Short Communication

Multiplicity adjustments in parallel-group multi-arm trials sharing a control group: Clear guidance is needed

Síle F. Molloy [a],[*],[1], Ian R. White [b],[1], Andrew J. Nunn [b], Richard Hayes [c], Duolao Wang [d], Thomas S. Harrison [a]

[a] Institute for Infection and Immunity, St George's University of London, London, UK
[b] MRC Clinical Trials Unit at UCL, Institute of Clinical Trials and Methodology, University College London, London, UK
[c] Department of Infectious Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK
[d] Global Health Trials Unit, Department of Clinical Sciences, Liverpool School of Tropical Medicine, Liverpool, UK

A B S T R A C T

Multi-arm, parallel-group clinical trials are an efficient way of testing several new treatments, treatment regimens or doses. However, guidance on the requirement for statistical adjustment to control for multiple comparisons (type I error) using a shared control group is unclear. We argue, based on current evidence, that adjustment is not always necessary in such situations. We propose that adjustment should not be a requirement in multi-arm, parallel-group trials testing distinct treatments and sharing a control group, and we call for clearer guidance from stakeholders, such as regulators and scientific journals, on the appropriate settings for adjustment of multiplicity.

Multiplicity is a major consideration in the analysis of clinical trials. It occurs when multiple significance tests are carried out, increasing the family-wise error rate (FWER), the probability of a "false positive" statistically significant result or type 1 error. Multiplicity can arise for various reasons including use of multiple outcomes, repeated measures, interim analyses, multiple sub-groups, factorial designs and, the focus of this viewpoint, multi-arm clinical trials. It is widely accepted that control of multiplicity is essential in many situations, for example, early stopping of a trial based on interim analyses, or where more than one hypothesis is being tested within a family of hypotheses [1,2]. If not handled correctly, unsubstantiated claims for the effectiveness of a drug may be made due to an inflated rate of false positive conclusions, and this is especially important in confirmatory trials for licensed medications [1,3]. Various methods of control have been developed including hierarchical procedures, the Bonferroni Method and Dunnett's test [4]. However, if such adjustments are applied unnecessarily, potentially effective treatments may be discarded prematurely [5].

Multi-arm trial designs are valuable in clinical research. They allow a number of new treatments, or varying treatment regimens/doses, to be tested within a single trial, increasing efficiency and reducing costs associated with conducting several independent trials. A three-arm trial,

for instance, reduces the sample size that would be required for two independent trials by 25% due to efficient sharing of the control group. Some 20% of superiority trials registered in 2010–2012 had more than two groups [6]. However, there appears no consensus, across stakeholders such as regulators and scientific journals, on the necessity to control for a potentially inflated type 1 error rate when comparing distinct treatments to a shared control group in confirmatory parallel-group multi-arm trials [7] and this has become the subject of much recent debate among statisticians and trialists [5,8–11]. Many guidelines on multiplicity do not refer specifically to multi-arm trials [1–3], resulting in inconsistencies in the application of adjustment methods in published articles. A 2014 review found that 49% of published multi-arm trials reported using a multiple-testing adjustment, with adjustment more common in trials evaluating multiple doses or regimens of the same treatments (67%), but surprisingly there was little evidence of difference in adjustment between exploratory and confirmatory trials [10].

There is a general consensus that for multi-arm exploratory trials stringent multiple-testing adjustment is not required [10] as doing so may drop potentially effective treatments too early in the assessment process. Conversely, many authors agree with current guidance from the

---

FDA and EMA that for confirmatory trials where arms represent several doses or regimens of the same treatment, adjustments for multiplicity should be applied [5,12–15]. However, the literature is unclear on the necessity of adjustment in confirmatory trials where the different arms represent distinct treatments which are compared with a shared control [10]. A number of authors argue that adjustment is not always necessary, particularly where the results are not combined into one final conclusion and decision [10,16–18], with Parker et al. arguing that non-adjustment should be the default starting point in such situations [15]. By contrast, guidance from the New England Journal of Medicine requires adjustment in this scenario, even for exploratory analyses, though no rationale is given for applying the same rules to all types of multiplicity [2].

In many cases where adjustment for multiple testing is required, the tests are correlated: for example, when tests of multiple outcomes in the same trial are correlated. Correlation also arises between the multiple tests in a multi-arm trial when they share a common control group. This correlation has practical implications: it makes Dunnett's test the best way to control for multiplicity, it reduces the impact of multiplicity on the FWER [5,19,20], and it increases the probability of making multiple errors given that at least one error is made [11]. However, the correlation does not inform whether multiplicity should be controlled [5].

The key issue in determining whether to control for multiplicity is whether multiple tests are conceptually related: that is, how separate are the scientific questions or the claims to be made? If an issue of a journal publishes multiple clinical trials relating to different medical areas, then the overall type 1 error rate is increased, but no-one would suggest controlling for multiplicity. If multiple doses of the same drug are tested, on the other hand, a claim of efficacy of the drug could be made if any one dose shows benefit, so multiplicity should be controlled: this is true whether the doses are tested in the same trial or in separate trials. Similarly, if multiple treatments are investigated for the purpose of a single regulatory submission then this may be a reason to control the FWER [21]. If drugs with different mechanisms of action are evaluated in the same trial, we believe that control for multiplicity is not required, just as if they were evaluated in separate trials. Hence the "family" over which FWER should be controlled is usually a treatment, and the difficult question is whether closely related treatments should be included in the same family: for example, drugs of the same class, or similar multi-drug regimens [21].

An alternative way to account for multiple tests is to control the false discovery rate (FDR), the expected proportion of rejected null hypotheses that are actually true. The FDR is similar to the FWER in studies with small numbers of experimental treatments (e.g. up to three), but is less stringent than the FWER with larger numbers of experimental treatments (e.g. five or more) [22]. Controlling the FDR may be done using Benjamini–Hochberg procedures [22,23]. When multiple drugs are successful, the FDR has the advantage that it represents the expected proportion of inefficacious drugs among the successful drugs. Wason et al. recommend that sponsors and trialists consider use of the FDR for multi-arm trials testing distinct treatment arms [22], while others suggest the FDR as an appropriate control measure in the context of trials with a large number of treatment arms [24]. However, trials with many treatments may be less likely to have distinct treatments.

We have focussed on the parallel-group multi-arm trial design, but various perspectives on the need for multiplicity adjustments in more complex designs such as basket, umbrella and platform trials are also being debated [14,24–28]. Such new and adaptive designs are increasingly important, especially in the case of COVID-19 where adaptive, cost-effective and rapid trials are critical [24]. However, as discussed by Collignon et al. [25], these complex designs have raised challenging statistical questions around the need for control of multiple testing when adding arms or drugs over time.

In conclusion, increasing trial complexity makes addressing multiplicity more complex but also more important. Clearer guidance for trialists from stakeholders, such as regulators and scientific journals, on the appropriate settings for adjustment of multiplicity is required. We agree with others that the need to adjust or not should be well justified based on the complexity of design and the specific setting and objectives of each trial [15,21] and that control of the FDR should be considered for trials testing a large number of treatments [22]. We propose that, for simple parallel-group multi-arm trials of distinct treatments with a shared control, adjustment should not be a requirement. However, further clarity is needed to define what are distinct treatments [21,25].

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

References

[1] Food and Drug Administration, Multiple Endpoints in Clinical Trials Guidance for Industry 2017, Available from, https://www.fda.gov/regulatory-information/search-fda-guidance-documents/multiple-endpoints-clinical-trials-guidance-industry, 11 August 2020.

[2] D. Harrington, R.B. D'Agostino, C. Gatsonis, J.W. Hogan, D.J. Hunter, S.-L. T. Normand, et al., New guidelines for statistical reporting in the journal, N. Engl. J. Med. 381 (2019) 285–286.

[3] European Medicines Agency, Guideline on Multiplicity Issues in Clinical Trials 2017, Available from, https://www.ema.europa.eu/en/documents/scientific-guideline/draft-guideline-multiplicity-issues-clinical-trials_en.pdf, 11 August 2020.

[4] Food and Drug Administration, Adaptive Designs for Clinical Trials of Drugs and Biologics: Guidance for Industry 2019, Available from, https://www.fda.gov/media/78495/download, 11 August 2020.

[5] D.R. Howard, J.M. Brown, S. Todd, W.M. Gregory, Recommendations on multiple testing adjustment in multi-arm trials with a shared control group, Stat. Methods Med. Res. 27 (5) (2018) 1513–1530.

[6] M.K. Parmar, J. Carpenter, M.R. Sydes, More multiarm randomised trials of superiority are needed, Lancet. 384 (9940) (2014) 283–284.

[7] G. Baron, E. Perrodeau, I. Boutron, P. Ravaud, Reporting of analyses from randomized controlled trials with multiple arms: a systematic review, BMC Med. 11 (1) (2013) 84.

[8] J. Wason, D. Magirr, M. Law, T. Jaki, Some recommendations for multi-arm multi-stage trials, Stat. Methods Med. Res. 25 (2) (2016) 716–727.

[9] O. Collignon, C.F. Burman, M. Posch, A. Schiel, Collaborative platform trials to fight COVID-19: methodological and regulatory considerations for a better societal outcome, Clin. Pharmacol. Ther. 110 (2) (2021) 311–320.

[10] J.M.S. Wason, L. Stecher, A.P. Mander, Correcting for multiple-testing in multi-arm trials: is it necessary and is it done? Trials. 15 (1) (2014) 364.

[11] M.A. Proschan, D.A. Follmann, Multiple comparisons with control in a single experiment versus separate experiments: why do we feel differently? Am. Stat. 49 (2) (1995) 144–149.

[12] R. Bender, S. Lange, Adjusting for multiple testing—when and how? J. Clin. Epidemiol. 54 (4) (2001) 343–349.

[13] J.M.S. Wason, T. Jaki, N. Stallard, Planning multi-arm screening studies within the context of a drug development program, Stat. Med. 32 (20) (2013) 3424–3435.

[14] N. Stallard, S. Todd, D. Parashar, P.K. Kimani, L.A. Renfro, On the need to adjust for multiplicity in confirmatory clinical trials with master protocols, Ann. Oncol. 30 (4) (2019) 506–509.

[15] R.A. Parker, C.J. Weir, Non-adjustment for multiple testing in multi-arm trials of distinct treatments: rationale and justification, Clin. Trials. 17 (5) (2020) 562–566.

[16] K.J. Rothman, No adjustments are needed for multiple comparisons, Epidemiol. (Cambridge, Mass). 1 (1) (1990) 43–46.

[17] B. Freidlin, E.L. Korn, R. Gray, A. Martin, Multi-arm clinical trials of new agents: some design considerations, Clin. Cancer Res. 14 (14) (2008) 4368–4371.

[18] M.A. Proschan, M.A. Waclawiw, Practical guidelines for multiplicity adjustment in clinical trials, Control. Clin. Trials 21 (6) (2000) 527–539.

[19] N. Fernandes, A. Stone, Multiplicity adjustments in trials with two correlated comparisons of interest, Stat. Methods Med. Res. 20 (6) (2011) 579–594.

[20] C.W. Dunnett, New tables for multiple comparisons with a control, Biometrics. 20 (3) (1964) 482–491.

[21] F. Bretz, F. Koenig, Commentary on Parker and Weir, Clin. Trials. 17 (5) (2020) 567–569.

[22] J.M.S. Wason, D.S. Robertson, Controlling type I error rates in multi-arm clinical trials: a case for the false discovery rate, Pharm. Stat. 20 (1) (2021) 109–116.

[23] Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing, J. R. Stat. Soc. Ser. B Methodol. 57 (1) (1995) 289–300.

[24] O. Collignon, C.-F. Burman, M. Posch, A. Schiel, Collaborative platform trials to fight COVID-19: methodological and regulatory considerations for a better societal outcome, Clin. Pharmacol. Ther. 110 (2) (2021) 311–320.

[25] O. Collignon, C. Gartner, A.-B. Haidich, R. James Hemmings, B. Hofner, F. Pétavy, et al., Current statistical considerations and regulatory perspectives on the planning of confirmatory basket, umbrella, and platform trials, Clin. Pharmacol. Therap. 107 (5) (2020) 1059–1067.

[26] S.M. Berry, Potential statistical issues between designers and regulators in confirmatory basket, umbrella, and platform trials, Clin. Pharmacol. Ther. 108 (3) (2020) 444–446.

[27] C.C. Lu, X.N. Li, K. Broglio, P. Bycott, Q. Jiang, X. Li, et al., Practical considerations and recommendations for master protocol framework: basket, umbrella and platform trials, Therap. Innov. Regul. Sci. 55 (6) (2021) 1145–1154.

[28] X. Bai, Q. Deng, D. Liu, Multiplicity issues for platform trials with a shared control arm, J. Biopharm. Stat. 30 (6) (2020) 1077–1090.