



Published in final edited form as:

Psychol Assess. 2018 October ; 30(10): 1356–1366. doi:10.1037/pas0000580.

The Brief Death Implicit Association Test: Scoring Recommendations, Reliability, Validity, and Comparisons with the Death Implicit Association Test

Alexander J. Millner^a, Daniel D.L. Coppersmith^a, Bethany A. Teachman^b, and Matthew K. Nock^a

^aDepartment of Psychology, Harvard University, 33 Kirkland St. Cambridge, Massachusetts, 02138, USA

^bDepartment of Psychology, University of Virginia, 485 McCormick Road Charlottesville, VA 22903m USA

Abstract

Assessing suicidal thoughts and behaviors is difficult because at-risk individuals often fail to provide honest or accurate accounts of their suicidal thoughts or intentions. Research has shown that the Death Implicit Association Test (D-IAT), a behavioral test that measures implicit (i.e. outside of conscious control) associations between oneself and death concepts, can differentiate among people with different suicidal histories, such as those with different severity or recency of suicidal behaviors. We report here on the development and evaluation of a shorter and simpler version of the D-IAT called the Death Brief Implicit Association Test (D-BIAT). We recruited large ($n > 1,500$) samples of participants to complete the original D-IAT and shorter D-BIAT via a public web-based platform. We evaluated different scoring approaches, assessed the reliability and validity of the D-BIAT and compared it with the D-IAT. We found that the D-BIAT was reliable, provided significant group differences with effect sizes on par with the D-IAT, as well as similarly sized classification metrics (i.e., receiver operator characteristics). Although the D-IAT was non-significantly better on most outcomes, the D-BIAT is 1 – 1 ½ minutes shorter and provided larger effect sizes for distinguishing between past year and lifetime attempters. Thus, there is a trade-off between administration time and improved outcomes associated with increased data. The D-BIAT should be considered for use where time or participant burden needs to be minimized, such as in clinical settings.

Keywords

Suicidal behavior; suicide; implicit cognitions; Implicit Association Test; IAT

Suicide is a leading cause of death globally with nearly a million suicide deaths every year (World Health Organization, 2017). In the U.S., government agencies have initiated campaigns and increased resources to reduce suicide (Office of the Surgeon General (US) &

National Action Alliance for Suicide Prevention (US), 2012), yet the suicide rate reached a 20 year peak in 2014 (Curtin, Warner, Hedegaard, & others, 2016). Thus, suicide prevention remains a pressing public health problem.

One of the most important issues in enhancing suicide prevention efforts is improving suicide risk assessment in clinical settings (Large et al., 2016). However, risk assessment is among the most challenging problems faced by both clinicians, who seek to keep their individual patients safe, and researchers, who seek to understand how to best predict and prevent suicide on a macro scale. One important concern is that the dominant form of risk assessment is asking patients to report their own level of risk. Although, self-report provides crucial information regarding suicidal thoughts and past behaviors, it is limited because people may not accurately report their current or near future suicidal risk (Carter et al., 2017), either because they are motivated not to be forthcoming or they do not have conscious awareness of their future risk level. A recent meta-analysis found that the three-month period following discharge from a psychiatric facility represents a high-risk period with a pooled suicide rate (1,132 per 100,000) approximately 100 times the global suicide rate (Chung et al., 2017). Studies have shown, however, that large majorities of inpatients that die by suicide shortly after they are discharged reported no suicidal ideation at last communication with a clinician (Berman, 2017; Busch, Fawcett, & Jacobs, 2003; Ribeiro et al., 2017).

To overcome the limitations of self-report and improve risk assessment, researchers have been seeking more objective markers of suicidal risk via computerized behavioral tests, such as the Implicit Association Test (IAT). The IAT uses the reaction time of keyboard (or touch screen) responses to measure implicit cognitions that are relatively outside of conscious control. Implicit measures are promising because they may provide access to risk levels that are not attainable through self-report. Indeed, across several studies, including one large-scale study ($n > 2,000$), implicit cognitions on self-harm IATs have been found to track the degree of severity of suicidal behaviors (i.e., distinguish between those with a history of suicide attempts and those with a history of ideation) and the degree of recency (i.e., differentiating those with a past year suicide attempt versus those with a lifetime attempt) (Glenn et al., 2016). In smaller-scale studies, implicit cognitions have also prospectively predicted suicidal behaviors above commonly assessed risk factors and clinician prediction (Barnes et al., 2017; Nock et al., 2010; Nock & Banaji, 2007; Randall, Rowe, Dong, Nock, & Colman, 2013). Thus, assessing implicit cognitions with the IAT appears to be a promising approach to improving suicide risk assessment.

One area of concern with the potential translation of behavioral tasks, such as the IAT, to applied clinical settings, is the patient burden associated with completing such tasks, in terms of time and cognitive demand. Although the Death IAT (D-IAT) is brief (~3-4 min including instructions), shortening the task may enhance its clinical utility. For example, if patients were asked to complete the D-IAT as part of an assessment battery in a busy emergency department or following weekly therapy sessions or multiple times a week or day as part of ongoing risk assessment, reducing the task by a minute or two may increase both feasibility and participation rates. Thus, a version of the D-IAT that is shorter but produces similar results as the full D-IAT could provide enhanced clinical value.

In terms of cognitive demand, the D-IAT requires participants to categorize words into four different categories. Although relatively simple, the D-IAT may be challenging for some high-risk patients, particularly given that patients with a prior suicide attempt show reduced executive functioning and attention (Keilp et al., 2001, 2013). Thus, there is also a need for a simpler version of the D-IAT that reduces the number of categories participants have to keep in mind.

The Brief IAT (BIAT; Nosek, Bar-Anan, Sriram, Axt, & Greenwald, 2014) can address these two issues because the BIAT requires about half the time of the regular IAT and, also, it is a simpler task, requiring participants to categorize words into only two categories. A Death version of the BIAT (D-BIAT) will only be useful if, at minimum, it reproduces effects from the regular D-IAT, such as tracking the severity and recency of suicidal behaviors. If the D-BIAT also produced similar magnitude effects as the regular D-IAT, then, given its shorter duration and potential reduced cognitive demands, it could find more widespread use in clinical settings and research studies as well, which could provide additional understanding of implicit cognitions about death and suicidal behaviors.

Given that the D-BIAT has never been previously tested, there are multiple steps required to establish whether it has utility, in either clinical or research contexts. Thus, there were two sequential aims of the current study. First, we sought to test the best approach to scoring the D-BIAT by comparing the group difference effect sizes, classification metrics, and reliability of different D-BIAT scoring methods. For this aim, we also tested whether, compared with a two-block version, an additional block improved results enough to warrant its inclusion and increased task duration (approximately 25 seconds longer). Second, in a separate sample, we compared the D-BIAT, using the scoring approach from the first aim that maximized its effects, and the regular D-IAT to determine whether the D-BIAT provided similar results as the regular D-IAT in less time and in a less cognitively demanding test.

Method

Participants

Between November 2014 and April 2017, 4,550 adult volunteers consented to participate at the Project Implicit Mental Health (PIMH) research website (www.ImplicitMentalHealth.com) and selected a self-harm IAT among several mental health-related IATs. Participants that selected to complete a self-harm-related IAT were randomly assigned to the D-IAT ($n = 2,212$) and, beginning in January 2015, the D-BIAT ($n = 2,338$). Publically available, web-based versions of the IAT were first administered in 1998. Since that time millions of people have participated through the Project Implicit and PIMH websites and the validity this administration approach has been well-supported (Nosek et al., 2007; Nosek, Banaji, & Greenwald, 2002; Nosek, Greenwald, & Banaji, 2005). Participants were excluded from analyses if: (1) they consented but did not complete the D-IAT ($n = 87$) or D-BIAT ($n = 203$), did not complete the demographic information (IAT: $n = 229$, D-BIAT: $n = 222$), surpassed the error (IAT: $n = 11$, D-BIAT: $n = 162$) or latency criteria (IAT: $n = 47$, D-BIAT: $n = 97$; see below for criteria), resulting in a final sample size of 3,586 (D-BIAT: $n = 1731$; D-IAT: $n = 1855$; some participants were excluded

for more than one criterion). This study was approved by the Harvard University and University of Virginia institutional review boards.

Measures

Sociodemographics—Sociodemographic data was collected from participants, including age, sex, ethnicity, race, education, and country of residence.

Death Implicit Association Test—Implicit cognition about suicide was measured using death-related versions of the IAT (Greenwald et al., 1998) and BIAT (Nosek et al., 2014). In both tasks, participants classify stimuli presented in the center of the screen into categories. The category labels are located on the left and right of the screen and participants classify stimuli by selecting keys corresponding to the left or right. By measuring reaction times (RTs) as participants classify the different stimuli, the IAT and BIAT can reveal implicit mental associations (explained in further detail below). Both the D-IAT and D-BIAT were examined to measure the extent to which participants associated death-related constructs with the self.

In the regular D-IAT, the categories correspond to concepts, such as “death” and “life” and, attributes, such as “me” and “not me.” Each side of the screen always contains one concept category and one attribute category. For example, in one category pairing condition, participants may see “death” and “me” paired and “life” and “not me” paired. On each trial, as stimuli related to each of these categories appear, one at a time, in the center of the screen, participants sort them as quickly as possible into their corresponding categories by pressing the “E” key, corresponding to the categories on the left, or the “I” key, corresponding to the categories on the right. Later in the task, the attribute categories switch sides and more trials are administered. Requiring the same response for concept and an attribute, implicitly links the two categories (e.g. “death and me” and “life and not-me”). RTs are assumed to be faster if the concept and attribute pairings are more automatically associated in memory. An implicit bias is thus revealed if RTs during one set of concept-attribute pairings (e.g., “death and me” and “life and not me”) are faster than reactions times during the opposite set (e.g., “death and not me” and “life and me”).

The D-IAT contains 7 blocks. To provide practice and introduction to the task, three of the blocks, each with 20 trials, are single dimension classification blocks (e.g., categorizing stimuli as fitting into either the death or life categories, without also categorizing the me/not me stimuli). Data from these blocks are discarded. Thus, there are four blocks scored. The first block of each pairing contains 20 trials and the second block contains 40 trials but they are scored equally, meaning the first twenty trials are weighted heavier than the later 40 trials. Weighing trials in this way is in accordance with standard IAT scoring procedures (Greenwald, Poehlman, Uhlmann, & Banaji, 2009), which also recommends that all error trials are retained in analyses and that trials with latencies greater than 10000 ms (0.3% of all trials) or less than 400 ms (2.6% of all trials) are discarded and participants with more than 40% errors or more than 10% trials with long or short latencies are excluded (see the Participants section for numbers of participants that exceeded these criteria (none had greater than 10% trials with 10000ms latencies)). Mean reaction times from corresponding

blocks are subtracted and divided by the standard deviation over all trials from those blocks, resulting in *D* scores that are then averaged to create an overall *D* score. *D* scores were calculated so that positive values indicate a stronger association between the self and death. See the Appendix of Glenn et al. (2016) for the complete stimulus set.

Death Brief Implicit Association Test—The D-BIAT is identical to the regular D-IAT with three important exceptions. First, like the D-IAT, there are four categories of words presented; however, only two categories labels are presented and participants only have to focus on categorizing words under these two labels. For example, participants are instructed to press the “I” key to categorize “death” or “me” words and to press the “E” key to categorize any other stimulus. Thus, rather than having to keep track of four categories as in the D-IAT, participants only keep track of two. Second, the “me” label appears in every block and the “death” and “life” labels are presented every other block. The “other” label is never presented. Third, the D-BIAT is shorter than the D-IAT; there are fewer blocks and there are no single dimension classification blocks, meaning, unlike the D-IAT no data are discarded. Based on Nosek et al. (2014), we recoded latencies shorter than 400 ms (2.7% of all trials) to 400 ms and latencies greater than 2000 ms (5.6% of all trials) to 2000 ms, removed trials greater than 10000 ms (0.2% of all trials) and excluded participants with more than 10% of trials under 300 ms or for whom 30% of trials were errors (see the Participants section for numbers of participants that exceeded these criteria). Blocks consisted of 20 trials.

The BIAT has shown high consistency and sensitivity for multiple topics (Petróczi et al., 2010; Yang, Shi, Luo, Shi, & Cai, 2014), but no study has ever examined the D-BIAT. Although, Nosek et al. (2014) provided scoring recommendations for the BIAT, the D-BIAT has several differences with the version of the BIAT analyzed in their paper. The BIAT analyzed in this prior study included 4 trials at the beginning of each block that contained only attribute categories, whereas the D-BIAT contained concept and attribute categories on all trials. Second, the BIAT from the prior study contained a 16-trial practice block whereas the D-BIAT did not contain any practice. Third, the BIAT from Nosek et al. (2014) contained four total blocks with usable data whereas the D-BIAT contained six blocks (three blocks of each pairing). Therefore, one goal of the current study was to test different approaches to including and excluding data (e.g., excluding the first four trials) from the scoring of the D-BIAT.

Self-Injurious Thoughts and Behaviors Interview—Participants’ history of self-injurious thoughts and behaviors was assessed using an abbreviated (20-item) self-report version of the Self-Injurious Thoughts and Behaviors Interview (SITBI; Nock, Holmberg, Photos, & Michel, 2007). Questions assessed history of non-suicidal self-injury, suicidal thoughts, suicide plans, and suicide attempts. To identify the subset of attempts requiring medical attention, “If you have ever attempted suicide, did any attempt result in an injury, poisoning, or overdose that had to be treated by a doctor or nurse?” was also asked. Individuals rated the frequency of each behavior within their lifetime, the past year, and the past week. The interview version of the SITBI has shown high reliability and validity in multiple samples (Fischer et al., 2014; García-Nieto, Blasco-Fontecilla, Paz Yepes, & Baca-

García, 2013). The study also included a risk management protocol, which asked participants questions about their current desire to harm themselves and their desire to die. Although, there are limitations to collecting these data with self-report, studies have found that people are more comfortable reporting sensitive information in online surveys, even when not fully anonymized (Shapiro, Chandler, & Mueller, 2013). The fact that the current study was both online and anonymous, may have made participants more open to reporting suicidal thoughts and behaviors, compared with face-to-face interviews in a lab or clinical setting. This online, self-report version of the SITBI has not been validated.

Procedure

The PIMH website (<https://implicit.harvard.edu/implicit/user/pimh>) was advertised through posts on internet websites, including reddit and similar message boards. It is also cited in many researchers' presentations and news coverage, and it is likely that some visitors come to the site through search engines or a class assignment. Unfortunately, we do not collect data on how people find the PIMH website.

Once on the PIMH website, participants are given preliminary information regarding implicit associations in general and then could select among several mental health-related IATs (e.g. depression, views of people with mental illness, self-esteem, alcohol, anxiety, and others (see <https://implicit.harvard.edu/implicit/user/pimh>)). After selecting to complete a self-harm IAT, they completed an informed consent agreement and then were randomly assigned to either the D-IAT or the D-BIAT or another self-harm IAT. The D-IAT or D-BIAT, the SITBI, and the demographics questionnaire were completed in a random order. Following the task and questionnaires, participants were provided information about the purpose of the study. Participants were then given the option of viewing their D-IAT/D-BIAT feedback on the next page (e.g., "Your responses show that you sorted words much faster when DEATH and ME were paired on the same key (relative to DEATH and NOT ME), which suggests that you may have a strong implicit association between death and yourself.") (Approximately 3% of participants declined to view the feedback and this was the case within both D-IAT and D-BIAT samples). If a participant responded to any item regarding current desire to die/hurt oneself with "extremely," he or she received a message of concern and additional mental health resource information.

Data Analysis

The first goal of the current study is to test different approaches to scoring the D-BIAT. We split the D-BIAT sample into two independent samples – one to test different D-BIAT scoring approaches and one to compare with the D-IAT. The sample used to compare different D-BIAT scoring approaches contained 33% of all participants that completed the D-BIAT ($n = 570$). We tested two different scoring parameters: different numbers of early trials removed, and number of blocks scored. Nosek et al. (2014) compared different BIAT scoring approaches and recommended removing the first four trials of each block. However, the version of the BIAT analyzed by these authors contained four "warm-up" trials at the onset of each block that contained only concept categories whereas the D-BIAT contained regular trials from the onset of the block. Therefore, we compared removing the first four trials, as recommended by Nosek et al. (2014), with removing just the first trial of each

block (based on a reviewer's comment, we also tested each block with all available trials; results, which were generally inferior to removing the first trial of each block, are available upon request). The BIAT analyzed in Nosek et al. (2014) also contained four total blocks, whereas the D-BIAT contained six total blocks. Therefore, we compared whether using data from all six blocks (i.e., three blocks of each category pairing condition) was superior to using the data from just the first four (i.e., two blocks of each category pairing condition) or two (i.e., one block of each category pairing condition) blocks. If using differing number of blocks produced identical results, we favored the shorter version, given that it reduces participant burden.

We used the following metrics to compare the different approaches to analyzing the D-BIAT: effect sizes of group differences, reliability, and classification metrics (i.e., area under the curve for the receiver operator characteristic; AUC-ROC) to discriminate among different groups. For group difference effect sizes and classification metrics, we used three groupings previously shown to be significantly different on the D-IAT (Glenn et al., 2016) and an additional grouping based on a reviewer's comment that we include a group relevant to risk assessment situations: (i) people with a lifetime suicide attempt versus those with no suicide attempt, (ii) people with a lifetime suicide attempt versus those with lifetime suicidal ideation but no attempt, and (iii) those with a past year suicide attempt versus those with a lifetime attempt (i.e. detecting past year attempts among lifetime attempters) and (iv) people with past week suicidal ideation and a past year suicide attempt from those with past week suicidal ideation and no lifetime history of a suicide attempt (i.e. detecting past year attempts among current ideators). To assess reliability, for each participant, we separated trials by whether they were odd or even trials, calculated two D scores, and correlated the D scores across subjects. AUC-ROCs is a classification metric that demonstrates how well a diagnostic test classifies positive and negative cases. In this case, AUC-ROCs provided metrics on how well the D-IAT or D-BIAT differentiated attempters from non-attempters, attempters from ideators, and past year attempters from lifetime attempters. We tested two AUC-ROCs for each outcome; one was based on whether the D score was above or below zero (i.e., zero threshold) and the other was based on a non-zero D score (i.e., the criterion threshold could be set at any D score value). Of note, using the zero threshold distinguishes between people that have an implicit bias for death over life (i.e. positive D scores) and those that with an implicit bias for life over death (i.e. negative D scores). For all metrics, we provide 95% bootstrapped confidence intervals with 2000 samples.

The second goal of this study was to compare the D-BIAT with the original D-IAT to test whether the shorter, simpler D-BIAT could produce similar outcomes to the D-IAT. The sample of D-BIAT participants used to compare the D-BIAT and D-IATs consisted of 66% of participants that completed the D-BIAT ($n = 1161$). There were 1855 participants that completed the D-IAT, none of which were included in a prior study using a similar sample (Glenn et al., 2016). There were similar sociodemographic and clinical outcomes among the two samples that were administered the D-BIAT (i.e., for the first and second aims) and the sample administered the D-IAT (see Table 1). For Aim 2, we compared the D-BIAT and D-IAT on the same previously mentioned outcomes used to assess different scoring approaches to the D-BIAT.

A potential problem when comparing two independent samples is that demographic or self-harm variables may differ between two groups in one sample but not in the other sample. For example, compared to ideators, attempters might be significantly older in the sample administered the D-IAT but, for the sample administered the D-BIAT, attempters might be significantly younger. This type of inconsistency in group differences between samples could confound comparisons between the D-IAT and D-BIAT. Therefore, to assess the presence of this issue we tested whether, for each sociodemographic and clinical variable, the group difference effect size in one sample was significantly different than the corresponding group difference effect size in the other sample. To accomplish this, we bootstrapped Cramer's V effect sizes from chi-square tests for group differences in sex, race (White vs. non-White), ethnicity, level of education, past week suicidal ideation, and lifetime presence of a plan within each group comparison (i.e., attempters versus non-attempters, attempters versus ideators, past year attempters versus lifetime attempters), separately within each sample (and bootstrapped Cohen's D for age differences). To statistically test this, we assessed whether, for any given group difference (e.g., age differences), the effect size from one sample (e.g., D-BIAT) fell outside the 95% confidence interval of the corresponding effect size from the other sample (e.g., D-IAT). If, on a given variable, the mean effect size from one sample was within the 95% confidence interval of the effect size for the other sample, then group differences on sociodemographic variables were not significantly different between the samples.

We then compared the D-BIAT and D-IAT on the same outcomes used to assess different D-BIAT scoring approaches. For the zero-threshold D score classification, we also provide sensitivity (proportion with a prior suicide attempt correctly identified by the D-IAT) and specificity (proportion without a prior suicide attempt correctly identified by the D-IAT) metrics for the D-BIAT and D-IAT. Finally, we also provide the task duration and percentage of errors on the D-BIAT and D-IAT (for percentage of errors, we focused on the four critical blocks of the D-IAT). We hypothesized that the D-BIAT is less cognitively demanding than the D-IAT and, therefore, should result in fewer errors. A factor that makes it difficult to directly compare the D-IAT and D-BIAT error rates is that the D-IAT has three practice blocks that might reduce the D-IAT task demands compared with a D-BIAT, which has no such practice.

Results

Aim 1: Comparing different D-BIAT scoring approaches

We compared scoring six, four, and two blocks of the D-BIAT after removing either the first trial or first four trials from each block. Results revealed that using data from all six blocks after removing only the first trial from each block produced the best (although not significantly better than using two or four blocks) metrics on eight out of 13 outcomes, and was tied for the highest on another two outcomes (see Table 2). Specifically, this approach provided slightly larger effect sizes and classification metrics than the other approaches tested when testing differences between lifetime suicide attempters and non-attempters between those with a lifetime suicide attempt and lifetime suicide ideation and between

those with past week suicidal ideation and past year attempts and those with past week ideation and no lifetime attempt. The six block approach also had the highest reliability.

Notably, across the different outcomes, almost all point estimates were contained within the 95% confidence intervals for the other approaches, meaning these differences were not significantly different. The exception to this was with reliability, where the point estimate for the 6-block approach fell outside of the 95% confidence interval for most of the other approaches. However, there was no case where the 95% confidence intervals did not overlap, which would have indicated conclusive statistical significance.

Looking across the 6-, 4-, and 2-block metrics, it is notable that the largest improvements in effect size and classification metrics are observed when going from 2- to 4-block tests, with only a minimal (or no) improvement going from 4- to 6-block tests. Given that the 4-block test is only slightly inferior to the 6-block test and takes approximately 30 seconds less to complete (6-block approach: $M = 1$ min 52 sec, $SD = 33$ sec; 4-block approach: $M = 1$ min 19 sec, $SD = 27$ sec; 2-blocks: $M = 44$ sec, $SD = 21$ sec), we examined both the 6- and 4-block versions of the D-BIAT in subsequent analyses.

Aim 2: Comparing the D-BIAT and the D-IAT

We found no instances where the sample administered the D-BIAT exhibited any group differences on sociodemographic and self-harm variables that were significantly different than the corresponding group difference in the sample administered the D-IAT. Thus, differences between the D-BIAT and D-IAT are unlikely to be driven by a sociodemographic or clinical group difference present in one sample but not in the other.

We compared effect sizes and classification metrics for the 6-block and 4-block versions of the D-BIAT with those for the traditional D-IAT. Results revealed that the D-IAT outperformed both the 6- and 4-block D-BIAT on most outcomes, including Cohen's D and classification metrics distinguishing lifetime suicide attempters from non-attempters, lifetime attempters from ideators and past-week ideators with a past year attempt from past-week ideators without an attempt history (Table 3). The D-IAT also showed higher reliability than the D-BIAT. The one area where the D-BIAT outperformed the D-IAT was for the Cohen's D effect size and classification metrics distinguishing lifetime and past year attempters. Thus, overall, the D-IAT's superiority mainly involved distinguishing attempters from non-attempters. The D-IAT was only modestly better differentiating attempters from ideators and the D-BIAT produced better outcomes separating past year from lifetime attempters. Without taking into account the time to read the task instructions, the D-IAT was approximately twice as long as the 4-block version of the D-BIAT and 50 seconds longer than the 6-block version (D-IAT: $M = 2$ min 43 sec, $SD = 40$ sec; 6-block D-BIAT: $M = 1$ min 54 sec, $SD = 33$ sec; 4-block D-BIAT: $M = 1$ min 20 sec, $SD = 26$ sec). The error rates on the D-BIAT and D-IAT were similar (D-BIAT 6-blocks: $M = 11.7\%$, $SD = 6.7$; D-BIAT 4-blocks: $M = 12.1\%$, $SD = 7.6$; D-IAT 4-critical blocks: $M = 10.1\%$, $SD = 7.7$; D-IAT all 7 blocks: $M = 8.7$, $SD = 5.8$). In addition to comparisons with the D-IAT, we also replicated the prior analysis comparing different approaches to scoring the D-BIAT by showing that scoring six blocks provided modestly (but not significantly) better outcomes than scoring four blocks (see Table 3).

We note that for the comparisons of groups with past-week ideators with and without past-year suicide attempts, the D-BIAT effect size estimates and classification metrics differ greatly between the Aim 1 and Aim 2 samples (see Table 2 and 3, respectively). Furthermore, these same estimates show large confidence intervals. We examined the data and did not find any outliers or other irregularities and therefore conclude that the high degree of variability is likely due to the relatively small sample sizes of these groups.

Discussion

The current study had two aims: to compare different approaches to scoring the D-BIAT and to compare the D-BIAT with the D-IAT. The advantage of the D-BIAT is that, compared with the D-IAT, it is a shorter and simpler task, and therefore less burdensome. Therefore, it could be helpful in assessing implicit cognitions about death in patients in clinical settings and study participants in research settings. For the first aim, we found that only eliminating one (vs. four) trials from each block resulted in better outcomes and that the 6-block and 4-block tests were largely comparable (and not significantly different), with the 6-block version having minimally better outcomes and the 4-block version taking 30 fewer seconds to complete.

In comparing the D-BIAT with the more traditional D-IAT, we found that the D-IAT produces slightly (and again non-significantly) larger group difference effect sizes, improved classification metrics and reliability compared with both the 6- and 4-block versions of the D-BIAT, with the exception of distinguishing between past year and lifetime attempters, where the D-BIAT showed improved metrics. However, the D-IAT is nearly a minute longer than the 6-block version and more than twice as long as the 4-block D-BIAT. Thus, in general, across the different D-BIAT and D-IAT versions, we find a tradeoff whereby longer tests that use more data show slightly improved metrics, but take more time to complete. Given that the stated goal of this study is to find the D-BIAT version that approximates the D-IAT results while minimizing administration time, we recommend the use of the 4-block version of the D-BIAT because it offered the best trade-off of administration time to strength of effects. In instances where administration time is not a concern, we suggest the D-IAT, given that, generally, it produced the larger effects than the D-BIAT. Several findings warrant additional comment.

First, when comparing effect sizes, classification metrics and reliability among the different D-BIAT scoring approaches and the D-BIAT and D-IAT, nearly all were non-significantly different. This was the case, even comparing the 2-block version of the D-BIAT with the 6- and 4-block versions. Thus, even though we recommended the 4-block version, none were conclusively significantly superior on any outcome. Similarly, when comparing the 4-block and 6-block versions of D-BIAT and D-IAT, the confidence intervals for each outcome overlapped across the D-BIAT and D-IAT scores, indicating that none were definitively statistically different. The tasks did not result in significantly different percentage of errors. Thus, we did not find evidence supporting the idea that the D-BIAT was less cognitively demanding than the D-IAT. However, the D-IAT has three practice blocks to familiarize participants with the task, which might have reduced the task demands, whereas the D-BIAT

has no practice blocks. Thus, the fact that the D-BIAT produced a similar error rate as the D-IAT, suggests that the D-BIAT does not necessarily require a practice block.

Although effects were relatively similar across the different tests examined here, it should be noted that even small effect size differences can have large ramifications for sample size requirements for future studies. For example, based on the effect sizes found here, adequately powering a study (assuming an alpha level of .05 and a power of .80) using the D-BIAT to detect a group difference between attempters and ideators would require an additional 44 participants compared with using the D-IAT, even though the effect sizes are fairly similar (D-IAT: Cohen's $D = 0.50$; D-BIAT: Cohen's $D = 0.43$) (calculations conducted with G-Power (Faul, Erdfelder, Lang, & Buchner, 2007)). For smaller effect sizes, smaller differences can still require many more participants to achieve adequate power. Thus, even though the differences among effect sizes comparing attempters and ideators (D-BIAT; Cohen's $D = 0.33$ vs. D-IAT: Cohen's $D = 0.36$) are very small, the D-BIAT would require an additional 46 total participants to achieve the same power as the D-IAT. In light of these findings, the traditional D-IAT may be of greater use when studying a relatively small high-risk sample in which suicide risk is the primary focus. The D-BIAT may be of greater utility in large-scale studies in which suicide risk is one of several aims and the researcher is unwilling to devote more time to the assessment of suicide risk. Future research should also assess whether task duration affects participation, attrition, and adherence rates, particularly in clinical settings, where one might be willing to trade-off some statistical power if a larger percentage of people participated in the assessment. Overall, the appropriate test may depend on several considerations, including the research question, context, and sample.

Second, given that chance classification produces AUC-ROCs of 0.5 both the D-BIAT and the D-IAT provided fairly poor ability to classify members of the various groups (e.g., attempters versus non-attempters, etc.). As has been previously hypothesized, (Glenn et al., 2016), using non-zero rather than a zero threshold provided better classification metrics. For the D-IAT or D-BIAT to effectively improve risk assessment, these tests need to better distinguish among differing risk levels. It could be the case that, although the current results show that the D-IAT and D-BIAT do track both severity (suicide attempts among ideators) and recency (past year attempts among attempters), the participants in this study may have been a relatively moderate or low level of risk, causing these implicit measures to show poor classification performance.

Interestingly, we found that, although the D-BIAT and D-IAT provided similar AUCs for ROC, they did so with different ratios of sensitivity and specificity. Across each group difference, the D-IAT provided high specificity and low sensitivity, whereas the D-BIAT was more balanced between specificity and sensitivity. This suggests that the D-BIAT and D-IAT may tap different aspects of implicit cognitions about death and may each account for unique variance.

Most prior studies of the D-IAT (and in suicide research in general) have compared groups based on lifetime suicidal thoughts or behaviors. More important in clinical settings is the ability to distinguish those who are at acute (not lifetime) risk of such outcomes. In the current study, we examined whether the D-IAT and D-BIAT could identify those with past

year attempts among lifetime attempters and among past week ideators and the results were mixed. Specifically, the D-BIAT was better at differentiating past year attempters from those with a more distant history of attempt, whereas the D-IAT was better at differentiating those with a past-year attempt amongst past-week ideators. Ultimately, additional prospective studies among high-risk populations are required to continue to assess whether implicit measures can be helpful in predicting suicidal behaviors.

Ultimately, additional prospective studies (Barnes et al., 2017; Nock et al., 2010; Nock & Banaji, 2007; Randall et al., 2013) among high-risk populations are required to continue to assess whether implicit measures can be helpful in predicting suicidal behaviors. In addition to prospective studies, future studies can also assess whether other manipulations might improve the power of the D-IAT and D-BIAT to classify risk-level, such as using personally relevant stimuli (e.g., words related to pills or medication for people that have previously attempted suicide by overdosing on pills), inducing a negative mood prior to the test, or first reminding participants of their suicidal history.

Prior studies have sought to uncover whether group differences on the D-IAT suggest that suicidal participants have a lower implicit bias for life or a higher implicit bias for death, compared with controls (Harrison, Stritzke, Fay, Ellison, & Hudaib, 2014). In general, suicidal groups show, on average, negative D scores, indicating reduced implicit bias for life, consistent with prior interpretations (Harrison et al., 2014). However, the D-IAT presents two concept categories (i.e. “life” and “death”) with two attributes (i.e. “me” and “not-me”) meaning average RT and D scores represent some combination of implicit bias for the concept paired with “me” and the concept paired with “not-me.” The D-BIAT only presents “me” with the concept categories (i.e. “not-me” is never presented) meaning the D-BIAT may represent a “purer” representation of the relationship between oneself and life or death. Interestingly, past-year attempters (both with and without past-week ideation) show positive score D scores on the D-BIAT (see top right half of Figure 1), suggesting that they may have a stronger association with themselves and death rather than just a weaker association between themselves and life.

These results highlight several important directions for future research. First, as mentioned, there is a crucial question of how to maximize the clinical utility of the D-BIAT or D-IAT. This includes practical concerns, such as the method in which the task is delivered (e.g., whether results from a D-IAT administered on a smartphone or tablet is comparable to results from laptop administration) and whether the length of the task affects participation rates. Second, because the D-BIAT and D-IAT were administered in separate populations in this study, we could not assess the correlation between the two measures. Given that the measures provide different balances of specificity and sensitivity to achieve similar AUCs, they may be helpful if used in conjunction. Third, the short duration of the D-BIAT may be helpful if clinicians wanted repeated assessments, for example, on a patient’s smartphone after hospital discharge. However, little work has assessed repeated assessment of the D-IAT or D-BIAT, and whether it increases or decreases predictive validity. Fourth, an important direction is research into using the D-BIAT or D-IAT in conjunction with other metrics of risk, such as medical records and self-report measures, along with advanced statistical procedures (Barak-Corren et al., 2017), to improve prediction. Fifth, recent reports suggest

that most people that die by suicide have some contact with a health care provider, if not mental health professionals (Ahmedani et al., 2014). Therefore, another possibility is that the D-BIAT be included in a mental health battery screener given to all patients when they visit their primary care physician. However, the threshold cutoff for “high-risk” will have to be determined and, like any detection tool, will be based on many considerations, including the consequences of acting on potential false-positives versus false-negatives.

Ultimately, the D-IAT or D-BIAT is unlikely to be a panacea for suicide-risk prediction and prevention but there are several reasons why it may be a helpful tool. First, given that many suicide attempts and deaths occur among people not currently in mental health care, it is important the D-IAT be able to detect risk outside of acute clinical settings. The current study, along with a prior study with more than 6,000 web-based participants (Glenn et al., 2016), suggest that the D-IAT can distinguish among suicidal and non-suicidal groups in non-clinical settings (assuming that web-based participants are unlikely to be completing the test in clinical care). Second, given the widespread reach of the internet (88% of Americans have internet access (Pew Research Center, 2017)) along with minimal computer requirements and the automatically returned scores, the D-IAT can easily be used across many clinical and non-clinical settings. Not all people at risk of suicide have internet access, for example homeless people may lack internet access, but the fact that the D-IAT can be administered via the web means that the possible scale for administration is quite broad. Furthermore, the fact that the score is reaction time differences between two conditions means that possible confounds like computer speed or individual differences are controlled for. Third, it is currently unclear how the D-IAT should be used in conjunction with other assessments, including self-report measures and clinicians’ expertise to determine who is at greatest risk of suicidal behaviors and needs to be hospitalized. This is a crucially important question and one that we intend to begin to address in the near future.

There are several limitations to the current study. First, all suicidal thoughts and behaviors were measured with self-report. The SITBI is more in-depth and reliable than several measures of suicide risk used in other epidemiological work (i.e., single items) (Nock, Borges, Bromet, Cha, et al., 2008) and similar to other widely used measures, such as the Columbia Suicide Severity Rating Scale (Posner et al., 2011). Still the reliance on self-report has the potential to introduce error through either misinterpretation (Millner, Lee, & Nock, 2015) or the motivation to conceal suicide risk. Second, all participation in the study was anonymous, which may have allowed participants to take the D-IAT multiple times. Given the large sample size (>3,000), it is unlikely that this would have significantly altered results. Third, all data are cross-sectional. Therefore, these results demonstrate that the D-BIAT and D-IAT can predict group status (i.e., history of suicidal behaviors), but can make no claims about predicting future suicidal behaviors. Fourth, the vast majority of analyses focus on suicidal behavior rather than thoughts. While behaviors are of the greatest interest to clinicians and likely to produce the largest between-group effects, testing the D-BIAT for suicidal ideation and suicidal plans could also be useful. Expanding the D-BIAT to predict suicidal thoughts along with prospective prediction of suicide attempts could prompt novel research questions and allow for the potential early detection of risk. Fifth, we utilized bootstrapped confidence intervals to determine whether differences between the D-BIAT and D-IAT were significant. However, the statistical significance of several outcomes, such as

past year attempt vs lifetime attempt, were left inconclusive (i.e., point estimates fell outside of the confidence intervals but the intervals overlapped to some degree, limiting our ability to draw conclusions about these particular outcomes).

Finally, the study sample is not representative of the general population and reported higher rates of suicidal thoughts and behaviors than prior community studies. For example, in analyses for the first aim, 32% reported a lifetime suicide attempt, whereas, the estimated prevalence in the general population is 2.7% (Nock, Borges, Bromet, Alonso, et al., 2008). This raises the possibility that the participants self-selected into the study potentially due to a personal interest in the topic.

Overall, we find the D-BIAT produces effects only slightly inferior to the full D-IAT and takes about the half the time. For clinical settings or other circumstances where time is limited but large samples can be acquired, the D-BIAT may be the preferred test to obtain implicit self-harm biases.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported in part by NIMH grants (R34MH106770 and R01MH113752) to B. Teachman and by the Norlien Foundation and John D. and Catherine T. MacArthur Foundation to Matthew K. Nock. Note, Teachman has a significant financial interest in Project Implicit, Inc., which provided services in support of this project under contract with the University of Virginia. We thank the Project Implicit Mental Health collaborators, including Brian Nosek, Mahzarin Banaji, and Tony Greenwald. We would also like to thank Frederick Smyth, Emily Umansky, and Nauder Namaky for their technical assistance.

References

- Ahmedani BK, Simon GE, Stewart C, Beck A, Waitzfelder BE, Rossom R, Solberg LI. Health Care Contacts in the Year Before Suicide Death. *Journal of General Internal Medicine*. 2014; 29(6): 870.doi: 10.1007/s11606-014-2767-3
- Barak-Corren Y, Castro VM, Javitt S, Hoffnagle AG, Dai Y, Perlis RH, Reis BY. Predicting suicidal behavior from longitudinal electronic health records. *The American Journal of Psychiatry*. 2017; 174(2):154–162. DOI: 10.1176/appi.ajp.2016.16010077 [PubMed: 27609239]
- Barnes SM, Bahraini NH, Forster JE, Stearns-Yoder KA, Hostetter TA, Smith G, Nock MK. Moving Beyond Self-Report: Implicit Associations about Death/Life Prospectively Predict Suicidal Behavior among Veterans. *Suicide and Life-Threatening Behavior*. 2017; 47(1):67–77. DOI: 10.1111/sltb.12265 [PubMed: 27387836]
- Berman AL. Risk Factors Proximate to Suicide and Suicide Risk Assessment in the Context of Denied Suicide Ideation. *Suicide and Life-Threatening behavior*. 2017.
- Busch KA, Fawcett J, Jacobs DG. Clinical Correlates of Inpatient Suicide. *The Journal of Clinical Psychiatry*. 2003; 64(1):14–19.
- Carter G, Milner A, McGill K, Pirkis J, Kapur N, Spittal MJ. Predicting suicidal behaviours using clinical instruments: systematic review and meta-analysis of positive predictive values for risk scales. *The British Journal of Psychiatry*. 2017; 210(6):387–395. DOI: 10.1192/bjp.bp.116.182717 [PubMed: 28302700]
- Chung DT, Ryan CJ, Hadzi-Pavlovic D, Singh SP, Stanton C, Large MM. Suicide Rates After Discharge From Psychiatric Facilities: A Systematic Review and Meta-analysis. *JAMA Psychiatry*. 2017.

- Curtin SC, Warner M, Hedegaard H, et al. Increase in suicide in the United States, 1999–2014. NCHS Data Brief. 2016; 241:1–8.
- Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior Research Methods. 2007; 39(2):175–191. DOI: 10.3758/BF03193146 [PubMed: 17695343]
- Fischer G, Ameis N, Parzer P, Plener PL, Groschwitz R, Vonderlin E, Kaess M. The German version of the Self-Injurious Thoughts and Behaviors Interview (SITBI-G): a tool to assess non-suicidal self-injury and suicidal behavior disorder. BMC Psychiatry. 2014; 14(1):1.
- García-Nieto R, Blasco-Fontecilla H, Paz Yepes M, Baca-García E. Translation and validation of the “Self-Injurious Thoughts and Behaviors Interview” in a Spanish population with suicidal behaviour. Revista De Psiquiatría Y Salud Mental. 2013; 6(3):101–108. DOI: 10.1016/j.rpsm.2012.07.001 [PubMed: 23084799]
- Glenn JJ, Werntz AJ, Slama SJK, Steinman SA, Teachman BA, Nock MK. Suicide and Self-Injury-Related Implicit Cognition: A Large-Scale Examination and Replication. Journal of Abnormal Psychology. 2016.
- Greenwald AG, Poehlman TA, Uhlmann EL, Banaji MR. Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. Journal of Personality and Social Psychology. 2009; 97(1):17–41. DOI: 10.1037/a0015575 [PubMed: 19586237]
- Harrison DP, Stritzke WGK, Fay N, Ellison TM, Hudaib AR. Probing the implicit suicidal mind: Does the Death/Suicide Implicit Association Test reveal a desire to die, or a diminished desire to live? Psychological Assessment. 2014; 26(3):831–840. DOI: 10.1037/pas0000001 [PubMed: 24611787]
- Keilp JG, Gorlyn M, Russell M, Oquendo MA, Burke AK, Harkavy-Friedman J, Mann JJ. Neuropsychological function and suicidal behavior: attention control, memory and executive dysfunction in suicide attempt. Psychological Medicine. 2013; 43(03):539–551. DOI: 10.1017/S0033291712001419 [PubMed: 22781400]
- Keilp JG, Sackeim HA, Brodsky BS, Oquendo MA, Malone KM, Mann JJ. Neuropsychological dysfunction in depressed suicide attempters. American Journal of Psychiatry. 2001; 158(5):735–741. [PubMed: 11329395]
- Large M, Kaneson M, Myles N, Myles H, Gunaratne P, Ryan C. Meta-Analysis of Longitudinal Cohort Studies of Suicide Risk Assessment among Psychiatric Patients: Heterogeneity in Results and Lack of Improvement over Time. PLOS ONE. 2016; 11(6):e0156322.doi: 10.1371/journal.pone.0156322 [PubMed: 27285387]
- Millner AJ, Lee MD, Nock MK. Single-item measurement of suicidal behaviors: Validity and consequences of misclassification. PLoS ONE. 2015; 10(10):e0141606.doi: 10.1371/journal.pone.0141606 [PubMed: 26496707]
- Nock MK, Banaji MR. Prediction of suicide ideation and attempts among adolescents using a brief performance-based test. Journal of Consulting and Clinical Psychology. 2007; 75(5):707–715. DOI: 10.1037/0022-006X.75.5.707 [PubMed: 17907852]
- Nock MK, Borges G, Bromet EJ, Alonso J, Angermeyer M, Beautrais A, Williams D. Cross-national prevalence and risk factors for suicidal ideation, plans and attempts. The British Journal of Psychiatry. 2008; 192(2):98–105. DOI: 10.1192/bjp.bp.107.040113 [PubMed: 18245022]
- Nock MK, Borges G, Bromet EJ, Cha CB, Kessler RC, Lee S. Suicide and suicidal behavior. Epidemiologic Reviews. 2008; 30(1):133–154. DOI: 10.1093/epirev/mxn002 [PubMed: 18653727]
- Nock MK, Holmberg EB, Photos VI, Michel BD. Self-Injurious Thoughts and Behaviors Interview: Development, reliability, and validity in an adolescent sample. Psychological Assessment. 2007; 19(3):309–317. DOI: 10.1037/1040-3590.19.3.309 [PubMed: 17845122]
- Nock MK, Park JM, Finn CT, Deliberto TL, Dour HJ, Banaji MR. Measuring the suicidal mind: Implicit cognition predicts suicidal behavior. Psychological Science. 2010; 21(4):511–517. DOI: 10.1177/0956797610364762 [PubMed: 20424092]
- Nosek BA, Banaji MR, Greenwald AG. Harvesting implicit group attitudes and beliefs from a demonstration web site. Group Dynamics: Theory, Research, and Practice. 2002; 6(1):101–115. DOI: 10.1037/1089-2699.6.1.101

- Nosek BA, Bar-Anan Y, Sriram N, Axt J, Greenwald AG. Understanding and Using the Brief Implicit Association Test: Recommended Scoring Procedures. *PLOS ONE*. 2014; 9(12):e110938.doi: 10.1371/journal.pone.0110938 [PubMed: 25485938]
- Nosek BA, Greenwald AG, Banaji MR. Understanding and Using the Implicit Association Test: II. Method Variables and Construct Validity. *Personality and Social Psychology Bulletin*. 2005; 31(2): 166–180. DOI: 10.1177/0146167204271418 [PubMed: 15619590]
- Nosek BA, Smyth FL, Hansen JJ, Devos T, Lindner NM, Ranganath KA, Banaji MR. Pervasiveness and correlates of implicit attitudes and stereotypes. *European Review of Social Psychology*. 2007; 18(1):36–88. DOI: 10.1080/10463280701489053
- Office of the Surgeon General (US), & National Action Alliance for Suicide Prevention (US). 2012 National Strategy for Suicide Prevention: Goals and Objectives for Action: A Report of the US Surgeon General and of the National Action Alliance for Suicide Prevention. Washington (DC): US Department of Health & Human Services (US); 2012. Retrieved from <http://www.ncbi.nlm.nih.gov/books/NBK109917/>
- Petróczi A, Aidman EV, Hussain I, Deshmukh N, Nepusz T, Uvacsek M, Naughton DP. Virtue or Pretense? Looking behind Self-Declared Innocence in Doping. *PLOS ONE*. 2010; 5(5):e10457.doi: 10.1371/journal.pone.0010457 [PubMed: 20463978]
- Pew Research Center. Record shares of Americans now own smartphones, have home broadband. Washington, D.C.: 2017. Retrieved from <http://www.pewresearch.org/fact-tank/2017/01/12/evolution-of-technology/>
- Posner K, Brown GK, Stanley B, Brent DA, Yershova KV, Oquendo MA, Shen S. The Columbia–Suicide Severity Rating Scale: Initial validity and internal consistency findings from three multisite studies with adolescents and adults. *American Journal of Psychiatry*. 2011; 168(12): 1266–1277. [PubMed: 22193671]
- Randall JR, Rowe BH, Dong KA, Nock MK, Colman I. Assessment of self-harm risk using implicit thoughts. *Psychological Assessment*. 2013; 25(3):714–721. DOI: 10.1037/a0032391 [PubMed: 23647043]
- Ribeiro JD, Gutierrez PM, Joiner TE, Kessler RC, Petukhova MV, Sampson NA, Nock MK. Health care contact and suicide risk documentation prior to suicide death: Results from the Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS). *Journal of Consulting and Clinical Psychology*. 2017; 85(4):403–408. DOI: 10.1037/ccp0000178 [PubMed: 28333538]
- Shapiro DN, Chandler J, Mueller PA. Using Mechanical Turk to Study Clinical Populations. *Clinical Psychological Science*. 2013; 1(2):213–220. DOI: 10.1177/2167702612469015
- World Health Organization. Suicide. WHO; 2017. Retrieved from <http://www.who.int/mediacentre/factsheets/fs398/en/>
- Yang J, Shi Y, Luo YLL, Shi J, Cai H. The Brief Implicit Association Test is Valid: Experimental Evidence. *Social Cognition*. 2014; 32(5):449–465. DOI: 10.1521/soco.2014.32.5.449

Public Significance Statement

The Death Implicit Association Test (D-IAT) has revealed that people with a history of suicidal behaviors unconsciously associate themselves with the concept of death more than people without such a history. Here we tested a brief version of the D-IAT, which may be more applicable to clinical settings because it takes approximately half the time to complete, and find that it shows similar effects as the full D-IAT.

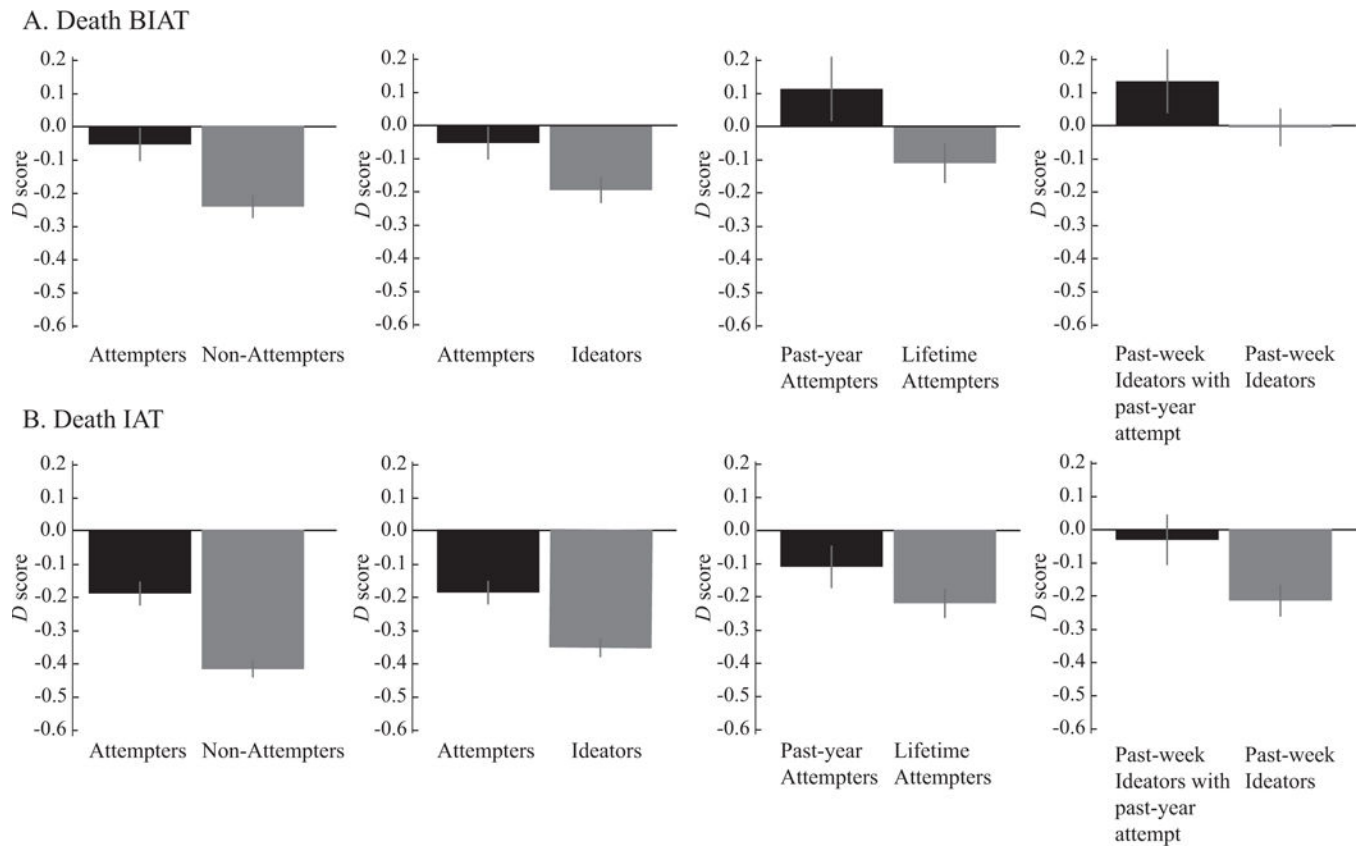


Figure 1. Group differences for the Death BIAT (4-block version) and Death IAT. Error bars represent 95% confidence intervals.

Table 1

Sociodemographic and clinical characteristics

	Aim 1 Death BIAT n = 570	Aim 2 Death BIAT n = 1161	Aim 2 Death IAT n = 1855	
Mean (<i>SD</i>) age in years	41.9 (10.8)	41.9 (10.8)	41.1 (9.9)	F(2,3583) = 2.5
Sex (% Female)	69.3	67.6	69.6	$\chi^2(2) = 4.9$
Citizenship (%)				$\chi^2(4) = 16.9$
U.S.	68.4	62.2	59.3	
Non-U.S.	28.8	35.4	37.6	
Did not report	2.8	2.4	3.1	
Ethnicity (%)				$\chi^2(4) = 5.0$
Non-Hispanic	79.8	81.7	79.0	
Hispanic	9.6	8.2	10.1	
Unknown or Did Not Report	10.5	10.2	10.9	
Race (%)				$\chi^2(12) = 14.5$
Caucasian	69.3	72.2	69.3	
Asian	7.2	7.3	7.3	
African American	3	3.3	3.1	
American Indian/Alaska Native	0.2	0.5	0.5	
Native Hawaiian/Pacific Islander	0.4	0.5	0.2	
Multiracial	7.7	7.5	9.5	
Other or Unknown	7.2	5.3	6.6	
Did Not Report	5.1	3.4	3.6	
Education (%)				$\chi^2(4) = 10.0$
Graduate Degree	15.8	16	14.7	
Bachelor's Degree	22.3	23.5	23	
Some college	40.7	37.3	37	
High school	10.5	6.3	7.8	
Less than high school	7.2	3.7	4.9	
Did not report	3.5	16	14.7	
Lifetime history of self-harm (%)				
Non-suicidal self-injury (NSSI)	62.5	64.2	63.6	$\chi^2(2) = 0.5$
Suicide ideation	84.9	84.5	82.8	$\chi^2(2) = 2.5$
Suicide plan	39.6	42.2	41	$\chi^2(2) = 1.2$
Suicide attempt	31.9	32	32.4	$\chi^2(2) = 0.02$
Any self-harm thought/behavior	87.0	88.5	88.0	$\chi^2(2) = 0.8$

Table 2
Comparisons between different approaches to scoring the Death Brief Implicit Association Test (BIAT)

		Remove 1st trial			Remove 1st 4 trials		
		6 Blocks	4 Blocks	2 Blocks	6 Block	4 Block	2 Block
LT-SA vs LT-No SA	Cohen's <i>D</i>	0.43 (0.25 - 0.62)	0.39 (0.21 - 0.58)	0.30 (0.12 - 0.47)	0.41 (0.23 - 0.60)	0.39 (0.21 - 0.58)	0.32 (0.14 - 0.49)
	AUC-ROC non-zero threshold	0.62 (0.56 - 0.67)	0.61 (0.55 - 0.66)	0.58 (0.52 - 0.63)	0.61 (0.55 - 0.66)	0.60 (0.55 - 0.65)	0.58 (0.53 - 0.64)
	AUC-ROC zero threshold	0.59 (0.55 - 0.64)	0.59 (0.55 - 0.63)	0.56 (0.51 - 0.60)	0.57 (0.53 - 0.62)	0.59 (0.54 - 0.63)	0.56 (0.52 - 0.60)
LT-SA vs LT-SI	Cohen's <i>D</i>	0.31 (0.13 - 0.52)	0.30 (0.11 - 0.50)	0.22 (0.03 - 0.41)	0.29 (0.10 - 0.50)	0.30 (0.10 - 0.49)	0.24 (0.05 - 0.43)
	AUC-ROC non-zero threshold	0.58 (0.53 - 0.64)	0.58 (0.53 - 0.63)	0.56 (0.51 - 0.61)	0.57 (0.52 - 0.63)	0.58 (0.53 - 0.63)	0.56 (0.51 - 0.62)
	AUC-ROC zero threshold	0.57 (0.53 - 0.62)	0.57 (0.53 - 0.62)	0.54 (0.49 - 0.58)	0.55 (0.50 - 0.59)	0.57 (0.52 - 0.61)	0.55 (0.50 - 0.59)
PY-SA vs. LT-SA	Cohen's <i>D</i>	0.55 (0.20 - 0.87)	0.51 (0.26 - 0.95)	0.51 (0.26 - 0.95)	0.54 (0.21 - 0.87)	0.50 (0.27 - .97)	0.50 (0.27 - .97)
	AUC-ROC non-zero threshold	0.65 (0.57 - 0.74)	0.66 (0.58 - 0.75)	0.64 (0.55 - 0.73)	0.65 (0.58 - 0.74)	0.66 (0.58 - 0.76)	0.64 (0.55 - 0.73)
	AUC-ROC zero threshold	0.61 (0.52 - 0.69)	0.62 (0.54 - 0.70)	0.61 (0.53 - 0.69)	0.60 (0.52 - 0.68)	0.62 (0.54 - 0.71)	0.62 (0.53 - 0.70)
PY-SA vs. PW-SI	Cohen's <i>D</i>	0.74 (0.33 - 1.19)	0.68 (0.29 - 1.11)	0.55 (0.18 - 0.96)	0.70 (0.29 - 1.17)	0.65 (0.25 - 1.10)	0.56 (0.17 - 0.98)
	AUC-ROC non-zero threshold	0.72 (0.62 - 0.82)	0.70 (0.60 - 0.79)	0.65 (0.55 - 0.75)	0.71 (0.61 - 0.81)	0.69 (0.59 - 0.79)	0.66 (0.55 - 0.76)
	AUC-ROC zero threshold	0.67 (0.58 - 0.75)	0.64 (0.55 - 0.72)	0.62 (0.53 - 0.71)	0.62 (0.54 - 0.71)	0.64 (0.55 - 0.72)	0.63 (0.54 - 0.72)
Odd-Even Reliability		0.73 (0.69 - 0.77)	0.68 (0.63 - 0.73)	0.64 (0.58 - 0.68)	0.68 (0.63 - 0.72)	0.64 (0.59 - 0.69)	0.59 (0.53 - 0.64)

Note: LT-SA vs LT-No SA = Lifetime suicide attempt ($n=182$) vs. No lifetime suicide attempt ($n=388$); LT-SA vs. LT-SI = Lifetime suicide attempt ($n=182$) vs. Lifetime suicidal ideation (with no history of a suicide attempt) ($n=302$); PY-SA vs. LT-SA = Past year suicide attempt ($n=37$) vs. Lifetime (but not past year) suicide attempt ($n=134$); PY-SA vs. PW-SI = Past week suicide ideation and past year suicide attempt ($n=37$) vs. past week suicide ideation (with no lifetime history of attempts) ($n=121$). AUC-ROC = Area under the curve for the receiver operating characteristic. Parentheses contain 95% bootstrapped confidence intervals.

Table 3

Comparing the Death Brief Implicit Association Test (BIAT) and the Death Implicit Association Test (IAT)

		Death BIAT <i>n</i> =1161		Death IAT <i>n</i> =1855
		6 Blocks	4 Blocks	
LT-SA vs LT-No SA	Cohen's <i>D</i>	0.43 (0.30 – 0.55)	0.38 (0.25 – 0.51)	0.50 (0.40 – 0.60)
	AUC-ROC non-zero threshold	0.62 (0.59 – 0.66)	0.61 (0.57 – 0.64)	0.64 (0.61 – 0.66)
	AUC-ROC zero threshold	0.58 (0.55 – 0.61)	0.57 (0.55 – 0.61)	0.58 (0.56 – 0.60)
	Sensitivity zero threshold	0.46	0.45	0.34
	Specificity zero threshold	0.70	0.69	0.82
LT-SA vs. LT-SI	Cohen's <i>D</i>	0.33 (0.19 – 0.46)	0.29 (0.18 – 0.45)	0.36 (0.26 – 0.46)
	AUC-ROC non-zero threshold	0.59 (0.56 – 0.63)	0.58 (0.55 – 0.62)	0.60 (0.57 – 0.63)
	AUC-ROC zero threshold	0.56 (0.53 – 0.60)	0.56 (0.53 – 0.59)	0.56 (0.54 – 0.59)
	Sensitivity zero threshold	0.47	0.46	0.35
	Specificity zero threshold	0.65	0.66	0.78
PY-SA vs. LT-SA	Cohen's <i>D</i>	0.46 (0.23 – 0.69)	0.47 (0.19 – 0.68)	0.20 (0.02 – 0.36)
	AUC-ROC non-zero threshold	0.63 (0.57 – 0.69)	0.63 (0.56 – 0.69)	0.56 (0.51 – 0.61)
	AUC-ROC zero threshold	0.57 (0.52 – 0.63)	0.60 (0.52 – 0.63)	0.56 (0.52 – 0.60)
	Sensitivity zero threshold	0.59	0.62	0.43
	Specificity zero threshold	0.57	0.60	0.69
PY-SA vs. PW-SI	Cohen's <i>D</i>	0.33 (0.08 – 0.70)	0.33 (0.07 – 0.71)	0.38 (0.19 – 0.37)
	AUC-ROC non-zero threshold	0.59 (0.52 – 0.67)	0.59 (0.52 – 0.67)	0.61 (0.55 – 0.66)
	AUC-ROC zero threshold	0.55 (0.49 – 0.61)	0.58 (0.52 – 0.64)	0.58 (0.54 – 0.63)
	Sensitivity zero threshold	0.47	0.46	0.35
	Specificity zero threshold	0.65	0.66	0.78
Odd-Even Reliability		0.74 (0.71 – 0.77)	0.71 (0.67 – 0.74)	0.76 (0.74 – 0.78)

Note: LT-SA vs LT-No SA = Lifetime suicide attempt (BIAT: *n*=372; IAT: *n*=598) vs. No lifetime suicide attempt (BIAT: *n*=789; IAT: *n*=1257); LT-SA vs. LT-SI = Lifetime suicide attempt (BIAT *n*=598; IAT *n*=605) vs. Lifetime suicidal ideation (BIAT *n*=613; IAT *n*=941); PY-SA vs. LT-SA = Past year suicide attempt (BIAT *n*=103; IAT *n*=198) vs. Lifetime (but not past year) suicide attempt (BIAT *n*=269; IAT *n*=400); PY-SA vs. PW-SI = Past year suicide ideation and past year suicide attempt (BIAT *n*=76; IAT *n*=147) vs. past week suicide ideation (with no lifetime history of attempts) (BIAT *n*=218, IAT *n*=370). AUC-ROC = Area under the curve for the receiver operating characteristic. Parentheses contain 95% bootstrapped confidence intervals.