Efficient designs for three-sequence stepped wedge trials with continuous recruitment

Richard Hooper[1], Olivier Quintin[1], Jessica Kasza[2]

[1] Wolfson Institute of Population Health, Queen Mary University of London, London, UK

[2] School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia

Corresponding author:

Richard Hooper, Wolfson Institute of Population Health, Queen Mary University of London, Yvonne Carter Building, 58 Turner Street, London E1 2AB;

telephone 044 (0)20 7882 7324; email r.l.hooper@qmul.ac.uk

Running header: Three-sequence stepped wedge trials

Words: 4095

Abstract

Background/Aims

The standard approach to designing stepped wedge trials that recruit participants in a continuous stream is to divide time into periods of equal length. But the choice of design in such cases is infinitely more flexible: each cluster could cross from the control to the intervention at any point on the continuous time-scale. We consider the case of a stepped wedge design with clusters randomised to just three sequences (designs with small numbers of sequences may be preferred for their simplicity and practicality), and investigate the choice of design that minimises the variance of the treatment effect estimator under different assumptions about the intra-cluster correlation.

Methods

We make some simplifying assumptions in order to calculate the variance: in particular that we recruit the same number of participants, m, from each cluster over the course of the trial, and that participants present at regularly spaced intervals. We consider an intra-cluster correlation that decays exponentially with separation in time between the presentation of two individuals from the same cluster, from a value of ρ for two individuals who present at the same time, to a value of $\rho\tau$ for individuals presenting at the start and end of the trial recruitment interval. We restrict attention to three-sequence designs with centrosymmetry – the property that if we reverse time and swap the intervention and control conditions then the design looks the same. We obtain an expression for the variance of the treatment effect estimator adjusted for effects of time, using methods for generalised least squares estimation, and we evaluate this expression numerically for different designs, and for different parameter values.

Results

There is a two-dimensional space of possible three-sequence, centrosymmetric stepped wedge designs with continuous recruitment. The variance of the treatment effect estimator for given ρ and τ can be plotted as a contour map over this space. The shape of this variance surface depends on τ and on the parameter $m\rho/(1-\rho)$, but typically indicates a broad, flat region of close-to-optimal designs. The "standard" design with equally spaced periods and 1:1:1 allocation rarely performs well, however.

Conclusions

In many different settings a relatively simple design can be found (for example, one based on simple fractions) that offers close-to-optimal efficiency in that setting. There may also be designs that are robustly efficient over a wide range of settings. Contour maps of the kind we illustrate can help guide this choice. If efficiency is offered as one of the justifications for using a stepped wedge design, then it is worth designing with optimal efficiency in mind.

Keywords

Stepped wedge trial, optimal design, continuous recruitment, centrosymmetry, decaying intra-cluster correlation

Introduction

Stepped wedge trials are longitudinal cluster randomised trials where clusters are randomised, not to treatment conditions, but to *sequences* which dictate when each cluster will cross over uni-directionally from the control condition to the intervention condition.[1,2] Since the seminal discussion of stepped wedge trials by Hussey and Hughes,[3] methodological work has tended to treat prospective time as a series of discrete periods, and we now understand a great deal about the optimal design of this kind of stepped wedge trial.[4-9]

But in many stepped wedge trials, including the first to be published with this label,[10] participants from a cluster are recruited/identified, exposed and assessed in one, long, uninterrupted stream. This is known as a continuous recruitment design.[11,12] The standard approach to the design of these stepped wedge trials is to divide time into periods of equal length, with cross-overs at the boundaries between periods. With time on a continuous scale, however, there are no canonical cross-over times, and the number of sequences is limited only by the total number of clusters: each cluster could cross from the control to the intervention at any point on the continuous time-scale.

While this means that designs in continuous time can become quite complicated,[13,14] there may be practical advantages, from the point of view of trial conduct, in choosing a design that has some parsimony, symmetry, or other simplicity of form. In this article we are motivated in particular by an interest in designing continuous recruitment stepped wedge trials with small numbers of randomised sequences. We keep this investigation simple by concentrating on designs with just three sequences. (Previous work has considered the case of a two-sequence design where one sequence remains in the control condition throughout.[15]) Aside from their simplicity, designs with small numbers of sequences also make it easier to balance the randomisation according to cluster characteristics.

The problem of optimal design in this case is quite different to the problem of designing an optimal stepped wedge trial with discrete time periods and an equal number of participants in each period.[4–9] In the continuous recruitment design problem we can effectively move the cross-over time in each sequence with a slider control: by moving it to the right on the time axis we steadily increase the number of control participants in each cluster in that sequence, but at the cost of steadily decreasing the number of intervention participants by a corresponding amount. A continuous time-scale also allows us to model an intra-cluster correlation (ICC) that varies smoothly as a function of separation in time: the further apart we recruit two participants from the same cluster, the weaker we might expect the correlation between their outcomes to be.[12]

Three-sequence designs are relatively simple to characterise, particularly if we focus attention on designs that have *centrosymmetry*. A centrosymmetric design has the property that if we run time backwards, and swap intervention and control, then we arrive at the same design.[16,17] Given the number of clusters, duration of the recruitment period, and rate of recruitment at each cluster, the things that we can control in a centrosymmetric, three-sequence stepped wedge trial are: (a) the proportion of clusters allocated to the middle sequence , w, and (b) the timing of the first cross-over, s (Figure 1).

Here we investigate the choice of design that minimises the variance of the intervention effect estimator under different assumptions about the correlation of outcomes within the same cluster, and consider whether there might be simple design choices that are robustly efficient in this sense, even when there is a degree of uncertainty about these unknown correlation parameters. The lower the variance of the intervention effect estimator, the higher the power to detect a given intervention effect at a certain significance level.

We implicitly assume a large number of clusters are available to be randomised. We use generalised least squares methods and asymptotic approximations to calculate the variance of the intervention effect estimator, and we assume, whatever the number of clusters, that we can allocate them to different sequences in whatever allocation ratios we happen to be discussing. A surprisingly high proportion of published stepped wedge trials are conducted with fewer than 10 clusters (33% according to one review).[18] Designing stepped wedge trials with very small numbers of clusters goes against most methodological guidance,[19,20] but for practical reasons we can expect many stepped wedge trials to include only moderate numbers of clusters. We include simulations for one such trial design scenario to illustrate how empirical statistical power matches the power derived from our large-sample formula in this case.

Motivating example

PATHWEIGH is a weight loss intervention for use in a primary care setting, that uses tools built into the electronic medical record to improve workflow and address various barriers to prioritising weight management.[21] Suresh and colleagues published a protocol for a stepped wedge trial of PATHWEIGH where the clusters are 57 family and internal medicine clinics in a large health system in Colorado, USA.[21] Patients are identified over a four-year period beginning March 17 2020, and are eligible to be included in the trial if they are aged 18 years or over and overweight (BMI ≥25 kg/m^2) at an initial, index visit.

The investigators estimate that a minimum of 30 patients per clinic per year will be identified. The primary outcome measure is weight loss 6 months after the index visit, extracted from the electronic medical record. PATHWEIGH, being an electronic intervention, can be "turned on" at a clinic at any time, and a patient's intervention status is defined according to whether the clinic they attend was in the control (routine care) condition or intervention condition at the patient's index visit.

The trial randomises clinics 1:1:1 to three sequences, in which the intervention is turned on after one year, two years or three years, respectively. In terms of the design parameters of a centrosymmetric, three-sequence stepped wedge design, the timing of the first cross over, $s = 1/4$ and the proportion of clusters allocated to the middle sequence, $w = 1/3$ in this case. This kind of regular spacing is traditional in stepped wedge designs. But what if we were willing to consider alternative allocation ratios, or alternative timings for the first and last cross-overs in a three-sequence design? Is there a more efficient choice of s and w?

Methods

Statistical model

To describe the general scenario that we consider, we suppose that time is re-scaled so that the recruitment/identification period runs from time 0.0 to time 1.0. For simplicity we assume that in each cluster we recruit the same number of individuals, m, over the course of this unit time interval.

We assume a continuous outcome Y_{ij} for individual $i = 1, \dots, m$ from cluster $j = 1, \dots, J$, with

$$Y_{ij} = \delta . H\left(t_{ij} - t_j^*\right) + T\left(t_{ij}, \boldsymbol{\beta}\right) + \varepsilon_{ij}, \qquad \varepsilon_{ij} \sim \mathrm{N}(0, \sigma^2), \tag{1}$$

where t_{ij} is the time at which individual i from cluster j is recruited, t_j^* is the time at which cluster j crosses over to the intervention, $H(x)$ is a unit step function taking value 0 if $x < 0$ and 1 if $x \geq 0$, and $T(t, \boldsymbol{\beta})$ is a function that describes the underlying effect of time on expected outcome (assumed to be the same in all clusters), parameterised by the vector $\boldsymbol{\beta}$. Errors ε_{ij} are identically distributed but not all independent of one another (see below). In a centrosymmetric, three-step design, all of

the cross-over times $t_j^*, j = 1 \ldots J$ belong to the set $\{s, 0.5, 1-s\}$, for some s with $0 \leq s < 0.5$, and clusters are allocated to these sequences in the respective proportions $(1-w)/2$, w and $(1-w)/2$, for some w with $0 \leq w < 1$ (Figure 1).

The parameter δ is the intervention effect we would like to estimate. Having a single parameter for this intervention effect, expressed through the step function, H, models a situation where the intervention has an instantaneous effect that is then maintained at a constant level, and is common to all the clusters.

We assume that ε_{ij} and $\varepsilon_{i'j'}$ are uncorrelated for all $j \neq j'$, for all i and i', that is, outcomes in different clusters are uncorrelated. Within a cluster there may be correlation between outcomes, measured by the ICC. Under a model of exchangeability, this ICC is constant:

$$\text{Corr}\left(\varepsilon_{ij}, \varepsilon_{i'j}\right) = \rho \ \ \forall\, i \neq i', j.$$

In this article we consider a more general model that allows us to investigate what happens if the ICC decays with increasing separation in time:[15]

$$\begin{aligned} \text{Corr}\left(\varepsilon_{ij}, \varepsilon_{i'j}\right) &= \rho\tau^{|t_{ij}-t_{i'j}|} \ \ \forall\, i \neq i', j; \\ \text{Corr}\left(\varepsilon_{ij}, \varepsilon_{i'j'}\right) &= 0 \ \ \forall\, i, i', j \neq j'. \end{aligned} \tag{2}$$

The parameter ρ is the ICC for two participants sampled from the same cluster at the same time (the time-specific ICC). The ICC for two participants sampled at times 1.0 apart is $\rho\tau$. The parameter τ is therefore the factor by which the ICC decays over the duration of the recruitment period.

Finally, to make headway with deriving a variance for the treatment effect estimator, we make the simplifying assumption that the times at which individuals are recruited from each cluster are regularly spaced, at intervals $1/m$ apart, rather than being a random process:

$$t_{ij} = i/m \quad \forall\, j. \tag{3}$$

Time effect

The variance of the treatment effect estimator will be adjusted for the time effect, under the assumption that this is correctly specified in the analysis model. Exactly which design minimises this variance would seem to depend on the form of the time effect, T in (1), so how are we to choose this? Here, we consider a time effect that is piecewise constant, with discontinuities that coincide with the cross-over times. That is,

$$T(t, \beta_1, \beta_2, \beta_3, \beta_4) = \beta_1 + \beta_2 H(t - s) + \beta_3 H(t - 0.5) + \beta_4 H(t - 1 + s). \tag{4}$$

While it might seem strange at first sight to choose a model for the time effect that depends on the design, we make this choice for a number of related reasons. Firstly, the regularly spaced recruitment times at each cluster, $i/m, i = 1, \dots, m$, and the fact that the cross-over times align perfectly with the discontinuities in an otherwise constant time effect, mean that the variance of the treatment effect estimator under the four-parameter functional form for T in (4) is identical to the variance if we were to assume a completely general time effect that was different at each of the times $i/m, i = 1, \dots, m$. (This follows from a more general invariance result proved elsewhere.)[23] Secondly, experience with two-sequence designs suggests that the optimal design solution under (4) may also be a reasonable approximation to the solution under other models with a smooth time effect, such as a polynomial function of sufficiently high degree.[15] Thirdly (and more qualitatively),

adjusting for an underlying time effect that has step-changes at exactly the times the intervention is switched on in certain clusters feels like it ought to offer a "strong" test of a genuine intervention effect.

In fact, if we were faced with a real-life dataset with $m = 100$, say, it is unlikely that we would use as many as 100 parameters for modelling the effect of time, or that we would assume the time effect was completely smooth. It is more likely that we would allow for step-changes at the cross-overs, and otherwise model time in some parsimonious but plausible way that fitted the data and our prior beliefs. As long as this analysis model for the time effect is correctly specified, then the invariance result mentioned above shows that the variance of the treatment effect estimator will be the same as the variance calculated to hold under model (4).[23]

Variance of the treatment effect estimator

If we write outcomes Y_{ij} as a single column vector, $\boldsymbol{Y}$, and rewrite (1), (2), (3), (4) as

$$\boldsymbol{Y} = \boldsymbol{X}\begin{pmatrix}\delta\\ \boldsymbol{\beta}\end{pmatrix} + \boldsymbol{\varepsilon}, \qquad \boldsymbol{\varepsilon} \sim \mathrm{N}(\boldsymbol{0}, \boldsymbol{V}),$$

Then the generalised least squares estimator for the parameters is obtained as

$$\begin{pmatrix}\hat{\delta}\\ \hat{\boldsymbol{\beta}}\end{pmatrix} = (\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{Y},$$

and the variance of $\hat{\delta}$ is the first diagonal element of the variance-covariance matrix given by

$$Var\begin{pmatrix}\hat{\delta}\\ \hat{\boldsymbol{\beta}}\end{pmatrix} = (\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}. \quad (5)$$

The results presented in this article were obtained with the help of numerical matrix inversion, making use of the fact that the matrix $\boldsymbol{V}$ is block diagonal, owing to the independence of outcomes in different clusters. We used Mata, the matrix programming language within Stata, which is a C-style language that is compiled to produce fast-running code (StataCorp, College Station, Texas USA). Our code can be accessed online (https://github.com/richard-hooper/SW-3sequence-continuous-recruitment).

Scenarios

Previous work on optimal stepped wedge design suggests, in the case $\tau = 1$ at least, that the optimal design choice depends on the total number of participants in each cluster, m, and time-specific ICC, ρ, via the single parameter $m\rho/(1-\rho)$.[4,5] This parameter is key because it determines the correlation between mean outcomes in within-cluster comparisons. If clusters are randomised to a sequence that crosses over from the control to the intervention condition half-way through the recruitment period, for example, then the correlation between the mean outcome in a cluster before cross-over ($m/2$ participants) and the mean outcome in that cluster after cross-over ($m/2$ participants) is given by the following expression, which is a function of $R := m\rho/(1-\rho)$:[25]

$$\frac{\frac{m}{2}\rho}{1+\left(\frac{m}{2}-1\right)\rho} = \frac{R}{2+R}.$$

For given $\tau < 1$ we hypothesise a similar role for the parameter $m\rho/(1-\rho)$ in determining the optimal design. Previous work on two-sequence designs with continuous recruitment proposed design rules of thumb based on the product $m\rho$.[15] Since the parameter $m\rho$ is somewhat simpler,

and for small ρ the parameters $m\rho$ and $m\rho/(1-\rho)$ are almost identical, we consider a range of values for m and $m\rho$. In order to limit the set of parameter values and to follow a systematic approach, we consider $m \in \{50, 200, 1{,}000\}$ and, for each m, values of ρ for which $m\rho \in \{0.2, 0.5, 1, 2, 5, 10, 20\}$. These scenarios are investigated for $\tau \in \{1.0, 0.5, 0.1\}$. By presenting results for fixed $m\rho$ side by side, we can see how closely the design recommendations match for different m. In the online supplement we also show what results look like when we keep $m\rho/(1-\rho)$ fixed for different m, to confirm the intuition that this is the parameter that strictly determines the optimal design.

Contour plots

We transform the variance of the treatment effect estimator to a log scale, and draw contour plots of the log-variance over the design parameter space $0 \leq s < 0.5$ and $0 \leq w < 1$, for each of the scenarios considered. Contour lines are separated by log(1.1), so that moving from one contour to the next represents a 10% increase in the variance. The lowest contour value is set at the minimum of the log-variance surface. Thus, the innermost contour is the region where the variance is within 10% of the minimum attainable. Because we are particularly interested in regions of low variance, and to reduce the amount of ink used, we have not drawn all of the higher variance contour lines. The variance of the treatment effect estimator will be a multiple of σ^2/J, σ^2 being the variance of each outcome Y_{ij}, and J being the number of clusters. Hence the appearance of the contour plot as described does not depend on either σ^2 or J.

Results

Contour plots for $\tau = 1.0$, $\tau = 0.5$, and $\tau = 0.1$ are shown in Figures 2(a), 2(b) and 2(c). For given τ and $m\rho$, the plots for different m are remarkably similar, confirming that the shape of the variance

surface, and therefore the choice of optimal design, depends on m and ρ primarily through their product. The greatest divergence between the plots in a given panel of these Figures is for $\tau = 0.1$, $m\rho = 20.0$ (bottom row of Figure 2(c)). Supplemental Figure 1 (in the online supplement) shows contour plots for $\tau = 0.1$, $m\rho/(1-\rho) = 20$, $m \in \{50, 200, 1{,}000\}$. These three plots are almost identical, illustrating that the dependence is strictly via $m\rho/(1-\rho)$. To present a relatively simple overview of our results below, with intuitive rules of thumb, we will continue to talk in terms of $m\rho$.

When $m\rho$ is zero the optimal design gives zero weight to the middle sequence and puts the first cross-over at time zero. If $m\rho$ is zero then observations within clusters are independent, so this design – a two-sequence, parallel groups design with one sequence in the intervention condition throughout and the other in the control condition throughout (analogous to an individually randomised parallel groups design) – is exactly what we would expect to see. As $m\rho$ starts to increases from 0.0 the optimal weight for the middle sequence increases, but the optimal time for the first cross-over remains at zero. This means there are clusters that remain in the control condition or the intervention condition for the entire trial. In the world of discrete-period stepped wedge designs these are known as hybrid designs, because they are a hybrid of a parallel groups design and a classic stepped wedge design where all the clusters cross over at some point.[5] Somewhere between $m\rho = 2.0$ and $m\rho = 5.0$ (at least for $\tau \geq 0.1$) the optimal time for the first cross-over breaks away from zero. The shape of the plotted surface depends on τ, although there are qualitative similarities between plots for given $m\rho$ and different τ.

The variance surfaces in the plots have relatively flat bottoms, suggesting, first, that for any given $m\rho$ and τ there may be a close-to-optimal solution where s and w are simple fractions, and second, that it may be possible to find a solution that is close-to-optimal over a range of values for $m\rho$ and τ, thus accommodating some uncertainty in these parameters. Look at the solution with $s = 0$, $w = 1/3$, for example, which is one of the solutions marked with a "+" in Figure 2. This simple, hybrid

design allocates clusters in equal proportions to three sequences. For any $m\rho$ in the range 0.2 to 5.0 and any τ between 0.1 and 1.0, this design falls within the innermost contour of the plot, *i.e.* the region where the variance is within 10% of the minimum attainable, suggesting it could be an acceptable design in all these scenarios. Or consider the solution with $s = 0.15$, $w = 1/3$, which is the second solution marked with a "+". This is perhaps not a "natural" choice of design, but nevertheless falls within the innermost contour of the plot for any $m\rho$ of 5.0 or greater and any τ between 0.1 and 1.0, making it a robustly efficient choice in this broad range of cases.

A design with equally spaced steps and equally weighted sequences ($s = 0.25$, $w = 1/3$), on the other hand, which is the third solution marked in Figure 2 with a "+", is borderline at best (at least in the scenarios plotted), achieving its most favourable performance when $m\rho = 20$ and $\tau = 0.5$.

Design and sample size in the motivating example

In our motivating example, the PATHWEIGH trial, the assumption was that 30 patients would be recruited per year from each clinic over 4 years, so that $m = 120$. The power calculation for PATHWEIGH assumed an ICC somewhere between 0.02 and 0.05. If we take this to reflect an interval estimate of the time-specific ICC, ρ, then we want a design that is robustly efficient with $m\rho$ in the range 2.4 to 6.0. The PATHWEIGH power calculation did not discuss the decay in the ICC, but let us suppose, for the sake of argument, that we want to allow for a decay of up to 0.5 in the ICC over the four-year trial period, so that $0.5 \leq \tau \leq 1$. We should remember that a prior estimate of the decay in the ICC over 4 years may be much harder to come up with than a prior estimate of the time-specific ICC. (Of course, when presented with an estimate of the ICC that does not mention the decay over time at all, as in the PATHWEIGH example, there is a question as to whether this is really an estimate of the time-specific ICC, or some average of the decaying ICC over the entire

recruitment period.[26] This speaks to the importance of specifying assumptions about the ICC when reporting sample size calculations for stepped wedge trials.[27])

Looking at our contour plots, we might then choose s intermediate between 0 and 0.15 – $s = 1/12$, say – and $w = 1/3$, as a solution that is robustly efficient. In this design the intervention is switched on after 4 months, 2 years, and 3 years and 8 months in the respective sequences.

Recall from the Methods that the variance of the treatment effect estimator is a multiple of σ^2/J, where σ^2 is the variance of each outcome, and J is the number of clusters. Let us call this $\theta\sigma^2/J$. To calculate the number of clusters needed to demonstrate an effect δ^* with power $1-\beta$ at the 5% significance level, we can therefore apply the "fundamental equation" of sample size:[28]

$$J \geq \left(z_{0.975} + z_{1-\beta}\right)^2 \left(\frac{\sigma}{\delta^*}\right)^2 \theta. \tag{6}$$

Where z_p is the p^{th} percentile of the standard normal distribution.

Table 1 shows the variance of the treatment effect estimator for the design $s = 1/12$, $w = 1/3$, at the four corners of the selected parameter range for $m\rho$ and τ. The variance is calculated from the matrix expression in the Methods. Table 1 also shows the total number of clusters (allocated equally to the three sequences) required to detect a mean difference in weight loss of 1kg, 1.25kg, or 1.5kg with 80% power at the 5% significance level, assuming (as the PATHWEIGH investigators did) that the standard deviation of weight loss is 10.7kg. The number of clusters is rounded up to the next multiple of three. For comparison, Table 1 shows the same requirements for the "standard" stepped wedge design (the design adopted for PATHWEIGH) with $s = 0.25$, $w = 1/3$. The non-standard design always out-performs the standard one – in some scenarios, considerably so.

The theoretical power (from equations (5) and (6)) of each of the designs presented in Table 1 was compared with the empirical power estimated by simulation. For each scenario and design, 1,000 replicated datasets were generated in R, and the continuous time decay model in equations (1) to (4) was fitted using the glmmTMB package in R.[28] This analysis did not include any small-sample correction to mitigate inflation of the nominal Type I error rate and power that might result from the moderate number of clusters. Such corrections are becoming more widely available in software implementations for the analysis of cluster-randomised trials,[30] but also add considerably to the processing time, which can make large-scale simulations challenging. Our code can be accessed online (https://github.com/richard-hooper/SW-3sequence-continuous-recruitment). Supplemental Figure 2 displays the results in a nested loop plot. Empirical power closely matched theoretical power when the time-specific ICC was 0.05. Empirical power was inflated when the time-specific ICC was 0.02, and this was particularly evident with the non-standard design. The performance of the glmmTMB package (and alternatives) for analysing data from longitudinal cluster randomised trials with a continuous time decay model for the ICC warrants further investigation.

Discussion

We have illustrated optimal designs for three-sequence stepped wedge trials with continuous recruitment, under different assumptions about the correlation of outcomes from the same cluster. We conclude that under given assumptions there may be a relatively simple design that offers close-to-optimal efficiency, and that there may be designs that are robustly efficient over a wide range of assumptions. If efficiency is offered as one of the justifications for using a stepped wedge design over a parallel groups design, then we should design with optimal efficiency in mind. The focus of this article has been on design, informed by theory. Suitable approaches to analysis that can handle

a continuous time model (including the decay in the ICC) and also control the Type I error rate when the number of clusters is moderate or small need further evaluation and comparison.

The way in which the ICC changes over time matters to the design, and it is important to articulate these assumptions when reporting sample size calculations for stepped wedge trials.[27] We assumed a particular parametric form for the decay in the ICC to help us understand the more general impact of this kind of decay on optimal design. Other models for the ICC could, of course, be investigated. As in earlier work,[15] we simplified considerably in assuming that eligible participants present at regular, fixed intervals rather than as a random continuous-time process, but assuming that the arrival rate is constant over time we would expect arrival times in a sample to become increasingly uniformly spread as the rate increases. Simulation studies that have investigated the impact of unevenly spaced arrival times on precision of the treatment effect estimator in the context of stepped wedge designs suggest that this impact is small.[22]

Our focus has been on the optimal design of three-sequence trials, but our findings may also offer clues about optimal design with larger numbers of sequences. With more sequences there are more degrees of freedom to the design space for centrosymmetric designs, which becomes correspondingly harder to visualise and requires more effort to search exhaustively. Nevertheless, previous work on optimal design in the discrete time case, with no decay in the intracluster correlation, has shown that the "internal" sequences (that is, sequences other than the first and last) should all be given equal weight.[4,5] There may be similar simplifications when we move over to considering designs with many sequences in the continuous time setting. Ultimately, however, we may prefer a design with fewer sequences for its greater simplicity and practicality.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Declaration of conflicting interests

The authors declare that there is no conflict of interest.

References

1. Hemming K, Haines TP, Chilton PJ, Girling AJ, Lilford RJ. The stepped wedge cluster randomised trial: rationale, design, analysis and reporting. BMJ 2015;350:h391
2. Hooper R. Key concepts in clinical epidemiology: Stepped wedge trials. J Clin Epi 2021;137:159-162
3. Hussey MA, Hughes JP. Design and analysis of stepped wedge cluster randomized trials. Contemp Clin Trials 2007;28(2):182-191
4. Girling AJ, Hemming K. Statistical efficiency and optimal design for stepped cluster studies under linear mixed effects models. Stat Med 2016;35(13):2149-66
5. Lawrie J, Carlin JB, Forbes AB. Optimal stepped wedge designs. Stat Probab Lett 2015;99:210-214
6. Li F, Turner EL, Preisser JS. Optimal allocation of clusters in cohort stepped wedge designs. Stat Probab Lett 2018;137:257–263
7. Thompson JA, Fielding K, Hargreaves J, Copas A. The optimal design of stepped wedge trials with equal allocation to sequences and a comparison to other trial designs. Clin Trials 2017; 14(6):639-647
8. Moerbeek M. Optimal allocation to treatment sequences in individually randomized stepped-wedge designs with attrition. Clin Trials 2023; 20(3):242-251
9. Watson S, Girling A, Hemming K. Optimal study designs for cluster randomised trials: An overview of methods and results. Stat Methods Med Res 2023;32(11):2135-2157
10. The Gambia Hepatitis Study Group. The Gambia Hepatitis Intervention Study. Cancer Res 1987;47:5782-5787
11. Copas AJ, Lewis JJ, Thompson JA, Davey C, Baio G, Hargreaves JR. Designing a stepped wedge trial: three main designs, carry-over effects and randomisation approaches. Trials 2015;16:352

12. Hooper R, Copas A. Stepped wedge trials with continuous recruitment require new ways of thinking. J Clin Epi 2019;116:161-166
13. Grantham KL, Kasza J, Heritier S, Hemming K, Litton E, Forbes AB. How many times should a cluster randomized crossover trial cross over? Stat Med 2019;38(25):5021-5033
14. Hooper R, Kasza J, Forbes A. The hunt for efficient, incomplete designs for stepped wedge trials with continuous recruitment and continuous outcome measures. BMC Med Res Methodol 2020;20:279
15. Hooper R, Copas AJ. Optimal design of cluster randomised trials with continuous recruitment and prospective baseline period. Clin Trials 2021;18(2):147-157
16. Kasza J, Forbes A. Information content of cluster-period cells in stepped wedge trials. Biometrics 2019;75(1):144-152
17. Bowden R, Forbes AB, Kasza J. On the centrosymmetry of treatment effect estimators for stepped wedge and related cluster randomized trial designs. Stat Probab Lett 2021;172:109022
18. Grayling MJ, Wason JMS, Mander AP. Stepped wedge cluster randomized controlled trial designs: a review of reporting quality and design features. Trials 2017;18:13
19. Kahan BC, Forbes G, Ali Y, Jairath V, Bremner S, Harhay MO, Hooper R, Wright N, Eldridge SM, Leyrat C. Increased risk of type I errors in cluster randomised trials with small or medium numbers of clusters: a review, reanalysis, and simulation study. Trials 2016;17(1):438
20. Taljaard M, Teerenstra S, Ivers NM, Fergusson DA. Substantial risks associated with few clusters in cluster randomized and stepped wedge designs. Clin Trials 2016;13(4):459-463
21. Suresh K, Summers Holtrop J, Dickinson LM, Willems E, Smith PC, Gritz RM, Perreault L. PATHWEIGH, pragmatic weight management in adult patients in primary care in Colorado, USA: study protocol for a stepped wedge cluster randomized trial. Trials 2022;23:26
22. Grantham KL, Kasza J, Heritier S, Hemming K, Forbes AB. Accounting for a decaying correlation structure in cluster randomized trials with continuous recruitment. Stat Med 2019;38(11):1918-1934

23. Grantham KL, Forbes AB, Heritier S, Kasza J. Time parameterizations in cluster randomized trial planning. Am Stat 2020;74(2):184–189
24. Kerry SM. The intracluster correlation coefficient in cluster randomisation. BMJ 1998;316:1455
25. Feldman HA, McKinlay SM. Cohort versus cross-sectional design in large field trials: Precision, sample size, and a unifying model. Stat Med 1994;13(1):61-78
26. Kasza J, Hemming K, Hooper R, Jns Matthews JNS, Forbes AB. Impact of non-uniform correlation structure on sample size and power in multiple-period cluster randomised trials. Stat Methods Med Res 2019;28:703-716
27. Hemming K, Taljaard M, McKenzie JE, Hooper R, Copas A, Thompson JA, Dixon-Woods M, Aldcroft A, Doussau A, Grayling M, Kristunas C, Goldstein CE, Campbell MK, Girling A, Eldridge S, Campbell MJ, Lilford RJ, Weijer C, Forbes AB, Grimshaw JM. Reporting of stepped wedge cluster randomised trials: extension of the CONSORT 2010 statement with explanation and elaboration. BMJ 2018;363:k1614
28. Machin D, Campbell MJ, Tan SB, Tan SH. Sample size tables for clinical studies. 3rd ed. Chichester, UK: Wiley-Blackwell; 200926. Grantham K, Kasza J, Heritier S, Hemming K, Forbes AB. Accounting for a decaying correlation structure in cluster randomised trials with continuous recruitment. Stat Med 2019; 38: 1918–1934
29. Brooks ME, Kristensen K, van Benthem KJ, Magnusson A, Berg CW, Nielsen A, Skaug HJ, Maechler M, Bolker BM (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. R Journal 2017;9(2):378-400
30. Leyrat C, Morgan KE, Leurent B, Kahan B. Cluster randomized trials with a small number of clusters: which analyses should be used? Int J Epidemiol 2018;47(1):321-331

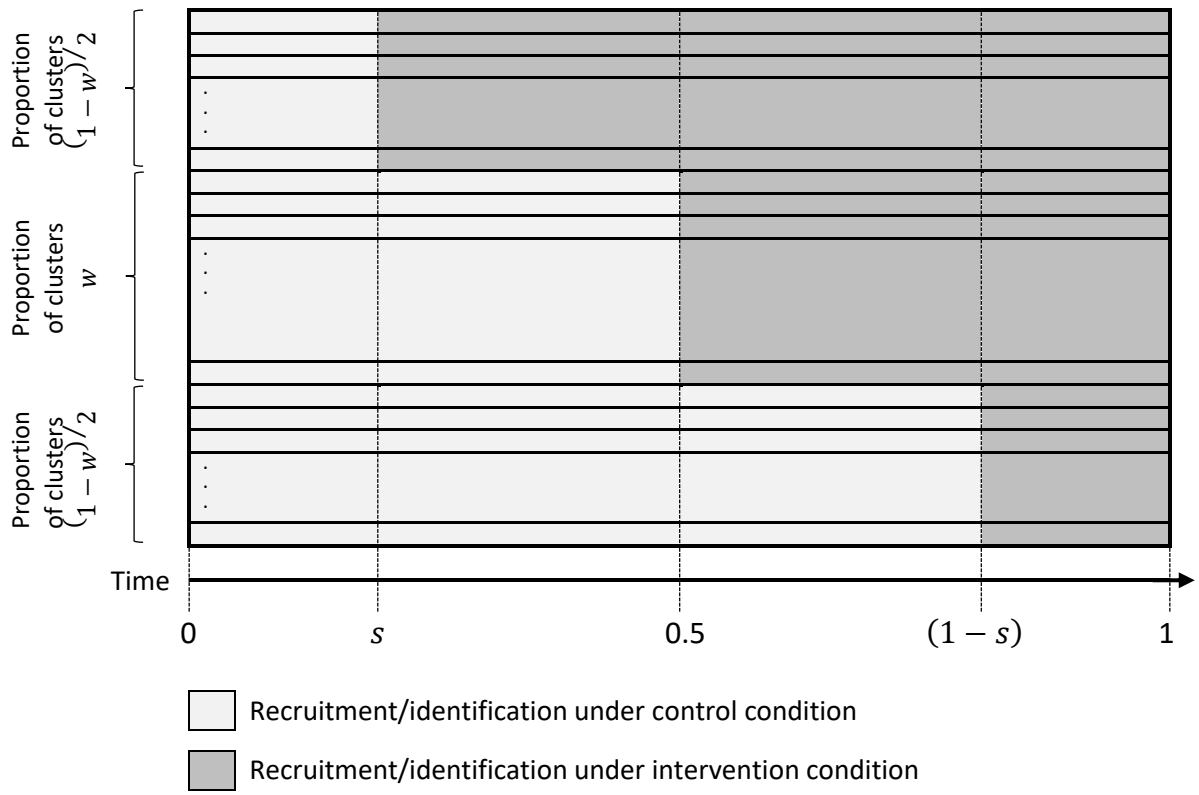
Table 1. Total number of clusters needed in the PATHWEIGH example (see text) to achieve 80% power at the 5% significance level to detect different treatment effects, under different scenarios concerning correlations between outcomes from the same cluster.

Scenarios		Non-standard design ($s = 1/12, w = 1/3$)				Standard design ($s = 1/4, w = 1/3$)			
		Variance of the treatment effect estimator	Number of clusters needed to detect a difference in mean weight loss of			Variance of the treatment effect estimator	Number of clusters needed to detect a difference in mean weight loss of		
ρ	τ		1kg	1.25kg	1.5kg		1kg	1.25kg	1.5kg
0.02	1.0	$0.0793\sigma^2/J$	72	48	33	$0.1002\sigma^2/J$	93	60	42
0.02	0.5	$0.0820\sigma^2/J$	75	48	33	$0.1054\sigma^2/J$	96	63	45
0.05	1.0	$0.0928\sigma^2/J$	84	54	39	$0.1054\sigma^2/J$	96	63	45
0.05	0.5	$0.1093\sigma^2/J$	99	63	45	$0.1217\sigma^2/J$	111	72	51

Figure legends

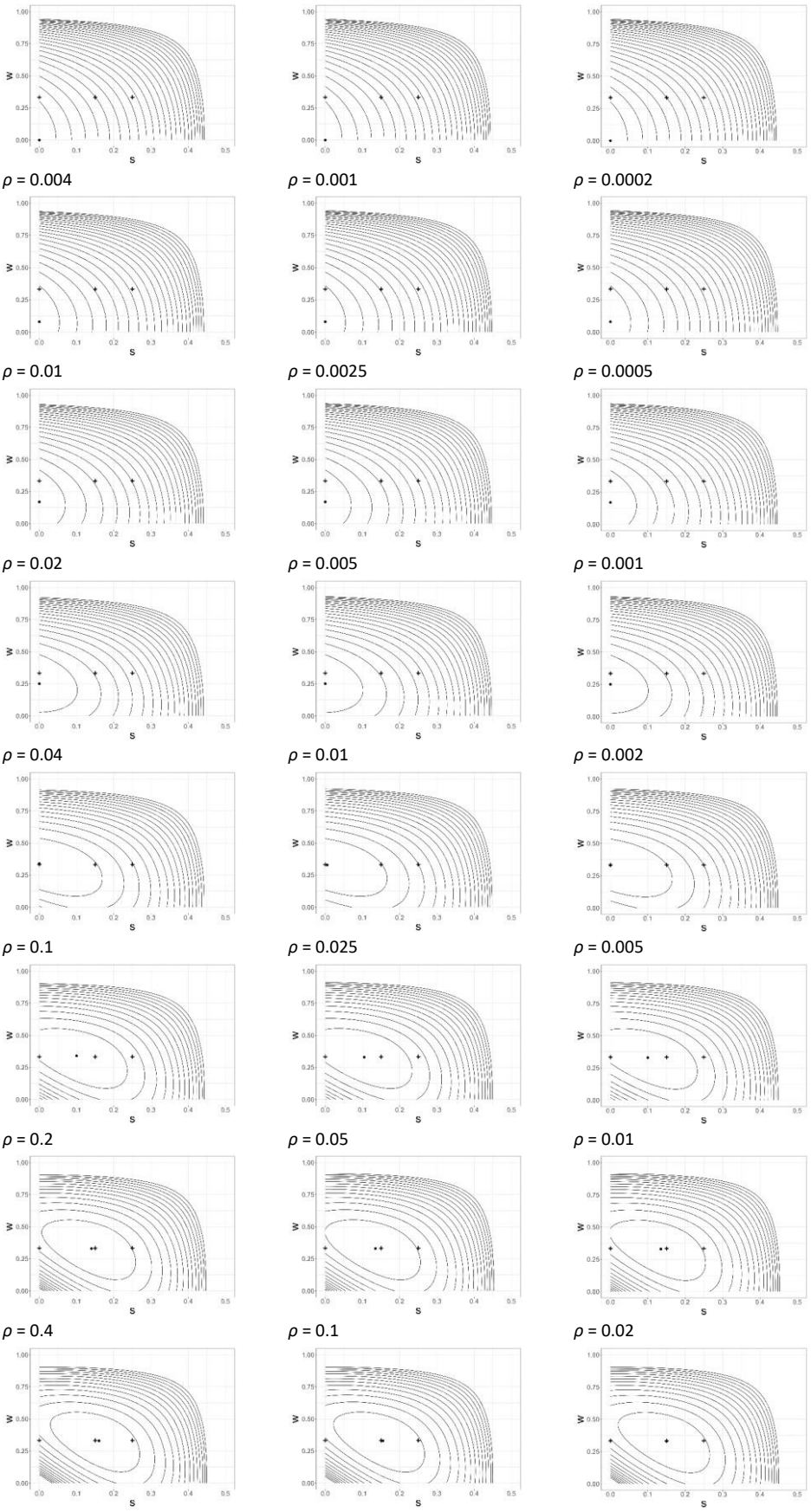
Figure 1. Schematic for a centrosymmetric, three-sequence stepped wedge trial with continuous recruitment. Centrosymmetry is the property that if we run time backwards, and swap intervention and control, then we arrive at the same design.

Figure 2. Contour plots of the log of the variance of the treatment effect estimator for different $m\rho$ and m, where m is the recruitment rate at a cluster, and ρ is the intra-cluster correlation. Time is scaled from 0 to 1 over the recruitment period. Contour plots are drawn over the design parameter space $0 \leq s < 0.5$ and $0 \leq w < 1$, where s is the first cross-over time and w is the proportion of clusters allocated to the middle sequence. The solutions $s = 0$, $w = 1/3$, $s = 0.15$, $w = 1/3$, and $s = 0.25$, $w = 1/3$ (see text) are marked with a "+" symbol. Contour lines are separated by log(1.1), so that moving from one contour to the next represents a 10% increase in the variance. The lowest contour value is set at the minimum of the log-variance surface (the small, circular mark on each plot marks this minimum). The factor, τ, by which the intra-cluster correlation decays over unit time is (a) $\tau = 1.0$; (b) $\tau = 0.5$; (c) $\tau = 0.1$.

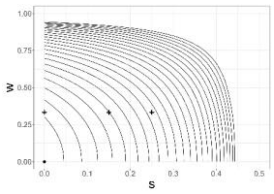
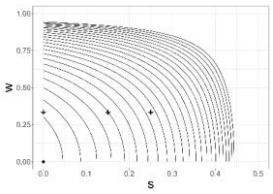
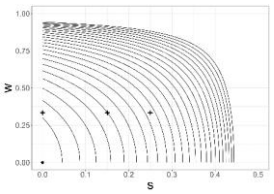
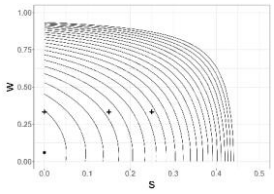
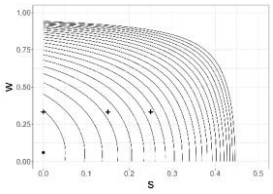
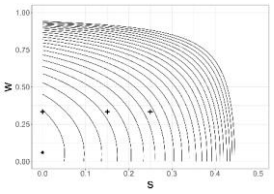
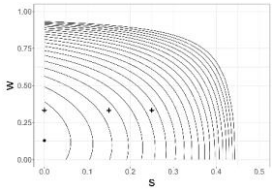
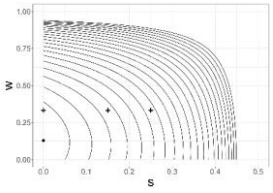
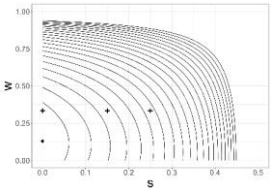
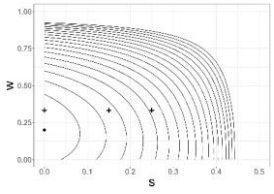
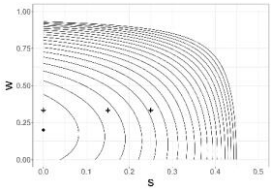
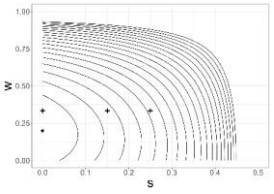
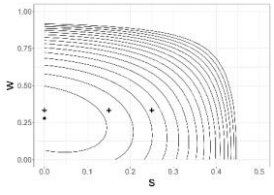
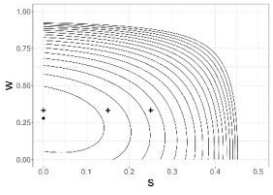
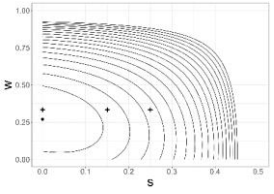
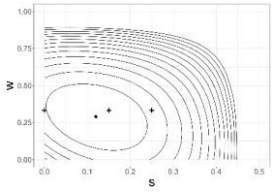
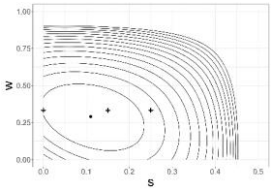
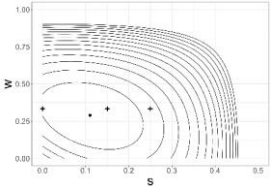
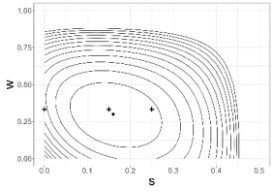
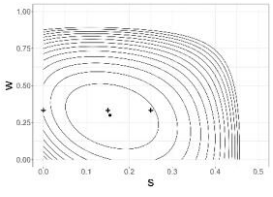
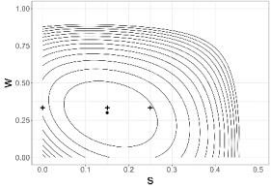
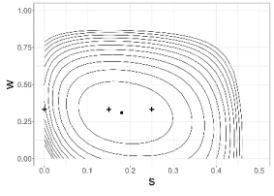
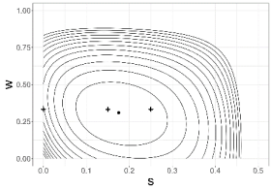
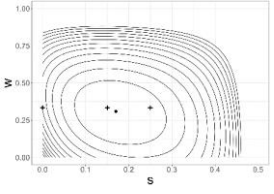

Proportion of clusters $(1-w)/2$
Proportion of clusters w
Proportion of clusters $(1-w)/2$
Time
0
s
0.5
$(1-s)$
1
Recruitment/identification under control condition
Recruitment/identification under intervention condition

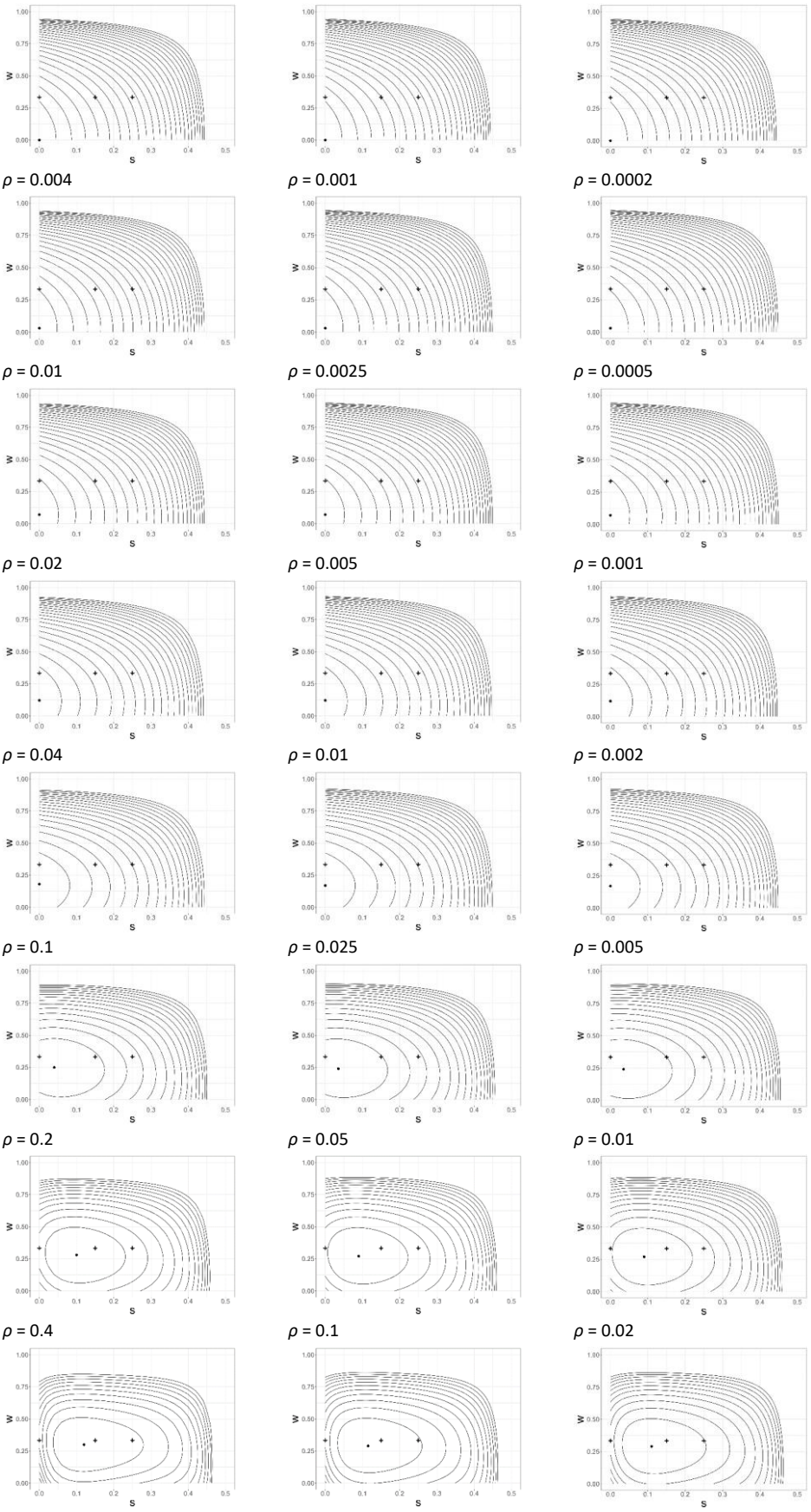
(a) $\tau = 1.0$

	$m = 50$	$m = 200$	$m = 1{,}000$
$m\rho = 0.0$	$\rho = 0.0$	$\rho = 0.0$	$\rho = 0.0$
$m\rho = 0.2$	$\rho = 0.004$	$\rho = 0.001$	$\rho = 0.0002$
$m\rho = 0.5$	$\rho = 0.01$	$\rho = 0.0025$	$\rho = 0.0005$
$m\rho = 1.0$	$\rho = 0.02$	$\rho = 0.005$	$\rho = 0.001$
$m\rho = 2.0$	$\rho = 0.04$	$\rho = 0.01$	$\rho = 0.002$
$m\rho = 5.0$	$\rho = 0.1$	$\rho = 0.025$	$\rho = 0.005$
$m\rho = 10.0$	$\rho = 0.2$	$\rho = 0.05$	$\rho = 0.01$
$m\rho = 20.0$	$\rho = 0.4$	$\rho = 0.1$	$\rho = 0.02$



(b) $\tau = 0.5$

	$m = 50$	$m = 200$	$m = 1{,}000$
$m\rho = 0.0$	$\rho = 0.0$	$\rho = 0.0$	$\rho = 0.0$
			
$m\rho = 0.2$	$\rho = 0.004$	$\rho = 0.001$	$\rho = 0.0002$
			
$m\rho = 0.5$	$\rho = 0.01$	$\rho = 0.0025$	$\rho = 0.0005$
			
$m\rho = 1.0$	$\rho = 0.02$	$\rho = 0.005$	$\rho = 0.001$
			
$m\rho = 2.0$	$\rho = 0.04$	$\rho = 0.01$	$\rho = 0.002$
			
$m\rho = 5.0$	$\rho = 0.1$	$\rho = 0.025$	$\rho = 0.005$
			
$m\rho = 10.0$	$\rho = 0.2$	$\rho = 0.05$	$\rho = 0.01$
			
$m\rho = 20.0$	$\rho = 0.4$	$\rho = 0.1$	$\rho = 0.02$
			

(c) $\tau = 0.1$

	$m = 50$	$m = 200$	$m = 1{,}000$
$m\rho = 0.0$	$\rho = 0.0$	$\rho = 0.0$	$\rho = 0.0$
$m\rho = 0.2$	$\rho = 0.004$	$\rho = 0.001$	$\rho = 0.0002$
$m\rho = 0.5$	$\rho = 0.01$	$\rho = 0.0025$	$\rho = 0.0005$
$m\rho = 1.0$	$\rho = 0.02$	$\rho = 0.005$	$\rho = 0.001$
$m\rho = 2.0$	$\rho = 0.04$	$\rho = 0.01$	$\rho = 0.002$
$m\rho = 5.0$	$\rho = 0.1$	$\rho = 0.025$	$\rho = 0.005$
$m\rho = 10.0$	$\rho = 0.2$	$\rho = 0.05$	$\rho = 0.01$
$m\rho = 20.0$	$\rho = 0.4$	$\rho = 0.1$	$\rho = 0.02$