HHS Public Access

Author manuscript

J Clin Epidemiol. Author manuscript; available in PMC 2017 November 30.

Published in final edited form as:

J Clin Epidemiol. 2015 September ; 68(9): 1068–1075. doi:10.1016/j.jclinepi.2014.11.001.

Covariate adjustment had similar benefits in small and large randomised controlled trials

Douglas D. Thompson[a], **Hester F. Lingsma**[b], **William N. Whiteley**[a,c], **Gordon D. Murray**[a], and **Ewout W. Steyerberg**[b]

[a]Edinburgh Hub for Trials Methodology Research, Centre for Population Health Sciences University of Edinburgh UK [b]Centre for Medical Decision Sciences, Department of Public Health, Erasmus MC Rotterdam, the Netherlands [c]Centre for Clinical Brain Sciences, University of Edinburgh, UK

Abstract

Objective—Covariate adjustment is a standard statistical approach in the analysis of randomised controlled trials. We aimed to explore whether the benefit of covariate adjustment on statistical significance and power differed between small and large trials, where chance imbalance in prognostic factors necessarily differs.

Study Design and Setting—We studied two large trial datasets (GUSTO-I, N=30,510 and IST, N=18,372) repeatedly drawing random samples (500,000 times) of sizes 300 and 5000 per arm and simulated each primary outcome using the control arms. We empirically determined the treatment effects required to fix power at 80% for all unadjusted analyses and calculated the joint probabilities in the discordant cells when cross-classifying adjusted and unadjusted results from logistic regression models (i.e., $p<0.05$ vs. $p\geq0.05$).

Results—The power gained from an adjusted analysis for small and large samples was between 5–6%. Similar proportions of discordance were noted irrespective of the sample size in both the GUSTO-I and the IST datasets.

Conclusion—The proportions of change in statistical significance from covariate adjustment of strongly prognostic characteristics were the same for small and large trials with similar gains in statistical power. Covariate adjustment is equally recommendable in small and large trials.

Keywords

Simulation; Logistic regression analysis; Covariate Adjustment; Randomised trial

1. Introduction

Randomised clinical trials (RCTs) provide key information for the assessment of treatment efficacy. At its simplest, eligible patients are randomly allocated to receive either an active intervention or a placebo; they are then followed for a specified period of time at the end of which their outcomes are suitably compared. Randomisation ensures that on average observed and unobserved patient characteristics are similar between treatment arms. It is an allocation mechanism with no preferential favouring toward treatment (1). However, it does not ensure full balance. Differences in baseline prognostic factors can arise between the treatment groups simply by chance. In the presence of such imbalance the comparability of the treatment groups may be questioned, especially with regards to any known important prognostic factors (e.g., severity of disease or age). Covariate adjustment can be used to account for such imbalances of known prognostic variables, although, even in the presence of balance, there are still gains to be made with regards statistical power (2–5). The CONSORT (Consolidated Standards of Reporting Trials) statement recommends that the *unadjusted* average treatment effect and the *adjusted* treatment effect be reported (Item No.'s 12b and 18) (6). Despite these recommendations and the known benefits, covariate adjustment is often underutilised in randomised trials (7).

Other things being equal, small trials are subject to a greater chance of imbalance than large trials. It may be anticipated that correcting for such imbalance through covariate adjustment may therefore be more of a concern in smaller trials than larger trials. On the other hand, the stability of fitting a regression model will be less in a small trial. We sought to determine whether covariate adjustment led to similar power gains as in larger trials; and whether chance imbalance in small trials would lead to more discordance in statistical significance than in large trials contrasting unadjusted and adjusted treatment effects. We hereto devised a simulation study with small and large sample sizes drawn from two real trial datasets: the Global Use of Strategies to Open Occluded Coronary Arteries (GUSTO-I) trial and the first International Stroke Trial (IST).

2. Methods

2.1 Trial datasets

We obtained data from two large randomised trials, both have been described in detail elsewhere (8–10). We chose these trials for their size and availability only. The Global Use of Strategies to Open Occluded Coronary Arteries (GUSTO-I) trial recruited patients with acute myocardial infarction (MI) and randomised them to receive either: tissue plasminogen activator (tPA); or streptokinase in combination with heparin. To meet the inclusion criteria patients had to be within six hours of symptom onset, suffered at least 20 minutes of chest pain and had the appropriate electrocardiographic signals. The primary outcome was patient mortality by 30 days. The first International Stroke Trial (IST) was a large randomised, 2×3 factorial designed, open control, trial of patients who had suffered an acute ischaemic stroke within 48 hours of randomisation. The uncertainty principle was used to randomise patients to one of the treatment options. The uncertainty principle states that if an eligible patient presents for which the treating clinician is either certain that benefit or harm will come from giving the active treatment then the clinician should treat accordingly; otherwise, if the

clinician is suitably unsure of the likely response, then the patient may be randomised. Patients with a small chance of treatment benefit or a high risk of treatment harm were excluded. Patients were randomised to receive either aspirin or no aspirin in combination with two separate doses of heparin or no heparin. The primary outcome was six month death or disability. We restricted the IST data to definite or probable ischaemic strokes only. In addition, missing baseline data in IST were imputed once from a single run of a multiple-imputation model producing one complete data set for analysis. For all simulations with the IST data we compared aspirin with no aspirin.

2.2 Adjusted and unadjusted treatment effects

Logistic regression can be used to model the relationship between a binary outcome and one or more prognostic risk factors. Multivariable binary logistic regression is often applied in clinical trials when the primary outcome is dichotomous. There are a number of known important prognostic risk factors both for the risk of 30 day mortality following MI and for the risk of six month death or disability after acute ischaemic stroke (11, 12). For each dataset we considered three covariate adjustment strategies using logistic regression. The first two were the same for both GUSTO-I and IST: Model 0, an unadjusted model fitting treatment only; and Model 1, an age adjusted model. For GUSTO-I we defined Model 2a, a fully adjusted model, which included six important covariates, fitting: age; Killip class; systolic blood pressure (mmHg); pulse (beats/min); infarct location (anterior vs. inferior/other); and previous myocardial infarction (MI) (11). Killip class is a strictly ordinal measure with four unique levels summarising left ventricular function, though was fit as a linear term in Model 2a. Both systolic blood pressure and pulse were fit as piecewise linear functions split at 120mmHg and 50 beats/min respectively. In each case no effect is assumed below the specified cut-point with a linear association fit above the cut-point. For IST our fully adjusted model, Model 2b, included five important covariates: age, consciousness at randomisation (drowsy/coma vs. alert), visible infarct on CT scan, evidence of any arm or leg deficit and prior history of atrial fibrillation. Note that whilst decisions made in model development for the purpose of prediction are important they were not of any direct interest in the current paper. For the full model fits – Models 2a and 2b – variable selection was pre-specified and in the case of the GUSTO-I data, functional form (i.e., handling of continuous variables) was based on existing modelling work already published on this trial (13, 14).

2.3 Simulation

We simulated *small* (300 patients per arm, with approximately 29/600 events for GUSTO-I and 346/600 events for IST) and *large* (5000 patients per arm, approximately 665/10000 events for GUSTO-I and 6188/10000 for IST) samples from both the GUSTO-I (N=30,510) and the IST (N=18,372) trial datasets. Additionally, we explored an even smaller trial size using the IST data simulating a sample size of 50 patients per arm (approximately 50/100 events). In each scenario we assumed a comparison of two treatment arms of equal size. In order that the simulated trials had the same unadjusted type II error rate we empirically searched for the unique treatment effects such that an estimated power (1-type II error) of 80% was obtained for each trial sample size. Although a closed form solution for the asymptotic sample size formulae with sample size and type I and II error rates fixed is not tractable, an approximate solution can be reached through iterative computation. By varying

treatment effect across a range of values we were able to reach the treatment effects which obtained an estimated power of 80% for unadjusted analyses in each of the trial size scenarios. We then simulated 500,000 draws without replacement from each of the trial datasets at each of the sample sizes. We quantified the degree of chance imbalance in simulation for small and large trials by retaining from each draw the absolute difference in proportions for binary categorical variables between treatment and control and similarly the absolute difference in the mean for continuous variables. We then calculated the proportion of chance imbalance as those that exceeded a fixed difference across each of the simulated sample sizes.

We used those randomised to receive control (N=20,162 in GUSTO-I and N=9,189 in IST) to model patient outcome in each of the simulated replications. A separate simulation model was developed within the control arms for each of the trials using the covariates specified in Models 2a and 2b. From each replication we retained the estimated treatment effect and its estimated standard error (SE). We estimated the type I error rate under the null hypothesis (H0) of no treatment effect (an odds ratio of one) and the power under the alternative hypothesis (H1) with the empirically determined treatment effects. The type I error rate and the power were calculated as the ratio of significant results ($p<0.05$) to the total number of simulations. Significance was calculated using the Wald statistic (the ratio of the estimated coefficient and its estimated standard error). We quantified the proportion of discordant entries when cross-classifying the unadjusted and adjusted results defined at the conventional $p<0.05$ significance level, i.e., an absolute Wald statistic which exceeds a standard normal deviate of 1.96.

We carried out all simulations and analyses in R version 3.0.1 with the add-on package rms (15).

3. Results

Patient characteristics were well balanced between the treatment arms (Table 1). A formal likelihood ratio test of imbalance for each of the measured baseline covariates is provided for illustrative purposes only – none were significant at the 5% level.

The full multivariable simulation model for GUSTO-I included all six risk factors for 30 day mortality as used in Model 2a (Table 2). The relative contribution of each risk factor to the overall Chi-square (χ^2) amongst those 20,162 patients randomised to receive control showed that age was the most important covariate followed by SBP. Model 2a had an R^2 of 23% suggesting a good amount of explained variation. The event rate amongst those randomised to receive the control treatment was 7.0%. The empirically determined ORs were: 0.27 (an approximate absolute risk reduction (ARR) of 5%) and 0.77 (ARR of 1.5%) for small and large trials respectively. Similarly for IST patient age and alertness were amongst the most important predictors of six month death or disability – Model 2b explained a large amount of the variation in the IST data with an R^2 of 25%. The event rate amongst those randomised to receive control was 50.5%. The empirically determined ORs were: 0.25 (ARR of 30.2%), 0.57 (ARR of 13.7%) and 0.87 (ARR of 3.5%) for extra small, small and large trials respectively. The probability of imbalance in both GUSTO-I and IST quantified using the

absolute difference measure highlights the greater probability of imbalance that can be expected in a small trial contrast to a large trial. Joint imbalance between covariates was not explored. Note that different thresholds where adopted in each trial as this was a within trial comparison.

Table 3 shows a cross-tabulation of our findings from the GUSTO-I data. This illustrates the way the proportions of discordance were obtained. Similar results were found using the IST data. The findings from our simulation are provided in full in Table 4.

In the absence of a true treatment effect (OR=1) the type I error rate was slightly conservative for small trials (4.6% without adjustment for covariates) and close to 5% for large trials (Table 4). The power to detect a true treatment effect using the unadjusted model, Model 0, was empirically fixed at about 80% for trial size. The gain in power under the two adjustment strategies was approximately the same for both small and large trials in GUSTO-I (+2% for Model 1 and +5% for Model 2a) and for IST (+2% for Model 1 and +6% for Model 2b). The overall performance of each model under each sample size is summarised through the average R^2 illustrating the added benefit of adjusting for strong prognostic variables (Models 1, and 2a/b) describing the heterogeneity within each of the populations. Nagelkerke's R^2 depends on the size of the sample as noted in Table 4 (16).

With a true treatment effect (under H1), the proportion of discordance between a significant unadjusted treatment effect (Model 0) and a non-significant adjusted treatment effect (Model 2a and Model 2b) result was similar for both small and large trials in GUSTO-I and IST (approximately 2%). The proportion of discordance between a non-significant unadjusted treatment effect (with Model 0) and a significant adjusted treatment effect (with Model 2a and Model 2b) was approximately 7% in GUSTO-I and 9% in IST. Hence, if the alternative hypothesis was true, a covariate-adjusted model correctly concluded significance where an unadjusted model would not in more trials than the other way around. A similar degree of equivalence was found between small and large sample sizes: when contrasting Model 0 with Model 1; under H0 (with approximately equivalent proportions); and with a smaller sample size (in IST).

There was an indication that covariate adjustment in the smaller trial sizes (i.e., 300 per arm and 50 per arm) introduced a bias in the treatment effect estimates resulting in a somewhat overestimated treatment effect relative to the true known effect under H1 (Table 4). A similar bias was observed under the unconditional model when adjusting for un-associated variables (data not presented) indicating that this bias in small samples was likely caused by adjustment.

4. Discussion

Our simulation suggests similar power gains for small and large RCTs through covariate adjustment for strongly predictive risk factors. Contrary to what we anticipated, the impact of chance imbalance in small versus large trials did not result in more discordance between unadjusted and adjusted statistical significance. Covariate adjustment is hence equally recommendable in small and large trials.

We explored covariate adjustment using binary logistic regression in randomised trial data. In contrast to linear models, the impact of covariate adjustment on the estimated treatment effect through non-linear regression models such as binary logistic, and Cox proportional hazards regression, is somewhat counterintuitive (17, 18). Adjusting for important covariates in a linear model has no impact on the estimated treatment effect but will be met by a reduction in the SE (a gain in precision and efficiency). In logistic regression, however, the SE is actually inflated by covariate adjustment (as seen in Table 4) although this is offset by an increase in the expected size of the treatment effect. Hauck *et al.* point out that the estimated ORs from a conditional and unconditional model are informative of the effect of treatment in separate reference populations. The moment additional information is included via covariate adjustment the definition of the reference population, and hence the odds ratio, changes and takes on an interpretation specific to that population (19). For this reason the treatment effect estimate from an adjusted non-linear model can be viewed as a move towards a more patient oriented estimate of the treatment effect accommodating for the individual's risk profile through the addition of prognostic covariates (20, 21).

Hence, the benefit of covariate adjustment is two-fold. It accommodates for the impact caused by chance imbalance as well as accounting for baseline prognostic risk. Conditioning on this extra information increases the statistical power and yields estimates of treatment effects that are more relevant to the individual patient. The gains from covariate adjustment in the context of continuous outcomes can be interpreted through the strength of the correlation between predictor and outcome, whilst with a binary response the R^2 may be used for a similar interpretation (14, 22). In our simulation we observed that with lower R^2 values (i.e., with Model 1) there was little benefit from adjustment, but with models adjusting for a number of strongly predictive covariates (i.e., Models 2a and 2b) there were substantial gains to be made. Based on our findings we suggest that in practice the gain made in statistical power from adjusting for strong prognostic factors (with R^2 values ≥20%) may exceed that of correcting for the impact of chance imbalance alone. Additionally, with equivalent gains in power in both small and large trials similar reductions in sample size should be anticipated (5). The reduction in required sample size is approximately equal to the R^2 of the prognostic model (23).

The observed number of events drives the reliability of estimation of parameters in a binary logistic regression model (i.e., the smallest frequency in a binary response). Some authors recommend a minimum of 10 events per variable (EPV) (24), while an EPV of 5 may be sufficient in the context of adjusting for confounders (25). We do not advocate a blithe approach to covariate adjustment in randomised trials. Indeed others have shown that adjusting for too many variables in small trials can in fact increase the type I error rate (26). Given the anticipated incidence of the outcome trialists would be wise to only consider adjusting for covariates that are known to be of prognostic importance and therefore predefined within the trial analysis plan (27). This practice is advocated by the European drug regulatory authority (EMA) (28). In their current guidance on covariate adjustment the EMA places emphasise on including a small number of prespecified and important prognostic characteristics within a model of simple form; only utilising non-linear transformations when there is strong *a priori* evidence to support it. Alternatively if an existing prognostic model was available trialists could adjust for predicted risk therefore

saving on the degrees of freedom spent on re-estimating parameters as well as minimising the risk of false positives by avoiding multiple subgroup analyses (29). We stress that covariate adjustment is as important in small trials as it is in large trials. A large trial has a smaller chance of imbalance but this is not a justification for ignoring patient characteristics; neither is a small trial too small to consider adjustment for important prognostic factors.

A similar study to ours considered the impact that chance imbalance in a single binary prognostic factor had on the estimation of treatment effect (30). This was a more artificial simulation than ours, exploring the impact of several trial level parameters, including: the size of the risk factor effect; the treatment effect; sample size; prevalence of the risk factor; and the incidence of the outcome. They identified an underestimation of a moderate treatment effect with sample sizes of 50 patients per arm or less (where the probability of chance imbalance was high) when failing to adjust for an important prognostic factor. Above this sample size estimation was unbiased. For a Normally distributed response and baseline measure, each with known variance and correlation Senn has demonstrated that on theoretical grounds covariate imbalance shares equal concern in both small and large trials where the effect on size is independent of the sample size (31). Our findings fit well with both of these studies. By fixing statistical power at 80% and exploring the operation of covariate adjustment exclusively on statistical significance our findings place further emphasis on the importance of multivariable covariate adjustment in the analysis of clinical trial data, yielding greater statistical power and – under H1 – a greater proportion of correct significant conclusions irrespective of sample size.

Although small in contrast to our *large trial* size, our definition of a *small trial* size was rather arbitrary and may be regarded by some as ‘large’. Indeed, it has been suggested that for trials with fewer than 50 patients per arm where chance imbalance could be more of a concern, trialists might consider employing minimisation – a dynamic approach to treatment allocation which ensures good balance between treatment arms for a set of important covariates (1, 32). An important point which is frequently overlooked by those who adopt minimisation though is that the same conditions imposed during randomisation must also be accounted for during analysis. Without appropriate adjustment the standard error of the obtained treatment effect is biased upwards, the statistical power is reduced and the type I error rate is low (33). Furthermore, the incidence of 30-day mortality in GUSTO-I was relatively low (7.0%). This implied that in simulation the required treatment effect for a trial containing fewer than 300 patients per arm for a fixed power of 80% would be so large that it would prevent all observed events in the treated arm and would preclude the estimation of a treatment effect. It was for this reason that we repeated our simulation method using data from the First International Stroke Trial (IST) (10). The primary outcome in IST had a far greater incidence permitting us to explore much smaller trial sizes. However, despite this there was little difference in discordance, at least of any practical importance.

Our study has some limitations. The simulation model imposes the assumption of linearity and additivity in the predictor outcome associations without formal testing; it is entirely possible that more complex associations underlie this data. However, without a loss in generality, the hypothesis that we have explored depends more on the perceived imbalance between patient characteristics and this would be the same irrespective of the underlying

simulation model. These findings are conditional upon randomisation being conducted correctly and considered differences that arise between treatment arms through random chance only. Although the assumption that randomisation has been undertaken in good faith seems the most plausible and reasonable default position to take, indeed anything else would be rather artificial (34).

We conclude that covariate adjustment through binary logistic regression is as important for small trials as it is for large trials. The impact of chance imbalance causes little practical difference in the proportion of discordance when comparing small and large trials. The size of the trial alone should hence not be a justification for avoiding implementing covariate adjustment.

Acknowledgments

DDT was supported by an MRC HTMR grant (G0800803). WNW was supported by an MRC Clinician Scientist Fellowship G0902303. EWS was supported by a U award (1U01-NS086294-01, Value of Personalized Risk Information). We acknowledge the GUSTO-I and IST collaborators for their efforts in conducting the trials and for making their raw data available for secondary analysis.

References

1. Senn, SJ. Statistical issues in drug development. 2. Wiley; 2007.
2. Steyerberg EW, Bossuyt PMM, Lee KL. Clinical trials in acute myocardial infarction: Should we adjust for baseline characteristics? American Heart Journal. 2000; 139(5):745–51. [PubMed: 10783203]
3. Turner EL, Perel P, Clayton T, Edwards P, Hernández AV, Roberts I, et al. Covariate adjustment increased power in randomized controlled trials: an example in traumatic brain injury. Journal of Clinical Epidemiology. 2012; 65(5):474–81. [PubMed: 22169080]
4. Kahan B, Jairath V, Dore C, Morris T. The risks and rewards of covariate adjustment in randomized trials: an assessment of 12 outcomes from 8 studies. Trials. 2014; 15(1):139. [PubMed: 24755011]
5. Hernández AV, Steyerberg EW, Habbema JDF. Covariate adjustment in randomized controlled trials with dichotomous outcomes increases statistical power and reduces sample size requirements. Journal of Clinical Epidemiology. 2004; 57(5):454–60. [PubMed: 15196615]
6. Schulz K, Altman D, Moher D. Group tC. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. BMC Medicine. 2010; 8(1):18. [PubMed: 20334633]
7. Yu L-M, Chan A-W, Hopewell S, Deeks J, Altman D. Reporting on covariate adjustment in randomised controlled trials before and after revision of the 2001 CONSORT statement: a literature review. Trials. 2010; 11(1):59. [PubMed: 20482769]
8. The GUSTO Investigators. An International Randomized Trial Comparing Four Thrombolytic Strategies for Acute Myocardial Infarction. New England Journal of Medicine. 1993; 329(10):673–82. [PubMed: 8204123]
9. IST Collaborative Group. The International Stroke Trial (IST): a randomised trial of aspirin, subcutaneous heparin, both, or neither among 19435 patients with acute ischaemic stroke. The Lancet. 1997; 349(9065):1569–81.
10. Sandercock P, Niewada M, Czlonkowska A, the International Stroke Trial Collaborative Group. The International Stroke Trial database. Trials. 2011; 12(1):101. [PubMed: 21510853]
11. Lee KL, Woodlief LH, Topol EJ, Weaver WD, Betriu A, Col J, et al. Predictors of 30-Day Mortality in the Era of Reperfusion for Acute Myocardial Infarction: Results From an International Trial of 41 021 Patients. Circulation. 1995 Mar 15; 91(6):1659–68. 1995. [PubMed: 7882472]

12. Veerbeek JM, Kwakkel G, van Wegen EEH, Ket JCF, Heymans MW. Early Prediction of Outcome of Activities of Daily Living After Stroke: A Systematic Review. Stroke. 2011 May 1; 42(5):1482–8. 2011. [PubMed: 21474812]
13. Califf RM, Woodlief LH, Harrell FE Jr, Lee KL, White HD, Guerci A, et al. Selection of thrombolytic therapy for individual patients: Development of a clinical model. American Heart Journal. 1997; 133(6):630–9. [PubMed: 9200390]
14. Steyerberg, EW. Clinical Prediction Models: A practical approach to development, validation, and updating. Springer; 2009.
15. Harrell, FE. rms: Regression Modeling Strategies. R package version 3.6-3. 2013. http://CRANR-projectorg/package=rms
16. Nagelkerke NJD. A note on a general definition of the coefficient of determination. Biometrika. 1991 Sep 1; 78(3):691–2. 1991.
17. Gail MH, Wieand S, Piantadosi S. Biased estimates of treatment effect in randomized experiments with nonlinear regressions and omitted covariates. Biometrika. 1984 Dec 1; 71(3):431–44. 1984.
18. Robinson LD, Jewell NP. Some Surprising Results about Covariate Adjustment in Logistic Regression Models. International Statistical Review / Revue Internationale de Statistique. 1991; 59(2):227–40.
19. Hauck WW, Neuhaus JM, Kalbfleisch JD, Anderson S. A consequence of omitted covariates when estimating odds ratios. Journal of Clinical Epidemiology. 1991; 44(1):77–81. [PubMed: 1986061]
20. Hauck WW, Anderson S, Marcus SM. Should We Adjust for Covariates in Nonlinear Regression Analyses of Randomized Trials? Controlled Clinical Trials. 1998; 19(3):249–56. [PubMed: 9620808]
21. Ford I, Norrie J. The role of covariates in estimating treatment effects and risk in long-term clinical trials. Stat Med. 2002; 21(19):2899–908. [PubMed: 12325106]
22. Pocock SJ, Assmann SE, Enos LE, Kasten LE. Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems. Stat Med. 2002; 21(19):2917–30. [PubMed: 12325108]
23. Hernández AV, Steyerberg EW, Butcher I, Mushkudiani N, Taylor GS, Murray GD, et al. Adjustment for strong predictors of outcome in traumatic brain injury trials: 25% reduction in sample size requirements in the IMPACT study. Journal of neurotrauma. 2006; 23(9):1295–303. [PubMed: 16958582]
24. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. Journal of Clinical Epidemiology. 1996; 49(12): 1373–9. [PubMed: 8970487]
25. Vittinghoff E, McCulloch CE. Relaxing the Rule of Ten Events per Variable in Logistic and Cox Regression. American Journal of Epidemiology. 2007 Mar 15; 165(6):710–8. 2007. [PubMed: 17182981]
26. Kahan B, Morris T. Adjusting for multiple prognostic factors in the analysis of randomised trials. BMC Medical Research Methodology. 2013; 13(1):99. [PubMed: 23898993]
27. Harrell, FE. Regression Modeling Strategies: With applications to linear models, logistic regression, and survival analysis. Springer; 2001.
28. European Medicines Agency. [cited 2014 04/07] Guideline on adjustment for baseline covariates. 2014. Available from: http://www.ema.europa.eu/ema/
29. Kent D, Rothwell P, Ioannidis J, Altman D, Hayward R. Assessing and reporting heterogeneity in treatment effects in clinical trials: a proposal. Trials. 2010; 11(1):85. [PubMed: 20704705]
30. Chu R, Walter SD, Guyatt G, Devereaux PJ, Walsh M, Thorlund K, et al. Assessment and Implication of Prognostic Imbalance in Randomized Controlled Trials with a Binary Outcome - A Simulation Study. PLoS ONE. 2012; 7(5):e36677. [PubMed: 22629322]
31. Senn SJ. Covariate imbalance and random allocation in clinical trials. Stat Med. 1989; 8(4):467–75. [PubMed: 2727470]
32. Altman DG. Comparability of Randomised Groups. Journal of the Royal Statistical Society Series D (The Statistician). 1985; 34(1):125–36.
33. Kahan BC, Morris TP. Improper analysis of trials randomised using stratified blocks or minimisation. Stat Med. 2012; 31(4):328–40. [PubMed: 22139891]

34. Senn S. Testing for baseline balance in clinical trials. Stat Med. 1994; 13(17):1715–26. [PubMed: 7997705]

What is New?

- Empirical evidence to support the hypothesis that covariate adjustment leads to similar gains in statistical power irrespective of the size of the trial.
- Contrary to what may be anticipated by many trialists the relative impact of the larger expected imbalance in small trials compared to large trials does not lead to a differential change in the effect of covariate adjustment. The proportions of discordance in statistical significance between adjusted and unadjusted analyses were comparable in both small and large trials.
- Our findings support the use of covariate adjustment in randomised trials using multivariable regression analyses in both small and larger trials.

Table 1

Summary of baseline characteristics and outcome by treatment arm for the GUSTO-I (N=30,510) and the IST trials (N=18,372).

Trial/Characteristics	Treatment (n)	Control (n)	P-value
GUSTO-I	***tPA (10,348)***	***Control (20,162)***	
Outcome			
30-day mortality	653 (6.31)	1475 (7.32)	0.0010
Baseline measurements *			
Age (years)	61.57 (52.33–70)	61.58 (52.02–69.78)	0.2434
Killip class (I to IV)			0.4692
I	8798 (85.02)	17209 (85.35)	-
II	1328 (12.83)	2529 (12.54)	-
III	142 (1.37)	275 (1.36)	-
IV	80 (0.77)	149 (0.74)	-
SBP (mmHg)	129 (112–144)	129 (112–144)	0.4239
Pulse	73 (62 to 86)	73 (62 to 86)	0.6595
Infarct location (anterior vs. inferior/other)	4022 (38.87)	7844 (38.91)	0.9493
Previous MI	1738 (16.80)	3320 (16.47)	0.4649
IST	***Aspirin (9183)***	***Control (9189)***	
Outcome			
6m death or disability	5693 (49.47)	5815 (50.53)	0.0713
Baseline measurements *			
Age (years)	73 (65–80)	73 (65–80)	0.8122
Drowsy/coma vs. alert	2121 (23.10)	2134 (23.22)	0.8391
Infarct visible on CT	3128 (34.06)	3203 (34.86)	0.2575
Any arm or leg deficit	8053 (87.69)	8105 (88.20)	0.2897
Prior Atrial Fibrillation	1628 (17.73)	1566 (17.04)	0.2197

Note: Median and IQR presented for continuous measures, frequency and percentage for categorical variables.

* P-values for imbalance are presented strictly for illustrative purposes; all are from a likelihood ratio test

J Clin Epidemiol. Author manuscript; available in PMC 2017 November 30.

Table 2

Design of the simulation study for GUSTO-I and IST, all included covariates were significant at the $P<0.0001$ level (see text for full details).

	Multivariable Simulation Model			Relative Importance	Estimated Imbalance (%)[a,b]		
Model/Variable	**Coefficient**	**S.E.**	**OR (95% CI)**	**Chi-square (χ^2)**	**Extra small trial**	**Small trial**	**Large trial**
Model 2a in GUSTO-I, control arm n= 1475, N=20162							
Intercept	−3.910	0.317	-	-	-	-	-
Age[c]	0.075	0.003	1.08 (1.07 to 1.08)	647	-	62	4
Killip class[c]	0.624	0.044	1.87 (1.71 to 2.04)	199	-	0	0
SBP (<120)[c]	−0.039	0.002	0.96 (0.96 to 0.97)	301	-	55	2
Pulse (>50bpmin)[c]	0.017	0.002	1.02 (1.01 to 1.02)	137	-	72	14
Infarct location (anterior vs. inferior/other)	0.570	0.060	1.77 (1.57 to 1.99)	91	-	61	5
Previous MI	0.458	0.068	1.58 (1.38 to 1.80)	46	-	55	1
Simulated Treatment effect[d]							
Small Trial (300 per arm)	−1.324	-	0.27 (−)	-	-	-	-
Large Trial (5000 per arm)	−0.264	-	0.77 (−)	-	-	-	-
Model 2b in IST control arm, n=5815, N=9189							
Intercept	−4.462	0.176	-	-	-	-	-
Age[c] (per 10 years)	0.523	0.022	1.69 (1.62 to 1.76)	554	39	3	0
Drowsy/coma vs. alert	1.658	0.076	5.25 (4.52 to 6.09)	477	72	36	0
Infarct visible on CT	0.205	0.052	1.23 (1.11 to 1.36)	16	75	44	0
Any arm or leg deficit	0.991	0.072	2.69 (2.34 to 3.10)	190	64	27	0
Prior Atrial Fibrillation	0.514	0.074	1.67 (1.45 to 1.93)	49	69	31	0
Simulated Treatment effect[d]							
Extra small trial (50 per arm)	−1.405	-	0.25 (−)	-	-	-	-
Small trial (300 per arm)	−0.569	-	0.57 (−)	-	-	-	-
Large trial (5000 per arm)	−0.141	-	0.87 (−)	-	-	-	-

[a]GUSTO-I: Fraction of simulations (%) exceeding a difference of 2% between treatment and control for binary and a difference of 0.5 unit for continuous variables;

[b]IST: Fraction of simulations (%) exceeding a difference of 3% between treatment and control for binary and a difference of 2 years for age;

[c]These variables were all fit as continuous linear terms; and

[d] Treatment effect ascertained through empirical iteration for 80% power.

Table 3

Cross tabulation of the results from the GUSTO-I simulation comparing the obtained treatment P-values from unadjusted (Model 0) and adjusted (Model 2a) analyses in both small and large sample sizes under H1.

	Adjusted, Model 2a (n = 2×300)			Adjusted, Model 2a (n = 2×5000)		
Unadjusted	**P<0.05**	**P≥0.05**	**Total**	**P<0.05**	**P≥0.05**	**Total**
P<0.05	389,515	10,344	399,859	391,141	8,520	399,661
	(77.9)	(2.1)	(80.0)	(78.2)	(1.7)	(80.0)
P≥0.05	35,644	64,497	100,141	37,450	62,889	100,339
	(7.1)	(12.9)	(20.0)	(7.5)	(12.6)	(20.0)
Total	425,159	74,841	500,000	428,591	71,409	500,000
	(85.0)	(15.0)		(85.7)	(14.3)	

Note: cell entries are the observed frequencies and both joint and marginal probabilities (%) from 500,000 replications

Table 4

Results from simulation studies using GUSTO-I and IST trial datasets (500,000 replications).

Estimates	Extra small trial (n=2×50)			Small trial (n=2×300)			Large trial (n=2×5000)		
	Model 0	**Model 1**	**Model 2a/b**	**Model 0**	**Model 1**	**Model 2a/b**	**Model 0**	**Model 1**	**Model 2a/b**
Mean effect from GUSTO-I simulation									
Treatment coefficient	-	-	-	−1.21	−1.26	−1.43	−0.23	−0.24	−0.26
Treatment SE	-	-	-	0.48	0.48	0.53	0.08	0.08	0.09
Nagelkerke R^2 (%)	-	-	-	4.7	15.5	28.6	0.2	11.8	23.1
Mean effect from IST simulation									
Treatment coefficient	−1.15	−1.27	−1.54	−0.46	−0.51	−0.58	−0.12	−0.13	−0.14
Treatment SE	0.42	0.45	0.53	0.17	0.18	0.19	0.04	0.04	0.05
Nagelkerke R^2 (%)	10.9	22.0	39.3	1.9	13.5	28.1	0.1	11.7	25.2
Conditional Probability (%) GUSTO-I									
Size of test (type I error)	-	-	-	4.6	4.7	5.0	5.1	5.1	5.1
Power (1-type II error)	-	-	-	80.0	82.0	85.0	79.9	82.0	85.7
Conditional Probability (%) IST									
Size of test (type I error)	4.9	5.0	5.4	4.9	5.0	5.0	5.0	5.0	5.0
Power (1-type II error)	80.6	82.0	85.8	80.0	82.6	86.8	79.5	83.1	87.0
Joint Probability under H1 (%) GUSTO-I									
$p_{unadj}<0.05$ vs. $p_{adj}\geq 0.05$	-	-	-	-	1.4	2.1	-	1.6	1.7
$p_{unadj}\geq 0.05$ vs. $p_{adj}<0.05$	-	-	-	-	3.4	7.1	-	3.7	7.5
Joint Probability under H0 (%) GUSTO-I									
$p_{unadj}<0.05$ vs. $p_{adj}\geq 0.05$	-	-	-	-	1.0	1.7	-	1.1	1.8
$p_{unadj}\geq 0.05$ vs. $p_{adj}<0.05$	-	-	-	-	1.1	2.0	-	1.1	1.8
Joint Probability under H1 (%) IST									
$p_{unadj}<0.05$ vs. $p_{adj}\geq 0.05$	-	2.5	2.8	-	1.9	1.8	-	1.7	1.6
$p_{unadj}\geq 0.05$ vs. $p_{adj}<0.05$	-	3.7	8.0	-	4.9	9.0	-	5.3	9.1
Joint Probability under H0 (%) IST									
$p_{unadj}<0.05$ vs. $p_{adj}\geq 0.05$	-	1.3	2.0	-	1.3	1.9	-	1.4	2.0
$p_{unadj}\geq 0.05$ vs. $p_{adj}<0.05$	-	1.4	2.5	-	1.4	2.1	-	1.4	2.0

Note: Model 0 included no covariates; Model 1 included age as a single, continuous covariate; while Model 2a adjusted for all baseline characteristics listed in Table 2 for GUSTO-I and Model 2b adjusted for all baseline characteristics listed in Table 2 for IST