Covariate adjustment in cardiovascular randomised controlled trials: its value, current practice and need for improvement

Running title: Covariate adjustment in cardiovascular randomised controlled trials

Leah Pirondini MSc [a], John Gregson PhD [a], Ruth Owen MSc [a] , Tim Collier MSc [a], Stuart Pocock PhD [a]

[a] *Department of Medical Statistics, London School of Hygiene and Tropical Medicine*

Word Count: 5,043 (From Introduction, including references)

Disclosures: John Gregson has received consultancy fees from Boehringer-Ingelheim.

Address for correspondence

Professor Stuart Pocock
Department of Medical Statistics
London School of Hygiene and Tropical Medicine
Keppel Street
London
WC1E 7HT
United Kingdom
Stuart.pocock@lshtm.ac.uk

Twitter: @pocock_stuart
Tweet: Covariate adjustment in cardiovascular randomised controlled trials: its value, current practice and need for improvement (121 characters with spaces)

Abstract

In randomised controlled trials, patient characteristics are expected to be well balanced between treatment groups. However, adjustment for characteristics which are prognostic can still be beneficial with a modest gain in statistical power. Nevertheless, previous reviews show that many trials use unadjusted analyses. In this article, we review current practice regarding covariate adjustment in cardiovascular trials among all 84 randomised controlled trials relating to cardiovascular disease published in the *New England Journal of Medicine*, *The Lancet*, and the *Journal of the American Medical Association* during 2019. We identify trials where use of covariate adjustment led to a change in the trial conclusions. By using these trials as case studies, along with data from the CHARM trial and simulation studies we demonstrate some of the potential benefits and pitfalls of covariate adjustment. We discuss some of the complexities of using covariate adjustment, including how many covariates to choose, how covariates should be modelled, how to handle missing data for baseline covariates, and how adjusted analyses are viewed by regulators. We conclude that contemporary cardiovascular trials do not make best use of covariate adjustment and that more frequent use could lead to improvements in the efficiency of future trials.

Keywords: Statistics, randomized controlled trials, covariate adjustment

Abbreviations

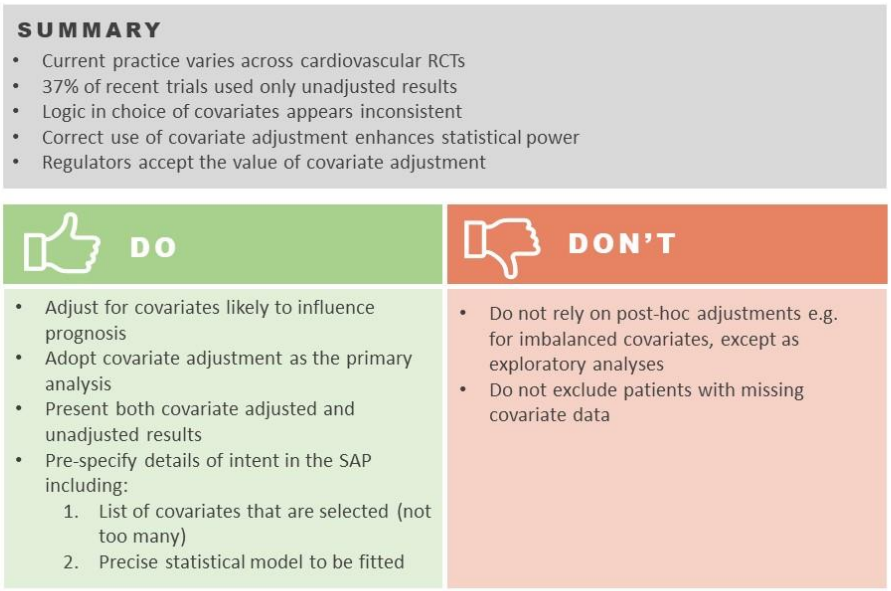
RCT	Randomized controlled trial
NT-proBNP	N-terminal pro b-type Natriuretic Peptide
EMA	European Medicines Agency
US FDA	United States Food and Drug Administration
SAP	Statistical analysis plan
Trial acronyms	
CHARM	Candesartan in Heart failure: Assessment of Reduction in Mortality and Morbidity
EXTEND	Extending the Time for Thrombolysis in Emergency Neurological Deficits
SYNTAXES	Synergy between PCI with Taxus and Cardiac Surgery Extended Survival
REGROUP	Randomized Endovein Graft Prospective
ARTEMIS	Affordability and Real-World Antiplatelet Treatment Effectiveness After Myocardial Infarction Study

Highlights

- Too many contemporary cardiovascular trials do not use covariate adjustment in the primary analysis
- Adjustment for a limited number of prognostic covariates is simple, has few risks and is viewed as appropriate by regulators
- Covariates used for adjustment should be pre-specified prior to unblinding
- Adjustment for prognostic covariates can offer a meaningful gain in statistical power

Central illustration

The value of covariate adjustment in clinical trials



Introduction

Randomised controlled trials (RCTs) are considered the gold-standard for assessing the effect of a new treatment. A major reason for this is that randomisation ensures that patients receiving the treatment or control do not differ systematically with respect to their characteristics. It therefore follows that statistical adjustment for baseline covariates is not necessary to obtain an unbiased comparison between treatment groups. However, such adjustment for baseline covariates can often be beneficial, as it tends to result in a gain in statistical power [1] [2] [3] [4]. Although the theoretical benefit of covariate adjustment is well established within the statistical community [5] [6], previous surveys conducted between 2000-2013 showed that a substantial proportion of major RCTs do not use covariate adjustment [7] [8] [9].

In this article we aim to review how covariate adjustment is being used in contemporary major RCTs by performing a survey of clinical trials published in major medical journals in 2019. We then illustrate some of the benefits and difficulties of using covariate adjustment. To do this we use case studies identified from our survey, individual patient data from the CHARM (Candesartan in Heart failure: Assessment of Reduction in Mortality and morbidity) trial, and simulation studies.

Survey of cardiovascular trials

We surveyed all randomised controlled trials relating to cardiovascular disease published in *New England Journal of Medicine*, *The Lancet*, and *Journal of the American Medical Association (JAMA)* during 2019 using a standard proforma. We collected information on the study design, type of covariate adjustment used and number of covariates adjusted for using a standardised form. All studies were assessed by two authors (LP, RO).

We identified 84 trial reports in *New England Journal of Medicine* (n=41), *The Lancet* (n=23) and *JAMA* (n=20). The majority of trials (81 out of 84) used individual patient randomisation; the remaining three were cluster randomised trials. 61 were superiority trials, 18 were non-inferiority trials, and the remaining 5 examined both. 52 trials performed time-to-event analyses using Cox proportional hazards models. Other common outcome types were binary (n=16) and continuous (n=14), which were typically analysed using logistic or linear regression, respectively. The most common primary outcomes were a composite of major cardiac events (n=42) or mortality (n=7).

Of the 84 trials surveyed, 31 performed only a simple treatment comparison, unadjusted for baseline covariates, whereas 53 used covariate adjustment as part of their primary or secondary analyses (table 1). Of these 53 trials, 39 gave primary (or equal) emphasis to the adjusted results. Covariate adjustment was performed using either multivariable or stratified regression modelling. The number of covariates included in the adjusted analysis varied, with a median of two and a maximum of 13. Reasons for the choice of covariates were not always clearly explained but common themes were adjusting for variables used in the randomisation procedure, covariates expected to be predictive of the outcome, adjusting for the baseline value of a continuous outcome, and, to a lesser extent, covariates imbalanced between groups. 66 of the 84 trials reported using some form of stratified randomisation; of these 28 stratified by centre only. Of the 66 trials using stratified randomisation, 51 (77%) performed a covariate-adjusted analysis which adjusted for all (or all but centre) of the covariates used in the stratified randomisation (one trial was unclear). Of 53 trials where covariate-adjustment was performed, 22 adjusted exclusively for variables used in the stratified randomisation. A further 8 trials adjusted exclusively for the baseline value of a continuous outcome, or a combination of the baseline value of a continuous outcome and the variables used in stratified randomisation.

Of 22 trials reporting both unadjusted and adjusted results, there was a smaller p-value for 16 trials with covariate adjustment and 6 trials with unadjusted analyses (**Supplementary Table 1**). In two trials the adjustment affected whether the result was significant at the conventional $p<0.05$ threshold, with both examples showing a significant result only after covariate adjustment.

Case studies: We now present two examples of trials where covariate adjustment altered the significance of the treatment effect and two examples of trials which adjusted for a large number of covariates.

The Extending the Time for Thrombolysis in Emergency Neurological Deficits (EXTEND) trial [10] randomly assigned 225 patients to intravenous alteplase or placebo between 4.5 and 9 hours after the onset of stroke. The primary outcome was a binary 90-day score of 0 or 1 on the modified Rankin scale (mRS). The pre-specified primary analysis used a modified Poisson regression with robust error estimation [11], with age and clinical severity of stroke (National Institute of Stroke Severity (NIHSS) score) at baseline as covariates, with an equivalent unadjusted analysis also performed. The authors prespecified these covariates because they were expected to be predictive of the outcome. By chance, the mean age in the alteplase group was higher than in the placebo group (73.7 vs 71.0 years) and the median NIHSS score was also higher (12.0 vs 10.0). Using an unadjusted analysis, the proportion of patients with an mRS score of 0 or 1 was 1.20 times higher in the treatment group than in the placebo group (95% CI: 0.82-1.76, p=0.35), but in the pre-specified adjusted primary analysis this ratio was statistically significant: 1.44 (95% CI: 1.01—2.06, p=0.04). This trial provides an example where appropriate use of covariate adjustment made a marked difference to the results and trial conclusions. The particularly strong impact of covariate adjustment in this instance occurred because the covariates were both strongly predictive and were, by chance, imbalanced such that, on average, patients in the alteplase group had a worse prognosis at baseline.

The Synergy between PCI with Taxus and Cardiac Surgery Extended Survival (SYNTAXES) study [12] is a 10-year follow-up of the SYNTAX trial which compared percutaneous coronary intervention (PCI) vs coronary artery bypass graft (CABG) in 1,800 patients with three-vessel or left-main coronary artery disease. The primary outcome was all-cause death at 10 years and was analysed using Cox proportional hazards models in 1,689 patients with available data. The unadjusted results showed no significant effect: HR for PCI vs. CABG of 1.17 (95% CI: 0.97-1.41, p=0.092). An exploratory sensitivity analysis was performed using a multivariable Cox model with adjustment for 8 prognostic covariates identified using forward stepwise selection. It is unclear whether or not these covariates were selected using treatment-blinded data. Following adjustment the results were statistically significant (HR 1.23, 95% CI: 1.02-1.48, p=0.028) but as they were not pre-specified they were not

emphasized by the investigators. We consider in more detail how to pre-specify covariate adjusted analyses in the discussion section.

The Randomized Endovein Graft Prospective (REGROUP) trial [13], which randomly assigned 1150 patients undergoing CABG to either open or endoscopic vein-graft harvesting provides an example of the practical difficulty caused by missing data in covariate adjusted analyses. The primary outcome was the first occurrence of a major adverse cardiac event (a composite of death from any cause, non-fatal myocardial infarction and repeat revascularisation). The primary analysis used the log-rank test (unadjusted), and a sensitivity analysis used a multivariable Cox proportional-hazards model, adjusted for other potentially influential baseline characteristics. Neither the unadjusted analysis (HR 1.12, 95% CI 0.83-1.51) nor the adjusted analysis (HR 1.07, 95% CI 0.71-1.62) showed a significant difference between the two treatments. However, the adjusted results were based on a model including complete records only (n=714), thereby excluding over a third of patients with missing data for any of the 13 covariates (n=436), which reduces statistical power and may affect the generalisability of results. We consider how to handle missing data on covariates in the Discussion.

The Affordability and Real-World Antiplatelet Treatment Effectiveness After Myocardial Infarction Study (ARTEMIS) [14] was a cluster randomised trial carried out in the United States which investigated the effect of medication co-payment vouchers on the co-primary endpoints of persistence (continuation of treatment for the prescribed duration) of P2Y12 inhibitors and MACE at 1 year. In cluster randomised trials the unit of randomisation is the entire cluster (each hospital was one cluster in ARTEMIS), and because there are fewer clusters than patients there is less opportunity for baseline characteristics to “balance out” between groups. Some baseline differences were therefore expected between treatment groups and the primary model was adjusted for 9 patient and hospital characteristics (including a propensity score containing 50 covariates prospectively selected based on their clinical relevance). Co-payment vouchers significantly improved persistence of P2Y12 inhibitors, although adjustment for patient and hospital characteristics decreased the strength of the effect (unadjusted OR 1.31 [95% CI 1.12-1.54, p=0.001], adjusted OR 1.19 [95% CI 1.02-1.40. p=0.03]). There was no statistically significant effect on the MACE outcome before or after adjustment (unadjusted HR 0.96 [95% CI 0.80-1.15, p=0.65], adjusted HR 1.07 [95% CI 0.93-1.25. p=0.35]).

Adjusted and unadjusted analyses of the CHARM trial

To demonstrate in more detail the impact of covariate adjustment in the context of a large cardiovascular trial, we used individual patient data from the CHARM trial.

CHARM was a programme of 3 clinical trials designed to examine the effect of candesartan on mortality and heart failure hospitalisation in patients with chronic heart failure (CHF) [15]. The CHARM-Overall programme contains the data from all 3 trials and included 3803 patients randomised to candesartan and 3796 patients randomised to placebo. Patients were followed for a median of 37.7 months and the primary outcome was all-cause mortality, which occurred in 886 (23%) patients in the candesartan group and 945 (25%) patients in the placebo group. Published results give an unadjusted hazard ratio of 0.91 [95% CI 0.83-1.00], p=0.055; and a covariate adjusted hazard ratio of 0.90 [95% CI 0.82-0.99], p=0.032. Unadjusted results were calculated using a Cox proportional hazards model, with stratification only by component trial. Covariate-adjusted results were reported to include all baseline covariates presented in Table 1 of the published report referenced above. Here we explore the impact of covariate adjustment (i) adjusting for each single covariate (one-at-a-time) and (ii) adjusting for multiple covariates via a stepwise addition of covariates in order of their predictive strength (from strongest to weakest). This analysis is intended to demonstrate the effect of covariate adjustment. We do not propose that one needs to adjust for as many covariates as are given here, nor that they should be included based on their predictive strength in univariate analyses.

Many baseline covariates were significantly associated with all-cause mortality (Supplementary Table 2), and the vast majority were well balanced between the treatment groups (Supplementary table 3). Table 2 provides hazard ratios and p-values before and after covariate adjustment, and the results are largely consistent with the expected effects of covariate adjustment. Overall, adjustment for prognostic risk factors tended to increase statistical power. This is reflected by the gradual reduction in p-value with stepwise addition of covariates, with p=0.055 in an unadjusted analysis to p=0.017 in an analysis containing all covariates. Adjustment for the most highly prognostic covariates had the greatest impact on the results. This can most easily be seen by comparing results when adjusting for only one covariate at a time. For 6 of the 10 most prognostic covariates (at the top left side of Table 2) the p-value after adjustment was less than 0.05, whereas this was the case for only 6 of the remaining 23 covariates. However, our analysis also shows how, for any particular covariate, the impact of covariate adjustment depends not only on its prognostic strength, but also on any slight difference in covariate values between treatment groups. As an example, age is the most highly prognostic covariate, but adjustment for age increases the p-value from 0.055 to 0.07.

This occurs because the average age is (non-significantly) higher in the control group, and therefore on the basis of age alone we would expect death to occur slightly less often in treated patients (i.e. making the observed lower death rate less surprising under the null hypothesis of no treatment effect). In contrast, there are slightly more patients with diabetes in the treated group, and so adjustment for diabetes reduces the p-value (to 0.036). In general, adjustment for the most highly prognostic covariates has the greatest impact on the reduction in p-value, however, for any particular covariate, slight chance imbalances between treatment groups also play a role.

Simulation study

To gain a greater understanding of the potential gain in statistical power from covariate adjustment in cardiovascular trials, we performed simulation studies to mimic realistic clinical trials. We simulated 1:1 RCTs of either 2,000 or 8,000 patients. A treatment effect of hazard ratio 0.75 was chosen for the N=2,000 trial and 0.85 for the N=8,000 trial to represent realistic assumptions for moderate and large heart failure trials. We simulated a single continuous normally distributed prognostic covariate which was associated with the primary outcome with hazard ratios ranging from 1.25 to 2.25 per 1-SD increase in the covariate. Survival times were generated from an exponential distribution with a rate chosen to achieve a target power of approximately 80% in the N=2,000 trial and 90% in the N=8,000 trial (see Statistical Appendix for further details). For simplicity we assumed a fixed follow-up time of one year with no early censoring. For each treatment and covariate effect size, we generated 10,000 datasets, and for each dataset we used adjusted and unadjusted Cox proportional hazards models to calculate hazard ratios and their 95% confidence intervals and p-values. For each scenario we report the mean hazard ratio and 95% confidence interval (averaged across all simulations), statistical power (the proportion of simulations with a significant treatment effect at $p<0.05$), and the equivalent effective increase in sample size corresponding to the gain in power (see Statistical Appendix for details).

Table 3A shows simulation results for the 2,000 patient trial where approximately 20% of control patients have the primary outcome and treatment is associated with a hazard ratio of 0.75. In each scenario, the adjusted analysis was more statistically powerful than the unadjusted analysis, and the gain in statistical power was larger when the prognostic covariate was more strongly associated with outcome. The gains in statistical power from using covariate adjustment in the 2,000 patient trial were 0.5%, 1.4%, 3.0%, 4.9% and 7.0% for the covariate's standardised hazard ratios of 1.25, 1.5, 1.75, 2.0 and 2.25, respectively. Another way to express this is to instead consider how many more patients we would need to recruit in a trial not using covariate adjustment to achieve the same statistical power as a trial using covariate adjustment. This 'effective increase in sample size' was 1.0%, 3.6%, 7.5%, 12.7% and 19.1% for the covariate's standardised hazard ratios of 1.25, 1.5, 1.75, 2.0 and 2.25, respectively. Table 3B shows very similar results for the simulated trials of 8,000 patients.

Discussion

Our survey, case studies and simulations studies demonstrate the potential benefit that can be achieved by adjusting for prognostic covariates in cardiovascular trials.

Our survey found that over half of trials published in 2019 did not use covariate adjustment in their main analysis of the primary outcome and fewer than 10% of studies made it clear that the basis for adjustment was that the covariate was expected to be prognostic. The value of covariate adjustment can be seen in the EXTEND study where unadjusted results were by some margin not statistically significant, whereas they became significant in the pre-specified primary analyses which adjusted for baseline disease severity and age. Likewise results from CHARM-Overall were statistically significant following covariate adjustment whereas unadjusted results were not. HF-ACTION is another analogous example of a cardiovascular trial where covariate adjustment altered the trial conclusions [16]. Our results from CHARM also show how adjustment for the covariates most strongly prognostic of the outcome has the greatest impact on the results.

Previous studies have demonstrated the benefit of adjustment for prognostic covariates on statistical power for trials using continuous, binary and time-to-event outcomes [5] [17] [3]. Our simulation study presents a simple example tailored to a typical cardiovascular randomised trial, and is consistent with previous work. For continuous outcomes an increase in statistical power occurs because adjustment reduces the standard error of the estimated treatment effect, whereas for binary and time-to-event outcomes the increase in power occurs due to an increase in the estimated treatment effect (i.e. odds ratios or hazard ratios further from the null value of 1) with a very slight increase in the standard error of the estimated treatment effect, leading overall to a smaller p-value. We found the benefits in terms of statistical power can be substantial when strong prognostic markers exist, as is often the case in cardiovascular trials. For example in recent trials of SGLT2 inhibitors, each 1-SD higher NT-proBNP is associated with a hazard ratio for cardiovascular death or heart failure hospitalisation of roughly 1.75 [18] [19]. Inferring from our simulation results, statistical adjustment for NT-proBNP in trials with 80-90% power would therefore be expected to further increase power by around 3%, which is equivalent to recruiting around 7.5% more patients, i.e. 300 more patients in a trial of 4,000 patients. We also found that the benefit of covariate adjustment was similar in simulated trials of size 2,000 or 8,000. This seems to suggest, at least for reasonably large trials, that the benefits of covariate adjustment are not strongly linked to trial size which is in line with previous research on this topic [5].

Recently, the European Medicines Agency (EMA) [20] and US FDA [21] have both published draft guidance regarding the use of covariate adjustment in clinical trials. Both regulators emphasise the

need to pre-specify unambiguously which covariates will be included in the statistical model and what their form will be (e.g. linear, non-linear forms such as log-transformed, or categorical) in order to maintain credibility and to control the risk of false positive results. This information should be recorded in the SAP which should be registered publicly at sites such as clinicaltrials.gov. They also recommend avoiding analyses which either use post-hoc adjustment to correct for chance imbalances in baseline covariates, or analyses which adjust for covariates measured after randomisation as these may be influenced by the treatment. The EMA recommends that covariates that are used for stratified randomisation should also be adjusted for unless the strata were chosen purely for "administrative purposes". Factors used for stratification are often chosen to balance out characteristics which are expected to be strongly related to outcome, so failure to adjust for such factors may lead to analyses which are unnecessarily conservative [22]. There is also guidance about the number of covariates to adjust for. The EMA suggests only adjusting for a few covariates, whereas the FDA simply states that the number of covariates should be "small relative to the size of the trial". Our results show that it is adjustment for the most highly prognostic covariates which matters most, and these are often small in number. In our experience, there would be little gain in adjusting for more than a handful of strongly prognostic covariates in the majority of cardiovascular trials.

It is important to consider the additional complexities caused by using covariate adjustment, although it is notable that these are usually easy to handle. The sensitivity analysis of the REGROUP trial shows that missing data has the potential to complicate the use of covariate adjustment. Missing data in covariates can be minimised by pre-specifying adjustment only for covariates where missing data is expected to be rare. However, in large trials some missing data is inevitable so a clear plan of how to handle missing data is needed. Patients with missing data on baseline covariates should not be excluded from intention-to-treat analyses [20], and covariates with missing data should not be removed from the statistical model unless the extent of missingness is substantial, and this has been clearly pre-specified in the SAP. Instead, one can perform imputation in order to incorporate such patients and/or covariates. Very simple strategies (e.g. imputing the mean value or using the missing-indicator method [23]) have been shown to yield unbiased estimates for randomised comparisons [24] [25] [26]. This makes missing data less problematic than in observational studies where a more sophisticated approach may be required. A further issue is how to choose the correct form of adjustment (e.g. linear, log-transformed or categorical). To gain maximum benefit from covariate adjustment, one needs to choose the form which best represents the true association between the covariate and outcome. For strongly prognostic covariates the correct form may be well-established from previous studies. However, even if the wrong form is

chosen or the covariate turns out not to be prognostic, for large trials this does not cause bias or inflation of the risk of Type I error [27]. As it is difficult to know in advance the impact of covariate adjustment on statistical power, and hence sample size, we do not recommend amending sample size calculations, but rather allowing covariate adjustment to provide a boost to detect a true treatment effect.

In our example using CHARM we demonstrated inclusion of covariates using a stepwise addition of covariates based on their predictive strength in the trial database. Although we are unaware of any reason to suspect this approach would be unreliable in similar trials, it is difficult to pre-specify exactly the form of each covariate and one would need to wait for events to occur in order to know which covariates to fit. Given that any very strongly prognostic covariates are usually known or suspected prior to enrolment, stepwise addition of covariates would be unlikely to offer any meaningful benefit to statistical power beyond our favoured approach, which is to select covariates based on pre-existing clinical knowledge.

Our study has limitations. Our survey was limited to cardiovascular trials published in three medical journals and therefore may not be representative of all trials in a broader class of diseases and journals. We demonstrated the detailed effects of covariate adjustment using only one trial example (CHARM), but the principals are generic. Our simulation study was limited to a simple example using one continuous covariate, and a time-to-event outcome with fixed follow-up. However, previous studies have reached broadly similar conclusions when using continuous or binary outcomes. Previous studies have also suggested that non-informative censoring does not materially affect the impact of covariate adjustment on statistical power for time-to-event outcomes. Our study is limited to investigating the impact of prognostic covariates. Previous studies have investigated adjustment for imbalances in baseline covariates, but we avoided such comparisons because such analyses are difficult to pre-specify and tend not to have an important impact on statistical power when covariates are not prognostic. Our simulations and case studies examined the impact of covariate adjustment in superiority trials. For non-inferiority trials covariate adjustment is more technically challenging when margins are defined on an absolute scale. The benefit of covariate adjustment on statistical power is also unlikely to transfer well to non-inferiority margins if the margin is defined on a relative scale, because of the tendency of adjusted estimates to be both further from the null and have slightly larger standard errors.

Conclusions: Covariate adjustment for prognostic covariates is simple to perform and can result in meaningful benefits in terms of statistical power. Despite this it is often not used in contemporary

cardiovascular RCTs. More widespread use of covariate adjustment could lead to more efficient and/or powerful RCTs in the future.

References

[1] Steyerberg E W, Bossuyt P M, Lee K L, Clinical trials in acute myocardial infarction: should we adjust for baseline characteristics?, *Am. Heart J.*, 2000; 139:745-51.

[2] Turner E L, Perel P, Clayton T, *et al.*, Covariate adjustment increased power in randomized controlled trials: an example in traumatic brain injury, *J. Clin. Epidemiol.*, 2012; 65:474-481.

[3] Kahan B C, Jairath V, Doré C J, Morris T P, The risks and rewards of covariate adjustment in randomized trials: An assessment of 12 outcomes from 8 studies, *Trials*, 2014; 15:1-7.

[4] Pocock S J, McMurray J J V, and Collier T J, 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

[5] Hernández A V, Steyerberg E W, Habbema J D F, Covariate adjustment in randomized controlled trials with dichotomous outcomes increases statistical power and reduces sample size requirements, *J. Clin. Epidemiol.*, 2004; 57:454-460.

[6] Pocock S J, Assmann S E, Enos L E, Kasten L E, Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: Current practice and problems, *Stat. Med.*, 2002; 21: 2917-30.

[7] Assmann S F, Pocock S J, Enos L E, Kasten L E, Subgroup analysis and other (mis)uses of baseline data in clinical trials, *Lancet*, 2000; 355:1064-69.

[8] Austin P C, Manca A, Zwarenstein M, Juurlink D N, Stanbrook M, A substantial and confusing variation exists in handling of baseline covariates in randomized controlled trials: a review of trials published in leading medical journals, *J. Clin. Epidemiol.*, 2010; 63:142-153.

[9] Saquib N, Saquib J, Ioannidis P A, Practices and impact of primary outcome adjustment in randomized controlled trials: Meta-epidemiologic study, *BMJ*, 2013; 347:1-12.

[10] Ma H, Campbell B C V, Parsons M W, *et al.*, Thrombolysis Guided by Perfusion Imaging up to 9 Hours after Onset of Stroke, *N. Engl. J. Med.*, 2019; 380:1795-1803.

[11] Zou G, A modified poisson regression approach to prospective studies with binary data, *Am. J. Epidemiol.*, 2004; 159:702-706.

[12] Thuijs D J F M, Kappetein A P, Serruys P W, *et al.*, Percutaneous coronary intervention versus coronary artery bypass grafting in patients with three-vessel or left main coronary artery disease: 10-year follow-up of the multicentre randomised controlled SYNTAX trial, *Lancet*, 2019; 394:1325-34.

[13] Zenati M A, Bhatt D L, Bakaeen F G *et al.*, Randomized Trial of Endoscopic or Open Vein-Graft Harvesting for Coronary-Artery Bypass, *N. Engl. J. Med.*, 2019; 380:132-141.

[14] Wang T Y, Kaltenbach L A, Cannon C P, *et al.*, Effect of Medication Co-payment Vouchers on P2Y12 Inhibitor Use and Major Adverse Cardiovascular Events among Patients with Myocardial Infarction: The ARTEMIS Randomized Clinical Trial, *JAMA - J. Am. Med. Assoc.*, 2019; 321:44-55. [15] Pfeffer M A, Swedberg K, Granger C B, *et al.*, Effects of candesartan on mortality and morbidity in patients with chronic heart failure: The CHARM-overall programme, *Lancet*, 2003; 362:759-766.

[16] O'Connor C M *et al.*, Efficacy and Safety of Exercise Training in Patients With Chronic Heart Failure: HF-ACTION Randomized Controlled Trial, *JAMA*, 2009; 301:1439-50.

[17] Thompson D D, Lingsma H F, Whiteley W N, Murray G D, Steyerberg E W, Covariate adjustment had similar benefits in small and large randomised controlled trials, *J. Clin. Epidemiol.*, 2015; 68:1068-75.

[18] Packer M, Anker S D, Butler J, *et al.*, Cardiovascular and Renal Outcomes with Empagliflozin in Heart Failure, *N. Engl. J. Med.*, 2020; 383:1413-24. [19] Anker S D, Butler J, Filippatos G, *et al.*, Effect of Empagliflozin on Cardiovascular and Renal Outcomes in Patients With Heart Failure by Baseline Diabetes Status: Results From the EMPEROR-Reduced Trial, *Circulation*, 2021; 143:337-349.

[20] European Medicines Agency, "Committee for Medicinal Products for Human Use (CHMP) Guideline on adjustment for baseline covariates in clinical trials," 2015. https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-adjustment-baseline-covariates-clinical-trials_en.pdf.

[21] Food and Drug Administration, "Adjusting for Covariates in Randomized Clinical Trials for Drugs and Biologics with Continuous Outcomes Guidance for Industry," 2021. https://www.fda.gov/media/148910/download.

[22] Kahan B C, Morris T P, Reporting and analysis of trials using stratified randomisation in leading medical journals: Review and reanalysis, *BMJ*, 2012; 345:1-8.

[23] Groenwold R H H, White I R, Donders A R T *et al.*, Missing covariate data in clinical research: when and when not to use the missing-indicator method for analysis, *C. Can. Med. Assoc. J.*, 2012; 184:1265.

[24] White I R, Thompson S G, Adjusting for partially missing baseline measurements in randomized trials, *Stat. Med.*, 2005; 24:993-1007. [25] Schemper M and Smith T L, Efficient evaluation of treatment effects in the presence of missing covariate values, *Stat. Med.*, 1990; 9: 777-784.

[26] Sullivan T R, White I R, Salter A B, Ryan P, and Lee K J, Should multiple imputation be the method of choice for handling missing data in randomized trials?, *Stat. Methods Med. Res.*, 2018; 27: 2610-26.

[27] Wang B, Ogburn E L, Rosenblum M, Analysis of covariance in randomized trials: More precision and valid confidence intervals, without model assumptions, *Biometrics*, 2019; 75:1391-1400.

[28] Bender R, Augustin T, Blettner M, Generating survival times to simulate Cox proportional hazards models, *Stat. Med.*, 2005; 24:1713-23.

Statistical appendix

Simulations: For each patient, an uncensored event times (measured in years), t, was randomly generated from an Exponential model using the following formula [28]:

$$t = -\frac{\log(u)}{\lambda} \exp\left[-(\beta_x x + \beta_z z)\right]$$

where:

- u are random numbers generated from a Uniform distribution in the range [0,1]
- β_x is the true treatment coefficient (log-hazard ratio)
- x is the treatment indicator, taking values of 0 for the placebo group and 1 for the active treatment group, randomly generated with a 50% probability of being in the treatment or placebo arm
- β_z is the true covariate coefficient (log-hazard ratio)
- z is the covariate value, randomly generated from a standard normal distribution N(0,1)
- λ is the rate parameter for the Exponential distribution. The rate parameter λ was set so that, for a patient in the control arm (x = 0) with covariate value of zero (z = 0), the probability of an event within 1 year was 20% . To do this we set the rate parameter to λ=0.223, based on the relationship under an exponential survival model: $S_0(t) = e^{-\lambda t}$. The actual number of simulated events in the control arm will vary depending on the covariate effect size, with larger effect sizes leading to a greater number of high risk simulated patients, and hence more events.

We considered administrative censoring only, patients with a survival time t, greater than 1 year were considered to be censored at 1 year (the maximum follow-up time).

Effective increase in sample size: We define the number of patients required for the unadjusted and adjusted analyses as n_u and n_a, respectively. These sample sizes can be expressed in terms of the type I error rate, α, target power, β, baseline event proportion, and estimated treatment hazard ratio. From our simulation results, we define Z_u as the mean standardised unadjusted log-treatment effect (i.e. $logHR/SE$ averaged over all simulations) and Z_a as the equivalent mean standardised adjusted log-treatment effect. Therefore, rewriting the sample size formulae n_u and n_a in terms of these standardised treatment effects, for the same target power, we obtain the approximate result:

$$\frac{n_a}{n_u} \approx \left(\frac{Z_u}{Z_a}\right)^2$$

Table 1: Use of covariate adjustment in cardiovascular trials published in three major medical journals in 2019

	No. trials
Were primary outcome analyses done with covariate adjustment?	
No	31
Yes	53
Which analyses received more emphasis?	
Adjusted	38
Unadjusted	14
Equal emphasis	1
Only unadjusted done	31
Did covariate adjustment alter the trial conclusions compared with unadjusted analyses?	
Yes	2
No	22
Only adjusted given	29
Number of covariates included	
1	18
2	11
3	9
4	3
5 - 7	6
8 – 10*	3
11 – 13	2
Not clear	1
Reasons for choice of covariates†	
Covariates used in stratified randomisation	24
Centre or country adjusted for	9
Covariates were (or expected to be) prognostic	8
Baseline value of quantitative outcome	8‡
Covariates imbalanced between groups	4
Other treatment factor in a factorial design	2
No reason given	8

*One trial used 9 covariates where 1 covariate was a propensity score incorporating 50 variables
†More than one reason in some trials
‡Out of a total of 14 trials where outcome was continuous/quantitative

Table 2: Re-analysis of the CHARM-Overall trial

Hazard ratios, 95% CIs and p-values for all-cause mortality in the CHARM-Overall trial using (a) models adjusting for a single covariate and (b) models adjusting for the stepwise addition of covariates. The table is ordered from the most strongly prognostic covariates at the top to the least strongly prognostic at the bottom, as defined by the size of likelihood ratio test statistic for inclusion. Stepwise addition of covariates includes all covariates on that row of the table and above (e.g. for NYHA, it includes age, LVEF and NYHA). All Cox proportional hazards models are additionally stratified by component trial.

	(a) Single covariate models			(b) Stepwise adding of covariates		
Covariate	**Hazard ratio (Candesartan vs. placebo)**	**95% CI**	**p-value**	**Hazard ratio (Candesartan vs. placebo)**	**95% CI**	**p-value**
None	-	-	-	0.914	(0.834 , 1.002)	0.055
Age (years)	0.919	(0.838 , 1.007)	0.070	0.919	(0.838 , 1.007)	0.070
Left ventricular ejection fraction (%)	0.914	(0.834 , 1.002)	0.054	0.915	(0.835 , 1.003)	0.057
NYHA class	0.909	(0.829 , 0.996)	0.041	0.916	(0.836 , 1.004)	0.060
Diastolic blood pressure (mmHg)	0.919	(0.838 , 1.007)	0.070	0.911	(0.831 , 0.999)	0.046
Body mass index (kg/m^2)	0.910	(0.830 , 0.997)	0.044	0.908	(0.828 , 0.995)	0.039
Digitalis glycoside	0.911	(0.832 , 0.999)	0.047	0.905	(0.826 , 0.992)	0.033
B-blocker	0.911	(0.831 , 0.998)	0.046	0.899	(0.82 , 0.985)	0.023
Diabetes Mellitus	0.907	(0.827 , 0.994)	0.036	0.903	(0.824 , 0.99)	0.030
Previous hospitalisation for CHF	0.910	(0.831 , 0.998)	0.045	0.902	(0.822 , 0.988)	0.027
Lipid-lowering drug	0.911	(0.832 , 0.999)	0.048	0.902	(0.823 , 0.988)	0.027
Spironolactone	0.913	(0.833 , 1.000)	0.050	0.899	(0.82 , 0.986)	0.024
Percutaneous coronary revascularisation	0.914	(0.834 , 1.002)	0.056	0.893	(0.815 , 0.979)	0.016
Other vasodilators	0.916	(0.836 , 1.004)	0.061	0.899	(0.82 , 0.985)	0.023
Atrial fibrillation	0.914	(0.834 , 1.002)	0.056	0.899	(0.82 , 0.986)	0.024
Systolic blood pressure (mmHg)	0.915	(0.835 , 1.003)	0.058	0.898	(0.819 , 0.984)	0.022
Antiarrhythmic agent	0.913	(0.833 , 1.001)	0.052	0.899	(0.82 , 0.986)	0.023
Previous myocardial infarction	0.907	(0.827 , 0.994)	0.036	0.900	(0.821 , 0.987)	0.025
Pacemaker implanted	0.914	(0.834 , 1.002)	0.055	0.900	(0.821 , 0.987)	0.025
ACE inhibitors	0.915	(0.834 , 1.002)	0.056	0.900	(0.821 , 0.987)	0.025
Heart rate (beats/min)	0.912	(0.832 , 1.000)	0.050	0.900	(0.821 , 0.987)	0.025
Ethnicity	0.913	(0.833 , 1.001)	0.053	0.892	(0.813 , 0.978)	0.015
Oral anticoagulant	0.914	(0.834 , 1.002)	0.054	0.892	(0.814 , 0.978)	0.015
Stroke	0.914	(0.834 , 1.002)	0.054	0.892	(0.814 , 0.979)	0.015
Calcium channel blocker	0.911	(0.832 , 0.999)	0.047	0.892	(0.814 , 0.978)	0.015
Diuretics	0.916	(0.835 , 1.004)	0.059	0.893	(0.814 , 0.979)	0.016
Sex	0.910	(0.830 , 0.997)	0.044	0.893	(0.815 , 0.98)	0.017
Cancer	0.911	(0.831 , 0.998)	0.046	0.894	(0.815 , 0.98)	0.017
Aspirin	0.914	(0.834 , 1.002)	0.054	0.893	(0.815 , 0.98)	0.016
Other anti-platelet agent	0.923	(0.842 , 1.011)	0.086	0.893	(0.814 , 0.979)	0.016
Coronary artery bypass grafting	0.912	(0.832 , 1.000)	0.049	0.894	(0.815 , 0.98)	0.017
Hypertension	0.914	(0.834 , 1.001)	0.054	0.894	(0.815 , 0.981)	0.017
Current smoker	0.912	(0.832 , 0.999)	0.048	0.892	(0.813 , 0.978)	0.015
Implanted cardioverter defibrillator	0.914	(0.834 , 1.002)	0.054	0.892	(0.814 , 0.978)	0.015
Angina Pectoris	0.914	(0.834 , 1.002)	0.054	0.891	(0.813 , 0.977)	0.014

Table 3: Results when using unadjusted or unadjusted analyses in 10,000 simulated clinical databases

A: When simulating trial datasets with 2,000 patients, an event rate of 20% in the control arm, and a hazard ratio for treatment of 0.75*

Hazard ratio per 1-SD increase in prognostic covariate	Adjusted / Unadjusted	Mean estimated treatment effect HR (95% CI)	Power (%)	Effective increase in sample size vs. unadjusted analysis†	Mean no. events (treated)	Mean no. events (control)	Mean no. events (total)
1.25	Unadjusted	0.751 (0.609, 0.925)	77.0%	-	157	203	360
	Adjusted	0.749 (0.608, 0.924)	77.5%	1.0%			
1.50	Unadjusted	0.753 (0.614, 0.924)	77.8%	-	164	211	375
	Adjusted	0.750 (0.611, 0.920)	79.2%	3.6%			
1.75	Unadjusted	0.757 (0.620, 0.925)	78.1%	-	172	221	393
	Adjusted	0.750 (0.614, 0.915)	81.1%	7.5%			
2.00	Unadjusted	0.762 (0.627, 0.926)	77.8%	-	182	231	413
	Adjusted	0.749 (0.617, 0.911)	82.7%	12.7%			
2.25	Unadjusted	0.768 (0.635, 0.929)	77.3%	-	191	241	432
	Adjusted	0.750 (0.620, 0.907)	84.3%	19.1%			

B: When simulating trial datasets with 8,000 patients, with an event rate of 20% in the control arm and a hazard ratio for treatment of 0.85*

Hazard ratio per 1-SD increase in prognostic covariate	Adjusted / Unadjusted	Mean estimated treatment effect HR (95% CI)	Power (%)	Effective increase in sample size vs. unadjusted analysis†	Mean no. events (treated)	Mean no. events (control)	Mean no. events (total)
1.25	Unadjusted	0.851 (0.769, 0.941)	88.1%	-	704	814	1517
	Adjusted	0.850 (0.768, 0.940)	88.3%	1.0%			
1.50	Unadjusted	0.852 (0.772, 0.941)	88.8%	-	733	845	1577
	Adjusted	0.850 (0.770, 0.938)	90.0%	3.7%			
1.75	Unadjusted	0.855 (0.776, 0.942)	88.8%	-	769	883	1653
	Adjusted	0.850 (0.772, 0.936)	91.1%	7.8%			
2.00	Unadjusted	0.859 (0.781, 0.944)	88.9%	-	809	924	1733
	Adjusted	0.850 (0.773, 0.934)	92.3%	13.3%			
2.25	Unadjusted	0.862 (0.786, 0.946)	88.5%	-	848	965	1813
	Adjusted	0.850 (0.775, 0.932)	93.5%	20.0%			

*20% event rate by the end of the simulated study. Event rate applies to patients in the control arm in whom the prognostic covariate is equal to 0 (the mean value). Further details in the Statistical Appendix.
†Effective increase in sample size corresponding to the increase in statistical power. Further details in the Statistical Appendix