Journal of Clinical Epidemiology 113 (2019) 147−158

Journal of Clinical Epidemiology

ORIGINAL ARTICLE

Several reasons explained the variation in the results of 22 meta-analyses addressing the same question

Assem M. Khamis[a], Mohamad El Moheb[b], Johny Nicolas[b], Ghida Iskandarani[b], Marwan M. Refaat[c], Elie A. Akl[a,c,d,*]

[a]*Clinical Research Institute, American University of Beirut Medical Center, Beirut, Lebanon*
[b]*Faculty of Medicine, American University of Beirut Medical Center, Beirut, Lebanon*
[c]*Department of Internal Medicine, American University of Beirut Medical Center, Beirut, Lebanon*
[d]*Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Ontario, Canada*

Accepted 22 May 2019; Published online 29 May 2019

Abstract

Objective: The objective of this study was to assess and to investigate the reasons for the variations between the results of meta-analyses addressing the same question.

Study Design and Setting: We included systematic reviews, and the trials that they included, on the use of implantable cardiac defibrillator (ICD) in patients with nonischemic cardiomyopathy. We assessed the variation between meta-analyses pooled effect estimates by calculating the percentage of absolute difference. We developed a list of 10 reasons for variations between the results of the meta-analyses clustered in four overarching categories.

Results: We identified 21 systematic reviews including six trials and reporting on 22 eligible meta-analyses. The percentage of absolute difference between each of the 22 pooled effect estimates (included odds ratio, risk ratio, hazard ratio) and their median value had an average of 3.2%. The number of trials for which the following categories of reasons for variations applied were as follows: (1) different decision to include or exclude trials ($n = 3$); (2) differences in analytical approaches ($n = 6$); (3) errors in conducting meta-analyses ($n = 5$); and (4) unclear reason ($n = 1$).

Conclusion: Few of the observed variations between the results of the 22 meta-analyses could lead clinicians or guideline development organizations to adopt different courses of actions. Variations were most frequently related to both errors and variations in trial eligibility and analytical approaches.

Keywords: Statistics methods; Systematic review; Meta-analysis; Time-to-event; Meta-research; Quality of research

1. Introduction

Systematic reviews have a key role in evidence-based clinical practice and in the development of clinical practice guidelines. Meta-analysis is an essential component of systematic reviews that provides a pooled effect estimate using data from individual clinical studies [1]. Traditional pairwise meta-analyses were developed over the past 30 years to address challenges such as statistical heterogeneity [2], confounders, and effect modifiers [3,4].

The number of systematic review publications has increased significantly over the past few years [5,6]. The number of meta-analyses increased by a factor of 20 times in 2014 compared to 1994 [6]. This increase was accompanied by many overlapping reviews, with about half of the reviews not mentioning the overlaps [6−8]. The duplication of systematic reviews is related to the fact that the conduct, writing, and editorial processing may take place without the respective authors or editor being aware [6]. As a startling example, Riaz et al. found 17 meta-analyses were duplicated using the results of only three randomized clinical trials on the management of patent foramen oval [6].

The conduct of meta-analyses has uncovered a number of challenges in their application or interpretation [9,10]. Riley et al. explored the statistical methods used in the

Conflicts of interest: All authors declared no financial conflict of interest. All authors contributed to a Cochrane review addressing the same research question.

Funding: Not funded.

* Corresponding author: Department of Internal Medicine, American University of Beirut Medical Center, P.O. Box: 11-0236 Riad-El-Solh, Beirut 1107 2020, Lebanon. Tel.: 00961 1 374374 ext. 5490.

E-mail address: ea32@aub.edu.lb (E.A. Akl).

https://doi.org/10.1016/j.jclinepi.2019.05.023
0895-4356/

What is new?

Key findings

- We identified 21 systematic reviews including six trials and reporting on 22 eligible meta-analyses addressing our clinical question of interest.
- The percentage of absolute difference between each of the 22 pooled effect estimates and their median value had an average of 3.2%.
- The number of trials for which the following categories of reasons for variations applied were as follows: (1) different decision to include or exclude trials ($n = 3$); (2) differences in analytical approaches ($n = 6$); (3) errors in conducting meta-analyses ($n = 5$); and (4) unclear reason ($n = 1$).

What this adds to what was known?

- This is the first study to systematically investigate reasons for variations between the results of meta-analyses addressing the same research question.
- The variations between the results of the meta-analyses addressing the same question could lead clinicians or guideline development organizations to adopt different courses of actions.

What is the implication and what should change now?

- Systematic reviewers should have detailed, explicit, and preplanned meta-analytical approaches.
- Systematic reviewers need to involve reviewers with experience in meta-analytical methods and consider conducting meta-analyses in duplicate and independently.

Cochrane pregnancy and childbirth reviews [10]. They found that more than third of the reviews used fixed-effect model in spite of the heterogeneity (I^2) being >25%; and around two-thirds of the reviews did not explore moderate to high heterogeneity [10]. Another study found discrepancy between planned interaction analyses (treatment by covariate) and applied analyses, and none of the reviews provided an external evidence for the covariates selection [11]. A recent study revealed that, even when given the same data set, 29 teams of 61 analysts had substantive differences in both the analytical approaches used and the resulting estimated effect sizes [12].

As an illustrative example, meta-analysis of time-to-events data is particularly challenging because individual studies do not frequently report all the needed statistical details (e.g., hazard ratios [HRs]) to conduct meta-analysis [13]. To address those challenges, Parmar et al. developed detailed methodology of HR estimation [14], and then Tierney et al. provided a supplementary excel worksheet to facilitate the calculations for nonstatisticians [13]. The PRISMA statement (Preferred Reporting Items for Systematic reviews and Meta-Analyses) highlighted this issue and emphasized the importance of reporting the methodology used for HR estimation [15].

We recently conducted a Cochrane systematic review of the effects of implantable cardiac defibrillators (ICDs) in comparison with usual care patients with nonischemic cardiomyopathy [16]. As part of that process, we identified 20 other systematic reviews addressing the same clinical question with the same primary outcome, that is, mortality. Although most meta-analyses appeared to include the same primary studies, the trials effect estimates and the pooled effect estimates varied. There is an increasing interest in exploring the causes of discrepant results between multiple systematic reviews addressing the same question [17]. The objective of this study was to assess and investigate the reasons for the variations between the results of meta-analyses addressing the same question.

2. Methods overall design

2.1. Eligibility criteria

We included systematic reviews of randomized trials comparing the use of ICD to optimal medical therapy in patients with nonischemic cardiomyopathy. We did not exclude systematic reviews that included comparisons other than the one of interest (e.g., use of ICD in patients with ischemic cardiomyopathy). We included meta-analyses for ''all-cause mortality,'' the only outcome for which all eligible systematic reviews reported at least one meta-analysis. We excluded sensitivity and subgroup meta-analyses. We assessed for eligibility trials included in at least one of the eligible meta-analyses. This study did not require an ethical approval and there was no patient or consumer involvement at any level of the study.

2.2. Search

We used the same search strategy we used for conducting our Cochrane systematic review to identify other relevant systematic reviews. We searched the following electronic databases: MEDLINE, EMBASE, Web of Science Core collection, and the Cochrane library. Appendix 1 includes the detailed search strategy. After running the initial search in August 2017, we used the same search strategy to set up search alerts in all databases. In addition, we searched in July 2018 Epistemonikos, the ''world's largest systematic review database,'' for relevant systematic reviews published in the 2016–2018 time range [18].

2.3. Study selection and data extraction

Four reviewers screened titles and abstracts, screened full texts, and abstracted data into structured Microsoft Office Excel sheets. We assessed the quality of eligible systematic reviews published as full texts using AMSTAR two instrument [19]. Two external reviewers assessed the quality of our own Cochrane review. We conducted all aforementioned steps in duplicate and independently, except for abstracting statistical data, which we conducted in triplicate. We resolved disagreements through discussion, or with the help of an additional reviewer if needed.

We abstracted data from eligible systematic reviews, eligible meta-analyses, and trials included in those meta-analyses. For each eligible systematic review, we collected data on the following characteristics:

- Type of document (full text vs. conference abstract).
- Specialty of the publication journal according to the Web of Science (cardiology vs. noncardiology).
- Journal impact factor.
- Date of the literature search.
- Number of included trials.

For each eligible meta-analysis, we collected data on the following analytical characteristics:

- Number of included trials.
- Number of participants included in the analysis.
- Statistical software used.
- Type of effect measure (e.g., HR, risk ratio [RR], odds ratio [OR]).
- Number of participants randomized to each arm.
- Type of effect measure (HR, RR, OR).
- Relative effect estimate (point estimate, upper and lower boundaries of the confidence interval).

2.4. Analysis

2.4.1. Descriptive analyses

We conducted descriptive analyses for the characteristics of included systematic reviews and meta-analyses. We used frequency and percentages for categorical variables, and median and interquartile range (IQR) for continuous variables.

2.4.2. Variation in trial effect estimates used across meta-analyses

For each trial-meta-analysis dyad, we calculated the percentage of absolute difference between the relative effect estimate as extracted from the trial (trial EE) report and the effect estimate as extracted from the meta-analysis forest plot (trial EE in MA) as follows:

$$\%\ \text{of absolute difference} = \frac{|\textit{trial EE in MA} - \textit{trial EE}|}{\textit{trial EE}} \times 100$$

For trials that did not report a relative effect estimate, we used instead the median of the effect estimates extracted from the meta-analyses and calculated the percentage of absolute difference as follows:

$$\%\ \text{of absolute difference} = \frac{|\textit{trial EE in MA} - \textit{median EE across MA}|}{\textit{median EE across MA}} \times 100$$

- Analysis model (i.e., fixed effect, random effects).
- Any justification for the choice of the analysis model.
- Statistical method (e.g., inverse variance, Mantel-Haenszel).
- Whether additional analyses (i.e., subgroup, sensitivity) were conducted.
- Pooled relative effect estimate (point estimate, lower and upper boundaries of the CI) (as displayed in the forest plot).
- Absolute risk difference (point estimate, lower and upper boundaries of the CI) and baseline risk used as reported in the systematic review.

For each eligible trial, we collected data on the following characteristics from the trial report itself and from each of the 22 included meta-analyses:

- Whether the study was included in the meta-analysis.
- Duration(s) of outcome follow-up.
- Number of events in each arm.

Then, to show the variation for each trial, we present the mean and range of the percentage of absolute difference across meta-analyses. We conducted similar analysis to show the variation across pooled effect estimates. In addition, we used scatter plots to illustrate the aforementioned variations.

2.4.3. Investigating reasons for variation in used trial effect estimates

To investigate the variation in trial effect estimates used across meta-analyses, we used an iterative process of feedback and refinement through face-to-face meetings and poster presentation at Cochrane colloquium for further feedback during which we identified all potential reasons. Box 1 lists our tentative list of reasons for variations for relative trial effect estimates.

To assess reason #3 (Use of different measures of effect estimates [RR vs. OR vs. HR]), we attempted to guess which of these measures the meta-analysis used as follows: for each trial, and whenever the numbers of events in both

Box 1 Reasons different meta-analyses use different relative effect estimates for the same trial

Different decision to include or exclude one or more trial(s)

1. Different inclusion/exclusion decision.
   - Trial was not available at the time of search.
   - Trial was available, but it was not included by the systematic review (missed or excluded).
   - The trial was included in the systematic but excluded from the comparison of interest.
   - The trial was included in the comparison of interest, but only in a secondary meta-analysis.
   - The trial was included in the meta-analysis of interest.

Different analytical approaches

2. Use of data for different follow-up times.
3. Use of different measures of effect estimates (e.g., one meta-analysis uses RR while another one uses HR for the same trial).
4. Use of different methods to calculate HR (when not reported by trial).

Errors in conducting meta-analysis

5. Use of data for the inappropriate population.
6. Use of data for the inappropriate comparators.
7. Nonconsideration of the alpha value (type 1 error) used to calculate the effect estimate confidence interval by the trialists (e.g., 97.5% CI).
8. Use of the inappropriate trial effect measure in meta-analysis (e.g., using an HR estimate from the trial in a meta-analysis using OR).
9. Error in data abstraction/entry.

Unclear reason

arms were reported, we calculated the different effect measures (OR, HR, and RR). Then, we compared those measures with the effect estimates reported for that trial in each meta-analysis.

Our tentative list of potential reasons for variations in trial absolute effect estimates included:

- Time frame for which the absolute risk difference was calculated.
- Baseline risk used in calculating.
- Effect estimate (95% CI) used in calculating absolute risk reduction (ARR).

When trying to figure out how the systematic reviewers calculated their absolute effects based on HR, we used the formula proposed by Altman and Anderson [20].

3. Results

We identified 21 eligible systematic reviews that addressed the comparison of ICD with standard care or optimal medical therapy. These reviews were published in either 2017 or 2018 [16,21−40]. The number of studies included in these 21 systematic reviews ranged from 4 trials to 11 trials. These 21 reviews reported a total of 22 eligible meta-analyses for ‘‘all-cause mortality’’ outcome, that is, one review reported two eligible meta-analyses (using RR and HR, respectively) [23]. The 22 meta-analyses included a total of seven trials, out of which we judged six to be eligible. These six trials were published between 2002 and 2016 [41−47]. We excluded the seventh trial because 10 out of its 19 participants with nonischemic cardiomyopathy were patients with heart transplantation [46]. Only one of the 21 reviews had included that trial in one of two meta-analyses it reported [23].

3.1. Characteristics of systematic reviews

Table 1 reports the characteristics of the 21 included systematic reviews. Most reviews were published as full-text publications (76%), and in cardiology journals (86%). The median impact factor of publishing journals was 6.059 (IQR 2.773−19.896).

Figure A1 summarizes the AMSTAR two assessment of the methodological quality of the 16 systematic reviews published as full texts. The assessment showed serious limitations to different degrees for the included systematic reviews. The vast majority of reviews failed to explain their

Table 1. General characteristics of the eligible systematic reviews ($N = 21$)

Characteristics	*n* (%)
Type of document	
Full text	16 (76)
Conference abstract	5 (24)
Journal specialty	
Cardiology	18 (86)
Noncardiology	3 (14)
Journals impact factor [median (IQR)]	6.059 (2.773−19.896)
Date of the literature search	
Not reported	5 (24)
2016	12 (57)
2017	4 (19)
Number of included trials in the systematic review	
11 trials	2 (10)
8 trials	2 (10)
6 trials	9 (42)
5 trials	6 (28)
4 trials	2 (10)

Abbreviation: IQR, interquartile range.

Table 2. Analytical characteristics of the eligible meta-analyses (summary data) ($N = 22$)

Analytical characteristics	*n* (%)
Number of included trials[b]	
6 trials	11 (49)
5 trials	9 (41)
4 trials	2 (10)
Total number of participants [median (range)]	2917 (2573 – 2992)
Statistical software used[a]	
Review Manager (RevMan)	9 (41)
Comprehensive Meta-Analysis (CMA)	3 (14)
Stata	3 (14)
R	2 (10)
Not reported	6 (27)
Type of effect measure used[a]	
Risk ratio	10 (45)
Hazard ratio	9 (41)
Odds ratio	3 (14)
Analysis model	
Random effects	15 (68)
Fixed effect	6 (27)
Both	1 (5)
Justified using random-effects model	9 (41)
Statistical method[a]	
Inverse variance	9 (41)
Mantel-Haenszel	6 (27)
Not reported	7 (32)
Additional analyses	
No	12 (54)
Subgroup analysis	5 (22)
Meta-regression	2 (10)
Both	3 (14)

[a] More than one choice could apply.
[b] One review included two comparisons for the same trial.

selection of the study designs for inclusion in the review (95%), report on the "a priori establishment of the review methods" (90%), and report on the sources of funding (90%) for the studies included in the review.

3.2. Characteristics of meta-analyses

Table 2 summarizes the analytical characteristics of the 22 included meta-analyses (Table A1 details the characteristics for each meta-analysis). Most meta-analyses (90%) included either six trials or five trials in their primary meta-analyses. The pooled effect measures used were RR (45%), HR (41%), or OR (14%). Two-thirds of the reviews (68%) used the random-effects model.

3.3. Variation in relative effect estimates

Table A2 provides for each trial (listed in the top row) the point effect estimates and the associated confidence intervals as extracted from the forest plot of each meta-analysis (listed in the first column). In addition, Table A2 provides the pooled effect estimate and the associated confidence interval in each included meta-analysis. Figures 1 and 2 present the same information in graphical format for the pooled relative effect estimate and the individual trials' relative effect estimates.

The last row of Table A2 presents for each trial and for the meta-analysis the mean and range of the percentage of absolute difference. The range of means of percentage of absolute difference of the six trials was 1.1–11.4%. The percentage of absolute difference for the pooled effect estimates was 3.2%.

3.4. Explanation of variation in relative effect estimates

A number of variables included in Table 2 reflect the variation in the analytical characteristics of the eligible meta-analyses that could have contributed to the variation in their pooled relative effect estimates.

Tables 3 and A3-A8 present respectively for each of the six trials: (1) the detailed statistical information used in each of the meta-analyses that included that trial; (2) the statistical information as abstracted by our team from the trial report (third row); (3) our explanations of any variations between the effect estimate reported by the trial and the one displayed in the forest plot of the respective meta-analyses (last column). These explanations are categorized according to the "reasons different meta-analyses use different effect estimates for the same trial" provided in Box 1. The first four reasons reflect differences in decisions that systematic reviewers may reasonably make when conducting meta-analysis. The next five reasons reflect errors that could be due to either the lack of a clear and specific analysis plan or a lack of attention when executing that plan. Unfortunately, in two cases, the reason for variation remained unclear even after thorough investigation.

All eligible trials were available at the time the included systematic reviews conducted their literature searches. However, three trials were either excluded or missed by at least one systematic review, respectively. One trial was excluded, respectively, from one systematic review, and from the primary meta-analysis of another systematic review [43]. One trial reported two follow-up time points for the primary outcome (1 year and 7 years) and then had the 1-year follow-up used in only one meta-analysis. For each of the six trials, different effect measures of effect estimates were either extracted (HR) or calculated (RR, OR, HR) from the trials and then used in the meta-analyses. When an HR was calculated for two of the six trials, the systematic reviews used different methodologies (e.g., Parmar or Guyot) [14,48]. For each of the six trials, authors of a varying number of systematic reviews appear to have made errors in conducting their meta-analyses.

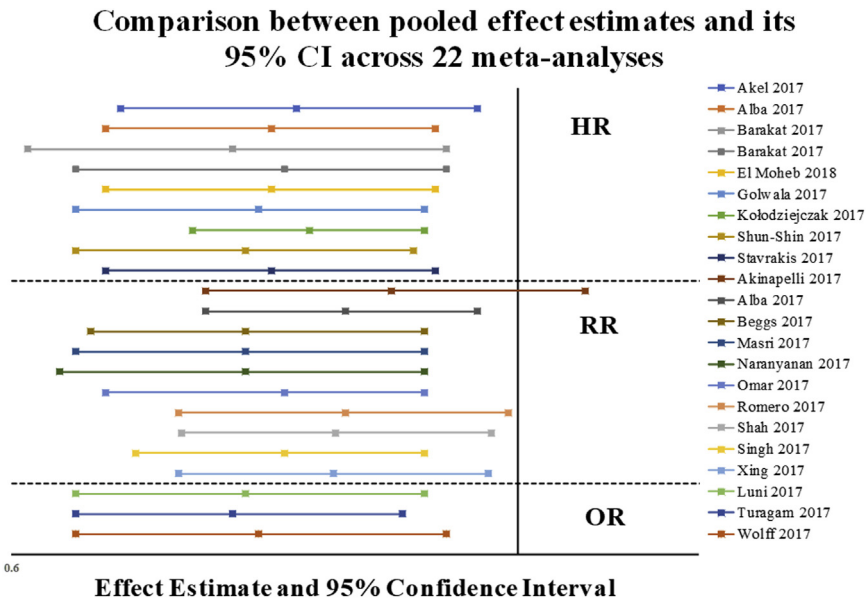

Fig. 1. The variation in pooled effect estimates and its 95% CI across the 22 meta-analyses.

3.5. Variation in absolute risk reduction

Four systematic reviews reported on the ARR for all-cause mortality outcome [16,23,25,33]. Table 4 presents the relevant statistical information as well as the possible explanations for variation in ARR. None of the systematic review reported on the time frame for which the absolute risk difference was calculated. Two systematic reviews reported the baseline risks used to estimate ARR and used GRADEpro to calculate ARR. The two other reviews were unclear about how the baseline risk they used or how they estimated ARR.

4. Discussion

4.1. Summary of findings

We identified 21 systematic reviews including six trials and reporting on 22 eligible meta-analyses addressing our clinical question of interest. The objective of this study was to assess and investigate the reasons for variations between the results of these 22 meta-analyses. The percentage of absolute difference between each of the 22 pooled effect estimates and their median value had an average of 3.2%. Similar differences were observed for each of the six trials when calculating the

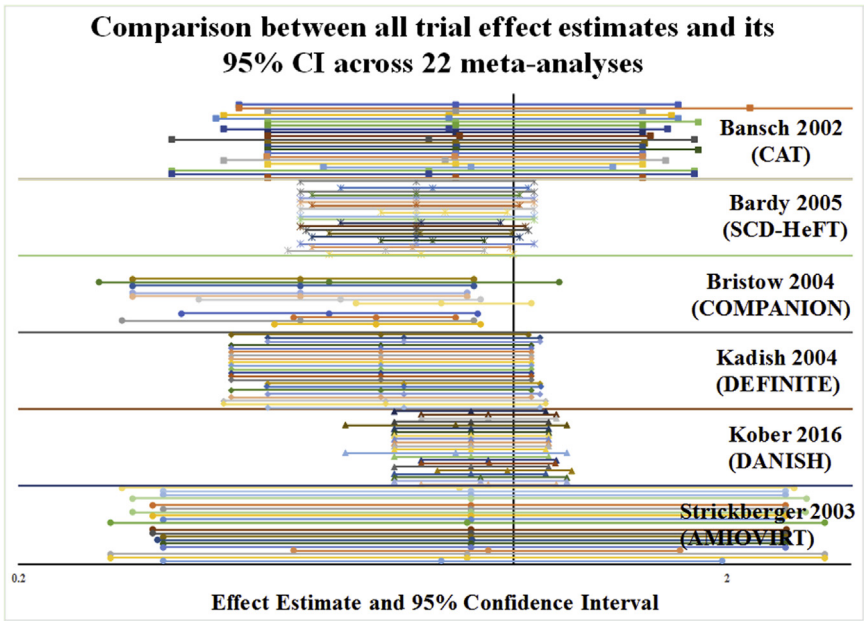

Fig. 2. The variation in relative effect estimates and its 95% CI of each trial across the 22 meta-analyses.

Table 3. Possible explanations for variation in trial effect estimates used in the 22 meta-analyses, by trial

Possible explanations	Bänsch 2002 (CAT)	Bardy 2005 (SCD-HeFT)	Bristow 2004 (COMPANION)	Kadish 2004 (DEFINITE)	Køber 2016 (DANISH)	Strickberger 2003 (AMIOVIRT)
Different decision to include or exclude one or more trial(s)	n/N (%)	n/N (%)	n/N (%)	n/N (%)	n/N (%)	n/N (%)
1. Differences in trial inclusion						
Trial was not available at the time of search	-	-		-	-	-
Trial was available, but it was not included by the systematic review (missed or excluded)	-	1/22 missed	2/22 missed 8/22 excluded	-	-	2/22 excluded
The trial was included in the systematic but excluded from the comparison of interest	-	-	1/22	-	-	
The trial was included in the comparison of interest, but only in a secondary meta-analysis			1/22			
The trial was included in the meta-analysis of interest	22/22	21/22	10/22	22/22	22/22	20/22
Differences in analytical approaches	n/N (%)	n/N (%)	n/N (%)	n/N (%)	n/N (%)	n/N (%)
2. Use of data for different follow-up times						
1-yr follow-up	1/22 (5)	NA	NA	NA	NA	NA
7-yr follow-up	21/22 (95)	NA	NA	NA	NA	NA
3. Use of different measures of effect estimates (RR vs. OR vs. HR)						
HR	5/22 (23)	11/21 (52)	8/10 (80)	14/22 (64)	14/22 (64)	3/20 (15)
RR	13/22 (55)	1/21 (5)	-	6/22 (27)	5/22 (23)	13/20 (65)
OR	3/22 (14)	-	-	2/22 (9)	2/22 (9)	3/20 (15)
Unclear	1/22 (5)	9/21 (43)	2/10 (20)	-	1/22 (5)	1/20 (5)
4. Use of different methods to calculate HR (when not reported by trial)		N/A (HR reported)		N/A (HR reported)	N/A (HR reported)	
Parmar or Tierney	3/5 (60)	NA[a]	-	NA[a]	NA[a]	2/3 (67)
Guyot	2/5 (40)	NA[a]	-	NA[a]	NA[a]	1/3 (33)
Indirect method	NA	NA[a]	1/1 (100)	NA[a]	NA[a]	NA
Errors in conducting meta-analyses	n/N (%)	n/N (%)	n/N (%)	n/N (%)	n/N (%)	n/N (%)
5. Use of data for the inappropriate population	-	-	2/10 (20)	-	-	-
6. Use of data for the inappropriate comparators	-	-	7/10 (70)	-	2/22 (9)	-
7. Nonconsideration of the alpha used to calculate the effect estimate confidence interval by the trialists	N/A (only 95% CI reported)	-	N/A (only 95% CI reported)	N/A (only 95% CI reported)	N/A (only 95% CI reported)	N/A (only 95% CI reported)

(*Continued*)

Table 3. Continued

Possible explanations	Bänsch 2002 (CAT)	Bardy 2005 (SCD-HeFT)	Bristow 2004 (COMPANION)	Kadish 2004 (DEFINITE)	Køber 2016 (DANISH)	Strickberger 2003 (AMIOVIRT)
95% CI recalculated		3/11 (27)				
97.5% as is		8/11 (73)				
8. Use the inappropriate effect measure in meta-analysis	-	3/21 (14)	3/10 (30)	4/22 (18)	4/22 (18)	-
9. Error data abstraction/ entry	-	-	-	-	-	1/20 (5)
Unclear	2/22 (9)[b]	-	-	-	-	-

Abbreviations: HR, hazard ratio; OR, odds ratio; RR, risk ratio.
[a] NA as HR was reported.
[b] Asymmetrical 95% CI.

percentage of absolute difference between the relative effect estimate as extracted from the trial report and the effect estimate as extracted from the meta-analysis forest plot. The number of trials affected, respectively, by the three overarching reasons for variations in the findings of 22 meta-analyses were as follows: (1) different decision to include or exclude trials ($n = 3$); (2) differences in analytical approaches ($n = 6$); (3) errors in conducting meta-analyses ($n = 5$); and (4) unclear reason ($n = 1$).

4.2. Strengths and limitations

This is one of few studies we are aware of that assessed and investigated the reasons for variations between the results of meta-analyses addressing the same research question [49,50]. Unlike our study, these two studies were descriptive and did not systematically investigate the reasons for variability. One study identified and described the conflicting meta-analyses certain topics [49], and the other reported on sources of discrepancies such as outcome selection, subgroup analyses, and interpretations of results [50].

In addition, this is one of few studies we are aware of, that highlight the flaws in the application of meta-analyses in systematic reviews [9]. We followed rigorous methodology, for example, evaluated the quality of included systematic reviews in duplicate and abstracted statistical data in triplicate. A major outcome of this study is the elaboration of a list of potential reasons of variation in effect estimates across meta-analyses. One limitation of this study is that we did not contact the systematic reviews' authors in situations where the reason for variation was unclear.

4.3. Interpretation of findings

We found an average of 3.2% percentage of absolute difference for the pooled effect estimates. Although interpreting such value is not intuitive, a relevant approach would consider whether the different pooled effect estimates could lead to different courses of clinical action. For example, the confidence interval for the pooled effect estimate for most meta-analyses suggests benefit, that is, the upper limit of the confidence interval indicates a clinically meaningful reduction in mortality. However, this is not the case for three meta-analyses where the upper limit of the confidence interval indicates either harm (e.g., Akinpelli 2017) [22] or a negligible benefit (e.g., Romero 2017, Shah 2017)([33,34]). This would become critical if clinicians or guideline development organizations adopt different courses of actions because the underlying synthesized evidence was based on different eligibility criteria, different analytical approaches, or errors.

Our findings showed multiple reasons for variation between the results of these 22 meta-analyses. First, the most common reason appeared to be the variation in the measure of effect estimate used by the different meta-analyses. This problem was compounded by the fact that most of the reviews did not clearly report which effect measures they used for the trials included in their meta-analysis. In many cases, we had to guess the effect measure used by the meta-analysis, by calculating the trial's OR, RR, and HR with 95% CI and compared them with the effect estimate used for that trial in the meta-analysis of interest.

Second, some reviews used for one or more trials an effect measure different than the one used in the meta-analysis, which led to pooling of HR and RR or HR and OR within the same meta-analysis. Although this might be related to the lack of experience in conducting meta-analyses, it is made worse by the poor reporting of how the reviewers handled the statistical data.

A third reason of variability between the results of the 22 meta-analyses was the different approaches to dealing with a trial that did not directly report data for the comparison of interest (i e., ICD vs. control). For example, the COMPANION study did not report effect estimates for the comparison of interest, but for two comparisons (i.e., CRT-D vs. pharmacological therapy & CRT vs. pharmacological

Table 4. Possible explanations for variation in absolute risk reduction, by systematic review

Study ID	Number of studies	Time frame for which absolute risk difference calculated	Baseline risk used in calculating ARR	Effect estimate (95% CI) used in calculating ARR	Absolute risk difference (95% CI)	Explanations
Alba 2017	5	Not reported	24.0%	HR 0.78 (0.66–0.92)	47/1000 (17 fewer to 74 fewer)	• Baseline risk calculated based on the number of events when reported by the studies, otherwise by estimating the number of events from Kaplan–Meier mortality curves • Absolute risk difference calculated in GRADEpro (using HR and calculated baseline risk)
Barakat 2017	5	Not reported	Not clear (although paper reported a "weighted incidence" of 18.0%)	HR 0.79 (0.64–0.93)	30/1000	• Not clear which baseline risk was used • Not clear how absolute risk difference was calculated
El Moheb 2018	4	Not reported	21.6%	HR 0.78 (0.66–0.92)	43/1000 (15 fewer to 68 fewer)	• Baseline risk calculated from the total number of events over the total number of participants in the control group, from the four trials that reported the number of events from nonischemic subgroups or groups • Calculated absolute risk difference in GRADEpro (using HR and the calculated baseline risk)
Romero 2017	5	Not reported	Not clear	RR 0.84 (0.71–0.99)	38/1000	• Not clear which baseline risk was used • Not clear how absolute risk difference was calculated

Abbreviations: ARR, absolute risk reduction; HR, hazard ratio; RR, risk ratio.

therapy) that could allow an indirect comparison (i.e., CRT-D vs. CRT, which is equivalent to ICD vs. control). While one review conducted an indirect comparison, 10 reviews excluded the trial, while the rest included the effect estimate for the inappropriate comparison (i.e., CRT-D vs. pharmacological therapy). This reason reflects problems with both attention to the comparison of interest and the need for statistical expertise to run indirect comparisons.

An additional reason of variation was the differences in the trial eligibility criteria across systematic reviews [7,51]. In our study, we found that half of the meta-analyses did not include all the six trials. Doak et al. identified nine types of differences in inclusion and exclusion criteria. Examples of those differences include the inclusion of non–peer review articles, inclusion of participants with different risk group, and inclusion of pilot studies [51].

The variation in the ARR is caused by both the variation of the pooled relative effect estimates used and the variation in the baseline risk used. Out of four systematic reviews reporting ARR, only two used the same pooled effect estimate while none of the reviews used the same baseline risk as the others. The variability in ARR can have a negative impact on interpreting the evidence, particularly in the setting of guideline development. There is a need for following standardized approaches to calculating ARRs. That includes the choice of the baseline risk time frame, the choice of the source of baseline risk, the choice of the relative effect, and the use of appropriate statistical methods, particularly when using time-to-event data [20].

Our study sheds the light on a broader challenge in the field of evidence synthesis, that is, the replication of systematic reviews [6–8]. In the case addressed in this study, the ratio of 21 systematic reviews to six original trials is stunning. The latest RCT (the DANISH trial) published in 2016 was a long-awaited trial given it had the largest number of participants up to that point [45]. We believe that was the main trigger for the subsequent publication of 21 systematic reviews; it is possible that more could have been conducted. Irrespective of the cause of this systematic review replication, it does reflect the enormous inefficiency and waste of research efforts, time, and money.

5. Conclusions

5.1. Implication for practice

We do not believe that our findings have major implications for clinical practice. That is because the differences

among effect estimates are clinically negligible (percentage of absolute difference between each of the 22 pooled effect estimates and their median value had a relatively low average of 3.2%) and have similar clinical implications.

However, the findings do have implications for the practice of conducting systematic reviews. Indeed, the quality assessment of a systematic review should reflect whether its findings are valid. Practically, quality assessment tools (e.g., the AMSTAR tool) focus on process aspects (e.g., performing study selection in duplicate), as surrogates for the validity of findings. This study provides evidence that statistical methods are critical for the validity of the results of a systematic review. Unfortunately, AMSTAR does not address in enough details the statistical methods used in systematic reviews and subsequently is not sensitive to these methods' variations as described in this study.

Investigations of variability between the results of meta-analyses addressing the same question need to consider the potential reasons listed in Box 1. There is evidence that variation in results of analyses may be difficult to avoid ''even by experts with honest intentions'' [12]. Therefore, systematic review authors should publish detailed, explicit, and preplanned meta-analytical approaches, for example, through protocol registration, in the international prospective register of systematic reviews (PROSPERO). Systematic review registration can also reduce the problem of replication of systematic reviews. Additional considerations are to involve reviewers with experience in meta-analytical methods and use quality assurance ''checks'' when conducting the meta-analysis; this could entail conducting meta-analyses in duplicate and independently for critical analyses or when resources are available. Major organizations producing systematic reviews (e.g., Cochrane and Campbell Collaboration) can facilitate the implementation of these recommendations by including them in their methodological standards to conducting reviews.

5.2. *Implications for future research*

Future research should investigate how best to measure the extent of variability between the results of meta-analyses addressing the same question, and the impact of that variability on the interpretation and translation into practice. More research is needed to better understand the reasons for such variability. In addition, future revisions of AMSTAR should consider process issues related to meta-analysis reported by the reviewers. Similarly, research work is needed to investigate the reasons for variation between systematic reviews, beyond variation in the results of meta-analyses including the same trials (i.e., variation in eligibility criteria, date of search, included studies). Finally, the evidence synthesis community needs to address the problem of replication of systematic reviews and develop rules for when it is warranted and when it is not.

CRediT authorship contribution statement

Assem M. Khamis: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Validation, Visualization, Writing - original draft, Writing - review & editing. **Mohamad El Moheb:** Data curation, Investigation, Validation, Writing - review & editing. **Johny Nicolas:** Data curation, Investigation, Validation, Writing - review & editing. **Ghida Iskandarani:** Data curation, Investigation, Validation, Writing - review & editing. **Marwan M. Refaat:** Investigation, Writing - review & editing. **Elie A. Akl:** Conceptualization, Methodology, Formal analysis, Investigation, Methodology, Supervision, Visualization, Writing - original draft, Writing - review & editing.

Acknowledgments

The authors would like to acknowledge Ms. Lara Kahale and Ms. Amena El-Harakeh for assessing the methodological quality of the Cochrane review using AMSTAR 2. The study received no funding.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jclinepi.2019.05.023.

References

[1] Higgins J, Green S. Cochrane handbook for systematic reviews of interventions version 5.1.0. The Cochrane Collaboration, Confidence intervals 2011. Available at: www.handbook.cochrane.org.
[2] Higgins JP, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. J R Stat Soc Ser A Stat Soc 2009; 172(1):137−59.
[3] Borenstein M, Higgins JP. Meta-analysis and subgroups. Prev Sci 2013;14(2):134−43.
[4] Thompson SG, Higgins JP. How should meta-regression analyses be undertaken and interpreted? Stat Med 2002;21:1559−73.
[5] Starr M, Chalmers I, Clarke M, Oxman AD. The origins, evolution, and future of the Cochrane database of systematic reviews. Int J Technol Assess Health Care 2009;25(S1):182−95.
[6] Riaz IB, Khan MS, Riaz H, Goldberg RJ. Disorganized systematic reviews and meta-analyses: time to systematize the conduct and publication of these study overviews? Am J Med 2016;129(3):339.e11−8.
[7] Campbell J, Bellamy N, Gee T. Differences between systematic reviews/meta-analyses of hyaluronic acid/hyaluronan/hylan in osteoarthritis of the knee. Osteoarthritis Cartilage 2007;15:1424−36.
[8] Pieper D, Antoine S-L, Mathes T, Neugebauer EA, Eikermann M. Systematic review finds overlapping reviews were not mentioned in every other overview. J Clin Epidemiol 2014;67:368−75.
[9] Page MJ, Altman DG, McKenzie JE, Shamseer L, Ahmadzai N, Wolfe D, et al. Flaws in the application and interpretation of statistical analyses in systematic reviews of therapeutic interventions were common: a cross-sectional analysis. J Clin Epidemiol 2018;95:7−18.
[10] Riley RD, Gates S, Neilson J, Alfirevic Z. Statistical methods can be improved within Cochrane pregnancy and childbirth reviews. J Clin Epidemiol 2011;64:608−18.

[11] Donegan S, Williams L, Dias S, Tudur-Smith C, Welton N. Exploring treatment by covariate interactions using subgroup analysis and meta-regression in Cochrane reviews: a review of recent practice. PLoS One 2015;10:e0128804.
[12] Silberzahn R, Uhlmann EL, Martin D, Anselmi P, Aust F, Awtrey EC, et al. Many analysts, one dataset: making transparent how variations in analytical choices affect results. Adv Methods Pract Psychol Sci 2018;1:337−56.
[13] Tierney JF, Stewart LA, Ghersi D, Burdett S, Sydes MR. Practical methods for incorporating summary time-to-event data into meta-analysis. Trials 2007;8(1):16.
[14] Parmar MK, Torri V, Stewart L. Extracting summary statistics to perform meta-analyses of the published literature for survival end-points. Stat Med 1998;17:2815−34.
[15] Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS Med 2009;6(7): e1000100.
[16] El Moheb M, Nicolas J, Khamis AM, Iskandarani G, Akl EA, Refaat M. Implantable cardiac defibrillators for people with non-ischaemic cardiomyopathy. Cochrane Database Syst Rev 2018; 12(12):CD012738.
[17] Katsura M, Kuriyama A, Tada M, Yamamoto K, Furukawa TA. Redundant systematic reviews on the same topic in surgery: a study protocol for a meta-epidemiological investigation. BMJ open 2017; 7(8):e017411.
[18] Rada G, Pérez D, Capurro D. Epistemonikos: a free, relational, collaborative, multilingual database of health evidence. MedInfo 2013;2013:486−90.
[19] Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. BMJ 2017;358:j4008.
[20] Altman DG, Andersen PK. Calculating the number needed to treat for trials where the outcome is time to an event. BMJ 1999;319:1492−5.
[21] Akel T, Lafferty J. Implantable cardioverter defibrillators for primary prevention in patients with nonischemic cardiomyopathy: a systematic review and meta-analysis. Cardiovasc Ther 2017;35(3):e12253.
[22] Akinapelli A, Patel A, Saadi A, Azzouz MS, Kanmanthareddy A, Saurav A, et al. Role of implantable defibrillators in primary prevention of mortality in non-ischemic cardiomyopathy patients. J Am Coll Cardiol 2017;69:385.
[23] Alba AC, Foroutan F, Posada JD, Battioni L, Schofield T, Alhussein M, et al. Implantable cardiac defibrillator and mortality in non-ischaemic cardiomyopathy: an updated meta-analysis. Heart 2017;104:230−6.
[24] Al-Khatib SM, Pokorney SD. Primary prevention implantable cardioverter defibrillators in patients with nonischemic cardiomyopathy: diminishing returns with advancing age? Am Heart Assoc 2017; 136:1781−3.
[25] Barakat AF, Saad M, Elgendy AY, Mentias A, Abuzaid A, Mahmoud AN, et al. Primary prevention implantable cardioverter defibrillator in patients with non-ischaemic cardiomyopathy: a meta-analysis of randomised controlled trials. BMJ open 2017;7(6): e016352.
[26] Beggs SA, Jhund PS, Jackson CE, McMurray JJ, Gardner RS. Non-ischaemic cardiomyopathy, sudden death and implantable defibrillators: a review and meta-analysis. Heart 2017;104:144−50.
[27] Golwala H, Bajaj NS, Arora G, Arora P. Implantable cardioverter-defibrillator for nonischemic cardiomyopathy: an updated meta-analysis. Circulation 2017;135:201−3.
[28] Kołodziejczak M, Andreotti F, Kowalewski M, Buffon A, Ciccone MM, Parati G, et al. Implantable cardioverter-defibrillators for primary prevention in patients with ischemic or nonischemic cardiomyopathy: a systematic review and meta-analysis. Ann Intern Med 2017;167:103−11.
[29] Luni FK, Singh H, Khan AR, Malik SA, Khawaja O, Riaz H, et al. Mortality effect of ICD in primary prevention of nonischemic cardiomyopathy: a meta-analysis of randomized controlled trials. J Cardiovasc Electrophysiol 2017;28(5):538−43.
[30] Masri A, Hammadah M, Adelstein E, Jain S, Saba S. Implantable cardioverter defibrillator in non-ischemic cardiomyopathy: a meta-analysis of randomized controlled trials. Cardiovasc Diagn Ther 2017;7(4):397.
[31] Narayanan MA, Vakil K, Reddy YN, Baskaran J, Deshmukh A, Benditt DG, et al. Efficacy of implantable cardioverter-defibrillator therapy in patients with nonischemic cardiomyopathy: a systematic review and meta-analysis of randomized controlled trials. JACC Clin Electrophysiol 2017;3(9):962−70.
[32] Omar WA, Parashar A, Khera R, Pandey A. Implantable cardioverter defibrillators for the prevention of mortality in patients with non-ischemic cardiomyopathy: a meta-analysis of randomized control trials. J Am Coll Cardiol 2017;69:904.
[33] Romero J, Chaudhary R, Garg J, Lupercio F, Shah N, Gupta R, et al. Role of implantable cardioverter defibrillator in non-ischemic cardiomyopathy: a systematic review and meta-analysis of prospective randomized clinical trials. J Interv Card Electrophysiol 2017;49(3): 263−70.
[34] Shah R, Heckle M, Rashid A, Flatt D, Slomka T, Bondy BR, et al. Prophylactic defibrillator implantation in patients with non-ischemic cardiomyopathy: an updated comprehensive meta-analysis of randomized controlled trials. J Am Coll Cardiol 2017;69:321.
[35] Shun-Shin MJ, Zheng SL, Cole GD, Howard JP, Whinnett ZI, Francis DP. Implantable cardioverter defibrillators for primary prevention of death in left ventricular dysfunction with and without ischaemic heart disease: a meta-analysis of 8567 patients in the 11 trials. Eur Heart J 2017;38:1738−46.
[36] Singh S, Mirhosseini A, Arora R. Reduced mortality benefit with prophylactic implantable cardioverter defibrillator in nonischemic cardiomyopathy: a systematic review and meta-analysis. J Am Coll Cardiol 2017;69:412.
[37] Stavrakis S, Asad Z, Reynolds D. Implantable cardioverter defibrillators for primary prevention of mortality in patients with nonischemic cardiomyopathy: a meta-analysis of randomized controlled trials. J Cardiovasc Electrophysiol 2017;28(6):659−65.
[38] Turagam M, Velagapudi P, Vuddanda V, Parikh V, Jazayeri M-A, Lavu M, et al. Implantable cardioverter defibrillator for the prevention of mortality in non-ischemic cardiomyopathy: a meta-analysis of randomized controlled trials. J Am Coll Cardiol 2017;69:360.
[39] Wolff G, Lin Y, Karathanos A, Brockmeyer M, Wolters S, Nowak B, et al. Implantable cardioverter/defibrillators for primary prevention in dilated cardiomyopathy post-Danish: an updated meta-analysis and systematic review of randomized controlled trials. Clin Res Cardiol 2017;106(7):501−13.
[40] Xing Z, Tang L, Chen C, Huang J, Zhu Z, Hu X. Effectiveness of implantation of cardioverter-defibrillators therapy in patients with non-ischemic heart failure: an updated systematic review and meta-analysis. Braz J Cardiovasc Surg 2017;32(5):417−22.
[41] Bänsch D, Antz M, Boczor S, Volkmer M, Tebbenjohanns J, Seidl K, et al. Primary prevention of sudden cardiac death in idiopathic dilated cardiomyopathy: the Cardiomyopathy Trial (CAT). Circulation 2002; 105:1453−8.
[42] Bardy GH, Lee KL, Mark DB, Poole JE, Packer DL, Boineau R, et al. Amiodarone or an implantable cardioverter−defibrillator for congestive heart failure. New Engl J Med 2005;352(3):225−37.
[43] Bristow MR, Saxon LA, Boehmer J, Krueger S, Kass DA, De Marco T, et al. Cardiac-resynchronization therapy with or without an implantable defibrillator in advanced chronic heart failure. New Engl J Med 2004;350(21):2140−50.
[44] Kadish A, Dyer A, Daubert JP, Quigg R, Estes NM, Anderson KP, et al. Prophylactic defibrillator implantation in patients with nonischemic dilated cardiomyopathy. New Engl J Med 2004;350(21): 2151−8.

[45] Køber L, Thune JJ, Nielsen JC, Haarbo J, Videbæk L, Korup E, et al. Defibrillator implantation in patients with nonischemic systolic heart failure. New Engl J Med 2016;375(13):1221−30.

[46] Pezawas T, Grimm M, Ristl R, Kivaranovic D, Moser FT, Laufer G, et al. Primary preventive cardioverter-defibrillator implantation (Pro-ICD) in patients awaiting heart transplantation. A prospective, randomized, controlled 12-year follow-up study. Transpl Int 2015; 28(1):34−41.

[47] Strickberger SA, Hummel JD, Bartlett TG, Frumin HI, Schuger CD, Beau SL, et al. Amiodarone versus implantable cardioverter-defibrillator: randomized trial in patients with nonischemicdilated cardiomyopathy and asymptomaticnonsustained ventricular tachycardia—AMIOVIRT. J Am Coll Cardiol 2003;41:1707−12.

[48] Guyot P, Ades A, Ouwens MJ, Welton NJ. Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival curves. BMC Med Res Methodol 2012;12:9.

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

[50] Lucenteforte E, Moja L, Pecoraro V, Conti AA, Conti A, Crudeli E, et al. Discordances originated by multiple meta-analyses on interventions for myocardial infarction: a systematic review. J Clin Epidemiol 2015;68:246−56.

[51] Doak C, Heitmann B, Summerbell C, Lissner L. Prevention of childhood obesity−what type of evidence should we consider relevant? Obes Rev 2009;10(3):350−6.