Dynamic wait-listed designs for randomized trials: new designs for prevention of youth suicide
C Hendricks Brown, Peter A Wyman, Jing Guo and Juan Peña
Clin Trials 2006; 3: 259
DOI: 10.1191/1740774506cn152oa

The online version of this article can be found at:
http://ctj.sagepub.com/cgi/content/abstract/3/3/259
Dynamic wait-listed designs for randomized trials: new designs for prevention of youth suicide

C Hendricks Browna, Peter A Wymanb, Jing Guoa and Juan Peñab

Background The traditional wait-listed design, where half are randomly assigned to receive the intervention early and half are randomly assigned to receive it later, is often acceptable to communities who would not be comfortable with a no-treatment group. As such this traditional wait-listed design provides an excellent opportunity to evaluate short-term impact of an intervention. We introduce a new class of wait-listed designs for conducting randomized experiments where all subjects receive the intervention, and the timing of the intervention is randomly assigned. We use the term "dynamic wait-listed designs" to describe this new class.

Purpose This paper examines a new class of statistical designs where random assignment to intervention condition occurs at multiple times in a trial. As an extension of a traditional wait-listed design, this dynamic design allows all subjects to receive the intervention at a random time. Motivated by our search for increased statistical power in an ongoing school-based trial that is testing a program of gatekeeper training to identify suicidal youth and refer them to treatment, this new design class is especially useful when the primary outcome is a count or rate of occurrence, such as suicidal behavior, whose rate can fluctuate over time due to uncontrolled factors.

Methods Statistical power is computed for various dynamic wait-listed designs under conditions where the underlying rate of occurrence is allowed to vary non-systematically. We also present as an example a large ongoing trial to evaluate a gatekeeper training suicide prevention program in 32 schools which we initially began as a classic randomized wait-listed design. The primary outcome of interest in this study is the count of the number of children who are identified by the school system as having suicidal thoughts or behaviors who are then validated as being suicidal by mental health professionals in the community.

Results A general result shows that dynamic wait-listed designs always have higher statistical power over a traditional wait-listed design. This power increase can be substantial. Efficiency gains of 33% are easy to obtain for situations where the intervention has a small effect and the variation in rate across time is quite high. When the rate variation for an outcome is very low or the intervention effect is large, efficiency gains approach 100%. A small increase in the number of times where random assignment occurs from 2 – for the standard wait-listed design, to say 4 – can provide a large reduction in variance. Efficiency gains can also be high when converting standard wait-listed design to a dynamic one half-way into the study.

Limitations As with all wait-listed designs, dynamic wait-listed designs can only be used to evaluate short-term impact. Since all subjects eventually receive the intervention, no comparison can be made after the end of the random assignment period. The statistical power benefits are primarily limited to outcomes that can be treated as count or rate to event data.

a Department of Epidemiology and Biostatistics, University of South Florida, Tampa, Florida, USA. b Department of Psychiatry, University of Rochester, Rochester, New York, USA

Author for correspondence: C Hendricks Brown, Department of Epidemiology and Biostatistics, University of South Florida, Tampa, Florida, USA

© Society for Clinical Trials 2006 10.1191/1740774506cn152oa
Introduction

In this paper we present a generalization of the wait-listed design and examine its applicability to the evaluation of an ongoing suicide prevention program. The classic wait-listed randomized trial often provides a convenient and scientifically rigorous way to evaluate an intervention's short-term impact. In such a design half of the units – whether they be individuals, families, classrooms, schools or communities – are randomly chosen to receive the intervention in the first phase of the study; the other half receives the intervention in the second phase. During the first phase of the study, a legitimate comparison can be made on an outcome variable between intervention and control conditions, and this difference can be attributed to the true effect of the intervention. Data from the second phase cannot be used to assess intervention impact because there is no control group to compare over that period of time.

A wait-listed design is especially useful when a community, school system or governmental agency has already decided that everyone in a specific population will eventually receive a new intervention or program, especially when that intervention is widely viewed as being beneficial, even though there may be little empirical evidence to back up this perception. Community leaders as well as individuals in the community often feel that the use of a random process to determine who receives the intervention first is fair and ethical, as long as everyone receives the intervention within a reasonable amount of time. There may also be a compensatory advantage for receiving the intervention in the later group, since knowledge about the program's implementation in the first phase can often lead to improved implementation in the second phase.

The major limitation of a wait-listed design is well-known – it cannot be used to evaluate anything but short-term impact since by the end of the phase 2, no subjects remain in the control condition. Thus the long-term effects of preventive interventions cannot be assessed with such designs. Nevertheless, due to their inherent advantages, various types of wait-listed designs have appealed to researchers and communities alike.

One motivating example for our work in developing the new class of dynamic wait-list designs is the Mpowerment community-based preventive intervention. That intervention aimed to prevent HIV/AIDS by changing norms and empowering individuals to change their risky sexual behavior [15,17,18]. Since the Mpowerment intervention required intensive onsite training in a community, the study team could only train one community at a time. The design stipulated that two communities be randomized to either early or later intervention. With baseline measures collected on both communities at the start of the study and the same data collected on both communities after the early intervention community had received the intervention, an unbiased estimate of intervention effect on this pair of communities could be obtained.

This project actually used multiple baselines and follow-ups on the same two communities. However, as originally intended, this project was designed to expand to additional communities that would be randomly wait-listed at the time they entered the study. After both of the original two communities had received the intervention, a matched pair of two new communities could then be randomized to the same intervention. The process can then be repeated, taking a new pair of communities and randomly assigning them to receive the same intervention immediately or later. Even if the communities were not perfectly matched, the random assignment would allow one to assess intervention impact without potential bias from community readiness or other factors that could easily confound any intervention effect in a
non-randomized trial. Randomizing a number of pairs of such communities to either immediate intervention or a wait-listed condition would then over time provide unbiased data on intervention effect. While such a continued design for evaluating the Powerment intervention has not occurred to date, the novelty of the design has prompted us to investigate how similar ideas can be used in practice [4].

In this paper, we begin by describing the rationale for an ongoing trial where a classic wait-listed design has been used to evaluate a gatekeeper training suicide prevention program in a large public school district. To date, only a handful of suicide prevention programs have ever been evaluated with rigorous randomized trials such as the one we describe here [11,13,22,28]. Epidemiologic evidence is provided that indicates there is currently a low rate of referral of suicidal youth by school staff. This provides a potential for a gatekeeper training program, such as the one being tested in the current trial, to enhance the referral of suicidal youth. The primary goal of this prevention program is to successfully identify more children who are suicidal so that they can be referred to the existing mental health system for intervention. Like other population-based interventions, the evaluation requires us to consider that the level of suicidal behavior as well as the rate of referral may well be influenced by uncontrolled external events – such as well-known contagion effects from a celebrity committing suicide [12,21,29]. The potential influence of these uncontrollable fluctuations makes it imperative to have a design involving randomization that also blocks on time; that is, every time interval that provides useful data on intervention impact must have both intervention and control units.

Our new contribution is to introduce the dynamic wait-listed design. This design allows for blocking on a number of smaller time units. We demonstrate using both linear and log-linear random Poisson regression models that this new design is always more efficient than that of the classic wait-listed design. The linear model provides an analytic expression for evaluating efficiency gain as a function of the number of time intervals. The log-linear model is more standard for the field and is used to assess the quantitative effects of changes in the number of time intervals, intervention effect and degree of variation across time. We conclude that dynamic wait-listed designs generally provide large gains in efficiency. Furthermore, such designs are logistically easier to implement in settings where significant training resources are required to implement an intervention. Finally, we return to the gatekeeper training trial for suicide prevention. Calculations are used to project the improvement in statistical power as well as logistical training effort provided by changing to a dynamic wait-listed design midway through the study.

**Youth suicide and the Georgia gatekeeper project**

Despite some recent information indicating that youth suicide in the United States has declined over the last decade (eg, from seven to less than five per 100000 aged 14–19) [5,24,36], suicide is still the third most common cause of death for youth. Nationally, nearly 10% of young people report having attempted to commit suicide [14]. Every death by suicide can be exceptionally painful to family members and friends. However, the loss of a child through suicide imparts a special burden to a family and community, including increased risk for stressful events, family distress and relationship problems [3,23,33], all of which may increase the suffering of surviving family and friends whenever a youth suicide occurs. With the latest reported rates of suicide deaths of 9.9 per 100000 among those aged 15–25 [24] and mortality rates ranging between 5–10% for middle and high-school youth [5,6], even modest size communities are likely to confront a youth suicide nearly every year.

Our current scientific knowledge base about suicide prevention is typified by well identified epidemiologic risk factors, developed from psychological autopsy studies, retrospective case control studies and prospective longitudinal studies [11]. Several treatments are considered “promising” for reducing suicide among high-risk clinical groups [26], but there are very few population-based preventive interventions that have received a scientific evaluation using randomized trials or other high quality designs [13], despite a national priority for increased scientific rigor [35]. One of the primary difficulties with conducting such trials is that suicides in the general population are, despite the enormous suffering they bring, relatively less frequent in occurrence compared to other targets of prevention programs. This so-called “low base rate” condition makes it much more difficult to test preventive interventions aimed at reducing youth suicide.

In 2003 the present authors were invited by a large school district in Georgia to help evaluate a school-based gatekeeper training program for suicide prevention. Drawing on funds provided by the State of Georgia Legislature for suicide prevention, the school district had already decided to train all its middle and high school staff members in a program called QPR [30], which stands for “question, persuade and refer”. The QPR gatekeeper program is designed to enhance each adult staff member’s ability to recognize signs of youth who

www.SCTjournal.com
are contemplating suicide, to give them skills to approach such children and question them openly about suicide, and to refer them to treatment using the centralized referral system in a school district. It is recognized that adults who are in a youth's formal network, eg, a teacher or informal network, eg, a cafeteria worker, can serve as either positive or negative roles in linking a distressed youth to mental health services [34]. By training all staff in a school regarding appropriate steps for a suicidal youth, gatekeeper training programs aim to increase the linkage to beneficial services. This particular school district is well-suited for a study of gatekeeper training. The existing system of services for youth demonstrating suicidal or other life threatening behavior in this school system is backed up by an excellent relationship with the local mental health system. Free mental health assessments are offered to all families where a crisis response is needed, and these are generally completed within one day.

In a collaborative study design process, the school district agreed to implement the QPR training using a randomized wait-listed design. We received funding from the National Institute of Mental Health in 2004 to begin a trial in 32 middle and high schools with 60,000 children. Currently the district is in the process of QPR training for the 2500 staff in the first half of 16 schools assigned to the early intervention condition.

Epidemiology of suicide and suicidality in this school district

Over the last 16 years, this school district has experienced an average of four child deaths each year with the number ranging from 0 to 10. With 60,000 school children between the ages of 11 and 18 in this district, this rate is relatively comparable with that of the current national average [24]. Even with this large size school district, the comparatively “low” frequency of youth suicides makes it unlikely to find a statistically significant reduction in suicides in a one or two-year study. Because suicide is comparatively a rare event, we have chosen to study suicidal behavior and ideation as the key target for intervention. As we show below, there is a substantial number of suicidal youths who are not being identified by a school. One general prevention strategy is to train all school staff to identify and refer such youth. We are now testing in a randomized trial whether QPR increases the identification and referral of suicidal youth.

For the first time this school district has begun anonymous surveys measuring suicidal behavior in eighth and tenth grade students. Although only a modest fraction of these students did talk to them about their thoughts and feelings, but relatively few staff felt they could identify signs of suicidality, understood what sources were available to assist suicidal youths, or felt they would be effective in referring students for help. Those latter constructs are key targets of QPR training. In prior implementations of QPR, data from pre and post-tests have shown that adults increase their knowledge as well as self-efficacy with gatekeeper training [30].

The school district has had a protocol and centralized system to identify and handle children demonstrating life threatening behavior. In the latter half of the 2003–2004 school year, 127 children were identified in the 32 study middle and high schools who required immediate crisis response due to suicidality, homicidal or other life threatening behavior. Approximately 15% of these were considered to be at such high risk that they required immediate inpatient services.

Overall, the survey data reveal that there are significant numbers of students who are suicidal but are not known to the school district. Based on the 6% prevalence of reported attempts to commit suicide in the last year, we would anticipate that 3600 students could be harboring significant thoughts and/or plans about suicide in this
population of 60,000 middle and high school students. It is estimated that only 5% (193/3600) of such suicidal children are currently identified and referred by the school staff. Thus a gatekeeper training program that helps identify such students and get them into treatment earlier would likely have an impact on reducing suicide attempts and, potentially, deaths through suicide.

We can use these data to make a crude assessment of what impact a gatekeeper training program could have on increasing appropriate referrals. It is believed that most youth who go on to commit suicide tell at least one other person in the week or two before this final act or otherwise manifest observable warning signs [5]. Let us define \( p_0 \) to be the probability that a single staff member in the school becomes aware that a student is suicidal and successfully refers that suicidal child in the school to appropriate crisis response, in the absence of any gatekeeper training program. It is possible to obtain a crude estimate of \( p_0 \) based on the previous data, the number of staff in the school, which averages 162, and some simplifying assumptions. If all staff happen to have the same probability of referring a suicidal child and the event of each staff member referring or not referring a single suicidal child is nearly independent, then the probability of a suicidal child being successfully referred is a function of the individual staff probability of referral, \( p_0 \), and the number of staff, \( N_{\text{staff}} \):

\[
Pr(\text{Referred}) = 1 - (1 - p_0)N_{\text{staff}}
\]

Solving for \( p_0 \) and setting \( Pr(\text{Referred}) \) to our estimate of 5%, a straightforward estimate of the individual's probability of successfully identifying and referring a suicidal subject is

\[
\hat{p}_0 = 1 - (1 - 0.05)^{1/162} = 0.00032
\]

This minute number suggests there is large room for improvement, and a gatekeeper training program is designed specifically to improve such rate. Defining \( \gamma \) as the increase in odds of a single trained person being able to effectively refer a suicidal child, and \( f \) as the fraction of staff at a school being trained, the probability of having a suicidal child referred is then

\[
Pr(\text{Referred}) = 1 - (1 - \hat{p}_0)^{f}\gamma N_{\text{staff}} (1 - p_0)N_{\text{staff}}
\]

where \( p_1 \) is the probability of a single staff member trained in the gatekeeper training program successfully referring a suicidal child,

\[
\frac{\hat{p}_1}{1 - \hat{p}_1} = \gamma \frac{\hat{p}_0}{1 - \hat{p}_0}
\]

Figure 1 uses these formulae to determine the predicted effect on the probability of referring a suicidal child as a function of the level of training saturation at each of the schools (abscissa) and different hypothesized effectiveness of the gatekeeper training, measured by the factor that measures training effectiveness. Note from the topmost curve that if the individual level odds of referring a child is increased tenfold by training (thus \( p_1 = 10 \times 0.00032 = 0.0032 \)), and 90% of the staff are trained, then the potential for identifying a suicidal youth can increase from 5% to 40%. This shows that even a small increase in everyone's ability to detect and refer suicidal individuals due to a successful gatekeeper training program can have a very large effect in identifying those who would not normally be detected.

In January of 2004, we implemented a wait-listed randomized trial to evaluate the QPR gatekeeper training model in this school district. The 32 eligible middle and high schools, which had never received QPR training, were stratified based on middle or high school and level of crisis referrals in the previous year (split at the median). Within each of these four strata, half of the schools were randomly selected to receive training in the first phase of the study. Thus staff in one-half (16) of all schools were designated to receive QPR training during the 2003–2004 school year. These schools are called “early intervention” schools. The other one-half of the schools were placed on a “wait list” and served as a control group during this first school year. During the next school year, these wait-listed schools will receive QPR training. We also took steps to ensure that no school or child ever receives less support than now provided by the school district. In our analyses, we will examine the effectiveness of the gatekeeper training in improving surveillance of students at risk for suicide and timely referral of these students for mental health service. The primary outcome of this prevention trial is the rate of detection of children who have been identified by the school as suicidal.
Training demand in a wait-listed design

For a typical wait-listed design, half of the units are immediately assigned to intervention after baseline data are collected. Thus, half of all training begins at the same time. It can be burdensome for school systems and other organizations to manage all this training at once, and it normally takes some time for such training to be scheduled and begin at all. In the Georgia gatekeeper trial, one can see that no training occurred until approximately 40 days into the study (see Figure 2). Even though it may take some time to begin training and to provide all the training to half the units, we would typically treat the date of randomization or date of first training as the beginning of the trial.

This tradition, inherent in the "intent to treat" approach to randomized trials [10,7,9,16], can have the effect of attenuating differences between intervention and control conditions since little or no differences exist until training is actually provided. During the first 125 days of training, the school district was able to train 1387 out of 2498 (56%) staff in the 16 early intervention schools. The level of training over time for all these schools was not uniform as shown in Figure 3. Virtually all training occurred within schools within a few days, so that no additional training occurred over time once training began. Also, schools that began training earlier had higher completion rates than schools starting later, as shown in Figure 4. These figures indicate that training continued to occur throughout most of this time; for single schools most of the training occurred over a short period of time, and these start times for the different schools were dramatically different. This level of incompleteness in training has persuaded us that we need to extend training in these early intervention schools over a longer period of time through the current school year, to provide booster training for those staff trained in the previous year, and to revise the system of training for the wait-listed schools. The major implication of these data is that inefficiencies appear to exist regarding the timing of training. In the dynamic wait-listed design described below, the training schedule is likely to better match the logistical challenges inherent in a large school system or other community setting with multiple sites for training.

Dynamic wait-listed designs

We introduce a generalization of the wait-listed randomized design that is well suited to examining intervention impact when intervention condition
is assigned at either the group (eg, school) or individual level. In this new design, all units begin in the control condition and at specified time intervals are switched to active intervention. If the number of total units to be assigned is N and the number of time intervals in the design is τ, then at the beginning of each new time interval, an additional \( m = N/\tau \) units are randomly selected from those still in the control condition and are assigned to start the intervention at that time. This process continues until all units are assigned to intervention. Table 1 compares a standard wait-listed design, where \( \tau \) is two, with one where \( \tau \) is eight. Both of these designs take place over the same total time interval, which we have set to two years. The primary difference, however, is that a wait-listed design can compare intervention versus control only in the first year whereas the dynamic wait-listed design continues to allow comparisons of intervention versus control for all time units but the last time interval when all units are finally in the active intervention condition.

**Variance in intervention impact estimate and power for a dynamic wait-listed design**

By dividing the total time of the study into \( \tau \) time intervals, the dynamic wait-listed design permits a comparison of intervention and control conditions at every time interval except the last, since only in this last time interval are all units allocated to the same active intervention condition. This provides an advantage over the classic wait-listed design which provides a legitimate comparison only in the first half of the study as seen in Table 1. On the other hand, it may appear that the statistical power of a dynamic wait-listed design is adversely affected by an imbalance in the number of intervention and control units at each time period as well as the shortened time blocks for assigning new units to active intervention. As we show below, however, for virtually all cases, the dynamic wait-listed design actually increases statistical power despite these last two countervailing factors.

We now consider power calculations for examining intervention impact when the outcome variable is based on a rate, such as the number of referred suicidal children per fixed unit of time. Our development below begins with a linear random effects Poisson model since in that case we can obtain an exact expression for the general least squares solution. We then provide similar results for the more traditional log-linear random effects Poisson model.

Let \( T \) be the total time in the entire study and partition the time interval from 0 to \( T \) into \( \tau \) equal time intervals. At each time interval \( \{t-1\}t/\tau, t/T \) indexed by \( t, t = 1, \ldots, \tau - 1 \), let \( X_t \) represent the time adjusted rate of referred suicidal children in those units who have not yet been assigned to the intervention by time \( j \) and \( Y_t \) be the same time adjusted rate of referred suicidal children in those units that have not yet been assigned to intervention by time \( j \). Note that each of these rates are comprised of the sums of reported rates for each unit in the two intervention conditions over that particular time interval. Specifically, let \( S_k \) be the random number of referred suicidal children in time interval \( t \) from the \( k \)th unit (here school), \( k = 1, \ldots, N \). Also let \( A_k \) be the random assignment time for the \( k \)th unit, \( k = 1, \ldots, N \).

A simple model we consider first is one of a constant intervention effect over all units along with an additive random Poisson time effect to allow for time variations in referrals. It turns out that a closed form solution exists for the additive Poisson model but not for the more traditional multiplicative Poisson model. Thus we begin with the additive Poisson model where we show that the variance of the intervention effect estimated by generalized least square quickly decreases with \( \tau \). In this additive Poisson model let \( \gamma \) be the random effect at time \( t \), with a mean of zero, and let the Poisson rate parameter \( \lambda_0 \) refer to the control condition and \( \lambda_1 \) refer to the intervention condition

\[
S_k = \text{Poisson}(\lambda_0 + \gamma(t)/\tau) \quad \text{if } A_k > t \quad (5)
\]

\[
- \text{Poisson}(\lambda_1 + \gamma(t)/\tau) \quad \text{if } A_k \leq t \quad (6)
\]

**Table 1** Unit assignment for standard wait-listed and dynamic wait-listed designs

<table>
<thead>
<tr>
<th>Year</th>
<th>Time block</th>
<th>Wait-listed design</th>
<th>Dynamic wait-listed design</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Intervention</td>
<td>Control</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>N/2</td>
<td>N/2</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>N/2</td>
<td>N/2</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>N/2</td>
<td>N/2</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>N/2</td>
<td>N/2</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>N</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Year</th>
<th>Time block</th>
<th>Interventio n</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>m</td>
<td>N - m</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>2m</td>
<td>N - 2m</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>3m</td>
<td>N - 3m</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>4m = N/2</td>
<td>N - 4m = N/2</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>5m</td>
<td>N - 5m</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>6m</td>
<td>N - 6m</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>7m</td>
<td>N - 7m</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>8m = N</td>
<td>N - 8m = 0</td>
</tr>
</tbody>
</table>

www.SCTjournal.com

*Clinical Trials* 2006; 3: 259–271
with \( \gamma_t \) being independent at each time interval and having zero mean and variance given by \( \sigma^2 \). This model is equivalent to treating time as a blocking factor. Then the observed average rates per time \( T \) of suicidal referrals in the intervention and control conditions at time \( t \) are given respectively by

\[
X_t = \frac{\sum_{k=A,T} S_{kt}}{mt} \times \tau
\]

(7)

\[
Y_t = \frac{\sum_{k=A,T} S_{kt}}{N-mt} \times \tau
\]

(8)

Using the independence assumption,

\[
X_t - \lambda m t \times \text{Poisson}(\lambda_1 + \gamma_t / \tau)
\]

and similarly

\[
Y_t - \lambda (N - mt) \times \text{Poisson}(N - mt) / (\lambda_0 + \gamma_t / \tau)
\]

(10)

These have conditional expectation \( \lambda_1 + \gamma_t \) and \( \lambda_0 + \gamma_t \) respectively, and their conditional variances are given by \( \text{Var}(X_t | Y_t) = (\tau / N) (\lambda_1 + \gamma_t) \) and \( \text{Var}(Y_t | Y_t) = (\tau / (N - mt)) (\lambda_0 + \gamma_t) \) respectively. In this model we have assumed that \( \gamma_t \) has zero mean so the unconditional variance of the difference \( X_t - Y_t \) under the null model of no intervention effect \( (\lambda_0 = \lambda_1 = \lambda) \) is

\[
\text{Var}(X_t - Y_t) = (1 / t + 1 / (N - mt)) \times \lambda \tau
\]

(11)

To obtain an efficient combined estimate of the difference between intervention and control conditions over all the \( \tau - 1 \) time blocks, one should obtain a linear estimate where the weights are inversely proportional to each variance, that is, letting \( w_t = (1 / t + 1 / (\tau - t))^{-1} \), the best unbiased weighting for \( \lambda_1 - \lambda_0 \) is

\[
\theta = \frac{\sum_{t=1}^{\tau-1} w_t (X_t - Y_t)}{\sum_{t=1}^{\tau-1} w_t}
\]

(12)

The variance of \( \hat{\theta} \) under the null model of no intervention effect is then

\[
\text{Var}(\hat{\theta}) = \frac{1}{\sum_{t=1}^{\tau-1} w_t} \lambda \tau / m
\]

(13)

This can be simplified by noting that

\[
\sum_{t=1}^{\tau-1} w_t = \sum_{t=1}^{\tau-1} \frac{1}{t + 1 / (\tau - t)}^{-1}
\]

\[
= \sum_{t=1}^{\tau-1} \frac{1}{t (\tau - t)}
\]

\[
= \sum_{t=1}^{\tau-1} \frac{\tau - 1 - t^2}{\tau t}
\]

\[
= (\tau - 1) / \tau - 2 (\tau - 1)(\tau - 1) / 6
\]

\[
= (\tau^2 - 1) / 6
\]

(14)

Inserting this last expression into Equation (13), we obtain,

\[
\text{Var}(\hat{\theta}) = \frac{6 \lambda \tau}{m \tau^2 - 1)
\]

\[
= \frac{6 \sigma^2}{m \tau^2 - 1)
\]

(15)

with the last expression resulting from the relationship \( m = N / \tau \). Since only the first factor depends on \( \tau \), this factor \( \tau^2 / (\tau^2 - 1) \) determines the relationship between the number of time intervals and the variance of the resulting estimate. This factor is decreasing in \( \tau \) when \( \tau = 2, 3, \ldots, N \), implying that the variance of the estimator decreases with increasing number of time intervals. The relative reduction in variance is most pronounced for small \( \tau \), for \( \tau = 3 \) the improvement in precision is 18%, for \( \tau = 4 \) the gain is 25%, and the limiting value is 33% gain in efficiency when \( \tau = N \) and each unit is randomly assigned to intervention individually.

Results for marginal maximum likelihood estimation of intervention impact under a gamma mixture of Poissons shows an even more pronounced gain in efficiency, especially with increasing intervention effect. Specifically, if we assume that the count for each time point and each unit has a Poisson distribution depending on intervention status and a random factor at each time point,

\[
S_{kt} \sim \text{Poisson}(\mu_{kt}) \quad \text{if } \lambda_k > t
\]

\[
\sim \text{Poisson}(\mu_{kt}) \quad \text{if } \lambda_k = t
\]

(16)

(17)

with

\[
\mu_{kt} = \gamma \lambda_k^t / \tau
\]

(18)

\[
\mu_{kt} = \gamma \lambda_k^t / \tau
\]

(19)

where the random factors \( \gamma_k \) are centered around 1 and have independent and identically distributed gamma distributions, \( \gamma_k \sim \Gamma(\alpha, \rho)\), \( t = 1, \ldots, \tau - 1 \) with mean \( \alpha / \rho = 1 \) and variance \( \alpha / \rho^2 \).

In Figure 5 we present an examination of the efficiency or ratio of asymptotic variances of marginal maximum likelihood estimates intervention impact, \( \lambda_k / \lambda_0 \) for the dynamic wait-listed design compared to the standard wait-listed design. Plotted on the abscissa is the number of time intervals, with 2 corresponding to the classic wait-listed design. This figure demonstrates the efficiency gain when there is both small and large variation in rates over time, as well as the impact of differing levels of intervention effectiveness. Beginning at the bottom most curve for no intervention impact and high variability in rates across time, this shows the rapid increasing efficiency to an upper limit of 1.3 just as we presented for the general weighted least squares result in the previous section. Next note that in this
case of high variability in rates across time, the efficiency increases substantially with increasing effect size, from 2, to 4, to 8. The topmost curve indicates the efficiency gain at the other extreme when there is low variation in the rates across time. Note that efficiency increases to a maximum of 2, and there is no difference in efficiency as a function of effect size. We can thus conclude that efficiency gains can vary from 1.3 to 2.0, with larger efficiencies with stronger intervention impact and smaller variation in these rates across time.

Use of a dynamic wait-listed design in the ongoing Georgia gatekeeper project

The previous findings of gains in efficiency have implications for our current design of the Georgia gatekeeper project even though we are already in phase 1 of a classic randomized wait-listed design. Traditionally, the second phase of a wait-listed design, during which staff in the wait-listed schools would become trained, is not used for evaluating intervention impact. However, it is still possible to randomly assign these remaining wait-listed schools to start at different time intervals in the second half of the study. That is, all the wait-listed schools are randomly assigned to two or more start times during the last half of the study. We present the effects on statistical power afforded by this design change in Figure 6. In this figure the abscissa corresponds to the efficiency, in terms of the overall school rate, of the intervention to identify suicidal youth relative to control. Note that with the traditional wait-listed design ($\tau = 2$), there is 80% power to detect a 32% increase in referral rates in the intervention condition compared to controls. For the planned modification where schools continue to be randomized to intervention at different times in the second phase, there is 80% power to detect a 23% increase in referral rates. This gain would correspond to adding six schools to the current design.

We plan to continue to randomize at the group (ie, school) level to immediate or delayed intervention. However, unlike a wait-listed design, which we used at the start of the Georgia gatekeeper trial, we do not require that half of the groups wait a fixed amount of time, ie, until the second year. This change will allow the school district to handle the scheduling of training and supervision necessary for a new intervention; it will also increase statistical power over the wait-listed group yet provide the same amount of protection in random assignment as does the original wait-listed design. To carry out this design, a blinded, randomized list of schools is generated; this determines the sequence of training of each school. At a specified advance time, the next block of schools in this random sequence is notified of its upcoming training so that scheduling can occur. Once this training in a particular school is complete, the next school on the list begins their training. This dynamic assignment of the sequence of schools or other groups continues until all schools are completely trained.

Discussion

The dynamic wait-listed design is a new variant of the classic randomized wait-listed design. We note three advantages, statistical efficiency, training efficiency and ability to improve statistical modeling. We have concluded that nearly every dynamic wait-listed design is much more efficient than that used in the traditional wait-listed design. Gains in
efficiency are substantial and range from 33% to 100% for most situations. The corresponding gains in statistical power due to this gain in efficiency can also be substantial. In our current trial, the switch to a dynamic wait-listed design in the second phase is equivalent to adding another six schools to the study. The power calculations in this paper are all based on a Poisson model, but this is not a requirement. As long as the variance in the counts is inversely proportional to the length of the time interval, the same results hold. If the variance in the outcome is not a function of the time interval, say a single follow-up self-report measure of suicidality or time to a suicide attempt, then the power for this wait-listed design is expected to do no better or no worse than that of a classic wait-listed design.

In many instances where training resources are limited, a dynamic wait-listed design is expected to speed up the training process and result in more complete training. A dynamic wait-listed design requires that finite training resources be allocated repetitively to a small number of units. In our particular trial that is testing the QPR gatekeeper program, this concentration on the training of staff in a few schools at a time would make it easier to approach full saturation in each school soon after being assigned to active intervention status. Complete saturation of training should also translate into higher levels of referral of suicidal children, as shown in Figure 1.

A third advantage to using a dynamic wait-listed design is that it allows for more general statistical modeling of intervention impact that is likely to reflect the realities of any intervention.

For example, we have omitted any discussion of variation in intervention effectiveness, including the potential for the effects of training to wear off over time. In the classic wait-listed design there is no way to unconfound the actual time of training with that of selection factors such as readiness for training or organizational climate. We have seen in our trial evidence that schools that train their staff early are more successful in achieving nearly complete training. By allowing the actual time of training to vary in a dynamic wait-listed design, however, we can model whether referrals drop off as a function of length of time since training occurred. Such information would be useful in determining if and when booster training is needed.

In a sufficiently large trial, it would also be possible using a dynamic wait listed design to determine if an intervention had differential effectiveness at specific periods during the year. For example, rates of suicide tend to be higher in the winter and spring [19,27,31]. By allowing time of training to vary in a dynamic wait-listed design, it would be possible to test if the effectiveness of gatekeeper training was more effective at certain phases of the year (eg, several months before elevated suicidal behaviors). Such information could enable communities and schools to use finite training resources more effectively by focusing training during periods of maximal impact.

One could ask which of the dynamic wait-listed designs is most efficient. The solution to this is rather straightforward. Simply conduct a standard wait-listed design with half of the units immediately assigned to intervention and the remaining units assigned to intervention at the last possible time. Thus the most efficient dynamic wait-listed design approaches that of a standard wait-listed design that lasts twice as long, hence the upper bound of 100% efficiency gain. However, this is not a particularly useful design. It is unlikely that communities which are interested in having all units in a study receive the intervention would favor waiting twice as long for any in the second half to receive the intervention. The design change that we have proposed in the Georgia gatekeeper project, which converts from a standard wait-listed design to that of a dynamic wait-listed design in the second phase is more appealing and still increases power substantially.

While most randomized trials still involve assignment of individuals to condition, group-based or place-based randomized trials have recently become popular [1,32]. This new dynamic wait-listed design adds a third dimension for randomization, that of time. There is symmetry in this respect with the epidemiologic complex of person, place, and time. Now all three, or combinations of these three, can be used in random assignment to create an efficient design. In our trial we have randomized at the school (place) level and are now randomizing across multiple time periods in a dynamic wait-listed design. Similarly, randomization of individuals can take place at different times as well. The traditional RCT that stratifies subjects, say on age and gender, and blocks on the sequential order of subjects is just one example of a design that randomizes individuals across time. We suggest that combinations of these three factors of person, place, and time can also be useful in evaluating interventions thought to have different impact across both subject characteristics and contextual characteristics. This could be useful in testing the effect of timing of an intervention on a disorder that is affected by gene by environment interactions.

Being a true randomized trial, the dynamic wait-listed design has clear advantages over interrupted time series designs. The first of these designs, introduced by Cook and Campbell [37], did not specifically involve randomization, but such
opportunities were soon recognized [38]. There are, however, somewhat related multiple baseline designs that do involve randomization similar to that presented here [8,39]. Such multiple baseline designs are often used in education with a small number of subjects (typically four or so) and a large number of times for testing (typically more than 15). The scores of each person prior to their random transition time to the intervention is considered baseline and is compared to those scores after the transition time. While the multiple baseline design is generally used for assignment of a small number of individuals who are measured on a continuous outcome, this same strategy of randomly assigning of time of intervention, is applicable to group randomization, to larger studies, and the modeling of low frequency counts with individual level as well as group level covariates as in the example presented in this paper.

Designs that randomize intervention time may be quite valuable in evaluating the effectiveness of programs as they are expanded or "scaled up" in a community. It is one thing to demonstrate effectiveness of an intervention with a randomized trial using 12 schools, for example; it is another thing to implement this intervention system-wide. Scalability is of fundamental importance for dissemination and implementation research since this is the final step in implementing evidence-based or empirically supported prevention programs in drug abuse, mental health and social services, and education. To date few scalability studies have considered randomly assigning the time that classes, schools, families or agencies begin intervention. We believe that randomization has an important role in the design of such studies, and such designs can be informed by the results described here.

These dynamic wait-listed designs are also appropriate for testing early detection screening approaches for cancer, HIV and other medical conditions as well. For example, to test the effectiveness of identifying lung cancer, one could randomize the time at which individual clinics switch from the usual practice where chest X-rays are occasionally ordered by physicians, compared to routine use of helical CT. At each interval of time, all eligible patients who visit the clinic can be tested with the new screening procedure. Rates of identification of lung cancer by cancer stage can then be compared, along with false positive rates and cost effectiveness evaluations. Such a design would work whether or not the screening protocol requires retesting at specified intervals or only a single test. In the latter case, as long as there are patients who continue to come for their first visit after a clinic begins the new screening procedure, the results of these screening tests can be used to assess intervention impact. Modeling over time would of course need to take into account patient characteristics and length-biased sampling since patients who attend more frequently are likely to differ from those coming to the clinic less frequently. Similar ideas were suggested by Braver and Braver [39].

There are situations where this dynamic wait-listed design should not be used. In particular, in those instances where it is not possible to vary training schedules, such as training in a new reading curriculum where all teachers need to be prepared at the beginning of the year, random start times throughout the year would be inappropriate. However, a dynamic wait-listed design could potentially be used over multiple years to accomplish a rolling-in of a new curriculum. In this way supervision or coaching during the year, rather than initial training, could be concentrated on a limited number of newly trained teachers each year. A second set of circumstances that would make this new design inappropriate is under a condition where those randomized to wait may end up withdrawing from the study before they are assigned to active intervention. For example, in clinical trials of interventions to reduce suicide among individuals with prior histories of suicide attempts [26], it may be inappropriate to extend waiting periods for some individuals given the high-risk status of all trial participants. A classic wait-listed design that trains everyone eligible soon after assignment would have a smaller average wait-time to intervention compared to the dynamic wait-listed design. Also, in a dynamic wait-listed design the variance in time to beginning the intervention increases with the number of time intervals, so those towards the end of the line may be more likely to withdraw or find an alternative active intervention. Of course, these differences in time to training of a standard versus dynamic wait-listed design diminish if much of the training cannot be accomplished until towards the end of the study.

Acknowledgements

Work on this paper was supported by the National Institute of Mental Health, the National Institute on Drug Abuse and the Centers for Disease Control and Prevention under the main grant and supplements of RO1 MH40859, as well as from the National Institute of Mental Health under grants R34 MH071189 and P20 MH071897. We thank our colleagues in the Prevention Science and Methodology Group (PSMIG), especially Dr Wei Wang, for their comments, for support from our colleagues at the University of Rochester Center for Public Health and Population Interventions for the Prevention of Suicide, as well as Drs Kevin
Cain, Elaine Thompson and other members of our Data Safety and Monitoring Committee for initial discussions that led to the development of this design. We also want to thank Jerry and Elsie Weyrauch without whom the Georgia Gatekeeper project would never have taken place. They have worked tirelessly for the development of rigorous scientific studies in the prevention of suicide. Special thanks go to our colleagues in the school district, who remain anonymous for reasons of school policy, for their dedication to their students as well as to this evaluation.

References

Appendix. Derivation of design efficiency for dynamic wait-listed designs based on a $\Gamma$ mixture of Poisson counts

In this appendix we present the general formulae for the variance for maximum likelihood estimators of two Poisson rate parameters when different numbers of units are assigned to intervention or control conditions over time. For time $t$, $t = 1, \ldots, T-1$, let $m_{it}$ be the number of units assigned to the intervention condition and $m_{it}$ be the number assigned to the control condition. Define $N_{it} = \sum_{k:A_k=t} S_k$ and $N_{it} = \sum_{k:A_k=t} S_k$ as the total number of observed counts in intervention units and control units respectively at time $t$. Let $N_t = N_{it} + N_{it}$. Also define $b_t = (m_{it} \Lambda_1 + m_{it} \Lambda_0) / \tau$, the expected number of reported events in interval $t$ when $Y_t = 1$. The joint density of the counts at time $t$ is then given by

$$f(S_{t1}, \ldots, S_{tT}) \propto \frac{\mu^S}{\Gamma(\alpha)} \exp(-\mu) \times \prod_{i=1}^{T} \left( \exp \left( \frac{m_{it} \Lambda_1 + m_{it} \Lambda_0}{\tau} \right) \right) \times e^{Y_t (m_{it} \Lambda_1 + m_{it} \Lambda_0) / \tau} d\mu^N_{it} d\mu^N_{it} \tag{20}$$

The log likelihood is then

$$\log L(\Lambda_1, \Lambda_0) = \text{const} + \sum_{t=1}^{T-1} (\alpha + N_t) \mu_{it} \log(\rho + b_t)$$

$$+ N_{it} \log \Lambda_1 + N_{it} \log \Lambda_0$$

The score equations are

$$\frac{\partial \log L}{\partial \Lambda_1} = \sum_{t=1}^{T-1} \frac{(\alpha + N_t) m_{it}}{(\rho + b_t)^2} \tau + \frac{N_{it}}{\Lambda_1}$$

$$\frac{\partial \log L}{\partial \Lambda_0} = \sum_{t=1}^{T-1} \frac{(\alpha + N_t) m_{it}}{(\rho + b_t)^2} \tau + \frac{N_{it}}{\Lambda_0}$$

The information matrix is then given by

$$I = \begin{pmatrix}
-\sum_{t=1}^{T-1} \frac{(\alpha + N_t) m_{it}^2}{(\rho + b_t)^2} \tau^2 & \sum_{t=1}^{T-1} \frac{(\alpha + N_t) m_{it} m_{it}}{(\rho + b_t)^2} \tau^2 \\
\sum_{t=1}^{T-1} \frac{(\alpha + N_t) m_{it} m_{it}}{(\rho + b_t)^2} \tau^2 & -\sum_{t=1}^{T-1} \frac{(\alpha + N_t) m_{it}^2}{(\rho + b_t)^2} \tau^2
\end{pmatrix} \tag{25}$$

The expected information is obtained by noting that

$$E(N_{it}) = E(N_{it}) E(S_k | Y_t)$$

$$= m_{it} \Lambda_1 E(Y_t)$$

$$= m_{it} \Lambda_1 / \rho$$

$$E(N_{it}) = m_{it} \Lambda_0 / \rho$$

The final step is to calculate the variance of the MLE for $\Lambda_1 / \Lambda_0$ (or its logarithm). This can be obtained through an application of the multivariate Delta method.