Researchers should convince policy makers to perform a classic cluster randomized controlled trial instead of a stepped wedge design when an intervention is rolled out

ELSEVIER

Journal of Clinical Epidemiology 65 (2012) 1255−1256

Journal of Clinical Epidemiology

Researchers should convince policy makers to perform a classic cluster randomized controlled trial instead of a stepped wedge design when an intervention is rolled out

Daniel Kotz[a,*], Mark Spigt[a,b], Ilja C.W. Arts[c,d], Rik Crutzen[e], Wolfgang Viechtbauer[f]

[a]*Department of General Practice, CAPHRI School for Public Health and Primary Care, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands*

[b]*Department of General Practice, Tromsø University, Romssa universitehta, N-9037 Tromsø, Norway*

[c]*Department of Epidemiology, CAPHRI School for Public Health and Primary Care, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands*

[d]*Department of Epidemiology, CARIM School for Cardiovascular Diseases, Maastricht University, PO Box 616, 6200 MD Maastricht, The Netherlands*

[e]*Department of Health Promotion, CAPHRI School for Public Health and Primary Care, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands*

[f]*Department of Psychiatry and Psychology, MHeNS School for Mental Health and Neuroscience, Maastricht University, PO Box 616, 6200 MD Maastricht, The Netherlands*

Accepted 10 June 2012; Published online 8 September 2012

We thank Mdege et al. [1] for their response to our critical appraisal [2] of the stepped wedge design (SWD). They agree with us that the SWD has many disadvantages compared with the cluster randomized controlled trial (cluster RCT) and that the cluster RCT is preferable in most circumstances. However, they also maintain that the SWD may be useful when the alternative is to conduct no randomized trial at all and that some of the cluster RCT variants that we suggest in our article are in fact variants of the SWD. To clarify the latter and better understand the discussion about the advantages and disadvantages of the SWD compared with the cluster RCT, it is important to point out the key difference between the two designs.

So, what actually differentiates the SWD from the cluster RCT? In our view, it is the repeated measurement of data from all clusters at each step. As we explained in our critical appraisal, other features of the SWD are not unique to this design but can also be part of a cluster RCT: rolling out an intervention to all clusters and sequential implementation of the intervention [2]. We also explained that rolling out the intervention to control clusters after the final data collection will result in a shorter duration to conduct a cluster RCT and, in addition, has the advantage that the intervention will only be rolled out if proven to be effective [2]. The repeated measurement of data from all clusters at each step, which is the key characteristic that differentiates the SWD from the cluster RCT, has two very important drawbacks: it puts a heavy burden on patients, caregivers, and researchers, and increases the risk of contamination and attrition [2]. Thus, in our opinion, the SWD should therefore only be considered in the absence of these two drawbacks.

Then when can the SWD be used instead of the cluster RCT? We agree with Mdege et al. that the SWD could be applied under some circumstances. However, the set of circumstances in which the SWD may be the preferred alternative (i.e., in the absence of the two aforementioned drawbacks) are rather limited. For one thing, all necessary data must be routinely collected at the appropriate time intervals (i.e., at the inclusion of each additional cluster) and in a reliable and valid fashion without additional burden for patients, caregivers, and researchers. Unfortunately, situations in which this applies are rather exceptional, which in our opinion severely limits the applicability of the SWD. Often, the routine collection of data is incomplete and not all necessary data are collected (in particular relevant covariates).

There may be one other situation in which the SWD may be preferable to a cluster RCT: when the number of clusters is extremely low and the researcher has no control over the number of clusters. In both designs, the power increases much more with an increasing number of clusters than with

DOI of original article: 10.1016/j.jclinepi.2012.06.003.

* Corresponding author. Tel.: +31-43-38-82893; fax: +31-43-36-19344.

E-mail address: d.kotz@maastrichtuniversity.nl or http://www.daniel-kotz.de (D. Kotz).

http://dx.doi.org/10.1016/j.jclinepi.2012.06.016

an increasing number of subjects within clusters [3]. Thus, a researcher who is in full control of designing a study would prefer as many clusters as possible. However, when policy makers decide to roll out an intervention, the researcher may not have an influence on the number of clusters. When there are only very few clusters, the SWD may be advantageous because the repeated measurement of data (if possible) may compensate at least to some extent for a lack of power.

We would like to finish this discussion with an appeal to the research community. Researchers can indeed be faced with the decision of policy makers to roll out an intervention to the population despite the lack of evidence of its efficacy and cost-effectiveness from scientific research. Mdege et al. [1] provide an illustrative example. In such a situation, the cluster RCT should still be the preferred choice of design and researchers should do their best to convince policy makers regarding its advantages. In other words, researchers should not advocate the use of the SWD but only use it as a last resort. Here are some arguments that may help to convince policy makers.

Compared with the SWD, the cluster RCT

- will take less time to complete and will therefore provide the policy maker with results about the intervention's efficacy, cost-effectiveness, and potential harms more quickly
- will put a much lower burden on patients, caregivers, and researchers for data collection and will therefore be more feasible and likely to be conducted successfully
- will reduce the risk of contamination and attrition of both patients and caregivers and will therefore yield less biased results.

We hope that this discussion was useful to the research community and provides valuable arguments for choosing the most appropriate design in situations in which policy makers have a strong influence on the implementation of an intervention.

Acknowledgments

This commentary resulted from a meeting of the Journal Club Mesch, which is a discussion group of epidemiologists, statisticians, and primary care researchers from the Maastricht University Medical Centre (www.journalclubmesch.nl). The authors do not have a conflict of interest regarding this manuscript.

References

[1] Mdege ND, Man MS, Taylor (nee Brown) CA, Torgerson DJ. There are some circumstances where the stepped wedge cluster randomised trial is preferable to the alternative: no randomised trial at all. Response to the commentary by Kotz and colleagues. J Clin Epidemiol 2012;65:1253−4.

[2] Kotz D, Spigt M, Arts ICW, Crutzen R, Viechtbauer W. Use of the stepped wedge design cannot be recommended: a critical appraisal and comparison with the classic cluster randomised controlled trial design. J Clin Epidemiol 2012;65:1249−52.

[3] Killip S, Mahfoud Z, Pearce K. What is an intracluster correlation coefficient? Crucial concepts for primary care researchers. Ann Fam Med 2004;2:204−8.