medRxiv

Accounting for external factors and early intervention adoption in the design and analysis of stepped-wedge designs: Application to a proposed study design to reduce opioid-related mortality

Lior Rennert[a]*, Moonseong Heo[a], Alain H. Litwin[b,c,d], Victor De Gruttola[e]

Background: Stepped-wedge designs (SWDs) are currently being used to investigate interventions to reduce opioid overdose deaths in communities located in several states. However, these interventions are competing with external factors such as newly initiated public policies limiting opioid prescriptions, media awareness campaigns, and social distancing orders due to the COVID-19 pandemic. Furthermore, control communities may prematurely adopt components of the proposed intervention as they become widely available. These types of events induce confounding of the intervention effect by time. Such confounding is a well-known limitation of SWDs; a common approach to adjusting for it makes use of a mixed effects modeling framework that includes both fixed and random effects for time. However, these models have several shortcomings when multiple confounding factors are present.

Methods: We discuss the limitations of existing methods based on mixed effects models in the context of proposed SWDs to investigate interventions intended to reduce mortality associated with the opioid epidemic, and propose solutions to accommodate deviations from assumptions that underlie these models. We conduct an extensive simulation study of anticipated data from SWD trials targeting the current opioid epidemic in order to examine the performance of these models under different sources of confounding. We specifically examine the impact of factors external to the study and premature adoption of intervention components.

Results: When only external factors are present, our simulation studies show that commonly used mixed effects models can result in unbiased estimates of the intervention effect, but have inflated Type 1 error and result in under coverage of confidence intervals. These models are severely biased when confounding factors differentially impact intervention and control clusters; premature adoption of intervention components is an example of this scenario. In these scenarios, models that incorporate fixed intervention-by-time interaction terms and an unstructured covariance for the intervention-by-cluster-by-time random effects result in unbiased estimates of the intervention effect, reach nominal confidence interval coverage, and preserve Type 1 error, but may reduce power.

Conclusions: The incorporation of fixed and random time effects in mixed effects models require certain assumptions about the impact of confounding by time in SWD. Violations of these assumptions can result in severe bias of the intervention effect estimate, under coverage of confidence intervals, and inflated Type 1 error. Since model choice has considerable impact on study power as well as validity of results, careful consideration needs to be given to choosing an appropriate model that takes into account potential confounding factors.

Copyright © 2020 The authors

medRxiv

1. Introduction

Stepped wedge designs (SWD) are a uni-directional crossover design in which clusters switch from the control to the intervention condition at varying time points. The first phase is usually a baseline period in which no clusters receive intervention. During the second phase, clusters are randomly assigned to intervention at pre-selected time points until all clusters receive the intervention. The third phase corresponds to the follow-up period in which all clusters receive the intervention. An example of a SWD is provided in Figure 1 and discussed in Section 1.1.

SWDs can be useful in public health settings where rolling out the intervention to all clusters at once is infeasible; they also ensure that all clusters in the study eventually receive the intervention.[1,2] The SWD is particularly suitable for implementing and evaluating complex health interventions.[3–5] The current COVID-19 pandemic and opioid epidemic provide settings in which such designs may prove useful; effective combinations of interventions are needed as soon as possible, but rolling them out to each community in a short time period may not be feasible. In this paper, we will focus on recently proposed SWDs to combat the opioid epidemic. In 2019, the National Institute on Drug Abuse (NIDA) awarded Kentucky and Ohio roughly $100 million each to implement an integrated set of interventions in high-risk communities using a SWD, with the objective of reducing opioid overdose deaths by 40% over a three-year period.[6–9]

Like most clinical trial designs, SWDs may be impacted by external factors that influence the primary outcomes. The term "rising tide" has been used to describe the situation when there is a drift towards improvement in the outcomes due to factors external to the study.[10] A rising tide may be seen in the current opioid epidemic, where the severity of this crisis has led to new public policies, media awareness campaigns, and external interventions that will improve outcomes in concurrence with any proposed interventions. For example, public policies were implemented to limit opioid prescriptions in both Kentucky and Ohio in the summer of 2017.[11] In addition, the Center for Disease Control and Prevention launched intensive media awareness campaigns on the dangers of opioids in these states in September of 2017.[12] In the current COVID-19 pandemic, interventions rolled out at the community level will compete with social distancing measures, novel treatments, and other interventions aimed at improving patient care. Such external factors may confound the intervention effect estimate in SWDs. This occurs when (i) factors external to the study (e.g., new public health policy) influence the primary outcomes over time, and (ii) the proportion of communities exposed to the intervention also increases with calendar time. This situation has been referred to as confounding by calendar time.[2,13]

In certain scenarios, confounding factors may differentially impact intervention and control communities. For example, while all communities may be exposed to mandated state-wide policies and media awareness campaigns, control communities may choose to prematurely adopt components of the intervention prior to the scheduled roll-out time, should they become available. The latter scenario may arise from external forces, such as implementation of intervention components in control communities by policymakers, or easy access to intervention components by participants in control communities. Therefore, premature adoption of intervention components, which is a version of cluster treatment non-compliance, can also be viewed as a special case of confounding by time which impacts control communities only.

Failure to account for confounding by time in these settings can result in severely biased intervention effect estimates,[2,3,13–19] and may lead to both Type 1 and Type 2 errors. Hussey and Hughes suggested the use of mixed effects models for analyzing data from SWD.[16] To account for potential confounding by time, they recommend the incorporation of fixed time effects. However, models that incorporate secular trends common to all clusters may not be appropriate in

[a] *Department of Public Health Sciences, Clemson University, Clemson, U.S.A.*
[b] *University of South Carolina School of Medicine, Greenville, SC, USA*
[c] *Prisma Health, Department of Medicine, Greenville, SC, USA*
[d] *Clemson University School of Health Research, Clemson, SC, USA*
[e] *Department of Biostatistics, Harvard T.H. Chan School of Public Health, Harvard University, Boston, U.S.A.*
[*] *Correspondence to: Lior Rennert, Department of Public Health Sciences, Clemson University, Clemson, U.S.A. E-mail: liorr@clemson.edu*

medRxiv

this setting, because only some of the communities may be exposed to the external factors. Furthermore, the impact of these external factors on the outcome will likely differ in each community.

For these reasons, Girling and Hemming, Hooper et al., and Hemming et al. have suggested incorporating random cluster-by-time effects into the mixed effects models.[13,17,20] This class of models captures confounding by time through these random effects. This is especially useful when the timing and level of exposure to confounding factors are unknown. However, these models implicitly assume that the random effect variance is the same for all clusters across all time points, and that the random time effects are independent within each cluster. Kazca et al. propose models with more general within-cluster correlation structures.[21] These models are more applicable to the scenarios discussed here, since outcomes within a community are more likely to be similar for time periods before or after exposure to confounding factors. Furthermore, since exposure to confounding factors may increase with time, variation in the outcome may change with time as well.

The models described above assume that the random effects are identically distributed across clusters. This assumption may not appropriate in settings where the impact of confounding factors systematically differs between intervention and control communities. Kazca et al showed that the random effect covariance structure can have an important impact on the sample size and power,[21] and that misspecification of this covariance structure can lead to biased estimates of the intervention effect.[22]

In this paper, we discuss confounding in the context of a proposed SWD to reduce opioid-related mortality in South Carolina, and show how trial conclusions can differ based on the choice of the fixed time effects and random effect covariance structure in mixed effects models. Our objectives are to (1) consider the implications of assumptions regarding random effects in mixed models in the setting of confounding by time– with a focus on our motivating example; (2) extend these models to accommodate different sources of confounding by time; (3) conduct an extensive simulation study to examine the performance of mixed effects models under different model assumptions regarding fixed and random time effects. Performance is assessed based on the bias, confidence interval coverage, power, and Type 1 error of the intervention effect estimate. An important component of our simulation study is the generation of latent confounding factors from different distributions than that assumed by the model, i.e. normal distribution for random effects. This allows for investigation of the robustness of our methods to violations of model assumptions.

The outline of this paper is as follows. Section 1.1 introduces the motivating example. In Section 2, we examine existing mixed effects models for SWD, discuss their limitations, and introduce alternative models that improve robustness to confounding by time. Section 3 conducts an extensive simulation study to examine the adequacy of these mixed-effects models under different scenarios of confounding by time. Discussion, extensions, and concluding remarks are provided in Section 4.

1.1. Motivating example: HEALing Communities Study

The HEALing Communities Study: Developing and Testing an Integrated Approach to Address the Opioid Crisis[7] is used as an example to illustrate confounding by time in SWD. The purpose of this initiative was to develop and integrate a set of evidence based interventions using cluster randomized trials to reduce opioid overdose fatalities by 40% over a 3-year period in states nationwide. In response to this research funding announcement, our research team proposed a SWD to implement a comprehensive external facilitation intervention in 18 of South Carolina's counties that were hit hardest by the opioid crisis.

The proposed SWD design is displayed in Figure 1. Each cluster consists of all individuals in a given county. Each time interval corresponds to a 3-month study period. The pre-intervention phase consists of two time periods between (study) month 0 and month 6, in which all clusters are in the control condition. The roll-out phase consists of nine time periods between months 6 through 33, where two clusters are randomly assigned to receive the intervention at the beginning of each time period. The post-intervention phase, in which all clusters receive the intervention, occurs between months 33 through 39. Data pertaining to the communities considered in our proposal are provided in Supplementary Table 1. The

medRxiv

outcome, recorded at the end of each time period, consists of the total number of opioid overdose deaths in each cluster during the 3-month time interval.

Phase	Pre-intervention		Roll-out									Post-intervention	
Study Month	0	3	6	9	12	15	18	21	24	27	30	33	36
Cluster 1	C	C	I	I	I	I	I	I	I	I	I	I	I
Cluster 2	C	C	I	I	I	I	I	I	I	I	I	I	I
Cluster 3	C	C	C	I	I	I	I	I	I	I	I	I	I
Cluster 4	C	C	C	I	I	I	I	I	I	I	I	I	I
Cluster 5	C	C	C	C	I	I	I	I	I	I	I	I	I
Cluster 6	C	C	C	C	I	I	I	I	I	I	I	I	I
Cluster 7	C	C	C	C	C	I	I	I	I	I	I	I	I
Cluster 8	C	C	C	C	C	I	I	I	I	I	I	I	I
Cluster 9	C	C	C	C	C	C	I	I	I	I	I	I	I
Cluster 10	C	C	C	C	C	C	I	I	I	I	I	I	I
Cluster 11	C	C	C	C	C	C	C	I	I	I	I	I	I
Cluster 12	C	C	C	C	C	C	C	I	I	I	I	I	I
Cluster 13	C	C	C	C	C	C	C	C	I	I	I	I	I
Cluster 14	C	C	C	C	C	C	C	C	I	I	I	I	I
Cluster 15	C	C	C	C	C	C	C	C	C	I	I	I	I
Cluster 16	C	C	C	C	C	C	C	C	C	I	I	I	I
Cluster 17	C	C	C	C	C	C	C	C	C	C	I	I	I
Cluster 18	C	C	C	C	C	C	C	C	C	C	I	I	I

Figure 1. Proposed SWD for 18 South Carolina communities. 'C' indicates cluster receives control and 'I' indicates cluster receives intervention. All clusters are in the control condition during the pre-intervention phase (months 0 through 6). During the roll-out phase (months 6 through 33), two clusters crossover to the intervention condition at the beginning of each time period. In the follow-up phase (months 33-39), all clusters receive the intervention.

The proposed external facilitation intervention included the following components: 1) Integration of screening, intervention, and referral to treatment within health care settings; 2) implementation of programs and providers prescribing medication-assisted treatment and linkage to such treatment for people with opioid use disorder (OUD); 3) implementation of evidence-based school and community-based OUD prevention programs; and 4) increasing availability and use of naloxone by first responders and the community. Given the amount of effort and resources currently directed to the fight against the opioid epidemic, there is potential for other events to affect the outcome. For example, in 2018 Executive Order No. 2017-43 was passed in South Carolina; this order set a 5-day limit for certain opioid prescriptions.[23] Also in 2018, the South Carolina Division of Alcohol and Other Drug Abuse Services rolled out a media campaign to raise community awareness of opioid addiction.[24] This campaign included digital, social, and traditional media tactics, and was intended to cover all counties in South Carolina. These external factors are part of a rising tide of interventions and are expected to reduce the opioid-related death count over time. Failing to account for this rising tide in an analysis will cause an upward bias in the estimation of the intervention effect. Similarly, an influx of synthetic opioids into the population will likely be associated with an increase in the death rate over time. A failure to account for this will cause underestimation of the intervention effect and may lead to Type II error. Another possibility is that control communities prematurely adopt components of the intervention before the scheduled roll-out time. For example, the opioid overdose reversal medication naloxone may become widely available in control communities prior to the scheduled roll-out time. Therefore premature adoption of a successful intervention will improve the outcome in control communities and attenuate the estimate of the intervention effect if unaccounted for.

2. Methods

We first introduce some notation. We denote Y_{ij} as the summary measure of the outcome for cluster i during time period j, $i = 1, ..., N$ and $j = 1, ..., n$, where N denotes the number of clusters and n denotes the number of time periods in the study, which are assumed to be common to all clusters. In the motivating example, Y_{ij} is an aggregate count of opioid deaths in county i between months j and $j + 1$. Using summary measures for the outcome in each cluster has implications

medRxiv

when estimating the intervention effect under certain distributional assumptions for the outcome. We discuss this further in Section 2.4.

We assume that the expected outcomes, denoted by $\mu_{ij} = E[Y_{ij}]$, come from a generalized linear mixed effects model (GLMM) with link function g. In the motivating example we consider $\mu_{ij} = E[Y_{ij}/O_i]$, where Y_{ij} assumes a Poisson distribution ($g = \log$ link), and O_i is an offset for the population size of cluster i. We also assume that all clusters receive the full intervention effect immediately after the scheduled implementation, i.e., that the intervention effect does not change with time. We denote the intervention effect by θ, and set the corresponding design matrix $X_{ij} = 1$ if clusters i receives intervention at time j, and 0 otherwise.

2.1. Basic model

To adjust for confounding by calendar time, Hussey and Hughes recommended incorporation of a time effect in the GLMM:[16]

$$g(\mu_{ij}) = \alpha + \theta \times X_{ij} + \beta_j + b_{0i}, \tag{1}$$

where α is the intercept, θ is the intervention effect, β_j is the discrete time effect, and $b_{0i} \overset{\text{iid}}{\sim} N(0, \sigma_{b_0}^2)$ for $i = 1, ..., N$ are the random intercepts for each cluster.

In some cases, the timing and location of confounding factors are known in each cluster and their effects can be modeled. Often, investigators are unaware of all factors affecting the outcome. Throughout this paper we assume exposure to these factors are unknown. By incorporating only a fixed effect for time, the standard Hussey and Hughes model requires that the effects of time are common to all clusters, and that the correlation between any two observations in the same cluster are independent of time.

2.2. Random period models under confounding due to external factors

Since misspecification of the time effects can lead to biased estimates of the intervention effect and its standard error, it has been recommended to incorporate a random cluster-by-time period interaction effect.[13, 17, 20] This formulation, referred to as the Hooper/Girling model,[21, 22] allows the random intercept for each cluster to vary by time period. This model is discussed in Section 2.2.1 below.

2.2.1. Random cluster-by-discrete time effect (uncorrelated, with single variance)

$$g(\mu_{ij}) = \alpha + \theta \times X_{ij} + \beta_j + b_{ij}, \tag{2}$$

where $b_{ij} \overset{\text{iid}}{\sim} N(0, \sigma_b^2)$ are the time-varying random intercepts for each cluster i at time period j, $i = 1, ..., N$ and $j = 1, ..., n$. These cluster-specific random effects are intended to capture the effects of external factors on the outcome by allowing unique secular trends for each cluster.

There are limitations to the model above that arise from the distributional assumptions regarding the random effects b_{ij}: Constant variance over time, independence of the random effects, and identical correlation between two observations in the same cluster between any two time periods. If exposure to confounding factors increases over time, the variance of the random effect may increase as well. The assumption of independence of the random effects may be violated in the scenario where confounding factors impact clusters at all time intervals after exposure. Such a scenario could lead to greater similarity of random effects during time intervals that are entirely before or entirely after exposure than for random effects for a mix of intervals that are both before and after exposure. Furthermore, while the Hooper/Girling model allows both within-cluster and within-period variation, it requires that the correlation between two observations in the same cluster be identical within and across two time intervals.

2.2.2. Random cluster-by-discrete time effect (unstructured covariance) To account for the limitations of model 2, we propose an unstructured covariance for the random cluster-by-discrete time interaction terms:

$$g(\mu_{ij}) = \alpha + \theta \times X_{ij} + \beta_j + b_{ij}^*, \tag{3}$$

where $\boldsymbol{b}_i^* = (b_{i1}^*, ..., b_{in}^*) \stackrel{\text{iid}}{\sim} N(\boldsymbol{0}, \Sigma_{\boldsymbol{b*}})$ for $i = 1, ..., N$. Similar to model 2, the cluster-specific random effects b_{ij}^* allow unique secular trends for each cluster. However, the unstructured covariance matrix $\Sigma_{\boldsymbol{b*}}$ imposes no restrictions on the variance across time periods nor on the correlation between the latent time effects within each cluster, albeit with the assumption that the correlation structure is the same for all clusters. While model 3 imposes less restrictions than model 2, it requires the estimation of $\frac{(n+1) \times n}{2}$ covariance parameters, which may greatly reduce the power to detect an intervention effect and may lead to issues of identifiability of regression parameter estimates.

2.2.3. Random cluster by linear time effect An alternative to models 2 and 3 is to include a random slope for time in each cluster:

$$g(\mu_{ij}) = \alpha + \theta \times X_{ij} + \beta_j + b_{0i} + b_{1i} \times t_j, \tag{4}$$

where $(b_{0i}, b_{1i}) \sim N(0, \Sigma_{\boldsymbol{b_0}, \boldsymbol{b_1}})$ and $\Sigma_{\boldsymbol{b_0}, \boldsymbol{b_1}}$ is the 2×2 covariance matrix for (b_{0i}, b_{1i}), $i = 1, ..., N$, and t_j is the time between the study start date and the beginning of time period j. The random effects b_{1i} allow for unique linear time trends in each cluster. Model 4 assumes the variance in the outcome increases with time and imposes restrictions on the correlation between observations within a cluster, and is therefore more restrictive than model 3.

2.3. *Random period models under confounding due to external factors and early adoption*

We consider the scenario where control clusters, unbeknownst to the investigator, prematurely adopt components of the intervention before the scheduled roll-out time. We refer to this scenario as early adoption. In SWDs, there are a greater number of control clusters at risk for early adoption in the beginning of the study compared to later on. In addition, the probability that a control cluster prematurely adopts aspects of the intervention may change over time. Because exposure to early adoption is dependent on calendar time, this can be viewed as a special case of confounding by time in the control clusters only.

The models described in Section 2.2 are insufficient for this setting where confounding differentially impacts control and intervention clusters. To account for this scenario, such as early adoption by control communities, we consider group×time interaction models which incorporate distinct fixed and random time effects for control clusters. Unlike group×time interaction models for time-varying treatments, these models are intended to account for situations where secular trends systematically differ between intervention and control clusters.

In models 5 through 7, we incorporate a fixed discrete time effect, β_j, that is common to all clusters. Because incorporating a group-by-discrete time interaction may lead to loss of power and issues of identifiability, we assume a fixed linear time effect in the control group by incorporating the term $\gamma \times t_j \times 1_{\{X_{ij}=0\}}$ in models 5 through 7. In this setting, incorporation of a linear time effect in the control group assumes that early adoption of a beneficial intervention causes overdose death rates to monotonically decrease with time. We discuss these models in Sections 2.3.1 through 2.3.3.

2.3.1. Random cluster-by-intervention-by-discrete time effect (uncorrelated, with single variance for each group)

$$g(\mu_{ij}) = \alpha + \theta \times X_{ij} + \beta_j + \gamma \times t_j \times 1_{\{X_{ij}=0\}} + b_{ij} + c_{ij} \times 1_{\{X_{ij}=0\}}, \tag{5}$$

where $\gamma \times t_j$ is the difference in the time effect between control and intervention clusters at time period j, and $c_{ij} \sim N(0, \sigma_c^2)$ is a random time effect for control cluster i at time period j for $i = 1, ..., N$ and $j = 1, .., t_i^I$, where t_i^I

is the time period in which the intervention was scheduled to be rolled-out to cluster i. The (control) cluster-specific random effects, c_{ij}, are intended to capture the effects of premature exposure to intervention components by control clusters. Control clusters that prematurely adopt intervention components are therefore expected to have larger c_{ij} (in magnitude). Model 5 is an extension to Model 2 (Hooper/Girling), in which the variance of the cluster-by-time period random effects systematically differ between intervention and control clusters. This model is subject similar limitations as those discussed in Section 2.2.1.

2.3.2. Random cluster by-intervention-by-discrete time effect (unstructured covariance for each group) To account for the limitations of model 5, we propose allowing an unstructured covariance for the random cluster-by-discrete time interaction terms in each group:

$$g(\mu_{ij}) = \alpha + \theta \times X_{ij} + \beta_j + \gamma \times t_j \times 1_{\{X_{ij}=0\}} + b_{ij}^* + c_{ij}^* \times 1_{\{X_{ij}=0\}}, \tag{6}$$

where $\boldsymbol{c}_i^* = (c_{i1}^*, ..., c_{i,t_i^I}^*) \overset{\text{iid}}{\sim} N(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{c}_i^*})$ for $i = 1, ..., N$. Here $\boldsymbol{\Sigma}_{\boldsymbol{c}_i^*}$ consists of the first t_i^I rows and columns of $\boldsymbol{\Sigma}_{\boldsymbol{c}^*}$, where $\boldsymbol{\Sigma}_{\boldsymbol{c}^*}$ is the unstructured covariance matrix for a control cluster scheduled to receive the intervention during the final period of the roll-out phase. Similar to model (5), the cluster-specific random effects, c_{ij}^*, are intended to capture the effects of premature exposure to intervention components. This is an extension of model 3, which assumes the random cluster-by-discrete time effects shared a common (unstructured) covariance matrix. While the random effect covariance structure in model 6 has less restrictions than in model 5, it requires the estimation of $\frac{(n+1) \times n + (n_\tau - 1) \times (n_\tau - 2)}{2}$ covariance parameters and is subject to similar limitations as discussed in Section 2.2.2. Here we define n_τ as the number of time periods prior to the follow-up phase of the study.

2.3.3. Random cluster by-intervention-by-linear time effect Similar to model 4, we can impose a random slope for time in each group.

$$g(\mu_{ij}) = \alpha + \theta \times X_{ij} + \beta_j + \gamma \times t_j \times 1_{\{X_{ij}=0\}} + b_{0i} + b_{1i} \times t_j + (c_{0i} + c_{1i} \times t_j) \times 1_{\{X_{ij}=0\}}, \tag{7}$$

where $(c_{0i}, c_{1i}) \sim N(0, \boldsymbol{\Sigma}_{\boldsymbol{c_0}, \boldsymbol{c_1}})$ and $\boldsymbol{\Sigma}_{\boldsymbol{c_0}, \boldsymbol{c_1}}$ is the 2×2 covariance matrix for (c_{0i}, c_{1i}), $i = 1, ..., N$. The cluster-specific random effects c_{1i} allow for unique linear time trends in each control cluster and are intended to capture the effects of premature adoption of intervention components.

2.3.4. Fixed effects for linear time in all clusters To further limit loss of power and potential issues of identifiability due to estimation of too many parameters, we replace the fixed discrete time effect in models 5 through 7, β_j, with a fixed linear time effect, $\beta \times t_j$, in models 8 through 10. This strategy may be useful in situations when the combined effects of confounding factors result in monotonic changes in the outcome over time. For example, the implementation of a new policy intended to reduce opioid prescriptions in conjunction with media awareness campaigns may decrease the level of opioid overdose deaths with time.

$$g(\mu_{ij}) = \alpha + \theta \times X_{ij} + \beta \times t_j + \gamma \times t_j \times 1_{\{X_{ij}=0\}} + b_{ij} + c_{ij} \times 1_{\{X_{ij}=0\}} \tag{8}$$

$$g(\mu_{ij}) = \alpha + \theta \times X_{ij} + \beta \times t_j + \gamma \times t_j \times 1_{\{X_{ij}=0\}} + b_{ij}^* + c_{ij}^* \times 1_{\{X_{ij}=0\}} \tag{9}$$

$$g(\mu_{ij}) = \alpha + \theta \times X_{ij} + \beta \times t_j + \gamma \times t_j \times 1_{\{X_{ij}=0\}} + b_{0i} + b_{1i} \times t_j + (c_{0i} + c_{1i} \times t_j) \times 1_{\{X_{ij}=0\}} \tag{10}$$

2.4. Estimation

These models can be fit by specifying the Poisson family in the *glmer* function (package: *lme4*) in **R**.[25] Of note, when the outcome is a summary measure for each cluster (as is the case in our motivating example), we ran into estimation

medRxiv

difficulties when specifying random cluster-by-discrete time effects when the outcome distribution was assumed to be Gaussian. This is due to incorporation of a residual term for each cluster for each time period, which yields unidentifiable random effects. This issue does not arise under Poisson distributional assumptions, since the variance is directly proportional to the mean and thus residual terms are not estimated. To accommodate the Gaussian distribution assumption in this scenario, multiple communities per cluster[13,17,20] or multiple intervals per time period would be needed.

It is important to distinguish between the impact of calendar time on the intervention assignment effect and on the intervention effect itself. Because of randomization, the intervention assignment effect is not confounded by calendar time; a randomized comparison will yield an unbiased estimate of the effect of assignment to the intervention initiation time. This effect includes all consequences of starting at that time, including the initiation of treatment. Nonetheless, our target of estimation, the effect of intervention alone, is confounded by calendar time.

In the models above, the fixed time effects β_j and β and the random time effects b_{ij}, b_{ij}^*, and b_{1i} are modeled to capture confounding factors that influence all clusters. The fixed time effects γ and random time effects c_{ij}, c_{ij}^*, and c_{1i} are modeled to capture confounding factors that influence control communities only. These effects are all identifiable because the design matrices are full rank. Without additional assumptions, however, the above models cannot distinguish among the effects of external factors, early intervention adoption, or of other time-varying factors. Rather, they model temporal trends that may be a consequence of these factors; such models allow for assessment of the impact of such trends on bias, power, and Type 1 error. When used in analysis, our simulation studies demonstrate that these models can reduce bias in estimation of intervention effects, in the settings appropriate for their use.

3. Simulation study

We conduct a simulation study to investigate the impact of confounding factors on the bias, coverage probability of 95% confidence intervals, power, and Type 1 error of the intervention effect estimate. The data are simulated based on the motivating example described in Section 1.1. In this setting, we have $N = 18$ clusters and $n = 13$ time periods. The intervention is rolled out to each cluster according to the time line provided in Figure 1. In all data generating models, the outcomes Y_{ij} are simulated from a Poisson distribution, and represent the number of opioid overdose deaths in cluster i during time period j. Here g is the log link, and the population size of cluster i, O_i, is included as an offset in all models. That is, $g(\mu_{ij}) = log(E[Y_{ij}]/O_i)$. The intercept α is set to -10 and the standard deviation of the random intercept, σ_{b_0}, is set to 0.30. These numbers were determined using the opioid death counts in the 18 South Carolina clusters between 2016 and 2018. The data are provided in Supplementary Table 1. The intervention effect θ is set to log(0.6), which represents the target 40% reduction in opioid overdose deaths due to intervention (Section 1.1).

3.1. Simulation scenarios

We apply models 1 through 10 under four general scenarios: (1) standard, (2) presence of external factors, (3) early intervention adoption, and (4) presence of external factors and early adoption. The data generation process for each scenario is summarized in Table 1, and is described in more detail below.

In the first scenario (standard) we assume no confounding is present; all clusters receive the intervention during the scheduled time period and no external factors influence the outcome. The data are generated according to scenario 1 in Table 1. In the second scenario, all clusters not currently exposed to external factors by time period j have a 1 in N chance of exposure. Once exposed, each cluster continues to be exposed through the remainder of the study period. This is intended to reflect the situation where external factors have long-lasting effects, such as new public policies limiting opioid prescriptions and media awareness campaigns. The number of clusters exposed to external factors at each time period j, n'_j, is simulated from a $Binomial(N'_j, 1/N)$ distribution, where N'_j is the total number of clusters unexposed to the external factors prior to time period j.

Prepared using simauth.cls

Rennert, Heo, Litwin, and De Gruttola

medRxiv

Table 1. Simulation scenarios.

Scenario	Data generating model and scenario description	Effect of confounding factors	Index
Standard	$\log(\mu_{ij}) = \alpha + \theta \times X_{ij} + b_i$ **Description**: No confounding present	None	1
External factors	$\log(\mu_{ij}) = \alpha + \theta \times X_{ij} + \beta_{ij} + b_i$ **Description**: At each time period j, n'_j clusters randomly exposed to external factors for remainder of study period; $n'_j \sim Binomial(N'_j, 1/N)$, where N'_j is total number of clusters unexposed to external factors prior to time period j.	$\beta_{ij} \sim unif[-1,0]$ if cluster i exposed to external event during time period j, and 0 otherwise	2.1
		$\beta_{ij} \sim unif[0,1]$ if cluster i exposed to external event during time period j, and 0 otherwise	2.2
Early adoption	$\log(\mu_{ij}) = \alpha + \theta_{ij} \times 1_{\{X_{ij}=0\}} + \theta \times X_{ij} + b_i$ **Description**: At each time period j, n^*_j control clusters prematurely adopt the intervention; $n^*_j \sim Binomial(N^*_j, N^*_j/N)$, where N^*_j is the number of control clusters not receiving the intervention prior to time period j.	$\theta_{ij} \sim unif[\theta,0]$ if control cluster i prematurely adopts intervention at time period j, and equals 0 otherwise	3
External factors & Early adoption	$\log(\mu_{ij}) = $ $\alpha + \beta_{ij} + \theta_{ij} \times 1_{\{X_{ij}=0\}} + \theta \times X_{ij} + b_i$ **Description**: At each time period j, n'_j cluster randomly exposed to external event(s) and n^*_j control clusters prematurely adopt the intervention, where n'_j and n^*_j are defined in the descriptions for scenarios 2 and 3, respectively.	$\beta_{ij} \sim unif[-1,0]$ if cluster i exposed to external event during time period j, and 0 otherwise, and θ_{ij} is defined as above	4.1
		$\beta_{ij} \sim unif[0,1]$ if cluster i exposed to external event during time period j, and 0 otherwise, and θ_{ij} defined as above	4.2

Data is simulated under 4 general scenarios. The data generating model for each simulation scenario is displayed in the second column. Here μ_{ij} is the expected rate of opioid overdose deaths in cluster i during time period j, θ is the intervention effect and is set to log(0.6), and X_{ij} is an indicator of whether cluster i is scheduled to receive intervention during time period j and is based on the SWD represented by Figure 1. The fixed intercept α is set to -10 and the random intercept b_i is simulated from a $N(0, 0.30)$ distribution. A description of the selection process for exposure to confounding factors (e.g., external factors and/or early adoption) is provided in the second column below the data generating model. The impact of confounding factors on the outcome is detailed in the third column. In scenarios 2 and 4, we allow the combined effects of external factors to have either a positive impact on the outcome (scenarios 2.1 and 4.1) or a negative impact on the outcome (scenarios 2.2 and 4.2).

Once exposed, the effect of external factors on the outcome in cluster i is set to vary uniformly at each time period j. Specifically, we let t'_i denote the time period in which cluster i is exposed to external factors. We simulate β_{ij} under 2 settings. For $j \geq t'_i$, we simulate $\beta_{ij} \sim unif[-1, 0]$, corresponding to a positive effect of external factors on the outcome opioid overdose deaths (scenario 2.1). This is intended to represent a rising tide of events aimed at improving outcomes. We set $\beta_{ij} = 0$ for $j < t'_i$, indicating that cluster i has yet to be exposed to external factors. To represent a negative impact of external factors on the outcome, we simulate $\beta_{ij} \sim unif[0, 1]$ for $j \geq t'_i$ and $\beta_{ij} = 0$ for $j < t'_i$ (scenario 2.2).

In the third scenario (early adoption), the number of clusters prematurely adopting the intervention during each time period j, n^*_j, is simulated from a $Binomial(N^*_j, \frac{N^*_j}{N})$ distribution. Here N^*_j is the number of control clusters which have not adopted any intervention components by time period j. This set up reflects the nature of SWD, where the number of control clusters at risk for early adoption decreases with time as more clusters crossover to the intervention group by design. However, the probably of exposure to early adoption for a control cluster likely increases with time as interventions become more widely available. Once exposed to early adoption, we assume that exposure continues until the scheduled roll-out of the intervention, at which point the cluster receives the full intervention effect.

For all time periods j in which a cluster is subjected to early adoption, the magnitude of the early intervention adoption effect is set to vary uniformly. Denote t^*_i as the time period of early adoption for cluster i and denote t^I_i as the time period in which the intervention was scheduled to be rolled-out to cluster i. We simulate $\theta_{ij} \sim unif[\theta, 0]$ for $j = t^*_i, ..., t^I_i - 1$, and $\theta_{ij} = 0$ otherwise. Thus the maximum effect of early intervention adoption on the outcome for exposed control clusters is set to θ. In this scenario, the confounding factor (early adoption) has a positive effect on the outcome in the control group only.

medRxiv

In the fourth scenario, the effects of external factors and early adoption are simulated in the same manner as in scenario 2 and scenario 3, respectively. In this scenario, the data are generated according to scenarios 4.1, where external factors have a positive effect on the outcome, and scenario 4.2, where external factors have a negative effect on the outcome. Additional detail is provided in Table 1. We note that in the latter scenario (4.2), the negative impact of external factors on the outcome is partially offset by the positive impact of early adoption on the outcome.

3.2. Results

Simulation results in are presented in Table 2. Under scenario 1 (standard), all models yield unbiased estimates of the intervention effect, reach nominal confidence interval coverage rates, and preserve Type 1 error. Model 1, which does not include any random effects for time, yields unbiased estimates of the intervention effect in scenario 2 (external factors), where early adoption is not present and therefore confounding does not differentially impact control and intervention clusters. However, coverage probabilities are below the nominal level of 0.95 and Type 1 error is inflated. Model 1 is heavily biased in scenarios 3 (early adoption) and 4 (external factors & early adoption), where confounding differentially impacts intervention and control clusters due to the early adoption.

Table 2. Simulation Results.

Model index	Time effects Main fixed	random	Interaction fixed	random	Scenario 1 Standard %bias	SD	SE	cov	pwr	α	Scenario 2.1 External factors (+) %bias	SD	SE	cov	pwr	α	Scenario 2.2 External factors (-) %bias	SD	SE	cov	pwr	α
1	disc				0.4	0.08	0.08	0.94	1	0.05	-0.5	0.11	0.08	0.83	1	0.20	0.7	0.10	0.09	0.92	1	0.10
2	disc	H/G			0.3	0.08	0.08	0.94	1	0.04	-0.4	0.10	0.10	0.93	1	0.07	0.7	0.10	0.09	0.93	1	0.07
3	disc	UN			0.9	0.10	0.11	0.96	1	0.03	0.6	0.12	0.12	0.95	0.99	0.07	2.3	0.12	0.12	0.95	0.99	0.04
4	disc	lin			0.5	0.08	0.08	0.93	1	0.06	0.1	0.11	0.08	0.82	1	0.21	0.5	0.10	0.09	0.90	1	0.12
5	disc	H/G	lin	H/G	0.6	0.25	0.24	0.93	0.6	0.05	8.0	0.28	0.26	0.93	0.53	0.05	4.2	0.28	0.26	0.93	0.53	0.07
6	disc	UN	lin	UN	1.5	0.29	0.31	0.96	0.38	0.03	9.3	0.32	0.33	0.95	0.39	0.05	6.8	0.32	0.35	0.96	0.32	0.04
7	disc	lin	lin	lin	0.9	0.25	0.24	0.93	0.59	0.05	9.2	0.29	0.23	0.87	0.62	0.13	5.0	0.28	0.26	0.93	0.54	0.09
8	lin	H/G	lin	H/G	-0.4	0.14	0.14	0.95	0.96	0.05	-4.7	0.16	0.16	0.94	0.85	0.08	4.5	0.17	0.16	0.92	0.92	0.08
9	lin	UN	lin	UN	-1.5	0.16	0.17	0.96	0.85	0.02	-2.7	0.18	0.19	0.96	0.79	0.04	5.3	0.19	0.2	0.95	0.8	0.04
10	lin	lin	lin	lin	-0.8	0.14	0.14	0.94	0.95	0.05	-9.8	0.17	0.13	0.81	0.88	0.18	2.2	0.18	0.15	0.92	0.92	0.12

Model index	Time effects Main fixed	random	Interaction fixed	random	Scenario 3 Early adoption %bias	SD	SE	cov	pwr	α	Scenario 4.1 External factors (+) & Early adoption %bias	SD	SE	cov	pwr	α	Scenario 4.2 External factors (-) & Early adoption %bias	SD	SE	cov	pwr	α
1	disc				-33.6	0.09	0.09	0.49	0.98	0.04	-33.1	0.12	0.08	0.44	0.94	0.20	-32.8	0.11	0.09	0.53	0.95	0.10
2	disc	H/G			-33.6	0.09	0.09	0.51	0.97	0.04	-33.4	0.11	0.10	0.56	0.89	0.07	-32.8	0.11	0.10	0.58	0.92	0.07
3	disc	UN			-32.7	0.10	0.11	0.7	0.88	0.03	-32.3	0.13	0.12	0.72	0.79	0.06	-31.8	0.13	0.13	0.76	0.78	0.05
4	disc	lin			-33.4	0.09	0.08	0.46	0.98	0.04	-33.0	0.12	0.08	0.43	0.94	0.20	-32.8	0.11	0.09	0.49	0.95	0.09
5	disc	H/G	lin	H/G	-8.0	0.24	0.24	0.95	0.49	0.04	2.1	0.30	0.27	0.92	0.49	0.08	-4.0	0.31	0.27	0.91	0.47	0.10
6	disc	UN	lin	UN	-3.9	0.28	0.32	0.98	0.29	0.01	4.8	0.34	0.34	0.94	0.33	0.04	2.9	0.37	0.37	0.95	0.29	0.04
7	disc	lin	lin	lin	-6.8	0.24	0.24	0.95	0.47	0.04	4.2	0.31	0.24	0.85	0.59	0.14	-2.4	0.31	0.26	0.9	0.48	0.11
8	lin	H/G	lin	H/G	-3.0	0.14	0.14	0.95	0.93	0.03	-6.6	0.18	0.16	0.90	0.83	0.07	2.9	0.17	0.16	0.95	0.90	0.07
9	lin	UN	lin	UN	-1.9	0.16	0.18	0.97	0.83	0.02	-1.8	0.20	0.19	0.93	0.76	0.05	2.5	0.20	0.20	0.96	0.73	0.05
10	lin	lin	lin	lin	-2.9	0.14	0.14	0.94	0.94	0.05	-10.8	0.18	0.13	0.81	0.87	0.21	1.0	0.17	0.15	0.93	0.92	0.11

% bias ($\frac{\hat{\theta}-\theta}{\theta}$), standard deviation (SD), estimated standard error (SE), coverage rate of 95% confidence interval (cov), power (pwr), and Type 1 error (α) of estimated intervention effect. Rows correspond to fitting models 1 through 10 under the scenarios described in Table 1; +/- imply positive/negative effect of external factors on outcome. Models include discrete (disc) or linear (lin) fixed time effects for all clusters (main effect) and/or control clusters (interaction effect). H/G , UN, and lin indicate Hooper/Girling, unstructured, and linear random effect structure for time effects.

Models 2 through 4, which include random cluster-by-time effects, yield unbiased estimates of the intervention effect when confounding does not differentially impact control and intervention clusters (i.e., scenarios 1 and 2). Model 3, which assumes an unstructured covariance for the random cluster-by-time interactions, performs the best with regard to coverage probabilities of the 95% confidence intervals and Type 1 error preservation. Model 2, which assume a single variance for the random cluster-by-time interactions (Hooper/Girling model), perform slightly worse on these metrics. Model 4, which assume a random slope for time for each cluster, has the highest inflation in Type 1 error with coverage probabilities well below the nominal level of 0.95. Models 2 through 4 are overpowered for scenarios 1 and 2. These models perform poorly

when early adoption is present in scenarios 3 and 4. The estimated intervention effect is reduced by 31.8% to 33.6% in these scenarios.

The performance of models with and without intervention-by-time interactions is compared in Figure 2. Models without intervention-by-time interactions are displayed in the left column, where the blue shapes correspond to models 2 through 4. For comparison, we include models which replace the discrete main effect for (fixed) time in models 2 through 4, β_j, with the linear fixed effect $\beta \times t_j$. These models are labeled by the red shapes in the left column of Figure 2. The models in the right column of Figure 2 include fixed and random intervention-by-time interactions. These models correspond to models 5 through 10.
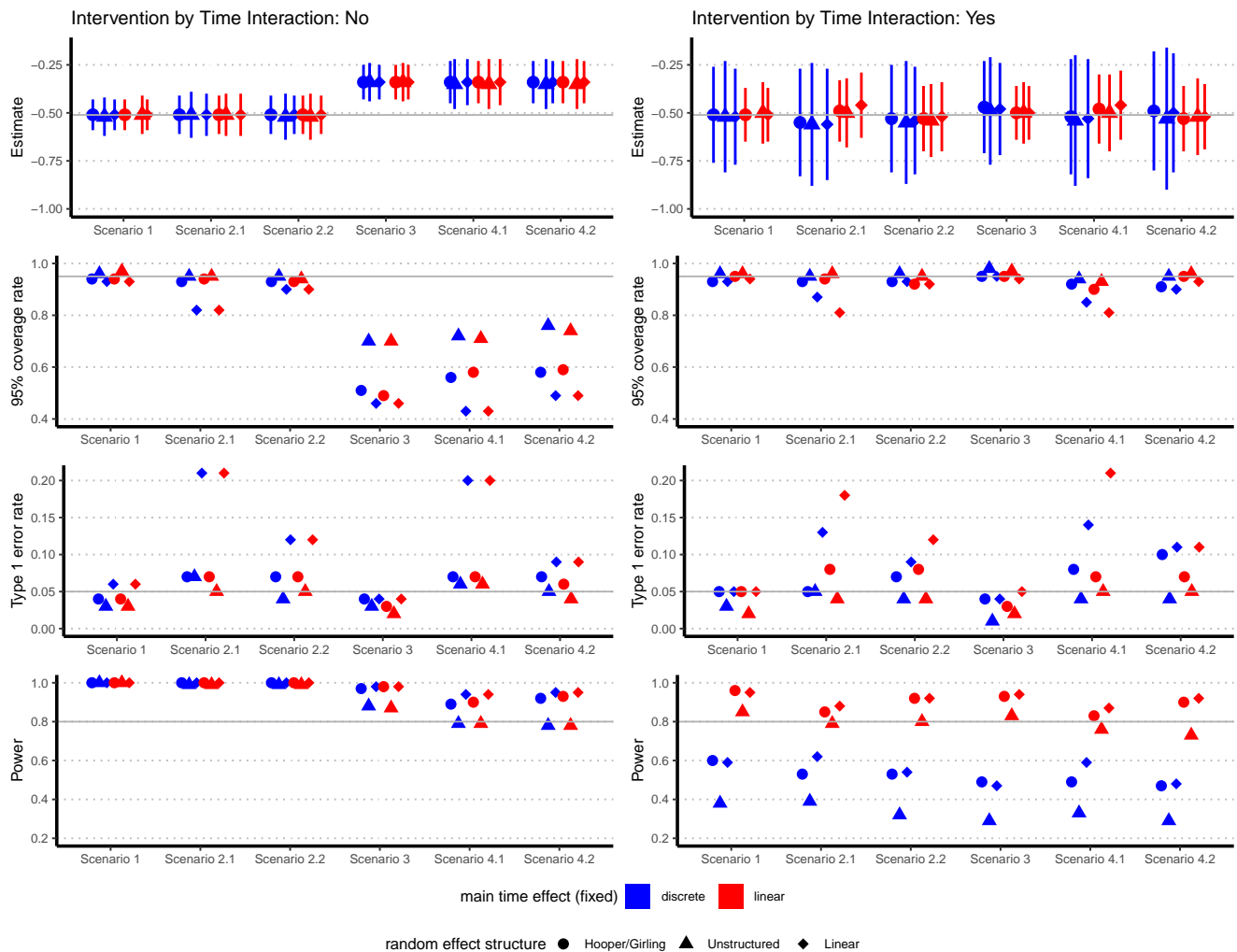
Figure 2. Performance of models with and without fixed and random intervention-by-time interactions. Models are compared across scenarios listed in Table 1. First row compares intervention effect estimates ± empirical standard error. Horizontal gray line: true intervention effect $\theta = \log(0.6)$. Second, third, and fourth rows compare models on coverage rate of 95% confidence intervals, Type 1 error rate, and power, respectively. Horizontal gray lines indicate 95% coverage rate, 0.05 Type 1 error rate, and a power of 0.80 in the second, third, and fourth rows, respectively. Covariance structure for cluster-by-time random effects: Hooper/Girling labeled by circles, unstructured labeled by triangles, and linear labeled by diamonds. Models which incorporate a discrete term for the main effect for time (fixed) are labeled by blue shapes; models which incorporate a linear term are labeled by red shapes. Models without intervention-by-time interactions displayed in left column, where blue shapes correspond to models 2 through 4 in Section 2. Models with intervention-by-time interactions displayed in right column, and correspond to models 5 through 10 in Section 2. Models 5 through 10 include a linear time effect for the fixed intervention-by-time interaction term.

Models 5 through 10 perform well in all simulation scenarios. When no confounding or early adoption is present (scenario 1), these models yield little bias, reach nominal coverage probabilities for 95% confidence intervals, and preserve Type 1 error. When only confounding due to external factors is present (scenario 2), all models have less than 10%

medRxiv

bias. These models generally reach nominal confidence interval coverage probabilities and preserve Type 1 error. The exceptions are models 7 and 10, which include random cluster-by-group-by-linear time interaction effects.

When confounding differentially impacts intervention and control clusters due to early adoption (scenarios 3 and 4), models 5 through 10 greatly reduce the bias in the intervention effect estimate compared to models which do not include intervention-by-time interactions. Models 6 and 9, which assume an unstructured covariance for the random cluster-by-group-by-discrete time interactions, generally yield the lowest bias, while reaching nominal coverage probabilities of the 95% confidence intervals and preserving Type 1 error. Models 5 and 8 (Hooper/Girling models) perform slightly worse on these metrics. Models 7 and 10 have the highest inflation in Type 1 error and coverage probabilities well below the nominal level.

The choice of a discrete or linear term for the fixed time effect does not impact power in models without intervention-by-time interactions. This result is expected given the findings of Grantham et al.[26] For models which include intervention-by-time interaction terms (i.e., models 5 through 10), the strongest correlate of power is the fixed effect for time. Models with discrete time effects, which incorporate a parameter for each time period, have much lower power compared to models which incorporate a single parameter corresponding to a linear time effect. Models with a random slope for time achieve the highest power (models 7 and 10), while models which assume an unstructured covariance for random cluster-by-time effects achieve the lowest power (models 6 and 9).

4. Discussion

SWDs and alternative cluster randomized trials are currently being used to implement interventions to reduce opioid overdose deaths in communities across several states. However, these interventions are competing with newly initiated public policies and media awareness campaigns. Furthermore, control communities may adopt components of proposed interventions as they become readily available. These scenarios induce confounding by time. In this paper, we have demonstrated the consequences of inadequately accounting for time effects in these scenarios. Given the difficulties in capturing the timing and exposure levels of such confounding factors, we considered mixed-effects models which incorporate random effects to capture these latent confounding factors. We discussed the limitations of commonly used models in the context of proposed SWDs to combat the opioid epidemic, and proposed solutions to accommodate deviations from these assumptions.

Through our simulation study, we showed that recently proposed mixed effects models are sensitive to the type of factors which induce confounding by time. The Hooper/Girling model[17, 20] offers an improvement over the standard Hussey and Hughes model,[16] but may result in severely biased estimates of the intervention effect when secular trends systematically differ between intervention and control clusters. Even in scenarios where these models do capture the secular trend, we demonstrated how incorrect specification of the cluster-level covariance over time can yield under coverage of confidence intervals and inflation of Type 1 error. Similar conclusions have been reached in other studies.[3, 22, 22, 27, 28]

Alternatively, models which allow secular trends to systematically differ between the intervention and control clusters through incorporation of fixed and random group-by-time effects offer a major improvement in terms of bias reduction. Our simulation studies confirmed that models incorporating unstructured cluster-level covariances for the random intervention-by-cluster-by-time interaction terms yielded nominal confidence interval coverage rates and preserved Type 1 error (i.e., models 6 and 9). However, these models may be underpowered for certain parameterizations of the fixed time effects.

The proposed mixed-effects modeling framework discussed in this paper treats external factors and early intervention adoption as latent processes; these confounding factors are accounted for through the incorporation of random effects. This modeling strategy is useful in practice because investigators are often unaware of all confounding factors affecting the outcome, let alone each cluster's level of exposure to these factors. Furthermore, as evidenced by the results of our

medRxiv

simulation study, these models have a certain level of robustness to the assumed distribution of the random effects.

4.1. Limitations and Future Research

An important limitation of the models we discussed in this paper is that they assume the intervention effect does not vary with time. Models attempting to account for both secular trends which differ by intervention group, and time-varying treatment effects, may lead to unidentifiable intervention effect estimates. Furthermore, the approach discussed here is limited to GLMM. Generalized estimating equations (GEE) also allow for correlation in the outcomes, and are robust to misspecification of the covariance structure.[29] Ren et al. show that GEE is more robust to model misspecification than linear mixed models when the random intercepts differ by intervention group. Future work is needed to explore the performance of GEE models in the context considered here, where secular trends in the control and intervention clusters arise from different mechanisms, and differentially impact clusters within each group. Several nonparametric methods have also been proposed that use within-period[3,27,30] or between period[3] comparisons to account for confounding by time. While these models are robust to misspecification of the random effects, the performance of these models when confounding factors differentially impact intervention and control clusters, such as in the case of early intervention adoption, has not been explored.

Although our paper focused on stepped-wedge designs, the secular trends in outcomes induced by rising tides and early intervention adoption may also be present in other types of cluster randomized trials. Simulation studies are needed to determine the impact of external factors and early adoption on the bias of the intervention effect estimate, Type 1 error, and power in these settings. Grantham et al. establish a sufficient condition for when the choice of time parameterization does not impact the variance of the estimated intervention effect in cluster randomized trials.[26] This information can be useful in the planning of such trials. While Grantham et al establish that categorical or linear fixed time effects do not impact the variance estimator of the intervention effect in SWD, this was not the case when group-by-time interaction terms were modeled (as demonstrated by our simulation study). Future investigation is needed to establish sufficient conditions in the presence of interactions.

4.2. Conclusion and Recommendations

Stepped-wedge designs are particularly suitable for epidemics and pandemics, where complex health interventions are needed as soon as possible. Given the urgency of such situations, other policies and interventions aimed at improving outcomes will likely be implemented in concurrence with any proposed interventions. This induces confounding by time, which we have shown can severely bias the intervention effect estimate in SWDs when not properly accounted for. Studies must therefore consider these scenarios during the planning stages. We have shown that models which incorporate fixed and random group-by-time effects are effective in reducing bias when confounding factors differentially impact intervention and control clusters. However, incorporation of such time effects may impact power and must be accounted for in sample size calculations. Models incorporating an unstructured covariance for the random intervention-by-cluster-by-time interaction effects are most effective in reducing bias, reaching nominal confidence interval coverage, and preserving Type 1 error; but they can lead to severely under powered studies when used in conjunction with discrete fixed time effects. One strategy is to use fixed parametric time effects and an unstructured covariance; these models perform reasonably well in the scenarios considered here. Alternatively, one can impose a more restrictive covariance structure such as exponential decay over time, Hooper/Girling covariance, or random slopes for time. To account for potential under coverage of confidence intervals and Type 1 error inflation, randomization based inference should be used.[31,32] Simulation studies may be needed to estimate sample size and power for certain models, as formulas to calculate them are not currently available for many random effect covariance structures.

5. Declarations

Availability of data and materials: The dataset supporting the conclusions of this article is included within the web appendix of this article. We are making the computer code available. Once accepted, our code will be deposited to github, and will be linked to the final published article.

Authors' contributions: LR, MH, and VG conceived the methods and LR drafted the manuscript. VDG and MH helped write the paper. VDG, MH, and AHL critically reviewed the contents of the paper. AHL designed the study used as a motivating example in this paper. LR carried out the simulation study.

Competing interests: The authors declare that they have no competing interests.

Consent for publication: Not applicable.

Ethics approval and consent to participate: No ethical approval or consent was required for this simulation study based on a proposed trial design.

References

1. Brown CA, Lilford RJ. The stepped wedge trial design: a systematic review. *BMC Medical Research Methodology.* 2006;6(1):54.
2. Hemming K, Haines TP, Chilton PJ, Girling AJ, Lilford RJ. The stepped wedge cluster randomised trial: rationale, design, analysis, and reporting. *BMJ.* 2015;350.
3. Kennedy-Shaffer L, De Gruttola V, Lipsitch M. Novel methods for the analysis of stepped wedge cluster randomized trials. *Statistics in Medicine.* 2020;39(7):815–844.
4. Mdege ND, Man MS, Taylor Nee Brown CA, Torgerson DJ. Systematic review of stepped wedge cluster randomized trials shows that design is particularly used to evaluate interventions during routine implementation. *Journal of Clinical Epidemiology.* 2011;64(9):936–948.
5. Keriel-Gascou M, Buchet-Poyau K, Rabilloud M, Duclos A, Colin C. A stepped wedge cluster randomized trial is preferable for assessing complex health interventions. *Journal of Clinical Epidemiology.* 2014;67(7):831–833.
6. Becker WC, Fiellin DA. When Epidemics Collide: Coronavirus Disease 2019 (COVID-19) and the Opioid Crisis. *Annals of Internal Medicine.* 2020;. Publisher: American College of Physicians.
7. *National Institute of Health. RFA-DA-19-016: HEALing Communities Study: Developing and Testing an Integrated Approach to Address the Opioid Crisis (Research Sites) (UM1 - Clinical Trial Required).* 2018.
8. *NIH RePORTER. OPTIMIZING HEALING IN OHIO COMMUNITIES (OHIO).* 2019.
9. *NIH RePORTER. KENTUCKY CAN HEAL (COMMUNITIES AND NETWORKS HELPING END ADDICTION LONG-TERM).* 2019.
10. Bion J, Richardson A, Hibbert P, et al. 'Matching Michigan': a 2-year stepped interventional programme to minimise central venous catheter-blood stream infections in intensive care units in England. *BMJ Quality & Safety.* 2013;22(2):110–123.
11. Davis CS, Lieberman AJ, Hernandez-Delgado H, Suba C. Laws limiting the prescribing or dispensing of opioids for acute pain in the United States: A national systematic legal review. *Drug and Alcohol Dependence.* 2019;194:166–172.
12. *Center for Disease Control and Prevention.CDC launches campaign to help states fight prescription opioid epidemic.* 2017.
13. Hemming K, Taljaard M, Forbes A. Analysis of cluster randomised stepped wedge trials with repeated cross-sectional samples. *Trials.* 2017;18(1):101.
14. Thompson JA, Fielding KL, Davey C, Aiken AM, Hargreaves JR, Hayes RJ. Bias and inference from misspecified mixed-effect models in stepped wedge trial analysis. *Statistics in Medicine.* 2017;36(23):3670–3682.
15. Taljaard M, Teerenstra S, Ivers NM, Fergusson DA. Substantial risks associated with few clusters in cluster randomized and stepped wedge designs. *Clinical Trials: Journal of the Society for Clinical Trials.* 2016;13(4):459–463.
16. Hussey MA, Hughes JP. Design and analysis of stepped wedge cluster randomized trials. *Contemporary Clinical Trials.* 2007;28(2):182–191.
17. Hooper R, Teerenstra S, Hoop Ed, Eldridge S. Sample size calculation for stepped wedge and other longitudinal cluster randomised trials. *Statistics in Medicine.* 2016;35(26):4718–4728.

medRxiv

18. Barker D, McElduff P, D'Este C, Campbell MJ. Stepped wedge cluster randomised trials: a review of the statistical methodology used and available. *BMC Medical Research Methodology.* 2016;16(1):69.

19. Nickless A, Voysey M, Geddes J, Yu LM, Fanshawe TR. Mixed effects approach to the analysis of the stepped wedge cluster randomised trial—Investigating the confounding effect of time through simulation. *PLoS ONE.* 2018;13(12).

20. Girling AJ, Hemming K. Statistical efficiency and optimal design for stepped cluster studies under linear mixed effects models. *Statistics in Medicine.* 2016;35(13):2149–2166.

21. Kasza J, Hemming K, Hooper R, Matthews J, Forbes A. Impact of non-uniform correlation structure on sample size and power in multiple-period cluster randomised trials. *Statistical Methods in Medical Research.* 2019;28(3):703–716.

22. Kasza J, Forbes AB. Inference for the treatment effect in multiple-period cluster randomised trials when random effect correlation structure is misspecified. *Statistical Methods in Medical Research.* 2019;28(10-11):3112–3122.

23. McMaster H, Governor SCO. *Executive order no. 2017-43.* 2017.

24. South Carolina DAODAS Library Catalog: www.daodas.sc.gov; .

25. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models using lme4. *Journal of Statistical Software.* 2015;67(1):1–48. arXiv: 1406.5823.

26. Grantham KL, Forbes AB, Heritier S, Kasza J. Time Parameterizations in Cluster Randomized Trial Planning. *The American Statistician.* 2020;74(2):184–189. Publisher: Taylor & Francis.

27. Thompson JA, Davey C, Fielding K, Hargreaves JR, Hayes RJ. Robust analysis of stepped wedge trials using cluster-level summaries within periods. *Statistics in Medicine.* 2018;37(16):2487–2500.

28. Ren Y, Hughes JP, Heagerty PJ. A Simulation Study of Statistical Approaches to Data Analysis in the Stepped Wedge Design. *Statistics in Biosciences.* 2019;.

29. Fitzmaurice G, Laird NM, Ware JH. *Applied longitudinal analysis.* John Wiley & Sons; 2012.

30. Hughes JP, Heagerty PJ, Xia F, Ren Y. Robust Inference for the Stepped Wedge Design. *Biometrics.* 2019;:biom.13106.

31. Wang R, De Gruttola V. The use of permutation tests for the analysis of parallel and stepped-wedge cluster-randomized trials. *Statistics in Medicine.* 2017;36(18):2831–2843.

32. Ji X, Fink G, Jacob Robyn P, S. Small D. Randomization inference for stepped-wedge cluster-randomized trials: An application to community-based health insurance. *The Annals of Applied Statistics.* 2017;11:1–20.