Prevention Science
https://doi.org/10.1007/s11121-023-01587-1

Sample Size Calculations for Stepped Wedge Designs with Treatment Effects that May Change with the Duration of Time under Intervention

James P. Hughes[1] · Wen-Yu Lee[2] · Andrea B. Troxel[2] · Patrick J. Heagerty[1]

Accepted: 11 September 2023

Abstract

The stepped wedge design is often used to evaluate interventions as they are rolled out across schools, health clinics, communities, or other clusters. Most models used in the design and analysis of stepped wedge trials assume that the intervention effect is immediate and constant over time following implementation of the intervention (the "exposure time"). This is known as the IT (immediate treatment effect) assumption. However, recent research has shown that using methods based on the IT assumption when the treatment effect varies over exposure time can give extremely misleading results. In this manuscript, we discuss the need to carefully specify an appropriate measure of the treatment effect when the IT assumption is violated and we show how a stepped wedge trial can be powered when it is anticipated that the treatment effect will vary as a function of the exposure time. Specifically, we describe how to power a trial when the exposure time indicator (ETI) model of Kenny et al. (Statistics in Medicine, 41, 4311–4339, 2022) is used and the estimand of interest is a weighted average of the time-varying treatment effects. We apply these methods to the ADDRESS-BP trial, a type 3 hybrid implementation study designed to address racial disparities in health care by evaluating a practice-based implementation strategy to reduce hypertension in African American communities.

Keywords Sample size · Stepped wedge · Exposure time indicator model

Introduction

In prevention research, clinical research, and implementation science, the stepped wedge design is often used to evaluate interventions as they are rolled out across schools, health clinics, communities, or other clusters (Hussey & Hughes, 2007; Copas et al., 2015; Hemming et al., 2017; Hemming et al., 2018). In a stepped wedge design, all clusters typically start in a control or standard of care condition and then, at pre-selected intervals, the intervention is introduced in one or more randomly selected clusters, until all clusters receive the intervention. Outcome measurements are usually collected in each cluster in each interval, although other designs are possible (Hooper & Burke, 2015; Kasza et al., 2022). A key motivation for the use of the stepped wedge design is to study alternative implementation strategies, tailored to the local context, for evidence-based interventions.

Most analyses of data from stepped wedge trials assume that the intervention effect is "instantaneous" (reaches full effect within the time interval in which it is introduced) and constant (does not vary as a function of time since introduction—the "exposure time"). Most papers on sample size calculations for stepped wedge designs (Hemming & Taljaard, 2016; Hooper et al., 2016; Xia et al., 2021) and software for sample size calculations for stepped wedge designs incorporate this instantaneous treatment (IT) effect assumption. Some of these programs also allow one to prespecify a transition period or pre-specified fractional treatment effects (Hughes et al., 2015) and some allow for a linearly increasing or decreasing treatment time effects (see Ouyang et al., 2022 for a review of packages).

More generally, however, the intervention effect may vary as an arbitrary function of exposure time. Importantly, Kenny et al. (2022) show that assuming the intervention effect is instantaneous and constant; when it is not can lead to extremely misleading estimates of the intervention effect (see, for example, Fig. 2 in Kenny et al. (2022)). As an

✉ James P. Hughes
jphughes@uw.edu

[1] Department of Biostatistics, University of Washington, Seattle, WA 98195, USA

[2] Department of Population Health, Division of Biostatistics, New York University, New York, NY, USA

Published online: 20 September 2023

	Study period (year/month)													
	Year 1				Year 2				Year 3				Year 4	
Sequence	3	6	9	12	3	6	9	12	3	6	9	12	3	6
1	OP/TAU	TAU	TAU	TAU	P1	P2	P3	P4	F5	F6	F7	F8	F9	F10
2		OP/TAU	TAU	TAU	TAU	P1	P2	P3	P4	F5	F6	F7	F8	F9
3			OP/TAU	TAU	TAU	TAU	P1	P2	P3	P4	F5	F6	F7	F8
4				OP/TAU	TAU	TAU	TAU	P1	P2	P3	P4	F5	F6	F7
5					OP/TAU	TAU	TAU	TAU	P1	P2	P3	P4	F5	F6

Fig. 1 ADDRESS-BP trial design. Each row corresponds to a sequence and there are five health care facilities per sequence. No data are collected in empty cells. In each sequence there are 4 "control" periods (OP, on-boarding period; TAU, training as usual) and up to 10 "exposed" periods (i.e., up to 10 exposure times). Exposure to the PATCH implementation strategy begins at exposure time P1 and there is scientific interest in assessing both the short-term (P#, PATCH implementation strategy at exposure time #) and long-term (F#, follow-up periods at exposure time #) effects of the strategy

alternative, Kenny et al. (2022) propose an "exposure time indicator" (ETI) model for analysis of stepped wedge trials. In this manuscript, we recap the ETI model of Kenny et al. (2022) and illustrate how one can do sample size calculations for various estimands under an ETI model using the R software package, `swCRTdesign`.

Our motivating example is the ADDRESS-BP trial, a type 3 hybrid implementation study. The ADDRESS-BP study addresses a critical health issue: Black adults have the highest rate of hypertension in the USA, nearly twice the risk of fatal stroke (Egan et al., 2010; Centers for Disease Control and Prevention, 2000), and a 50% higher rate of cardiovascular disease mortality (Giles et al., 1995; Klag et al., 1997; Pavlik et al., 1997; Singh et al., 1996) compared to the general population (Gyamfi et al., 2022). Barriers to hypertension control exist at multiple levels, including the patient (e.g., poor adherence and lack of patient engagement), the physician (e.g., clinical inertia), and the health system (e.g., poor integration of clinical decision support tools into care). ADDRESS-BP is designed to evaluate a novel implementation strategy to reduce hypertension in African Americans and address racial disparities in cardiovascular and pulmonary health, disease, and disease risk factors in high-burden communities in the USA.

The ADDRESS-BP study uses a stepped-wedge design (see Fig. 1) with five sequences and 14 periods to evaluate PATCH, a tailored practice facilitation, and community health worker implementation strategy designed to promote adoption of an evidence-based intervention (Practice support and Community Engagement—PACE) for the treatment of uncontrolled hypertension. PACE is delivered to patients by nurse case managers and community health workers, who address patients' social risk factors (described as "specific adverse social conditions that are associated with poor health") for management of chronic conditions, including hypertension. The PATCH implementation strategy will be compared to a training-as-usual (TAU) implementation strategy; key outcomes are facility adoption of PATCH and blood pressure control by individuals treated at these facilities.

In this manuscript we first review the statistical methodology necessary to conduct power and sample size evaluation for stepped-wedge designs similar to ADDRESS-BP. We then illustrate the use of the developed tools using the ADDRESS-BP design with a goal of facilitating the use of these methods for future prevention and intervention studies.

Methodology

Let Y_{ijk} represent the outcome measurement for person k ($k = 1 ,\ldots, m_{ij}$) in cluster i ($i = 1 ,\ldots, I$) at time period j ($j = 1 ,\ldots, J$) (Table 1 summarizes the notation used in this section). Mixed effects models are convenient for sample size calculations in stepped-wedge trials as they allow one to explicitly specify a correlation structure for the observations (Xia et al., 2021). Therefore, we use the following general model for analysis of stepped wedge design trials:

$$g\left(\mu_{ijk}|b_{ij}\right) = X_{ij}\beta + Z_{ij}b_{ij}$$

where μ_{ijk} is the mean of Y_{ijk}, g is a link function, and X_{ij} is a design matrix for the fixed effects, $\beta = (\Gamma, \delta)^T$, which are partitioned into a vector of parameters capturing the underlying trend in study time (Γ) and a scalar or vector of parameters capturing the intervention effect (δ) (more on this below). Finally, we use Z_{ij} (a design matrix) and

Table 1 Notation

Notation	Description
Y_{ijk}	outcome measurement for person k ($k = 1, \ldots, m_{ij}$) in cluster i ($i = 1, \ldots, I$) at time j ($j = 1, \ldots, J$)
μ_{ijk}	mean of Y_{ijk}
$g(\ )$	link function in the generalized linear models framework
X_{ij}	design matrix for the fixed effects β
Z_{ij}	design matrix for the random effects $b_{i\bullet}$
$\beta = (\Gamma, \delta)^T$	β is a vector of parameters composed of a vector of parameters that capture the underlying trend in study time (Γ) and a scalar (δ) or vector ($\delta = (\delta_1, \ldots, \delta_S)$) of parameters capturing the intervention effect over S exposure times
$b_{i\bullet}$	Vector of random effects
$H = \left(h_1, h_2, \ldots, h_S\right)$	Weighting vector (with constraint $\sum h_i = 1$) used with δ to form the estimand Ψ
$\Psi = \sum_{s=1}^{S} h_s \delta_s$	Summary intervention effect of interest
Σ	Variance of $\hat{\beta}$
$\mathrm{Var}_0\left(\hat{\Psi}\right), \mathrm{Var}_a\left(\hat{\Psi}\right)$	Variance of $\hat{\Psi}$ under the null and alternative hypotheses, respectively

$b_{i\bullet} \sim N(0, D)$ (a vector of random effects) to quantify the correlation among observations in cluster i. Random effects commonly included in the analysis of stepped-wedge trials include cluster, cluster by time, cluster by treatment, and, for cohort designs, an individual random effect (Hooper et al., 2016). Decaying auto-correlation structures have also been proposed, although these cannot be represented with random effects (Kasza et al., 2017). Li et al (2021) provide a comprehensive overview of the mixed-effect model framework for stepped-wedge designs. Ouyang et al. (2023) show the relationship between the random effects parameterization given above and the correlation coefficient (i.e., intra-cluster correlations) parameterization for specifying dependence in the data.

Hughes et al. (2015) discuss approaches to eliciting variances of the random effects for the purposes of power calculations. For binomial or Poisson endpoints, these variance components are typically expressed on the proportion or count scale, respectively, but may be transformed to the scale corresponding to the link function g using the delta method[1].

Let $Var\left(\hat{\beta}\right) = \Sigma$. Note that Σ is a function of the study design and the variances of the random effects, quantities that must be specified during the design of the study to determine power. Following Xia et al. (2021), Σ may be expressed as

$$\Sigma = \left(X^T V^{-1} X\right)^{-1}$$

where

$$V = W + ZDZ^T.$$

And W is a diagonal matrix with entries $w_i = \phi a_i v\left(\mu_i\right)\left[g'\left(\mu_i\right)\right]^2$ (see Table 2). For a normal distribution, $w_i = \sigma^2$. Σ may be partitioned as

$$\Sigma = \begin{pmatrix} \Sigma_{\Gamma\Gamma} & \Sigma_{\Gamma\delta} \\ \Sigma_{\Gamma\delta}^T & \Sigma_{\delta\delta} \end{pmatrix}.$$

As noted in the introduction, most models for analysis and sample size calculations assume that the intervention effect,δ, is a scalar and that the corresponding element of X_{ij} is a simple 0/1 measure, indicating whether the intervention is turned off or on in cluster i at time j (this is the IT model mentioned above). Alternatively, Kenny et al. (2022) assume that $\delta = (\delta_1, \ldots, \delta_S)$, where δ_s is the effect of the intervention after a cluster has been exposed to the intervention for s time units (exposure time $= s$) (the ETI model) and S is

Table 2 Variance function values for selected distributions with canonical links

Distribution (link)	ϕ	a	$v(\mu)$	$g(\mu)$	$g\prime(\mu)$
Normal (identity)	σ^2	1	1	μ	1
Bernoulli (logit)	1	1	$\mu(1-\mu)$	$\log\left(\frac{\mu}{1-\mu}\right)$	$\frac{1}{\mu(1-\mu)}$
Poisson (log)	1	$\frac{1}{m}$	μ	$\log(\mu)$	$\frac{1}{\mu}$
Binomial (logit)	1	$\frac{1}{m}$	$\mu(1-\mu)$	$\log\left(\frac{\mu}{1-\mu}\right)$	$\frac{1}{\mu(1-\mu)}$

[1] Specifically, $Var_{g(x)} = Var_x g\prime(x)^2$. For example, for a binomial endpoint on the logit scale, we would apply the transformation $Var_{logit(p)} = \frac{Var_p}{p^2(1-p)^2}$ to random effect variances expressed on the proportion scale.

the maximum exposure time (note that $S = J - 1$ in a classic stepped wedge design where $J =$ the number of sequences + 1).

Under the ETI model, the entire vector δ may be of interest, or the investigator may define a function of δ as a summary of the intervention effect. For example, Kenny et al. (2022) define the time-averaged treatment effect over an exposure interval s_1 to s_2 as

$$\Psi\left[s_1, s_2\right] \equiv \frac{1}{s_2 - s_1 + 1} \sum\nolimits_{r=s_1}^{s_2} \delta_r.$$

Alternatively, interest may lie in the treatment effect at a specific time s (point treatment effect):

$$\Psi_s = \delta_s.$$

More generally, most summary estimands of interest can be written as a linear combination of the δ_s:

$$\Psi = \sum\nolimits_{s=1}^{S} h_s \delta_s. \tag{1}$$

for known constants h_s. Letting $H = \left(h_1, h_2, \dots, h_S\right)$ (with constraint $\sum h_i = 1$), $\widehat{\delta} = (\widehat{\delta}_1, \widehat{\delta}_2, \dots, \widehat{\delta}_S)$, and $\Sigma_{\delta\delta} = \mathrm{Var}(\widehat{\delta})$ we can write the variance of the estimated intervention effect, $\widehat{\Psi}$, as

$$\mathrm{Var}\left(\widehat{\Psi}\right) = H \Sigma_{\delta\delta} H^T.$$

Therefore, given a hypothesized intervention effect (ψ), weighting vector (H), pre-specified study design, and variance components, and assuming a two-sided type-1 error rate of α, the power of the study can be computed as

$$\Phi\left(\frac{|\Psi| - \Phi^{-1}\left(1 - {}^{\alpha}/_{2}\right) * \sqrt{\mathrm{Var}_0\left(\widehat{\Psi}\right)}}{\sqrt{\mathrm{Var}_{\mathrm{a}}\left(\widehat{\Psi}\right)}}\right) \tag{2}$$

where Φ is the normal cumulative distribution function and $\mathrm{Var}_0\left(\widehat{\Psi}\right)$ and $\mathrm{Var}_{\mathrm{a}}\left(\widehat{\Psi}\right)$ are the variances under the null and alternative hypotheses, respectively. For the normal distribution with identity link, power depends only on Ψ and not the individual δ_s (and $\mathrm{Var}_0\left(\widehat{\Psi}\right)$ and $\mathrm{Var}_{\mathrm{a}}\left(\widehat{\Psi}\right)$ are equal in this case). For non-identity links, power will depend on the individual δ_s through $\mathrm{Var}_{\mathrm{a}}\left(\widehat{\Psi}\right)$.

The R package `swCRTdesign` can be used to implement the methods described above for cross-sectional and closed cohort designs (specifically, `swCRTdesign` implements Eq. (2) which does not, in general, simplify to a design effect (Hooper et al., 2016)). `swCRTdesign` has two functions for computing study power, `swPwr()` and `swGlmPwr()`. `swPwr()` computes power using an identity link for data with a normal or binomial distribution. `swGlmPwr()` can compute power for binomial data with a logit link or Poisson data with a log link. Both functions can compute power for either an IT model or an ETI model with user-specified H vector (allowing specification of a time-averaged treatment effect, a point treatment effect, or other summaries) and both can incorporate random effects for cluster, cluster by time, cluster by treatment, and, for cohort designs, individual.

Data

In our motivating example, ADDRESS-BP, the unit of randomization is the practice facility. There are 25 practice facilities (5 per sequence) and three treatment teams which consist of one practice facilitator and one community health worker per team; each treatment team will work with multiple facilities. During an onboarding period, each facility will establish a closed cohort of approximately 20 individuals with uncontrolled hypertension. Facilities will then provide training as usual (TAU) for the PACE intervention until the PATCH implementation strategy is introduced, as illustrated in Fig. 1.

In this paper we focus on the clinical outcome of blood pressure control (yes/no) in the individual patient. During the TAU periods we expect 40% of participants will achieve blood pressure control. We expect the PATCH implementation strategy may take up to 6 months (2 time periods—each time period is 3 months) to achieve full effect so we propose to evaluate PATCH 6 to 12 months following introduction (exposure times P3 and P4 in Fig. 1). We believe PATCH will improve blood pressure control to 60% during this period. To evaluate the sustainability of this implementation strategy, we will also estimate the proportion of patients achieving/maintaining blood pressure control in exposure times 5–10 following PATCH introduction (the follow-up period—F5–F10 in Fig. 1).

Preliminary data are available from another NYU study, Advancing Medication Adherence for Latinos with Hypertension through a Team-based Care Approach, that also utilized a stepped-wedge cluster randomized design and practice facilitation to improve blood pressure control (Schoenthaler et al., 2021); these data enabled estimation of variance components for practice facility, practice facility by time, and individual. As noted above, each treatment team will serve multiple facilities and may represent an additional source of variation. However, due to software limitations, we were unable to include that source of variation in our power calculations.

Table 3 Parameter values used for power calculations for ADDRESS-BP trial. See Fig. 1 for definition of TAU, P3, P4

Parameter description	Value	
Number of observations per facility per time period	0 pre-onboarding; otherwise, $n = 20$	
Outcome percent during TAU period	40%	
Outcome percent during exposure times three (P3) and four (P4)	60%	
Time trend (on logit scale)	0.08/period	
	Variance (on logit scale)	Intra-cluster correlation[a]
Cluster variance/between-period ICC	0.1316	0.022
Cluster*time variance/within-period ICC	0.1974	0.054
Individual variance/within-individual ICC	2.5	0.430

ICC intra-cluster correlation

[a] ICCs are expressed on logit scale with $\sigma^2 = \pi^2/3 \approx 3.290$ (Eldridge et al., 2009). See Ouyang et al. (2023) table 1 for definitions of ICCs

Results

The R package `swCRTdesign` was used to compute power for the ADDRESS-BP study. Table 3 lists parameter values for the power calculations. R code for each result given below is included in the Appendix and referenced below as R# (e.g., R1, R2, etc.). Note that all the results from `swGlmPwr` (R1–R5) have been verified by simulation.

Figure 2 shows the power comparing the primary outcome of interest, the average blood pressure control over exposure times 3 and 4, shown as P3 and P4 in Fig. 1, versus TAU based on the ETI model of Kenny et al. (2022) (R1). Since the maximum number of exposure time periods is 10 (P1–F10 in Fig. 1), this comparison corresponds to simply averaging the responses in periods P3 and P4 (i.e., $H = (0, 0, 0.5, 0.5, 0, 0, 0, 0, 0, 0)$ in Eq. (1)). For the alternative of 60% of patients achieving blood pressure control, the power is 82% (R2). For comparison, an IT model that compares all PATCH periods to TAU would have power 99.9% (R3) although, as noted above, this model is susceptible to substantial bias and possible loss of power if the IT assumption is wrong.

Note that the power analysis shown above assumes that no data are available for cohort members prior to the onboarding period (Fig. 1). An alternative would be to form the cohorts in all clusters at the start of the study (start of year 1) so that data were available from all clusters in all time periods. In this case, the power for comparing exposure times 3 and 4 to TAU would increase to 92% (R4).

Interestingly, the power for comparing the average treatment effect over the follow-up period (exposure times F5–F10) versus TAU under the same alternative is only 39% (R5). Even though there is more data in the follow-up period, power is lower because (i) there are fewer cluster periods at later exposure times and (ii) there are no direct vertical (between cluster) comparisons between treatment and control in exposure times 5 through 10.

The ETI model is inefficient for estimating the treatment effect compared to the IT model when the treatment effect is constant. To recover some of the efficiency, but still reduce the risk of bias, one could assume a piecewise constant treatment effect (this is effectively a version of the spline approach discussed by Kenny et al. (2022)). If we assume piecewise constant treatment effects for exposure times P1–P2, exposure times P3–P4, and exposure times F5–F10 then the power to compare exposure times P3–P4 versus TAU is 94% and to compare exposure times F5–F10 to TAU is 75% (R6). Note that including data from exposure times P1 and P2 in the analysis improves power by contributing to estimation of the temporal trend component of the model, even though these periods are not part of the primary or secondary hypotheses.

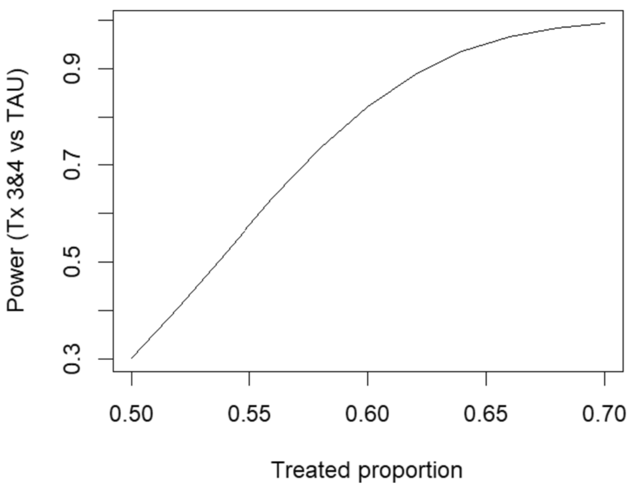

Fig. 2 Power for comparing treatment exposure times 3 and 4 versus TAU. Assumed proportion during TAU is 0.40

Discussion

We have presented a set of generalizable tools for designing stepped-wedge studies that can be used to evaluate the impact of interventions during real-world implementation, and we have demonstrated their use in the ADDRESS-BP study, a type 3 hybrid implementation study designed to address racial disparities in health care by evaluating a novel implementation strategy for a practice-based intervention to reduce hypertension in African American communities.

The ETI model of Kenny et al. (2022) relaxes the assumption of an immediate and constant response to treatment in step-wedge studies and allows the treatment effect to vary as a function of exposure time. In this article we have shown how power calculations for analysis of a stepped-wedge trial based on the ETI model can be conducted. The package, `swCRTdesign`, can be used to implement these calculations. Importantly, when the treatment effect varies as a function of exposure time, careful scientific thought must be used to define the estimand of interest and this choice can strongly influence power. This effectively means defining the weighting vector H (and the intervention effect size associated with H). However, this is the only additional requirement needed for sample size calculations for the ETI model compared to the traditional IT model. An alternative approach is discussed by Maleyeff et al. (2022) in which they assume that the treatment effect varies randomly over exposure time; they then provide power calculations for estimating the average treatment effect.

It is at present unclear how often the treatment effect varies significantly with exposure time but Kenny et al. (2022) recommend that the IT model should not be used unless the assumption of an immediate and constant treatment effect is justifiable based on contextual knowledge of the intervention. They also note a need to reanalyze data from past stepped-wedge trials to understand the prevalence of time-varying treatment effects.

ETI-based estimators are less efficient than estimates based on an IT model but are robust against the assumption of a constant treatment effect, violation of which can lead to an extremely misleading treatment effect estimate. One approach to recovering some efficiency and improving power is to model the treatment effect using a piecewise constant or higher-order spline function. The piecewise constant approach can be particularly useful when the period of interest corresponds directly to one of the constant pieces of the spline. However, whenever one assumes that the treatment is constant over multiple exposure times, there is a potential for the type of misleading treatment estimates observed by Kenny et al. (2022). In the case of the ADDRESS-BP trial, we believe it is unlikely that combining exposure times P3 and P4 would lead to any problems because (i) only two adjacent periods are being combined and (ii) those two periods have similar information content. In contrast, there is a greater risk in assuming that the treatment effect is constant over the entire follow-up period, i.e., exposure times F5–F10.

We have also seen in the ADDRESS-BP case study that, under the ETI model, power is greater for studying the treatment effect immediately after the transition from control to treatment compared to treatment effects at longer exposure times, consistent with Kasza and Forbes (2019). We also noted in the ADDRESS-BP trial that including data that does not directly inform the treatment effect estimate of interest can improve power by providing more precise estimation of other model parameters. More research is needed, however, to fully understand how other modeling choices (e.g., using spline-based models for the underlying time trend and/or the exposure time trend, extending follow-up, etc.) affect power for testing short-term and long-term treatment effects.

Interestingly, because of the partial collinearity between study time and treatment, as well as the complex correlation structure in the stepped-wedge design, standard statistical intuition regarding power may be misleading. For example, in the ADDRESS-BP study power analysis, the power for comparing exposure times P3 and P4 to TAU is greater than the power for comparing exposure times F5–F10 to TAU even though there is quantitatively more data for the latter comparison. This is a result of several interacting factors: (i) decreasing information (and, therefore, greater variation) for exposure times F7 and above; (ii) the lack of a between-cluster comparison between exposure times F5–F10 and TAU; (iii) the correlation structure of the model which affects the relative information provided by between-cluster versus within-cluster comparisons; and (iv) the need to estimate study time effects, which are partly collinear with treatment.

One limitation of the power analysis of the ADDRESS-BP study presented above is that we have not accounted for the effect of the treatment teams, each of which will work across multiple facilities. Davis-Plourde et al. (2023) describe sample size calculations for a multi-level stepped-wedge study with nested clusters (in this case, facilities nested within treatment teams) and their idea of adding additional random effects to account for additional levels of clustering could be used with both the IT and ETI models. However, due to software limitations, we are not able to account for multiple levels of clustering in this analysis although some insight is possible by including a nonzero `eta` argument in calls to `swGlmPwr.` `eta` characterizes the variation in the treatment effect between facilities (whereas, ideally, we would like to include variation in the treatment effect between treatment teams—see Brown et al., 2022). We found that power generally declines as eta increases, suggesting that inclusion of a treatment team effect may decrease power modestly. The effect of treatment teams should be included in the final analysis of the ADDRESS-BP data.

The ADDRESS-BP trial will evaluate a multi-component implementation strategy leveraging practice facilitation, nurse case managers, and community health workers to promote implementation of home blood pressure monitoring and incorporation of social determinants of health to improve blood pressure control (Odedosu et al., 2012; Pickering et al., 2008). The study will use the ETI model and is well powered to address a clinically meaningful increase in rate of blood pressure control at 12 months following initiation of PATCH.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11121-023-01587-1.

Funding This research was supported by NIH grants AI29168, UG3HL151310-01, and UH3HL151310-04.

Declarations

Ethics Approval Not applicable.

Consent to Participate Not applicable.

Conflict of Interest The authors declare no competing interests.

References

Brown, C. H., Hedeker, D., Gibbons, R. D., Duan, N., Almirall, D., Gallo, C., Burnett-Zeigler, I., Prado, G., Young, S. D., Valido, A., & Wyman, P. A. (2022). Accounting for context in randomized trials after assignment. *Prevention Science, 23*, 1321–1332.

Copas, A. J., Lewis, J. J., Thompson, J. A., Davey, C., Baio, G., & Hargreaves, J. R. (2015). Designing a stepped wedge trial: Three main designs, carry-over effects and randomization approaches. *Trials, 16*, 352.

Centers for Disease Control and Prevention. (2000). Age-specific excess deaths associated with stroke among racial/ethnic minority populations - United States, 1997. *Morbidity and Mortality Weekly Report, 49*(5), 94–7.

Davis-Plourde, K., Taljaard, M., & Li, F. (2023). Sample size considerations for stepped wedge designs with subclusters. *Biometrics, 79*, 98–112.

Egan, B. M., Zhao, Y., & Axon, R. N. (2010). US trends in prevalence, awareness, treatment, and control of hypertension, 1988–2008. *JAMA, 303*, 2043–2050.

Eldridge, S. M., Ukoumunne, O. C., & Carlin, J. B. (2009). The intra-cluster correlation coefficient in cluster randomized trials: A review of definitions. *International Statistical Review, 77*, 378–394.

Giles, W. H., Kittner, S. J., Hebel, J. R., Losonczy, K. G., & Sherwin, R. W. (1995). Determinants of Black-White differences in the risk of cerebral infarction. The National Health and Nutrition Examination Survey Epidemiologic Follow-up Study. *Archives of Internal Medicine, 155*, 1319–1324.

Gyamfi, J., Cooper, C., Barber, A., Onakomaiya, D., Lee, W., Zanowiak, J., Mansu, M., Diaz, L., Thompson, L., Abrams, R., Schoenthaler, A., Islam, N., & Ogedegbe, G. (2022). Needs assessment and planning for a clinic-community-based implementation program for hypertension control among blacks in New York City: A protocol paper. *Implementation Science Communications, 3*(1), 96.

Hemming, K., & Taljaard, M. (2016). Sample size calculations for stepped wedge and cluster randomised trials: A unified approach. *Journal of Clinical Epidemiology, 69*, 137–146.

Hemming, K., Taljaard, M., McKenzie, J. E., et al. (2018). Reporting of stepped wedge cluster randomized trials: Extension of the CONSORT 2010 statement with explanation and elaboration. *British Medical Journal, 363*,

Hemming, K., Taljaard, M., & Forbes, A. (2017). Analysis of cluster randomized stepped wedge trials with repeated cross-sectional samples. *Trials, 18*, 110.

Hooper, R., & Burke, L. (2015). Cluster randomised trials with repeated cross sections: Alternatives to parallel group designs. *British Medical Journal, 350*,

Hooper, R., Teerenstra, S., de Hoop, E., & Eldridge, S. (2016). Sample size calculation for stepped wedge and other longitudinal cluster randomized trials. *Statistics in Medicine, 35*, 4718–4728.

Hughes, J. P., Granston, T. S., & Heagerty, P. J. (2015). On the design and analysis of stepped wedge trials. *Contemporary Clinical Trials, 45*(Pt A), 55–60.

Hussey, M. A., & Hughes, J. P. (2007). Design and analysis of stepped wedge cluster randomized trials. *Contemporary Clinical Trials, 28*, 182–191.

Li, F., Hughes, J. P., Hemming, K., Taljaard, M., Melnick, E. R., & Heagerty, P. J. (2021). Mixed-effects models for the design and analysis of stepped wedge cluster randomized trials: An overview. *Statistical Methods in Medical Research, 30*, 612–639.

Kasza, J., Hemming, K., Hooper, R., Matthews, J. N. S., & Forbes, A. B. (2017). Impact of non-uniform correlation structure on sample size and power in multiple-period cluster randomised trials. *Statistical Methods in Medical Research, 28*, 703–716.

Kasza, J., & Forbes, A. B. (2019). Information content of cluster–period cells in stepped wedge trials. *Biometrics, 75*, 114–152.

Kazsa, J., Bowden, R., Hooper, R., & Forbes, A. B. (2022). The batched stepped wedge design: A design robust to delays in cluster recruitment. *Statistics in Medicine, 41*, 3627–3641.

Kenny, A., Voldal, E., Xia, F., Heagerty, P. J., & Hughes, J. P. (2022). Analysis of stepped wedge cluster randomized trials in the presence of a time-varying treatment effect. *Statistics in Medicine, 41*, 4311–4339.

Klag, M. J., Whelton, P. K., Randall, B. L., Neaton, J. D., Brancati, F. L., & Stamler, J. (1997). End-stage renal disease in African-American and White men. 16-year MRFIT findings. *JAMA, 277*, 1293–1298.

Maleyeff, L., Li, F., Haneuse, S., & Wang, R. (2022). Assessing exposure-time treatment effect heterogeneity in stepped wedge cluster randomized trials. *Biometrics*, in press (https://doi.org/10.1111/biom.13803).

Odedosu, T., Schoenthaler, A., Vieira, D. L., Agyemang, C., & Ogedegbe, G. (2012). Overcoming barriers to hypertension control in African Americans. *Cleveland Clinic Journal of Medicine, 79*, 46–56.

Ouyang, Y., Li, F., Pressier, J. F., & Taljaard, M. (2022). Sample size calculators for planning stepped-wedge cluster randomized trials: A review and comparison. *International Journal of Epidemiology, 51*, 2000–2013.

Ouyang, Y., Hemming, K., Li, F., & Taljaard, M. (2023). Estimating intra-cluster correlation coefficients for planning longitudinal cluster randomized trials: A tutorial. *International Journal of Epidemiology*, in press.

Pavlik, V. N., Hyman, D. J., Vallbona, C., Toronjo, C., & Louis, K. (1997). Hypertension awareness and control in an inner-city African-American sample. *Journal of Human Hypertension, 11*, 277–283.

Pickering, T. G., Miller, N. H., Ogedegbe, G., Krakoff, L. R., Artinian, N. T., & Goff, D. (2008). Call to action on use and reimbursement for home blood pressure monitoring: A joint scientific statement from the American Heart Association, American Society Of Hypertension, and Preventive Cardiovascular Nurses Association. *Hypertension, 52*, 10–29.

Schoenthaler, A., De La Calle, F., Soto, A., Barrett, D., Cruz, J., Payano, L., Rosado, M., Adhikar, S., Ogedegbe, G., & Rosal, M. (2021). Bridging the evidence-to-practice gap: A stepped wedge cluster randomized controlled trial evaluating practice facilitation as a strategy to accelerate translation of a multi-level adherence intervention into safety net practices. *Implementation Science Communications, 2*(1), 21. Erratum in: Implement Sci Commun. 2021 Oct 21;2(1):122.

Singh, G. K., Kochanek, K. D., & MacDorman, M. F. (1996). Advance report of final mortality statistics, 1994. Monthly Vital Statistics Report, 45:3, supp. Hyattsville, Maryland: National Center for Health Statistics.

Xia, F., Hughes, J. P., Voldal, E., & Heagerty, P. J. (2021). Power and sample size calculation for stepped-wedge designs with discrete outcomes. *Trials, 22*(1), 598.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.