On the Pooling and Subgrouping of Data from PCI versus CABG Trials: A Call to Circumspection

Freemantle N[ad], Ruel M[b], Gaudino MFL[c], Pagano D[d].

a Comprehensive Clinical Trials Unit, Institute for Clinical Trials and Methodology, University College London, London, UK

b University of Ottawa Heart Institute, Ottawa, Ontario, Canada

c Department of Cardiothoracic Surgery, Weill Cornell Medicine, New York, USA

d Quality and Outcomes Research Unit, University Hospital Birmingham NHS Foundation Trust, Birmingham B15 2PR, UK

In the modern era, treatment choice is guided by scientific evidence, usually gathered from well-conducted clinical trials, and often followed by the pooling of their data. In this article, we review the most recent pooled evidence regarding myocardial revascularisation strategies, and discuss how these meta-analyses have inherent shortcomings that should be better understood, prior to their purported conclusions potentially influencing clinical decisions.

Properly conducted randomised trials comparing PCI with CABG can provide unbiased estimates of treatment effects. However, the design of such trials has often involved a primary outcome that is a composite measure, and thus open to challenges with regard to the appropriate interpretation of each individual components.[1] Quantitative synthesis of such data from multiple trials can enable estimates of individual components of the composite outcomes (e.g. all-cause mortality). Where individual patient data are available, a full investigation of mediating effects, or subgroups analyses, may also be undertaken. Relevant to this article, we now have multiple meta analyses of trials available, [2, 3, 4] which provide an opportunity to assess the appropriateness of criteria for patient selection between PCI and CABG. Consequently, we will appraise below the robustness of these meta-analytic methods in answering the question at stake: does PCI provide equivalent results to CABG for the treatment of unprotected left main coronary artery stenosis and multivessel CAD?

Network Meta-Analyses

Windecker and colleagues[4] summarised the data from trials of PCI or CABG which compared any combination of PCI with alternative PCI strategies (eg different drug eluting stents), medical therapy, or CABG, pooling 93553 patients from 100 randomised trials in a network meta-analysis. The purpose of a network meta-analysis is to provide comparisons between all treatments which link through a chain of common comparators, whether or not there are randomised comparisons between them. Curiously, despite the choice for treatment to be most likely between two interventional strategies, the authors chose to describe the differences between intervention and medical therapy and did not report the difference between PCI and CABG. Of the significant findings on all-cause mortality, CABG reduced mortality with a rate ratio of 0.80 (95%CI 0.70 to 0.91) compared to medical therapy in the network analysis. The authors also found similar results in directly randomised trials. Among PCI treatments, only everolimus eluting stents (EES) with a rate ratio of 0.75 (95% CI 0.59 to 0.96) and, marginally, zotarolimus eluting (R-ZES) with a rate ratio of 0.65 (95% CI 0.42 to 1.00), showed evidence of superiority over medical therapy alone. These results were driven almost entirely by information derived from indirect comparison in trials that compared different PCI stents, with no direct comparison with CABG, and only one directly randomised comparison with medical therapy for EES, providing an underpowered relative risk of all-cause mortality of 0.33 (95% CI 0.03 to 3.16; p=0.34).

The limitation of network meta-analyses is that they require the strong, generally unrealistic, assumption that treatment works in the same way in each trial included in the analysis. This assumption is not required for individual trials, and individual patient meta-analyses.

Individual Patient Meta-Analyses

Individual patient meta-analyses are particularly useful, because i) they can be used to examine treatment effects for consistency; and ii) strategies for the examination of subgroups are well established in the methodological literature.[5; 6, 7] The identification of true subgroups can be challenging. As Salim Yusuf and colleagues noted:

"A key principle for interpretation of subgroup results is that quantitative interactions (differences in degree) are much more likely than qualitative interactions (differences in kind). Quantitative interactions are likely to be truly present whether or not they are apparent, whereas apparent qualitative interactions should generally be disbelieved…."[5]

The key statistical tool to differentiate between true and chance differences in effect within subgroups is the test for interaction, which indicates whether the observed differences are of a magnitude which we would consider unlikely to have occurred by the play of chance. Within individual trials such interaction tests are notoriously underpowered and may fail to identify real subgroup differences. This challenge may be overcome in individual patient meta-analyses. Indeed, regulatory agencies generally require an individual patient meta-analysis of all relevant trials in order to support the approval of an investigational medical product.

It is particularly welcome that an individual patient meta-analysis of 11 trials comparing PCI with CABG including some 11518 patients, of whom 976 died (8.5%), has recently been published.[2] The primary analysis from the Head et al [2] shows that CABG is superior to PCI, with PCI associated with a 20% increased relative risk of death (95% CI 6% to 37%; p=.004). The authors also undertook extensive testing of subgroups, the principal ones being described in Figure 1.

[Figure 1. Here]

In Figure 1. we observe that all of the point estimates are greater than 1 (e.g. indicating numerical superiority of CABG), although many have wide confidence intervals or are close to 1, thus showing no apparent convincing benefit. In this regard, if we drew a series of random subsets of the patients in the analysis by Head et al, we could find the play of chance expressed in the variation of the effects observed, as Richard Peto and colleagues famously did in ISIS-2 [8]. In that classic analysis, Peto et al. identified that patients with a star sign of Gemini or Libra appeared to experience increased mortality when taking aspirin, but that a strong benefit was associated with aspirin for other star signs. Of course nobody believes that star sign affects the action of aspirin, and the test for interaction was non-significant, but this example provides a timely reminder that we must combine statistical rigour not only with biological plausibility, but also with prospective definition of endpoints, in order to separate true subgroups effects from chance phenomena. [5]

Only the interaction between diabetes and outcome in Head et al. is nominally significant, with an at first glance impressive p value of .008. However, as Yusuf and colleagues note,[5] it is important to consider both the plausibility and the questions of multiplicity. The latter are quite straightforward here: Head et al likely performed many tests for interaction, but highlight these 10 as the principal ones (Figure 1). Each test conducted at a critical p value of 5% independently has a false positive rate of 1 in 20. Indeed, each test is analogous to rolling a 20 sided dice which has on one side is a tick and on 19 sides a cross, so a good strategy for gaining a tick on the dice is to roll it many times. Adjusting the p value for multiplicity using a simple Bonferroni adjustment for the 10 tests ('rolls of the dice') provides an adjusted critical p value of .005; in other words, diabetes does not achieve significance on the statistical criterion.

There are reasons to consider that patients with diabetes face increased risk with PCI,[9] and it could be construed that the benefits of treatment appear in the analysis to be numerically concentrated only in patients with diabetes. However, we do not also see the difference in patients who have conditions related to diabetes such as prior MI, hypertension, hypercholesterolemia or PVD, highlighting the plausibility of the play of chance as an explanation for the effect observed. This inconsistency is further highlighted by an analysis previously conducted on a subset of the trial data in non-diabetic patients with multivessel disease, where the Chang et al[3] found a substantial benefit for CABG over PCI.

The number of patients randomised in the trials included in the Head et al[2] analysis is impressive. However, in time to event analysis the power is driven by the number of events (eg deaths) rather than the number of patients randomised (eg 'at risk of death'). As only 8.5% of randomised patients died during the relatively short follow up (mean 3.8 years) reported in the Head et al analysis, the data are not suitable for the estimation of subgroup effects.

Another important point highlighted by Head's paper is the weak evidence in support of the use of the SYNTAX score for patients' stratification. In Head's analysis no significant treatment interaction with the three SYNTAX subgroups was demonstrated. Although the effects of treatment appears to increase with increasing subgroup score, the test for interaction shows that these apparent differences are not more than we would expect to see by chance alone (p=0.21). In this situation Yusuf et al[5] correctly remind us that we should favour the primary outcome of the analysis. In order to identify whether any interaction truly exists we need more events. While these could be gleaned from new randomised trials, a more efficient approach (and more relevant given the decision problem and the longer term experience of patients) would be to enhance follow up of the existing trial populations. This strategy should be pursued with some urgency and must be viewed as an important priority for all the cardiovascular community.

Finally, a curious aspect of the Head et al paper is that they separate Left Main and Multivessel disease, providing estimates within each subgroup, but do not report a test for interaction. Calculating the interaction test from the results provided in Head et al [2] reveals a nominal p value of 0.12, indicating that this interesting comparison in fact does not constitute a true subgroup effect.

Implications for clinical practice

Evidence from clinical trials and meta-analysis is summarised in clinical guidelines which provide a guide for the treatment of patients groups with similar characteristics. The treatment of the individual patient with coronary artery disease, is however decided within the individual Heart Teams who take into account several factors including local expertise and facilities, clinical outcomes and the patient's preference. How should the patient level meta-analysis by Head et al[2] and other recent evidence[3,4] influence existing guidelines?

Whilst the work by Windecker et al[4] provides convincing evidence that CABG is superior to medical treatment of Coronary artery disease as described above, the work by Head et al[2] demonstrates the superiority of CABG over PCI on all-cause mortality. The results of the primary analysis describe a substantial increase in mortality with PCI (HR 1.20; 95% CI 1.06 to 1.37, p=0.004) detectable at a medium term follow up (mean 3.8 years).

The next question is whether, in providing recommendations for treatment, we can identify subgroups of patients in whom the benefit of CABG over PCI varies?

SYNTAX score

The previous guidelines [10] divided patients on the basis of coronary anatomy complexity, using the SYNTAX Score. Although this has become an established algorithm in common practice, it is not underpinned by robust scientific evidence. No trial has used the SYNTAX score as a prospectively declared primary or secondary outcome. Head et al[2] found no significant interaction between the SYNTAX. Thus, based on the current evidence, CABG should be inferred to achieve better results than PCI in all patients regardless of Syntax Score. This conclusion applies both to Multivessel disease and left main.

Multivessel disease and Left main disease.

Head et al[2] provide an opportunity to compare the treatment of LMS with that of patients with multivessel disease although no statistical comparison was described in the paper precluding robust interpretation of these results. Fortunately, data provided in the manuscript allows the comparison of the treatment effects in these two groups, with a non significant P=0.12 for interaction, hence for the LMS subgroup we should infer the primary end point and a substantial disadvantage of PCI (HR 1.20, 95% CI 1.06 to 1.37, p=0.004). On this point, the Head et al[2] conclusions offered in their paper are scientifically inappropriate and should be corrected.

Conclusions

The recent evidence on CABG vs PCI provide an opportunity to rethink the way clinical guidelines and recommendations are made to guide best treatment.

The finding, that SYNTAX Score grouping is not supported by the scientific evidence, poses some challenges and invites a novel way of thinking and a break from the past. Although with insufficient events for subgroup analyses, these new data describe a substantial and convincingly statistically significant mortality benefit of CABG over PCI. Previous studies have suggested that this difference is driven by a reduction in cardiovascular deaths, a more appropriate outcome to evaluate differences in treatments, and in the SYNTAX Trial PCI had a higher cardiovascular mortality than CABG.[11] This finding demonstrates that CABG remains the gold standard treatment for coronary disease. These findings are applicable to patients with a low or acceptable risk for surgery and individual patient treatment should be decided by heart team approach.

An evolving area within Guidelines has been the treatment of Left main disease. The recent evidence suggests a significant superiority of CABG over PCI irrespective of coronary anatomy and there is no good evidence at this stage to suggest equivalence of PCI with CABG in this subgroup of patients.[2]

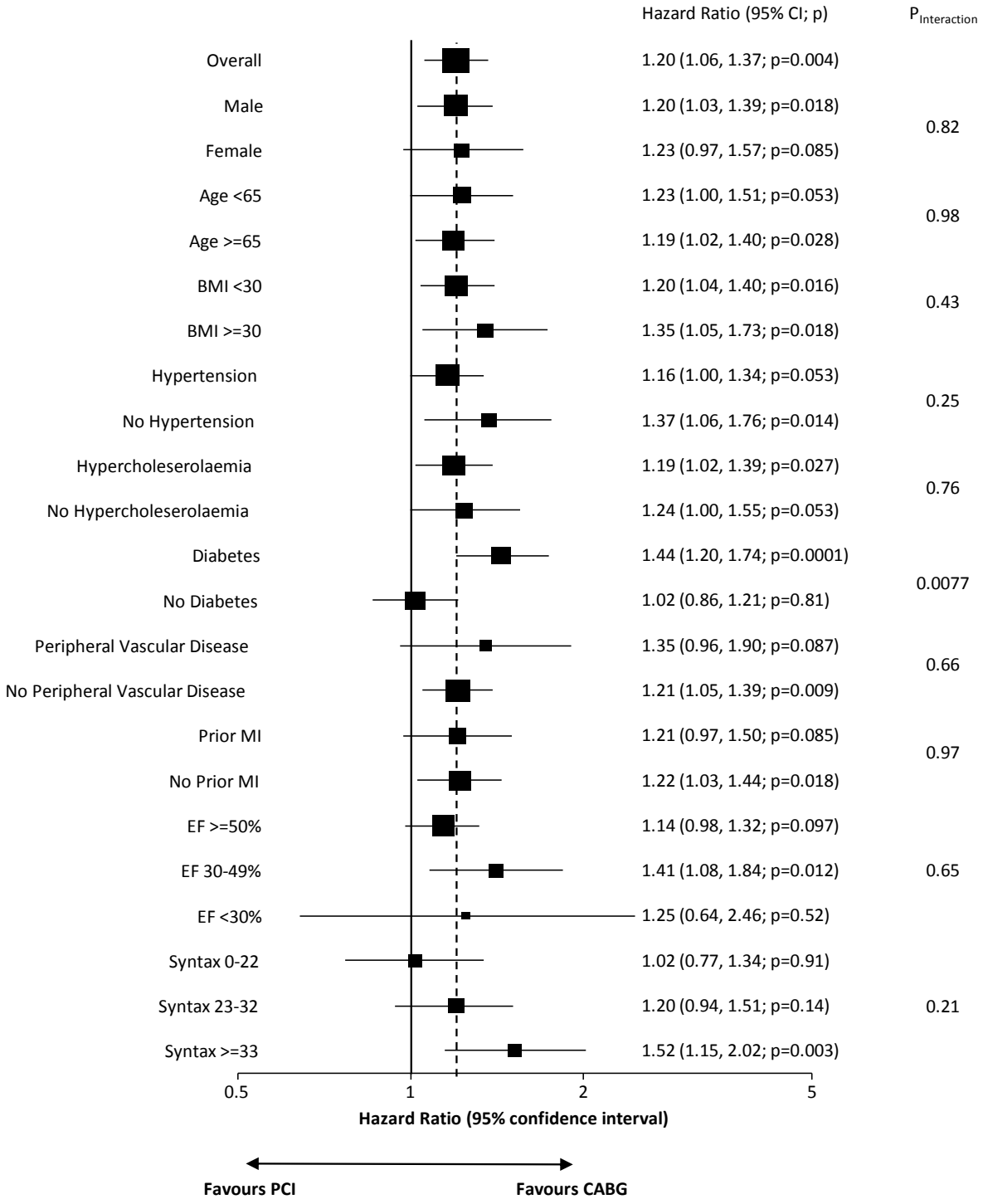
The limitations of the current evidence however, remains in two areas.

First, the end point of Head et al ‘s study is all-cause mortality, we also know that complications which may affect the quality of life are important determinants in the decision making for the individual patient’s treatment.

Second, the number of events upon which these analyses are based, remains low, and this invites caution. More accurate follow-up data is required in order to utilise the very valuable information presented in Head et al study’s with more scientific weight.

To conclude, based on the current evidence CABG achieves better results than PCI in the treatment of coronary disease independently from SYNTAX score and presence of left main or multivessel disease. While longer term data are needed, at the moment surgery should be considered the best therapeutic option in these settings among patients considered candidates for either procedure.

Figure 1. Mortality in Randomised Trials of PCI versus CABG after 5 years follow up, main effects and subgroups. Hazard ratios, 95% confidence intervals and interaction tests.



Note, the size of the marker is proportionate to the number of events in the analysis.

Adapted from Head et al 2018[2]

References

1. Freemantle N, Calvert M, Wood J, Eastaugh J, Griffin C. Composite outcomes in randomized trials – greater precision but with greater uncertainty? Journal of the American Medical Association, 2003; 289: 2554-9.
2. Head SJ, Milojevic M, Daemen J, Ahn JM, Boersma E, Christiansen EH et al. Mortality after coronary artery bypass grafting versus percutaneous coronary intervention with stenting for coronary artery disease: a pooled analysis of individual patient data. Lancet February 22, 2018 http://dx.doi.org/10.1016/S0140-6736(18)30423-9
3. Chang M, Ahn JM, Lee CW, Cavalcante R, Sotomi Y, Onuma Y, et al. Long-term mortality after coronary revascularization in nondiabetic patients with multivessel disease. J Am Coll Cardiol 2016;68:29–36
4. Windecker S, Stortecky S, Stefanini GG, daCosta BR, Rutjes AW, Di Nisio M, et al. Revascularisation versus medical treatment in patients with stable coronary artery disease: network meta-analysis. BMJ 2014;348:g3859 doi: 10.1136/bmj.g3859 (Published 23 June 2014)
5. Yusuf S, Wittes J, Probstfield J, Tyroler HA. Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials. JAMA. 1991; 266: 93-8
6. Freemantle N. Interpreting the results of secondary endpoints and subgroup analyses in clinical trials: should we lock the crazy aunt in the attic? British Medical Journal, 2001; 322: 989-91.
7. Sun X, Briel M, Walter SD, Guyatt GH. Is a subgroup effect believable? Updating criteria to evaluate the credibility of subgroup analyses. BMJ 2010;340:c117
8. ISIS-2 (Second International Study of Infarct Survival) Collaborative Group. Randomised trial of intravenous streptokinase, oral aspirin, both, or neither among 17 187 cases of suspected acute myocardial infarction: ISIS-2. Lancet 1988; 332: 349–60.
9. Kwon JS, Kim YS, Cho As et al, Origin of restenosis after drug eluting stent implantation in hyperglycemia is inflammatory cells and thrombus J Atheroscleror Thomb, 2011; 18 604-15
10. 2014 ESC/EACTS guidelines on myocardial Revascularisation. Kolh P, Windecker S, Alfonso F et al. European Journal of Cardio-Thoracic Surgery, Volume 46, Issue 4, 1 October 2014, Pages 517–592, https://doi.org/10.1093/ejcts/ezu366
11. Milojevic M, Head SJ, Parasca CA, Serruys PW, MD, Mohr FW, MD, Morice MC, 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