Published in final edited form as:

Stat Methods Med Res. 2019 February ; 28(2): 503–514. doi:10.1177/0962280217729843.

Population metrics for suicide events: A causal inference approach

Hua He¹, Naiji Lu², Brady Stephens³, Yinglin Xia⁴, Robert M Bossarte⁵, Cathleen P Kane³, Wan Tang⁶, and Xin M Tu⁷

¹Department of Epidemiology, Tulane University School of Public Health and Tropical Medicine, New Orleans, LA, USA

²School of Management, Harbin Institute of Technology, Harbin, P. R. China

³Center of Excellence at Canandaigua Veterans Administration Medical Center, Canandaigua, NY, USA

⁴Department of Medicine, University of Illinois at Chicago, Chicago, IL, USA

⁵Injury Control Research Center and Department of Behavioral Medicine and Psychiatry, West Virginia University, Morgantown, WV, USA

⁶Department of Global Biostatistics and Data Science, Tulane University School of Public Health and Tropical Medicine, New Orleans, LA, USA

⁷Division of Biostatistics and Bioinformatics, Department of Family Medicine and Public Health, University of California, San Diego, CA, USA

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 events 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 events within each of the risk levels, number of prior attempts, and then taking a weighted sum of the conditional probabilities. The metrics provide unbiased estimates of the risk of suicide events. 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

Reprints and permissions: sagepub.co.uk/journalsPermissions.nav

Corresponding author: Hua He, Department of Epidemiology, Tulane University School of Public Health and Tropical Medicine, 1440 Canal Street, Suite 2031, New Orleans, LA 70112-2703, USA. hhe2@tulane.edu.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

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 practices, more objective measures need to be created, particularly within a large healthcare system for low incidence events and large observable heterogeneity across reporting units. Among the most important challenges facing current public health strategies is the lack of 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, other suicide events (attempts and serious ideation resulting in hospitalization), and associated other cause mortality, and improve the overall quality of life of veterans.¹ However, there are few quantitative indicators with the degree of granularity 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.² 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 the 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; the greater the history of prior suicidal events the more likely of a future suicide event over a given period of time.² As a result, the risk level of a subject is dynamically changing with the intervention outcome and the interwovenness of the predictor and outcome calls for more carefully designed metrics to take into account such dynamic relationships between risk and suicide events.

In this paper, we develop a metric to address the aforementioned difficulties. The 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 events to illustrate the statistical issues and motivate our development. For simplicity, we consider only one risk factor, history of suicide event, and assume a categorical outcome with three levels: no suicide event, one suicide event, and two or more suicide events.

Consider two facilities F_1 and F_2 with the same number of subjects 3200. We are interested in comparing effect of some type of prevention intervention between the facilities. Note that in reality there are typically multiple ongoing suicide prevention programs at VHA facilities

at a given point in time. Thus “intervention” does not refer to a particular intervention, but rather strategies that incorporate national and facility-specific initiatives to reduce suicides. We present three different scenarios, depicting different distributions of number (proportion) of suicide event.

Shown in Table 1 are number of subjects (“Subj.”), number of events (“S.A.”), and number of re-event rates (“Rate”) of subjects who have suicide events in a time window such as six months broken down by number of prior suicide attempts (“History”). We assume that the same distribution of subjects with prior attempts and the same relationship between prior events and subjects with suicide events across all scenarios for F_1 . For F_2 , while the distribution of subjects with prior events remains the same, the relationship between prior events and subjects with suicide events varies across the scenarios.

In the first scenario, rates of suicide events are the same between the two facilities within each category of history of suicide event. The intervention performs equally well in both facilities. However, if we ignore history of event and simply compare the overall rate (total number of suicide events/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 events stratified by the history of suicide events. However as in the first scenario, F_2 again has a lower overall rate, although not significantly so, than F_1 (p value = 0.64) if the history of suicide event is ignored.

The seemingly “non-sensible” results in the two examples above are really the phenomenon of Simpson’s paradox, or selection bias in the lingo of causal inference.^{3–5} History of suicide event, which is highly correlated to the outcome, is not controlled for when performing comparisons. If we control for this confounder 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? This is the focus of the next section.

3 New metrics for comparing suicide attempts

We consider the problem of comparing two facilities in the third scenario as well as other issues and develop a general approach for program evaluation for episodic diseases such as suicide event by drawing upon the potential-outcome based causal inference paradigm.^{3,4,6,7} For simplicity, we focus on suicide attempt. We again consider two facilities, F_1 and F_2 , and assume a categorical risk factor with K levels, $H_1, H_2, \dots, H_K$, which can be history of suicide event alone or in combination with other variables such as age, gender, and medical conditions.

3.1 Potential outcome and binary outcome

For any randomly selected subject i , let y_{i1} (y_{i2}) denote the potential outcome of suicide event over a given period of time T , such as six months, if she/he receives the care within

facility F_1 (F_2). The potential outcome y_{i1} (y_{i2}) is a binary variable with value 1 indicating a suicide event and value 0 otherwise. Since the difference $\tau_i = y_{i1} - y_{i2}$ is based on the outcomes from the same individual, it must be the result of differential intervention effect between F_1 and F_2 . However, τ_i is not computable, since only one of y_{i1} and y_{i2} is observable; the one corresponding to the actual facility from which the individual receives the primary care. A large part of the causal inference literature centers around how to estimate the average, or population-level, treatment effect, $\tau = E(y_{i1} - y_{i2})$.^{3,5-9} In the absence of any selection bias such as randomly assigning subjects 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. $E(y_{i1}) = E(y_{i1}|i \in F_1)$ ($i = 1, 2$). Thus, $\tau = E(y_{i1} - y_{i2}) = E(y_{i1}) - E(y_{i2}) = E(y_{i1}|i \in F_1) - E(y_{i2}|i \in F_2)$, and we can estimate τ by estimating $E(y_{i1}|i \in F_1)$, the observed suicide events in facility F_1 (F_2). However, in the presence of selection bias such as non-random assignment of subjects to the two facilities, 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 τ by estimating the difference between $E(y_{i1}|i \in F_1)$ and $E(y_{i2}|i \in F_2)$ from observed data. For example, as illustrated in Section 2, none of the overall rates of suicide event can be used to compare the facilities in these examples, because history of suicide event, which is correlated with suicide event that occurs during the study, is not distributed evenly between the facilities.

To be able to estimate τ based on observed outcomes and information in the risk groups H_k , we assume, as in the literature, a strong ignorable assumption^{3-5,7}

$$(y_{i1}, y_{i2}) \perp (F_1, F_2) | H_k, \quad 1 \leq k \leq K \quad (3.1)$$

where $\perp$ denotes stochastic independence. The condition in (3.1) ensures that within each subgroup H_k , assignment of F_l is independent of prior suicide event. Thus, when restricted to H_k , there is no selection bias, implying

$$E(y_{il}|i \in H_k) = E(y_{il}|i \in H_k, i \in F_l), \quad 1 \leq l \leq 2, \quad 1 \leq k \leq K \quad (3.2)$$

The identity in (3.2) allows one to estimate $E(y_{il}|i \in H_k)$ using observed outcomes in each facility F_l .^{6,9} For example, if $K = 5$ and $H_1, H_2, \dots, H_5$ represent subjects with 0, 1, 2, 3, 4 prior events, we can estimate $E(y_{i1})$ by estimating $E(y_{i1}|i \in H_k, i \in F_1)$ within each H_k ($1 \leq k \leq 5$).

Based on the strongly ignorable condition,^{3,4,7} and by applying the rule of total probability, we can express $E(y_{i1})$ in terms of $E(y_{i1}|i \in H_k, i \in F_1)$ as

$$\phi_1 = E(y_{i1}) = \sum_{k=1}^K E(y_{i1}|i \in H_k) \Pr(i \in H_k) = \sum_{k=1}^K E(y_{i1}|i \in H_k, i \in F_1) \Pr(i \in H_k) \quad (3.3)$$

For the facility l , the metric ϕ_l defined in (3.3) is a weighted average of conditional probabilities of suicide events within each risk group H_k at facility F_l . The weights are calculated based on the distribution of H_k in the study population.

Since

$$E(y_{il}|i \in F_l) = \sum_{k=1}^K E(y_{il}|i \in H_k, i \in F_l) \Pr(i \in H_k|i \in F_l) \quad (3.4)$$

the unstratified rate $E(y_{il}|i \in F_l)$ differs from the proposed metric in (3.3) only in the weights, or the distribution of H_k , i.e. $\Pr(i \in H_k|i \in F_l)$ for the unstratified rate vs. $\Pr(i \in H_k)$ for the proposed metric. If the selection process is random, then $\Pr(i \in H_k) = \Pr(i \in H_k|i \in F_l)$ and (3.3) is identical to (3.4). Otherwise, $E(y_{il}|i \in F_l)$ is different from ϕ_l .

3.2 Inference

Inference about the proposed metric involves estimation of the probability of suicide event within each risk level, $E(y_{il}|i \in H_k, i \in F_l)$, and the weights, $\Pr(i \in H_k)$, from the study population. Since a subject may continue to attempt suicide during the study period, the risk changes over time. For simplicity, we first consider the case where each subject contributes at most one event in the calculation of the metric by defining the risk group H_k at baseline, the time of patient entry into the facility or a starting point that is appropriate for the research question considered. Although the metric based on the fixed risk membership defined at baseline still providing useful information for estimating re-event rates, such analysis does not utilize all available data, rendering it less efficient. A better alternative 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, even for this high-risk population and including all events allows one to obtain more efficient and reliable estimates.

We consider both approaches below and start with the relatively simpler one that defines H_k with history of suicide attempt at baseline.

3.2.1 Metric with fixed group membership—For a given facility F_l , let N_j denote the number of subjects in H_j and n_j be the number of subjects who have an reevent within a given time period T . Then $\Pr(y_{il} = 1|i \in F_l, i \in H_j)$ can be readily estimated by the observed proportion

$$\widehat{\Pr}(y_{il} = 1|i \in F_l, i \in H_j) = \frac{n_j}{N_j} \quad (3.5)$$

provided that the following assumptions are met.

Assumption (A1): No censoring: all subjects are followed up in the entire period T .

If the weights, $\Pr(i \in H_j)$, are known, substituting $\Pr(i \in H_j)$ and $\widehat{\Pr}(y_{i1} = 1 | i \in F_1, i \in H_j)$ into (3.3) yields an estimate of ϕ_1

$$\tilde{\phi}_1^f = \widehat{\Pr}(y_{i1} = 1) = \sum_{j=1}^K \widehat{\Pr}(y_{i1} = 1 | i \in H_j, i \in F_1) \Pr(i \in H_j) = \sum_{j=1}^K \frac{n_j}{N_j} \Pr(i \in H_j) \quad (3.6)$$

where the superscript f of ϕ_1^f indicates that the metric is estimated by fixing subjects' membership at baseline. Otherwise, we need to estimate $\Pr(i \in H_j)$. 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 H_j .

Under A2 we estimate $\Pr(i \in H_j)$ by $\widehat{\Pr}(i \in H_j) = \frac{M_j}{M}$. By substituting $\widehat{\Pr}(i \in H_j)$ in place of $\Pr(i \in H_j)$ (3.6), we obtain

$$\hat{\phi}_1^f = \widehat{\Pr}(y_{i1} = 1) = \sum_{j=1}^K \widehat{\Pr}(y_{i1} = 1 | i \in H_j, i \in F_1) \widehat{\Pr}(i \in H_j) = \sum_{j=1}^K \frac{n_j}{N_j} \frac{M_j}{M} \quad (3.7)$$

As a special case, if F_1 has a random sample of the study population, i.e., if there is no selection bias, then $\widehat{\Pr}(i \in H_j) = \widehat{\Pr}(i \in H_j | i \in F_1)$ and (3.7) simplifies to

$$\hat{\phi}_1^f = \sum_{j=1}^K \frac{n_j}{N_j} \frac{M_j}{M} = \sum_{j=1}^K \frac{n_j}{N_j} \frac{N_j}{\sum_{i=1}^K N_i} = \frac{\sum_{j=1}^K n_j}{\sum_{i=1}^K N_i} \quad (3.8)$$

which is just the observed proportion of subjects with suicide attempts over T in facility F_1 , i.e. no weighting adjustment is needed.

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

Theorem 1: Let $p_j = \Pr(y_{i1} = 1 | i \in F_1, i \in H_j)$ and $\hat{p}_j = \widehat{\Pr}(y_{i1} = 1 | i \in F_1, i \in H_j)$ defined in (3.5).

a. Under assumption A_1 , as $N_j \rightarrow \infty$, we have

$$\sqrt{N_j}[\hat{p}_j - p_j] \rightarrow N(0, p_j(1 - p_j))$$

- b. Let $w_j = \Pr(i \in H_j)$. Under assumptions A1, the asymptotic distribution of $\tilde{\phi}_1^f$ in (3.6) is given by

$$\tilde{\phi}_1^f - \phi_1 \rightarrow N(0, \sigma_1^2), \text{ as } N_j \rightarrow \infty \text{ for } j = 1, 2, \dots, K$$

where $\sigma_1^2 = \sum_{j=1}^K w_j^2 \frac{1}{N_j} p_j (1 - p_j)$, and can be estimated by

$$\hat{\sigma}_1^2 = \sum_{j=1}^K w_j^2 \frac{1}{N_j} \frac{n_i}{N_i} \frac{N_i - n_i}{N_i} \dots$$

- c. Let $\hat{w}_j = \widehat{\Pr}(i \in H_j)$. Under assumptions A1 and A2, the asymptotic distribution of $\hat{\phi}_1^f$ in (3.7) is given by

$$\hat{\phi}_1^f - \phi_1 \rightarrow N(0, \sigma_2^2), \text{ as } M \rightarrow \infty \text{ and } N_j \rightarrow \infty \text{ for } j = 1, 2, \dots, K$$

where

$$\sigma_2^2 = \sum_{j=1}^K p_j^2 \frac{1}{M} w_j (1 - w_j) + \sum_{j=1}^K w_j^2 \frac{1}{N_j} p_j (1 - p_j) + \sum_{j=1}^K \frac{1}{M} \frac{1}{N_j} p_j (1 - p_j) w_j (1 - w_j) - \sum_{j=1}^K \sum_{i \neq j} \frac{1}{M} p_i p_j w_i w_j$$

and can be estimated by replacing p_j and w_j with $\hat{p}_j = \frac{n_i}{N_i}$, $\hat{w}_j = \frac{M_j}{M}$.

Note. case (b) may be viewed as the special case of c) when M goes to infinite. The proof of Theorem 1 is provided in the Appendix 1.

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

$$\Pr(i \in H_1) = 50\%, \Pr(i \in H_2) = 30\%, \Pr(i \in H_3) = 20\%$$

Then the value of the metric for each facility is given by

$$\tilde{\phi}_1^f = 50\% \times 1\% + 30\% \times 2\% + 20\% \times 5\% = 0.021$$

$$\tilde{\phi}_2^f = 50\% \times 1.5\% + 30\% \times 3\% + 20\% \times 1\% = 0.0185$$

Since $\tilde{\phi}_2^f < \tilde{\phi}_1^f$, F_2 outperforms F_1 .

If some subjects are censored before T , then (3.5) would underestimate $\Pr(y_{i1} = 1 | i \in F_1, i \in H_j)$. We consider this along with more general situations with multiple suicide events next.

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 events 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 events. We denote the former by S and continue to denote the latter by T . For subjects with multiple suicide events, 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 event within a 12-month window starting from baseline ($S = 24$, $T = 12$). Suppose a subject, starting with group H_1 , i.e. with 1 event prior to baseline ($s = 0$), is followed up for 24 months and has two events: one at $s = 2$ and the other at $s = 16$ months. Since the subject has one event prior to the study, we can estimate $\Pr(\text{Suicide event}|H_1)$. Starting from baseline ($s = 0$), the subject commits one event at $s = 2$ (within the 12-month time window of interest), so the subject contributes one event for estimating $\Pr(\text{Suicide event}|H_1)$ and has $\text{ESS} = 1$. At the occurrence of the first observed event at $s = 2$, this subject has two events 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 this new 12-month time window starting at $s = 2$, the subject has no event (between $s = 2$ and $s = 14$), so the subject does not contribute any event in estimating $\Pr(\text{Suicide event}|H_2)$, even though the subject has $\text{ESS} = 1$ as she/he is 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 she/he does not have any additional event during the next eight-months (from $s = 16$ to $s = 24$), she/he does not contribute any event for estimating $\Pr(\text{Suicide event}|H_3)$. However, his/her $\text{EES} = \frac{8}{12} = \frac{2}{3}$ when estimating $\Pr(\text{Suicide event}|H_3)$.

Let N_j be the ESS with j suicide events 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 event 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 from the time when the $(k - 1)$ th event occurs. Table 2 summarizes the information regarding suicide events/re-events.

For a subject in the j th group at $s = 0$, his/her risk for suicide event is elevated after each event and his/her group membership changes to $(j + k)$ th group after the k th event. We use the information to estimate $\Pr(y_{1i}|i \in F_1, i \in H_{j+k})$, which increases the sample size and improves the efficiency of estimates.

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

$$\hat{p}_j = \widehat{\Pr}(y_{i1} = 1 | i \in F_1, i \in H_j) = \frac{n_j}{N_j} = \frac{\sum_{i=0}^{j-1} m_{i,(j+1-i)}}{\sum_{i=0}^j N_{j-i,i}}, \text{ and } \widehat{\Pr}(i \in H_j) = \frac{M_j}{M} \quad (3.9)$$

We define the metric as

$$\tilde{\phi}_1^d = \sum_{j=0}^K \frac{\sum_{i=0}^{j-1} m_{i,(j+1-i)}}{\sum_{i=0}^j N_{j-i,i}} \cdot \Pr(i \in H_j), \text{ or } \hat{\phi}_1^d = \sum_{j=0}^K \frac{\sum_{i=0}^{j-1} m_{i,(j+1-i)}}{\sum_{i=0}^j n_{j-i,i}} \frac{M_j}{M}$$

(3.10)

where the superscript d in $\tilde{\phi}_1^d$ indicates that the metric is estimated based on the dynamic risk membership model. Please note, if a patient commits multiple suicide attempts, this patient can be used multiple times. However the correlation between multiple observations on the patient can be ignored because conditional probabilities are used in (3.10).

For valid inference, we assume:

Assumption (B1): Random censoring: the follow-up time is independent of suicide attempt. This assumption eliminates the type of selection bias caused by the selection of subjects for follow-up based on their risk levels for suicide event.¹⁰

Assumption (B2): Random sampling: a 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 $\hat{p}_j$ and the metric $\tilde{\phi}_1^d$. The distributions have similar expressions their counterparts in Theorem 1, except that the sample size is replaced by ESSs.

Theorem 2

- a. Under the assumptions B1, the asymptotic distribution of $\hat{p}_j$ in (3.9) is given by

$$\sqrt{N_j}[\hat{p}_j - p_j] \rightarrow N(0, p_j(1 - p_j))$$

- b. Under the assumptions B1, the asymptotic distribution of $\tilde{\phi}_1^d$ (3.10) is given by

$$\tilde{\phi}_1^d - \phi_1 \rightarrow N(0, \tau_1^2), \text{ as } N_j \rightarrow \infty \text{ for } j = 1, 2, \dots, K$$

where $\tau_1^2 = \sum_{j=1}^K \frac{1}{N_j} w_j^2 p_j (1 - p_j)$ and can be estimated by

$$\hat{\tau}_1^2 = \sum_{j=1}^K \frac{1}{N_j} w_j^2 \hat{p}_j (1 - \hat{p}_j).$$

- c. Under the assumptions B1–B2, the asymptotic distribution of $\hat{\phi}_1^d$ (3.10) is given by

$$\hat{\phi}_1^d - \phi_1 \rightarrow N(0, \tau_2^2), \text{ as } M \rightarrow \infty \text{ and } N_j \rightarrow \infty \text{ for } j = 1, 2, \dots, K$$

where

$$\tau_2^2 = \sum_{j=1}^K p_j^2 \frac{1}{M} w_j (1 - w_j) + \sum_{j=1}^K w_j^2 \frac{1}{N_j} p_j (1 - p_j) + \sum_{j=1}^K \frac{1}{M} \frac{1}{N_j} p_j (1 - p_j) w_j (1 - w_j) - \sum_{j=1}^K \sum_{i \neq j} \frac{1}{M} p_i p_j w_i w_j$$

and τ_2^2 can be estimated by replacing p_j and w_j with $\hat{p}_j$ and $\hat{w}_j$ defined in (3.9).

Note. case (b) may be viewed as the special case of (c) when M goes to infinite. The proof of Theorem 1 is provided in the Appendix 1.

The above approach not only provides estimates of risk for suicide attempt ($\hat{p}_j$), but also a metric ($\hat{\phi}_1^d$) to compare such risks at the population level between different facilities. For example, if ϕ_1 and ϕ_2 denote the risk of suicide event for facility F_1 and F_2 , and $\hat{\phi}_1$ and $\hat{\phi}_2$ are the corresponding estimates based on (3.6) or (3.7) or (3.10), then $(\hat{\phi}_1 - \hat{\phi}_2) - (\phi_1 - \phi_2) \rightarrow N(0, \omega^2)$, where $\omega^2 = \omega_1^2 + \omega_2^2$, $(\hat{\phi}_1 - \phi_1) \rightarrow N(0, \omega_1^2)$, and $(\hat{\phi}_2 - \phi_2) \rightarrow N(0, \omega_2^2)$.

4 Simulation studies

Simulation studies are conducted to examine the performance of our proposed metrics. We consider a six-level history of suicide attempt defined by 0, 1, 2, 3, 4, and 5 or more events 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 events is the only risk factor for suicide attempt and apply the metric to assess the risk of such event 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 attempting suicide and the time to commit a new suicide event 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 $\phi = 0.1217$ under the assumed known distribution of the risk group and conditional probabilities of suicide attempt.

We consider sample sizes 100, 200, 500, 1000, and 10,000, with a Monte Carlo size 1000. To allow for varying follow-up times, we assume that the follow-up time S follows the distribution $S_j = 1 + 5u_j$ and $u_j \sim \text{Uniform}(0, 1)$, i.e. the follow-up time is uniformly distributed between 1 and 6 units time. We set T as one unit, i.e. $T = 1$. Repeated follow-up times to next suicide attempts are simulated independently from an exponential distribution based on the updated risk levels until the total follow-up time is reached, i.e. when sum of times between repeated suicide attempts (or baseline and first attempt), is greater than or equal to S_j . Every segment of such time interval corresponds to an event, except for the last one, which is between the last two attempts (if the last event occurs at S_j) or between the last event and the censoring time S_j (if the last event occurs prior to S_j).

Shown in Table 3 are the true metric, the mean of estimates of the metric (Est.), the empirical standard errors (SE_E) and the mean of the asymptotic standard errors (SE_A) for both the fixed and dynamically-changing risk group models. The point estimates of the metric for both models were close to the true metric, although the one based on the dynamically-changing risk group model was a bit closer. As the 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.

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 Table 4 are the various estimates resulting from this approach. The results were again very similar to their counterparts presented in Table 3, indicating that using the estimated population distribution of the risk factor seems to have little impact on the estimated metric. As in Table 3, the dynamically-changing risk group model again yielded more accurate estimates.

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 Veterans Integrated Service Network (VISN) 2 Center of Excellence for Suicide Prevention in the Department of Veterans Affairs (VA). Suicide Prevention Coordinators (SPCs) located at each VA medical center (VAMC) are responsible for completing a SPAN record for every known suicidal event by a Veteran. These events can be first reported through Suicide Behavior Reports by VHA clinicians, but can also be identified through other sources, such as through local community partners. Prior work suggests that SPAN does not necessarily capture all suicide events by veterans receiving health services within the VHA.

The SPAN database includes the category of suicidal event according to the Self-Directed Violence Classification System,^{1,11} which uses standard definitions for suicidal ideation, suicide attempt and suicide. The SPAN also captures individual identifiers, dates and locations (e.g. facilities) of Veteran suicidal events, allowing measurement of re-events over

time and for specific geographic areas. For the analyses presented here, baseline events are comprised of veterans' first suicidal events between 1 October 2012 and 30 September 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.^{1,12–14} Since death from suicide remains a comparatively rare event, we removed them before the analysis. The motivation for developing this re-event rate is to monitor the less rare non-fatal attempts over shorter time periods and for smaller units of analysis such as a VHA facility. Monitoring non-fatal attempts in this way provides more timely feedback about risk of suicide over a catchment area such as a VHA facility. Fatal events would be monitored in a parallel process using standard methods. 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 provide bootstrapped variance estimates to compare with the asymptotic results.

As shown in Table 5, the overall risk for suicide re-events 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 5 summarizes the results by VISN in order of increasing re-event 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.

6 Discussion

We developed an approach to assess risk for suicide attempt by drawing upon the causal inference paradigm to address heterogeneity in the study population, such as history of such events. The approach not only provides a meaningful and informative measure of risk of suicide event for a heterogeneous study population, but also allows for comparing different groups for such risk, such as the multiple VHA service networks within the current context. Estimation of the metric with fixed risk membership is analogous to standardization based on histories of suicide attempts using the whole population as the study population. While the metric with dynamic risk membership considers richer information because it integrates dynamic risk and new event, providing more precise estimate of suicide risk. By accounting for history of suicide attempt, the dynamically-changing risk model allows one to include all events to improve power, which is especially important when dealing rare events such as suicide. The approach performs well as evidenced by both real and simulation studies.

When accounting for prior events, we do not distinguish reported from observed suicide events. This assumption may not hold, if there is recall bias, especially if the reported events occurred a long time before the study. The proposed approach may yield estimates that are biased or may not even have a clear interpretation, if this assumption fails.

Although motivated by and developed for suicide event, the proposed approach may be applied to any episodic disease in which risks for the disease changes as events 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 composite variable by grouping different levels of the different variables. If the 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 variables.

Acknowledgments

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: in part by NIH grants R33 DA027521 and R01GM108337, and research fundings from the Center of Excellence for Suicide Prevention, Canandaigua VA Medical Center, Canandaigua, NY.

References

1. Xia Y, Lu N, Katz I, et al. Models for surveillance data under reporting delay: applications to US veteran first-time suicide attempters. *J Appl Stat.* 2015; 42:1861–1876.
2. Britton PC, Brady S, Wu J, et al. Comorbid depression and alcohol use disorders and prospective risk for suicide attempt in the year following inpatient hospitalization. *J Affect Disord.* 2015; 187:151–155. [PubMed: 26339924]
3. Rubin DB. Causal inference using potential outcomes: design, modeling, decisions. *J Am Stat Assoc.* 2005; 100:322–331.
4. Tang, W, He, H, Tu, X. Applied categorical and count data analysis. Boca Raton, FL: Chapman & Hall/CRC; 2012.
5. Wu P, Han Y, Chen T, et al. Causal inference for Mann–Whitney–Wilcoxon rank sum and other nonparametric statistics. *Stat Med.* 2014; 33:1261–1271. [PubMed: 24132928]
6. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika.* 1983; 70:41–55.
7. Holland PW. Statistics and causal inference (with discussion). *J Am Stat Assoc.* 1986; 81:945–970.
8. Robins JM, Greenland S, Hu FC. Estimation of the causal effect of a time-varying exposure on the marginal mean of a repeated binary outcome (with discussion). *J Am Stat Assoc.* 1998; 94:687–700.
9. DAgostino R. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Stat Med.* 1998; 17:2265–2281. [PubMed: 9802183]
10. Miller, RG. Survival analysis. John Wiley & Sons; 2011.
11. Brenner LA, Breshears RE, Betthausen LM, et al. Implementation of a suicide nomenclature within two VA healthcare settings. *J Clin Psychol Med Settings.* 2011; 18:116–128. [PubMed: 21626353]
12. Pagano M, Tu XM, DeGruttola V, et al. Analysis of censored and truncated data: estimating the reporting delay distribution and current AIDS incidence. *Biometrics.* 1994; 50:1203–1214. [PubMed: 7787003]
13. Kalbfleisch JD, Lawless JF. Inference-based retrospective ascertainment: an analysis of the data on transfusion-related AIDS. *J Am Stat Assoc.* 1989; 84:360–372.
14. Brookmeyer R, Gail MH. A method of obtaining short-term projection and lower bounds on the size of AIDS epidemic. *J Am Stat Assoc.* 1988; 83:301–308.

8 Appendix 1

8.1 Technical details

Proof of Theorem 1

- a. Let $p_j = \Pr(y_1 | i \in F_1, i \in H_j)$, and $\hat{p}_j = \widehat{\Pr}(y_1 | i \in F_1, i \in H_j) = \frac{n_j}{N_j}$. Since $\hat{p}_j$ 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 $\tilde{\phi}_1^f = \sum_{j=1}^K \hat{p}_j \cdot \Pr(H_j)$, a summation of the proportions for independent samples, it follows that

$$E(\tilde{\phi}_1^f) = E\left[\sum_{j=1}^K w_j \hat{p}_j\right] = \sum_{j=1}^K w_j E[\hat{p}_j] = \sum_{j=1}^K w_j p_j = \phi_1$$

and

$$\sigma_1^2 = \text{Var}(\tilde{\phi}_1^f) = \text{Var}\left[\sum_{j=1}^K w_j \hat{p}_j\right] = \sum_{j=1}^K w_j^2 \text{Var}[\hat{p}_j] = \sum_{j=1}^K w_j^2 \frac{1}{N_j} p_j (1 - p_j)$$

- c. Since p_j can be estimated by $\hat{p}_j = \widehat{\Pr}(y_1 | i \in F_1, i \in H_j) = \frac{n_j}{N_j}$, the σ_1^2 can be estimated by $\hat{\sigma}_1^2 = \sum_{j=1}^K w_j^2 \frac{1}{N_j} \frac{n_j}{N_j} \frac{N_j - n_j}{N_j}$. Hence the asymptotic variance in Theorem 1(b) has been proved.

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

$$E(\hat{\phi}_1^f) = E\left[\sum_{j=1}^K \hat{w}_j \hat{p}_j\right]$$

and two independent procedures for estimating w_j and p_j are independent, we have

$$E(\hat{\phi}_1^f) = E\left[\sum_{j=1}^K \hat{w}_j \hat{p}_j\right] = \sum_{j=1}^K E[\hat{w}_j] E[\hat{p}_j] = \sum_{j=1}^K w_j p_j = \phi_1$$

The variance of $\hat{\phi}_1^f$ can be derived as follows

$$\begin{aligned}
\sigma_1^2 &= \text{Var}(\hat{\phi}_1^f) = \text{Var}\left[\sum_{j=1}^K \hat{w}_j \hat{p}_j\right] = E\left[\sum_{j=1}^K \hat{w}_j \hat{p}_j\right]^2 - \left[E\left(\sum_{j=1}^K \hat{w}_j \hat{p}_j\right)\right]^2 \\
&= E\left[\sum_{j=1}^K \hat{w}_j^2 \hat{p}_j^2 + \sum_{j=1}^K \hat{w}_j \hat{p}_j \sum_{i \neq j} \hat{w}_i \hat{p}_i\right] - \left[\sum_{j=1}^K w_j p_j\right]^2 \\
&= E\left[\sum_{j=1}^K \hat{w}_j^2 \hat{p}_j^2 + \sum_{j=1}^K \sum_{i \neq j} \hat{w}_j \hat{p}_j \hat{w}_i \hat{p}_i\right] - \left[\sum_{j=1}^K w_j^2 p_j^2 + \sum_{j=1}^K \sum_{i \neq j} w_j p_j w_i p_i\right] \\
&= E\left[\sum_{j=1}^K \left(\hat{w}_j^2 \hat{p}_j^2 - w_j^2 p_j^2\right)\right] + \sum_{j=1}^K E\left[\sum_{i \neq j} (\hat{w}_j \hat{p}_j \hat{w}_i \hat{p}_i - w_j p_j w_i p_i)\right]
\end{aligned}$$

Since

$$\begin{aligned}
E\left[\sum_{j=1}^K \left(\hat{w}_j^2 \hat{p}_j^2 - w_j^2 p_j^2\right)\right] &= E\left[\sum_{j=1}^K \left(\hat{w}_j^2 \hat{p}_j^2 - w_j^2 \hat{p}_j^2 + w_j^2 \hat{p}_j^2 - w_j^2 p_j^2\right)\right] \\
&= \sum_{j=1}^K E\left[\hat{w}_j^2 \hat{p}_j^2 - w_j^2 \hat{p}_j^2\right] + E\left[w_j^2 \hat{p}_j^2 - w_j^2 p_j^2\right] = \sum_{j=1}^K \left[E\left(\hat{p}_j^2 (\hat{w}_j^2 - w_j^2)\right) + E\left(w_j^2 (\hat{p}_j^2 - p_j^2)\right)\right] \\
&= \sum_{j=1}^K \left(p_j^2 + \frac{1}{N_j} p_j (1 - p_j)\right) \frac{1}{M} w_j (1 - w_j) + \sum_{j=1}^K w_j^2 \frac{1}{N_j} p_j (1 - p_j) \\
&= \sum_{j=1}^K p_j^2 \frac{1}{M} w_j (1 - w_j) + \sum_{j=1}^K \frac{1}{M} \frac{1}{N_j} p_j (1 - p_j) + w_j (1 - w_j) + \sum_{j=1}^K w_j^2 \frac{1}{N_j} p_j (1 - p_j)
\end{aligned}$$

and

$$\begin{aligned}
\sum_{j=1}^K E\left[\sum_{i \neq j} (\hat{w}_j \hat{p}_j \hat{w}_i \hat{p}_i - w_j p_j w_i p_i)\right] &= \sum_{j=1}^K \sum_{i \neq j} E[(\hat{w}_j \hat{p}_j \hat{w}_i \hat{p}_i - \hat{w}_i \hat{w}_j p_i p_j + \hat{w}_i \hat{w}_j p_i p_j - w_j p_j w_i p_i)] \\
&= \sum_{j=1}^K \sum_{i \neq j} \{E[\hat{w}_i \hat{w}_j (\hat{p}_i \hat{p}_j - p_i p_j)] + E[p_i p_j (\hat{w}_i \hat{w}_j - w_i w_j)]\} = \sum_{j=1}^K \sum_{i \neq j} -\frac{1}{M} p_i p_j w_i w_j
\end{aligned}$$

we have

$$\begin{aligned}
\sigma_1^2 &= \text{Var}(\hat{\phi}_1^f) = \text{Var}\left[\sum_{j=1}^K \hat{w}_j \hat{p}_j\right] = \sum_{j=1}^K p_j^2 \frac{1}{M} w_j (1 - w_j) + \sum_{j=1}^K w_j^2 \frac{1}{N_j} p_j (1 - p_j) \\
&\quad + \sum_{j=1}^K \frac{1}{M} \frac{1}{N_j} p_j (1 - p_j) w_j (1 - w_j) - \sum_{j=1}^K \sum_{i \neq j} \frac{1}{M} p_i p_j w_i w_j
\end{aligned}$$

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

As $M \rightarrow 1$, $\text{Var}(\sum_{j=1}^K \hat{w}_j \hat{p}_j) \rightarrow \sum_{j=1}^K w_j^2 \frac{1}{N_j} p_j (1 - p_j)$, and the variance is reduced to that for

$\tilde{\phi}_1^f$ in part (b). On the other hand, as $N_j \rightarrow \infty$, i.e. if the rates for all the levels are known,

$\text{Var}(\sum_{j=1}^K \hat{w}_j \hat{p}_j) \rightarrow \sum_{j=1}^K p_j^2 \frac{1}{M} w_j (1 - w_j) = \text{Var}(\sum_{j=1}^K p_j \hat{w}_j)$. Note that in the proof we used

the independence between $\hat{p}_j$ and $\hat{w}_j$. 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.

Table 1

Three scenarios for the hypothetical example.

Scenario	Facility F_1				Facility F_2			
	Subj.		Rate (%)		Subj.		Rate (%)	
	1-3	1-3	1-3	1-3	1-3	1-3	1-3	1-3
History	1-3	1-3	1-3	1-3	1-3	1-3	1-3	1-3
0	1000	10	1.0	3000	30	45	45	1.5
1	2000	40	2.0	100	2	3	3	3.0
2	200	10	5.0	100	5	6	1	1.0
Total	3200	60	-	3200	37	54	49	-

S.A.: number of events.

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

H_j	ESS at baseline	1st event within T	ESS at 1st attempt	2nd event within T	...	ESS at $(L - 1)$ th attempt	L th event within T
0	N_{00}	m_{01}	N_{01}	m_{02}	...	$N_{0(L-1)}$	m_{1L}
1	N_{10}	m_{11}	N_{11}	m_{12}	...	$N_{1(L-1)}$	m_{1L}
2	N_{20}	m_{21}	N_{21}	m_{22}	...	$N_{2(L-1)}$	m_{2L}
...	...	...	...	...	...	...	...
$K - 1$	$N_{(K-1)0}$	$m_{(K-1)1}$	$N_{(K-1)1}$	$m_{(K-1)2}$	...	$N_{(K-1)(L-1)}$	$m_{(K-1)L}$
K	N_{K0}	m_{K1}	N_{K1}	m_{K2}	...	$n_{K(L-1)}$	n_{KL}

ESS: effective sample size.

The mean of the estimated metrics (Est.), the empirical standard errors (SE_E), the asymptotic standard errors (SE_A), and the 95% coverage probability (CP) when the population distribution of the risk factor is known.^a

Table 3

Sample size	Fixed membership				Dynamic change membership			
	Est.	SE_E	SE_A	CP (%)	Est.	SE_E	SE_A	CP (%)
100	0.1224	0.0322	0.0305	91.8	0.1214	0.0265	0.0266	93.5
200	0.1224	0.0232	0.0223	94.1	0.1216	0.0195	0.0191	93.8
500	0.1219	0.0143	0.0143	94.3	0.1213	0.0119	0.0121	95.7
1000	0.1220	0.0102	0.0101	94.4	0.1216	0.0086	0.0086	94.8
5000	0.1218	0.0045	0.0046	95.0	0.1216	0.0039	0.0038	95.3
10,000	0.1218	0.0032	0.0032	95.7	0.1215	0.0027	0.0027	95.1

^aThe true metric is 0.1217.

The mean of the estimated metrics (Est.), the empirical standard errors (SE_E), the asymptotic standard errors (SE_A), and the 95% coverage probability (CP) when the population distribution of the risk factor is estimated.^a

Table 4

Sample size	Fixed membership				Dynamic change membership			
	Est.	SE_E	SE_A	CP (%)	Est.	SE_E	SE_A	CP (%)
100	0.1210	0.0330	0.0305	90.9	0.1209	0.0277	0.0266	93.5
200	0.1224	0.0226	0.0222	94.8	0.1220	0.0192	0.0191	94.2
500	0.1218	0.0145	0.0143	94.4	0.1215	0.0121	0.0121	94.7
1000	0.1219	0.0101	0.0101	94.4	0.1217	0.0088	0.0086	94.8
5000	0.1217	0.0046	0.0046	94.5	0.1216	0.0039	0.0038	94.9
10,000	0.1218	0.0033	0.0032	94.3	0.1216	0.0027	0.0027	95.1

^aThe true metric is 0.1217.

Table 5

The estimated risk for suicide re-events across all the VISNs.

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

CI: confidence interval; SE: standard error; VHA: Veterans Health Administration; VISN: Veterans Integrated Service Networks.