been had the study reported 15 unrelated outcomes. Regardless of statistical values, absolute numbers/percentages are shown in *Table 3* and are of clear clinical significance (a reduction in the overall complication rate from 55 to 28 per cent, a reduction in the infective complication rate from 37 to 20 per cent, and a reduction in the surgical-site infection rate from 23 to 10 per cent). This was, however, a relatively small study with approximately 100 patients in each group, and clearly not powered to detect significant differences in less frequently observed complications including deep surgical-site infections and anastomotic leaks.

As the above outcomes of interest are likely to be related, a Bonferroni correction is perhaps an overly conservative way of correcting for multiple testing. Given the interrelationship of our outcomes, a different analysis to correct for multiple testing such as the Benjamini–Hochberg procedure may be more appropriate[3]. Indeed, we have now carried out such a *post hoc* analysis. Using this correction, all of the outcomes reported as statistically significant in *Table 3* remained so when the false discovery rate was set at 5 per cent, and the majority remained statistically significant when the false discovery rate was sent at 10 per cent. Therefore, the suggestion of '*P*-hacking' is unlikely to be the case and is supportive of the peer review process.

Hartrick and colleagues state in their letter that the use of oral antibiotics and mechanical bowel preparation in resectional colorectal surgery is an important issue requiring further prospective research in the form of large prospective RCTs. As acknowledged in the final paragraph of the Discussion section of our article ('This strategy is worthy of further investigation'), we are in clear agreement. Indeed, we look forward to the reporting of those trials currently underway, in particular the COLONPREP trial (EudraCT no. 2017-002542-72). This is of particular interest given the recent negative findings of the MOBILE trial[4], contrary to most of the published meta-analyses[5–7] in the field.

Disclosure

The authors declare no conflict of interest.

A. M. Golder and co-authors
Academic Unit of Surgery, University of Glasgow, Level 2, New Lister Building, Glasgow Royal Infirmary, Glasgow G31 2ER, UK
allan.golder@glasgow.ac.uk

1 Golder AM, Steele CW, Conn D, MacKay GJ, McMillan DC, Horgan PG *et al*. Effect of preoperative oral antibiotics in combination with mechanical bowel preparation on inflammatory response and short-term outcomes following left-sided colonic and rectal resections. *BJS Open* 2019; **3**: 830–839.
2 Watt DG, McSorley ST, Park JH, Horgan PG, McMillan DC. A postoperative systemic inflammation score predicts short- and long-term outcomes in patients undergoing surgery for colorectal cancer. *Ann Surg Oncol* 2017; **24**: 1100–1109.
3 Glickman ME, Rao SR, Schultz MR. False discovery rate control is a recommended alternative to Bonferroni-type adjustments in health studies. *J Clin Epidemiol* 2014; **67**: 850–857.
4 Koskenvuo L, Lehtonen T, Koskensalo S, Rasilainen S, Klintrup K, Ehrlich A *et al*. Mechanical and oral antibiotic bowel preparation *versus* no bowel preparation for elective colectomy (MOBILE): a multicentre, randomised, parallel, single-blinded trial. *Lancet* 2019; **394**: 840–848.
5 McSorley ST, Steele CW, McMahon AJ. Meta-analysis of oral antibiotics, in combination with preoperative intravenous antibiotics and mechanical bowel preparation the day before surgery, compared with intravenous antibiotics and mechanical bowel preparation alone to reduce surgical-site infections in elective colorectal surgery. *BJS Open* 2018; **2**: 185–194.
6 Rollins KE, Javanmard-Emamghissi H, Acheson AG, Lobo DN. The role of oral antibiotic preparation in elective colorectal surgery: a meta-analysis. *Ann Surg* 2019; **270**: 43–58.
7 Koullouros M, Khan N, Aly EH. The role of oral antibiotics prophylaxis in prevention of surgical site infection in colorectal surgery. *Int J Colorectal Dis* 2017; **32**: 1–18.

[Correction added on 17 April 2020, after first online publication: The article title was previously missing and has been inserted in this current version.]

Cluster-randomized crossover trial of chlorhexidine–alcohol *versus* iodine–alcohol for prevention of surgical-site infection (SKINFECT trial)

DOI: 10.1002/bjs5.50285

We read with interest the work of Charehbili and colleagues[1], which 'aimed to investigate whether there is a superiority of chlorhexidine–alcohol over iodine–alcohol for preventing SSI'.

This cluster-randomized crossover trial was conducted in five hospitals and 3665 patients were included. The authors found that the incidence of surgical-site infection (SSI) was not different between the groups: 3·8 per cent among patients in the chlorhexidine–alcohol group *versus* 4·0 per cent in those in the iodine–alcohol group (odds ratio 0·96, 95 per cent c.i. 0·69 to 1·35).

We commend the authors for performing this interesting study, as these results are useful for the choice of the most appropriate preoperative antiseptic. However, we have several statistical suggestions and queries that we would like to communicate to the authors.

The authors concluded that 'Preoperative skin disinfection with chlorhexidine–alcohol is similar to that for iodine–alcohol with respect to reducing the risk of developing an SSI'. This may be due to an underpowered study.

In fact, sample size was estimated by simulation. Although this approach is

efficient, the authors do not provide enough details on the parameters they used. Thus it is not easy to replicate calculations. As mentioned by the CONSORT statement[2]: 'the reports of cluster randomized trials should state the assumptions used when calculating the number of clusters and the cluster sample size'.

The authors mention R software for the simulations that led to the final estimation of sample size. But several R packages are available and one may suppose that a package such as clusterPower was used[3]. Not knowing which package was used does not permit the analysis to be replicated. Moreover, algorithms used may vary between packages and lead to different estimations of sample size.

In a cluster-randomized crossover trial, a sequence of interventions is assigned to a cluster (group) of individuals. Each cluster receives each intervention in a separate period of time and this leads to 'cluster periods'[4]. There is usually a correlation between patients in the same cluster. In addition, within a cluster, patients within the same period may be more similar to one another than to patients in other periods[5].

In a cluster-randomized crossover trial, the sample size estimated by not taking into account the above-mentioned features must be multiplied by a defined inflation factor. The latter can be approximated by $(1+(n-1)\rho)-\eta$[6,7], where n is the average number of patients in a cluster during one of the periods, ρ is the intraclass correlation (ICC), and η the interperiod correlation. See, for example, Turner *et al.*[8] or Moerbeek and Teerenstra[9] (p. 94) for other approaches to estimate sample size in this context.

Parameters ρ and η can be retrieved from literature or estimated using assumptions or approximations[10] (p. 203), for instance: the logarithm of the ICC can be approximated by the logarithm of the prevalence of disease (here, the SSI rate)[11]; the interclass correlation is intrinsically lower than the ICC[12].

Data were analysed using a multilevel model, which is appropriate. Treatment period was considered as a fixed effect and hospitals as random effect. Treatment period could also be considered as a random effect. In their simulations, Morgan *et al.*[5] actually demonstrated that 'hierarchical models without random effects for period-within-cluster, which do not account for any extra within-period correlation, performed poorly with greatly inflated Type I errors in many scenarios'.

The authors did not report variance components of outcomes: within- and between-participant variance, the ICC, as recommended by some authors[13].

In a cluster-randomized crossover trial, three components of variation are available: variation in cluster mean response; variation in the cluster period mean response; and variation between individual responses within a cluster period[4]. Small changes in the specification of the within-cluster–within-period correlation, or the within-cluster–between-period correlation, can increase the required number of clusters[4]. Thus, as the above-mentioned correlation parameters were not reported by Charehbili *et al.*[1], the number of clusters required may be larger than that used in the study.

A simulation study showed an association between an increase in cluster size variability and a decrease in statistical power[14]. The authors did not address this point.

In summary, the results of this study are interesting, but readers should interpret them with caution, according to the statistical methods used for design and analysis of cluster-randomized crossover trials.

Disclosure

The authors declare no conflict of interest.

L. S. Aho Glélé[1], P. Ortega-Deballon[2], A. Guilloteau[1], O. Keita-Perse[4], K. Astruc[1] and D. Lepelletier[3]
[1]Epidemiology and Infection Control Department and [2]Digestive Surgery Department, Dijon University Hospital, 14 Rue Paul Gaffarel, 21000 Dijon, and [3]Epidemiology and Infection Control Department, Nantes University Hospital, Nantes, France, and [4]Epidemiology and Infection Control Department, Monaco Hospital, Monaco
ludwig.aho@chu-dijon.fr

1 Charehbili A, Koek MBG, de Mol van Otterloo JCA, Bronkhorst MWGA, van der Zwaal P, Thomassen B *et al.* Cluster-randomized crossover trial of chlorhexidine–alcohol *versus* iodine–alcohol for prevention of surgical-site infection (SKINFECT trial). *BJS Open* 2019; **3**: 617–622.
2 Campbell MK, Piaggio G, Elbourne DR, Altman DG, CONSORT Group. Consort 2010 statement: extension to cluster randomised trials. *BMJ* 2012; **345**: e5661.
3 Reich NG, Myers JA, Obeng D, Milstone AM, Perl TM. Empirical power and sample size calculations for cluster-randomized and cluster-randomized crossover studies. *PLoS One* 2012; **7**: e35564.
4 Arnup SJ, McKenzie JE, Hemming K, Pilcher D, Forbes AB. Understanding the cluster randomised crossover design: a graphical illustration of the components of variation and a sample size tutorial. *Trials* 2017; **18**: 381.
5 Morgan KE, Forbes AB, Keogh RH, Jairath V, Kahan BC. Choosing appropriate analysis methods for cluster randomised cross-over trials with a binary outcome. *Stat Med* 2017; **36**: 318–333.
6 Donner A, Birkett N, Buck C. Randomization by cluster. Sample size requirements and analysis. *Am J Epidemiol* 1981; **114**: 906–914.
7 Giraudeau B, Ravaud P, Donner A. Sample size calculation for cluster randomized cross-over trials. *Stat Med* 2008; **27**: 5578–5585.
8 Turner EL, Li F, Gallis JA, Prague M, Murray DM. Review of recent methodological developments in group-randomized trials: part 1 – design. *Am J Public Health* 2017; **107**: 907–915.
9 Moerbeek M, Teerenstra S. *Power Analysis of Trials with Multilevel Data*. CRC Press: Boca Raton, 2015.

10 Machin D, Campbell MJ, Tan SB, Tan SH. *Sample Sizes for Clinical, Laboratory and Epidemiology Studies*. John Wiley: Hoboken, 2018.
11 Gulliford MC, Adams G, Ukoumunne OC, Latinovic R, Chinn S, Campbell MJ. Intraclass correlation coefficient and outcome prevalence are associated in clustered binary data. *J Clin Epidemiol* 2005; **58**: 246–251.
12 Donner A, Klar N, Zou G. Methods for the statistical analysis of binary data in split-cluster designs. *Biometrics* 2004; **60**: 919–925.
13 Arnup SJ, Forbes AB, Kahan BC, Morgan KE, McKenzie JE. The quality of reporting in cluster randomised crossover trials: proposal for reporting items and an assessment of reporting quality. *Trials* 2016; **17**: 575.
14 Lauer SA, Kleinman KP, Reich NG. The effect of cluster size variability on statistical power in cluster-randomized trials. *PLoS One* 2015; **10**: e0119074.