Dear Author,

Please, note that changes made to the HTML content will be added to the article before publication, but are not reflected in this PDF.

Note also that this file should not be used for submitting corrections.

AUTHOR QUERY FORM

ELSEVIER	**Journal:** JCE **Article Number:** 8747	**Please e-mail or fax your responses and any corrections to:** **E-mail: corrections.esi@elsevier.tnq.co.in** **Fax: +31 2048 52789**

Dear Author,

Please check your proof carefully and mark all corrections at the appropriate place in the proof (e.g., by using on-screen annotation in the PDF file) or compile them in a separate list. Note: if you opt to annotate the file with software other than Adobe Reader then please also highlight the appropriate place in the PDF file. To ensure fast publication of your paper please return your corrections within 48 hours.

For correction or revision of any artwork, please consult http://www.elsevier.com/artworkinstructions.

Any queries or remarks that have arisen during the processing of your manuscript are listed below and highlighted by flags in the proof.

Location in article	Query / Remark: Click on the Q link to find the query's location in text Please insert your reply or correction at the corresponding line in the proof
	If there are any drug dosages in your article, please verify them and indicate that you have done so by initialing this query
Q1	Please provide the department name for affiliations "b," "h," "i," and "j," if any, and also provide complete address (eg, road, street, state, ZIP code) for all the affiliations.
Q2	Please provide the grant number for "Italian Ministry of Education, University and Research" if any.
Q3	As per journal style, 6–10 keywords are allowed. Please provide one or two more keywords.
Q4	Please provide a definition for the abbreviation "ORs" in the sentence "In cluster 1…"
Q5	Please provide a definition for the abbreviation "INR" in the sentence "A potential explanation…'
Q6	Please provide the first column heading for Table 2.
Q7	Please note that the part labels "A–H" presented in artwork of Fig. 2, but not in the caption. Please check and correct.
Q8	Please provide publisher location details for the Ref. 16 and 21.
Q9	Please confirm that given names and surnames have been identified correctly.

(*continued on next page*)

	Please check this box or indicate your approval if you have no corrections to make to the PDF file ☐

Thank you for your assistance.

ELSEVIER

Journal of Clinical Epidemiology ■ (2014) ■

Journal of Clinical Epidemiology

REVIEW ARTICLE

Discordances originated by multiple meta-analyses on interventions for myocardial infarction: a systematic review

Ersilia Lucenteforte[a,*], Lorenzo Moja[b,c], Valentina Pecoraro[b], Andrea A. Conti[d], Antonio Conti[e], Elena Crudeli[d], Alessio Galli[f,g], Gian Franco Gensini[d], Martina Minnelli[d], Alessandro Mugelli[a], Riccardo Proietti[h,i], Jonida Shtylla[d], Roberto D'Amico[j], Elena Parmelli[j,k], Gianni Virgili[l]

[a]*Dipartimento di Neuroscienze, Psicologia, Area del Farmaco e Salute del Bambino, Università degli Studi di Firenze, Firenze, Italia*
[b]*Unità di Epidemiologia Clinica, IRCCS Istituto Ortopedico Galeazzi, Milano, Italia*
[c]*Dipartimento di Scienze Biomediche per la Salute, Università degli Studi di Milano, Milano, Italia*
[d]*Dipartimento di Medicina Sperimentale e Clinica, Università degli Studi di Firenze, Italia*
[e]*Dipartimento di Scienze Biomediche, Sperimentali e Cliniche, Università degli Studi di Firenze, Italia*
[f]*Dipartimento di Scienze Cliniche "L. Sacco", Università degli Studi di Milano, Italia*
[g]*Dipartimento CardioToracoVascolare "A. De Gasperis", Ospedale Niguarda Ca' Granda, Milano, Italia*
[h]*Laboratorio di Elettrofisiologia Cardiaca, Ospedale Luigi Sacco, Milano, Italia*
[i]*Montreal General Hospital McGill University Montreal, Montreal, Canada*
[j]*Centro Cochrane Italiano, Dipartimento di Medicina Diagnostica, Clinica e di Sanità Pubblica, Università degli Studi di Modena e, Reggio Emilia, Italia*
[k]*Dipartimento Epidemiologia del S.S.R. - ASL RME, Regione Lazio, Italia*
[l]*Dipartimento di Chirurgia e Medicina Traslazionale, Università degli Studi di Firenze, Italia*

Accepted 5 November 2014; Published online xxxx

Abstract

Objectives: To clarify the impact of multiple (covering the same population, intervention, control, and outcomes) systematic reviews (SRs) on interventions for myocardial infarction (MI).

Study Design and Setting: Clinical Evidence (BMJ Group) sections and related search strategies regarding MI were used to identify multiple SRs published between 1997 and 2007. Multiple SRs were classified as discordant if they featured conflicting results or interpretation of them.

Results: Thirty-six SRs (23.5% of 153 on the treatment or prevention of MI) were classified as multiple and grouped in 16 clusters [ie, at least two SRs with the same PICO (population, condition/disease, intervention, control) and at least one common outcome] exploring angioplasty, angiotensin-converting enzyme inhibitors, anticoagulants, antiplatelets, β-blockers, and stents. Complete agreement on statistically significant differences between interventions was found in 7 of 10 clusters with a shared composite outcome. Agreement was reduced when single outcomes were considered. Despite substantial variation and limited agreement in reporting of major outcomes, SRs agreed in their conclusions on the superiority of either the intervention or control in 14 of 16 clusters. Sources of minor discrepancies were found in terms of study and outcome selection, subgroup analyses, and interpretation of findings.

Conclusion: Multiple SRs agreed in their qualitative conclusions but not on reporting and on analyses of hard outcomes. Discordance on significance of treatment effects was due to a combination of variation in design with inclusion of different studies and lack of precision for single hard outcomes compared with a composite outcome. Such inconsistencies among SRs could potentially slow the translation of SRs' results to clinical and public health decision making and suggest the need for a broader methodological and clinical agreement on their design.

Keywords: Cardiovascular agents; Dissent and disputes; Meta-analysis; Myocardial infarction; Policy making

Conflict of interest: We declare that L.M. and A.C. received research grant for the submitted work from Italian Ministry of Education, University and Research. The conclusions in this report are those of the authors and do not necessarily reflect the scientific judgment of its funding agencies. There are no other relationships or activities that could appear to have influenced the submitted work.

* Corresponding author. Tel.: +39-(0)554271333; fax: +39-(0) 554271280.
E-mail address: ersilia.lucenteforte@unifi.it (E. Lucenteforte).

http://dx.doi.org/10.1016/j.jclinepi.2014.11.004
0895-4356/

What is new?

- A sizable proportion of published systematic reviews (SRs) in cardiovascular medicine overlaps in terms of population, intervention, comparator, and outcomes.
- Overlapping multiple reviews on myocardial infarction interventions were found to agree in their qualitative conclusions, but disagreements in the reporting and the analysis of hard outcomes are common.
- Because outcome selection and reporting is inconsistent across SRs exploring the same objective, multiple reviews can be seen as useful confirmation of findings rather than a waste of resources in evidence synthesis production.

1. Introduction

The number of systematic reviews (SRs) published has risen in the past decade with approximately 2,500 new publications indexed annually on MEDLINE [1]. This increases the likelihood of finding multiple overlapping reviews [2] that present conflicting results.

The term ''systematic'' implies that the process and methodology of generating a review, in addition to being exhaustive, are valid, transparent, and well reported such that other independent researchers following the same methods can replicate the same results and, therefore, arrive at the same conclusions [3,4]. In 1997, Jadad et al. [5] first addressed the issue of how to appraise discordant evidence originating from multiple but similar SRs. Similar SRs have the same objective but might differ at the level of the results or their interpretation. In such cases, the review featuring the most comprehensive literature search, explicit and reproducible selection of studies, and a quality assessment of the primary studies should be preferred. There are some examples of discordant SRs that have generated controversy [6–8]. One emblematic case study explored the results of 10 multiple SRs focused on the role of acetylcysteine in the prevention of contrast-associated nephropathy [9]. Five reviews recommended routine use of acetylcysteine, whereas the others were more cautious and called for further trials. There exists little empirical evidence about the multiplicity cumulative phenomenon.

The aim of this article was to assess the scientific validity and reproducibility of results in multiple SRs examining the health effects of interventions for a cardiovascular disease such as myocardial infarction (MI). We focused on a cohort of SRs published between January 1997 and December 2007. These SRs are the potential, key drivers of actual clinical practice and can help elucidate the impact of discordances on practice for cardiovascular diseases.

2. Methods

We have described the rationale, design, and methods in detail in a previous publication [10]. Briefly, the eligibility of studies was assessed, independently, by two reviewers across all phrases in accordance to standard rules and implementing ad hoc forms. Disagreements were resolved by consensus; arbitration with a third reviewer was possible when necessary. We kept double entry of all details to ensure data quality.

Our study was carried out in six phases.

Phase 1. We carried out a systematic search process starting from Clinical Evidence search strategy process, as described elsewhere [10]. Clinical Evidence is the BMJ Group medical textbook synthesizing biomedical evidence on a wide range of globally important clinical conditions [11]. We focused on the MI Clinical Evidence chapter presenting interventions encompassing the primary and secondary prevention on MI. The Clinical Evidence search strategy was based on strategy process and outputs developed a priori and performed by BMJ Evidence Centre information specialists (Clinical Evidence Study design search filters, available at http://clinicalevidence.bmj.com/x/set/static/ebm/learn/665076.html). The search strategy was adapted to MI prevention and applied to MEDLINE, Embase, and Cochrane Register of Controlled Trials (Central) to capture all potential overlapping SRs. We checked, also, the reference list of all selected SRs to identify other relevant SRs.

Phase 2. We screened the titles and abstracts to identify SRs investigating the efficacy of treatments for coronary artery disease, including both pharmacological and procedural interventions. We included publications that mentioned the terms ''systematic review'' or ''meta-analysis'' in the title or abstract or reported to have searched at least one bibliographic database. We searched SRs published from January 1997 to December 2007 in English, Italian, Spanish, or French.

Phase 3. We extracted information from the title and abstract to identify potential, multiple SRs. Potential multiplicity was defined as at least two independent SRs sharing the same population condition/pathology and intervention, irrespective of differences in outcomes and controls. Reasons for exclusion, such as duplicate publication or narrative nature of review, were documented.

Phase 4. We accessed the full text of reviews, analyzing the overlap between controls and outcomes. Thus, we identified clusters of at least two SRs with the same PICO (population, condition/disease, intervention, control and at least one outcome) and objective. These were classified as clinically homogeneous (ie, multiple SRs). SRs considering several drugs or interventions could have been grouped in more than one cluster.

Phase 5. We completed an analytical characteristic report for each SR within a cluster, determining the clinical features, design (eg, inclusion criteria), methods, and the publication history. We used the checklist by Oxman and Guyatt [4] to assess the quality of included SRs. When a previous review was cited, we abstracted the authors' rationale for repeating the review, if reported [12].

Phase 6. We examined the concordance or discordance for direction and statistical significance of meta-analysis results. We considered ''qualitative interaction'' and ''quantitative interaction'' [13]. Qualitative interaction exists if the direction of effect is reversed. Quantitative interaction exists when the magnitude of the effect, but not of the intervention, varies such that an intervention is beneficial to different degrees in different subgroups. We subcategorized quantitative interaction in: (1) ''significance cliché'': two effect sizes were in the same direction, but only one reached statistical significance, which we called; (2) ''quantitative interaction'': there was a statistical criterion to identify interaction [14].

As a final health-policy step, we presented the results of each outcome according to patterns that are relevant to clinicians and policy makers who seek advice from SRs to inform practice.

Concordant significant superiority of one treatment: the direction of the effects agreed, and findings were statistically significant for all SRs. We a priori reasoned that the identification of statistically significant heterogeneity, defined as chi-squared test P-value <0.1, would mostly affect estimates of the amount of benefit but not of its likeness. In this scenario, decision makers finding multiple reviews would choose one intervention and discard the other and move on to balance the efficacy against the harms and costs to make a final recommendation.

Concordant nonsignificant difference between treatments: no SR was able to demonstrate statistically significant superiority for the effectiveness or harm. SRs could be concordant or discordant about direction, but none found significant differences between the intervention and the control. Meta-analyses were classified in this category irrespective of statistically significant heterogeneity. Our definition might be simplistic because we did not consider the optimal information size [15] or the concept of ruling out the possibility of an important effect, such as those explored in equivalence or noninferiority studies [16]. In such scenario, decision makers finding multiple reviews would not be able to decide on which treatment to adopt based on the efficacy.

Discordant significance of treatment effects: at least one SR was discordant from the others for statistical significance. Decision makers finding such multiple discordant SRs would not be supported by consistent evidence syntheses when deciding in favor or against the adoption of a treatment. Ambiguity would be enhanced if statistical significant heterogeneity of effect estimates is found among these SRs.

Each cluster was categorized according to these patterns (ie, concordant significance, concordant nonsignificance, discordant significance), considering the outcomes in the following hierarchical order: overall mortality, which is most relevant and unbiased; single cardiovascular components; and composite outcomes, which are more prone to reporting and other biases [17].

Finally, we reviewed all multiple SRs and recorded all concluding statements in the abstract and, secondarily, in the main text addressing the efficacy of the intervention in modifying outcomes. A concluding statement had to explicitly include the intervention and remarks about the causal relationship, with or without the outcome [18]. We excluded comments about implications for practice.

We used multinomial logit regression, with SRs as a unit of analysis and clusters as random effects, to compute the relative risk ratio that single-component outcomes (mortality, MI, stroke) fall into the nonsignificant effect or outcome not reported category compared with the significant effect category. P-values less than 0.05 were considered statistically significant. The degree of heterogeneity between meta-analyses' effect measures was identified if the chi-squared test significance P-value was less than 0.1. The I^2 was not used because this is estimated with great uncertainty when less than five studies are included in a meta-analysis, as found in our review. These analyses were conducted using Stata 12.1 software (StataCorp, College Station, TX, USA).

3. Results

3.1. Included studies

The Clinical Evidence search strategy identified 1,503 records (Fig. 1). After an initial screening of the titles and abstracts, 1,350 studies were excluded because they assessed an intervention not considered in any other SR or were narrative reviews focusing on the disease or interventions rather than the efficacy, safety, and MI. We obtained the full text of 153 articles. Among these, two SRs were excluded because they were duplicate publications, 40

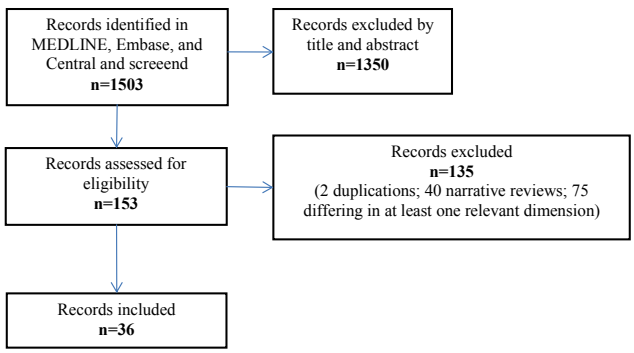

Fig. 1. Study flow diagram for screened, included, and excluded articles.

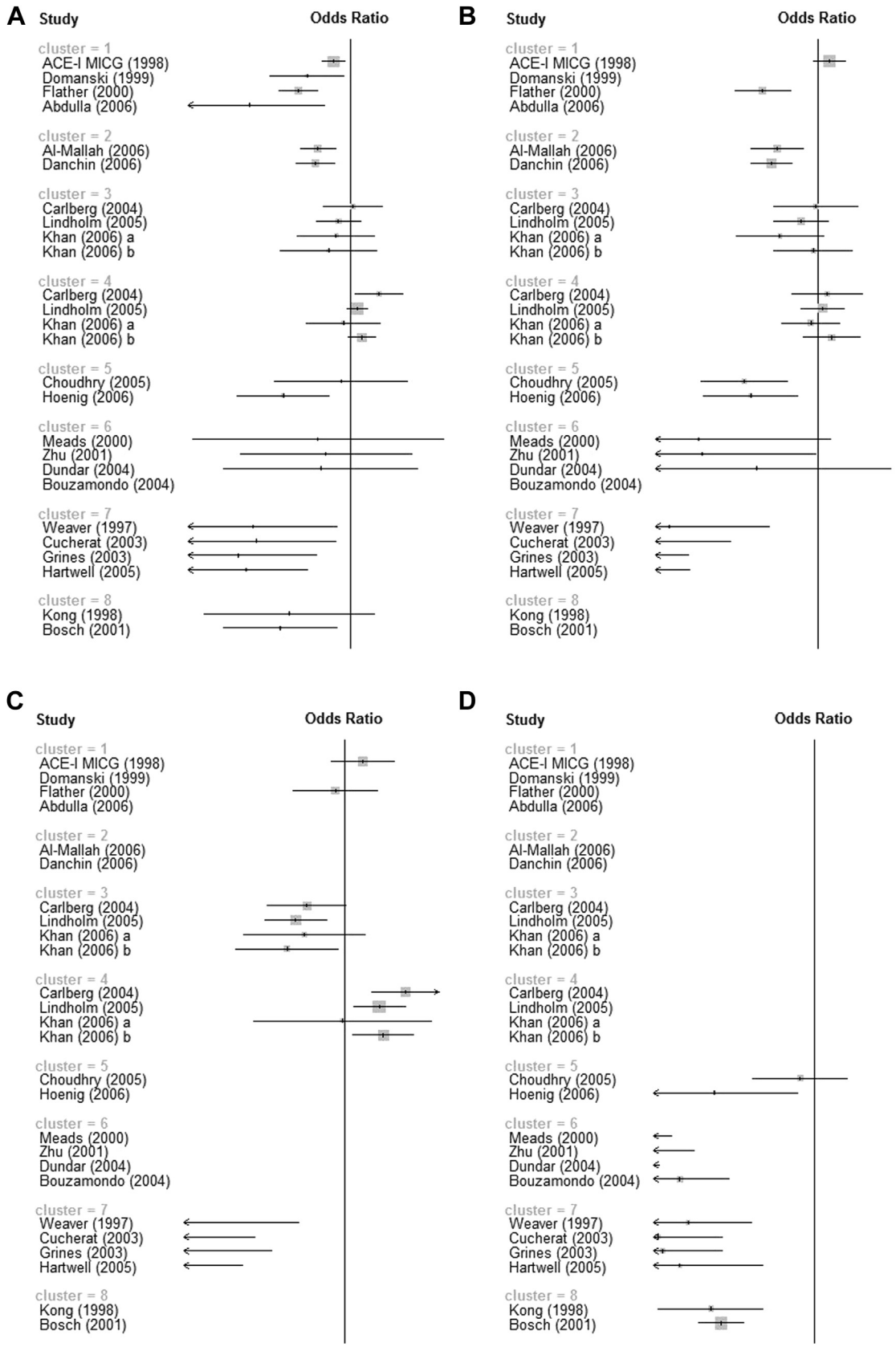

Fig. 2. Forest plots enabling the inspection of all meta-analyses results for single-component and composite outcomes shared by two or more SRs in each cluster. ACE, angiotensin-converting enzyme; ACS, acute coronary syndrome; CAD, coronary artery disease; LMWH, low molecular weight heparin; LVSD, left-ventricular systolic dysfunction; MI, myocardial infarction; NSTEMI, non-ST elevation acute coronary syndrome; PCI, percutaneous coronary intervention; UA, unstable angina; UFH, unfractionated heparin.

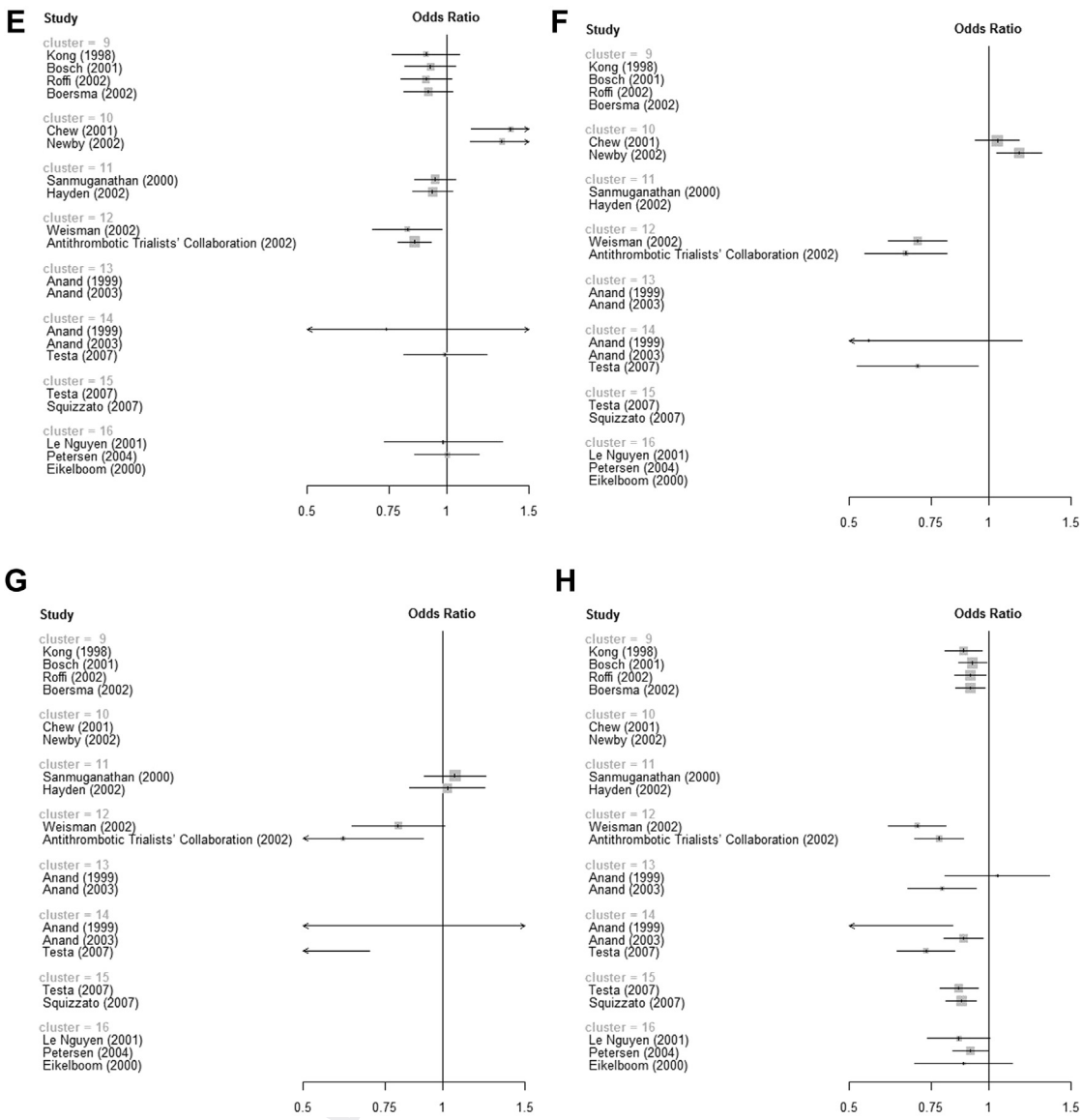

Fig. 2. (*continued*).

narrative reviews, and 75 differing in at least one relevant dimension (ie, population, intervention, or comparator). Finally, we included 36 SRs from 19 different journals; most journals published only one of the selected SRs (10 of 19). Four sources were identified as top publishers: Lancet ($n = 5$), The Cochrane Database of Systematic Reviews ($n = 4$), Circulation ($n = 3$), and JAMA ($n = 3$). The SRs investigated angioplasty, angiotensin-converting enzyme (ACE) inhibitors, anticoagulants, antiplatelets, β-blockers, and stents. Their general characteristics and methodological quality are reported in eTables 1 and 2, /Apendix on the Journal's website at www.jclinepi.com respectively. Overall SR quality was better for the inclusion criteria reporting satisfied (34 of 36, 94.4%), meta-analysis methods described (36 of 36, 100%), and conclusions supported by data (35 of 36, 97.2%). The lowest quality items were selection bias avoided (12 of 36, 33.3%), study validity methods reported (15 of 36, 41.7%), and study validity methods used for inclusion and analysis (9 of 36, 25%). The quantitative results and conclusions of each SR are presented in eTable 3/Apendix on the Journal's website at www.jclinepi.com.

3.2. Concordance of meta-analytic results

The 36 identified SRs were categorized in 16 clusters of multiple SRs. Fig. 2 present forest plots enabling the inspection of all meta-analyses' results for single-component and composite outcomes shared by two or more SRs in each cluster. Individual cluster ratings using defined criteria for quantitative and qualitative discordance are shown in Table 1, where they are grouped according to a decision-maker perspective. Cluster ratings for authors' conclusions are shown in Table 2.

Table 1. Individual cluster ratings for quantitative and qualitative discordance stratified by outcome

Cluster	Mortality	MI	Stroke	Composite	Composite type
#1. ACE inhibitors *versus* placebo/control in patients with MI	H	HD	D		
#2. ACE inhibitors *versus* placebo/control in patients without LVSD after MI					
#3. β-blockers *versus* placebo in patients with hypertension	D				
#4. β-blockers *versus* active control in patients with hypertension	D	D			
#5. Invasive *versus* conservative strategy in patients with UA or NSTEMI	D				Mortality, MI
#6. Stent *versus* no stent during PCI for MI					Mortality, MI, revascularization
#7. PCI *versus* thrombolysis for MI					Mortality, MI
#8. Platelet Glycoprotein IIb/IIIa antagonists *versus* placebo in patients treated with PCI for MI					Mortality, MI
#9. Platelet Glycoprotein IIb/IIIa antagonists *versus* control in patients with NSTEAC					Mortality, MI
#10. Oral Platelet Glycoprotein IIb/IIIa antagonists *versus* aspirin in patients treated with PCI for ACS					
#11. Aspirin *versus* control for primary prevention of cardiovascular events					
#12. Antiplatelet therapy *versus* control for secondary prevention of cardiovascular events					Cardiovascular mortality, MI, Stroke
#13. Moderate-to-High-dose oral anticoagulant *versus* aspirin in patients with CAD					Mortality, MI, Stroke
#14. Moderate-to-High-dose oral anticoagulant plus aspirin *versus* aspirin in patients with ACS				H	Mortality, MI, Stroke
#15. Aspirin plus clopidogrel *versus* aspirin in patients at high risk of cardiovascular events					Mortality, MI, Stroke
#16. LMWH *versus* UFH in patients with UA or NSTEMI					Mortality, MI

Abbreviations: MI, myocardial infarction; ACE, angiotensin-converting enzyme; LVSD, left-ventricular systolic dysfunction; UA, unstable angina; NSTEMI, non-ST elevation acute coronary syndrome MI; PCI, percutaneous coronary intervention; NSTEACS, non-ST elevation ACS; ACS, acute coronary syndrome; CAD, coronary artery disease; LMWH, low molecular weight heparin; UFH, unfractionated heparin.

Green: concordant statistical significant superiority of either one treatment or the other in the cluster; yellow: concordant statistical nonsignificant difference between treatments; red: discordant statistical significance of treatment effects; D: discordance on direction for at least one SR of the cluster vs. others; H: heterogeneity chi-square $P < 0.10$. (For interpretation of references to color in this table legend, the reader is referred to the web version of this article.)

Table 2. Summary results on cluster ratings for quantitative and qualitative discordance stratified by outcome

	Mortality	MI	Stroke	Composite outcome
Concordant statistical significant superiority of either one treatment over the other	Five clusters (1, 2, 7, 12: benefit; 10: harm)	Four clusters (2, 5, 7, 12: benefit)	One cluster (7: benefit)	Seven clusters (6, 7, 8, 9, 12, 14, 15: benefit)
Concordant statistical nonsignificant difference between treatments	Six clusters (3, 6, 9, 11, 14, 16)	Three clusters (3, 4, 6)	Two clusters (1, 11)	—
Discordant statistical significance of treatment effects	Three clusters (4, 5, 8)	Three clusters (1, 10, 14)	Four clusters (3, 4, 12, 14)	Three clusters (5, 13, 16)
Outcome not reported	Two clusters (13, 15)	Six clusters (8, 9, 11, 13, 15, 16)	Nine clusters (2, 5, 6, 8, 9, 10, 13, 15, 16)	Six clusters (1, 2, 3, 4, 10, 11)

Abbreviation: MI, myocardial infarction.

Agreement was found in 11 of the 14 clusters reporting mortality, the hardest outcome measure: five (36%) agreed on either benefit (4 clusters) or harm (1 cluster) and six (43%) agreed on the inability to detect a difference. The remaining three (21%) included SRs with discordant results, for instance, one SR finding a benefit, whereas the others did not find statistical significant effects.

Agreement on concordant significance was found more often for reported composite outcomes (7 of 10 clusters, 70%) compared with single-component outcomes (10 of 31, 32%; $P = 0.035$). In only one cluster (7), of seven, composite outcomes fully agreed with single component; in all other clusters, composite outcome was consistently in favor of the intervention.

When the composite outcome was not reported (six clusters), the other single-component outcomes agreed in two clusters showing concordant significance for mortality and MI (2) or concordant nonsignificant for mortality and stroke (11).

There was very good agreement on the summary estimate favoring the experimental intervention. In fact, discordance on the direction of the overall estimate was found in only six outcomes across four clusters, which did not find a statistical significant difference (Table 1). Furthermore, statistical significant heterogeneity was found in only three clusters, all concordant on benefit.

At the cluster level, the concordance across outcomes, regarding the three concordance patterns plus the lack of reporting as a fourth category, was poor. In fact, mortality (5, 7, 12), MI (7, 11, 12), and stroke (2, 7, 10) fell into the same pattern as the composite outcome in only 3 of 16 clusters each.

Multinomial logit regression confirmed that single-component outcomes, that is, mortality, MI, and stroke, were more likely to show a statistically nonsignificant as opposed to a significant effect compared with composite outcomes [relative risk ratio (95% confidence interval): 6.4 (1.9, 21); 6.2 (1.7, 22); 5.1 (1.3, 21) in favor of composite outcomes]. MI and stroke [3.1 (1.0, 9.4); 5.6 (1.8, 17)], but not mortality (0.78 [0.23-2.5]), were more likely to be unreported than reported as statistically significant compared with composite outcomes.

Finally, year of publication could not be shown to increase the probability of reaching nominal statistical significance, adjusting by type of outcome [0.89 (0.72, 1.08)].

3.3. Concordance of authors' conclusions

Table 3 summarizes the concordance within each cluster regarding authors' conclusions as presented in the abstract or, where the abstract was unclear, in the discussion. The 36 SRs showed very good concordance because as many as 14 of 16 clusters agreed on the superiority of the treatment (13 clusters) or control (one cluster, see eTable 2/Apendix on the Journal's website at www.jclinepi.comfor details). The only discordant clusters were 3 and 4, in which the same three SRs conducted two comparisons. Two SRs agreed in their results and conclusions and did not find significant differences between β-blockers and the control, whereas one SR conducted subgroup analyses on age, finding heterogeneity of effect. However, no test for interaction was reported to confirm an effect modification by age, and it was unclear whether the subgroup analysis was prespecified.

In 6 of 16 clusters, sources of discordance did not affect the final agreement between the conclusions. In cluster 5, an earlier SR included three present era studies published from 1995 to 1998, whereas the other SRs excluded them. This led to more encouraging results across all outcomes favoring the intervention. In cluster 6, only one review presented the effect on the composite outcome, which led the authors to a stronger statement compared with the other three SRs, which found that the benefit of stenting was inflated by revascularization rates, an unblinded outcome component that was considered to be at risk of bias, thereby deflating the conclusion remark. Two clusters, 9 and 13, agreed in favor of one treatment, but one SR in each cluster argued for the lack of benefit on specific single outcomes. Two other clusters, 11 and 16, concluded balancing the benefits and harms.

Table 3. Individual cluster ratings for authors' conclusions

Cluster	Overall agreement in conclusions	Comment from a methodological / clinical point of view and minor sources of discordance
#1. ACE inhibitors *versus* placebo/control in patients with MI	 Effective (4 SRs)	Full overall agreement on the use of ACE inhibitors after MI, although one SR discussed effect modification according to NYHA class.
#2. ACE inhibitors versus placebo/control in patients without LVSD after MI	 Effective (2 SRs)	Small treatment effect valued differently in two SRs (stress on statistical significance vs. stress on amount of benefit).
#3. β-blockers versus placebo in patients with hypertension	 Cannot show difference (2/3 SRs). Effective in younger age subgroup (1 SR)	One SR conducted subgroup analysis, but unclear if preplanned and statistical test for subgroup differences not provided. The same three SRs assessed both comparisons.
#4. β-blockers versus active control in patients with hypertension		
#5. Invasive versus conservative strategy in patients UA or NSTEMI	 Effective (2 SRs)	Weaker conclusion in one earlier SR including pre-stent studies: this SR found a benefit limited to one outcome as compared to the other SR which found benefits across more outcomes and excluded pre-stent studies.
#6. Stent versus no stent during PCI for MI	 Effective (4 SRs)	3 SRs with caveats regarding the use of composite outcomes, as opposed to 1 SR which does not use composite components.
#7. PCI versus thrombolysis for MI	 Effective (4 SRs)	Full overall agreement on the superiority of PCI over thrombolysis; two SR highlighted that these were short-term effects.
#8. Platelet Glycoprotein IIb/IIIa antagonists versus placebo/control in patients treated with PCI for MI	 Effective (2 SRs)	Full agreement.
#9. Platelet Glycoprotein IIb/IIIa antagonists versus placebo/control in patients with NSTEACS	 Effective (4 SRs)	4 SRs agree on benefit on mortality/MI, but 1 SR highlights the lack of benefit on mortality as single component and 2 SRs reports moderate effects.
#10. Oral Platelet Glycoprotein IIb/IIIa antagonists versus aspirin in patients treated with PCI for ACS	 Harmful (2/2)	Full agreement.
#11. Aspirin versus control for primary prevention of cardiovascular events	 Effective (2 SRs)	Both SRs discuss the balance benefit/harm according to subgroups for different baseline risks.
#12. Antiplatelet therapy versus control for secondary prevention of cardiovascular events	 Effective (2 SRs)	Full agreement; both SRs place thsi information in the context of the balance between benefits and harms for different doses.
#13. Moderate-to-High-dose oral anticoagulant versus aspirin in patients with CAD	 Effective (2 SRs)	SRs discordant on statistical significance of the shared composite outcome, but the earlier review concludes on several outcomes. Both SRs balance benefits and harms.
#14. Moderate-to-High-dose oral anticoagulant plus aspirin versus aspirin in patients with ACS	 Effective (3 SRs)	Full agreement; both SRs place thsi information in the context of the balance between benefits and harms for different doses.
#15 Aspirine plus clopidogrel versus aspirin in patients at high risk of cardiovascular events	 Effective (2 SRs)	Full agreement; both SRs place thsi information in the context of the balance between benefits and harms for different doses.
#16. LMWH versus UFH in patients with UA or NSTEMI	 Effective (3 SRs)	Non-inferiority in all 3 SRs supporting favorable profile in benefit/harm balance.

Abbreviations: ACE, angiotensin-converting enzyme; MI, myocardial infarction; SRs, systemic reviews; LVSD, left-ventricular systolic dysfunction; UA, unstable angina; NSTEMI, non-ST elevation acute coronary syndrome MI; PCI, percutaneous coronary intervention; NSTEACS, non-ST elevation ACS; ACS, acute coronary syndrome; CAD, coronary artery disease; LMWH, low molecular weight heparin; UFH, unfractionated heparin.

Green indicates concordance; red indicates discordance. (For interpretation of references to color in this table legend, the reader is referred to the web version of this article.)

3.4. Study overlapping and sources of quantitative discordance

We assessed overlapping of included studies and causes of discordance on significance of treatment effects, that is, for outcomes with a red flag in each cluster in Table 1.

Q4 In cluster 1 (four SRs), two individual patient data (IPD) meta-analyses yielded very different ORs regarding outcome MI, both for direction and for significance, and this was related to the planned inclusion of mutually exclusive patients' subgroups, each of four studies, by the same collaborative IPD review group. In fact, the two reviews assessed the effect of ACE inhibitors initiated in the acute phase as compared with the late phase of MI, and detecting discordant treatment effects was an aim of the authors. The other two reviews in this cluster did not report on outcome MI.

Stroke in cluster 3 and stroke and mortality in cluster 4, which included the same three SRs, showed discordance on significance due to (1) different inclusion criteria: one included only four atenolol studies and another included additional studies on other β-blockers and (2) conclusions based on subgroup analysis: the third SR split results by younger vs. older patients, whereas the former did not. Study overlapping among the three reviews was not a problem once inclusion criteria were taken into account.

In cluster 5 (two SRs), disagreement on mortality and on the composite outcome was also due to different inclusion criteria because the abstract of one SR reported an OR based on two studies undertaken "in the stent era," whereas the other included five more studies and yielded nonsignificant results.

In cluster 8 (two SRs), the more recent SR, published in 2001, comprised all but 1 of 10 studies included in the other, published in 1998, plus 20 additional studies published later, thus reaching statistical significance with no substantial difference in OR point estimate.

In cluster 10 (two SRs), the same four studies were included in both SRs, although one SR specifically commented that the other had overlooked time and dose dependence of harm related to oral glycoprotein IIb/IIIa antagonists because of the arbitrary 30-day end point used.

In clusters 12 (two SRs), study overlap and, therefore, causes of discordance could not be investigated as the included studies could not be traced for each outcome in an overview of 287 studies including several clinical questions.

In cluster 14 (three SRs), two SRs included three or two studies for MI and stroke, whereas the third did not specify which of 14 RCTs included where used for this outcomes. Q5 A potential explanation is that this last review highlighted a beneficial OR of MI restricted to studies targeting at INR between two and three, as compared with a nonsignificant meta-analysis of all studies (number of studies not reported).

4. Discussion

4.1. Summary of findings

Our findings suggest that in a 10-year period (1997 to 2007), the phenomenon of multiple overlapping SRs investigating treatments for MI is not rare. The results of these multiple SRs and their interpretation were consistent across SRs. Despite strengths and weaknesses regarding completeness and concordance of outcome reporting, the authors managed discordances so that the conclusions of SRs agreed, for or against the experimental treatment, in as many as 14 of 16 clusters. Concordance regarding the most commonly reported outcome components such as the overall mortality, MI, and stroke was more limited compared with composite outcomes, which may be related to the greater statistical power allowed by composite outcomes. Moreover, although mortality data were consistently presented by at least two SRs in almost all clusters, we found selective reporting of MI or stroke in published manuscripts. Regarding outcome components, SRs agreed well in terms of the direction of effects but not on their statistical significance.

The overall discordance pattern among outcomes suggests that small effects are difficult to measure precisely, even by means of SRs that include several studies. This is why many reviewers are prone to select composite outcomes: they can increase the statistical precision [19]. Because mortality and other adverse events are relatively infrequent in patients included in RCTs, lack of power can be a cause of inconsistency in the direction and significance of meta-analysis results. Other causes could be different precisions in measuring effects; for example, all-cause mortality can be measured more precisely compared with MI recurrence. Finally, treatments could have truly different effects on each outcome component, an issue that might largely depend on the clinical and biological background of the disease. In other words, multiple outcomes refer to an overall disease process, an assumption that can be met sporadically. This led Freemantle et al. [20] to conclude that a benefit regarding the composite outcome does not necessarily translate to a benefit regarding each component. Accordingly, in 2007, Ferreira-Gonzalez et al. [[17]] found that end points of least importance to patients typically contributed the most events of composite end points used in cardiovascular trials but often showed much larger effects than hard outcome such as mortality. Reassuringly, most multiple SRs included in our review used composite outcomes based on hard components in addition to death, such as MI and stroke.

Poor study overlapping for outcomes that disagreed on statistical significance of treatment effects was mostly related to differences in SR inclusion criteria and use of subgroups analyses. We also found that discordance was more common for single hard outcomes compared with composite outcomes, suggesting that both lack of precision and design differences were causes of discordance.

Our study supports current guidance for systematic reviewers, such as in the Cochrane Handbook [21] and GRADE [22], that authors and panelists must prespecify the primary outcome(s), clarifying the approach used to value primary and secondary outcomes when drawing conclusions on comparative treatment effect. SRs are not immune from selective outcome reporting bias [23]: they can be at risk of bias due to selective inclusions of RCTs and selective reporting of results when there are multiple outcomes.

If we adopt the point of view of policy makers and research investors, our findings might appear more negative. In fact, even limited discordances, which do not affect agreement on conclusions of treatment effect, might lead to doubts regarding the real value of an intervention and, consequently, deter decisions to enforce it. The implications of such discordances are difficult to interpret. Scientists might be perceived as conflicting professionals and science as an unreliable process. Critiques to research that generate discordant results are not new [24] but relatively uncommon at the level of cumulative studies. Parallel to a study by Siontis et al. [2], our study confirms that there is a waste in the production and reporting of research synthesis with many topics covered by multiple overlapping meta-analyses[2].

Our study reinforces the importance of projects such as PROSPERO for the registration of meta-analysis protocols [25] and COMET, which aims to standardize the use of predefined outcomes in each research field that fulfill the requirements of relevance and usability [26]. Furthermore, patient-reported outcomes are seldom used in cardiology and should be considered because problems may exist not only with composite outcomes, but also with the balance between benefits and harms [27].

In 1997, Jadad et al. [5] proposed a theoretical model, which hypothesized that discordance between SRs can originate from differences in quantitative results (direction, magnitude, or significance) and/or their interpretation[5]. Few other studies, mainly anecdotic, have investigated sources of discordance among SRs: different interpretation of identical quantitative results [8]; different interpretation due to noncomparable analyses because of, mainly unmotivated, subgroup analyses; different selection of primary or secondary outcomes [6]; different inclusion criteria [6]; and different meta-analytical approaches [9,28]. Our systematic attempt to explore the theoretical model proposed by Jadad et al. [5], although limited to the cardiologic field, confirms many of the sources of discordance hypothesized. In addition, we found that study overlapping was good once design differences among SRs are taken into account, meaning that the systematic search of evidence was not a problem, and that an agreement on the PICO should be reached to plan SRs that answer questions that are relevant to clinical practice.

The strength of our study lies in the extensive search of SRs using validated strategies, the duplicate and independent assessment of multiple SRs based on PICO, and a caution adopted at each step, including the methodological quality assessment, data extraction, and selection of primary statements in each article.

A limitation of our review is that despite the large number of SRs assessed, we were unable to correlate the occurrence of discordance with either the SR quality or year of publication. These can be factors of interest given that the quality of SRs have improved over the last years [29]. Although the cardiovascular disease field is one of the widest area of investigation, we found too few multiple SRs to explore these issues.

5. Conclusion

Our study in the cardiologic field suggests that SRs addressing the same objective substantially agree in both their results and interpretation, despite the fact that they were conducted years apart and did not include the same studies. This finding is reassuring for those who use SRs to inform clinical and public health decision making. In clusters of multiple SRs, some reviews might disagree with others because of biases in their conduct. To safely use existing SRs for decision making, we suggest that users: (1) do not rely only on their conclusions but, as a minimum, examine quantitative results such as effect estimates and assess their coherence with interpretation; (2) rely on prespecified and widely agreed primary composite outcomes but also examine their components, especially hard outcomes such as mortality; and (3) do not use subgroup analysis as a basis for recommendations, unless prespecified and strongly supported by analyses. Readers should ask themselves whether additional outcomes that are relevant to their own question or practice have been transparently reported and can be included in the clinical decision making. Because outcome selection, collection, and reporting is often inconsistent across SRs exploring the same objective, multiple reviews can be seen as useful confirmation of findings by independent groups rather than a waste of resources in evidence synthesis production. Discordance on significance of treatment effects was due to a combination of variation in design with inclusion of different studies and lack of precision for single hard outcomes compared with a composite outcome. Such inconsistencies among SRs could potentially slow the translation of SRs' results to clinical and public health decision making and suggest the need for a broader methodological and clinical agreement on their design.

Acknowledgments

We are grateful to Alessandro Liberati who conceived this study and helped to start it. Unfortunately, he passed away before seeing the completion of the review. We would also like to acknowledge the other members of the Discordant Team (Silvia Minozzi and so on). This study used search strategies kindly provided by Clinical Evidence,

published by the BMJ Evidence Centre, London, UK (RERF).

Supplementary data

Supplementary data related to this article can be found online at http://dx.doi.org/10.1016/j.jclinepi.2014.11.004.

References

[1] Moher D, Tetzlaff J, Tricco AC, Sampson M, Altman DG. Epidemiology and reporting characteristics of systematic reviews. Plos Med 2007;4(3):e78.

[2] Siontis KC, Hernandez-Boussard T, Ioannidis JP. Overlapping meta-analyses on the same topic: survey of published studies. BMJ 2013; 347:f4501.

[3] Antman EM, Lau J, Kupelnick B, Mosteller F, Chalmers TC. A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts. Treatments for myocardial infarction. JAMA 1992;268:240−8.

[4] Oxman AD, Guyatt GH. The science of reviewing research. Ann N Y Acad Sci 1993;703:125−33. discussion 133−4.

[5] Jadad AR, Cook DJ, Browman GP. A guide to interpreting discordant systematic reviews. CMAJ 1997;156(10):1411−6.

[6] Pladevall-Vila M, Delclos GL, Varas C, Guyer H, Brugues-Tarradellas J, Anglaa-Arisa A. Controversy of oral contraceptives and risk of rheumatoid arthritis: meta-analysis of conflicting studies and review of conflicting meta-analyses with special emphasis on analysis of heterogeneity. Am J Epidemiol 1996;144:1−14.

[7] Polderman KH, Ely EW, Badr AE, Girbes AR. Induced hypothermia in traumatic brain injury: considering the conflicting results of meta-analyses and moving forward. Intensive Care Med 2004;30(10):1860−4.

[8] Vavken P, Dorotka R. A systematic review of conflicting meta-analyses in orthopaedic surgery. Clin Orthop Relat Res 2009; 467(10):2723−35.

[9] Biondi-Zoccai GG, Lotrionte M, Abbate A, Testa L, Remigi E, Burzotta F, et al. Compliance with QUOROM and quality of reporting of overlapping meta-analyses on the role of acetylcysteine in the prevention of contrast associated nephropathy: case study. BMJ 2006;332:202−9.

[10] Moja L, Fernandez del Rio MP, Banzi R, Cusi C, D'Amico R, Liberti A, et al. Multiple systematic reviews: methods for assessing discordances of results. Intern Emerg Med 2012;7(6):563−8.

[11] Henderson RA, O'Flynn N, Guideline Development Group. Management of stable angina: summary of NICE guidance. Heart 2012; 98(6):500−7.

[12] Poolman RW, Abouali JA, Conter HJ, Bhandari M. Overlapping systematic reviews of anterior cruciate ligament reconstruction comparing hamstring autograft with bone-patellar tendon-bone autograft: why are they different? J Bone Joint Surg Am 2007;89:1542−52.

[13] Yusuf S, Wittes J, Probstfield J, Tyroler HA. Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials. JAMA 1991;266:93−8.

[14] Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003;327:557−60.

[15] Guyatt GH, Oxman AD, Kunz R, Brozek J, Alonso-Coello P, Rind D, et al. GRADE guidelines 6. Rating the quality of evidence−imprecision. J Clin Epidemiol 2011;64:1283−93.

[16] EPC W. Assessing equivalence and noninferiority. In Edition Agency for Healthcare Research and Quality; 2012.

[17] Ferreira-Gonzalez I, Busse JW, Heels-Ansdell D, Montori VM, Akl EA, Bryant DM, et al. Problems with use of composite end points in cardiovascular trials: systematic review of randomised controlled trials. BMJ 2007;334:786.

[18] Li LC, Moja L, Romero A, Sayre EC, Grimshaw JM. Nonrandomized quality improvement intervention trials might overstate the strength of causal inference of their findings. J Clin Epidemiol 2009;62: 959−66.

[19] Freemantle N, Calvert M, Wood J, Eastaugh J, Griffin C. Composite outcomes in randomized trials: greater precision but with greater uncertainty? JAMA 2003;289:2554−9.

[20] Freemantle N, Calvert M. Composite and surrogate outcomes in randomised controlled trials. BMJ 2007;334:756−7.

[21] Higgins JGS. Cochrane handbook for systematic reviews of interventions. The Cochrane Collaboration. Wiley; 2011.

[22] Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ 2008;336: 924−6.

[23] Page MJ, McKenzie JE, Forbes A. Many scenarios exist for selective inclusion and reporting of results in randomized trials and systematic reviews. J Clin Epidemiol 2013;66:524−37.

[24] Ioannidis JP. Contradicted and initially stronger effects in highly cited clinical research. JAMA 2005;294:218−28.

[25] Booth A, Clarke M, Ghersi D, Moher D, Petticrew M, Stewart L. An international registry of systematic-review protocols. Lancet 2011; 377:108−9.

[26] Williamson P, Altman D, Blazeby J Clarke M, Gargon E. Driving up the quality and relevance of research through the use of agreed core outcomes. J Health Serv Res Policy 17: 1−2.

[27] Rahimi K, Malhotra A, Banning AP, Jenkinson C. Outcome selection and role of patient reported outcomes in contemporary cardiovascular trials: systematic review. BMJ 2010;341:c5707.

[28] Hopayian K. The need for caution in interpreting high quality systematic reviews. BMJ 2001;323:681−4.

[29] Wen J, Ren Y, Wang L, Li Y, Liu Y, Zhou M, et al. The reporting quality of meta-analyses improves: a random sampling study. J Clin Epidemiol 2008;61:770−5.