**Poor allocation concealment methods are associated with heterogeneity in age and statistical significance of the primary outcome: review of recent trials published in four general medical journals**

Alex Mitchell1, Thirimon Moe Byrne2, Rachel Cunningham-Burley1, Alexandra Dean1, Aditi Rangan3, Jenny Roche1, David J Torgerson1

1. **York Trials Unit, Department of Health Sciences, University of York, UK**
2. **Department of Health Sciences, University of York, UK**
3. **Newcastle Medical School, University of Newcastle, UK**

Corresponding author: Alex Mitchell (alex.mitchell@york.ac.uk, 01904 32 (1756))

**Objective:** To assess the association of the quality of allocation concealment with heterogeneity in age, the p-value of the primary outcome and statistical significance of the primary outcome.

**Study Design and Setting:** We extracted data from articles published in four major medical journals in 2017 and 2018 that reported the results of randomised controlled trials. The outcome measures were the quality of allocation concealment used in the trial, the p-value of the primary outcome, whether the p-value of the primary outcome was statistically significant and the level of heterogeneity in age between the treatment groups (measured using the I2 statistic). The association between the quality of allocation concealment and the p-value of the primary outcome was assessed using a kernel density plot, while the association between the quality of allocation concealment and whether the p-value was statistically significant was assessed using logistic regression.

**Results:** Trials that used inadequate concealment methods were more likely to report statistically significant findings than trials that used good or adequate methods (OR 1.90; 95% CI: 0.91 to 3.95; p=0.09).The values of I2 for trials that used good, adequate, inadequate and unclear concealment methods were 0%, 1.0%, 32.6% and 93.8% respectively.

**Conclusion:** There is evidence of an association between poor allocation concealment methods and statistical significance of the primary outcome. Trials that use inadequate allocation concealment methods are more likely to have statistically significant p values compared with trials using good or adequate allocation concealment methods.

**Keywords:** Randomised controlled trial, allocation concealment, bias

**1. Introduction**

A well-conducted randomised controlled trial is the gold standard for assessing the effectiveness of healthcare interventions. When done correctly, the use of random allocation in trials removes the risk of selection bias by ensuring researchers cannot allocate patients with certain characteristics to a particular treatment group1, 2. However, randomisation will only remove the risk of selection bias if the allocation schedule is concealed from the investigator recruiting the patient to the trial. Otherwise the investigator may use their knowledge of the next allocation to subvert the randomisation3. As a result, a range of methods including central randomisation and sealed envelopes have been used in trials in an attempt to conceal the allocation schedule4. However, some allocation concealment methods are better than others. A study by Schulz and colleagues published in 1995 assessed the quality of allocation concealment of 250 trials5. The authors categorised the allocation concealment as adequate if the trial used central randomisation or sealed envelopes. Trials that used methods such as alternation or reference to date of birth were categorised as having inadequate concealment, while those that did not report a method of allocation concealment were categorised as unclear. The study found that trials with inadequate or unclear concealment reported larger estimated treatment effects than trials with adequate concealment. Central randomisation is the most methodologically sound method of concealment, but can be subverted when the block sizes are small and fixed. However, the widespread practice of using sealed opaque envelopes is, in our view, questionable. For example, two case studies have demonstrated that envelopes are not conducive to secure concealment. First, a cardiovascular trial published in 1999 using sealed envelopes reported small but highly statistically significant baseline imbalances in height, weight and blood pressure6, which led Peto to suspect the randomisation process had been violated7. Second, a historical case study of a trial also conducted in the 1990s found strong evidence that the randomisation process in a surgical trial that used sealed envelopes was subverted8. Furthermore, Hewitt and colleagues reviewed 234 individually randomised trials published in 2002 in four high impact medical journals (the BMJ, JAMA, The Lancet, and the New England Journal of Medicine)9. They found that trials that reported the use of inadequate allocation concealment (largely using envelopes) were more likely to report statistically significant findings compared with trials that reported adequate allocation concealment9. Recently a review of a sample of RCTs published in major medical journals in 2015 found that 19% of trials continued to use inadequate concealment methods10. In addition, Pildal and colleagues found a relationship between inadequate allocation concealment and effect estimates in meta-analyses11. Hopewell and colleagues also found that as of 2006, allocation concealment is poorly reported12. In this paper we assess the quality of allocation concealment in recently published RCTs and assessed whether there was any statistical evidence of an association between allocation concealment quality and the following: heterogeneity in age, p-value of the primary outcome and statistical significance. This review was not pre-registered.

**2. Materials and Methods**

We hand searched the BMJ, JAMA, The Lancet, and the New England Journal of Medicine for individually randomised controlled trials published in the years 2017 and 2018. We designed and piloted a data extraction form, which was then used to extract the study title, journal name, allocation concealment method, summary statistics of age for each treatment group, number of participants in each treatment group and the p-value of the primary outcome. The search and selection processes were integrated, and no second checking was done. Authors inexperienced in the hand searching method were trained in this method before the study began. Throughout the extraction phase of the study, authors consulted AM and DT if they had any queries regarding data extraction. Articles were included if it was reported that participants were randomised and the trial was published as a full report with the results of the main analyses. We excluded articles that reported the results of trials that were cluster randomised, were phase I or phase II studies, had more than two arms, were crossover trials or did not report summary statistics amenable to meta-analysis for age at baseline by treatment group (either mean and standard deviation, or median and interquartile range)13.

For each article, we extracted the mean and standard deviation of the age for each treatment group. If the mean and standard deviation were not reported, we extracted the median and interquartile range, from which we estimated a standard deviation14. For each article we extracted the p-value of the primary outcome obtained from the primary analysis. If an article reported multiple primary outcomes, we took the p-value of the primary outcome that was stated first. Articles that reported confidence intervals instead of p-values could still be included in the study, as the p-value can be derived from the confidence interval15. We categorised articles according to whether the allocation concealment was good (trial was placebo controlled, and/or independent third party allocation method, such as web-based or telephone allocation, and used large block sizes, or used minimisation), adequate (used an independent third party allocation system used small blocks (i.e. < 6) but did not stratify by site), inadequate (used envelopes not opened by a third party, used small block sizes stratified by site or used small block sizes and was a single site study.) or were unclear. The extracting author initially categorised the articles. The categorisation was then second checked by DT. For those trials that did not report clearly their allocation concealment method we emailed the corresponding author for more details.

We used a kernel density plot to investigate the relationship between quality of allocation concealment and the p-value of the primary outcome16, 17. If there is no association between quality of allocation concealment and the p-value of the primary outcome, then the distribution of p-values for each category should be similar. We compared the likelihood of studies reporting a statistically significant result (p<0.05) for each pairwise comparison of concealment quality by carrying out a logistic regression model controlling for journal and total sample size. We took clustering by journal into account by using robust standard errors. We also compared the likelihood of studies that used inadequate concealment methods reporting a statistically significant result to studies that used good or adequate methods using the same methods. We undertook a fixed effects meta-analysis of age of randomised participants of each group of trials18, 19. If the randomisation is unbiased then we would expect the heterogeneity of age to be zero as the difference between the intervention arm and the control arm should only differ by chance. Therefore, there should be no heterogeneity in a meta-analysis as all trials are trying to estimate the same difference, which is zero. The presence of heterogeneity suggests that some of the trials have a biased estimate of an age difference, which would support the view that the randomisation method is suboptimal. We used the I2 statistic to assess evidence for heterogeneity. For papers that reported the interquartile range, we followed the recommendations given in the Cochrane Handbook of Systematic Interventions and obtained an estimate of the standard deviation by dividing the width of the interquartile range by 1.3513.

**3. Results**

We identified 338 eligible papers, of which 214 (63.3%) were rated good, 66 (19.5%) were rated adequate, 47 (13.9%) were rated inadequate and 11 (3.3%) were unclear. Table 1 provides a breakdown of the number of RCTs extracted for each journal by year.

[TABLE 1]

As the search and selection processes were integrated it is not possible to give information on the number of studies that were screened and the number that were ineligible. As the distribution of p-values was highly skewed, we used the logit function to transform the p-values before producing the kernel density plot. Figure 1 is the kernel density plot of the logit transformation of the p-values, displayed by the quality of allocation concealment. The average p-value was 0.21 (SD 0.29) for trials that used a good method of allocation concealment, 0.26 (SD 0.31) for trials that used an adequate method, 0.16 (SD 0.28) for trials that used an inadequate method and 0.39 (SD 0.36) for trials whose method was unclear. Table 2 displays the likelihood of reporting a statistically significant result for each pairwise quality comparison. We found that trials that used inadequate concealment methods were more likely to report statistically significant findings than trials that used good or adequate methods (OR 1.90; 95% CI: 0.91 to 3.95; p=0.09).

[TABLE 2]

With respect to heterogeneity we found that for the trials with good allocation concealment that the I2 value was zero, and that for trials with adequate allocation I2 was equal to 1.0%. However, for the inadequate allocation methods there was 32.6% heterogeneity, suggesting that imbalances in age in some trials were not due to chance. For trials that did not clearly report the method of allocation concealment used, I2 was equal to 93.8%, which may be due to these trials being a mixture of studies that used good, adequate or unclear concealment methods.

**4. Discussion**

We found evidence of a relationship between poor allocation concealment methods and the p-value of the primary outcome. There was a statistically significant difference in the likelihood of reporting a p-value less than 0.05 between trials with adequate concealment and trials with inadequate concealment. Trials with good concealment were also more likely than trials with inadequate concealment to report a p-value less than 0.05, however this difference was not statistically significant. Since Schulz reported the relationship between poor randomisation practice and outcomes, methods of randomisation have improved: we did not find any trials where future randomisations were ‘open’ (e.g., pinned on a notice board). However, significant numbers of trials still use opaque sealed envelopes or stratify by centres using fixed small block sizes, which allow treatment prediction. Hewitt and colleagues observed, in the same group of journals in 2002, that 18% of the trials used inadequate methods, whilst Clark and colleagues in 2016 and in the same journals noted that 19% of trials used inadequate methods, whilst we found about 14% using inadequate methods. This suggests relatively limited progress in avoiding the use of poor allocation concealment methods. Clark and colleagues argued in 2016 that journals should announce that trials with poor randomisation practices would no longer be published, as is the case with unregistered trials: and we support this view as our current report suggests that there is still a relationship between poor allocation concealment methods and outcomes.

One limitation of our study is that the results are only applicable to studies published in the four general medical journals that we searched, when there is evidence that most trials are published in smaller, more specialised journals20. Randomised controlled trials published in other journals may differ systematically from trials published in the BMJ, JAMA, The Lancet, and the New England Journal of Medicine. In addition, the fact that we only included individually randomised controlled trials means our results cannot be extrapolated to cluster RCTs. Finally, the observational nature of study means that our results may have been affected by unmeasured confounders of the relationship between quality of allocation concealment and the likelihood of reporting a statistically significant result.

In conclusion, we have found a relationship between poor allocation methods and the statistical significance of the primary outcome. Trials that use sealed opaque envelopes or small block sizes are more likely to have statistically significant p-values compared with trials using good or adequate allocation methods.

**Funding**

No funding was received for this study.

**Data Sharing**

A list of eligible studies and the extracted characteristics is available from the corresponding author on request.

**Acknowledgements**

We would like to thank the anonymous peer reviewers for their helpful and constructive comments.

**References**

1. Altman DG, Bland JM. Treatment allocation in controlled trials: why randomise? BMJ. 1999;318(7192):1209-.

2. Attia A. Bias in RCTs: confounders, selection bias and allocation concealment. Middle East Fertility Society Journal. 2005;10(3):258.

3. Torgerson DJ, Roberts C. Randomisation methods: concealment. Bmj. 1999;319(7206):375-6.

4. Dettori J. The random allocation process: two things you need to know. Evidence-based spine-care journal. 2010;1(03):7-9.

5. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA. 1995;273(5):408-12.

6. Hansson L, Lindholm LH, Niskanen L, Lanke J, Hedner T, Niklason A, et al. Effect of angiotensin-converting-enzyme inhibition compared with conventional therapy on cardiovascular morbidity and mortality in hypertension: the Captopril Prevention Project (CAPPP) randomised trial. The Lancet. 1999;353(9153):611-6.

7. Peto R. Failure of randomisation by “sealed” envelope. The Lancet. 1999;354(9172):73.

8. Kennedy AD, Torgerson DJ, Campbell MK, Grant AM. Subversion of allocation concealment in a randomised controlled trial: a historical case study. Trials. 2017;18(1):204.

9. Hewitt C, Hahn S, Torgerson DJ, Watson J, Bland JM. Adequacy and reporting of allocation concealment: review of recent trials published in four general medical journals. BMJ. 2005;330(7499):1057-8.

10. Clark L, Fairhurst C, Torgerson DJ. Allocation concealment in randomised controlled trials: are we getting better? Bmj. 2016;355:i5663.

11. Pildal J, Hrobjartsson A, Jørgensen K, Hilden J, Altman D, Gøtzsche P. Impact of allocation concealment on conclusions drawn from meta-analyses of randomized trials. International journal of epidemiology. 2007;36(4):847-57.

12. Hopewell S, Dutton S, Yu L-M, Chan A-W, Altman DG. The quality of reports of randomised trials in 2000 and 2006: comparative study of articles indexed in PubMed. Bmj. 2010;340:c723.

13. Higgins JP, Green S. Cochrane handbook for systematic reviews of interventions. 2008.

14. Wan X, Wang W, Liu J, Tong T. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. BMC Medical Research Methodology. 2014;14(1):135.

15. Altman DG, Bland JM. How to obtain the P value from a confidence interval. Bmj. 2011;343:d2304.

16. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria. 2017.

17. Wickham H. ggplot2: Elegant Graphics for Data Analysis: Springer-Verlag New York; 2016.

18. Clark L, Fairhurst C, Hewitt CE, Birks Y, Brabyn S, Cockayne S, et al. A methodological review of recent meta-analyses has found significant heterogeneity in age between randomized groups. Journal of Clinical Epidemiology. 2014;67(9):1016-24.

19. Trowman R, Dumville JC, Torgerson DJ, Cranny G. The impact of trial baseline imbalances should be considered in systematic reviews: a methodological case study. Journal of Clinical Epidemiology. 2007;60(12):1229-33.

20. Chan A-W, Altman DG. Epidemiology and reporting of randomised trials published in PubMed journals. The Lancet. 2005;365(9465):1159-62.

**Conflict of interest**

We have no conflicts of interest to declare.

**Tables**

**Table 1:** Number of extracted articles for each journal, presented by year of publication.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **2017**  **(n=160)** | **2018**  **(n=178)** | **Total**  **(n=338)** |
| **BMJ** | 7 (4.4) | 12 (6.7) | 19 (5.6) |
| **The Lancet** | 48 (30.0) | 53 (29.8) | 101 (29.9) |
| **JAMA** | 48 (30.0) | 44 (24.7) | 92 (27.2) |
| **NEJM** | 57 (35.6) | 69 (38.8) | 126 (37.3) |

**Table 2:** The likelihood of reporting a statistically significant result at the 5% level for the primary outcome, for each pairwise quality of allocation concealment comparison.

|  |  |  |
| --- | --- | --- |
|  | **OR (95% CI)** | **P-value** |
| **Inadequate vs Adequate** | 2.14 (1.35, 3.40) | 0.001 |
| **Inadequate vs Good** | 1.82 (0.78, 4.27) | 0.168 |
| **Adequate vs Good** | 0.85 (0.52, 1.38) | 0.508 |
| **Inadequate vs Unclear** | 4.56 (2.74, 7.59) | <0.0005 |
| **Adequate vs Unclear** | 2.13 (1.36, 3.33) | 0.001 |
| **Good vs Unclear** | 2.50 (1.04, 6.05) | 0.042 |

**Figure legends**

**Figure 1:** A kernel density plot showing the distribution of the logit of the p-values by quality of allocation concealment. The vertical dashed lines represent the mean value for each category.

