Galante et al. (2021) é uma extensa metanálise, pode lhe ajudar:

Out of all the analysis methods tested on our simulated data, we found that the GLMM approach with a random intercept was often the best analysis approach. Although, the bias was substantial when only three clusters with cell sizes of ten or fewer, especially when the ICC was higher and there were few subjects. When there were only 3 clusters and the cell sizes were 50 or more, the bias of the GLMM was much less, but the type I error rate was inflated. The major problems with the GEE are the inflated type I error rate and convergence failures when there are few clusters. If researchers wish to use a GEE when there are few clusters then we suggest that one of the corrections evaluated by Scott et al. [43] be considered.

When a time trend is adjusted for, it is more robust in general for it to be fitted as a categorical variable rather than as a continuous variable, which we have assumed for the sake of simplicity. The decision of how to adjust for time in the analysis can be informed by knowledge of the trial subject matter at hand; however, we note that current methods for calculating the power/sample size of an SW-CRT do so based on a model that adjusts for time as a categorical variable rather than a continuous one so that the type I error rate is correct [7, 29].

It is widely regarded that the SW-CRT is more powerful than a traditional cluster RCT [3, 7, 10, 11]. Although this has now been proven to be not universally the case [12], we suspect that this belief has contributed to the large number of stepped wedge studies with very few clusters. However, these same studies regularly use either a GEE or GLMM modelling approach for binary outcomes, which we have shown have at least one undesirable statistical property when there are few clusters. We also point out that these simulations reflect an ideal scenario where there are no missing data and the cluster sizes are equal. There is the distinct possibility that the number of clusters required will increase when the situation departs from these ideals or when the analysis increases in complexity, such as when additional random effects terms or interactions are added to the model.

There are also other problems with randomising very few clusters, which apply to SW-CRTs and CRTs alike. As Taljaard et al. point out, results from trials with few clusters may not be generalisable to wider populations [44]. A related concern is that the benefit of randomisation is potentially lost as the balance of known and unknown confounders depends on sufficient numbers of clusters being randomised [44, 45].

In summary we recommend that SW-CRTs with a limited number of clusters and binary outcomes should be analysed using a GLMM. Our strongest recommendation of all is that a cross-sectional SW-CRT with three steps should not randomise fewer than six clusters and that when few clusters are available there needs to be a large number of subjects per cluster per time.

Barker 2017 – é um estudo de simulação, pode ajudar a prever potenciais problemas por utilizar apenas dois clusters e um pequeno número de respondentes.

Christopher 2018 – bastante similar ao seu estudo, pode ser orientativo para futuros procedimentos e análises.

Algumas referências úteis que estão neste repositório:

Barker, D., D’Este, C., Campbell, M. J., & McElduff, P. (2017). Minimum number of clusters and comparison of analysis methods for cross sectional stepped wedge cluster randomised trials with binary outcomes: A simulation study. *Trials*, *18*(1). https://doi.org/10.1186/s13063-017-1862-2

Barker, D., McElduff, P., D’Este, C., & Campbell, M. J. (2016). Stepped wedge cluster randomised trials: A review of the statistical methodology used and available. *BMC Medical Research Methodology*, *16*(1). https://doi.org/10.1186/s12874-016-0176-5

Brown, C. A., & Lilford, R. J. (2006). The stepped wedge trial design: A systematic review. In *BMC Medical Research Methodology* (Vol. 6). https://doi.org/10.1186/1471-2288-6-54

Christopher, M. S., Hunsinger, M., Goerling, L. R. J., Bowen, S., Rogers, B. S., Gross, C. R., Dapolonia, E., & Pruessner, J. C. (2018). Mindfulness-based resilience training to reduce health risk, stress reactivity, and aggression among law enforcement officers: A feasibility and preliminary efficacy trial. *Psychiatry Research*, *264*, 104–115. https://doi.org/10.1016/j.psychres.2018.03.059

Galante, J., Friedrich, C., Dawson, A. F., Modrego-Alarcón, M., Gebbing, P., Delgado-Suárez, I., Gupta, R., Dean, L., Dalgleish, T., White, I. R., & Jones, P. B. (2021). Mindfulness-based programmes for mental health promotion in adults in nonclinical settings: A systematic review and meta-analysis of randomised controlled trials. *PLoS Medicine*, *18*(1). https://doi.org/10.1371/journal.pmed.1003481

Golden, M. R., Kerani, R. P., Stenger, M., Hughes, J. P., Aubin, M., Malinski, C., & Holmes, K. K. (2015). Uptake and Population-Level Impact of Expedited Partner Therapy (EPT) on Chlamydia trachomatis and Neisseria gonorrhoeae: The Washington State Community-Level Randomized Trial of EPT. *PLoS Medicine*, *12*(1), 1–22. https://doi.org/10.1371/journal.pmed.1001777

Hargreaves, J. R., Copas, A. J., Beard, E., Osrin, D., Lewis, J. J., Davey, C., Thompson, J. A., Baio, G., Fielding, K. L., & Prost, A. (2015). Five questions to consider before conducting a stepped wedge trial. In *Trials* (Vol. 16, Issue 1). BioMed Central Ltd. https://doi.org/10.1186/s13063-015-0841-8

Hemming, K., Haines, T. P., Chilton, P. J., Girling, A. J., & Lilford, R. J. (2015). The stepped wedge cluster randomised trial: Rationale, design, analysis, and reporting. *BMJ (Online)*, *350*. https://doi.org/10.1136/bmj.h391

Hemming, K., Kasza, J., Hooper, R., Forbes, A., & Taljaard, M. (2020). A tutorial on sample size calculation for multiple-period cluster randomized parallel, cross-over and stepped-wedge trials using the shiny CRT calculator. *International Journal of Epidemiology*, *49*(3), 979–995. https://doi.org/10.1093/IJE/DYZ237

Hemming, K., & Taljaard, M. (2020). Reflection on modern methods: When is a stepped-wedge cluster randomized trial a good study design choice? *International Journal of Epidemiology*, *49*(3), 1043–1052. https://doi.org/10.1093/IJE/DYAA077

Hemming, K., Taljaard, M., McKenzie, J. E., Hooper, R., Copas, A., Thompson, J. A., Dixon-Woods, M., Aldcroft, A., Doussau, A., Grayling, M., Kristunas, C., Goldstein, C. E., Campbell, M. K., Girling, A., Eldridge, S., Campbell, M. J., Lilford, R. J., Weijer, C., Forbes, A. B., & Grimshaw, J. M. (2018). Reporting of stepped wedge cluster randomised trials: Extension of the CONSORT 2010 statement with explanation and elaboration. *BMJ (Online)*, *363*. https://doi.org/10.1136/bmj.k1614

Hooper, R. (2021). Key concepts in clinical epidemiology: Stepped wedge trials. *Journal of Clinical Epidemiology*, *137*, 159–162. https://doi.org/10.1016/j.jclinepi.2021.04.003

Hooper, R., & Eldridge, S. M. (2021). Cutting edge or blunt instrument: How to decide if a stepped wedge design is right for you. In *BMJ Quality and Safety* (Vol. 30, Issue 3, pp. 245–250). BMJ Publishing Group. https://doi.org/10.1136/bmjqs-2020-011620

Hussey, M. A., & Hughes, J. P. (2007). Design and analysis of stepped wedge cluster randomized trials. *Contemporary Clinical Trials*, *28*(2), 182–191. https://doi.org/10.1016/j.cct.2006.05.007

Scott, J. M., Decamp, A., Juraska, M., Fay, M. P., & Gilbert, P. B. (2017). Finite-sample corrected generalized estimating equation of population average treatment effects in stepped wedge cluster randomized trials. *Statistical Methods in Medical Research*, *26*(2), 583–597. https://doi.org/10.1177/0962280214552092

Voldal, E. C., Hakhu, N. R., Xia, F., Heagerty, P. J., & Hughes, J. P. (2020). swCRTdesign: An R Package for Stepped Wedge Trial Design and Analysis. *Computer Methods and Programs in Biomedicine*, *196*. https://doi.org/10.1016/j.cmpb.2020.105514