A systematic assessment of the characteristics of randomised controlled trials cited by acute coronary syndrome clinical practice guidelines

Maribel Gonzalez-Del-Hoyo[1,2], MD, PhD; Caterina Mas-Llado[1,3,4], MD; Joan Siquier-Padilla[2], MD; Laura Blaya-Peña[2] MD; JJ Coughlan[5], MB BCh, MD; Vicente Peral[1,2,4], MD, PhD; Xavier Rossello[1,2,4,6], MD, PhD

[1] *Health Research Institute of the Balearic Islands (IdISBa), 07122 Palma, Spain*

[2] *Cardiology Department, Hospital Universitari Son Espases, 07122 Palma, Spain*

[3] *Cardiology Department, Hospital de Manacor, 07500 Manacor, Spain*

[4] *Faculty of Medicine, Universitat de les Illes Balears (UIB), 07122 Palma, Spain*

[5] *Cardiovascular Research Institute, Mater Private Network, Dublin, D07 WKW8, Ireland*

[6] *Centro Nacional de Investigaciones Cardiovasculares (CNIC), 28029 Madrid, Spain*

Address for correspondence:

Xavier Rossello, Cardiology Department, Hospital Universitari Son Espases, Health Research Institute of the Balearic Islands (IdISBa), 07122, Palma de Mallorca, Spain.

Phone number: +34 871 20 50 50 (ext. 64527). Fax: +34 871 20 68 68.

Email: fjrossello@ssib.es

ABSTRACT

Aims: The aim of this study was to describe the methodological features of the randomised controlled trials (RCTs) cited in American and European clinical practice guidelines (CPGs) for ST elevation myocardial infarction (STEMI) and non-ST elevation acute coronary syndrome (NSTE-ACS).

Methods and results: Out of 2128 non-duplicated references cited in the 2013 and 2014 ACC/AHA and 2017 and 2020 ESC CPGs for STEMI and NSTE-ACS, we extracted data for 407 RCTs (19.1% of total references). The majority were multicenter studies (81.8%), evaluated pharmacological interventions (63.1%), and had a 2-arm (82.6%), superiority (90.4%) design. Most RCTs (60.2%) had an active comparator, and 46.2% were funded by industry. The median observed sample size was 1001 patients (84.2% of RCTs achieved ≥80% of the intended sample size). Most RCTs had a single primary outcome (90.9%), which was a composite in just over half (51.9%). Amongst the RCTs testing for superiority, 44.0% reported a p-value of ≥0.05 for the primary outcome and 61.9% observed a risk reduction of >15%. The observed treatment effect was lower-than-expected in 67.6% of RCTs, with 34.4% having at least a 20% lower-than-expected treatment effect. The calculated post hoc statistical power was ≥80% for 33.9% of cited RCTs.

Conclusions: This analysis demonstrates that RCTs cited by CPGs can still have significant methodological issues and limitations, highlighting that a better understanding of the methodological aspects of RCTs is crucial in order to formulate recommendations relevant to clinical practice.

Keywords: Randomised clinical trials; Evidence-based medicine; Clinical Practice Guidelines, Acute coronary syndrome; Recommendations.

INTRODUCTION

Clinical practice guidelines (CPGs) influence the care provided to millions of patients worldwide(1–5) and their recommendations are central to the evidence-based medicine paradigm(6,7). Randomised controlled trials (RCTs) (and meta-analysis of RCTs) are pivotal to set recommendations in CPGs(8,9). An understanding of RCT design, statistical limitations, and methodological issues is crucial in order to translate RCT-based scientific evidence into sound clinical recommendations(10,11).

The American College of Cardiology/American Heart Association (ACC/AHA) and the European Society of Cardiology (ESC) have separate CPGs for ST-segment elevation myocardial infarction (STEMI), and non-ST-segment elevation acute coronary syndrome (NSTE-ACS)(12–15). When interpreting these recommendations, it is important to consider the strengths and the limitations of the supporting randomised evidence (16,17). This should include an assessment of methodological issues, including the statistical power, the balance between the observed and the expected event rate, the absence of blinding, and the type and number of primary and secondary endpoints(10), as well as the type of funding of the study.

In order to investigate this, we performed a systematic assessment of the methodological characteristics of the RCTs cited in the ACC/AHA and ESC CPGs for STEMI and NSTE-ACS. The main aims of this analysis were: (1) to define the features of RCTs cited in ACS CPGs with regard to their design, sample size, and type of outcome; (2) to assess the consistency between the expected and observed sample sizes, primary event rates and treatment effects, (3) to explore the associations between RCT characteristics and the type of guideline recommendations (i.e., as per the class or level of evidence [LOE]) and (4) to compare the association between the design features of the cited RCTs and the types of recommendations.

METHODS

Selection of Randomised Controlled Trials

Using the ACC and ESC websites, we reviewed the latest versions of the CPGs for ACS as of January 1, 2022. The following CPGs were therefore included in our analysis: 2013 ACC/AHA

Guideline for the management of STEMI(15), 2014 ACC/AHA guideline for the management of patients NSTE-ACS(12), 2017 ESC guideline for the management of STEMI(13), and 2020 ESC guideline for the management NSTE-ACS(14).

Screening and selection process of references

Data collection and analysis was performed according to the PRISMA guidelines(18). **Supplementary Figure 1** summarises the screening and selection process. From each CPG document, two reviewers independently screened and extracted all non-duplicated references using an electronic form. To identify duplicates, we used the PubMed Identifier (PMID). References were classified by another reviewer into 5 categories according to their study design: (i) meta-analyses, (ii) primary publications of RCTs, (iii) secondary publications of RCTs using randomised data(19), (iv) non-randomised studies (either observational or secondary RCTs not using randomised data), and (v) other papers (consensus papers, other CPGs, reviews). The category of each study was revised and confirmed by a second reviewer. Any disagreements were resolved by consensus after discussion between the three reviewers. To avoid overrepresenting studies with multiple secondary publications, only the primary publications of RCTs were included in this analysis(20).

Data Extraction for RCTs

Once all primary publications of RCTs cited in the CPGs were identified, one of the authors used an electronic form to extract data from the full publications, such as the type of hypothesis (e.g., superiority, non-inferiority), number of primary and secondary outcomes, or type of outcome. Details about the statistical analysis plan were also extracted, including: the expected treatment effect, statistical power, and the estimated sample size. We also extracted data on the observed outcome data (event numbers, treatment effects, outcome-values in RCTs testing for superiority). For trials testing a single primary outcome, we manually calculated the expected and the observed relative risks for the primary endpoint, and conducted a post hoc analysis in order to estimate the actual statistical power, based on a two-sided type 1 error of 0.05. We used relative risks when making comparisons across trials in order to homogenise the observed and expected treatment effects.

Data extraction for CPGs recommendations

The recommendations provided by each guideline were extracted by two independent reviewers and were linked to the supporting references. In the CPGs, each recommendation is presented using a statement (recommendation text), and an associated class and LOE(12–15). By linking the supporting evidence to each recommendation, we were able to crossmatch the characteristics of the RCTs at a study-level with the information obtained at recommendation-level (class and LOE).

Statistical Analysis

At recommendation-level, RCTs features were compared by class of recommendation (I *vs*. II *vs*. III) and LOE (LOE A *vs*. LOE B). Continuous variables were presented as medians with IQRs, and categorical variables as frequencies with percentages. Visual inspection of the data showed that most continuous variables were non-normally distributed; therefore, nonparametric methods were used to compare the groups. The Mann-Whitney U test was used for comparisons between 2 groups, whereas the Kruskal-Wallis test was used for comparisons between 3 or more groups. Categorical variables were compared using $\chi 2$ test, or Fishers exact test when appropriate.

To investigate if trial design features were independently associated with class I, LOE A recommendations (relative to Class II, LOE A recommendations), we produced a multivariate model adjusted for 12 trial design characteristics (industry funding, type of intervention, superiority hypothesis design, international study, multicenter cohort, number of arms, active control group, blinding, independent events adjudication committee, number of secondary outcomes, planned sample size, and mean follow-up). Adjusted association between variables and type of recommendations are presented as odds ratio (OR) with 95% confidence intervals (CIs). Two-sided significance testing was used and a p-value <0.05 was considered significant. All analyses were performed using Stata version 16.1 (StataCorp, Texas, USA).

RESULTS

Study selection of the references cited in the CPGs

Of 2525 references cited by the four CPGs, there were 407 non-duplicated primary publications of RCTs (**Supplementary Figure 1**). Most RCTs (74.9%) were cited in one guideline. The RCTs cited in ≥3 guidelines are presented in **Supplementary Table 1** and some basic data for all RCTs are reported in **Supplementary Table 2**.

Basic characteristics and sample size of RCTs

Table 1 summarises the characteristics of the cited RCTs. Events were adjudicated by an independent adjudication outcome committee in 57.5% of RCTs. The majority of RCTs evaluated a pharmacological intervention (63.1%), whilst 27.5% of RCTs assessed a management strategy (**Figure 1**). Further detail regarding the type of pharmacological treatment or management strategy in the active arms of these RCTs is provided in **Supplementary Figure 2**. The majority of RCTs had a superiority design (90.4%), with 2-arms (82.6%) and an active comparator (60.2%) as the control arm (**Figure 1**). In this cohort, 46.2% of the RCTs were funded exclusively by industry, 19% were partially funded by industry sources and 19.4% were fully funded by public sources. The majority of RCTs reported a statistical analysis plan (86.2%) and based their main analysis on the intention-to-treat principle (87.2%).

The median planned sample size was 1327 participants (IQR 445-4060). Based on their design, 60.5% of the RCTs had less than 85% statistical power. The median observed sample size was 1001 (IQR 263-3602) and 24.8% of the trials enrolled ≤250 participants. The majority of the RCTs (84.2%) achieved ≥80% of their planned sample size and 51.5% of RCTs achieved (or exceeded) their planned sample size.

Type of outcomes and follow-up

Most RCTs had a single primary outcome (90.9%), with the remaining 9.1% having two co-primary outcomes (**Table 2**). Among the RCTs with a single primary outcome, 51.9% of these were composite endpoints. For RCTs using two co-primary outcomes, 62.5% had a composite outcome. Composite endpoints had a median of 3 (IQR 3-4) individual components, including mortality

(87.1%), myocardial infarction (80.3%), stroke (37.8%) and unplanned revascularization (31.1%). The primary outcome was a surrogate outcome in 34.6% of the RCTs. **Supplementary Table 3** reports the different surrogate outcomes. Two thirds of the RCTs had more than 3 secondary outcomes (69.8%). The distribution of the follow-up time is shown in **Table 2**.

Observed vs. expected number of outcomes and relative risk

Among the 368 RCTs testing for superiority, 336 (91.3%) had a single primary endpoint **(Figure 2)**, whereas the remaining 8.7% had co-primary endpoints. Further characteristics of the RCTs testing for superiority on single primary endpoints (either binary or composite outcomes) are summarised in **Table 3**. The median number of observed outcomes was 210 (IQR 66-569), while the expected number of outcomes was 288 (119-669). Based on the number of outcomes in each group, the observed treatment effect was manually estimated using a relative risk for each RCTs assessing a single primary endpoint. The median (IQR) relative risk was 0.81 (0.61-0.91). A risk reduction of >15% was observed in 61.9% of the RCTs, regardless of their p-value (**Figure 3**). Furthermore, among all RCTs testing a single composite or binary primary endpoint for superiority, 44.0% had a p-value ≥0.05.

In addition to manually estimating the relative risk, we also evaluated whether there was a mismatch between the expected and the observed estimate for the treatment effect (whether it was an odds ratio, hazard ratio, incidence rate ratio, or any other estimate). Among the RCTs evaluating a single primary endpoint and reporting data to estimate the expected relative risk (n=185), the median (IQR) expected relative risk was 0.72 (0.60-0.80). However, the observed treatment effect was lower than expected in 67.6% of these RCTs. In 34.4% of these RCTs, the treatment effect was at least 20% lower than the expected treatment effect.

When considering all 336 (90.4%) RCTs testing for superiority and having a single primary endpoint, 43.8% reported a p-value of ≥0.05 for the primary endpoint. Based on *post hoc* power analysis, 33.9% of these RCTs had a calculated statistical power of ≥80%.

RCTs characteristics according to type of funding and year of publication

Further analysis describing trial characteristics according to the type of funding and the year of publication are provided in **Supplementary Table 4** and **Supplementary Table 5**.

RCTs characteristics according to class of recommendation and level of evidence

There were 600 recommendations across the four CPGs for STEMI and NSTE-ACS. Four hundred thirty-seven recommendations had supporting evidence (i.e., at least one study was cited). Amongst these recommendations, there were 257 (58.8%) class I, 135 (30.9%) class II and 45 (10.3%) class III recommendations. As per LOE, there were 113 (25.9%) LOE A, 266 (60.9%) LOE B and 58 (13.3%) LOE C recommendations. There were 274 recommendations that cited ≥1 RCT: 156 (56.9%), 89 (32.5%) and 29 (10.6%) were class I, II, and III, respectively, whereas 84 (30.7%), 177 (64.6%), and 13 (4.7%) were LOE A, B, and C, respectively. For the class I recommendations, 156 out of 349 (44.7%) recommendations cited at least one RCT. Similarly, 89 out of the 200 class II recommendations (44.5%) and 29 out of the 51 class III recommendations (56.9%) cited at least one RCT (**Figure 4**).

The characteristics of the RCTs according to the class of recommendations and LOE are summarised in **Table 4.** Compared with RCTs cited in class II recommendations, RCTs cited in class I and III had a larger sample size (median of 2184 vs. 1502 vs. 2166 for class I, II, III, respectively, $p<0.001$). Compared with RCTs cited in class III recommendations, RCTs cited in class I and II had larger treatment effects (estimated relative risk of 0.80 vs. 0.79 vs. 0.87, for class I, II, and III, respectively; $p<0.001$). There was a significantly lower number of RCTs with a p-value ≥0.05 among class I recommendations (27.0%), compared to class II (34.2%) and class III (58.8%) ($p<0.001$).

LOE comparisons were only performed comparing LOE A vs. LOE B recommendations. This is because LOE C recommendations are generally based on the consensus opinion of experts, or on small observational studies (i.e., data that is of limited value to compare treatments). RCTs cited in LOE A recommendations more frequently had a multicentre design (92.8% vs. 85.5%, $p=0.014$), and were more commonly blinded (56.5% vs 43.9%, $p=0.007$). In comparison to trials cited for LOE B recommendations, trials cited for LOE A also had a larger sample size (median sample of 2527

patients vs. 1810 patients, p=0.005). Whereas median p-values were significantly lower for RCTs cited for LOE A in comparison to RCTs cited for LOE B recommendations (median p value 0.004 vs. 0.009, p=0.035), the median observed treatment effect was not significantly different between RCTs cited in support of LOE A and LOE B recommendations (median relative risk 0.80 vs 0.82, p=0.998).

Association between class I, LOE A recommendations and trial features

We evaluated 12 potential determinants of Class I, LOE A recommendations by conducting a multivariate adjusted model (**Table 5**). Relative to Class II, LOE A recommendations, there were two trial features independently associated with a class I, LOE A recommendations (industry funding, and number of secondary outcomes). There were no significant associations with class I, LOE A recommendation among the remaining trial features.

DISCUSSION

RCTs are a key component in both the production and interpretation of CPGs(8,9). The main findings of our systematic assessment of the methodological characteristics of the RCTs supporting ACC/AHA and ESC CPGs for STEMI and NSTE-ACS are as follows: (1) the observed treatment effect was lower than expected in 67.6% of RCTs, with 34.4% of RCTs having a treatment effect that was at least 20% lower than expected; (2) on *post hoc* analysis, the statistical power was calculated to be ≥80% for 33.9% of cited RCTs; (3) around half of all recommendations were supported by at least one RCT; (4) there was no difference in the sample size or estimated treatment effect between trials supporting recommendations with LOE A *vs.* LOE B; and (5) class I recommendations were supported more often by trials with a larger treatment effect and a lower p-value in comparison to class III recommendations.

Understanding the design, statistical challenges, and methodological issues relevant to RCTs is crucial in order for both CPG task forces to translate RCT evidence into applicable recommendations, and for clinicians to interpret these recommendations and implement them in clinical practice(10,11). There are a paucity of data on the type of RCTs supporting ACC/AHA and

ESC CPGs for ACS(20). To our knowledge, this is the first systematic approach assessing the methodological features of the RCTs cited in these ACS CPGs.

Our results demonstrate that the ACS RCT landscape is dominated by trials testing a pharmacological intervention for superiority, comparing 2-arms and recruiting patients from multiple centers. Around half of cited RCTs were supported by industry and ~20% were exclusively funded by public sources, which is in line with a previous report addressing this issue in invasive cardiovascular RCTs(21) . In a separate study, Gaudino *et al.* reported a statistically significant difference in the primary outcome (p-value <0.05) among industry funded in comparison to those publicly funded trials (64.3% vs 48.5%). In our analysis, funding was not associated with a statistically significant difference in the primary outcome (industry funded vs public funded; 55.8% vs 58.2%). The difference between studies may be explained by the type of included trials, given that the study by Gaudino *et al.* focused on invasive cardiovascular intervention trials, whereas our analysis including trials supporting all recommendations from the ACS guidelines. The median sample size of the RCTs in our analysis was 1001 patients, though around 25% of RCTs enrolled less than 250 patients. The median sample size was around 30% below the planned sample size, but 84.2% of RCTs achieved at least 80% of the planned sample size. In line with our findings, a previous analysis reported a similar median sample size of 1192 participants when evaluating RCTs among the ACC/AHA and ESC guidelines for stable ischemic heart disease(22). Among the scarce related literature, other reports evaluating 21 RCTs supporting guidelines on venous thromboembolism prophylaxis(23), or assessing 399 RCTs published in high impact medical journals from 2004-2010(24), reported a lower median sample size of 400 patients and 682 patients, respectively.

Composite endpoints are commonly used in cardiovascular RCTs (25). In our cohort of RCTs, the primary endpoint was a composite in 51.9%, whereas it was a surrogate endpoint (i.e., not a clinical event) in 34.9%. Whilst the benefits of using composite endpoints in terms of efficiency are clear, there is a danger of oversimplifying the evidence if the individual components of the composite are of different magnitude or direction. For example, unplanned revascularization was included as part of a composite endpoint in 31.1% of the RCTs using a composite endpoint in our cohort. This tends to be the most frequently occurring (and often least clinically relevant) component of composite

endpoints in ACS trials and tends to be the driver of event rates (26). In our cohort, 69.8% of the RCTs had more than 3 secondary outcomes. The risk of type I error increases when assessing multiple endpoints(11). The interpretation of findings from multiple secondary outcomes is a statistical challenge that should be taken into account when interpreting the totality of the evidence supporting some recommendations. In a survey of all cardiovascular RCTs published in three top journals during 2019, 75% of the RCTs performed no formal statistical correction for multiple outcomes(11).

In our cohort, 56.2% of the RCTs testing for superiority met their primary endpoint (i.e., reported a significant p value). However, the interpretation of RCT results should always go beyond p-values(16,17), and focus on a more nuanced, sophisticated and balanced interpretation of trial evidence. In addition to statistical significance, a treatment difference should be large enough to matter in clinical practice (i.e., clinically meaningful). Of note, we reported that the observed treatment effect was lower than expected in 67.0% of RCTs, with 34.4% of RCTs reporting a treatment effect at least a 20% lower than expected. It should be highlighted that these analyses had a median relative risk of 0.81 regardless of their statistical significance, which translated into a substantial reduction in the incidence of the primary outcomes. However, our *post hoc* power analysis evaluating the statistical power based on the observed data revealed that the statistical power was higher than ≥80% for only 33.9% of cited RCTs, indicating that the risk of type II error was high in many RCTs. It should be acknowledged that *post hoc* power analysis has some important limitations. Estimations performed when outcomes have already been observed is theoretically and conceptually flawed and may also be analytically misleading (27,28). A p-value higher than 0.05 was reported in 43.8 % of the RCTs testing for superiority. Importantly, one needs to take into account that CPGs cite all types of RCTs in order to capture the totality of evidence and avoid bias. For instance, some of the RCTs might be underpowered for their primary endpoint, but might have provided other relevant information, and in some areas (e.g., cardiogenic shock), only small trials might be available.

The type of RCTs cited differed by class of recommendations. RCTs with a larger treatment effect and statistical significance were more frequently cited in class I ("should be performed") than class III recommendations ("is not indicated"). These observations can be partly explained by the current grading system. In order to not recommend an intervention, less evidence seems to be needed,

probably due to ethical and futility considerations. In addition, to receive a Class III recommendation, an intervention can either not demonstrate a benefit or may in fact be harmful, but for a Class I recommendation, an intervention must demonstrate a benefit. In addition, it is interesting to note that while there were no substantial differences in terms of sample size and treatment effect for RCTs supporting recommendations with LOE A or B, there were significant differences in trial design. For LOE A recommendations, cited RCTs were more commonly multicenter, placebo controlled and had some type of blinding, in comparison with RCTs cited for LOE B recommendations.

While there is no doubt regarding the important contribution of industry funding to guideline changing RCTs, previous studies have been stressed the differences in study design and potential reporting bias between industry funded and non-industry funded trials(29,30). Our data expands the knowledge in the field by not only focusing on industry trials, but by comparing trial determinants of class IA recommendations (as compared to class II LOE A). In our current cohort of RCTs, we were able to identify few trial factors associated with class IA recommendations. Of note, industry-funded studies were independently associated with class I, LOE A as compared to class II, LOE A. The association of industry-funded trials with a class I, LOE A recommendation ("is recommended") compared with a class II, LOE A recommendation ("should be considered") may be explained by class I recommendations more frequently dealing with pharmacological interventions (31), as trials of this type were more frequently industry funded (**Supplementary Table 4**, industry funded vs public funded, 65.7% vs 43% p=0.001).

LIMITATIONS

Our analysis should be interpreted in the light of several limitations. We focused only on ACC/AHA and ESC CPGs for ACS (because of their global impact), and did not evaluate CPGs from other scientific societies. Some RCTs might have been counted ≥2 times when performing comparisons between recommendations (e.g., one RCT might be part of a recommendation with a LOE A, and in a separate recommendation of LOE B)(7). Finally, we manually estimated relative risks to have a common estimate of the primary outcome for all RCTs. Nevertheless, it should be

acknowledged that the effect size cannot be interpreted in the same way for relative risks than other estimates (if an odds ratio is less than one, then it is always smaller than the relative risk). In the same vein, time-to-event was not considered when evaluating relative risks.

CONCLUSIONS

The majority of RCTs cited by ACC/AHA and ESC CPGs for STEMI and NSTE-ACS tested a pharmacological intervention for superiority in 2-arm trials recruiting patients from multiple centres. Most RCTs had an active comparator and around half were supported by industry. While the majority of RCTs achieved at least 80% of their expected sample size, ~one-third were calculated to have post hoc statistical power of >80%. The observed treatment effect was lower than expected in two-thirds of RCTs. This analysis demonstrates that RCTs cited by CPGs can still have significant methodological issues and limitations, highlighting that a better understanding of the methodological aspects of RCTs is crucial in order to formulate recommendations relevant to clinical practice.

ACKNOWLEDGMENTS

None.

FUNDING

This project has received funding to cover the electronic case report form from the Balearic Society of Cardiology (Sociedad Balear Cardiología) through the “Bernat Nadal Ginard” grant.

CONFLICT OF INTEREST

None declared.

DATA AVAILABILITY STATEMENT

The data underlying this article will be shared on reasonable request to the corresponding author.

REFERENCES

1. Rossello X, Medina J, Pocock S, Van de Werf F, Chin CT, Danchin N, et al. Assessment of quality indicators for acute myocardial infarction management in 28 countries and use of composite quality indicators for benchmarking. European Heart J Acute Cardiovascular Care. 2020;9(8):911–22.
2. Bebb O, Hall M, Fox KAA, Dondo TB, Timmis A, Bueno H, et al. Performance of hospitals according to the ESC ACCA quality indicators and 30-day mortality for acute myocardial infarction: national cohort study using the United Kingdom Myocardial Ischaemia National Audit Project (MINAP) register. European Heart Journal. 2017;38(13):974–82.
3. Mehta RH, Chen AY, Alexander KP, Ohman EM, Roe MT, Peterson ED. Doing the Right Things and Doing Them the Right Way: Association Between Hospital Guideline Adherence, Dosing Safety, and Outcomes Among Patients With Acute Coronary Syndrome. Circulation. 2015;131(11):980–7.
4. Vaucher J, Marques-Vidal P, Waeber G, Vollenweider P. Population impact of the 2017 ACC/AHA guidelines compared with the 2013 ESH/ESC guidelines for hypertension management. Eur J Prev Cardiolog. 2018;25(10):1111–3.
5. Sulman D, Zeitouni M, Silvain J, Kerneis M, Guedeney P, Barthélémy O, et al. ESC/EAS guidelines for the detection, prevention, and treatment of individuals at risk of a first myocardial infarction: effect of 5 years of updates and the new SCORE2. European Heart Journal - Cardiovascular Pharmacotherapy. 2022;8(6):633–43.
6. Sheridan DJ, Julian DG. Achievements and Limitations of Evidence-Based Medicine. J Am Coll Cardiol. 2016;68(2):204–13.
7. Mas-Llado C, González-Del-Hoyo M, Siquier-Padilla J, Blaya-Peña L, Coughlan JJ, de la Villa BG, et al. Representativeness in randomised clinical trials supporting acute coronary syndrome guidelines. Eur Heart J Qual Care Clin Outcomes. 2023 Jan 25;qcad007.
8. Ioannidis JPA. Why Most Published Research Findings Are False. PLoS Med. 2005;2(8):e124.
9. Rosenberg W, Donald A. Evidence based medicine: an approach to clinical problem-solving. BMJ. 1995;310(6987):1122–6.

10. Pocock SJ, McMurray JJV, Collier TJ. Making Sense of Statistics in Clinical Trial Reports: Part 1 of a 4-Part Series on Statistics for Clinical Trials. J Am Coll Cardiol. 2015;66(22):2536–49.
11. Pocock SJ, Rossello X, Owen R, Collier TJ, Stone GW, Rockhold FW. Primary and Secondary Outcome Reporting in Randomized Trials: JACC State-of-the-Art Review. J Am Coll Cardiol. 2021;78(8):827–39.
12. Amsterdam EA, Wenger NK, Brindis RG, Casey DE, Ganiats TG, Holmes DR, et al. 2014 AHA/ACC guideline for the management of patients with Non-ST-Elevation acute coronary syndromes: Executive summary: A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. J Am Coll Cardiol. 2014;64(24):2645–87.
13. Ibanez B, James S, Agewall S, Antunes MJ, Bucciarelli-Ducci C, Bueno H, et al. 2017 ESC Guidelines for the management of acute myocardial infarction in patients presenting with ST-segment elevation. European Heart Journal. 2018;39(2):119–77.
14. Collet JP, Thiele H, Barbato E, Barthélémy O, Bauersachs J, Bhatt DL, et al. 2020 ESC Guidelines for the management of acute coronary syndromes in patients presenting without persistent ST-segment elevation. Eur Heart J. 2021;42(14):1289–367.
15. O'Gara PT, Kushner FG, Ascheim DD, Casey DE, Chung MK, de Lemos JA, et al. 2013 ACCF/AHA guideline for the management of ST-elevation myocardial infarction: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. Circulation. 2013;127(4):e362-425.
16. Amrhein V, Greenland S, McShane B. Scientists rise up against statistical significance. Nature. 2019;567(7748):305–7.
17. Wasserstein RL, Lazar NA. The ASA Statement on *p* -Values: Context, Process, and Purpose. The American Statistician. 2016;70(2):129–33.
18. Moher D, Liberati A, Tetzlaff J, Altman DG, for the PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ. 2009;339(jul21 1):b2535–b2535.
19. Lai C, Sbidian E, Giraudeau B, Le Cleach L. Twenty percent of secondary publications of randomized controlled trials of drugs did not provide new results relative to the primary publication.

Journal of Clinical Epidemiology. 2020;117:20–8.
20. Gonzalez-Del-Hoyo M, Mas-Llado C, Blaya-Peña L, Siquier-Padilla J, Peral V, Rossello X. The Fragility Index in randomised clinical trials supporting clinical practice guidelines for acute coronary syndrome: measuring robustness from a different perspective. Eur Heart J Acute Cardiovasc Care. 2023 (ahead of print).
21. Gaudino M, Hameed I, Rahouma M, Khan FM, Tam DY, Biondi-Zoccai G, et al. Characteristics of Contemporary Randomized Clinical Trials and Their Association With the Trial Funding Source in Invasive Cardiovascular Interventions. JAMA Intern Med. 2020 ;180(7):993.
22. Gaudino M, Hameed I, Biondi-Zoccai G, Tam DY, Gerry S, Rahouma M, et al. Systematic Evaluation of the Robustness of the Evidence Supporting Current Guidelines on Myocardial Revascularization Using the Fragility Index. Circ: Cardiovascular Quality and Outcomes. 2019;12(12):e006017.
23. Edwards E, Wayant C, Besas J, Chronister J, Vassar M. How Fragile Are Clinical Trial Outcomes That Support the CHEST Clinical Practice Guidelines for VTE? Chest. 2018 ;154(3):512–20.
24. Walsh M, Srinathan SK, McAuley DF, Mrkobrada M, Levine O, Ribic C, et al. The statistical significance of randomized controlled trial results is frequently fragile: a case for a Fragility Index. Journal of Clinical Epidemiology. 2014;67(6):622–8.
25. Pocock SJ, McMurray JJV, Collier TJ. Statistical Controversies in Reporting of Clinical Trials: Part 2 of a 4-Part Series on Statistics for Clinical Trials. J Am Coll Cardiol. 2015;66(23):2648–62.
26. Pocock SJ, Clayton TC, Stone GW. Design of Major Randomized Trials: Part 3 of a 4-Part Series on Statistics for Clinical Trials. J Am Coll Cardiol. 2015;66(24):2757–66.
27. Heckman MG, Davis JM, Crowson CS. Post Hoc Power Calculations: An Inappropriate Method for Interpreting the Findings of a Research Study. J Rheumatol. 2022;49(8):867–70.
28. Zhang Y, Hedo R, Rivera A, Rull R, Richardson S, Tu XM. Post hoc power analysis: is it an informative and meaningful analysis? Gen Psych. 2019;32(4):e100069.
29. Als-Nielsen B, Chen W, Gluud C, Kjaergard LL. Association of Funding and Conclusions in

Randomized Drug TrialsA Reflection of Treatment Effect or Adverse Events? JAMA. 2003;290(7):921–8.

30. Djulbegovic B, Lacevic M, Cantor A, Fields KK, Bennett CL, Adams JR, et al. The uncertainty principle and industry-sponsored research. Lancet. 2000;356(9230):635–8.

31. Fanaroff AC, Fudim M, Califf RM, Windecker S, Smith SC, Lopes RD. Levels of evidence supporting drug, device, and other recommendations in the American Heart Association/American College of Cardiology guidelines. American Heart Journal. 2020;226:4–12.

TABLES

Table 1. Publication characteristics, study design and statistical approach of the set of 407 RCTs

Publication characteristics	N (%)
Year of publication	
≤2000	104 (25.6)
2000-2010	157 (38.6)
>2010	146 (35.9)
Source of funding	
Industry	188 (46.2)
Public	79 (19.4)
Both	77 (19.0)
Not reported	63 (15.5)
Design and statistics of RCTs	**N (%)**
Event driven trial	43 (10.6)
Type of primary statistical analysis	
Intention-to-treat	355 (87.2)
Per-protocol	1 (0.3)
Both	8 (2.0)
Not reported	42 (10.6)
Mention to planned sample size	328 (80.6)
Planned sample size	
≤1000	146 (44.5)
1000-3000	85 (25.9)
3000-5000	24 (7.3)
5000-10.000	33 (10.1)
≥10.000	40 (12.2)
Estimates for the primary outcome assessment in superiority RCTs	
Hazard ratio	56 (15.2)
Relative risk	147 (40.0)
Odds ratio	48 (13.0)
Incidence rate ratio	1 (0.3)
Not applicable (i.e., non-binary outcomes)	43 (14.4)
Not reported	69 (18.8)
Statistical power reported	332 (81.6)
Expected statistical power	
≤85%	201 (60.5)
>85%	131 (39.5)
Early stopping	
No	365 (89.9)
Safety	12 (3.0)
Efficacy	14 (3.5)
Futility	4 (1.0)
Low enrolment	11 (2.7)
Number of expected events	
≤100	43 (18.9)
100-200	40 (17.5)
200-400	51 (22.4)
400-600	31 (13.6)
≥600	63 (27.6)

Table 2. Outcomes and follow-up of the RCTs cited in ACC/AHA and ESC STEMI and NSTE-ACS CPGs.

Characteristics	N (%)
Type of primary outcome	
Single primary outcome	370 (90.9)
Two co-primary outcomes	37 (9.1)
Purpose of the single primary outcome	
Efficacy	327 (88.4)
Safety	19 (5.1)
Both	24 (6.5)
Surrogate variable in the primary outcome	141 (34.6)
Type of outcomes in composite primary outcomes	
All-cause death	113 (58.6)
CV death	58 (30.1)
MI	155 (80.3)
Unstable angina or refractory ischaemia	32 (16.6)
HF	24 (12.4)
Unplanned revascularization	60 (31.1)
Stroke	73 (37.8)
Bleeding	22 (11.4)
Secondary safety endpoint	188 (46.3)
Number of secondary outcomes, median (IQR)	5 (3-7)
Number of secondary outcomes	
≤3	112 (30.2)
4-8	202 (54.5)
≥9	57 (15.4)
Mean follow up time	
<1month	114 (28.1)
1-month	97 (23.8)
>1 month to 1 year	63 (15.5)
1-3 years	94 (23.1)
>3 years	39 (9.6)
Lost to follow up reported	346 (85.0)
Lost to follow up	
<1%	232 (67.1)
1-5%	65 (18.8)
≥5%	49 (14.6)

Table 3. Sample size, outcomes, relative risk and statistical power for RCTs testing for superiority in single primary endpoints (either binary or composite)

Characteristics	No. (%)
Total RCTs	**257**
Observed sample size, median (IQR)	1232 (421-4444)
Observed sample size	
≤1000	109 (42.4)
1000-4000	80 (31.1)
>4000	68 (26.5)
Consistency between planned and observed sample size	
<80% (lower than expected)	38 (17.7)
80-100%	58 (27.0)
>100% (higher than expected)	119 (55.4)
Number of observed outcomes, median (IQR)	210 (66-569)
Number of observed outcomes (categories)	
≤200	121 (49.0)
200-1000	90 (36.4)
>1000	36 (14.6)
Lower than expected observed relative risk (n=206 reporting data)	136 (66.0)
Percentage of mismatch for RCTs with lower-than-expected relative risk (n=136), n (%)	
≤10%	39 (26.7)
10-20%	54 (39.7)
>20%	43 (31.6)
Lower than expected observed relative risk (n=185 reporting data)	125 (67.6)
Percentage of mismatch for RCTs with lower-than-expected relative risk (n=125), n (%)	
≤10%	38 (30.4)
10-20%	44 (35.2)
>20%	43 (34.4)
Percentage of trials with a number of observed outcome events lower than expected (n=187 reporting data)	104 (53.5)
Percentage of mismatch for RCTs with lower-than-expected number of outcomes (n=104), n (%)	
≤25%	35 (33.7)
25-50%	26 (25.0)
>50%	43 (41.4)
Observed difference in the percentage of primary endpoints between the control and the intervention arm	
≤2 %	89 (36.0)
2-6%	79 (32.0)
>6 %	79 (32.0)
p-value for single primary outcomes	0.039 (0.003-0.339)
≤0.001	60 (23.4)
0.01-0.001	42 (16.3)
0.05-0.01	42 (16.3)
≥0.05	113 (44.0)

Table 4. Summary of the RCTs supporting the ACC/AHA and ESC STEMI and NSTE-ACS CPGs by class of recommendation and level of evidence

	Class of recommendation				Level of evidence*		
	I (n=291)	II (n=156)	III (n=40)	p-value	A (n= 207)	B (n=262)	p-value
Type of trial				0.082			0.151
Pharmacological	205(70.5)	94(60.3)	27(67.5)		147(71.0)	164(62.6)	
Management Strategy	74(25.4)	55(35.4)	9(22.5)		52(25.1)	83(31.7)	
Device & Others	12(4.1)	7(4.5)	4(10.0)		8(3.9)	15(5.7)	
Year of publication				<0.001			0.006
<2000	78(26.8)	20(12.8)	3(7.5)		45(21.7)	49(18.7)	
2000-2010	171(58.8)	80(51.3)	24(60.0)		101(48.8)	164(62.6)	
>2010	42(14.4)	56(35.9)	13(32.5)		61(29.5)	49(18.7)	
Funding				0.358			0.063
Partial or complete Industry	235 (87.4)	122 (82.4)	29(82.9)		178 (88.6)	194 (82.2)	
Non-Industry	34 (12.6)	26 (17.6)	6 (17.1)		23 (11.4)	42 (17.8)	
Any type of blinding	152(52.2)	71(45.5)	21(52.5)	0.380	117(56.5)	115(43.9)	0.007
Use of placebo in the comparison arm	88(30.2)	35(22.4)	16(40.0)	0.054	70(33.8)	62(23.7)	0.015
Surrogate endpoint	52(17.9)	34(21.8)	10(25.0)	0.415	33(15.9)	54(20.6)	0.197
Safety endpoint	145(49.8)	91(58.3)	26(65.0)	0.076	119(57.5)	135(51.5)	0.198
Independent events adjudication committee	191(61.6)	106(68.0)	30(75.0)	0.481	153(73.9)	164(62.6)	0.009
Mean follow up >1 year	116(39.9)	56(35.9)	10(25.0)	0.171	89(43.0)	9(34.4)	0.056
Observed sample size				<0.001			0.001
≤1500	106(36.6)	78(50.0)	16(40.0)		66(32.0)	125(47.7)	
1500-5000	90(31.0)	29(18.6)	7(17.5)		69(33.5)	51(19.5)	
5000-15.000	44(15.2)	31(19.9)	14(35.0)		41(19.9)	46(17.6)	
>15.000	50(17.2)	18(11.5)	3(7.5)		30(14.6)	40(15.3)	
Consistency between planned and observed sample size				0.359			0.343
<80% (lower than expected)	28(12.2)	12(8.6)	1(2.9)		21(12.0)	17(7.8)	
80-100%	70(30.4)	50(35.7)	11(31.4)		54(30.9)	75(34.6)	
>100% (higher than expected)	132(57.4)	78(55.7)	23(65.7)		100(57.1)	125(57.6)	
No. Observed events for single primary outcome	392(155-1301)	301(78-762)	347(211-1262)	0.038	405(164-982)	308(106-1262)	0.277
p-value for single primary outcomes in superiority RCTs				<0.001			0.261
≤0.001	97(42.9)	42(34.2)	4(11.8)		63(41.2)	70(32.9)	
0.01-0.001	35(15.5)	30(24.4)	9(26.5)		30(19.6)	41(19.3)	
0.05-0.01	33(14.6)	9(7.3)	1(2.9)		18(11.8)	24(11.3)	
≥0.05	61(27.0)	42(34.2)	20(58.8)		42(27.5)	78(36.6)	
Observed relative risk for single primary outcomes and binary /composite primary endpoint				0.001			0.980
≤0.65	46(23.2)	34(31.5)	2(6.7)		59(35.8)	76(36.7)	
0.65-0.75	33(16.7)	12(11.1)	1(3.3)		53(32.1)	66(31.9)	
0.75-0.85	72(36.4)	29(26.9)	9(30.0)				
>0.85	47(23.7)	33(30.6)	18(60.0)		53(32.1)	65(31.4)	
Lower-than-expected relative risk	99(64.7)	56(63.2)	23(88.5)	0.038	64(58.2)	111(73.3)	0.010

**Level of evidence (LOE) comparisons were performed between LOE A vs. LOE B given that LOE C is generally based on consensus opinion of experts, or small observational studies.*

Randomised clinical trials; ACC/AHA, American College of Cardiology/American Heart Association; ESC, European Society of Cardiology. CPGs, clinical practice guidelines.

Table 5. Multivariate logistic regression model for class I LOE A recommendation vs. class II LOE A.

	Class I LOE A vs class II LOE A Multivariate aOR 95% CI	p-value
Complete or partial industry funding	7.61 (1.70-34.00)	0.008
Type of intervention		
Pharmacological	1.45 (0.16-12.97)	0.741
Management Strategy	1.89 (0.21-17.22)	0.572
Device & Others	Ref.	
Superiority hypothesis design	1.10 (0.32-3.78)	0.880
International cohort	0.99 (0.32-3.07)	0.985
Multicenter cohort	1.22 (0.13-11.34)	0.864
>2 arm design	0.68 (0.22-2.11)	0.505
Active arm in the comparison arm	1.60 (0.57-4.53)	0.376
Any type of blinding	0.64 (0.19-2.18)	0.479
Independent events adjudication committee	0.85 (0.27-2.69)	0.778
No. of secondary outcomes		
≤3	4.32 (1.04-17.91)	0.044
4-8	5.34 (1.54-18.57)	0.008
>9	Ref.	
Planned sample size per 1000 patients	0.99 (0.95-1.03)	0.616
Mean follow up >1 year	1.24 (0.49-3.14)	0.647

aOR: adjusted odds ratio; Level of evidence (LOE)

FIGURE LEGENDS

Graphical abstract. Characteristics of randomised controlled trials cited by acute coronary syndrome clinical practice guidelines

Figure 1. Basic characteristics of the 407 RCTs cited in the ACC/AHA and ESC STEMI and NSTE-ACS CPGs

RCTs, randomised clinical trials; ACC/AHA, American College of Cardiology/American Heart Association; ESC, European Society of Cardiology; STEMI, ST elevation myocardial infarction; NSTE-ACS, non-ST elevation acute coronary syndrome; CPGs, clinical practice guidelines.

Figure 2. Outcomes and samples size of the 336 RCTs testing a single primary endpoint for superiority

The study population of panels A, C and E comprises all superiority RCTs with a single primary outcome, whereas for panel B comprises only those with a composite primary outcome and for panel D those RCTs with either a binary or a composite outcome.
RCTs, randomised clinical trials.

Figure 3. Observed relative risk and p-value of the 257 RCTs testing for superiority a single primary endpoint (either binary or composite)

RCTs, randomised clinical trials.

Figure 4. Association between class of recommendation and level of evidence and the presence of supporting RCTs

RCTs, randomised clinical trials; ACC/AHA, American College of Cardiology/American Heart Association; ESC, European Society of Cardiology

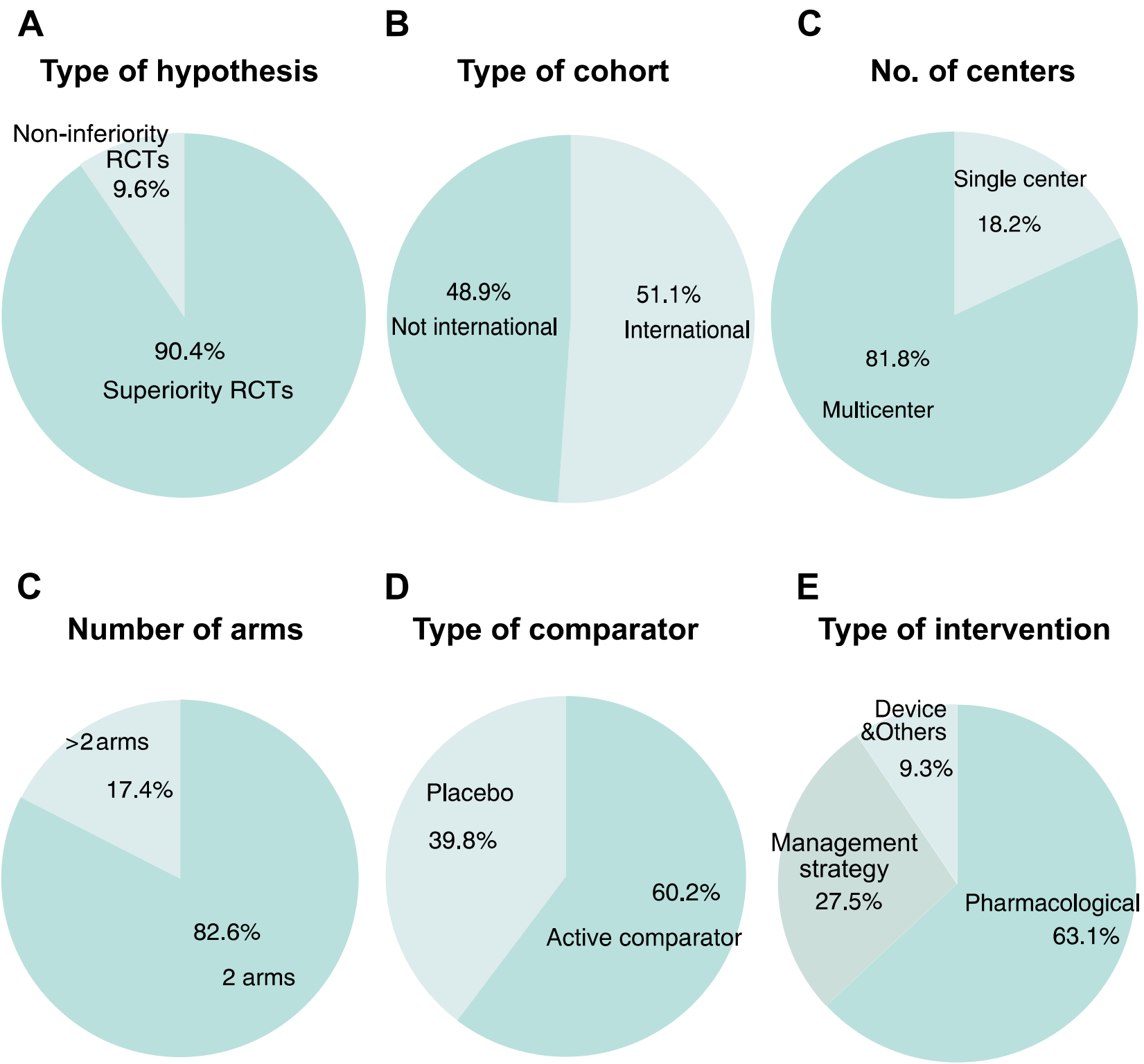
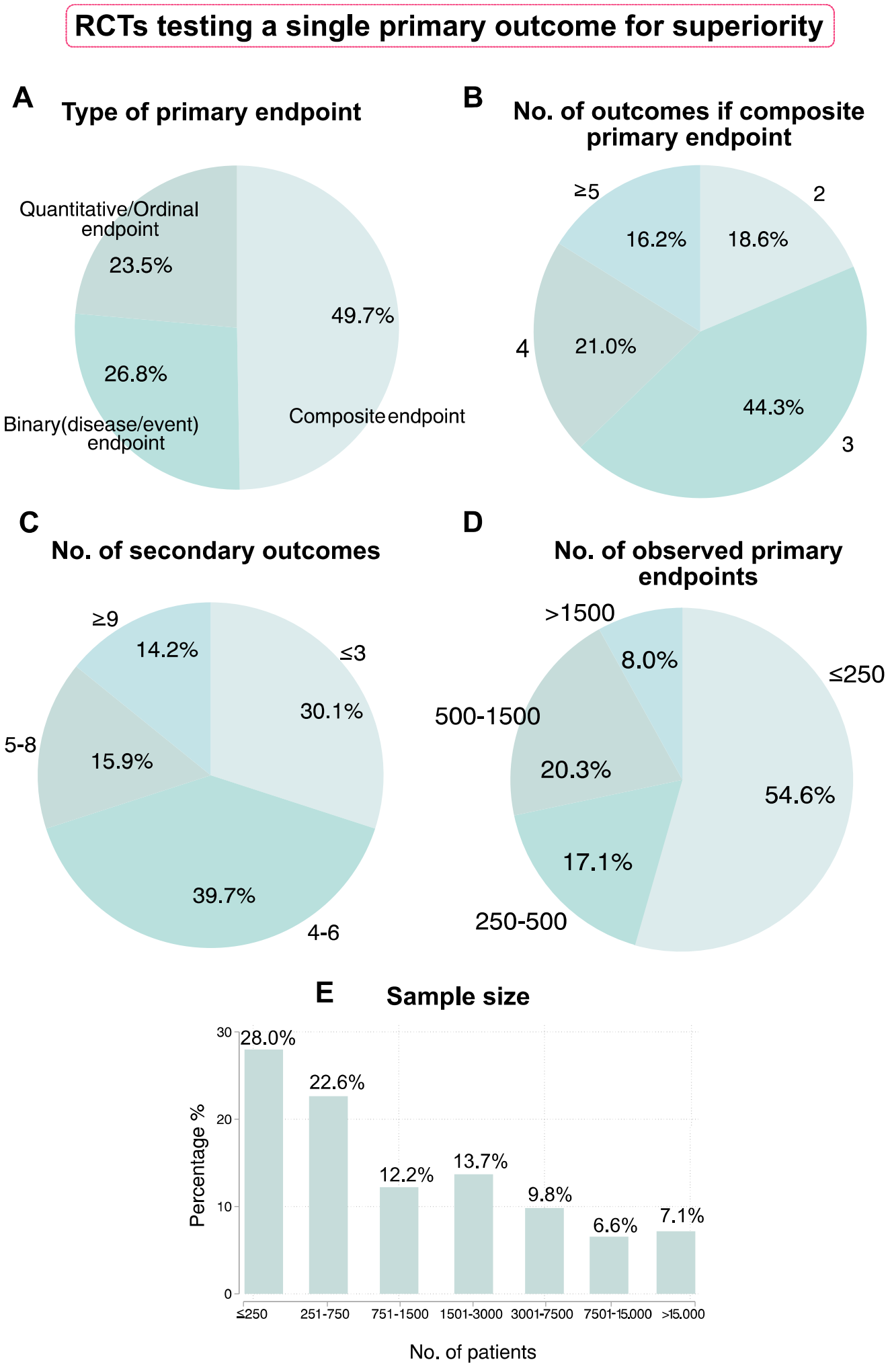
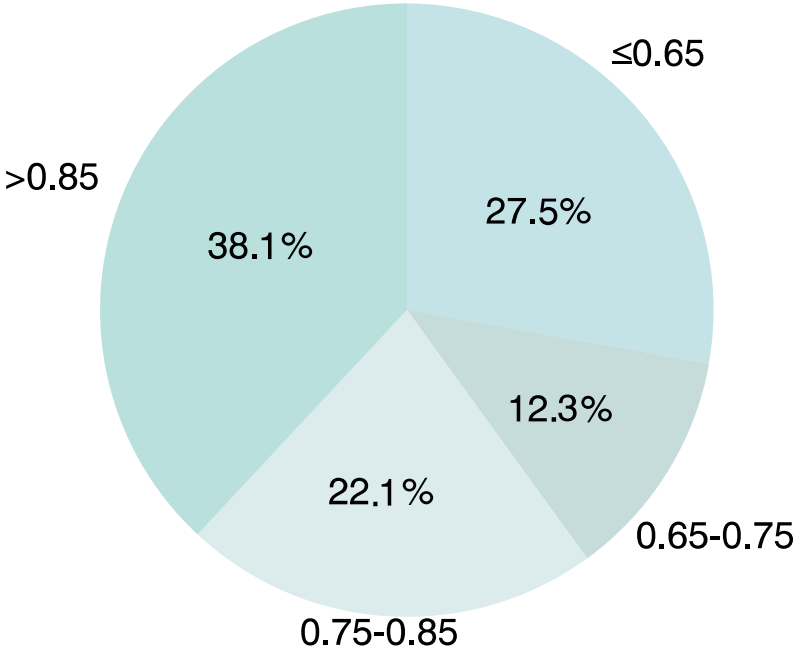

A
Type of hypothesis
Non-inferiority RCTs 9.6%
90.4% Superiority RCTs
B
Type of cohort
48.9% Not international
51.1% International
C
No. of centers
Single center 18.2%
81.8% Multicenter
C
Number of arms
>2 arms 17.4%
82.6% 2 arms
D
Type of comparator
Placebo 39.8%
60.2% Active comparator
E
Type of intervention
Device &Others 9.3%
Management strategy 27.5%
Pharmacological 63.1%

Figure 2

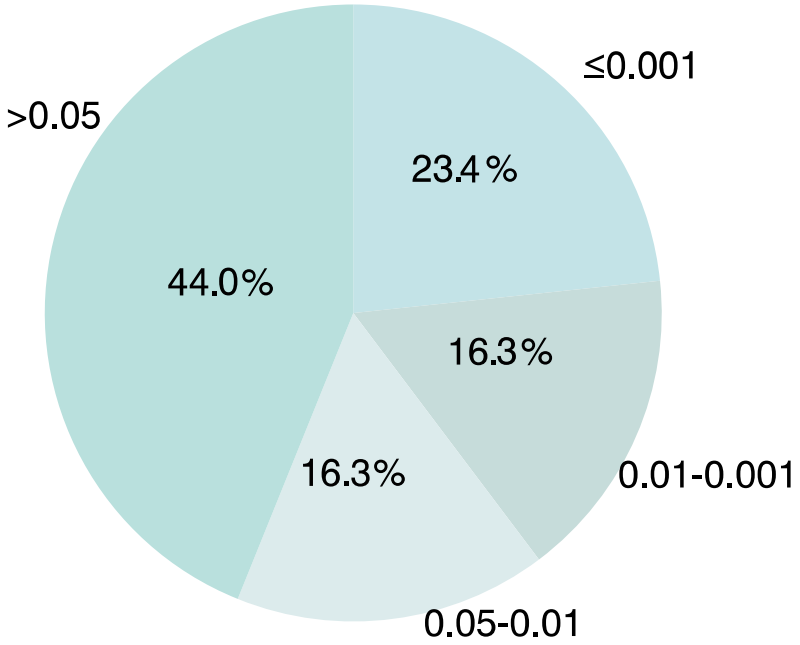

RCTs testing a single primary outcome for superiority
A
Type of primary endpoint
Quantitative/Ordinal endpoint
23.5%
49.7%
26.8%
Binary(disease/event) endpoint
Composite endpoint
B
No. of outcomes if composite primary endpoint
≥5
2
16.2%
18.6%
4
21.0%
44.3%
3
C
No. of secondary outcomes
≥9
14.2%
≤3
30.1%
5-8
15.9%
39.7%
4-6
D
No. of observed primary endpoints
>1500
8.0%
≤250
500-1500
20.3%
54.6%
17.1%
250-500
E
Sample size
Percentage %
30
20
10
0
28.0%
22.6%
12.2%
13.7%
9.8%
6.6%
7.1%
≤250
251-750
751-1500
1501-3000
3001-7500
7501-15.000
>15.000
No. of patients

RCTs testing for superiority in single primary endpoints

A Observed relative risk



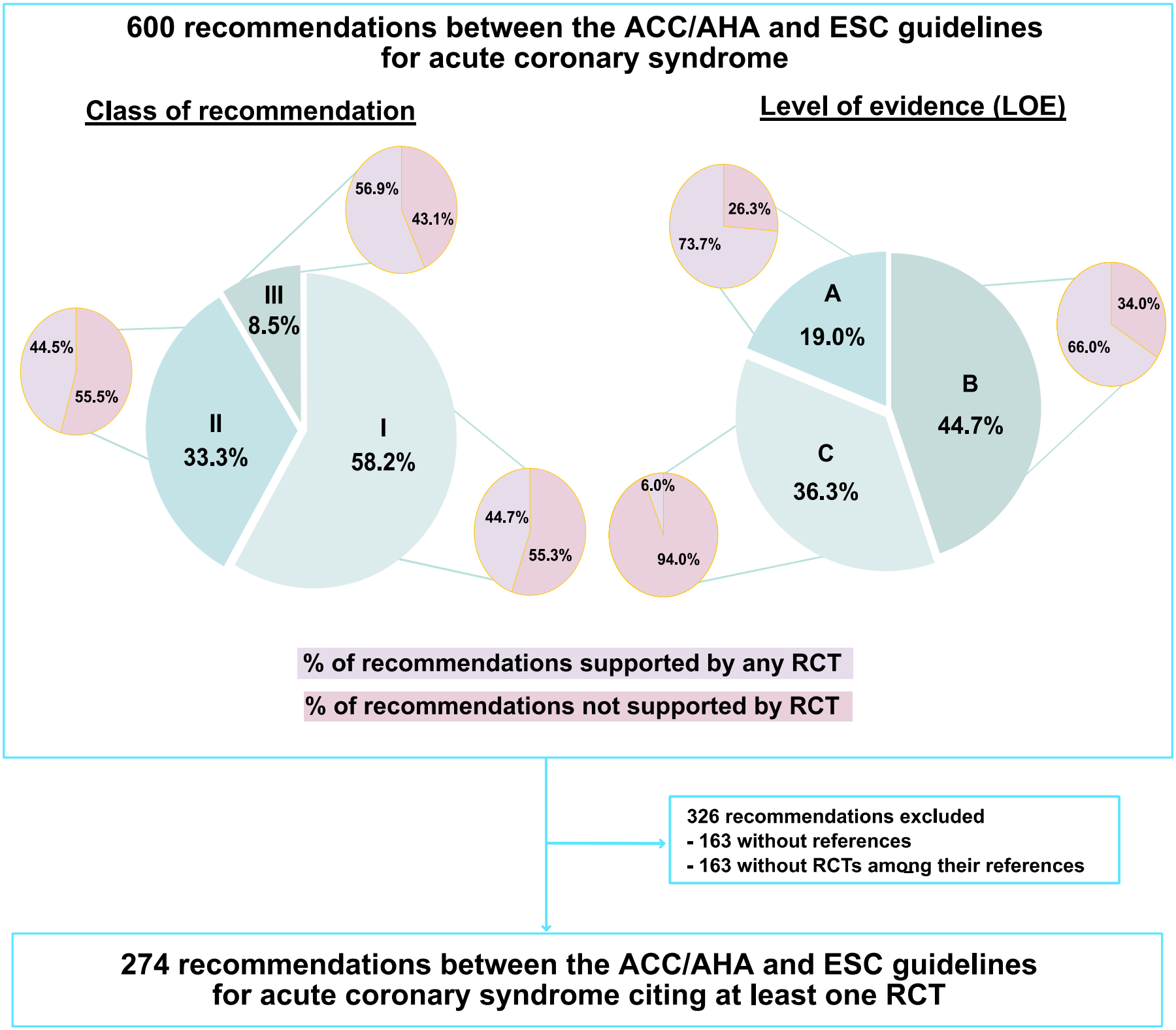
B p-value



C p-value by observed relative risk

	≤0.65 (n=69)	0.65-0.75 (n=30)	0.75-0.85 (n=55)	>0.85 (n=96)
≤0.001	29 (42.0)	10 (33.3)	15 (27.3)	3 (3.1)
0.01-0.001	21 (30.4)	9 (30.0)	7 (12.7)	5 (5.2)
0.05-0.01	13 (18.8)	5 (16.7)	9 (16.4)	14 (14.6)
≥0.05	6 (8.7)	6 (20.0)	24 (43.6)	74 (77.1)

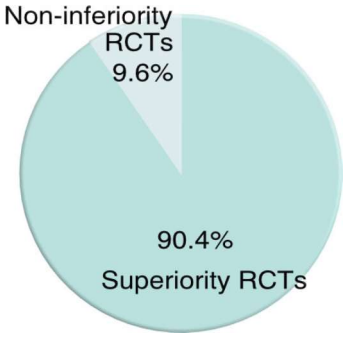
p <0.001


600 recommendations between the ACC/AHA and ESC guidelines for acute coronary syndrome
Class of recommendation
Level of evidence (LOE)
I 58.2%
II 33.3%
III 8.5%
A 19.0%
B 44.7%
C 36.3%
56.9%
43.1%
44.5%
55.5%
44.7%
55.3%
26.3%
73.7%
34.0%
66.0%
6.0%
94.0%
% of recommendations supported by any RCT
% of recommendations not supported by RCT
326 recommendations excluded
- 163 without references
- 163 without RCTs among their references
274 recommendations between the ACC/AHA and ESC guidelines for acute coronary syndrome citing at least one RCT

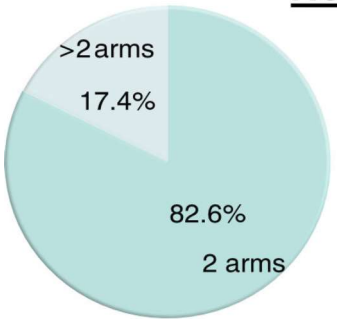
Characteristics of randomised controlled trials cited by acute coronary syndrome clinical practice guidelines

2128 non-duplicated references cited in 2013 and 2014 ACC/AHA and 2017 and 2020 ESC CPGs for STEMI and NSTE-ACS → 407 primary publications of RCTs (19.1%)

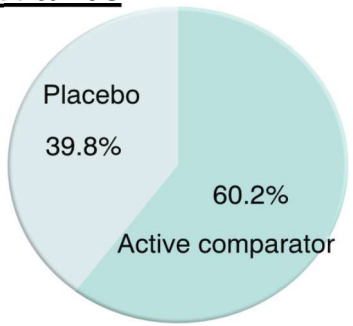
RCTs design features



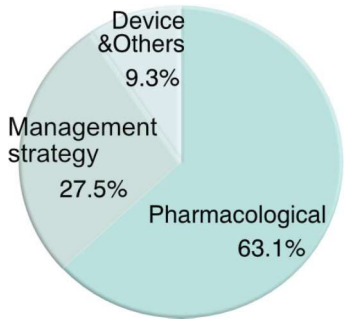
Type of hypothesis



Number of arms

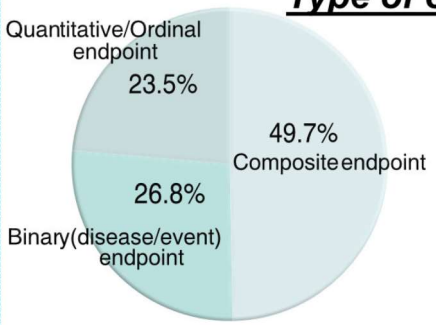


Type of comparator

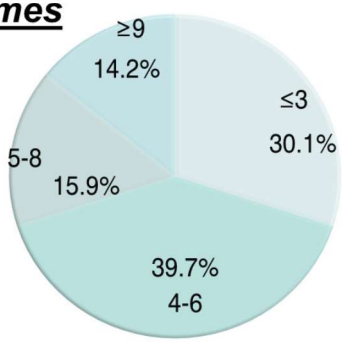


Type of intervention

Type of outcomes

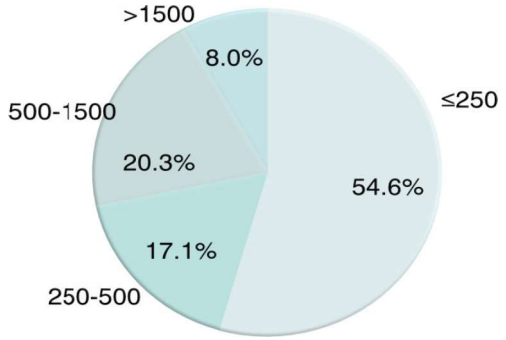


Type of primary endpoint

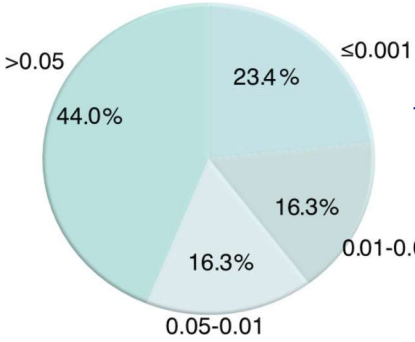


No. of secondary outcomes

No. of observed primary endpoints



Observed p-value



84.2%

RCTs achieving ≥80% of their planned sample size

Median sample size		***Median relative risk***
1327 participants (IQR 445-4060)	***Expected***	0.72 (IQR 0.60-0.80)
1001 participants (IQR 263-3602)	***Observed***	0.81 (IQR 0.61-0.91)

67.6%

RCTs with a lower than expected treatment effect