The quality of the reported sample size calculation in clinical trials on COVID-19 patients indexed in PubMed

Paul H. Lee, PhD

School of Nursing, Hong Kong Polytechnic University

Correspondence to: Dr Paul H. Lee (email: paul.h.lee@polyu.edu.hk, phone: +852-3400 8275, fax: +852-2364 9663), School of Nursing, GH527, Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

Manuscript word count: 1,498

Article type: Letter to the Editor

Keywords: protocol; sample size; statistics; trials; SARS-CoV-2

Given the utmost priority of COVID-19 research, many medical journals, especially the leading ones, expedited the review process of these papers. It was expected that the amount of submitted papers for peer review would raise sharply and the deadline of the review period at the COVID-19 outbreak will be tightened. Therefore, sample size calculation of these papers, a component that was being omitted in more than 40% of the published randomized controlled trials,[1] might be neglected during the peer review process. Most importantly, authors of these papers were also rushing to conduct their COVID-19 research and they might not have seek necessary statistical consultation regarding sample size calculation. The CONsolidated Standards Of Reporting Trials (CONSORT) statement recommends trial reports to provide all essential information on the determination of sample size, including the level of statistical significance, the desired level of power, and the estimated effect size of the treatment. However, the degree of compliance to the CONSORT statement regarding sample size calculation of the newly-published COVID-19 trial papers is unknown. Therefore, we reviewed all clinical trials on COVID-19 patients published from 1st January 2020 to 4th April 2020 indexed in PubMed.

We adopted the revised Cochrane highly sensitive search strategy [2] to search PubMed for papers reporting clinical trial involving COVID-19 patients, either randomized or non-randomized, indexed by 4th April 2020. The search term can be found in the supplementary material. A total of 521 papers were identified. Papers were excluded when they were 1) case studies or case series (*n*=25), 2) commentaries, editorials, or letters to the editor (*n*=189), 3) guidelines (*n*=61), 4) observational studies (*n*=108), 5) reviews (*n*=69), 6) studies on the virus instead of the patients (*n*=15), and 7) not about COVID-19 (*n*=50). Finally, four studies were included in this study. Details of these papers and the reason of exclusion are listed in Supplementary Table 2.

The following information, together with the relevant texts in the papers, were extracted: 1) whether the group allocation of patients was randomized, 2) the measurement type of the primary outcome that the sample size calculation based on, 3) the level of significance used, 4) the desired power, and 5) the estimated effect size and the relevant statistics to calculate this effect size.

For papers that provide all essential information to replicate their sample size calculation, including the level of significance, the desired power, and the estimated effect size, we calculated the sample size required to achieve the reported desired power given the level of significance suggested by the authors of these papers. Since all included studies were superiority trials, we will assume that the test was two-tailed if it was not specified in the text. We referred to the standard textbook formulas for calculating the sample size.[3]

Supplementary Table 1 shows the information of the four included papers.[4-7] Three of them were drug trials (hydroxychloroquine and azithromycin,[4] lopinavir and ritonavir,[6] and methylprednisolone[7]) and the remaining one examined the effectiveness of high-flow nasal-oxygenation-assisted fibreoptic tracheal intubation.[5] All four studies were a parallel trial. Three studies reported the level of significance, the desired power, and the estimated effect size,[4,6,7] and we replicated the sample size calculation for these three studies step-by-step in the following paragraphs.

In two trials,[4,7] the targeted level of significance ("*a type I error rate of 5%*" [4] and "*a level of 95% confidence*" [7]) and the desired power ("*a 85% power*"[4] and "*a power of 80%*" [7]) were reported. However, some information for calculating the treatment effect size were missing in these two papers. Zhou *et al*. (2020)[7] stated that "*detecting a reduction of 40% in SOFA scores between the treatment and nontreatment groups*",[7] however the SOFA (or the Sequential Organ Failure Assessment) score is a continuous variable. To calculate the sample

size, the post-treatment between-group mean difference and the pooled standard deviation of the primary outcome variable, or the Cohen's *d* effect size which equals the between-group mean difference divided by the pooled standard deviation, should be estimated. The sample size of 23 per group as stated by the authors ("*Considering a dropout rate of 5%, the sample size is estimated to be 24 cases in each group*") implied that they estimated the Cohen's *d* to be 0.83 (= $(z_{0.025}+z_{0.2})\times\sqrt{(2/n)} = (1.96+0.84)\times\sqrt{(2/23)}$).

In Gautret *et al*. (2020),[4] the authors only reported the effect of the treatment (hydroxychloroquine) group ("*Assuming a 50% efficacy of hydroxychloroquine in reducing the viral load at day 7*" [4]) but the effect of the control group was missing. Note that if we assume the effect of the control group is 0%, which is a common assumption in effect size estimation, the required sample size calculated using Fleiss' formula [8] would be $(z_{0.025}\sqrt{(2(p_1+p_2)(p_1-p_2))} + z_{0.2}\sqrt{(p_1(1-p_1)+p_2(1-p_2))})^2/(p_1-p_2)^2 = (1.96\sqrt{(2(0.5)(0.25))} + 0.84\sqrt{(0.5(1-0.5))})^2/(0.5)^2 = 12$ per group, and the required sample size using Fleiss' formula with continuity correction [8] would be $n(1+\sqrt{(1+4/n(p_1-p_2))})^2/4 = 12(1+\sqrt{(1+4/12(0.5))})^2/4 = 16$ per group. The authors, claiming that Fleiss' formula with continuity correction was used, calculated that 21 subjects per group were required ("*assume …10% loss to follow-up, we calculated that a total of 48 COVID-19 patients (ie, 24 cases in the hydroxychloroquine group and 24 in the control group) would be required for the analysis (Fleiss with CC)*" [4]), which implied that they assumed the effect of the control group would be around 5%, i.e., 1 out of the 24 in the control group would achieve a viral load clearance. This assumption deviated with the usual assumption that the control group has no effect, and the authors should have specify this assumption in the main text.

In the only trial that the all essential information for sample size calculation was provided,[6] the sample size calculated in the paper could not be reproduced. The paper reported that "*8 days in the median time to clinical improvement between the two groups, assuming that the*

median time in the standard-care group was 20 days".[6] Using the reported statistics, the hazard of the treatment (lopinavir-ritonavir) group and control group would be 0.0577 and 0.0347, respectively (hazard = ln(2) / median time of survival). Given the censoring time of 28 days ("*The time to clinical improvement was assessed after all patients had reached day 28, with failure to reach clinical improvement or death before day 28 considered as right-censored at day 28.*"),[6] the percentage of patients reaching clinical improvement should be the average of the percentage of the treatment group ($1-e^{(-28\times0.0577)} = 80.2\%$) and the percentage of the control group ($1-e^{(-28\times0.0347)} = 62.1\%$), that is, 71.1%. Note that this percentage was overestimated in the paper (75%).[6] According to the standard formula, the number of events required to achieve a significance level of 0.05 and a power of 0.8 would be $4\times (z_{0.025}+z_{0.2})^2/(\ln(HR))^2 = 4 \times (1.96+0.84)^2/\ln(0.0577/0.0347)^2 = 121$.[3] To accumulate 121 events, the total sample size required would be 121 / 71.1% = 170, i.e., 85 per group. The estimated sample size reported by the authors was 160, which had a 6% deviation with our calculated sample size, and the power of this sample size would be 78% instead of the desired 80%.

To conclude, the quality of sample size calculation in clinical trials on COVID-19 patients indexed in PubMed was not acceptable. Inappropriate sample size calculation will give a negative impression to its readers and leads to suspicion on the existence of other methodological flaws in the study. Some believed that studies with any sample size could still contribute to science [9] as the published results allow other readers to conduct a meta-analysis to obtain a more accurate estimate of treatment effect when further studies are available. However, in the time of rapid disease outbreak such as COVID-19, researchers are working around the clock to examine the effectiveness of potential treatments, and inappropriate sample size calculation will lead to adverse consequences. If the sample size is overestimated, researchers need to spend unnecessary, additional time to recruit study and

clean the data, so that the release of the results would be delayed. On the other hand, if the sample size was underestimated, the probability of committing a type II error, that is, unable to reject the null hypothesis while the treatment is effective, will be high. A null find will discourage clinicians to use an effective treatment, and further studies are required to support its effectiveness, again resulting in a delay.

Many existing, effective treatments on infections with coronavirus and other RNA-virus have been proposed to test for its effectiveness on COVID-19 patients. Given the fast-growing number of undergoing trials,[10] we expect that many papers on trials for COVID-19 patients will be published very soon. Since the sample size should be determined *a priori* before the recruitment of the first patient, we strongly suggest all research teams include a statistician as a team member or invite a statistician to evaluate the appropriateness of the sample size calculation. It would be irreversible if flaws in sample size calculation are identified when the trial has been completed or the reported has been submitted for peer review.

REFERENCES

1. Lee PH, Tse ACY. The Diagnosis Checking of Statistical Analysis in Randomised Controlled Trials Indexed in PubMed. *Eur J Clin Invest.* 2017;47(11):847-852.
2. Robinson KA, Dickersin K. Development of a highly sensitive search strategy for the retrieval of reports of controlled trials using PubMed. *Int J Epidemiol.* 2002;31:150-153.
3. Donner A. Approaches to sample size estimation in the design of clinical trials--a review. *Stat Med.* 1984;3(3):199-214.
4. Gautret P, Lagier JC, Parola P, et al. Hydroxychloroquine and azithromycin as a treatment of COVID-19: results of an open-label non-randomized clinical trial. *Int J Antimicrob Agents.* 2020:article in press.

5. Wu CN, Xia LZ, Li KH, et al. High-flow nasal-oxygenation-assisted fibreoptic tracheal intubation in critically ill patients with COVID-19 pneumonia: a prospective randomised controlled trial. *Br J Anaesth.* 2020(article in press).

6. Cao B, Wang Y, Wen D, et al. A Trial of Lopinavir-Ritonavir in Adults Hospitalized with Severe Covid-19. *New Engl J Med.* 2020:article in press.

7. Zhou YH, Qin YY, Lu YQ, et al. Effectiveness of glucocorticoid therapy in patients with severe novel coronavirus pneumonia: protocol of a randomized controlled trial. *Chin Med J.* 2020:article in press.

8. Fleiss JL, Tytun A, Ury HK. A Simple Approximation for Calculating Sample Sizes for Comparing Independent Proportions. *Biometrics.* 1980;36:343-346.

9. Bacchetti P, McCulloch C, Segal MR. Being 'underpowered' does not make a study unethical. *Stat Med.* 2012;31(29):4138-4139.

10. Zhang Q, Wang Y, Qi C, Shen L, Li J. Clinical trial analysis of 2019-nCoV therapy registered in China. *J Med Virol.* 2020:accepted.