Self-injurious thoughts and behaviors as risk factors for future suicide ideation, attempts, and death: a meta-analysis of longitudinal studies

J. D. Ribeiro1,2,3*, J. C. Franklin1,3, K. R. Fox1, K. H. Bentley4, E. M. Kleiman1, B. P. Chang5 and M. K. Nock1

1 Department of Psychology, Harvard University, Cambridge, MA, USA
2 Military Suicide Research Consortium, Florida State University, Tallahassee, FL, USA
3 Department of Psychological Sciences, Vanderbilt University, Nashville, TN, USA
4 Center for Anxiety and Related Disorders, Boston University, Boston, MA, USA
5 Department of Medicine, Columbia University, New York, NY, USA

Background. A history of self-injurious thoughts and behaviors (SITBs) is consistently cited as one of the strongest predictors of future suicidal behavior. However, stark discrepancies in the literature raise questions about the true magnitude of these associations. The objective of this study is to examine the magnitude and clinical utility of the associations between SITBs and subsequent suicide ideation, attempts, and death.

Method. We searched PubMed, PsycInfo, and Google Scholar for papers published through December 2014. Inclusion required that studies include at least one longitudinal analysis predicting suicide ideation, attempts, or death using any SITB variable. We identified 2179 longitudinal studies; 172 met inclusion criteria.

Results. The most common outcome was suicide attempt (47.80%), followed by death (40.50%) and ideation (11.60%). Median follow-up was 52 months (mean = 82.52, s.d. = 102.29). Overall prediction was weak, with weighted mean odds ratios (ORs) of 2.07 [95% confidence interval (CI) 1.76–2.43] for ideation, 2.14 (95% CI 2.00–2.30) for attempts, and 1.54 (95% CI 1.39–1.71) for death. Adjusting for publication bias further reduced estimates. Diagnostic accuracy analyses indicated acceptable specificity (86–87%) and poor sensitivity (10–26%), with areas under the curve marginally above chance (0.60–0.62). Most risk factors generated OR estimates of <2.0 and no risk factor exceeded 4.5. Effects were consistent regardless of sample severity, sample age groups, or follow-up length.

Conclusions. Prior SITBs confer risk for later suicidal thoughts and behaviors. However, they only provide a marginal improvement in diagnostic accuracy above chance. Addressing gaps in study design, assessment, and underlying mechanisms may prove useful in improving prediction and prevention of suicidal thoughts and behaviors.

Introduction

Suicidal behavior is one of the leading causes of injury and death worldwide. Presently, suicide accounts for nearly one million deaths each year (WHO, 2012). For every death, there are an estimated 25 non-fatal suicide attempts and millions more who experience suicidal thoughts (Nock et al. 2008; Crosby et al. 2011). Over the last five decades, there have been marked increases in research focused on the understanding, treatment, and prevention of suicidal behavior. Despite these efforts, suicide rates remain virtually unchanged (CDC, 2015).

Prior self-injurious thoughts and behaviors (SITBs) are often identified as some of the most robust predictors of future SITBs. Among SITBs, suicide attempt history is consistently cited as one of the strongest risk factors for future suicidal behaviors (Fawcett et al. 1990; Joiner et al. 2005). This claim has become widely accepted and highly influential – for instance, a recent WHO (2014) report stated that ‘by far the strongest risk factor for suicide is a previous suicide attempt’. Supporting this position, a large body of research has consistently linked suicide attempt history to later suicidal ideation (Miranda et al. 2012; Links et al. 2012), attempts (Borges et al. 2008; O’Connor et al. 2013), and death (Suokas et al. 2001; Wenzel et al. 2011). Some studies cite over a 70-fold increase in the
likelihood of a subsequent attempt (Sanchez-Gistau et al. 2013) and close to a 40-fold increase in the likelihood of death (Harris & Barraclough, 1997). However, in stark contrast to these findings, several studies have reported substantially smaller (Wenzel et al. 2011; Van Dulmen et al. 2013) or non-significant (Tejedor et al. 1999; Brådvik & Berglund, 2009) effects. These discrepancies raise doubt about the true effect of prior suicide attempts on future suicidal thoughts and behaviors.

Features of prior attempts [e.g., number, recency (i.e., time since last attempt), intent, lethality, method] have also been proposed to be important indicators of risk (Beck et al. 1979; Rudd et al. 1996; Joiner et al. 1997; Peruzzi & Bongar, 1999). Yet large discrepancies exist in this literature as well, with some studies reporting moderate-to-strong effects (Nock et al. 2010; Roaldset et al. 2012; Troister et al. 2013) and others small or non-significant effects (Sher et al. 2006; Keilp et al. 2010; Chen et al. 2013).

There also is disagreement about the effects of other SITBs – namely, suicidal ideation, suicide plans, and non-suicidal self-injury (NSSI) – on the likelihood of future suicidal thoughts and behaviors. Although it is generally accepted that ideation confers risk for later suicidal thoughts and behaviors, the magnitude of its effects varies widely (Ramchand et al. 2008; Anderson, 2011). The role of planning for suicide on later suicidal thoughts and behaviors also remains in question. Some researchers have proposed that planning for suicide is a necessary precursor of potentially lethal suicidal behavior (Witte et al. 2008; Van Orden et al. 2010) whereas others have argued that lack of planning and impulsivity confers greater risk (Maser et al. 2002; Simon et al. 2002). With respect to NSSI, early research focused heavily on distinguishing the behavior from suicidal outcomes (Brausch & Gutierrez, 2010; Joiner et al. 2012); however, emerging research indicates that the longitudinal effects of NSSI on suicidal behaviors may be much stronger than originally anticipated (Asarnow et al. 2011; Wilkinson et al. 2011).

The effects of indirect experiences with suicidal thoughts and behavior, such as a family history of suicidal thoughts and behaviors and exposure to the suicidal thoughts and behaviors of others (e.g., family, peers, etc.), also are unclear. Several studies report that family history of suicidal behavior is associated with a >10-fold increase in the likelihood of future suicidal behavior (Dahlggaard et al. 1998; Soloff & Chiappetta, 2012; Chan et al. 2014); however, many others have failed to find any effect (Valtonen et al. 2006; Soloff & Fabio, 2008; Dennenhy et al. 2011). Similar discrepancies exist in the literature examining the effects of exposure to others’ SITBs (Lewinsohn et al. 2001; Swanson & Colman, 2013), raising questions about suicide contagion and clustering as well as best practices for postvention.

Given all of these inconsistencies in the literature prompts the question: ‘What are the effects of prior SITBs on future suicidal thoughts and behaviors?’ The present meta-analysis addresses this crucial question and includes four primary aims. First, we provide a descriptive summary of the existing longitudinal literature addressing this question. We focus only on longitudinal studies because we are interested in determining whether prior SITBs confer risk for later suicidal thoughts and behaviors. Although cross-sectional and retrospective studies can provide insight about correlates (i.e., associated features) of a phenomenon, they are not useful in determining risk factors because temporal precedence cannot be inferred from these designs (Kraemer et al. 1997). Second, we examine what, if any, effects prior SITBs, features of prior SITBs, family history, and exposure to SITBs have on future suicidal ideation, attempts, and death. Third, we evaluate the potential moderating effects of sample age, sample severity, and study follow-up length. We do so because differences in methodology related to these issues may influence effects or prior SITBs. Fourth, we consider these findings in the context of their clinical utility, quantified as: (1) improvement in diagnostic accuracy above chance and (2) odds ratios considered in terms of absolute risk of suicide ideation, attempts, and death.

Method

Data sources, study selection, inclusion criteria


A total of 2179 unique papers were identified. Based on abstracts, 1578 studies did not meet inclusion criteria. Inclusion required that the paper include at least one longitudinal analysis predicting suicide ideation, attempts or death using any SITB variable in any population, any year, and any geographic location. Published peer-reviewed studies were excluded on the basis of three criteria: (1) analyses that were not longitudinal; (2) analyses that did not examine discrete
suicide-relevant outcomes; and (3) analyses that were conducted in the context of a primary treatment study. After reading the remaining articles in full, a total of 172 studies were retained (see Fig. 1 for PRISMA flowchart). Supplementary Table S1 includes details regarding each eligible study; Supplement S2 includes a complete reference list for all eligible studies.

We elected to focus on suicidal ideation, attempts, and death as outcomes for two reasons. First, we were primarily interested in determining the effects of prior SITBs on suicidal outcomes.¹ Accordingly, it was necessary that outcomes only include self-injurious thought or behavior outcomes that involved some level of suicidal intent. In other words, the thought or behavior must involve two key elements: (1) it is self-directed and (2) involves some non-zero level of intent to die. Therefore, variables that were not specific to suicidal SITBs [e.g. deliberate self-harm (i.e. self-directed injury with or without suicidal intent), parasuicide (i.e. self-directed injury with unclear suicidal intent), etc.] were not examined as outcome variables. Second, we were interested in understanding the specificity of effects on discrete suicide outcomes. As such, we required that outcome variables did not combine discrete forms of suicidal thoughts and behaviors. Of note, however, these variables (e.g. combined forms of SITBs, deliberate self-harm, parasuicide, etc.) were examined as predictors of suicide ideation, attempts, and death.

To ensure case independence, duplicate cases were removed (n = 8). This occurred when either the same data were re-analyzed across multiple publications and/or multiple follow-up assessments using the same predictors were included in a single study. In the former case, the most inclusive study was retained; in the latter, only estimates at the final assessment were retained.

Data extraction

All statistical tests were examined for each study. Any statistical test where a SITB variable was used to

---

¹ The notes appear after the main text.
predict suicide ideation, attempt or death outcome within a study was termed a ‘prediction case’ and retained for analysis. There were a total of 494 unique prediction cases.

Data extracted from each eligible study included: authors, publication year, follow-up length, number of SITB participants\(^2\), sample type (i.e. general population, clinical, self-injurious), sample age group (i.e. adult, adolescent, mixed), predictor variable, outcome type, and any statistics relevant to any longitudinal analyses using SITB to predict suicide ideation, attempts, or death. Regarding sample type, general population samples included participants not selected for a history of SITB or psychopathology; clinical samples included participants selected based on psychopathology; and SITB samples included participants selected based on SITB history. Category and subcategory codes were assigned to each predictor. Agreement was reached on all category assignments for each predictor by lead and co-authors.

We also considered the issue of study quality. Assessments of study quality are indicated when there is a substantial degree of methodological variability that may influence the accuracy of results. In particular, meta-analyses of treatment studies often warrant quality assessments as methodological differences among treatment studies are common. For instance, treatment studies can vary considerably with respect to design type (e.g. case-control, randomization, double-blindinging), sample (e.g. severity, representativeness), and treatment provision (e.g. fidelity, adherence, compliance), among many other factors. These common methodological differences across treatment studies can substantially influence the accuracy of results. However, in contrast to many treatment meta-analyses, the present meta-analysis includes studies that are highly uniform. This is because inclusion criteria in the present meta-analysis constrained the study pool by requiring that eligible studies share a common core design (i.e. longitudinal) and outcome (i.e. suicidal ideation, attempts, or death).

Although finer-grained methodological differences still exist (e.g. length of follow-up, population severity, etc.), there are no established objective criteria that could inform a priori hypotheses about how these differences may relate to study quality. For instance, there is no objective precedence for determining how the specific length of follow-up (e.g. shorter or longer) or type of sample (e.g. general community, clinical, or self-injurious) impacts accuracy of prediction. To explore possible influences of these differences on meta-analytic results, moderator analyses were conducted. Heterogeneity across studies was also accounted for by using random-effects models in analyses. Both moderator analyses and random-effects models are described in detail below.

**Statistical analyses**

Meta-analyses were performed with Comprehensive Meta-Analysis, version 2 (CMA, 2014); diagnostic accuracy statistics were analyzed with MetaDiSc, version 1.4 (Zamora et al. 2014). When available, unadjusted estimates were used.\(^1\) When odds ratios (ORs) were not reported, they were calculated based on correlations, independent group means, risk ratios, and 2 × 2 contingency tables. Hazard ratios (HRs) were analyzed independently.\(^1\)

Between-study heterogeneity was quantified by \(I^2\) tests. As high levels of between-study heterogeneity were expected, a random-effects model was used for all meta-analyses. Fixed-effects models assume that the underlying (i.e. true) effect size is identical across studies; any observed variance is assumed to be a result of chance. In contrast, random-effects models assume a distribution of similar (but not identical) effects across studies. Accordingly, the combined effect estimated by a random-effect model represents the mean of a distribution of true effects rather than a single true effect. Whereas fixed-effects models only estimate within-study variance random-effects models estimate both within- and between-study variance. Given that systematic between-study variance is common and expected across studies (e.g. due to differences in population, methodology, etc.), random-effects models are typically more appropriate than fixed-effects models in meta-analyses. Systematic between-study variance (i.e. heterogeneity) is accounted for in random-effects models in the weighting and calculation of each prediction case.

Moderator analyses were used to test potential moderating effects of sample age groups, sample severity, or study follow-up length. Slightly diverging from moderation analyses in primary studies, moderation analyses in meta-analyses test whether variation in effect size among studies is associated with differences in selected covariates (i.e. moderators). To test the effects of moderators on effect estimates, we employed meta-regression using a random-effects model using unrestricted maximum likelihood. For sake of interpretability, we provide means and confidence intervals for each analysis involving a categorical moderator (i.e. sample age groups, sample severity) and beta coefficients and slopes for analyses using a continuous moderator (i.e. follow-up length).

Publication bias was quantified using multiple indices, including classic fail-safe \(N\), Orwin’s fail-safe \(N\), Begg & Mazumdar’s rank correlation test, and Egger’s regression test, funnel plot symmetry, and Duval & Tweedie’s trim-and-fill test. Diagnostic accuracy was evaluated in terms of sensitivity, specificity, likelihood ratios (LR\(+\), LR\(−\)), pooled diagnostic OR, and area

https://doi.org/10.1017/S0033291715001804 Published online by Cambridge University Press
under the curve (AUC) using receiver-operated characteristic (ROC) curves. Estimates >3 standard deviations (S.D.) from the mean (n = 13; 2.63%) were omitted as outliers, resulting in a total of 481 unique prediction cases included in analyses. Of note, when outliers were included in analyses, results were highly consistent with those when outliers were excluded.

Results

Data description

Studies spanned 1965–2014. Suicide attempt was the most common outcome (47.80%), followed by death (40.50%) and ideation (11.60%). Nine cases (1.90%) were classified as ‘protective factors’ rather than risk factors; as such, they were excluded from further analysis. The median number of SITB participants across studies was 79 (mean = 710.53, S.D. = 4441.78, range = 1–48,649); however, 4.65% of studies failed to provide clear information about number of SITB participants. Adolescent-only samples comprised 19.50% of all cases, most of which examined attempts (58.50%) and ideation (35.10%).

The median follow-up length across all studies was 52 months (mean = 82.52, S.D. = 102.29, range = 1 week–648 months). The most frequent follow-up interval was 25–60 months (21.40%). Approximately 20% had follow-up intervals of 121 months or longer. Less than 6% had follow-ups shorter than 6 months and <1% had a follow-up shorter than 1 month. Death outcome cases were associated with the longest follow-up (median = 120, mean = 138.99 months, S.D. = 128.94, range = 1–648), followed by attempt (median = 24, mean = 49.53 months, S.D. = 55.34, range = 1–360) and ideation (median = 24 months, mean = 47.16 months, S.D. = 53.52, range = 1–168).

With respect to sample severity, the majority of cases involved either self-injurious (42.60%) or clinical (38.00%) samples; only 19.30% involved general population samples. Death outcomes most commonly involved self-injurious samples (75.90%), followed by clinical (18.97%) and general population (5.13%) samples. Clinical samples were most common for attempt (56.52%); self-injurious (22.17%) and general population (21.30%) samples were equally represented. For ideation, general population samples were the most common (60.71%) followed by clinical (28.57%) and self-injurious (10.71%) samples.

Overall prediction and publication bias

Ideaition

For overall prediction analyses, the estimate reflects the pooled effect of all prior SITBs, regardless of category type, on the outcome of interest. A total of 54 prediction cases were included, resulting in a weighted mean OR of 2.07 [95% confidence interval (CI) 1.76–2.43]. Between-study heterogeneity was extreme ($I^2 = 93.56\%$) and there was consistent evidence of publication bias. Accounting for publication bias, the effect would be reduced to an overall weighted mean OR of 1.33 (95% CI 1.10–1.61). See Table 1 and Fig. 2 for detailed publication bias statistics and funnel plots. No cases were available for diagnostic accuracy analyses.

Attempt

For OR analyses, 189 prediction cases were included, generating a weighted mean OR of 2.16 (95% CI 2.01–2.33). Heterogeneity was high ($I^2 = 90.76\%$). Multiple indices indicated significant publication bias. Accounting for publication bias, the weighted mean OR would reduce to 1.68 (95% CI 1.57–1.81). See Table 1 and Fig. 2.

A total of 57 cases included sufficient information for diagnostic accuracy analyses. Overall accuracy was slightly better than chance (AUC = 0.60). Sensitivity was 26% (95% CI 24–27), indicating extremely limited ability to identify true positive cases. Specificity was 87% (95% CI 87–88), suggesting the ability to identify true negative cases was acceptable; however, this is likely a statistical artifact resulting from the low-base rates of the outcome and predictors, which necessarily generate a high number of true negatives and low number of false negatives. The LR+ was 1.77 (95% CI 1.57–2.04), LR- was 0.84 (95% CI 0.80–0.88), and the pooled diagnostic OR was 2.25 (95% CI 1.85–2.73).

Death

OR analyses included 144 cases. Heterogeneity was high ($I^2 = 75.49\%$). Publication bias was indicated across multiple indices. The weighted mean OR was 1.54 (95% CI 1.39–1.71), and decreased to 1.51 (95% CI 1.36–1.68) when publication bias was considered. See Table 1 and Fig. 2.

Diagnostic accuracy analyses included 52 cases. Overall accuracy was poor (AUC = 0.62). This likely was a result of extremely poor sensitivity (10%, 95% CI 10–11). Specificity was acceptable (86%, 95% CI 86–96); however, this again is likely in part a statistical artifact. The LR+ was 1.57 (95% CI 1.37–1.81), LR- was 0.96 (95% CI 0.90–0.99) and pooled diagnostic OR was 1.76 (95% CI 1.45–2.15).

Prediction trends across years of research

The first studies using prior SITBs to predict ideation, attempts, or death were published in 1983 (Adam
When considered as a single construct, the weighted mean ORs were significant for attempt (weighted mean OR 1.58) and death (weighted mean OR 1.60) and not ideation (weighted mean OR 0.79).

The effects of specific features also were tested. Only estimates for attempts and death are presented as there were too few ideation cases. In the prediction of suicide attempts, number (weighted mean OR 1.92), recency (weighted mean OR 2.51), maximum lethality (weighted mean OR 1.54) and maximum intent (weighted mean OR 1.23) of past attempts were significant. General (rather than maximum) lethality and intent were not significant. Too few cases were available to provide reliable estimates for the remaining features [i.e. level of preparations (n=2), onset of first attempt (n=1), history of a violent attempt (n=2), and concerning reactions following attempts (n=3)].

In the prediction of suicide death, only history of a violent attempt (weighted mean OR 1.89) and intent (weighted mean OR 1.20) were significant. Number, lethality, level of preparations, degree of intoxication, and having a concerning reaction following an attempt were not. Reliable estimates could not be provided for attempt recency (n=1), reason for past attempt (n=3), and history of interrupted attempt (n=1).

NSSI

A prior history of NSSI was associated with significantly increased odds of suicide attempt (weight mean OR 4.27). There were insufficient cases to test associations with suicide ideation or death.

SITB and deliberate self-harm (DSH)

Several studies included more general predictors of SITB or DSH that did not specify what type of thought or behavior was present. A prior history of SITB was associated with significantly increased odds of suicide attempt (weighted mean OR 2.26) and death (weighted mean OR 1.30) but not ideation (weighted mean OR 0.79).

<table>
<thead>
<tr>
<th>Fail-safe N</th>
<th>Begg &amp; Mazumdar’s rank correlation</th>
<th>Egger’s test of the intercept</th>
<th>Duval and Tweedie’s trim &amp; fill</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Classic</td>
<td>Orwin’s</td>
<td>Missing cases</td>
</tr>
<tr>
<td>Suicidal ideation</td>
<td>9230</td>
<td>22</td>
<td>( r = 0.13, p &lt; 0.001 )</td>
</tr>
<tr>
<td>Suicide attempt</td>
<td>7391</td>
<td>177</td>
<td>( r = -0.27, p &lt; 0.001 )</td>
</tr>
<tr>
<td>Suicide death</td>
<td>8871</td>
<td>74</td>
<td>( r = 0.13, p &lt; 0.01 )</td>
</tr>
</tbody>
</table>

OR, Weighted mean odds ratio.

Classic and Orwin’s fail-safe N values represent the number of studies required to nullify the observed effects; Begg & Mazumdar’s rank correlation test computes the rank order correlation between effect estimates and standard error; Egger’s test of the intercept uses precision (i.e. the inverse of the standard error) to predict the standardized effect (i.e. effect size divided by the standard error). The size of the effect is reflected in the slope and bias is reflected in the intercept \( (B_0) \); Missing cases under Duval & Tweedie’s trim & fill are the number of cases estimated as missing below the mean.

et al. 1983), 1971 (Bagley & Greer, 1971), and 1965 (Motto, 1965), respectively. Meta-regressions based on 1-year intervals indicated no significant change in predictive ability across time for suicidal ideation \( (b = -0.05, p = 0.08) \), attempts \( (b = -0.006, p = 0.50) \), or death \( (b = 0.007, p = 0.15) \).

Risk factor category analyses

Suicidal ideation

The full results of the risk factor category analyses are presented in Table 2. As shown, prior suicide ideation significantly increases the likelihood of all outcomes, with the strongest estimate for subsequent ideation (weighted mean OR 3.12). Effects were weaker in the prediction of attempts (weighted mean OR 1.88) and death (weighted mean OR 1.95).

Suicide plans

A history of suicide plans was associated with significantly increased odds of suicide death (weighted mean OR 1.44). There were an insufficient number of cases to produce reliable estimates for other suicide-relevant outcomes.

Suicide attempt

A history of suicide attempt (presence or absence) was associated with significantly increased odds of all outcomes, with the strongest effects for attempt (weighted mean OR 3.61) and weakest for suicide ideation (weighted mean OR 1.58).

Suicide attempt features

When considered as a single construct, the weighted mean ORs were significant for attempt (weighted mean OR 1.54) and maximum intent (weighted mean OR 2.51), maximum lethality (weighted mean OR 1.54), and concerning reactions following attempts (n=3)].

In the prediction of suicide death, only history of a violent attempt (weighted mean OR 1.89) and intent (weighted mean OR 1.20) were significant. Number, lethality, level of preparations, degree of intoxication, and having a concerning reaction following an attempt were not. Reliable estimates could not be provided for attempt recency (n=1), reason for past attempt (n=3), and history of interrupted attempt (n=1).

NSSI

A prior history of NSSI was associated with significantly increased odds of suicide attempt (weight mean OR 4.27). There were insufficient cases to test associations with suicide ideation or death.

SITB and deliberate self-harm (DSH)
Fig. 2. Funnel plots. Open circles represent observed estimates; shaded circles represent imputed values estimated to be missing to the left of the mean (due to publication bias). Open diamond indicates unadjusted weighted mean odds ratio; shaded diamond indicates weighted mean odds ratio adjusted for publication bias.
Exposure to SITB of others (e.g. friends, family members, etc.) was associated with increased odds of suicide ideation (OR 2.13) and attempt (OR 1.58) but not death. Given the small number of prediction cases, results were not reliable for specific forms of family history (e.g. history of death vs. attempt, etc.).

**Family history of SITB**

Family history of SITB was associated with significantly increased odds of suicide ideation (OR 2.13) and attempt (OR 1.57) but not death. Given the small number of prediction cases, results were not reliable for specific forms of family history (e.g. history of death vs. attempt, etc.).

**SITB exposure**

Exposure to SITB of others (e.g. friends, family members, schoolmates, etc.) was associated with significantly increased odds of suicide ideation (weighted mean OR 1.56) and attempt (weighted mean OR 2.05). There were too few cases available to provide reliable estimates for more specific forms of exposure.

**Moderator analyses**

**Sample severity**

Across all outcomes, weighted mean ORs were statistically equivalent across clinical (ideation: OR 2.81, 95% CI 1.98–4.00; attempts: OR 2.45, 95% CI 2.20–2.71; death: OR 2.10, 95% CI 1.49–2.94) and general population samples (ideation: OR 2.17, 95% CI 1.82–2.59; attempts: OR 2.27, 95% CI 1.98–2.60; death: OR 2.82, 95% CI 1.65–4.81). Estimates were substantially
weaker among self-injurious samples (ideation: OR 0.84, 95% CI 0.17–4.17; attempts: OR 1.46, 95% CI 1.14–1.86; death: OR 1.42, 95% CI 1.31–1.54).

Sample age
Predicting ideation, weighted mean ORs for adult (OR 2.38, 95% CI 1.26–4.49) and adolescent (OR 1.94, 95% CI 1.70–2.20) samples were not significantly different; the estimate for mixed samples was not significant (OR 1.67, 95% CI 0.02–102.14). Predicting attempts, adult (OR 2.28, 95% CI 2.05–2.53), mixed (OR 1.96, 95% CI 1.26–3.05), and adolescent (OR 2.08, 95% CI 1.85–2.33) samples were statistically equivalent. For death, the effect was strongest for adolescent samples (OR 2.52, 95% CI 1.74–3.65); adult (OR 1.69, 95% CI 1.42–1.99) and mixed (OR 1.35, 95% CI 1.21–1.50) samples were statistically equivalent.

Follow-up length
There were no significant effects of follow-up length on the prediction of ideation ($b = -0.001, p = 0.59$) or death ($b < 0.001$, $p = 0.79$). For attempts, predictive ability significantly worsened as follow-up length increased, though the effect was small ($b = -0.002, p < 0.001$).

Discussion
Our findings indicate that prior SITBs are statistically significant risk factors for suicide ideation, attempts, and death; however, effects were considerably weaker than anticipated. Overall weighted mean ORs for ideation and attempts were only slightly above 2.0; for death the estimate was near 1.5. Adjusting for publication bias further reduced estimates. Effects were consistent regardless of sample severity, sample age groups or study follow-up lengths. Diagnostic accuracy analyses were in line with these findings, with prior SITBs being characterized by acceptable specificity (i.e. true negative rate), extremely poor sensitivity (i.e. true positive rate), and AUCs marginally above chance. We emphasize that these findings reflect the effects of prior SITBs when studied within the narrow methodological constraints of the existing literature.

Beyond overall predictive power, we also evaluated the effects of specific risk factor categories. Prior suicide ideation was the strongest predictor of later ideation; NSSI and suicide attempt history conferred the most risk for later suicide attempts; and suicide attempt history and suicide ideation were among the strongest predictors of suicide death. Surprisingly, features of past attempts were relatively weak predictors, with most features producing estimates close to 1.5 and several others generating non-significant effects. Family history of SITBs also conferred relatively little risk for later suicide ideation or attempts and had no significant effect on suicide death. Exposure to SITBs was a stronger predictor of suicide attempts than anticipated with effects comparable to factors like prior suicidal ideation and number of past attempts.

Although some predictors did emerge as stronger than others, it is critical to evaluate these results with respect to their clinical utility. We evaluated clinical utility on two domains. First, we considered the improvement of diagnostic accuracy. As noted above, knowledge of SITBs offered only slight improvement above chance. Second, we considered the magnitude of OR estimates in terms of the absolute risk of suicide ideation, attempts, and death. Suicidal thoughts and behaviors are rare. The likelihood of death by suicide in the United States in a given year is 12.5/1 000 000 (i.e. 0.000125); attempts are estimated to be 25 times more likely (i.e. 0.0031). The strongest predictor in this meta-analysis was NSSI (weighted mean OR 4.27) predicting future suicide attempts. Based on this estimate, the presence of NSSI would still only increase the odds of an attempt to 0.013 – a figure still nearly zero. Moreover, these calculations are based on prediction over a 1 year interval; most clinicians are tasked with determining risk over a period of days or weeks.

The present findings suggest that, in terms of absolute (rather than relative) odds, prior SITBs are weak risk factors for future suicide ideation, attempts, and death at least within the narrow methodological confines in which they have been studied to date. It is possible, however, that prior SITBs are powerful predictors when considered in the context of other risk factors (e.g. psychopathology). Results from sample severity moderation analyses offer some insight into this issue. Studies of self-injurious and clinical samples typically involve more stringent comparison groups, thereby controlling for a host of risk factors not accounted for in general samples. If prior SITB were indeed stronger in the context of other risk factors, we would expect that effects of prior SITB would be substantially stronger in general samples. However, sample severity moderation analyses generated only small effects, suggesting that, even in the context of other risk factors, prior SITBs remain weak risk factors in an absolute sense. Results of these analyses were surprising as we had anticipated more pronounced differences between sample severity groups; however, we note again that these results only reflect the evidence of the existing literature and should therefore be interpreted within the narrow methodological constraints of researched published to date. Studies that assess this issue more directly and extend beyond existing methodological confines are needed. For instance, approaches that involve machine learning algorithms...
to combine large number of risk factors over a short follow-up period may be particularly promising (Kessler et al. 2015).

Based on the present meta-analysis, we recommend four primary directions for future research. First, features of prior attempts, family history of SITB, and exposure to SITB have received very little empirical attention. Widespread consensus exists in the field that prior attempt features (e.g. lethality, number, intent, etc.) in particular are extremely important risk factors (Joiner et al. 1999; Peruzzi & Bongar, 1999). Our findings are inconsistent with this belief; however, we caution that results are based on remarkably few prediction cases within each specific feature category. As such, we strongly recommend prioritizing longitudinal studies of prior attempt features.

Second, there is a need for improved assessments, particularly for the assessment of intent and plans. The majority of intent cases used the Suicide Intent Scale (SIS; Beck et al. 1974); the majority of planning cases used the planning subscale of the same scale. The inclusion of planning within the SIS necessarily conflates these constructs, raising questions about the true effects of these features considered independently. Efforts focused on improving the understanding and measurement of specific features of attempts, particularly intent and planning, are critical.

Third, studies focused on acute or short-term prediction are needed. Presently, the average follow-up is nearly 7 years; yet, clinicians are tasked with determining risk over the period of hours or days. Our findings indicated that longer follow-ups did not improve predictive power and, in some cases, significantly weakened it. Moving forward, studies would benefit from incorporating shorter follow-ups with strong potential for producing clinically useful results (Glenn & Nock, 2014).

Fourth, clarifying the mechanisms underlying the relations between prior SITBs and future suicidal thoughts and behaviors would be informative. Viable mechanisms must account for the present pattern of findings. This includes (1) why prior SITB are generally weak predictors of future suicidal SITB and (2) why suicidal and non-suicidal SITBs confer statistically equivalent risk for future suicidal behaviors.

In sum, prior SITBs do confer risk for future suicidal ideation, attempts, and death; however, effects are substantially weaker than anticipated. When considered in the context of the extremely low prevalence of suicidal thoughts and behaviors and the clinical demands that call for very short prediction timeframes, prior SITBs may not improve prediction much beyond chance levels. Our findings highlight critical gaps in study design, assessment, and the mechanisms that lead prior SITBs to confer risk for future suicidal thoughts and behaviors. Future research must address these gaps in order to make significant progress in the prevention of suicidal thoughts and behaviors.

Supplementary material
For supplementary material accompanying this paper visit http://dx.doi.org/10.1017/S0033291715001804.

Acknowledgements
This work was supported in part by funding from the Military Suicide Research Consortium (MSRC), an effort supported by the Office of the Assistant Secretary of Defense for Health Affairs (Award No. W81XWH-10-2-0181). This work was supported in part through a training grant (T32MH18921) from the National Institute of Mental Health. Opinions, interpretations, conclusions and recommendations are those of the authors and are not necessarily endorsed by the MSRC or the Department of Defense.

Declaration of Interest
None.

Notes
1 Historically, debate about nomenclature has been common in the fields of suicidal and non-suicidal self-injury research. For the purposes of this meta-analysis, we drew from the terminology proposed by Nock (2010), as it could provide the most fine-grained information about specific self-injurious behaviors.
2 Of note, we elected to report number of SITB participants (i.e. the number of participants who reported any SITBs during the study) instead of sample size. This is because sample statistics were highly skewed given that some samples included population level data whereas others were substantially smaller. Number of SITB participants provides a less biased estimate and is also highly relevant to the aims of this study.
3 Only 12.5% of cases involved adjusted estimates. Results from adjusted estimates were in line with the results from analyses using unadjusted estimates.
4 HR results were highly consistent with those of OR analyses; however, due to space limitations, HR results are reported in the Supplementary Material (see Supplement S3).

References


