Population metrics for suicide events: a causal inference approach

He, H., Lu, N., Stephens, B., Xia, Y., Bossarte, R.M., Kane, C.P., Tang, W., Tu, X.M.

6/10/2019

Abstract

Large-scale public health prevention initiatives and interventions are a very important component to current public health strategies. But evaluating effects of such large-scale prevention/intervention faces a lot of challenges due to confounding effects and heterogeneity of study population. In this paper, we will develop metrics to assess the risk for suicide attempts based on causal inference framework when the study population is heterogeneous. The proposed metrics deal with the confounding effect by first estimating the risk of suicide attempts within each of the risk level and then taking a weighted sum of the conditional probabilities. The metrics provide unbiased estimates of the risk of suicide attempts. Simulation studies and a real data example will be used to demonstrate the proposed metrics.

Keywords: causal inference; effective sample size; metrics; multiple events; population heterogeneity; potential outcome

1 Introduction

Large-scale public health prevention initiatives and interventions are a very important component to current public health strategies. In order to properly evaluate and determine the most effective initiatives and interventions, more objective measures need to be created, particularly in the face of large-scale, uniform initiatives with observable heterogeneity across reporting units. Among the most important challenges facing current public health strategies is the lack of timely data and associated objective quantitative indicators that account for key aspects of population heterogeneity.

For example, in recent years, the Department of Veterans Affairs has implemented a series of clinical and preventive programs between and across Veterans Health Administration (VHA) medical facilities to reduce rates of suicide and associated other cause mortality, and improve the overall quality of life of Veterans. However, there are few quantitative indicators with the degree of granularity or timeliness required to evaluate the impact of newly-implemented programs and/or identify those areas in need of additional support. In suicide research, existing literature suggests that history of suicide attempt/serious ideation resulting in hospitalization confers significant risk for repeated suicide events and the resulting increase in risk persists for an extended period of time. Thus, an observed suicide event represents a cogent marker of risk for identifying subjects for preventive treatment. Observed numbers or rates of suicide events from VHA facilities cannot be directly used to assess the effect of newly-implemented suicide prevention efforts, because of heterogeneity among patients across VHA facilities.

For example, when a suicide attempt does not result in a death, literature suggests that a history of suicidal events is a significant risk factor for future suicide events, and the greater the history of prior suicidal events the more likely of a future suicide event over a given period of time. As a result, the risk level of a subject is dynamically changing with the intervention outcome and the interwoveness of the predictor and outcome calls for more carefully designed metrics to take into such dynamic relationships between risk and suicide attempt.
In this paper, we develop a metric to address the aforementioned difficulties. This approach is readily applied or expanded for program evaluation of other episodic diseases. We use both real and simulation studies to illustrate as well as assess the performance of the proposed approach.

2 Rationale for New Metrics

In this section, we use hypothetical example in suicide attempts to illustrate the statistical issues to motivate our development. For simplicity, we consider only one risk factor, history of suicide attempt and assume that history of suicide attempt is a categorical outcome with three levels: no suicide attempt, one suicide attempt, and two or more suicide attempts.

Consider two facilities $F_1$ and $F_2$ with the same number of subjects 3200. Both facilities receive some type of suicide prevention intervention. We present three different scenarios, depicting different distributions of number (proportion) of suicide attempt.

Table 1: Three scenarios for the hypothetical example

<table>
<thead>
<tr>
<th>History</th>
<th>Scenario</th>
<th>Facility $F_1$</th>
<th></th>
<th>Facility $F_2$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Subj. S.A. Rate(%)</td>
<td></td>
<td>Subj. S.A. Rate(%)</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1-3</td>
<td>1000 10 1.0</td>
<td></td>
<td>3000 30 45 45 1.0 1.5 1.5</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1-3</td>
<td>2000 40 2.0</td>
<td></td>
<td>100 2 3 3 2.0 3.0 3.0</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1-3</td>
<td>200 10 5.0</td>
<td></td>
<td>100 5 6 1 5.0 6.0 1.0</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>3200 60</td>
<td></td>
<td>3200 37 54 49</td>
<td></td>
</tr>
</tbody>
</table>

Show in Table 1 are number of subjects ("Subj."), and number ("S.A.") and rates ("Rate") of subjects who have suicide attempts in a time window such as 6 months broken down by number of prior suicide attempts ("History"). We assume that the distribution of subjects with prior attempts and the relationship between prior attempts and subjects with suicide attempts do not change across all scenarios for $F_1$. For $F_2$, while the distribution of subjects with prior attempt remains the same, the relationship between prior attempts and subjects with suicide attempts varies across the scenarios.

In the first scenario, rates of suicide attempts are the same between the two facilities within each category of history of suicide attempt. The intervention performs equally well in both facilities. However, if we ignore history of attempt and simply compare the overall rate (total number of suicide attempters/sample size), then $F_1$ is significantly worse than $F_2$ (p-value = 0.024). In the second scenario, $F_1$ performs uniformly better than $F_2$ based on the rates of suicide attempts stratified by the history of suicide attempt. However as in the first scenario, $F_2$ again has a lower overall rate than, although not significantly different from, $F_1$ (p-value = 0.64) if the history of attempt is ignored.

The seemingly “non-sensible” results in the two examples above are really the phenomenon of the Simpson’s paradox, or selection bias in the lingo of causal inference. History of suicide attempt, which is highly related to the outcome, is not controlled for when performing comparisons. If we control for this selection bias and compare the two facilities within each
category of the history variable, then results make perfect sense. However, it can still be difficult to compare facilities even if controlling for prior attempts.

In the third scenario, $F_1$ fares better than $F_2$ for categories 1 and 2, but worse in category 3. What can we say about the two facilities in terms of the intervention effect? If we assign more weights on the last category of the history outcome, $F_2$ would perform better. But, if we assign more weights on the first two categories, $F_1$ would be better. Thus to compare the facilities, we must use weights based on objective criteria. This is the focus of the next section.

3 New Metrics for Comparing Suicide Attempts

We develop an approach for program evaluation for episodic diseases such as suicide attempt by framing problems under the potential outcome based causal inference framework Rosenbaum and Rubin (1983); Rosenbaum and Rubin (1984); Rosenbaum (2002). For simplicity, we discuss the development by focusing on suicide attempt. We again consider two facilities, $F_1$ and $F_2$, and assume the population can be divided into $K$ homogenous groups, $H_1, H_2, \ldots, H_K$, based on risk factors such as history of suicide attempt, age, gender and medical conditions.

3.1 Potential-outcome and Binary Outcome

For any randomly selected subject, let $y_{i1}$ ($y_{i2}$) denote the potential outcome of suicide attempt over a given period of time $T$, such as one year, if s/he receives the intervention within Facility $F_1$ ($F_2$). The intervention effect for the subject is $\Delta_i = y_{i1} - y_{i2}$. Since this difference is based on the outcomes from the same individual, the difference must be the result of differential intervention effect between $F_1$ and $F_2$. Unfortunately, $\Delta_i$ is not computable, since only one of $y_{i1}$ and $y_{i2}$ is observable; the one corresponding to the facility from which the individual receives the intervention. A large part of the causal inference literature centers around how to estimate the average, or population-level, treatment effect, $\Delta = E(y_{i1} - y_{i2})$.

In the absence of any selection bias,

$$\Delta = E(y_{i1} - y_{i2}) = E(y_{i1}) - E(y_{i2}) = E(y_{i1}|F_1) - E(y_{i2}|F_2).$$

This is because if subjects are randomly assigned to $F_1$ or $F_2$, then the mean of the potential outcome is the same as the mean of its observed counterpart from the facility to which the subject is assigned, i.e., $E(y_{il}) = E(y_{il}|F_l)$ ($l=1,2$). Thus, we can estimate $\Delta$ by estimating $E(y_{il}|F_l)$, the observed suicide attempts in Facility $F_1$ ($F_2$). However, when the assignment is not random, the mean of potential outcome may become different from its observed counterpart, because subjects may be assigned to a particular facility based on features that are correlated with the outcome, as amply illustrated by the examples in Section 2. In this case, we cannot estimate $\Delta$ by estimating the difference between $E(y_{i1}|F_1)$ and $E(y_{i2}|F_2)$ from observed data. For example, as illustrated in Section 2, none of the overall rates of suicide attempt can be used to compare the facilities in these examples, because history of suicide attempt, which is correlated with suicide attempt that occurs during the study, is not distributed evenly between the facilities.

To be able to estimate $\Delta$ based on observed outcomes and information in the risk groups $H_k$, we assume, as in the literature, a strong ignorable assumption:

$$(y_{i1}, y_{i2}) \perp (F_1, F_2)|H_k, \quad 1 \leq k \leq K,$$

(1)
where \( \perp \) denotes stochastic independence. The condition in (1) ensures that within each subgroup \( H_k \), assignment of \( F_l \) is independent of suicide attempt. Thus, when restricted to \( H_k \), there is no selection bias, implying:

\[
E \left( y_{il} \mid H_k \right) = E \left( y_{il} \mid H_k, F_l \right), \quad 1 \leq l \leq 2, \quad 1 \leq k \leq K.
\]  
(2)

The identity in (2) allows us to estimate \( E \left( y_{il} \mid H_k \right) \) using observed outcomes in each facility \( F_l \).

For example, if history of suicide attempt is the only confounding factor and \( H_1, H_2, \ldots, H_5 \) represent subjects with 0, 1, 2, 3, \( \geq 4 \) prior attempts, we can estimate \( E \left( y_{il} \mid H_k \right) \) by estimating \( E \left( y_{il} \mid H_k \right) \) within each \( H_k \) \((1 \leq k \leq 5)\).

By applying the rule of total probability, we can express \( E \left( y_{il} \mid H_k \right) \) as:

\[
\phi = E \left( y_{il} \right) = \sum_{k=1}^{K} E \left( y_{il} \mid H_k \right) \Pr(H_k) = \sum_{k=1}^{K} E \left( y_{il} \mid H_k, F_l \right) \Pr(H_k). \tag{3}
\]

For each facility, the metric \( \phi \) defined in (3) is a weighted average of conditional probabilities of suicide attempts within each risk group \( H_k \). The weights are calculated based on the distribution of \( H_k \) in the study population. Since

\[
E \left( y_{il} \mid F_l \right) = \sum_{k=1}^{K} E \left( y_{il} \mid H_k, F_l \right) \Pr(H_k \mid F_l), \tag{4}
\]

the unstratified rate \( E \left( y_{il} \mid F_l \right) \) differs from the proposed metric only in the weights, or the distribution of \( H_k \); \( \Pr(H_k \mid F_l) \) for the unstratified rate vs. \( \Pr(H_k) \) for the proposed metric. This makes perfect sense, since \( \Pr(H_k) \) reflects the distribution of subjects with the history variable in the study population, while \( \Pr(H_k \mid F_l) \) is specific to the selected sample in \( F_l \). If the selection process is random, then \( \Pr(H_k) = \Pr(H_k \mid F_l) \) and (3) is identical to (4). Otherwise, \( E \left( y_{il} \mid F_l \right) \) is different from \( \phi \).

### 3.2 Inference

Inference about the proposed metric involves estimation of the probability of suicide attempt within each risk level, \( E \left( y_{il} \mid H_k, F_l \right) \), and the weights, \( \Pr(H_k) \). Since a subject may continue to attempt suicide during the study, the risk changes over time. We may ignore the dynamically-changing risk and define \( H_k \) simply based on prior attempts at baseline. In this case, each subject can only contribute at most one event in the calculation of the conditional probability for the metric, even though the subject may attempt suicide multiple times during the study. Although still providing useful information for estimating reattempt rates, such analysis does not utilize all available, rendering it less efficient. A better approach is to include all events by allowing the risk to change as new events occur during the study period and accounting for such changes in the metric. This is especially true within the current context, since suicide attempt is a rare event and including all events allows one to obtain more efficient and reliable estimates.

We consider both approaches below and start with the relatively simpler approach that defines \( H_k \) with history of suicide attempt at baseline.
3.2.1 Metric with Fixed Group Membership

For a given facility $F_1$, let $N_j$ denote the number of subjects in $H_j$ and $n_j$ be the number of subjects who attempt suicide within a given time period $T$. Then $\Pr(y_{i1}=1|F_1, H_j)$ can be readily estimated by the observed proportion:

$$\Pr(y_{i1}=1|F_1, H_j) = \frac{n_j}{N_j},$$

provided that the following assumptions are met.

**Assumption (A1).** No Censoring: All subjects have been followed up in the entire period $T$.

If the weights, $\Pr(H_j)$, are known, substituting $\Pr(H_j)$ and $\Pr(y_{i1}=1|F_1, H_j)$ into (3) yields an estimate of $\phi$:

$$\phi_j = \Pr(y_{i1}=1) = \sum_{j=1}^{K} \Pr(y_{i1}=1|H_j, F_1) \Pr(H_j) = \sum_{j=1}^{K} \frac{n_j}{N_j} \Pr(H_j).$$

(6)

Otherwise, we need to estimate $\Pr(H_j)$. To this end, we further assume:

**Assumption (A2).** Random Sampling: a random sample of $M$ subjects is drawn from the study population, with $M_j$ denoting the number of subjects in the $j^{th}$ group. We then estimate $\Pr(H_j)$ by $\Pr(H_j) = \frac{M_j}{M}$ and substitute $\Pr(H_j)$ in place of $\Pr(H_j)$ in (6) to obtain:

$$\phi_j = \Pr(y_{i1}=1) = \sum_{j=1}^{K} \Pr(y_{i1}=1|H_j, F_1) \Pr(H_j) = \sum_{j=1}^{K} \frac{n_j}{N_j} \frac{M_j}{M}.$$  

(7)

As a special case, if the sample composition is the same as that of the study population, i.e., $\Pr(H_j) = \Pr(H_j|F_1)$, (7) simplifies to:

$$\phi_j = \sum_{j=1}^{K} \frac{n_j}{N_j} \frac{M_j}{M} = \sum_{j=1}^{K} \frac{n_j}{N_j} \frac{N_j}{\sum_{i=1}^{K} N_i} = \sum_{j=1}^{K} \frac{n_j}{N_j} \frac{N_j}{\sum_{i=1}^{K} N_i},$$

(8)

which is the observed proportion of subjects with suicide attempts over $T$ in facility $F_1$.

Summarized in the Theorem below are the asymptotic properties of these estimates.

**Theorem 1** Let $p_j = \Pr(y_{i1}=1|F_1, H_j)$ and $p_j = \Pr(y_{i1}=1|F_1, H_j)$ defined in (5).

a). Under Assumption A1, as $N_j \to \infty$, we have

$$\sqrt{N_j} \left( p_j - p \right) \to N(0, \sigma_j^2),$$

as $N_j \to \infty$ for $j=1,2,...,K$

b). Let $w_j = \Pr(H_j)$. Under Assumptions A1, the asymptotic distribution of $\phi_j$ in (6) is given by

$$\phi_j - \phi \to N(0, \sigma_j^2),$$

where $\sigma_j^2 = \sum_{j=1}^{K} w_j \frac{1}{N_j} \frac{N_j}{N_i} \frac{n_i}{N_i} \frac{N_i-n_i}{N_i}$, and can be estimated by $\sigma_j^2 = \sum_{j=1}^{K} w_j \frac{1}{N_i} \frac{N_i}{N_i} \frac{n_i}{N_i} \frac{N_i-n_i}{N_i}$. 


c. Let \(w_j = \Pr(H_j)\). Under Assumptions A1 and A2, the asymptotic distribution of \(\varphi_j\) in (7) is given by
\[
\varphi_j - \varphi \rightarrow N(0, \sigma_j^2), \text{ as } M \rightarrow \infty \text{ and } N_j \rightarrow \infty \text{ for } j = 1, 2, \ldots, K
\]
where \(\frac{M_j}{M}\) and can be estimated by replacing \(\frac{M_j}{M}\) with \(\frac{M_j}{M}\).

Note: case b) may be viewed as the special case of c) when \(M\) goes to infinite.

We can use the above theorem to compare \(F_1\) and \(F_2\) for the third scenario discussed in Section 2. For example, suppose that
\[
\Pr(H_1) = 50\%, \Pr(H_2) = 30\%, \Pr(H_3) = 20\%.
\]
Then the value of the metric for each facility is given by:
\[
\varphi_1 = 50\% \times 1\% + 30\% \times 2\% + 20\% \times 5\% = 0.021,
\]
\[
\varphi_2 = 50\% \times 1.5\% + 30\% \times 3\% + 20\% \times 1\% = 0.0185.
\]
Since \(\varphi_2 < \varphi_1\), \(F_2\) outperforms \(F_1\).

If some subjects are censored before \(T\), then (5) would underestimate \(\Pr(y_{i1} = 1 | F_1, H_j)\).

3.2.2 Metric with Dynamic Risk Membership

In the proceeding section, we assume that everyone is observed over the same period \(T\). Such an assumption is unrealistic in practice, as subjects may drop out at different times. In addition, some subjects may experience multiple attempts during the study period. In this section, we extend the approach above to allow for time-varying censoring and multiple events.

When considering multiple events, we need to distinguish the observation (or follow-up or study) time for each subject from the time window of interest for the distribution of suicide attempts. We denote the former by \(S\) and continue to denote the latter by \(T\). For subjects with multiple suicide attempts, each event defines a \(T\), with the beginning of \(T\) set as the time when the event occurs. In addition, we introduce effective sample size (ESS), which is an adjusted sample size based on the fraction of \(T\) for which the subject is observed without any event.

For example, consider a study with a follow-up period of 24 months and suppose we are interested in estimating the risk of suicide attempts within a 12-month period starting from baseline \((S = 24, T = 12)\). Suppose a subject, starting with group \(H_1\), i.e., with 1 attempt prior to baseline \((s = 0)\), is followed up for 24 months and has two attempts: one at \(s = 2\) months and the other at \(s = 16\) months. Since the subject has 1 attempt prior to the study, we can estimate \(\Pr(\text{suicideattempt}|H_1)\). Starting from baseline \((s = 0)\), the subject commits one attempt at \(s = 2\) (within the 12-month time window of interest), so the subject contributes one event for estimating \(\Pr(\text{suicideattempt}|H_1)\) and has ESS=1. At the occurrence of the first observed attempt at \(s = 2\), this subject has two attempts and his membership is updated to \(H_2\). At the same time, a new beginning of the time window \(T\) for the future suicide risk is set to be at \(s = 2\). Within the 12-month time window starting at \(s = 2\), the subject has no attempt (between \(s = 2\) and \(s = 14\)), so the subject does not contribute any event in estimating \(\Pr(\text{suicideattempt}|H_2)\), even though
this subject has ESS=1 as s/he is indeed followed up for a full length of time $T$. At $s=16$, the subject has three events and his membership is then updated to $H_3$. Starting at $s=16$, since s/he does not have any additional attempt during the next 8-months (from $s=16$ to $s=24$), s/he does not contribute any event for estimating $\Pr(\text{suicide attempt}|H_3)$. However, his/her EES = $\frac{8}{12} = \frac{2}{3}$ when estimating $\Pr(\text{suicide attempt}|H_3)$.

Let $N_j$ be the ESS with $j$ suicide attempts at the beginning of the study, i.e., $s=0$, let $N_{jk}$ be the ESS in group $H_j$ who have $k^{th}$ suicide attempt during the study period $S$, and let $m_{jk}$ be the number of subjects in group $H_j$ who have $k^{th}$ attempt within the time $T$ (the origin of $T$, i.e., $t=0$, is at the time when the $(k-1)^{th}$ attempt occurs). Table 2 summarizes the information regarding suicide attempts/re-attempts.

Table 2: The summarized information for estimating the metrics when the membership is dynamically changing

<table>
<thead>
<tr>
<th>$H_j$</th>
<th>ESS at baseline</th>
<th>$1^{st}$ attempt within $T$</th>
<th>ESS at $1^{st}$ attempt within $T$</th>
<th>$2^{nd}$ attempt</th>
<th>...</th>
<th>ESS at $(L-1)^{th}$ attempt</th>
<th>$L^{th}$ attempt within $T$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$N_{00}$</td>
<td>$m_{01}$</td>
<td>$N_{01}$</td>
<td>$m_{02}$</td>
<td>...</td>
<td>$N_{0(L-1)}$</td>
<td>$m_{1L}$</td>
</tr>
<tr>
<td>1</td>
<td>$N_{10}$</td>
<td>$m_{11}$</td>
<td>$N_{11}$</td>
<td>$m_{12}$</td>
<td>...</td>
<td>$N_{1(L-1)}$</td>
<td>$m_{1L}$</td>
</tr>
<tr>
<td>2</td>
<td>$N_{20}$</td>
<td>$m_{21}$</td>
<td>$N_{21}$</td>
<td>$m_{22}$</td>
<td>...</td>
<td>$N_{2(L-1)}$</td>
<td>$m_{2L}$</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>$K-1$</td>
<td>$N_{(K-1)0}$</td>
<td>$m_{(K-1)1}$</td>
<td>$N_{(K-1)1}$</td>
<td>$m_{(K-1)2}$</td>
<td>...</td>
<td>$N_{(K-1)(L-1)}$</td>
<td>$m_{(K-1)L}$</td>
</tr>
<tr>
<td>$\geq K$</td>
<td>$N_{K0}$</td>
<td>$m_{K1}$</td>
<td>$N_{K1}$</td>
<td>$m_{K2}$</td>
<td>...</td>
<td>$n_{K(L-1)}$</td>
<td>$n_{KL}$</td>
</tr>
</tbody>
</table>

For a subject in the $j^{th}$ group at $s=0$, his/her risk for suicide attempt is elevated after each event and his/her group membership changes to $(j+k)^{th}$ group after the $k^{th}$ suicide attempt. We use the information to estimate $\Pr(y_{i1}|F_{1}, H_j)$, which increases the sample size and improves the efficiency of the estimates as all events are used in estimating the conditional probabilities.

Consider facility $F_1$, for a given $j$, we first estimate $\Pr(y_{i1}|F_1, H_j)$. Let $N_j = \sum_{i=0}^{j} N_{j-i,i}$ and $\eta_j = \sum_{i=0}^{j-1} m_{i,(j+1-i)}$. Among the $N_j$ at-risk subjects, $n_j$ of them commit suicide attempt within $T$. Thus, we have:

$$p_j = \Pr(y_{i1} = 1|F_1, H_j) = \frac{n_j}{N_j} = \frac{\sum_{i=0}^{j-1} m_{i,(j+1-i)}}{\sum_{i=0}^{j} N_{j-i,i}}, \quad \text{and} \quad \Pr(H_j) = \frac{M_j}{M}, \quad (9)$$

We define the metric as:
\[ \varphi_d = \sum_{j=0}^{K} \frac{\sum_{i=0}^{j-1} m_{i,j+1-i} \Pr(H_j)}{\sum_{i=0}^{j} N_{j-i,i}} , \quad \text{or} \quad \varphi_d = \sum_{j=0}^{K} \frac{\sum_{i=0}^{j-1} m_{i,j+1-i} M_j}{\sum_{i=0}^{j} N_{j-i,i}} \] (10)

For valid inference, we assume:

**Assumption (B1).** Random Censoring: the follow-up time is independent of the outcome. This assumption eliminates the type of selection bias caused by the selection of subjects for follow-up based on their risk levels for suicide attempt.

**Assumption (B2).** Random Sampling: a simple random sample of \( M \) subjects is selected from the study population, with \( M_j \) representing such a sample in the \( j^{th} \) subpopulation.

Summarized in Theorem 2 are the asymptotic distributions of estimated conditional probability \( p_j \) and the metric \( \varphi_d \). The distributions have similar forms as their counterparts in Theorem 1, except that the sample size is replaced by effective sample sizes.

**Theorem 2 a).** Under the assumptions B1, the asymptotic distribution of \( p_j \) in (9) is given by

\[ \sqrt{N_j} \left[ p_j - \hat{p}_j \right] \rightarrow \mathcal{N}(0, \sigma^2_{p_j}) \]

**b).** Under the assumptions B1, the asymptotic distribution of \( \varphi_d \) (10) is given by

Error!

where \( \tau_1^2 = \sum_{j=1}^{K} \frac{1}{N_j} w_j^2 \left( 1 - p_j \right) \) and can be estimated by \( \hat{\tau}_1^2 = \sum_{j=1}^{K} \frac{1}{N_j} w_j^2 \left( 1 - \hat{p}_j \right) \).

**c).** Under the assumptions B1-B2, the asymptotic distribution of \( \varphi_d \) (10) is given by

Error!

where \( \text{Error!} \) and \( \text{Error!} \) can be estimated by replacing \( \text{Error!} \) and \( \text{Error!} \) with \( \text{Error!} \) and \( w_j \) defined in (9).

Note: case b) may be viewed as the special case of c) when \( M \) goes to infinite.

The above approach not only provide estimates of risk for suicide attempt \( (p_j) \), but also a metric \( (\varphi_d) \) to compare such risks at the population level between different facilities. For example, if \( \varphi_1 \) and \( \varphi_2 \) denote the risk of suicide attempt for facility \( F_1 \) and \( F_2 \), and \( \varphi_1 \) and \( \varphi_2 \) are the corresponding estimates based on (6) or (7) or (10), then

\[ (\varphi_1 - \varphi_2) - (p_1 - p_2) \rightarrow \mathcal{N}(0, \omega^2) \], where \( \omega^2 = \sigma^2_{\varphi_1} + \sigma^2_{\varphi_2} \), \( (\varphi_1 - \varphi_1) \rightarrow \mathcal{N}(0, \omega^2_{\varphi_1}) \) and \( (\varphi_2 - \varphi_2) \rightarrow \mathcal{N}(0, \omega^2_{\varphi_2}) \).

### 4 Simulation Studies

Simulation studies are conducted to examine the performance of our proposed metrics. We consider a 6-level history of suicide attempts defined by 0, 1, 2, 3, 4, and 5 or more attempts at baseline of the study \( (s=0) \), denoted by \( H_0, H_1, H_2, H_3, H_4 \) and \( H_5 \). For notational brevity, we assume that prior attempts is the only risk factor for suicide attempt and apply the metric to assess the risk of suicide attempt over half a year. If there are more than one (categorical) risk
factor in real studies, one can combine such variables to create a new categorical variable by grouping different levels from different factors. We also assume subjects within the same risk group have the same likelihood of committing suicide attempt and the time to commit a new suicide attempt in the study period follows an exponential distribution.

In the simulation study, we assume the conditional probabilities of suicide attempt are 0.08, 0.12, 0.16, 0.23, 0.27, and 0.28 for the six risk groups $H_j (0 \leq j \leq 5)$. We first assume that the distribution of $H_j$ is known and given by: 0.50, 0.25, 0.12, 0.07, 0.04 and 0.02. We then relax this assumption and simulate the risk levels from a multinomial distribution with these cell probabilities. The true metric is $\varphi = 0.1217$ under the assumed known distribution of the risk group and conditional probabilities of suicide attempt.

We consider sample sizes 100, 200, 500, 1,000 and 10,000. The Monte Carlo size is 1,000. We assume one unit time for $T$, the time window of interest. The unit time could be 6-month, 1-year etc. To allow for varying follow-up times, we assume that the follow-up time $S$ follows the distribution $S = 1 + (u_i)^5$ and $u_i \sim Uniform(0, 1)$, i.e., the follow-up time is uniformly distributed between 1 and 6 units time. Each subject is simulated a subsequent suicide attempt based on the exponential distribution for the time to suicide attempt until the end of the follow-up time.

We estimated the metric for both the fixed and dynamically-changing risk group models discussed above.

Shown in Table 3 are the true metric, the mean of estimates of the metrics (Est.), the empirical standard errors (SE_E) and the mean of the asymptotic standard errors (SE_A). The point estimates of the metric for both models were close to the true metric, but the one based on the dynamically-changing risk group model was a bit closer. As sample size increased, estimates from both models improved and became closer to their respective true values. The estimated asymptotic standard errors were also very close to their empirical counterparts for both models, but the standard errors for the dynamically-changing risk group model were smaller, reflecting improved efficiency.

Table 3: The mean of the estimated metrics, the empirical standard errors and the asymptotic standard errors when the population distribution of the risk factor is known. The true metric is 0.127.

| Sample Size | Fixed Membership | | | Dynamic Change Membership | | |
|-------------|------------------|------------------|------------------|------------------|------------------|
|             | Est.  | SE_E  | SE_A  | Est.  | SE_E  | SE_A  |
| 100         | 0.1222 | 0.0316 | 0.0306 | 0.1213 | 0.0264 | 0.0266 |
| 200         | 0.1225 | 0.0226 | 0.0222 | 0.1215 | 0.0189 | 0.0190 |
| 500         | 0.1219 | 0.0143 | 0.0143 | 0.1216 | 0.0121 | 0.0121 |
| 1000        | 0.1220 | 0.0101 | 0.0101 | 0.1217 | 0.0086 | 0.0086 |
| 5000        | 0.1218 | 0.0047 | 0.0046 | 0.1216 | 0.0038 | 0.0038 |
| 10000       | 0.1218 | 0.0032 | 0.0032 | 0.1215 | 0.0027 | 0.0027 |

In the case with unknown distribution of risk levels, the probability of each risk level was estimated by the respective sample proportion and then substituted in place of the corresponding probability in the metric. Shown in 4 are the various estimates resulting from this approach. The results are very similar to their counterparts presented in Table 1, indicating that using the
estimated population distribution of the risk factor seems to have little impact on the estimated metric. As in Table 1, the dynamically-changing risk group model again yielded more accurate estimates.

Table 4: The mean of the estimated metrics, the empirical standard errors and the asymptotic standard errors when the population distribution of the risk factor is estimated. The true metric is 0.127.

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Fixed Membership</th>
<th>Dynamic Change Membership</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Est. SE_E SE_A</td>
<td>Est. SE_E SE_A</td>
</tr>
<tr>
<td>100</td>
<td>0.1222 0.0323 0.0307</td>
<td>0.1215 0.0269 0.0267</td>
</tr>
<tr>
<td>200</td>
<td>0.1225 0.0227 0.0223</td>
<td>0.1217 0.0189 0.0191</td>
</tr>
<tr>
<td>500</td>
<td>0.1211 0.0150 0.0143</td>
<td>0.1217 0.0125 0.0121</td>
</tr>
<tr>
<td>1000</td>
<td>0.1220 0.0013 0.0101</td>
<td>0.1216 0.0084 0.0086</td>
</tr>
<tr>
<td>5000</td>
<td>0.1217 0.0047 0.0046</td>
<td>0.1216 0.0039 0.0038</td>
</tr>
<tr>
<td>10000</td>
<td>0.1218 0.0031 0.0032</td>
<td>0.1216 0.0026 0.0027</td>
</tr>
</tbody>
</table>

5 A Case Study

Data for this case study is from the Suicide Prevention Application Network (SPAN) database maintained by the Office of the Mental Health Services Director, Suicide Prevention and the VISN 2 Center of Excellence for Suicide Prevention in the Department of Veterans Affairs (VA). Suicide Prevention Coordinators located at each VA medical center (VAMC) are responsible for completing a SPAN record for every known suicidal event by a Veteran. The SPAN database includes the category of suicidal event according to the Self-Directed Violence Classification System (brenner11), which uses standard definitions for suicidal ideation, suicide attempt and suicide. The SPAN also captures individual identifiers, dates and locations of Veteran suicidal events, allowing measurement of re-attempts over time and for specific geographic areas. For the analyses presented here, baseline events are comprised of Veterans’ first suicidal events between October 1, 2012 and September 30, 2013. Thus, \(S=12\) and \(T=6\) months.

For this example, we excluded fatal events and assumed that all Veterans had complete follow-up because the data were extracted several months after the end of follow-up and thus number of unreported cases due to delay in reporting was negligible. Since death from suicide remains a comparatively rare event and can occur without any indication of care related to suicidal thoughts or attempts, we removed them before the analysis. The proposed approach is applied to each of the 21 VISNs as well as the pooled national sample. In addition to asymptotic results, we also apply bootstrap to assess accuracy of asymptotic results.

Table 5: The estimated risk for suicide re-attempts across all the VISNs
As shown in Table 5, the overall risk for suicide re-attempts in this sample was 0.122, with a standard error of 0.004. The 95% CI by the bootstrap was 0.1145-0.1290, almost identical to the 95% CI, (0.1146-0.1291), by the asymptotic method. Table 3 summarizes the results by VISN in order of increasing re-attempt rate. The proposed metric seems to differentiate the VISNs, with a range from 0.062 to 0.196. The 95% CIs overlap with the national 95% CI, but the CIs for VISNs A and B are below that of the national CI, while the CI for VISN U is almost outside its national counterpart.

### Table 3: Geographic Area Re-attempt Metric and Confidence Intervals

<table>
<thead>
<tr>
<th>Geographic Area</th>
<th>Re-event Metric</th>
<th>Asymptotic SE</th>
<th>95% CI</th>
<th>Bootstrap 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>National VISN</td>
<td>0.122</td>
<td>0.004</td>
<td>0.1146-0.1291</td>
<td>0.1145-0.1290</td>
</tr>
<tr>
<td>A</td>
<td>0.062</td>
<td>0.016</td>
<td>0.031-0.092</td>
<td>0.028-0.082</td>
</tr>
<tr>
<td>B</td>
<td>0.075</td>
<td>0.017</td>
<td>0.042-0.108</td>
<td>0.044-0.101</td>
</tr>
<tr>
<td>C</td>
<td>0.082</td>
<td>0.023</td>
<td>0.037-0.127</td>
<td>0.042-0.129</td>
</tr>
<tr>
<td>D</td>
<td>0.089</td>
<td>0.015</td>
<td>0.058-0.119</td>
<td>0.058-0.121</td>
</tr>
<tr>
<td>E</td>
<td>0.089</td>
<td>0.012</td>
<td>0.065-0.114</td>
<td>0.048-0.095</td>
</tr>
<tr>
<td>F</td>
<td>0.091</td>
<td>0.012</td>
<td>0.067-0.114</td>
<td>0.064-0.109</td>
</tr>
<tr>
<td>G</td>
<td>0.092</td>
<td>0.019</td>
<td>0.055-0.129</td>
<td>0.047-0.110</td>
</tr>
<tr>
<td>H</td>
<td>0.096</td>
<td>0.021</td>
<td>0.055-0.136</td>
<td>0.039-0.117</td>
</tr>
<tr>
<td>I</td>
<td>0.116</td>
<td>0.015</td>
<td>0.086-0.145</td>
<td>0.080-0.134</td>
</tr>
<tr>
<td>J</td>
<td>0.121</td>
<td>0.016</td>
<td>0.089-0.153</td>
<td>0.082-0.139</td>
</tr>
<tr>
<td>K</td>
<td>0.122</td>
<td>0.020</td>
<td>0.083-0.160</td>
<td>0.080-0.157</td>
</tr>
<tr>
<td>L</td>
<td>0.123</td>
<td>0.020</td>
<td>0.084-0.163</td>
<td>0.079-0.160</td>
</tr>
<tr>
<td>M</td>
<td>0.125</td>
<td>0.018</td>
<td>0.091-0.160</td>
<td>0.084-0.151</td>
</tr>
<tr>
<td>N</td>
<td>0.130</td>
<td>0.011</td>
<td>0.109-0.151</td>
<td>0.105-0.147</td>
</tr>
<tr>
<td>O</td>
<td>0.132</td>
<td>0.028</td>
<td>0.077-0.187</td>
<td>0.078-0.188</td>
</tr>
<tr>
<td>P</td>
<td>0.133</td>
<td>0.019</td>
<td>0.096-0.170</td>
<td>0.096-0.173</td>
</tr>
<tr>
<td>Q</td>
<td>0.137</td>
<td>0.021</td>
<td>0.096-0.178</td>
<td>0.085-0.166</td>
</tr>
<tr>
<td>R</td>
<td>0.139</td>
<td>0.015</td>
<td>0.111-0.168</td>
<td>0.103-0.163</td>
</tr>
<tr>
<td>S</td>
<td>0.152</td>
<td>0.017</td>
<td>0.119-0.185</td>
<td>0.118-0.186</td>
</tr>
<tr>
<td>T</td>
<td>0.186</td>
<td>0.016</td>
<td>0.155-0.216</td>
<td>0.151-0.213</td>
</tr>
<tr>
<td>U</td>
<td>0.196</td>
<td>0.036</td>
<td>0.126-0.267</td>
<td>0.124-0.255</td>
</tr>
</tbody>
</table>

### 6 Discussion

We developed a metric to assess the risk for suicide attempts based on the causal inference framework to address heterogeneity in the study population, such as history of suicide attempt. The approach not only provides a meaningful and informative measure of risk of suicide attempt for a heterogeneous study population, but also allows for inference for comparing such risks between different subgroups, such as the VHA facilities within the current context. By accounting for history of suicide events, this approach allows one to include all events to improve power when comparing different groups for risks of suicide attempt. The approach performs well as evidenced by both real and simulation studies.

When accounting for prior events, we do not distinguish reported events from observed suicide attempts, especially those events that are reported prior to the study. This assumption may not hold, if there is recall bias, or if the reported suicide attempts occur a long time before the study. The proposed approach may yield estimates that are biased or may not have a clear interpretation, if this assumption fails.
Although motivated by and developed for suicide attempt, the proposed approach may be applied to any episodic disease in which risks for the disease changes as event occur. Also, we focused on one risk factor in this paper. If there are multiple categorical risk variables, one may combine the different risk factors to create a single variable by grouping different levels of the different variables. If number of risk factors is large, the approach may be difficult to apply and/or may not perform well. Research is underway to extend the current work to continuous risk factors.

7 Appendix: technical details

Proof of Theorem 1. (a). Let \( p_j = \Pr(y_i | F_1, H_j) \), and \( p_j = \Pr(y_i | F_1, H_j) = \frac{n_j}{N_j} \). Since \( p_j \) is just a standard estimation of simple proportion, by the Central Limit Theorem, Theorem 1(a) can be proved.

(b). Let \( w_j = \Pr(H_j) \). Since \( \phi_f = \sum_{j=1}^{K} p_j \cdot \Pr(H_j) \), a summation of the proportions for independent samples, it follows that

\[
E(\phi_f) = E \left[ \sum_{j=1}^{K} w_j p_j \right] = \sum_{j=1}^{K} w_j E[p_j] = \sum_{j=1}^{K} w_j p_j = \phi_f,
\]

and

\[
\sigma_1^2 = \text{Var}(\phi_f) = \text{Var} \left[ \sum_{j=1}^{K} w_j p_j \right] = \sum_{j=1}^{K} w_j^2 \text{Var}[p_j] = \sum_{j=1}^{K} w_j^2 \frac{1}{N_j} \frac{N-j}{N_i} (1-p_j).
\]

(c). Since \( p_j \) can be estimated by \( p_j = \Pr(y_i | F_1, H_j) = \frac{n_j}{N_j} \), the \( \sigma_1^2 \) can be estimated by

\[
\sigma_1^2 = \sum_{j=1}^{K} w_j^2 \frac{1}{N_i} \frac{N-n_j}{N_i} = \frac{M_j}{M} \text{Var}(\phi_f) = \frac{M_j}{M} \sum_{j=1}^{K} w_j^2 \frac{1}{N_j} \frac{N-j}{N_i} (1-p_j).
\]

Hence the asymptotic variance in Theorem 1(b) has been proved.

Let \( w_j = \Pr(H_j) = \frac{M_j}{M} \). If \( \Pr(H_j) \) need to be estimated, the asymptotic distribution can be similarly derived. Specifically, since

\[
E(\phi_f) = E \left[ \sum_{j=1}^{K} w_j p_j \right],
\]
and two independent procedures for estimating \( w_j \) and \( p_j \) are independent, we have

\[
E(\varphi_j) = E\left[ \sum_{j=1}^{K} w_j p_j \right]
= \sum_{j=1}^{K} E[w_j] E[p_j]
= \sum_{j=1}^{K} w_j p_j = \phi.
\]

The variance of \( \varphi_j \) can be derived as follows:

\[
\sigma_1^2 = \text{Var}(\varphi_j) = \text{Var}\left[ \sum_{j=1}^{K} w_j p_j \right]
= E\left[ \left( \sum_{j=1}^{K} w_j p_j \right)^2 \right] - \left[ E\left( \sum_{j=1}^{K} w_j p_j \right) \right]^2
\]

\[
= E\left[ \sum_{j=1}^{K} w_j^2 p_j^2 + \sum_{j=1}^{K} w_j p_j \sum_{i \neq j} w_i p_i \right] - \left[ E\left( \sum_{j=1}^{K} w_j p_j \right) \right]^2
\]

\[
= E\left[ \sum_{j=1}^{K} \left( w_j^2 p_j^2 - w_j p_j \sum_{i \neq j} w_i p_i \right) + \sum_{j=1}^{K} E\left[ \sum_{i \neq j} w_j w_i p_j p_i \right] \right] - \left[ E\left( \sum_{j=1}^{K} w_j p_j \right) \right]^2
\]

Since

\[
E\left[ \sum_{j=1}^{K} \left( w_j^2 p_j^2 - w_j p_j \sum_{i \neq j} w_i p_i \right) \right]
= E\left[ \sum_{j=1}^{K} \left( w_j^2 p_j^2 + w_j^2 p_j^2 - w_j^2 p_j^2 - w_j^2 p_j^2 \right) \right]
\]

\[
= \sum_{j=1}^{K} \left[ E\left( w_j^2 p_j^2 - w_j^2 p_j^2 \right) + E\left( w_j^2 p_j^2 - w_j^2 p_j^2 \right) \right]
\]

Error!

\[
= \sum_{j=1}^{K} p_j^2 + \frac{1}{N_j} \left( 1-p_j \right) \frac{1}{M} w_j (1-w_j) + \sum_{j=1}^{K} \frac{1}{N_j} p_j^2 \frac{1}{M} (1-p_j)
\]

\[
= \sum_{j=1}^{K} p_j^2 + \frac{1}{N_j} \left( 1-p_j \right) \frac{1}{M} w_j (1-w_j) + \sum_{j=1}^{K} \frac{1}{N_j} p_j^2 \frac{1}{M} (1-p_j)
\]

and
\[
\sum_{j=1}^{K} E \left[ \sum_{i \neq j} \left( w_i p_j w_j - w_j p_i w_j \right) \right]
\]

\[
= \sum_{j=1}^{K} \sum_{i \neq j} \left[ w_i p_j w_j - w_j p_i w_j + w_i p_j w_j - w_j p_i w_j \right]
\]

\[
= \sum_{j=1}^{K} \sum_{i \neq j} \left\{ E \left[ w_i w_j (p_j - p_i) \right] + E \left[ p_j (w_i w_j - w_j w_i) \right] \right\}
\]

\[
= \sum_{j=1}^{K} \sum_{i \neq j} \frac{1}{M} p_j w_i w_j
\]

we have

\[
\sigma_i^2 = \text{Var} \left( \phi_j \right) = \text{Var} \left[ \sum_{j=1}^{K} w_j \right]
\]

\[
= \sum_{j=1}^{K} p_j^2 \frac{1}{M} w_j (1-p_j) + \sum_{j=1}^{K} w_j^2 \frac{1}{N_j} p_j \left( 1-p_j \right)
\]

\[
+ \sum_{j=1}^{K} \frac{1}{M} \frac{1}{N_j} p_j \left( 1-p_j \right) w_i (1-w_j) - \sum_{j=1}^{K} \sum_{i \neq j} \frac{1}{M} p_j w_i w_j.
\]

Hence the asymptotic variance in Theorem 1(c) has been proved.

As \( M \to \infty \), \( \text{Var} \left( \sum_{j=1}^{K} w_j p_j \right) \to \sum_{j=1}^{K} w_j^2 \frac{1}{N_j} p_j \left( 1-p_j \right) \), and the variance is reduced to that for \( \phi_j \) in part b). On the other hand, As \( N_i \to \infty \), i.e., if the rates for all the levels are known, \( \text{Var} \left( \sum_{j=1}^{K} w_j p_j \right) \to \sum_{j=1}^{K} p_j^2 \frac{1}{M} w_j (1-w_j) = \text{Var} \left( \sum_{j=1}^{K} p_j w_j \right) \). Note that in the proof we used the independence between \( p_j \) and \( w_j \). In practice, it is common that we have a simple random sample from the whole population and \( w_j \) is estimated from the sample, and use the subsample of subjects in the facility to estimate \( p_j \).

**Proof of Theorem 2.** The derivative of the Theorem 2 is similar to that of Theorem 1. Based on the exponential assumption and random assumptions, we need to establish the unbiasedness. All the others are straightforward.

**References**

