Better evidence for programmatic approaches to routine syphilis testing among men living with HIV: Does a stepped wedge trial provide the answer?

Jennifer A Thompson*[a] & Suzanna C Francis [a]

* Corresponding author: Jennifer.thompson@lshtm.ac.uk

[a] Department of Infectious Disease Epidemiology, London School of Hygiene & Tropical Medicine, London, UK

Running title: Stepped wedge trials as pragmatic evaluations

Keywords: stepped wedge cluster randomised trial, pragmatic evaluation, risk of bias, level of evidence, study design

Abstract

Evidence of the real world impact of programmatic interventions often requires compromise. We explore whether a stepped wedge trial can provide a balance of pragmatism and robust design to assess the impact of programmatic interventions, and we appraise a trial of routine syphilis testing.

Article

Pragmatic evidence of the real world impact of programmatic interventions often requires compromise. Where it is possible to incorporate randomisation in to an evaluation, this can reduce the chance of confounding the intervention effect with other factors that affect the outcome. But while a highly controlled randomised trial can provide very strong evidence of intervention impact, this is often inappropriate for implementation science. The drawback is a lower level of evidence: not all evaluations that incorporate randomisation are equal.

Burchell et al's trial of routine syphilis screening demonstrates some of the benefits and limitations of pragmatic trials, particularly the stepped wedge cluster randomised trial [1]. In their trial, every 6 months, one randomly selected HIV clinic from a total of four clinics switched from syphilis testing

based on physician decision to routine testing of syphilis during HIV viral load monitoring. By the end of the trial, all four clinics were using routine testing. Such stepped wedge trial designs can be particularly well suited to such pragmatic evaluations, but are particularly challenging to implement well to provide robust evidence.

In this commentary, we will describe some design features of this trial that highlight strengths and weaknesses of the stepped wedge design. We will conclude with some remarks on how this should affect our interpretation of Burchell et al's results.

Accounting for time trends

In a recent systematic review of testing strategies for sexually transmitted infections among populations at risk, the majority of studies identified reported before and after designs [2]. Burchell et al highlight the limitations of this study design in the context of a strong trend of increasing syphilis incidence. Before and after studies have no way to separate out such time trends from the effect of increased testing for syphilis: is the increased rate of detection due to increasing incidence, or increased testing?

The stepped wedge trial is an improvement on this design as the staggered introduction of the intervention allows researchers to separately model changes over time and the effect of the intervention. For Burchell et al, this improved study design has resulted in very different results: previous before and after studies found a large increase in detection after introduction of routine testing, but Burchell et al found no evidence of a change.

At each time a clinic switches to the intervention, the design uses clinics that have not switched to estimate the change in incidence that is due to trends in the incidence. However, with only four clinics, can we really be sure that the average time trend in three clinics represents the time trend of the fourth when it switches? A clinic with an outlying time trend could cause spurious results. This

becomes less of a problem with a larger number of clusters (clinics) as each contributes relatively less.

A simple "check" compares the trial arms within each 6-month period: the rate of early syphilis detection between clinics that had switched to routine testing and those still using physician lead testing were either identical or lower with routine testing. For more formal analysis, the Hussey and Hughes analysis used by Burchell et al assumes that all clinics have the same change in incidence between each study time period [3]. This is unlikely to be the case for an infectious disease where the timing of outbreaks varying between populations, and failure to account for such variability in the analysis can lead to biased estimates of the effect of the intervention, confidence intervals that are too narrow, and p-values that are too small [4]. Alternative methods such as generalised estimating equations and non-parametric analysis methods take into account such between cluster variability and provide more robust results [5-7].

Risks with a small number of clusters

Using analysis to overcome the issue of time trends is also more challenging in trials with a small number of clusters. With only 4 clinics, it would be difficult to use the suggested alternative methods. Even the model used by Burchell et al requires small sample corrections to provide confidence intervals and p-values the correct size [8]. Such a correction was not reported in the analysis methods, so the estimated p-value is likely to be too small and confidence interval too narrow.

There are additional issues with conducting a trial with such a small number of clusters [9]. The benefit of randomising so few clusters is unclear: it will avoid a systematic bias, but a chance imbalance in the characteristics of the clusters between early and late switching is almost certain. Trials are also likely to be less generalizable. Even if adequate power can be achieved with such a small number of clusters, evaluators may want to consider these broader issues against the additional cost and complexity from increasing the number of clusters.

Risk of bias

Cluster randomised trials of any design are at a higher risk of bias than individually randomised trials [10]. Differential identification and recruitment of participants between the arms is a common concern for cluster randomised trials [11], and blinding in the assessment of outcomes can be more difficult in stepped wedge trials.

It is common in cluster randomised trials for individuals to be recruited into the study after clusters are randomised to either control or intervention conditions. If those recruiting patients are aware of the clusters' allocation, this could alter their identification and recruitment of patients causing an imbalance in the type of patients in the control and intervention conditions. Burchell et al had a low risk from this bias. The majority of the patient population, which is observed under both the control and intervention conditions, is identified before clinics are randomised.

The issue of blinding assessment of outcomes is pertinent to all research studies. When there is subjectivity in determining patient outcomes, those assessing the outcome should be blind to whether the observation was from the control or intervention condition to ensure that the outcome assessment is the same for both conditions. Blinding can be particularly challenging for a stepped wedge trial as assessors must also be blind to the timing of the outcome within the trial to keep them blind to the intervention status. Burchell et al did not report blinding those assessing whether a patient had a diagnosis of syphilis. The direction of bias from unblinded assessment could go either way, or it may be that there is no difference in assessment: blinding avoids the question being asked, and there is little cost or compromise to the pragmatic nature of the trial to blinding assessment.

Conclusions

Burchell et al is a well report trial that provides more robust evidence of the impact of routine syphilis testing than preceding studies. The pragmatic design enables an assessment of real world impact, and whilst the study has limitations these would generally not have changed the

interpretation of the trial - that the increase in testing did not lead to evidence of an increase in early diagnosis of syphilis. Many of the limitations described would have reduced the difference between the testing strategies further.

High prevalence of syphilis persists in this population, and higher levels of evidence are needed to inform programmatic decision making and policy guidelines for the prevention and control of syphilis and other sexually transmitted infections. Studies without a comparison group or those using before/after designs provide a low level of evidence, stepped wedge trials provide a higher level, and parallel cluster randomised trials and other more controlled trials a higher level of evidence still. While compromise is required for pragmatic research, well-designed, robust, and cost-effective studies are needed to ensure that only interventions that show impact are implemented.

Funding

JAT is jointly funded by the UK Medical Research Council (MRC) and the UK Department for International Development (DFID) under the MRC/DFID Concordat agreement and is also part of the EDCTP2 programme supported by the European Union [MR/R010161/1].

Acknowledgements

JAT has no potential conflicts.

References

1. Burchell AN, Tan DHS, Grewal R, et al. Routinised syphilis screening among men living with HIV: A stepped wedge cluster randomised controlled trial. Clinical Infectious Disease **2021**.
2. European Centre for Disease Prevention and Control. Technical Report: Technologies, strategies and approaches for testing populations at risk of sexually transmitted infections in the EU/EEA. Stockholm: ECDC, **2021**.

3. Hussey MA, Hughes JP. Design and analysis of stepped wedge cluster randomized trials. Contemporary clinical trials **2007**; 28(2): 182-91.

4. Thompson JA, Fielding K, C D, Aiken AM, Hargreaves J, Hayes RJ. Bias and inference from misspecified mixed-effect models in stepped wedge trial analysis. Statistics in Medicine **2017**; 36(23): 3670-82.

5. Thompson J, Hemming K, Forbes A, Fielding K, Hayes R. Comparison of small-sample standard-error corrections for generalised estimating equations in stepped wedge cluster randomised trials with a binary outcome: A simulation study. Statistical Methods in Medical Research **2020**; 0(0): 0962280220958735.

6. Thompson JA, Davey C, Fielding K, Hargreaves J, Hayes RJ. Robust analysis of stepped wedge trials using cluster-level summaries within periods. Statistics in Medicine **2018**; 37: 2487-500.

7. Thompson JA. SWPERMUTE: Stata module to compute Permutation tests for Stepped-Wedge Cluster-Randomised Trials. S458426 ed: Boston College Department of Economics, **2017**.

8. Li P, Redden DT. Comparing denominator degrees of freedom approximations for the generalized linear mixed model in analyzing binary outcome in small sample cluster-randomized trials. BMC Medical Research Methodology **2015**; 15(1): 38.

9. Taljaard M, Teerenstra S, Ivers NM, Fergusson DA. Substantial risks associated with few clusters in cluster randomized and stepped wedge designs. Clinical Trials **2016**; 13(4): 459-63.

10. Puffer S, Torgerson D, Watson J. Evidence for risk of bias in cluster randomised trials: review of recent trials published in three general medical journals. BMJ **2003**; 327(7418): 785-9.

11. Eldridge S, Kerry S, Torgerson DJ. Bias in identifying and recruiting participants in cluster randomised trials: what can be done? BMJ **2009**; 339: b4006.