Mdege, N. D., Man, M-S., Taylor, C. A., & Torgerson, D. J. (2012). There are some circumstances where the stepped-wedge cluster randomized trial is preferable to the alternative: no randomized trial at all. Response to the commentary by Kotz and colleagues. Journal of Clinical Epidemiology, 65(12), 1253-1254. https://doi.org/10.1016/j.jclinepi.2012.06.003
There are some circumstances where the stepped-wedge cluster randomized trial is preferable to the alternative: no randomized trial at all.

Response to the commentary by Kotz and colleagues

Noreen D. Mdege\textsuperscript{a,\ast}, Mei-See Man\textsuperscript{a}, Celia A. Taylor (nee Brown)\textsuperscript{b}, David J. Torgerson\textsuperscript{a}

\textsuperscript{a}Department of Health Sciences, York Trials Unit, University of York, Heslington, York YO10 5DD, UK
\textsuperscript{b}College of Medical and Dental Sciences, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK

Accepted 11 June 2012; Published online 9 September 2012

We thank Kotz et al.\textsuperscript{[1]} for the interesting commentary on our article\textsuperscript{[2]}. We agree with the authors in a number of areas. First, there is no question that in terms of the hierarchy of evidence, a classic cluster randomized controlled trial (CRCT) is preferred to the stepped wedge CRCT for many of the reasons the authors outline. Second, in an ideal world, an unevaluated intervention should not be given to all potential recipients because we do not know whether it is effective. However, there are some circumstances where the stepped wedge CRCT is preferable to the alternative: no randomized trial at all.

Perhaps, we can illustrate our argument through a real-life example as partly described by Pearson et al.\textsuperscript{[3]} In this study, two areas were about to implement a novel training package for probation officers dealing with offenders. The policy makers were convinced that the intervention was effective based on a small study using historical controls. The researchers were unconvinced by the evidence and argued strongly for a classic “waiting list” cluster trial, where the control clusters would receive the intervention after the final data collection. The policy makers did not see the conundrum of spending public funds on a potentially ineffective intervention. They had funding to implement the intervention immediately: implementation could not wait. Policy makers in one area refused any form of randomization and implemented the training package straight away, only allowing a “before and after” evaluation. However, policy makers in the remaining area did not have sufficient resources to implement the training package to all officers immediately. They decided to implement the training package in a sequential fashion and agreed to determine the order with which the officers would get the training package using random allocation. There was no additional burden of data collection as the main outcome, reoffending, was routinely collected on the Police National Computer. The ethical problem of implementing an unproven intervention to all is correct. However, policy makers did not accept this view and were going to implement the training package to all potential recipients anyway. In this case, the results of the stepped wedge trial could still be used to either encourage or prevent further implementation to other parts of the country—enabling an evidence base to be built when none of the alternatives suggested by Kotz et al.\textsuperscript{[1]} were feasible. When undertaking policy-type trials, researchers work in a different environment, politically, compared with evaluating usual health care treatments: politicians do not generally get involved with whether aspirin is effective or not. Consequently, in this instance, a stepped wedge trial was the only way the researchers could have achieved a robust evaluation.

We maintain our position, therefore, that in some circumstances the stepped wedge method is the only way to generate useful evaluation data. Furthermore, some of the suggested alternatives by Kotz et al.\textsuperscript{[1]} seem to us to be the versions of the stepped wedge. Thus, whether their second example shown in Fig. 1 here (figure 3 in Kotz et al.\textsuperscript{[1]}) is a “classic” CRCT is highly debatable. The only difference between this particular example and the stepped wedge design is that data are collected for some but not all clusters at each point where new cluster(s) receive the intervention. However, if the primary outcome is obtained from routinely collected data, it would be advantageous to use data from all the clusters at each point where new cluster(s) receive the intervention. This also applies to their third example illustrated in Fig. 2 here (figure 4 in Kotz et al.\textsuperscript{[1]}) where the only difference between this and the stepped wedge design is that there are no data collection after the last three clusters have received the intervention. In the trial described by Pearson et al.\textsuperscript{[3]} had they used the designs described in Fig. 1 or Fig. 2, they would have been “throwing away” data. The only advantage to Fig. 2 is that...
if the intervention had been harmful, half the clusters would not have received it. In the Pearson et al. [3] evaluation, non-implementation was not possible. Consequently, the stepped wedge design was superior both scientifically and ethically, as more data could have been obtained to inform whether or not other areas should implement the program.

From our review [2], we found that the stepped wedge design is used particularly “for interventions that have been shown to be effective in more controlled research settings (implementation research).” This indicates that researchers are aware of the fact that the design comes lower than standard randomized controlled trials in terms of hierarchy of evidence. It is not always that the design is used without evidence of effectiveness as Kotz et al. [1] seem to suggest when they say the most important disadvantage is an intervention is implemented in all clusters of the stepped wedge design while it has not yet been proven effective. There might be evidence from trials including classic CRCT and in such cases the stepped wedge design is sometimes used as an evaluative design when the intervention is actually being rolled out as routine practice.

In our article, we stated that for a stepped wedge design at the end of the study, all clusters would have received the intervention. In their counter argument, Kotz et al. [1] state that it is also possible to roll out an intervention to all individuals in a classic CRCT, where the control clusters can receive the intervention after the final data collection. Although this is true, this classic CRCT design would require more clusters than the stepped wedge design, which might be difficult if the finances are constrained. Sometimes the number of available clusters available is very limited and this may exclude the use of a classic CRCT. In the trial by Pearson et al. [3], there were only six clusters available, so the stepped wedge was more powerful than the waiting list approach. Moreover, where there are financial constrains, such as in the Pearson evaluation, or logistical problems, this might be a challenge because the intervention is implemented in more clusters at the start of the trial.

Kotz et al. [1] elaborate more on some of the drawbacks of the stepped wedge design that we discussed in our original review that are: (1) it involves data collection at each point where a new cluster receives the intervention. If the study is not using routinely or easily collected data, the cost of data collection can be substantial; (2) the design may require longer trial duration than parallel designs [1,2]. They also highlight the risk of contamination and attrition, with the possibility of patients moving from the late implementation cluster to the early implementation cluster if the intervention is believed to be superior to control. However, as they point out, this problem is not unique to the stepped wedge design. It may also occur in a classic CRCT.

We agree with Kotz et al. [1] that a classic CRCT is preferable in most circumstances [1]. However, we disagree with their view that the use of the stepped wedge design cannot be recommended because there are instances when a stepped wedge design is the best “acceptable” alternative to not doing a trial.

References