

# Suicidology Meets “Big Data”

Michael F. Grunebaum, MD

**S**uicide is the worst acute outcome in psychiatry, in part because it is preventable. Globally, nearly a million people die yearly by suicide,<sup>1</sup> yet despite research and prevention efforts, global suicide rates have increased in the past half century, and the World Health Organization projects they will rise more in the years ahead.<sup>2</sup> Suicide is overwhelmingly associated with mental illness, mostly depressive disorders,<sup>3,4</sup> though substance use and psychotic disorders are also significant risks. Beyond loss of life and other morbidity, suicidal behavior costs an estimated \$33 billion yearly in the United States.<sup>5,6</sup> The increase in the US suicide rate, from 10 deaths per 100,000 persons in 1955 to 12 per 100,000 in 2011,<sup>7</sup> shows the failure of prevention at the population level in the past half century.

## BARRIERS TO PREVENTION

Prevention of suicidal behavior is a challenge for many reasons. While research has identified long-term risk factors—mental illness, substance use disorder, social isolation, to name a few—the ability to predict whether an individual will attempt suicide in the near term (eg, the next day or two) does not exist. Suicide experts from 15 countries met in 2004, conducted a systematic review of studies of suicide prevention programs, and found that restriction of lethal means and education of physicians and gatekeepers on depression were the only interventions with evidence of effectiveness.<sup>3</sup>

Because suicide is relatively uncommon in the general population, studies to demonstrate that an intervention prevents it require very large samples and/or follow-up. An analysis found that, to have adequate statistical power to demonstrate that a specific intervention resulted in a 50% reduction in the US suicide rate, a trial would require close to 1 million person-years of participant follow-up, and even more for smaller effect sizes.<sup>8</sup> Compounding this difficulty, the suicidal phenotype is heterogeneous: some attempts are impulsive, while others are premeditated; some individuals make multiple attempts, while more than half of all suicides are a first attempt.<sup>9</sup> Finally, suicidal persons have historically been excluded from antidepressant clinical trials, akin to excluding patients with metastatic cancer from oncology trials, which is unheard of.

## THE ROLE OF BIG DATA

In this context, the study by de Araújo et al<sup>10</sup> in this issue of the *Journal*, with a sample of more than 48,000 persons, highlights a promising new frontier—using Web-based methods to harness the potential power of “Big Data” for research on suicide and its prevention. The data analyzed were collected as part of the Brazilian Internet Study on Temperament and Psychopathology. The study was advertised through national media, and participants gave electronic informed consent. The incentive for subjects to participate was “to receive a report on their temperament profile and on the likelihood of having a psychiatric disorder based on screening instruments.”<sup>10(pp e359–e360)</sup> The main instrument for the study was adapted from the Linehan Suicidal Behavior Questionnaire.<sup>11</sup> Participants accessed a Web site and filled out instruments from November 2010 to July 2011. To enhance data reliability, questions checking for attention were inserted into the instruments, and, at the end of the survey, subjects were asked how sincere and serious they had been in their answers. From an initial sample of more than 71,000 persons, those deemed unreliable were excluded, leaving a final sample of 48,569.

The study’s main results are largely consistent with published literature: suicide attempts were associated with female sex, lower education, unemployment, being nonreligious, family history of suicide attempt or suicide, hopelessness, sadness, and social isolation. Some results were surprising. For example, 60% of the sample reported at least “a passing thought” of suicide and 6.8% reported a past suicide attempt, whereas in a study of 17 countries, the cross-national lifetime prevalence of suicidal ideation was 9.2%, and 2.7% for suicide attempts.<sup>12</sup> Most likely, the divergent results are due to sampling characteristics.

With its use of Web-based technology, the de Araújo et al study highlights the exciting potential for access to very large study samples to conduct suicidology research. Not surprisingly, other groups have begun to test these waters. A South Korean study analyzed 153,107,350 social media blog posts during a 3-year time period and found strong statistical associations between frequency of use of the Korean words for *suicide* and especially *dysphoria* with national suicide rates.<sup>13</sup> Another group, with personnel from the Dartmouth engineering and medical schools and the US Department of Veterans Affairs (VA) and funding from the Defense Advance Research Project Agency (DARPA), has partnered with Facebook to analyze veterans’ social media posts as a means to predict suicide risk.<sup>14</sup> The study is named “The Durkheim Project,” after the sociologist whose 1897 publication, “Suicide,” was a pioneering work in the field.

Submitted: July 30, 2014; accepted July 31, 2014.

Corresponding author: Michael F. Grunebaum, MD, Columbia University Medical Center and New York State Psychiatric Institute, Department of Psychiatry, Box 42, 1051 Riverside Dr, New York, NY 10032 (mgruneb@nyspi.columbia.edu).

*J Clin Psychiatry* 2015;76(3):e383–e384  
 (doi:10.4088/JCP.14com09421).

© Copyright 2015 Physicians Postgraduate Press, Inc.

This work may not be copied, distributed, displayed, published, reproduced, transmitted, modified, posted, sold, licensed, or used for commercial purposes. By downloading this file, you are agreeing to the publisher's Terms & Conditions.

In a pilot study of The Durkheim Project's method, the Dartmouth group used machine-learning analysis ("a computerized system that can learn to recognize patterns associated with a known outcome") of word and phrase frequencies from VA electronic medical records to assess suicide risk.<sup>15</sup> The study data consisted of more than 11,500 unstructured clinical notes from doctors, nurses, and other clinicians on veterans in 3 cohorts (N = 70 in each): those who died by suicide, those who used mental health services but did not die by suicide, and those who did not use mental health services and did not die by suicide.<sup>15</sup> The most common single word associated with the suicide cohort was *agitation*, with *feeling frightened* being another prominent observation.<sup>15</sup> The predictive model classified the cohorts with 65% or greater accuracy, and while acknowledging the retrospective design, sample size, and other limitations, the authors conclude, "The resulting system could allow clinicians to potentially screen seemingly healthy patients at the primary care level, and to continuously evaluate the suicide risk among psychiatric patients."<sup>15</sup>

There is a great need for sufficiently sensitive and specific tools to enhance clinicians' ability to evaluate individual patients' risk of suicidal behavior, especially in the short term (ie, next few days). Assessment of risk for suicide, especially in the immediate future, is a research goal of the National Action Alliance for Suicide Prevention, a public-private partnership whose goal is to advance a national strategy for suicide prevention.<sup>16</sup> The 65% accuracy of prediction found in the veterans' study above is clearly not good enough for the clinic. Nonetheless, innovative use of Web-based strategies to study very large samples, analytic methods for "Big Data," and public-private partnerships such as those in some of the studies cited give a sense of promise.

Clearly there are challenges inherent to these approaches, including assurance of privacy, confidentiality and informed consent, validity of data obtained through Web-based collection, and sampling issues, to name just a few. As researchers develop creative solutions to overcome these hurdles, their efforts are likely to open up rich sources of data for the study of suicidal behavior and other challenging and unsolved mysteries of psychiatry.

**Author affiliations:** Department of Psychiatry, Molecular Imaging and Neuropathology Division, Columbia University and New York State Psychiatric Institute, New York.

**Potential conflicts of interest:** Dr Grunebaum has received grant/research support from the National Institute of Mental Health and the Brain and Behavior Research Foundation (formerly NARSAD).

**Funding/support:** None reported.

## REFERENCES

1. World Health Organization. Preventing suicide: a global imperative. World Health Organization 2014. [http://www.who.int/mental\\_health/suicide-prevention/world\\_report\\_2014/en/](http://www.who.int/mental_health/suicide-prevention/world_report_2014/en/). Accessed February 9, 2015.
2. World Health Organization. SUPRE: the WHO worldwide initiative for the prevention of suicide. World Health Organization 2014. [http://www.who.int/mental\\_health/management/en/SUPRE\\_flyer1.pdf?ua=1](http://www.who.int/mental_health/management/en/SUPRE_flyer1.pdf?ua=1). Accessed August 8, 2014.
3. Mann JJ, Apter A, Bertolote J, et al. Suicide prevention strategies: a systematic review. *JAMA*. 2005;294(16):2064–2074.
4. Knox KL, Caine ED. Establishing priorities for reducing suicide and its antecedents in the United States. *Am J Public Health*. 2005;95(11):1898–1903.
5. Centers for Disease Control and Prevention. The cost of violence in the United States. [www.cdc.gov](http://www.cdc.gov). Accessed May 4, 2012.
6. Corso PS, Mercy JA, Simon TR, et al. Medical costs and productivity losses due to interpersonal and self-directed violence in the United States. *Am J Prev Med*. 2007;32(6):474–482.
7. Centers for Disease Control and Prevention. WISQARS: fatal injury reports. [http://webappa.cdc.gov/sasweb/ncipc/mortrate10\\_us.html](http://webappa.cdc.gov/sasweb/ncipc/mortrate10_us.html). Accessed July 25, 2014.
8. Brown CH, Wyman PA, Brinales JM, et al. The role of randomized trials in testing interventions for the prevention of youth suicide. *Int Rev Psychiatry*. 2007;19(6):617–631.
9. Maris RW, Berman AL, Silverman MM. Suicide attempts and methods. In: *Comprehensive Textbook of Suicidology*. New York, NY: The Guilford Press; 2000:284–310.
10. de Araújo RMF, Mazzochi L, Lara DR, et al. Thinking about dying and trying and intending to die: results on suicidal behavior from a large Web-based sample. *J Clin Psychiatry*. 2015;76(3):e359–e365.
11. Linehan MM. Suicidal Behavior Questionnaire: SBQ-17. University of Washington: Behavioral Research & Therapy Clinics. Copyright 1996; <http://depts.washington.edu/brtc/files/SBQ-18%204%20Month.pdf>. Accessed July 29, 2014.
12. Nock MK, Borges G, Bromet EJ, et al. Cross-national prevalence and risk factors for suicidal ideation, plans and attempts. *Br J Psychiatry*. 2008;192(2):98–105.
13. Won H-H, Myung W, Song G-Y, et al. Predicting national suicide numbers with social media data. *PLoS ONE*. 2013;8(4):e61809.
14. The Durkheim Project. <http://www.durkheimproject.org/>. Accessed August 7, 2014.
15. Poulin C, Shiner B, Thompson P, et al. Predicting the risk of suicide by analyzing the text of clinical notes. *PLoS ONE*. 2014;9(1):e85733.
16. National Action Alliance for Suicide Prevention. Research Prioritization Task Force. A prioritized research agenda for suicide prevention: an action plan to save lives. National Institute of Mental Health and the Research Prioritization Task Force. <http://actionallianceforsuicideprevention.org/task-force/research-prioritization> Accessed August 7, 2014.