

openheart Critical appraisal of cardiology guidelines on revascularisation: clinical practice

David R Dobies,¹ Kimberly R Barber²

To cite: Dobies DR, Barber KR. Critical appraisal of cardiology guidelines on revascularisation: clinical practice. *Open Heart* 2018;**5**:e000779. doi:10.1136/openhrt-2018-000779

Received 10 January 2018
Accepted 6 February 2018

ABSTRACT

Evidence-based medicine (EBM) provides clinicians with beneficial information. Nonetheless, study findings are often arbitrary, speculative or provisional. The current state of misleading evidence exists in all applications, including those for guideline recommendations. We conducted an appraisal of the American College of Cardiology and European Society of Cardiology Guidelines for revascularisation of complex coronary anatomy to determine the veracity of the evidence that recommendations were based on. Study-specific critical appraisals were conducted by the authors on the 5-year Synergy between percutaneous coronary intervention with Taxus and cardiac surgery (SYNTAX) and future revascularisation evaluation in patients with diabetes mellitus: optimal management of multivessel disease (FREEDOM) Trials. Each appraisal was performed according to the standard EBM practices. A thorough design and analytic critique was performed for each study and the results presented and explained. The guideline recommendations were reviewed in terms of the veracity of the evidence cited. The relative difference in major adverse cardiac and cerebrovascular event (MAACE) rates between coronary artery bypass grafting (CABG) and percutaneous coronary intervention (PCI) are not the 30% level reported by the SYNTAX Trial but closer to 11% difference when study limitations are factored in. Similarly, the 30% effect size in MAACE rates between procedures from the FREEDOM Trial is closer to a non-significant 5% relative difference when limitations are adjusted for. Based on the actual findings of each study, outcomes from procedures by CABG or PCI for multivessel revascularisation are similar and contradict the conclusions of the study authors as well as the recommendations. These recommendations fail to inform current clinical practice.

INTRODUCTION

The standard for clinical practice has traditionally been derived from guideline recommendations and these standards are reinforced through medical board certification and qualification examinations. More recently, medical administrators and policy-makers have turned to guideline recommendations to advocate particular institutional policies.¹⁻³ An assumption regarding guidelines has persisted that best practices

are based on the best evidence and this evidence informs the recommendations in a rigorous and robust manner.^{2,3} This assumption, however, is false and has serious implications for patient safety and/or well-being.

Evidence-based medicine (EBM) does provide the clinician with beneficial information. Nonetheless, study findings are often arbitrary, speculative or provisional.⁴⁻⁶ The current state of misleading evidence exists in all applications, including those applied for guideline recommendations.^{4,5,7}

We undertook an examination of the American College of Cardiology/American Heart Association (ACC/AHA) and European Society of Cardiology (ESC) Guidelines for a specific cardiology practice, revascularisation of complex coronary anatomy, to determine the veracity of the evidence that the recommendations were based on. This review examines the evidence on complex, coronary disease treatment cited by the guidelines with a critical eye.

METHODS

The 2014 ACC/AHA Guidelines on revascularisation recommend selection of coronary arterial bypass grafting (CABG) over percutaneous coronary intervention (PCI) for patients with complex and multivessel acute coronary syndrome.⁸ The recommendation of CABG as the preferred intervention is made in three of six situations where patients have multivessel disease. These include specifically patients with:

- ▶ three-vessel coronary artery disease (CAD) with intermediate to high CAD burden (multiple diffuse lesions, presence of chronic total occlusion (CTO) or high SYNTAX Score),
- ▶ isolated left main stenosis,
- ▶ left main stenosis and additional CAD with low CAD burden (one-vessel to two-vessel additional involvement and low SYNTAX Score).



¹Cardiology Department, Regional Cardiology Associates, Grand Blanc, Michigan, USA
²Research Department, Ascension/Genesys Regional Research, Grand Blanc, Michigan, USA

Correspondence to
Professor Kimberly R Barber;
kimberly.barber@ascension.org

Table 1 ACCF/AHA/SCAI PCI guideline evidence definitions by class and level of recommendations^{9 10}

	Class I	Class IIa	Class IIb	Class III
Level A				
ACC	Sufficient evidence from multiple randomised trials or meta-analyses	Some conflicting evidence from multiple randomised trials or meta-analyses	Greater conflicting evidence from multiple randomised trials or meta-analyses	Sufficient evidence from multiple randomised trials or meta-analyses
ESC	Evidence/general agreement of efficacy Data from multiple randomised trials or meta-analyses	Weight of evidence/opinion in favour of efficacy Data from multiple randomised trials or meta-analyses	Efficacy less well established by evidence/opinion Data from multiple randomised trials or meta-analyses	Evidence/agreement of non-efficacy or possible harm Data from multiple randomised trials or meta-analyses
Level B				
ACC	Evidence from single randomised trial or non-randomised studies	Some conflicting evidence from a single randomised trial or non-randomised studies	Greater conflicting evidence from a single randomised trial or non-randomised studies	Evidence from a single randomised trial or non-randomised studies
ESC	Evidence/general agreement of efficacy Data from single randomised trial or large non-randomised studies	Weight of evidence/opinion in favour of efficacy Data from single randomised trial or large non-randomised studies	Efficacy less well established by evidence/opinion Data from single randomised trial or large non-randomised studies	Evidence/agreement of non-efficacy or possible harm Data from single randomised trial or large non-randomised studies
Level C				
ACC	Only expert opinion, case studies or SOC	Only diverging expert opinion, case studies or SOC	Only diverging expert opinion, case studies or SOC	Only expert opinion, case studies or SOC
ESC	Evidence/general agreement of efficacy Consensus of experts and/or non-randomised studies	Weight of evidence/opinion in favour of efficacy Consensus of experts and/or non-randomised studies	Efficacy less well established by evidence/opinion Consensus of experts and/or non-randomised studies	Evidence/agreement of non-efficacy or possible harm Consensus of experts and/or non-randomised studies
Recommendation	Is recommended/indicated	Can be useful/effective/beneficial or indicated	May be considered/reasonable	No Benefit/not recommended or potentially harmful

ACC (American College of Cardiology), ESC (European Society of Cardiology); PCI (percutaneous coronary intervention); SOC (standard of care).

PCI is ranked as a class IIa or class III/level B for all three of the above situations while CABG is ranked as a class I/level A or B in all three.⁸ The 2014 ESC guidelines make similar recommendations for patients with multivessel disease and intermediate to high CAD burden based on SYNTAX Score risk.⁹ These classifications suggest that CABG is a preferred intervention for these patients based on evidence defined in broad terms of the presence and number of studies available (table 1). Because research is not static and evidence is not absolute, a closer look at the recommendations cited evidence is warranted.

Evidence

The ACC and ESC Guideline rankings are based on the study findings from two major randomised controlled trials (RCTs): (1) the 5-year SYNTAX Trial and (2) the FREEDOM Trial.^{10–12} Other studies cited by the guidelines (Cardia, BARI, ARTS, MASS) pre-date the advancement of newer stents and/or are small and underpowered.

SYNTAX

The findings of the SYNTAX Trial require closer scrutiny. The study conclusions state that CABG should remain

the standard of care for patients with complex lesions (high or intermediate SYNTAX Scores).¹⁰ They base this conclusion on the findings from the randomised arm of the 5-year SYNTAX Trial which demonstrated that after a 5-year follow-up, MACCE occurred in 26.9% of patients in the CABG group and 37.3% of patients in the PCI group. This 27.9% relative difference was highly significant ($P < 0.0001$). There is no doubt that the data used in the analysis showed a significant difference in rates (figure 1A). The question is whether and to what extent the data represent precise and reproducible results.

The investigators enrolled and randomised a total of 1800 patients into the trial. Of these, 60 patients (3.3%) were lost to follow-up by the first year. A total of 124 (6.9%) were lost to follow-up before the 5-year survival analysis. More importantly, we see a differential in the dropout rate with the CABG group losing significantly more patients after randomisation than the PCI group (10.3% vs 3.5%). It is possible that none of these patients had events, but given the reported overall event rate of 32%, it is not likely. It is also possible that the group with the higher dropout had more events. This we do not know. Therefore, the appropriate intention-to-treat (ITT)

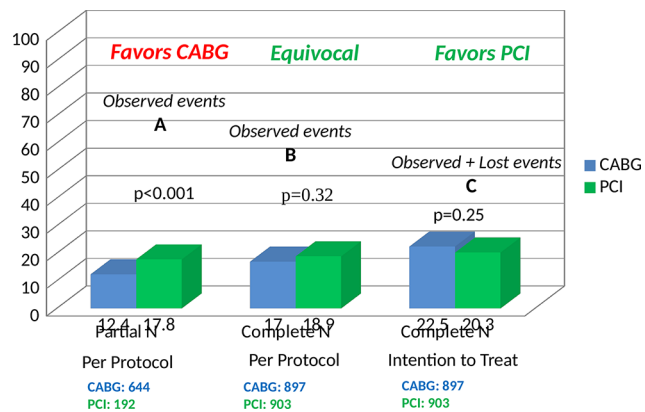


Figure 1 SYNTAX 5-Year Trial outcomes by procedure and analysis.

analysis is to transpose those data as events for their respective groups.^{5,6} This would add an additional 92 events to the CABG group and an additional 32 events to the PCI group. The revised MACCE rates (figure 1B) would be 35.5% in the CABG group compared with 40.0% in the PCI group (RRD=0.88; 95% CI 0.80 to 1.0; P=0.05). Thus, the revised relative difference is 11.2% not the almost 30% difference presented in the paper.

In addition, the study definition for myocardial infarction (MI) differed for each treatment group based on an arbitrary cut-off. The measure for a PCI-related MI was a creatine kinase-myocardial band (CK-MB) of 5 IU/L and for the CABG-related MI it was a CK-MB of 10 IU/L.¹³⁻¹⁵ Since enzymes can rise with no presentation of symptoms and since the majority of MACCE events in the PCI group were due to myocardial infarctions, a more accurate representation of MACCE composite should be based on EKG findings or patient-important outcomes such as symptoms.¹⁴ The conclusion that CABG and PCI should have different rankings because of differences in an equally weighted combined outcome misrepresents the legitimacy of the conclusion that CABG is superior to PCI in terms of MACCE. Furthermore, CK-MB test has been largely replaced by troponin t and troponin i, markers that are more specific to cardiac tissue.¹³

Although the study authors calculated an ITT analysis, they did so transposing the lost data as non-events. For this study population, that is highly unlikely and doing the analysis by only adding those patients into the denominator rather than both the numerator and denominator makes the findings misleading. The study authors also emphasised a non-ITT (per-protocol) analysis rather than an ITT (per assignment) analysis. A per-assignment ITT calculation results in a reversal of outcomes (figure 1C)—not significant but enlightening regarding the veracity of the study's conclusions. The per-protocol analysis conceals what the findings would be if none of the patients had been lost. That is because the lost patients destroy the similarity of the groups as the study proceeds. This is particularly important where there is a differential in lost outcomes from the groups. The per-protocol finding cannot provide an accurate comparison, simply

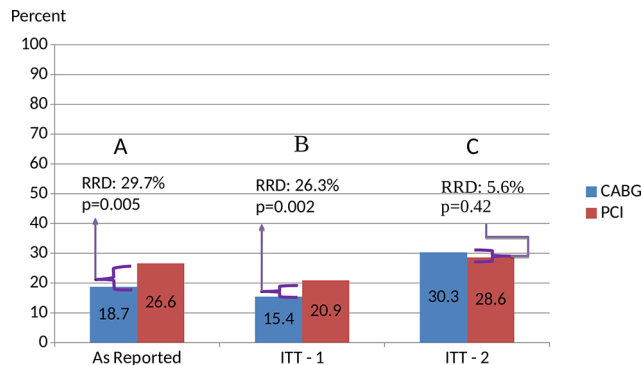


Figure 2 FREEDOM Trial results comparing ITT analyses. CABG, coronary arterial bypass graft surgery; ITT, intention-to-treat; PCI, percutaneous coronary intervention.

because the groups are no longer similar. *Once the randomisation is destroyed, the findings can be misleading.*

FREEDOM

The FREEDOM Trial is also widely cited in the guideline references.^{11,12} Authors proclaimed that this study would provide the 'definitive' answer to the controversy of the preferred revascularisation procedure for multivessel disease among diabetics. The study showed that the outcomes were significantly lower among patients randomised to CABG (18.7%) than patients randomised to PCI (26.6%) (figure 2A). A closer look at how these rates were derived is warranted. A total of 1900 patients (953 in the PCI group and 947 in the CABG group) were enrolled and randomised. However, for the 5-year outcome rates, the denominator was 752 for PCI and 781 for CABG. These numbers are not the group totals but rather the number of patients remaining at risk at the end of the study. The number of events and the number remaining at risk are independent of each other. A basic tenant of a rate is that the subjects in the numerator are included in the denominator. The percentage the authors report are not rates, they are ratios and are very misleading. Calculating the events among the number randomised in each group results in a relative difference of 26% that is less significant than reported (figure 2B).

We would have more confidence in these recalculated rates if the study included all subjects in the denominator and accounted for outcomes on all subjects. They do not include the 214 (11.3%) patients lost to follow-up for whom we have no outcome data. This study also experienced a significant differential in attrition by group. The CABG group had twice the patients lost to follow-up (14.9%) as the PCI group did (7.7%). Revising the comparison by adding in the lost patients as events and calculating it with an ITT analysis (attributing events to the group of original assignment), we get a very different picture for the 5-year outcome (figure 2C). The relative 5% difference is not significant (P=0.42). This finding is in line with the 2-year composite outcomes in which the study authors observed no difference in outcome rates

(13.0% vs 11.9%, $P=0.51$). *The 5-year finding is significantly biased by the differential attrition rate.*

Another concern is that the prominence of the SYNTAX Score in the recommendations. This is a score that was designed during the SYNTAX Trial and these arbitrary categories showed a correlation with general risk outcomes.¹⁰ However, its use in general practice requires validation (ie, confirmation) in a large independent dataset such as an all-inclusive, clinical registry which has not been conducted. The FREEDOM Trial conducted a subgroup analyses with the SYNTAX Score categories. This would not be the large clinical registry necessary to test the validity of the index, yet it is a dataset independent of the SYNTAX trial.¹¹ Using the cut-offs as created (≤ 22 , 23–32, >33), we see that the tool fails to predict outcomes in terms of a primary composite. There was no association between type of procedure and outcomes for the low and high categories and close to a null effect for the intermediate category. *At least among diabetics, either the SYNTAX Score is arbitrary or the finding as to a CABG benefit is inaccurate.*

DISCUSSION

There is growing awareness of a systematic problem with the accuracy of reported research, particularly with, but not limited to, RCTs.^{16 17} Because RCTs are held up as the gold standard of clinical practice change, it is important that clinicians understand whether study findings and the guideline recommendations they inform are valid. Threats to the validity of a study are more widespread than previously appreciated and require a critical eye by the clinician. Scientific journals have fallen short of being the arbiter of trustworthy research for the clinician, and guideline recommendations are not immune to research finding duplicity.¹⁸

The evidence cited for the guideline recommendations is seriously flawed. Limitations include the differential drop out, differential outcome definitions, the lack of a complete ITT analysis and the lack of rigour in the calculations on outcomes.

It highlights a widespread problem with the current review process, specifically the lack of expert critical appraisal on study findings, which leads to specious conclusions. Published research that appears elaborate puts forth findings that have not been thoroughly vetted. Instead, an article's conclusion is substituting for the study's findings because few understand how to do the deep dive into what the study data actually shows. Guideline committees are not immune to this problem. Recommendations are limited to the extent that the study is flawed. Clinically usable evidence requires a definition on the *quality* of the study (ie, strength of design, precision of data and reproducibility of findings) rather than a study's type or quantity.

We presented an example of a typical clinical decision in cardiology (surgery vs PCI) for which the answer has been distorted by study publications and perpetrated by

the guidelines. The SYNTAX Trial emphasised findings that differed significantly from the ITT analysis and did not account for all patients and the FREEDOM Trial presented a non-reproducible rate calculation. These findings have yet to be validated but are discoverable by reviewing the supplemental papers and visiting the ClinicalTrials.gov website. Studies *do exist* utilising the RCT design to address the clinical question (class I) and there is *more than one* RCT on the clinical question (level A).

However, they do not meet the spirit of evidence based on rigorous scientific findings that are clinically relevant and reproducible. These trials have fundamental and serious flaws with findings that are imprecise and uncertain. This critique demonstrates how the purported evidence for the superiority of CABG is weak when examined closely. Combining several flawed studies together does not strengthen the conclusion. Furthermore, when taking patient individuality and patient values into consideration, a recommended procedure may not be the more beneficial option.^{15 19}

BARBER-DOBIES EBM PRINCIPLES-IN-PRACTICE

There are general rules that are helpful for making decisions about changing practice or accepting recommendations based on evidence. Several rules-of-thumb can help apply a critical eye to the evidence. For example, complete follow-up of trial patients is ideal but not realistic and yet lost patients and missing data can completely sway the finding in a direction that does not reflect reality. A 5/20 rule can be used to determine whether the finding is sufficient for the conclusions.^{7 8} A study missing $\geq 20\%$ of patient outcome data can significantly change the finding and invalidate the conclusions. Losing participants during the conduct of a trial skews results in unpredictable ways. *The study can be rejected on this alone.* A loss of 5% or less of the data will not impact the findings and the critique should continue. A loss of data between 5% and 20% is a grey zone that may or may not impact the finding and a recalculation is in order. Our critique of the FREEDOM Trial demonstrates a significant shift in the finding based on missing data that was within this grey zone (11.3%). The busy clinician could restrict the rule to a 5/10 cut-off and would not likely miss real evidence capable of changing practice.

Another useful rule is to examine the 95% CIs surrounding the difference (mean) or comparison (rate) statistic.⁶ A CI that includes 0 for a mean difference (eg, $-5-0+5$) or includes 1 for a rate difference (eg, 0.7–1–2.0) is not a significant finding. The width of a 95% CI is also important. If the 95% CI is narrow (eg, 2.5 to 3.5) the study finding is indicating more precision to that statistic and we can be more confident that the finding, say 3.0, is close to the 3.0. However, if the 95% CI is wide (eg, 1.0 to 12.0) then a 3.0 ratio is very imprecise. Imprecision equates with greater uncertainty and is rarely reproducible. *Uncertainty in study data cannot represent certainty in practice and the study can be rejected.* Evidence that is highly

imprecise should not be applied to changing practice or following recommendations.

A more difficult concept to apply but important nonetheless is the ITT method. In a RCT, endpoints are attributed (ie, applied) to one group or the other to calculate and compare their rates. Depending on which group the endpoint is applied could change the rates and thereby shift the finding one way or the other. The only appropriate method to compare group outcomes is with ITT.^{7 20} Outcomes are attributed to the group that the patient was initially assigned to 'at randomisation'. Other methods of calculation (such as a per-protocol analysis) shift outcomes away from their originally assigned group and thus destroy the similarity of the groups. Group similarity is the major strength of the RCT and without it, there is uncertainty in the true reason for group differences in outcomes. *Study conclusions not based on ITT analysis can be rejected.* The non-ITT findings are not reproducible and should not be applied to decisions about changing practice.

Furthermore, the importance of careful scrutiny for combined outcomes cannot be overemphasised for clinical trials. Trial endpoints are often combined such as death, myocardial infarction and repeat revascularisation, which have differing clinical weights. One may be benign and temporary while another is malignant and permanent. Yet the study weights the items equally. Such benefit-risk assessments rely on data that represent a group experience, not the effect of the drug on individual patients. There are also other endpoints not included in the index analysis of SYNTAX that are of similar impact to the patient, including rehospitalisation and postprocedure atrial fibrillation, which further dilutes the use of its findings. Thus, clinical uncertainty is increased because of the widely varying weight when based on meaningful outcomes for patients.

When clinicians are not able to complete the most cursory evidence review, they can turn to a professional systematic review by EBM experts. A systematic approach that examines existing evidence judges its rigour and utility and puts the varying evidence together to provide an overall measure of effectiveness is necessary. Clinicians should look for such 'systematic reviews' (not to be confused with meta-analysis) that include individual study critiques and an overall assessment of reproducibility, generalisability and applicability to current practice. Systematic reviews may also have limitations or flaws. However, a systematic review that is performed expertly and transparently may provide the clinician with the closest understanding of best practice that we can get.²⁰

WHAT IS NEEDED

Guidelines should not only cite the evidence used for the recommendation but also the critical appraisals performed and make them available as a supplement to the guidelines. Critical appraisals need to be performed in a systematic, standardised, non-biased manner, similar

to the Grading of Recommendations Assessment, Development and Evaluation reporting process.^{21 22} Guideline committee endorsements of the evidence will strongly affect clinical practice.²¹ There is an expanding number of recommendations based on inconclusive evidence and citing evidence without publishing their appraisals seems to be merely a gratuitous advertisement for each study.²³

Future clinicians need to be versed in critical appraisal of published research. A more rigorous training for a prerequisite critical appraisal course is needed in medical schools.

In our current world of healthcare transformation, payment reform may soon be linked to guideline recommendations. Flawed clinical trial findings could be held out as recommendations which will become quality metrics and then inserted into policy²⁴. Following inaccurate guidelines will thus lead to compromised care.

CONCLUSION

Based on the actual findings of current trials, outcomes from procedures by CABG or PCI for multivessel revascularisation are similar and contradict the conclusions of the study authors as well as the recommended guidelines. These recommendations fail to inform current clinical practice.

Funding This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data available.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

© Article author(s) (or their employer(s) unless otherwise stated in the text of the article) 2018. All rights reserved. No commercial use is permitted unless otherwise expressly granted.

REFERENCES

1. Alamri HS, Alotaiby M, Almoghairi A, *et al.* Lessons from the SYNTAX trial. *J Saudi Heart Assoc* 2010;22:35–41.
2. Masoudi FA, Curtis JP, Desai NR. PCI Appropriateness in New York: if it makes it there, can it make it everywhere? *J Am Coll Cardiol* 2017;69:1243–69.
3. Hannan EL, Samadashvili Z, Cozzens K, *et al.* Changes in percutaneous coronary interventions deemed "inappropriate" by appropriate use criteria. *J Am Coll Cardiol* 2017;69:1234–42.
4. Howick J. *The philosophy of evidence-based medicine*. 1st edn: Blackwell Publishing Ltd, 2011.
5. Pocock SJ, McMurray JJ, 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;66:2536–49.
6. Pocock SJ, McMurray JJ, 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;66:2648–62.
7. Every-Palmer S, Howick J. How evidence-based medicine is failing due to biased trials and selective publication. *J Eval Clin Pract* 2014;20:908–14.
8. Levine GN, Bates ER, Blankenship JC, *et al.* 2011 ACCF/AHA/SCAI Guideline for Percutaneous Coronary Intervention: a report of the American College of Cardiology Foundation/American Heart

- Association Task Force on Practice Guidelines and the Society for Cardiovascular Angiography and Interventions. *Circulation* 2011;124:e574–651.
9. Windecker S, Kolh P, Alfonso F, *et al.* 2014 ESC/EACTS Guidelines on myocardial revascularization: The Task Force on Myocardial Revascularization of the European Society of Cardiology (ESC) and the European Association for Cardio-Thoracic Surgery (EACTS) Developed with the special contribution of the European Association of Percutaneous Cardiovascular Interventions (EAPCI). *Eur Heart J* 2014;35:2541–619.
 10. Milojevic M, Head SJ, Parasca CA, *et al.* Causes of Death Following PCI Versus CABG in Complex CAD: 5-Year Follow-Up of SYNTAX. *J Am Coll Cardiol* 2016;67:42–55.
 11. Dangas GD, Farkouh ME, Sleeper LA, *et al.* Long-term outcome of PCI versus CABG in insulin and non-insulin-treated diabetic patients: results from the FREEDOM trial. *J Am Coll Cardiol* 2014;64:1189–97.
 12. Farkouh ME, Domanski M, Sleeper LA, *et al.* Strategies for multivessel revascularization in patients with diabetes. *N Engl J Med* 2012;367:2375–84.
 13. Moussa ID, Klein LW, Shah B, *et al.* Consideration of a new definition of clinically relevant myocardial infarction after coronary revascularization: an expert consensus document from the Society for Cardiovascular Angiography and Interventions (SCAI). *J Am Coll Cardiol* 2013;62:1563–70.
 14. Idris H, Lo S, Shugman IM, *et al.* Varying definitions for periprocedural myocardial infarction alter event rates and prognostic implications. *J Am Heart Assoc* 2014;3:e001086.
 15. Pardo-Hernandez H, Alonso-Coello P. Patient-important outcomes in decision-making: a point of no return. *J Clin Epidemiol*. In Press. 2017;88:4–6.
 16. Cramer GW. Form and Function in the News: Large-scale Analysis Finds Majority of Trials Provide No Meaningful Evidence. *Off the Wire*. 2012 www.eurekalert.org/EurekaAlert
 17. Armstrong JJ, Goldfarb AM, Instrum RS, *et al.* Improvement evident but still necessary in clinical practice guideline quality: a systematic review. *J Clin Epidemiol* 2017;81:13–21.
 18. Cramer GW. Form and Function in the News: Editors Question if Modern Science has Become Dysfunctional. *Off the Wire*. 2012 www.eurekalert.org/EurekaAlert
 19. Pocock SJ, Stone GW. The Primary Outcome Is Positive - Is That Good Enough? *N Engl J Med* 2016;375:971–9.
 20. Gopalakrishnan S, Ganeshkumar P. Systematic Reviews and Meta-analysis: Understanding the Best Evidence in Primary Healthcare. *J Family Med Prim Care* 2013;2:9–14.
 21. Guyatt GH, Oxman AD, Montori V, *et al.* GRADE guidelines: 5. Rating the quality of evidence—publication bias. *J Clin Epidemiol* 2011;64:1277–82.
 22. Schünemann HJ, Schünemann AH, Oxman AD, *et al.* Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. *BMJ* 2008;336:1106–10.
 23. Tricoci P, Allen JM, Kramer JM, *et al.* Scientific evidence underlying the ACC/AHA clinical practice guidelines. *JAMA* 2009;301:831–41.
 24. Abrahá I, Cozzolino F, Orso M, *et al.* A systematic review found that deviations from intention-to-treat are common in randomized trials and systematic reviews. *J Clin Epidemiol* 2017;84:37–46.