects [15–18]. Therefore, as a secondary objective of our study, we assessed to what extent inclusion of different patient groups affects generalizability and thus results in different treatment effects.

2. Methods

2.1. Review of trials on secondary prevention of cardiovascular events

We conducted a review to assess the current practice in reporting and justification of exclusion criteria in RCTs. We focused on the rationale for excluding groups of subjects by extracting information on included subjects and reported exclusion criteria. Using the query described in Appendix at www.jclinepi.com, we searched MEDLINE (using PubMed) for articles indexed from October 1, 2010, till May 31, 2012. Based on the title and abstract, we identified
RCTs on secondary prevention of cardiovascular events, which were defined as trials including patients with a stroke, myocardial infarction (MI), heart failure, cardiac arrhythmia, peripheral vascular disease, or patients undergoing coronary artery bypass grafting (CABG) or percutaneous coronary intervention. In addition, participants had to be randomized to one of the following treatments: statins, platelet aggregation inhibitors, beta blockers, angiotensin-converting enzyme inhibitors, or angiotensin II receptor blockers and a placebo or active comparator. Furthermore, the articles needed to be written in English and describe a single trial. To allow for a fair comparison between different article types (eg, main analyses of trial results vs. post hoc analyses), we also searched for design articles and primary publications using crossreferences and trial registries. Information from different sources related to a single trial was combined into a single entry.

2.2. Data extraction

Of the included articles, data were extracted on applied exclusion criteria, the justification of exclusion criteria, baseline characteristics, and treatment effect estimates. The number of exclusion criteria reported by a single article is often large; to limit this number, data were extracted on 18 a priori defined criteria with the option to include more if relevant. In the case of inclusion criteria, we defined the opposite as an exclusion criterion. For example, if for a particular trial, age of 65 years and older was reported as an inclusion criterion, age below 65 years was considered as an exclusion criterion. Finally, we determined whether a rationale for exclusion criteria was provided. This was defined as any point-by-point explanation about why exclusion criteria were applied. This allowed us to differentiate between articles that mentioned very explicit criteria such as any contraindication or high risk for loss to follow-up, but otherwise offered no explanation, from studies that did offer justification. Exclusion criteria were divided into those needing justification and those that were self-explanatory. Criteria were judged to be self-explanatory when there was one obvious explanation for excluding these patients. For example, a self-explanatory exclusion criterion would be a contraindication to the medication under study (eg, allergy). The rationale behind excluding such patients is obviously safety concerns, and patients with contraindications would not be considered future users.

Obviously, RCTs exploring different treatments are also likely to differ regarding exclusion criteria and groups of participants included. Therefore, to allow for a fair comparison, we selected (from the larger overall review) three groups of trials that explored the effect of clopidogrel (n = 20), beta-blocker (n = 6), or statin therapy (n = 13). These (phase III) trials compared treatment (clopidogrel, beta-blocker, or statin therapy) to an active control or placebo, on (composite) outcomes including death, MI, or stroke. Most studies randomized participants to add-on treatment, for example, adding clopidogrel to aspirin treatment and comparing this with aspirin plus placebo or usual care.

2.3. Data analysis

All analyses were performed using R for Windows, version 3.0.2 (R Foundation for Statistical Computing, Vienna, Austria) [19]. The flowchart of the MEDLINE search was created using the Diagram Designer program [20]. Baseline characteristics and effect estimates were pooled and weighted by the number of subjects. To assess generalizability of treatment effect estimates, we compared baseline characteristics of study participants across trials and determined the heterogeneity of treatment effects. We chose not to (statistically) test for the presence of treatment heterogeneity. Instead, treatment heterogeneity was quantified using the $I^2$ statistic [21] and its precision by a 95% confidence interval (95% CI). The $I^2$ statistic represents the percentage of variation in effect estimates across studies explained by actual differences (ie, not due to chance). An $I^2$ value of 0–25%, 25–50%, 50–75%, and > 75% can be interpreted as no, low, moderate, or high heterogeneity, respectively [22]. Additionally, we explored if there were any signs of treatment effect modification by age and proportion of women (ie, if there was a trend of increasing or decreasing treatment effect dependent on age or gender) [23]. These baseline characteristics were chosen because we expected them to be uniformly reported. Although we did not expect a large number of trials to exclude subjects based on age or gender, we do expect that exclusion due to other reasons will impact the gender and age distribution. For example, if the number of comorbidities increases with age, excluding subjects based on any comorbidity will decrease the average age in the study sample. Therefore, mean age and gender are used as proxies for differences in the application of exclusion criteria. Finally, to evaluate the possibility that treatment heterogeneity was dependent on exclusion criteria and not (or not only) on baseline characteristics, we evaluated whether treatment effects changed when stratifying for three exclusion criteria. These criteria (exclusion due to any medication usage at baseline, nonnaïve for intervention, and opinion of physician) were selected (post-hoc) because they were applied around 50% of the times, thus ensuring approximately equally sized strata.

3. Results

3.1. Description of included trials

The MEDLINE search resulted in 3,001 potentially relevant articles, of which 113 were included (see Fig. 1 for the flow and Appendix at www.jclinepi.com for references of the 113 included articles). Among the 113 included RCTs, 17 articles (15%) reported only on the design of the study. Characteristics of trial participants were reported in 96 (85%) of the articles, which included a median of 447 subjects [interquartile range (IQR): 192–2,141].
3.2. Reporting and justification of exclusion criteria

Exclusion criteria were reported in 112 articles (99%), a median of six exclusion criteria were reported per article (IQR: 4–6; range: 0–12). The prevalence of different exclusion criteria is presented in Table 1 and stratified for criteria needing justification and those criteria that are self-explanatory. Self-explanatory exclusion criteria included exclusion due to high bleeding risk (56 of 112 = 50%), contraindications for the studied intervention (73 of 112 = 65%), and impaired renal (63 of 112 = 56%) or liver (59 of 112 = 53%) function. Exclusion criteria that potentially require justification are exclusion of pregnant or fertile women [reported by 53 (47%) of the trials] and exclusion of lactating women [30 (27%), all of which also excluded the category pregnant or fertile women]. Studies not excluding lactating or fertile/pregnant women reported a relatively high median age (64 years; IQR 62–67, for the first group and 64 years; IQR 62–67 for the latter group), indicating that only a small number of women would be affected by excluding lactating or fertile/pregnant women. Other criteria needing justification are exclusion due to impaired cognition (16 of 112 = 14%), physical disability (12 of 112 = 11%), medication use at baseline (80 of 112 = 71%), nonnaivety to the studied intervention (40 of 112 = 36%), specific indication for either treatment arm (13 of 112 = 12%), short life expectancy (45 of 112 = 40%), based on the opinion of the physician (28 of 112 = 25%), or exclusion due to an

Table 1. Reported exclusion criteria (with percentage and 95% confidence interval) in RCTs in secondary prevention of cardiovascular events (published October 2010–May 2012)"
increased risk of being lost to follow-up (9 of 112 = 8%). Furthermore, children (i.e., participants aged <18 years) were excluded in 78 trials (70%), and exclusion based on age, other than age <18 years, was reported in 41 trials (36%), often resulting in the inclusion of older subjects. A single trial (1%) mentioned excluding women. In a sensitivity analysis, we explored whether exclusion criteria differed between publications from journals with a high (>5) and low (<5) impact factor (Appendix at www.jclinepi.com). This revealed that trials published in journals with a higher impact factor tended to apply more exclusion criteria.

Because differences in exclusion criteria in such a large group of trials might occur because of differences in, for example, treatment or outcome under study, we also explored the exclusion criteria mentioned in a subset of trials, that is, trials on clopidogrel (n = 20), beta-blocker (n = 6), or statin therapy (n = 13). These trials were similar in the studied treatments and the outcomes (see Appendix at www.jclinepi.com for details). Given this similarity, one might also except that within each group of trials, similar exclusion criteria were applied. This was indeed the case for some criteria, for example, 12 of the 13 statin trials excluded subjects with liver impairment at baseline. Contrary to this, some exclusion criteria were more variably applied, within groups of trials. For example, within the group of clopidogrel trials, nine RCTs (45%) excluded nonnaive subjects, whereas 11 (55%) included such participants. In the group of statin trials, six trials (46%) excluded pregnant or fertile women. Similarly within the beta-blocker RCTs, three of six (50%) focused on a specific age group of adults.

Only one article reported a rationale for the applied exclusion criteria. This particular study assessed the effect of clopidogrel in patients undergoing CABG and used a nonfatal end point [24,25]. The authors explain that subjects with a current malignancy were excluded by stating: “Higher risk of early postoperative mortality” [25].

### 3.3. Generalizability

The baseline characteristics of included trials showed a large range in patient characteristics between trials (Table 2), also when focusing on the subsets of trials on

**Table 2.** Baseline characteristics of study participants in RCTs on beta-blocker, clopidogrel, or statin therapy in secondary prevention of cardiovascular events (published October 2010–May 2012)

<table>
<thead>
<tr>
<th>Baseline characteristic</th>
<th>Beta-blocker RCTs</th>
<th>Clopidogrel RCTs</th>
<th>Statin RCTs</th>
<th>All RCTs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Range (min–max)</td>
<td>No. of studies (N = 6)</td>
<td>Range (min–max)</td>
<td>No. of studies (N = 17)</td>
</tr>
<tr>
<td>Number of subjects</td>
<td>70–2,708</td>
<td>6</td>
<td>60–13,608</td>
<td>17</td>
</tr>
<tr>
<td>Women (%)</td>
<td>2.1–38.9</td>
<td>6</td>
<td>23.1–47.4</td>
<td>17</td>
</tr>
<tr>
<td>Mean age (yr)</td>
<td>46.6–75.7</td>
<td>6</td>
<td>59.0–68.6</td>
<td>16</td>
</tr>
<tr>
<td>Mean weight (kg)</td>
<td>84.1</td>
<td>1</td>
<td>87.8</td>
<td>1</td>
</tr>
<tr>
<td>Mean height (cm)</td>
<td>172.7</td>
<td>1</td>
<td>—</td>
<td>0</td>
</tr>
<tr>
<td>Mean BMI (kg/m²)</td>
<td>25.7–28.0</td>
<td>4</td>
<td>24.2–30.0</td>
<td>10</td>
</tr>
<tr>
<td>Currently smoking (%)</td>
<td>4.2–17.5</td>
<td>3</td>
<td>12.8–49.8</td>
<td>15</td>
</tr>
<tr>
<td>Previously smoking (%)</td>
<td>73</td>
<td>1</td>
<td>—</td>
<td>0</td>
</tr>
<tr>
<td>Never smoked</td>
<td>—</td>
<td>0</td>
<td>—</td>
<td>0</td>
</tr>
<tr>
<td>White race (%)</td>
<td>110</td>
<td>1</td>
<td>88.8–93.7</td>
<td>3</td>
</tr>
<tr>
<td>Black race (%)</td>
<td>24</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Asian race (%)</td>
<td>—</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Hispanic race (%)</td>
<td>6</td>
<td>1</td>
<td>—</td>
<td>0</td>
</tr>
<tr>
<td>Diabetes (%)</td>
<td>20.6–35.5</td>
<td>4</td>
<td>19.4–45.1</td>
<td>16</td>
</tr>
<tr>
<td>Hypercholesterolemia (%)</td>
<td>43.0–63.0</td>
<td>4</td>
<td>15.7–82.6</td>
<td>16</td>
</tr>
<tr>
<td>Hypertension (%)</td>
<td>59.0–82.9</td>
<td>4</td>
<td>40.1–88.8</td>
<td>15</td>
</tr>
<tr>
<td>Mean serum cholesterol (mmol/L)</td>
<td>4.8</td>
<td>1</td>
<td>4.0</td>
<td>1</td>
</tr>
<tr>
<td>Mean systolic BP (mm Hg)</td>
<td>113.6–140.9</td>
<td>6</td>
<td>129.4–130.0</td>
<td>2</td>
</tr>
<tr>
<td>Mean diastolic BP (mm Hg)</td>
<td>71.0–90.8</td>
<td>4</td>
<td>77.9–80.0</td>
<td>2</td>
</tr>
<tr>
<td>Mean heart rate (%)</td>
<td>73.0–81.5</td>
<td>6</td>
<td>75.0</td>
<td>1</td>
</tr>
<tr>
<td>History of MI (%)</td>
<td>40.0–50.3</td>
<td>3</td>
<td>3.8–53.6</td>
<td>12</td>
</tr>
<tr>
<td>History of ST (%)</td>
<td>10</td>
<td>1</td>
<td>2.0–7.7</td>
<td>6</td>
</tr>
</tbody>
</table>

**Abbreviations:** RCT, randomized clinical trial; min, minimum; max, maximum; BMI, body mass index; BP, blood pressure; MI, myocardial infarction; ST, stroke.

Displayed information is based on articles that allowed for extraction of patient characteristics. The range gives minimum and maximum mean or mean percentage if there were two or more RCTs included. If only one RCT reported on the respective characteristics, the point estimate of this RCT is presented.
clopidogrel, beta-blocker, or statin therapy. For example, the proportion of women included ranged from 23.1% to 47.4% for the clopidogrel, 2.1–38.9% for the beta-blocker, and 10.6–50.6% for the statin RCTs. Other examples could be differences in the proportion of subjects with, for example, diabetes, hypertension, and hypercholesterolemia (Table 2). Note that although one trial reported to exclude women, the minimum included proportion of women was 2.1%.

Among the clopidogrel trials, 13 studies (80%) allowed for extraction of the treatment effect. Of the seven trials not reporting outcome data, three were design articles, and in three trials, no outcomes were observed (e.g., due to the outcome being of secondary importance). The reported risk ratio (RR) for clopidogrel vs. active or placebo add-on therapies ranged between 0.13 and 0.99 (pooled RR = 0.77; 95% CI: 0.67, 0.88), for the composite end point of mortality, MI, stroke, revascularization, and stent thrombosis (see Appendix at www.jclinepi.com). Plotting the RR against the proportion of women or mean age of the included subjects did not show any dependency (Fig. 2). This is in line with the $I^2$ statistic, which indicated little heterogeneity ($I^2 = 16$%; 95% CI: 0, 35).

For the group of statin trials, treatment effects could be extracted from 12 articles (92%). The RR for the composite end point of mortality, MI, stroke, and revascularization ranged between 0.25 and 1.50 (pooled RR = 0.82; 95% CI: 0.75, 0.91). Graphics did not suggest any dependency between gender or age and the treatment effect (Fig. 3) and neither did the $I^2$ statistic ($I^2 = 12$%; 95% CI: 0, 31).

Among the beta-blocker trials, using data from 5 studies (83%), the RR for the mortality and/or MI end point ranged between 0.68 and 2.25 (pooled RR = 0.91; 95% CI: 0.68, 1.21). After excluding the most extreme observation of 2.25, the range was an RR of 0.68–0.94. As with the two previous examples, the graphical display (Fig. 4) and the $I^2$ statistic did not suggest any heterogeneity ($I^2 = 0$%, 95% CI: 0, 100). However, because of small sample size, the 95% CI was large, indicating a lack of precision.

Finally, we explored whether treatment effects were dependent on the following exclusion criteria: 

### Table

<table>
<thead>
<tr>
<th>Study</th>
<th>Age(SD)</th>
<th>Prop Women(N)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bhatt2010</td>
<td>69(4.7)</td>
<td></td>
</tr>
<tr>
<td>Valgми2012</td>
<td>68(11)</td>
<td></td>
</tr>
<tr>
<td>Wang2011</td>
<td>67(11.2)</td>
<td></td>
</tr>
<tr>
<td>Fernandez2011</td>
<td>64(10.8)</td>
<td></td>
</tr>
<tr>
<td>Lee2011a</td>
<td>64(9.3)</td>
<td></td>
</tr>
<tr>
<td>Ahn2010</td>
<td>64(11.4)</td>
<td></td>
</tr>
<tr>
<td>Patti2011</td>
<td>63(12.5)</td>
<td></td>
</tr>
<tr>
<td>Arad2012</td>
<td>62(7.8)</td>
<td></td>
</tr>
<tr>
<td>Good2012</td>
<td>62(11.1)</td>
<td></td>
</tr>
<tr>
<td>Lee2011b</td>
<td>62(9.1)</td>
<td></td>
</tr>
<tr>
<td>Ruff2012</td>
<td>62(5.5)</td>
<td></td>
</tr>
<tr>
<td>Steg2010</td>
<td>59(5.4)</td>
<td></td>
</tr>
<tr>
<td>Valgми2012</td>
<td></td>
<td>0.23(1970)</td>
</tr>
</tbody>
</table>

**Fig. 2.** Forest plot of the effect of clopidogrel on the composite end point of mortality, myocardial infarction, stroke, revascularization, and stent thrombosis, ordered by the proportion of women or mean age of the individual trials. Triangles indicate treatment effects (risk ratio). Horizontal bars indicate 95% confidence intervals of the risk ratios. $N$ reflects the sample size including both genders. One of the clopidogrel trials did not report mean age and was excluded. SD, standard deviation.
medication at baseline,” “nonnaıve for intervention,” or “opinion of physician”; see Appendix at www.jclinepi.com. No dependency between the treatment effect estimates and exclusion criteria was observed.

4. Discussion

Key findings of this study are that (1) a rationale for exclusion criteria is hardly ever reported and (2) the applied exclusion criteria differed considerably between studies exploring the same treatment, yet despite differences in baseline characteristics between these studies, there was no evidence for impaired generalizability. In the following, we will discuss these findings.

Although almost all RCTs in our review of secondary prevention of cardiovascular events reported exclusion criteria (112 of 113 = 99%), only one article provided reasons for the criteria that had been applied. Therefore, it is difficult to assess whether the inclusion and exclusion of trial participants in these trials was justified. We categorized exclusion criteria into those needing justification and those that are self-explanatory. Obviously, this is to some extent an arbitrary decision, and other categorizations are also possible. However, we expect that most would agree that justification is not needed for excluding patients due to safety reasons such as contraindications for the intervention under study. We deemed other criteria less self-explanatory, and these would require justification. For example, in some occasions, participants were excluded because of a “short life expectancy” (45 of 112 = 40%). Exclusion for this reason possibly has to do with statistical power when studying nonmortality outcomes. This can, however, also be interpreted as gatekeeping [26], meaning that some groups of subjects may have been eligible to participate but have nonetheless been excluded. Another example is “opinion of the physician” (28 of 112 = 25%). This criterion may imply that researchers in their roles as physicians have made individualized judgments for patients, which should typically be avoided in a research context. Other exclusion criteria, for example, those relating to age or pregnancy, might be seen as self-explanatory by some. However, we still viewed these as needing justification because there can be multiple nonexclusive reason for applying these criteria. For example, children could be excluded because treatment effectiveness was expected to differ but also because of the

Fig. 3. Forest plot of the effect of statins on the composite end point of mortality, myocardial infarction, and revascularization, ordered by the proportion of women or mean age of the individual trials. Triangles indicate treatment effects (risk ratio). Horizontal bars indicate 95% confidence intervals of the risk ratios. N reflects the sample size including both genders. SD, standard deviation.
(administrative) burden of including children. Similarly, excluding pregnant/fertile or lactating women could be due to expected adverse event or other reasons such as the need for closer monitoring, which might be infeasible during the trial. In our sample of trials, almost half of the articles excluded pregnant/fertile women and one-fourth excluded lactating women. However, because of the relatively older target population of studies of secondary prevention of cardiovascular events, it remains uncertain whether indeed women have been excluded inappropriately, and if so, how many have been unfairly excluded. In Section 1, we have already mentioned that these reasons for exclusion of potential trial participants are probably not published elsewhere. Hence, there is a potential risk of unjustifiable exclusion.

In three groups of RCTs of the same treatment, there was no uniform application of exclusion criteria, and baseline characteristics differed considerably between studies. Despite this, the observed treatment effects were similar across trials. This suggests that findings from these trials can be generalized across groups of participants; that is, inclusion of different groups of participants did not seem to impair generalizability of treatment effects. Generalizability was assessed using the $I^2$ statistics and by graphically determining whether there was a trend between treatment effect estimates and the baseline characteristics, age and gender; no trend was found. However, it could be possible that researchers did not a priori expect difference in treatment effects between age or gender subgroups, thus allowing for differences in exclusion rates for the different subgroups. On the other hand, it seems likely that age and gender are related to other patient characteristics such as frailty and polypharmacy. Thus, age and gender might still be used as proxies for treatment effect modification between treatment and other baseline characteristics. We focused, however, on age and gender because these patient characteristics were reported by almost all trials.

Despite careful considerations, this review potentially suffers from a few weaknesses. Our review focused on trials on secondary prevention of cardiovascular events. Our findings may therefore be only applicable to this particular clinical domain and not to other domains. In addition, we conducted our search using the MEDLINE database only. Hence, RCTs not indexed by MEDLINE were not included in our review. Most, if not all, journals with a high impact factor are indexed by MEDLINE. However, this is not the case for lower impact journals. Inclusion of more lower impact publications, from other databases, could possibly change our findings. As a sensitivity analysis, we therefore

![Forest plot](image-url)
stratified exclusion criteria by the impact factor (> 5 vs. \( \leq 5 \)), which suggested that lower impact publications reported less exclusion criteria. Thus, the percentage of reported exclusion criteria is possibly somewhat inflated by only using MEDLINE (PubMed). However, it seems unlikely that searching additional databases would markedly increase the percentage of articles justifying exclusion criteria. Similarly, we recognize that our MEDLINE search, of which 4% of the hits were included, might have been overly sensitive. Although inefficient, this might nevertheless reduce the likelihood of excluding relevant publications. In the present review, we observed differences between baseline characteristics of trial populations. It seems likely that these differences are not only explained by different application of exclusion criteria but also (partially) reflect differences in available patient populations. Regardless of the causes of these differences in baseline characteristics, we did not find any indication of treatment heterogeneity depending on these differences. Heterogeneity was assessed using \( I^2 \) and by determining whether there was a trend in the treatment effect estimates per study and the mean age and the proportion of women. Because of the relative small number of studies, precision of these methods was sometimes lacking. This was most pronounced in the beta-blocker example, in which the 95% CI of the \( I^2 \) included 0% and 100%. In the clopidogrel and statin examples, the precision was higher, indicating an upper level of 35% heterogeneity. Given these limitations, we cannot conclude that there is in fact no treatment heterogeneity but merely that we could not detect any. Another issue is that the pooling of baseline characteristics, treatment effect estimates, and the exploration of heterogeneity are based on meta-analysis methods. Given that our interest is not on estimating any clinically relevant treatment effect estimates, no assessment of risk of bias of the individual studies was performed. On the other hand, little heterogeneity was found between the three trial subgroups, indicating that if a bias assessment had been applied, results would not differ markedly.

Although ethical guidelines require justification of exclusion, this study shows that authors do not feel obliged to provide this rationale in their articles. In line with these guidelines, we emphasize that researchers should not only report exclusion criteria but also discuss why these exclusion criteria were used and to what extent exclusion of those subjects could affect generalizability of treatment effects. Explicitly reporting both exclusion criteria and rationales for these criteria may decrease the use of ambiguous exclusion criteria, or at the very least, readers can more easily judge whether exclusion of groups of patients was justified.

Acknowledgments

The authors gratefully acknowledge Mirjam Nielen for her detailed comments on a previous version of the manuscript.

Appendix

Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.jclinepi.2013.12.005.

References


