The Baby or the Bath Water?

Lessons Learned from the National Action Alliance for Suicide Prevention Research Prioritization Task Force Literature Review

Sherry Davis Molock, PhD, Janet M. Heekin, MLS, Samantha G. Matlin, PhD,
Crystal L. Barksdale, PhD, Ekwenzi Gray, PhD, Chelsea L. Booth, PhD

Context: The Research Prioritization Task Force of the National Action Alliance for Suicide Prevention conducted a comprehensive literature review of suicide prevention/intervention trials to assess the quality of the scientific evidence.

Evidence acquisition: A literature "review of reviews" was conducted by searching the most widely used databases for mental health and public health research. The quality of the reviews was evaluated using the Revised Assessment of Multiple Systematic Reviews system; the quality of the scientific evidence for the suicide preventions/interventions was assessed using U.S. Preventive Services Task Force criteria. The reviews were limited to peer-reviewed publications with human subjects published in English.

Evidence synthesis: Ninety-eight systematic reviews and 45 primary sources on suicide prevention/interventions published between January 2000 and September 2012 were evaluated. The results suggest that the quality of both the systematic reviews and the scientific evidence for suicide preventions/interventions were mixed. The majority of the systematic reviews and prevention/interventions were evaluated as fair to poor in quality.

Conclusions: There are many promising suicide prevention/intervention trials, but research findings are often inconclusive because of methodologic problems. Methodologic problems across systematic reviews include not conducting hand searches, not surveying gray literature, and being unable to aggregate data across studies. Methodologic problems with the scientific quality of the prevention/intervention trials include paucity of information on sample demographic characteristics, poorly defined outcomes, and excluding actively suicidal participants. Suggestions for ways to improve the quality of the systematic reviews and suicide preventions/interventions are provided.
(Am J Prev Med 2014;47(3S2):S115–S121) © 2014 American Journal of Preventive Medicine. All rights reserved.

Introduction

Globally, suicide represents an important public health concern in many countries. Each year, nearly 1 million people die by suicide, which translates into a global mortality rate of 16/100,000 deaths each year. Suicide death rates have increased by 60% in the last 45 years.[1] The most recent data indicate that suicide is the tenth-leading cause of death in the U.S.; there are more than 38,000 suicide deaths in the U.S. each year.[2] Suicide accounts for 1.5% of the global burden of disease, which represents 20 million years of health lost because of death and disability.[3]

The Research Prioritization Task Force (RPTF) of the National Action Alliance for Suicide Prevention (Action Alliance) conducted a comprehensive review of the literature on suicide preventions/interventions to help inform the prioritization of the research agenda for suicide prevention. The RPTF felt it was important to understand the current state of the science on suicide prevention to provide some context for future directions in suicide prevention research. Although there have been several systematic reviews on suicide preventions/interventions since the first U.S. National Strategy for Suicide Prevention[4] highlighted the need for more effective

From the Department of Psychology (Molock), George Washington University; Counseling Center (Gray), Howard University, Washington, District of Columbia; Substance Abuse and Mental Health Services Administration (Booth, Heekin), Bethesda, SRA International (Barksdale), Rockville, Maryland; and the Department of Behavioral Health and Intellectual Disability (Matlin), Philadelphia, Pennsylvania

Address correspondence to: Sherry D. Molock, PhD, Department of Psychology, George Washington University, Washington DC 20052. E-mail: smolock@gwu.edu.

suicide preventions/interventions, very few reviews have conducted a comprehensive review across different types of treatments (e.g., biological, psychosocial, community-based) with heterogeneous populations. Thus, this review attempted to focus on demographic groups with the greatest burden (e.g., elderly, individuals with substance abuse) to allow us to focus on (1) impact—how we can significantly reduce suicide rates; (2) "boundaried" populations—where at-risk groups typically receive treatment; (3) underserved populations—what particular groups are underserved in typical systems of care (e.g., specialty mental health services), where providers are most likely to reach such groups and provide quality care; and (4) pointing out gaps in treatment and identifying innovative ways to provide treatment.

The RPTF also searched gray literature and conducted hand searches of some of the premier peer-reviewed journals because recent literature on the methodology used in systematic reviews suggests that these two techniques can help to minimize selection, location, and publication bias.[5,6] Gray literature refers to papers, reports, and other documents that are not distributed or indexed by commercial publishers. Gray literature needs to be carefully scrutinized because it is not peer reviewed, but it can help to reduce publication bias as published studies in medicine and social sciences tend to only publish positive findings, which may result in inflated treatment and intervention effect sizes.[7] Systematic reviews typically rely on electronic sources from established databases (e.g., MEDLINE), which can result in location and selection bias. For example, studies that are incorrectly marked may be missed in electronic searches. Hand searching can help minimize these biases because it involves manually searching the entire content of a journal to identify all the research on a particular topic, whether it appears in an article, abstract, brief reports, or editorial comments (thecochranelibrary.com/).

Evidence Acquisition

The RPTF literature review team initially searched the Cochrane Library to identify relevant systematic reviews of suicide prevention trials. The Cochrane Library—an internationally recognized resource in evidence-based health practice research—is a collection of databases in human health care and health policy and includes the Database of Systematic Reviews, which contains systematic reviews and meta-analyses that summarize and interpret the results of intervention trials (thecochranelibrary.com/). The RPTF literature review team also searched the Cochrane Central Register of Controlled Trials (CENTRAL), which is derived from regular systematic searches of bibliographic databases including

MEDLINE; Excerpta Medica Database (EMBASE); PsycINFO; and Science Citation Index. As a crosscheck for all relevant literature, the team searched PubMed; EMBASE; PsycINFO; as well as the Web of Science (includes the Science Citation Index and Social Science Citation Index); Scopus; and the Cumulative Index to Nursing and Allied Health Literature (CINAHL). For comprehensive database searches, the key concepts were identified as *suicide, suicide attempts, suicidal ideation, suicide risk, self-harm, self-injurious behavior, intervention, prevention, systematic review, meta-analysis, controlled trials, cohort studies,* and *case control studies.* Relevant Medical Subject Headings (MeSH) were identified using the U.S. National Library of Medicine's MeSH Browser[8] and relevant keywords for searching the titles and abstracts of articles.

Search filters containing MeSH and keywords were iteratively developed and tested in PubMed, and subsequently adapted to search other databases. Articles were limited to peer-reviewed articles involving human subjects between January 1, 2000, and September 30, 2012; literature alerts were set up to identify new studies from October 1, 2012, forward. Additionally, the RPTF team searched for gray literature from relevant organizations or their websites and through consultation with key stakeholders and content experts. Additional citations were sought through the reference lists of relevant documents, as well as hand searching for primary studies in peer-reviewed journals that were targeted because they publish the highest percentage of empirical work on suicidology. These journals included *Suicide & Life Threatening Behavior, Crisis—The Journal of Crisis Intervention and Suicide Prevention, British Journal of Psychiatry, Journal of Affective Disorders, Acta Psychiatrica Scandinavica, Archives of Suicide Research,* and the *American Journal of Public Health.* The RPTF did not further search or review the references in the articles in the hand-searched material. Further details of the literature search protocol are available from the authors.

Abstracts were screened for relevance by doctoral-level researchers who were trained to conduct critical appraisals using the guidelines set forth by the Oxford Centre for Evidence Based Medicine (www.cebm.net/); the 1991 Oxman and Guyatt guidelines for systematic reviews[9]; the 2007 U.S. Preventive Services Task Force (USPSTF)[10] criteria for quality ratings of evidence-based interventions; and the 2006 National Institute of Clinical Excellence's (NICE)[11] guidelines for critical appraisal of evidence. Although there is some overlap in the three rating systems, they also have unique features that would allow the RPTF to conduct a more comprehensive review; the Assessment of Multiple Systematic Reviews (AMSTAR) and NICE guidelines address the methods

used to conduct systematic reviews, while the USPSTF provides a detailed tool to assess the scientific quality of other types of studies (e.g., RCTs). If an article met the selection criteria, the full paper was reviewed. Data were extracted using an extraction template developed for this

study and checked for completeness and accuracy by members of the review team (Table 1). The quality of the systematic reviews was evaluated using the Revised AMSTAR (R-AMSTAR) system, a widely used assessment tool that allows one to quantify the evaluation of

Table 1. Template for data extraction

Data fields
Quality of review (AMSTAR and USPSTF)   • A prior design used   • Duplicate study selection and data extraction   • Comprehensive literature search performed   • Status of publication (gray literature) used as inclusion criteria   • List of included and excluded studies provided   • Characteristics of included studies provided   • Scientific quality of included studies assessed and documented   • Scientific quality of included studies used appropriately in formulating conclusions   • Appropriate methods used to combine findings of studies   • Assessed likelihood of publication bias   • Conflict of interest stated   • Other criteria (assessed but not included in scoring):   ○ Inclusion of international and domestic peer-reviewed journals   ○ Search terms included   ○ Validity criteria reported   ○ Conclusions of review are warranted given evaluation of studies
Quality of scientific evidence (USPSTF criteria)   • Good: Includes well-designed, well-conducted studies in representative populations that directly assess effects on health outcomes   • Fair: Evidence is sufficient to determine outcomes but strength of evidence is limited by number, quality of consistency of individual studies, generalizability of intervention, or indirect nature of evidence on outcomes   • Poor: Insufficient evidence to assess the effects of outcomes because of limited number or power of studies, important flaws in the design of the study, gaps in chain of evidence, or lack of information on important outcomes
Demographic characteristics of participants   • Age, gender, race/ethnicity, sexual orientation, rural, urban, suburban communities
Mental health characteristics of participants   • Community-based sample, clinical sample, diagnoses
Characteristics of intervention   • Universal, selected, indicated   • Dose/duration of intervention   • Follow-up   • Intervention settings: medical facilities, outpatient mental health settings, schools, churches, communities
Outcomes   • Risk factors   • Suicide ideation   • Suicide attempts   • Suicide deaths
Feasibility of prevention/intervention
Generalizable to other settings/sites

AMSTAR, Assessment of Multiple Systematic Reviews; USPSTF, U.S. Preventive Services Task Force

Table 2. Characteristics of systematic reviews and hand-searched primary sources, *n* (%)

	Systematic reviews	Primary sources
Study type		
Cohort study	0 (0)	11 (23.9)
Gray literature	1 (1.0)	0 (0)
Literature reviews	0 (0)	3 (6.6)
Meta-analysis	18 (18.6)	0 (0)
Quasi-experimental	0 (0)	21 (45.6)
RCT	0 (0)	11 (23.9)
Systematic reviews	78 (80.4)	0 (0)

the methodologic quality of systematic reviews. The R-AMSTAR assesses each review based on 11 questions using a Likert-type scale that ranges from 1 (*satisfies none of the criteria*) to 4 (*satisfies all of the criteria*); scores range from 11 to 44.[12] Based on the R-AMSTAR scores, systematic reviews could be rated as excellent (meets 90%–100% of criteria); good (meets 80%–89% of criteria); fair (meets 60%–79% of criteria); or poor (meets <60% of criteria). The quality of the scientific evidence for the suicide prevention/interventions in both the systematic reviews and hand-searched articles were evaluated using the criteria recommended by the USPSTF.[13] The doctoral-level researchers were randomly assigned to code half of the systematic reviews so that each review was coded by two raters. The calculated inter-rater agreement across the reviews was 0.86; discrepancies between coders were resolved via discussion. The hand-searched articles were coded by two doctoral students in clinical psychology whose training focuses on suicide prevention research. The inter-rater agreement for coding the primary sources was 0.88 and discrepancies in their coding were also resolved by discussion.

Evidence Synthesis

Table 2 summarizes the results of the comprehensive reviews based on type of study. The majority of the retrieved studies were systematic reviews; the majority of the studies that were extracted via hand searching had quasi-experimental designs, followed by an equal number of cohort studies and RCTs. Table 3 summarizes the characteristics of the systematic reviews in terms of the use of inclusion/exclusion criteria, the geographic regions in which the studies were conducted, and how the reviews addressed selection biases in their studies.

Although most of the systematic reviews provided inclusion and exclusion criteria, 20% listed no criteria. The majority of systematic reviews surveyed international and U.S. studies and did not exhibit selection bias (e.g., did not include gray literature searches) in their reviews. Most of the international studies were conducted in the United Kingdom, Australia, and Japan. The most common reasons for selection bias in the systematic reviews included not conducting hand searches and conducting limited searches that only involved one to two databases. Using the R-AMSTAR criteria for evaluating reviews, 19% of the systematic reviews were evaluated as having excellent quality, 21% good quality, 37% fair quality, and 23% poor quality. The most common problem areas for the systematic reviews were not including gray literature in the searches, not listing included or excluded materials, and not reporting the methods used to combine studies in meta-analyses.

Table 4 summarizes the findings on the assessment of the quality of the scientific evidence for suicide preventions/interventions based on both the systematic reviews and hand-searched articles. The majority of the prevention/interventions were assessed as having fair to poor scientific evidence across seven types of interventions: access to treatment (75%); community-based programs (78.9%); biological treatments (59.5%); psychosocial treatments (66.1%); screenings (71.7%); and training providers (76.6%). Although few in number, the eighth type of study that focused on restricting access to lethal means (e.g., placing barriers in subway systems) showed stronger scientific evidence compared to the other types of interventions; a little over 83% (*n*=5) of these studies were rated as having good to fair scientific evidence.

Table 3. Characteristics of systematic reviews

Cited inclusion/exclusion criteria

- Inclusion: 17.2%
- Inclusion and exclusion: 54.8%
- No criteria included: 20.2%
- Not applicable: 5.1%

Location of reviewed studies

- International: 2.9%
- International and U.S.: 74%
- U.S. only: 5.8%
- No information: 17.3%

Selection bias

- Yes: 21%
- Unclear: 9.6%
- No: 65.4%
- Not applicable: 4.8%

Table 4. Quality of evidence for interventions, *n* (%)

Quality of evidence	Access to treatment	Means restriction	Community	Biological treatment	Psychosocial treatment	Screenings	Training
Good			4 (21.1)	7 (16.7)	11 (16.2)	1 (14.3)	2 (11.7)
Good to fair	1 (25)	5 (83.3)		10 (23.8)	12 (17.7)	1 (14.3)	2 (11.7)
Fair		1 (16.7)	11 (57.9)	15 (35.7)	33 (48.5)	3 (42.8)	11 (64.7)
Fair to poor	2 (50)		2 (10.5)	5 (11.9)	7 (10.3)	1 (14.3)	1 (5.8)
Poor	1 (25)		2 (10.5)	5 (11.9)	5 (7.3)	1 (14.3)	1 (5.8)
Total (*n*)	4	6	19	42	68	7	17

We were unable to focus the review on the demographic groups with the greatest burden of suicide or examine which interventions worked in boundaried versus undefined settings because very few of the reviews described the demographic characteristics of the study samples beyond age and sometimes gender. Similarly, the feasibility of implementing the interventions and generalizability of the interventions beyond the study site could not be assessed because this information was not described in the systematic reviews.

Although a detailed summary of each reviewed article is beyond the scope of this study, a more detailed summary of each article reviewed for this study can be found in Appendix A. This table lists the authors of the systematic reviews/primary reviews, year of publication, study type, quality of the systematic review/meta-analyses, a brief summary of the review/article, the quality of the scientific evidence of the interventions, and the type of interventions reviewed in the study.

Discussion

The RPTF literature review conducted a "review of reviews" to get a better sense of the current state of suicide prevention/intervention research to better inform recommendations for future directions for research. This review confirmed the difficulty and complexity in conducting research in the area of suicide. Although suicide ranges from the third- to 11th-leading cause of death among various age groups,[14] it is a relatively rare behavior. Additionally, patterns of suicidal behaviors are complex; for example, although suicide ideation is a known risk factor for suicide attempts and deaths, most people who experience suicide ideation do not go on to die by suicide. Recent research suggests that suicide ideators, suicide attempters, and those who die by suicide are three distinct groups.[15] Adding to the complexity of suicide prevention/intervention research is the fact that many of the studies do not have actively suicidal

participants in the study, for ethical and practical reasons. Many studies on suicide focus on the reduction of risk factors, which can be problematic because many of the risk factors associated with suicide (e.g., depression, substance use) are ubiquitous and not unique to suicide.

The findings from our review suggest that there may be many promising suicide prevention/intervention approaches, but the research findings are inconclusive because of methodologic problems. Of the eight types of surveyed interventions, restricting access to lethal means seemed particularly promising, but this is based on a small number of studies, almost all of which did not use a control comparison community.

Somewhat surprisingly, there were methodologic problems with the ways the systematic reviews were conducted. Systematic reviews are an arduous undertaking, but some of the authors seemed to be unaware that there are standards and guidelines that should be followed in conducting a systematic review. Common methodologic problems included not using hand searches and not surveying gray literature, which would actually give the field more accurate effect sizes because gray literature is more likely to report what does not work.

It was particularly puzzling that so few reviews reported the demographic characteristics of the samples included in the reviewed research. Given that suicide rates vary across gender, age groups, race/ethnicity, geographic regions, and nationality, it was surprising that very few reviews reported on the sociodemographic characteristics of the sample. This paucity of information on basic demographic characteristics makes it difficult to assess the relative degree of suicide burden across different communities, which in turn makes it difficult to prioritize how to allocate resources for suicide prevention/intervention. The most commonly reported demographic characteristic was age, followed by gender. The absence of information on race/ethnicity is somewhat understandable in some of the international studies

that may have taken place in racially homogenous communities, but none of the reviews using U.S. samples and very few of the primary sources that were retrieved from the hand searches reported racial/ethnic information from their samples. Similarly, most reviews did not report on the settings in which the preventions/interventions occurred, which makes it difficult to assess the feasibility or address concerns regarding the dissemination of successful programs. Again, most of the primary sources that were retrieved through hand searches often did not describe the settings for the prevention intervention programs.

Another problem in both the systematic reviews and primary sources was that outcomes were often poorly defined and there were no standard criteria for outcomes. The terms *suicide ideation*, *attempts*, and *completions* are often used interchangeably, and most of these terms are not defined or operationalized in a study.[16] As previously noted, most of the studies were not designed to directly assess or intervene with suicidal ideation and behaviors, but were designed to address risk factors commonly associated with suicide. Relatedly, most of the studies did not include actively suicidal participants because of concerns about client safety and medical liability. There are also concerns about the methodologic rigor of the studies, as it is difficult to recruit and retain enough participants to have an adequately powered study.

Recommendations

There are many steps that can be taken to improve the quality of the research so that more definitive statements can be made about what does and does not work in the area of suicide prevention/intervention. Suggestions to improve the quality of reviews include describing the demographic characteristics of study participants; describing intervention characteristics (e.g., intervention settings dose/duration of the intervention; and using a common set of risk/protective factors and outcomes to facilitate aggregation of data across studies;

The field should consider providing guidelines for minimum criteria needed to conduct research on suicide prevention/intervention These guidelines could include consensus on the operationalization of terms for suicide (e.g., suicide ideation, suicide intent) and for consensus on terms for interventions/treatment/prevention (e.g., is there a difference between help-seeking behaviors and treatment seeking?). It would also be helpful if there were better agreement on a set of measures that could be used to assess not only outcomes but risk/protective factors as well. For example, when examining the research on the role of depression and suicide, are researchers talking about depressive symptoms, a syndrome, acute disorders, or chronic disorders?

There is also a need for studies that are specifically designed to assess specific suicide-related outcomes—ideation, intent, and attempts that can be associated with various interventions; developing appropriate and feasible ways to link vital statistics with, for example, healthcare and criminal justice databases to facilitate measurement of suicide-related outcomes would be helpful as well. It would also be helpful to strongly encourage researchers to address the feasibility and generalizability of research findings in their studies.

Conclusions

As noted earlier, suicide prevention/intervention research is particularly challenging because it focuses on a relatively rare behavior for which the underlying mechanisms are not clearly identified. As such, it is difficult to design interventions for complex phenomena with underlying processes that are not always well understood. However, the take-home message from this review should not be that treatment does not work or that we should "throw the baby out with the bath water." Researchers need to improve the science so we can actually find out what works. There has been a recent emphasis in funding more collaborative research approaches across institutions that focus on rapidly advancing the science in the areas of cancer (e.g., the National Cancer Institute's Cancer Center Support Grants)[17] and depression (National Network of Depression Centers; nndc.org/). These collaborative centers often facilitate infrastructure changes in the allocation of resources and help change norms/values in how scientists conduct research because the focus is on collaboration rather than competition. These collaborative models may also help teams become more open about discussing and reporting both successes and failures, and help researchers take bigger risks. Given the challenges that suicide research currently faces (e.g., low base rate behavior, underpowered studies, few systematically tested theories), the use of more collaborative research models may yield more useful findings in the field.

References

1. WHO. Public health action for the prevention of suicide: a framework. Geneva: WHO, 2012. apps.who.int/iris/bitstream/10665/75166/1/9789241503570_eng.pdf.
2. CDC. Leading causes of death reports, national and regional, 1999–2010. webappa.cdc.gov/sasweb/ncipc/leadcaus10_us.html.
3. Mann JJ, Apter A, Bertolote J, et al. Suicide prevention strategies: a systematic review. JAMA 2005;294(16):2064–74.
4. USDHHS, Public Health Service. National strategy for suicide prevention: goals and objectives for action. Washington DC: USDHHS, Public Health Service, 2001. www.sprc.org/sites/sprc.org/files/library/nssp.pdf.
5. Armstrong R, Jackson N, Doyle J, Waters E, Howes F. It's in your hands: the value of handsearching in conducting systematic reviews of public health interventions. J Public Health 2005;27(4):388–91.
6. Egger M, Davey Smith G. Meta-analysis: bias in location and selection of studies. BMJ 1988;316:61–6.
7. Schlosser RW. Appraising quality of systematic reviews. Austin TX: Southwest Educational Development Laboratory, 2007. www.ktdrr.org/ktlibrary/articles_pubs/ncddrwork/focus/focus17/Focus17.pdf.
8. U.S. National Library of Medicine. MeSH Browser. nlm.nih.gov/mesh/mbinfo.html.
9. Oxman AD, Guyatt GH. Validation of an index of the quality of review articles. J Clin Epidemiol 1991;44(11):1271–8.
10. U.S. Preventive Services Task Force. uspreventiveservicestaskforce.org/uspstf08/methods/procmanual1.htm.
11. National Health Service, National Institute for Health and Clinical Excellence. Methods for development of NICE public health guidance. London: National Institute for Health and Clinical Excellence, 2006. www.nice.org.uk/nicemedia/pdf/CPHEMethodsManual.pdf.
12. Kung J, Chiappelli F, Cajulis OO, et al. From systematic reviews to clinical recommendations for evidence-based health care: validation of Revised Assessment of Multiple Systematic Reviews (R-AMSTAR) for grading of clinical relevance. Open Dent J 2010;4:84–91.
13. U.S. Preventive Services Task Force. Appendix A: How the U.S. Preventive Services Task Force grades its recommendations: guide to Clinical Preventive Services. ahrq.gov/professionals/clinicians-providers/guidelines-recommendations/guide/appendix-a.html.
14. CDC, National Center for Injury Prevention and Control. Web-Based Injury Statistics Query and Reporting System (WISQARS). cdc.gov/injury/wisqars.
15. Smith PN, Cukrowicz KC, Poindexter EK, Hobson V, Cohen LM. The acquired capability for suicide: a comparison of suicide attempters, suicide ideators, and non-suicidal controls. Depress Anxiety 2010;27(9):871–7.
16. Silverman MM, Berman AL, Sanddal ND, O'Carroll PW, Joiner TE. Rebuilding the Tower of Babel: a revised nomenclature for the study of suicide and suicidal behaviors. Part I: background, rationale, and methodology. Suicide Life Threat Behav 2007;37(3):248–63.
17. Policies and Guidelines for the P30 Cancer Center Support Grants. cancer.net/publications-and-resources/support-and-resource-links/general-cancer-organizations-and-resources/cancer-centers-coop-groups.

Appendix

Supplementary data

Supplementary material cited in this article is available online at http://dx.doi.org/10.1016/j.amepre.2014.05.023.