White Rose Research Online

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/144680/

Version: Published Version

Article:
Velupillai, S, Hadlaczky, G.., Baca-Garcia, E. et al. (2019) Risk Assessment Tools and Data-Driven Approaches for Predicting and Preventing Suicidal Behavior. Frontiers in Psychiatry, 10. 36. ISSN: 1664-0640

https://doi.org/10.3389/fpsyt.2019.00036

Reuse
This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here: https://creativecommons.org/licenses/

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

UNIVERSITY OF LEEDS
University of Sheffield
UNIVERSITY of York

frontiers in Psychiatry

PERSPECTIVE
published: 13 February 2019
doi: 10.3389/fpsyt.2019.00036

Risk Assessment Tools and Data-Driven Approaches for Predicting and Preventing Suicidal Behavior

Sumithra Velupillai[1,2,3], Gergö Hadlaczky[4,5], Enrique Baca-Garcia[6,7,8,9,10,11,12], Genevieve M. Gorrell[13], Nomi Werbeloff[14], Dong Nguyen[15,16], Rashmi Patel[1,3], Daniel Leightley[1], Johnny Downs[1,3], Matthew Hotopf[1,3] and Rina Dutta[1,3]*

[1] Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom, [2] School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, Stockholm, Sweden, [3] South London and Maudsley NHS Foundation Trust, London, United Kingdom, [4] National Center for Suicide Research and Prevention (NASP), Department of Learning, Informatics, Management and Ethics (LIME), Karolinska Institutet, Stockholm, Sweden, [5] National Center for Suicide Research and Prevention (NASP), Centre for Health Economics, Informatics and Health Services Research (CHIS), Stockholm Health Care Services (SLSO), Stockholm, Sweden, [6] Department of Psychiatry, IIS-Jimenez Diaz Foundation, Madrid, Spain, [7] Department of Psychiatry, Autonoma University, Madrid, Spain, [8] Department of Psychiatry, General Hospital of Villalba, Madrid, Spain, [9] CIBERSAM, Carlos III Institute of Health, Madrid, Spain, [10] Department of Psychiatry, University Hospital Rey Juan Carlos, Móstoles, Spain, [11] Department of Psychiatry, University Hospital Infanta Elena, Valdemoro, Spain, [12] Department of Psychiatry, Universidad Católica del Maule, Talca, Chile, [13] Department of Computer Science, University of Sheffield, Sheffield, United Kingdom, [14] Division of Psychiatry, University College London, London, United Kingdom, [15] Alan Turing Institute, London, United Kingdom, [16] School of Informatics, University of Edinburgh, Edinburgh, United Kingdom

OPEN ACCESS

Edited by:
Antoine Bechara, University of Southern California, United States

Reviewed by:
Miguel E. Rentería, QIMR Berghofer Medical Research Institute, Australia
Xavier Noel, Free University of Brussels, Belgium

***Correspondence:**
Sumithra Velupillai
sumithra@kth.se;
sumithra.velupillai@kcl.ac.uk

Specialty section:
This article was submitted to Psychopathology, a section of the journal Frontiers in Psychiatry

Received: *29 June 2018*
Accepted: *21 January 2019*
Published: *13 February 2019*

Citation:
Velupillai S, Hadlaczky G, Baca-Garcia E, Gorrell GM, Werbeloff N, Nguyen D, Patel R, Leightley D, Downs J, Hotopf M and Dutta R (2019) Risk Assessment Tools and Data-Driven Approaches for Predicting and Preventing Suicidal Behavior. Front. Psychiatry 10:36. doi: 10.3389/fpsyt.2019.00036

Risk assessment of suicidal behavior is a time-consuming but notoriously inaccurate activity for mental health services globally. In the last 50 years a large number of tools have been designed for suicide risk assessment, and tested in a wide variety of populations, but studies show that these tools suffer from low positive predictive values. More recently, advances in research fields such as machine learning and natural language processing applied on large datasets have shown promising results for health care, and may enable an important shift in advancing precision medicine. In this conceptual review, we discuss established risk assessment tools and examples of novel data-driven approaches that have been used for identification of suicidal behavior and risk. We provide a perspective on the strengths and weaknesses of these applications to mental health-related data, and suggest research directions to enable improvement in clinical practice.

Keywords: suicide risk prediction, suicidality, suicide risk assessment, clinical informatics, machine learning, natural language processing

TARGETED SUICIDE PREVENTION–TIME FOR CHANGE?

Suicide is a global public health concern, with more than 800,000 worldwide deaths, annually. The World Health Organization has set a global target to reduce the rates of suicides by 10% by 2020 (1, 2) The concept of suicidal behavior encapsulates thoughts, plans and acts an individual makes toward intentionally ending their own life (3).

For targeted suicide prevention strategies to be effective for those with mental health problems, high-quality and accessible data from health services is essential (1). With the increased availability of electronic data from public health services and patient-generated data online, advances in data-driven methods could transform the ways in which psychiatric health services are provided (4–6).

Here, we discuss and contrast the use of risk assessment tools and data-driven computational methods, such as the use of machine learning and Natural Language Processing (NLP) within the precision medicine paradigm to aid individualized care in psychiatry. Our aim is to convey the strengths, but also the limitations of current approaches, and to highlight directions for research in this area to move toward impact in clinical practice.

RISK ASSESSMENT TOOLS FOR SUICIDE PREDICTION AND PREVENTION

Tools developed for suicide risk assessment include psychological scales, e.g., the Beck Suicide Intent Scale, and scales derived from statistical models, e.g., the Repeated Episodes of Self-Harm score. Completing suicide risk assessments have become a mandatory part of clinical practice in psychiatry (7), absorbing a considerable proportion of time allocated to clinical care (8). Assessments are aimed at identifying treatable and modifiable factors, the premise being that identifying "high" suicide risk allows clinicians to enhance service provision, or to implement suicide prevention measures in specific, "high-risk" patient-groups, thus avoiding the implementation of inappropriate or costly interventions in "low-risk" patients.

Recently published meta-analyses suggest that the existing tools have inadequate reliability and low positive predictive value (PPV) in distinguishing between low and high-risk patients (9–13). For instance, the meta-analytically derived PPV of studies based on 53 samples was 5.5% over an average follow-up period of 63 months (12). The majority of suicides occurred in the patient groups categorized as "low-risk," as they vastly outnumber the "high-risk" group. Furthermore, most patients in the "high-risk" group did not die by suicide, because of the relative rarity of the outcome. In addition, the 5.5% risk of suicide for "high-risk" patients pertains to a time interval of more than 5 years; from a clinical perspective it is more helpful to identify those likely to die by suicide within much shorter time frames, namely weeks or months. No relationship between the precision of risk assessment tools and date of their publication has been found, suggesting no radical improvement over the past 50 years (10, 12).

Suicide is a rare outcome, even in individuals with severe mental health disorders. This presents several challenges when it comes to suicide risk assessment. Low base-rates demand an instrument with very strong predictive validity. The performance of almost all instruments fall short of this—a mean sensitivity of 56% and a specificity of 79% was derived from a recent meta-analysis (12). But, even a detection method with 90% sensitivity and 90% specificity, would still lead to low PPV-5% at a 1 year incidence of 500/100,000 in clinical populations at higher risk (14). In addition, a time interval of 1 year is not useful in practice. If the goal of prediction was to identify suicide risk within a fortnight or a month, these risk tools would offer little extra predictive power above chance.

Current methods of building suicide risk assessments stem from translating clinical observations, and theory, into either fixed scales or risk factors to inform statistical models. It is possible that these approaches have reached their limits, since there has been no improvement over several decades, and new scales are costly to implement within already stretched front-line services. Suicidal behavior is a complex phenomenon which is contextually dependent and can shift rapidly from 1 day to the next. Capturing these dynamics requires sophisticated measurement and statistical models (15).

One promising direction is to make use of information generated routinely over the course of everyday public service and research activity that deliver dynamic risk assessment at the point of care. This real-world data (RWD) can come from sources such as case reports, administrative and healthcare claims, electronic health records (EHRs), or public health investigation data. RWD show promise in generating new, previously unknown hypotheses with data-driven machine learning techniques, e.g., detection of previously unknown risk or mitigating factors, adverse effects or treatments for mental health disorders (16, 17).

DATA-DRIVEN APPROACHES

Machine learning techniques are methods that learn from and model large datasets using statistical and algorithmical approaches. They can be used to model risk factors, patterns of illness evolution and outcomes, on a speed and scale that is impossible for humans. These models use features to provide information on future events, such as the likelihood that a patient will attempt suicide within a given time interval, and can model complex relations between features and outcomes. Clinical databases such as EHRs typically contain a variety of data, of which structured data entries lend themselves well to computational analysis. As an alternative, or complementary approach to risk assessment tools, data mining techniques have been applied to the problem of identifying suicidal behavior and assessing suicide risk, using different levels of detail and cohorts (18–24), examples in **Table 1**. These findings indicate that machine learning approaches applied to RWD have potential and could be used to generate tools to improve e.g., medical decision-making and patient outcomes. Owing to the flexibility of these approaches, the models can be continuously updated to refine and improve their clinical applicability.

FREE-TEXT AND NATURAL LANGUAGE PROCESSING

One main advantage with EHR data is that it captures routine clinical practice, which may hold cues for suicidal behavior amongst individuals in contact with health services. Detailed clinical information in EHRs is predominantly recorded in free text fields (e.g., clinical case notes and correspondence). Text records contain rich descriptive narratives—describing symptoms, behaviors and changes experienced by patients, which are elicited during clinical assessment and follow-up (27). Criterion-based classification systems (e.g., ICD-10 and DSM-5) do not necessarily reflect the underlying etiology and pathophysiology at an individual patient level (28), and genetic and environmental risk factors are shared between different

TABLE 1 | Six example studies published between 2014 and 2017 that use data-driven approaches—machine learning and/or natural language processing (NLP)-for classifying or predicting suicide risk.

References	Task	Data source; Approach	Key results and findings
Barak-Corren et al. (19)	Prediction of patients' future risk of suicidal behavior	Partners Healthcare Research Patient Data Registry, US EHR 1998–2012; Bayesian machine learning	33–45% sensitivity, 90–95% specificity, and early (3–4 years in advance on average). The approach identified well-known risk factors (e.g., substance abuse) but also less conventional risk factors (e.g., certain injuries and chronic conditions)
Kessler et al. (22)	Prediction of suicides after psychiatric hospitalization	HADS: data from 38 Army/DoD administrative data systems, US; elastic net (regression trees, penalized regressions)	Higher risk of suicide within 12 months of hospital discharge compared to total Army. Strongest predictors included socio-demographics (male, late age of enlistment), criminal offenses (verbal violence, weapons possession), prior suicidality, aspects of prior psychiatric inpatient and outpatient treatment, and disorders diagnosed during the focal hospitalizations
McCoy et al. (25)	Prediction of suicide and accidental death after discharge	Massachusetts General Hospital and Brigham and Women's Hospital, Boston, US EHRs; NLP approach to characterize positive and negative valence (compared with model using only structured codes)	Positive valence reflected in narrative notes was associated with a 30% reduction in risk for suicide
Metzger et al. (26)	Epidemiological surveillance of suicide attempts	Lyon University Hospital Emergency Department, France; Random forest and naïve Bayes including NLP derived variables	Automatic detection of suicide attempts ranged from 70.4 to 95.3% F-measure. Improved quality of epidemiological indicators as compared to current national surveillance approaches.
Tran et al. (23)	Risk stratification using EHR data, compared with clinician assessments	Barwon Health, Australia, EHRs from inpatient admissions and ED visits; L1-penalized continuation-ratio model for ordinal outcomes	Clinicians using checklist predicted patients at high-risk in 3 months with AUC 0.58, 95% CIs: 0.50–0.66. The data-driven model was superior: AUC 0.79, 95% CIs: 0.72–0.84. Predictive factors included known risks for suicide, but also other information relating to general health and health service utilization
Walsh et al. (24)	Prediction risk of suicide attempt	Vanderbilt University Medical Center, US, BioVU Synthetic Derivative data repository; Random forest	Future suicide attempts were predicted with AUC 0.84, precision 0.79, recall 0.95, Brier score 0.14. Accuracy improved from 720 days to 7 days before the suicide attempt. Predictor importance shifted across time.

mental disorders (29). Thus, a richer and more reliable picture of what is documented in EHRs needs to include an analysis of the textual content, which is where NLP methods are important.

Recent years have seen an increase in use of NLP and text mining tools to extract clinically relevant information from EHR and other biomedical text (30–33). Information extraction is an established subfield within NLP seeking to automatically derive structured information from text. In the mental health domain, NLP has been used to extract and classify clinical constructs such as symptoms, clinical treatments and behavioral risk factors (34–41). Using NLP approaches to identify patients at risk of suicidal behavior in addition to, or in combination with, structured data can increase both precision and coverage (26, 42–45).

Other text-based aspects can also be important to the full understanding of suicide risk. For instance, frequent use of third-person pronouns in EHRs, indicating interpersonal distance, has been found to be discriminative for patients who died from suicide, with an increased relative frequency closer to the event (46). Positive valence in discharge summaries (e.g., terms like *glad*, *pleasant*) has also been associated with diminished risk of death by suicide (25).

LOOKING AHEAD: THE ROLE OF DATA-DRIVEN APPROACHES

The distinctive advantage of data-driven approaches is that they may be powerful even if the PPV of the predictions are low, because they can be deployed on a large scale. The usefulness is dependent on the cost and efficacy of the possible intervention. If an automated model reduced the suicide risk by just a fraction, it could save numerous lives cost-efficiently. If we accept that investment in machine learning and NLP approaches is needed to improve predictive and preventive measures for identifying suicide risk (10), focus should now be placed to make these methods applicable in clinical reality (6, 47).

Obtaining and Utilizing Quality Data

The success of machine learning and NLP approaches depends on several factors, such as data availability and task difficulty. EHRs are not easily shared due to confidentiality and governance constraints, thus method comparison, reliability analysis and generalizability studies are still uncommon. Suicidality represents a broad spectrum of actions and thought processes. There

is a wide range of clinical practice in labeling suicide-related phenomena within and across nations. With researchers struggling to settle on standardized nomenclature on non-fatal suicidal behaviors and uniformity in classifying "ideation," this presents considerable challenges to devise an inclusive but specific framework for using NLP to extract relevant material from text sources (e.g., defining appropriate suicide-related keywords). In order to gather a sufficient number of terms, a keyword search strategy on an entire EHR database is commonly used. Whilst effective for unambiguous concepts such as "anemia" or "migraine," this may result in an artificially simplified sample where synonymous terms are missed. For example, from a manually reviewed small EHR sample, suicidal ideation was expressed with alternative phrases such as "go to sleep and not wake up" or "jump off a bridge." However, generating high quality data is time-consuming and costly. Applying keyword matching methods on a large data sample may still result in high coverage (48). Methods to iteratively refine and extend appropriate keywords and data samples for generating high quality annotations on text data can help minimize development costs.

How Can Data-Driven Models be Explained?

While the effectiveness of data-driven approaches has been increasing rapidly due to both technological advancements (e.g., in deep neural networks) and the availability of larger and richer datasets, many approaches are overly opaque. The underlying prediction models are developed on large, complex datasets with a multitude of features and data points that are internally condensed into abstracted representations which are difficult for humans to interpret. Acceptance of data-driven risk prediction models by healthcare practitioners and patients, involves ensuring that the model output can be clinically trusted (49). The increasing interest in algorithmic accountability (50) is thus a welcome development. For example, in a project on evaluating the use of machine learning methods to predict the probability of death for pneumonia patients, neural network models were most accurate, but discarded in favor of simpler models, because they were more intelligible (49). Advanced machine learning methods rely on numerous parameters and configurations, which need to be made interpretable and understandable in order to support practitioners in judging the quality of the assessment, and help identify confounding factors in the decision process.

For example, a suicide risk model developed with an advanced machine learning approach using large numbers of features from EHRs, such as symptoms and behavioral patterns, will produce a model that outputs a risk score but without an explanation of how the score was derived. Making machine learning models comprehensible could be done in different ways (51). One alternative is to extract a more interpretable model, e.g., decision trees, from an underlying "black-box" model (52) by for instance visualizing the most important features and providing an interface to analyse these. Another approach could be to explain a particular predictive outcome rather than explaining the complete model (53), or by visualizing the strength of different model weights and features as in recent text applications (54–56). Further, recent advances in developing patient similarity models could be a valuable approach to develop visual representations and models for improved outcome prediction (57).

The concept of interpretability is not well-defined (58) and there is as yet no consensus on how to evaluate the *quality* of an explanation (59). Explanations should be tailored toward the specific task and the end users; employing and testing the output scores and explanations in a practical setting (60).

Characteristic	Risk assessment tools	Data-driven methods
Format	*Standardised template, pre-defined questionnaire*	*Computational models, features, latent variables*
Output	*Summative score, threshold score*	*Probabilities, labels*
Underlying data	*Theoretical, psychometric*	*Real-world data*
Administration and governance	*Easy to administer, roll-out*	*Requires governance models, data access procedures*
Transferability/generalizability	*Standardised, can be applied across range of settings / populations*	*Unclear – most current models built on data from one institution*
Customisable	*One-size-fits-all*	*Bespoke, precision medicine*
Usefulness in clinical practice	*Aide-mémoire, summary*	*May lead to clinical decision support*

FIGURE 1 | Summary of main characteristics (left) underlying suicide prediction and prevention models: format, output, underlying data, administration and governance, transferability/generalizability, customizable, usefulness in clinical practice. Risk assessment tools (middle) compared with data-driven models (right).

Toward Impact in Clinical Practice

Although recent studies using data-driven methods show promising results, there is still much more work to be done to improve their predictive *utility*, even within "high-risk" cohorts such as those who actually reach health care services. NLP and machine learning methods are still far from perfect. The need to account for the *longitudinal* nature of EHRs is challenging—e.g., establishing a pattern of behavior or treatment response where symptoms may fluctuate over time (61). *Changes* in symptoms, behaviors and healthcare service use prior to suicidality are often strong predictors and need to be appropriately modeled.

The main advantage with data-driven approaches compared to time-consuming risk assessment tools is that they can be continuously refined and updated, they are bespoke, and the data is already there. Access to computing power and data no longer requires huge investment (62, 63). An example of a decision support tool that would support a clinician in their daily work could be one that automatically generates a summary based on a patient's previous history, compared with a larger population trajectory. The tool could output a risk score, highlight which data elements were used to infer this score, and provide the clinician support to conduct further interactive analyses.

However, the main limiting factor for progress in deploying these types of models in clinical practice lies in the lack of clarity around data governance standards and large-scale solutions for patient consent, particularly cross-institutionally. Further, these methodological advances are fairly recent compared to risk assessment tools, and are still continuously being developed. Support for interdisciplinary environments where technical expertise alongside clinical is necessary to enable validation and deployment into clinical practice.

Preventing suicides on a national or even international scale requires multiple societal and health care service considerations (64, 65). To incorporate new technological support that may aid clinicians in their daily work, data-driven methods need to be developed in a way that they actually provide actionable and interpretable information. The main advantages of risk assessment tools in clinical practice are that they are standardized, easy to administer, learn, and interpret—but because they offer little or no predictive ability, they could be enhanced, adapted or complemented by data-driven models that better reflect the individual patient situation (**Figure 1**).

Beyond the Clinic

With the increased use of social media, there is a growing source of text online related to mental health, including suicidal behavior, that can be analyzed with data-driven methods (66–68). The growth of online support networks is an issue that could be integrated in research and health and social care processes (69). Deploying prevention systems that can also operate to improve public health and wellbeing is a another area of growing interest to researchers and policy makers (70, 71). A considerable number of suicides occur in people who have not received any prior mental health assessment or treatment (72). Reliable suicide detection from data generated outside of the healthcare setting is one way of addressing this issue. For instance, moderated online social media-based therapy has been successfully developed for first episode psychosis patients (73). With appropriate ethical research protocols in place (74), this approach could serve as inspiration for developing moderated intervention programmes open to the public based on retrospective large-scale, diverse non-clinical data sources.

CONCLUSIONS

Over the last decade there has been an important shift in medical care, with an active role for patients in their care. Clinicians are encouraged to sustain a reciprocal and collaborative relationship with their patients; enshrined in the 4Ps-predictive, preventive, personalized and participatory medicine (75). The ubiquity of IT technology, increase in education level, and maturation of digital natives have all contributed to an active role for patients. In fact, in 2013, 24% of adults in Europe were millennials aged 18–33 (76). Researchers need to be sensitive to not just the engagement of patients but also the ethical issues of using IT in novel strategies with potential patient benefit (5), so avoiding the public concern and mistrust which followed the introduction of *care.data* in England (77), and recent events with Cambridge Analytica and Facebook.

Today, we are in a unique position to utilize a vast variety of data sources and computational methods to advance the field of suicide research. To address the inherent complexity of suicide risk prediction, collaborative, interdisciplinary research environments that combine relevant knowledge and expertise are essential to ensure that the requisite clinical problem is addressed, that appropriate computational approaches are employed, and that ethical considerations are integrated in the research process when moving toward participatory developments.

AUTHOR CONTRIBUTIONS

SV and RD proposed the manuscript and its contents. All authors participated in the workshop The Interplay of Evaluating Information Extraction approaches and real-world Clinical Research that was held at the Institute of Psychiatry, Psychology and Neuroscience, King's College London, April 27 2017, and financially supported by the European Science Foundation (ESF) Research Networking Programme Evaluating Information Access Systems: http://elias-network.eu/. SV and RD outlined the first draft of the manuscript. Each author contributed specifically to certain manuscript sections: GH on risk assessment tools, EB-G on data-driven methods, GG on NLP, NW, JD, and RP on NLP specifically for mental health, DN on explainability of data-driven methods, DL on deployment and real-world implications, MH on the overall manuscript. All authors contributed to editing and revising the manuscript. SV incorporated edits of the other authors. All authors approved the final version.

FUNDING

This manuscript was written as a result of a workshop that was held at the Institute of Psychiatry, Psychology and Neuroscience, King's College London, financially supported by the European Science Foundation (ESF) Research Networking Programme

Evaluating Information Access Systems: http://elias-network.eu/. SV is supported by the Swedish Research Council (2015-00359) and the Marie Skłodowska Curie Actions, Cofund, Project INCA 600398. EB-G is partially supported by grants from Instituto de Salud Carlos III (ISCIII PI13/02200; PI16/01852), Delegación del Gobierno para el Plan Nacional de Drogas (20151073); American Foundation for Suicide Prevention (AFSP) (LSRG-1-005-16). NW is supported by the UCLH NIHR Biomedical Research Centre. DN is supported by The Alan Turing Institute under the EPSRC grant EP/N510129/1, with an Alan Turing Institute Fellowship (TU/A/000006). RP has received support from a Medical Research Council (MRC) Health Data Research UK Fellowship (MR/S003118/1) and a Starter Grant for Clinical Lecturers (SGL015/1020) supported by the Academy of Medical Sciences, The Wellcome Trust, MRC, British Heart Foundation, Arthritis Research UK, the Royal College of Physicians and Diabetes UK. DL is supported by the UK Medical Research Council under grant MR/N028244/2 and the King's Centre for Military Health Research. JD is supported by a Medical Research Council (MRC) Clinical Research Training Fellowship (MR/L017105/1). RD is funded by a Clinician Scientist Fellowship (research project e-HOST-IT) from the Health Foundation in partnership with the Academy of Medical Sciences. This paper represents independent research part funded by the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

REFERENCES

1. WHO. *Preventing Suicide: A Global Imperative*. (2014). Available online at: https://www.who.int/mental_health/suicide-prevention/world_report_2014/en/
2. WHO. *Mental Health Action Plan 2013–2020*. (2013). Available online at: https://www.who.int/mental_health/publications/action_plan/en/
3. Nock MK, Borges G, Bromet EJ, Cha CB, Kessler RC, Lee S. Suicide and suicidal behavior. *Epidemiol Rev*. (2008) 30:133–54. doi: 10.1093/epirev/mxn002
4. Chekroud AM. Bigger data, harder questions-opportunities throughout mental health care. *JAMA Psychiatry* (2017) 74:1183–4. doi: 10.1001/jamapsychiatry.2017.3333
5. McIntosh AM, Stewart R, John A, Smith DJ, Davis K, Sudlow C, et al. Data science for mental health: a UK perspective on a global challenge. *Lancet Psychiatry* (2016) 3:993–8. doi: 10.1016/S2215-0366(16)30089-X
6. Torous J, Baker JT. Why psychiatry needs data science and data science needs psychiatry: connecting with technology. *JAMA Psychiatry* (2016) 73:3–4. doi: 10.1001/jamapsychiatry.2015.2622
7. Hawley CJ, Littlechild B, Sivakumaran T, Sender H, Gale TM, Wilson KJ. Structure and content of risk assessment proformas in mental healthcare. *J Ment Health* (2006) 15:437–48. doi: 10.1080/09638230600801462
8. Hawley CJ, Gale TM, Sivakumaran T, Littlechild B. Risk assessment in mental health: staff attitudes and an estimate of time cost. *J Ment Health* (2010) 19:88–98. doi: 10.3109/09638230802523005
9. 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. *Br J Psychiatry* (2017) 210:387–95. doi: 10.1192/bjp.bp.116.182717
10. Franklin JC, Ribeiro JD, Fox KR, Bentley KH, Kleiman EM, Huang X, et al. Risk factors for suicidal thoughts and behaviors: a meta-analysis of 50 years of research. *Psychol Bull*. (2017) 143:187–232. doi: 10.1037/bul0000084
11. Large M, Myles N, Myles H, Corderoy A, Weiser M, Davidson M, et al. Suicide risk assessment among psychiatric inpatients: a systematic review and meta-analysis of high-risk categories. *Psychol Med*. (2018) 48:1119–27. doi: 10.1017/S0033291717002537
12. 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:e0156322. doi: 10.1371/journal.pone.0156322
13. Runeson B, Odeberg J, Pettersson A, Edbom T, Jildevik Adamsson I, Waern M. Instruments for the assessment of suicide risk: a systematic review evaluating the certainty of the evidence. *PLoS ONE* (2017) 12:e0180292. doi: 10.1371/journal.pone.0180292
14. Pokorny AD. Prediction of suicide in psychiatric patients. *Rep Prospect Study Arch Gen Psychiatry* (1983) 40:249–57. doi: 10.1001/archpsyc.1983.01790030019002
15. Knox KL, Bajorska A, Feng C, Tang W, Wu P, Tu XM. Survival analysis for observational and clustered data: an application for assessing individual and environmental risk factors for suicide. *Shanghai Arch Psychiatry* (2013) 25:183–94. doi: 10.3969/j.issn.1002-0829.2013.03.10
16. Dipnall JF, Pasco JA, Berk M, Williams LJ, Dodd S, Jacka FN, et al. Fusing data mining, machine learning and traditional statistics to detect biomarkers associated with depression. *PLoS ONE* (2016) 11:e0148195. doi: 10.1371/journal.pone.0148195
17. Harpaz R, DuMouchel W, Shah NH, Madigan D, Ryan P, Friedman C. Novel data-mining methodologies for adverse drug event discovery and analysis. *Clin Pharmacol Ther*. (2012) 91:1010–21. doi: 10.1038/clpt.2012.50
18. Baca-Garcia E, Perez-Rodriguez MM, Basurte-Villamor I, Saiz-Ruiz J, Leiva-Murillo JM, de Prado-Cumplido M, et al. Using data mining to explore complex clinical decisions: a study of hospitalization after a suicide attempt. *J Clin Psychiatry* (2006) 67:1124–32. doi: 10.4088/JCP.v67n0716
19. Barak-Corren Y, Castro VM, Javitt S, Hoffnagle AG, Dai Y, Perlis RH, et al. Predicting suicidal behavior from longitudinal electronic health records. *Am J Psychiatry* (2017) 174:154–62. doi: 10.1176/appi.ajp.2016.16010077
20. Blasco-Fontecilla H, Delgado-Gomez D, Ruiz-Hernandez D, Aguado D, Baca-Garcia E, Lopez-Castroman J. Combining scales to assess suicide risk. *J Psychiatr Res*. (2012) 46:1272–7. doi: 10.1016/j.jpsychires.2012.06.013
21. Delgado-Gomez D, Baca-Garcia E, Aguado D, Courtet P, Lopez-Castroman J. Computerized adaptive test vs. decision trees: development of a support decision system to identify suicidal behavior. *J Affect Disord*. (2016) 206:204–9. doi: 10.1016/j.jad.2016.07.032
22. Kessler RC, Warner CH, Ivany C, Petukhova MV, Rose S, Bromet EJ, et al. Predicting suicides after psychiatric hospitalization in US Army soldiers: the Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS). *JAMA Psychiatry* (2015) 72:49–57. doi: 10.1001/jamapsychiatry.2014.1754
23. Tran T, Luo W, Phung D, Harvey R, Berk M, Kennedy RL, et al. Risk stratification using data from electronic medical records better predicts suicide risks than clinician assessments. *BMC Psychiatry* (2014) 14:76. doi: 10.1186/1471-244X-14-76
24. Walsh CG, Ribeiro JD, Franklin JC. Predicting risk of suicide attempts over time through machine learning. *Clin Psychol Sci*. (2017) 5:457–69. doi: 10.1177/2167702617691560
25. McCoy THJ, Castro VM, Roberson AM, Snapper LA, Perlis RH. Improving prediction of suicide and accidental death after discharge from general hospitals with natural language processing. *JAMA Psychiatry* (2016) 73:1064–71. doi: 10.1001/jamapsychiatry.2016.2172
26. Metzger M-H, Tvardik N, Gicquel Q, Bouvry C, Poulet E, Potinet-Pagliaroli V. Use of emergency department electronic medical records for automated epidemiological surveillance of suicide attempts: a French pilot study. *Int J Methods Psychiatr Res*. (2017) 26:e1522. doi: 10.1002/mpr.1522
27. Perera G, Broadbent M, Callard F, Chang C-K, Downs J, Dutta R, et al. Cohort profile of the South London and Maudsley NHS Foundation Trust

Biomedical Research Centre (SLaM BRC) Case Register: current status and recent enhancement of an Electronic Mental Health Record-derived data resource. *BMJ Open* (2016) 6:e008721. doi: 10.1136/bmjopen-2015-008721
28. Kola I, Bell J. A call to reform the taxonomy of human disease. *Nat Rev Drug Discov*. (2011) 10:641–2. doi: 10.1038/nrd3534
29. Van Os J, Marcelis M, Sham P, Jones P, Gilvarry K, Murray R. Psychopathological syndromes and familial morbid risk of psychosis. *Br J Psychiatry* (1997) 170:241–6. doi: 10.1192/bjp.170.3.241
30. Meystre SM, Savova GK, Kipper-Schuler KC, Hurdle JF. Extracting information from textual documents in the electronic health record: a review of recent research. *IMIA Yearb*. (2008) 17:128–44. doi: 10.1055/s-0038-1638592
31. Névéol A, Zweigenbaum P. Clinical natural language processing in 2014: foundational methods supporting efficient healthcare. *Yearb Med Inf.* (2015) 10:194–8. doi: 10.15265/IY-2015-035
32. Stewart R, Davis K. 'Big data' in mental health research: current status and emerging possibilities. *Soc Psychiatry Psychiatr Epidemiol*. (2016) 51:1055–72. doi: 10.1007/s00127-016-1266-8
33. Velupillai S, Mowery D, South BR, Kvist M, Dalianis H. Recent advances in clinical natural language processing in support of semantic analysis. *IMIA Yearb Med Inform*. (2015) 10:183–93. doi: 10.15265/IY-2015-009
34. Downs J, Dean H, Lechler S, Sears N, Patel R, Shetty H, et al. Negative symptoms in early-onset psychosis and their association with antipsychotic treatment failure. *Schizophr Bull*. (2018) 45:69–79. doi: 10.1093/schbul/sbx197
35. Gorrell G, Oduola S, Roberts A, Craig T, Morgan C, Stewart R. Identifying first episodes of psychosis in psychiatric patient records using machine learning. *ACL* (2016) 2016:196. doi: 10.18653/v1/W16-2927
36. Gorrell G, Jackson R, Roberts A, Stewart R. Finding negative symptoms of schizophrenia in patient records. In: *Proceedings of the Workshop on NLP for Medicine and Biology Associated with RANLP*. Hissar (2013). p. 9–17.
37. Jackson RG, Patel R, Jayatilleke N, Kolliakou A, Ball M, Gorrell G, et al. Natural language processing to extract symptoms of severe mental illness from clinical text: the Clinical Record Interactive Search Comprehensive Data Extraction (CRIS-CODE) project. *BMJ Open* (2017) 7:e012012. doi: 10.1136/bmjopen-2016-012012
38. Patel R, Lloyd T, Jackson R, Ball M, Shetty H, Broadbent M, et al. Mood instability is a common feature of mental health disorders and is associated with poor clinical outcomes. *BMJ Open* (2015) 5:e007504. doi: 10.1136/bmjopen-2014-007504
39. Patel R, Jayatilleke N, Jackson R, Stewart R, McGuire P. Investigation of negative symptoms in schizophrenia with a machine learning text-mining approach. *Lancet* (2014) 383:S16. doi: 10.1016/S0140-6736(14)60279-8
40. Roberts A, Gaizauskas R, Hepple M. Extracting clinical relationships from patient narratives. In: *Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing*. Columbus, OH: Association for Computational Linguistics (2008). p. 10–18. doi: 10.3115/1572306.1572309
41. Wu C-Y, Chang C-K, Robson D, Jackson R, Chen S-J, Hayes RD, et al. Evaluation of smoking status identification using electronic health records and open-text information in a large mental health case register. *PloS ONE* (2013) 8:e74262. doi: 10.1371/journal.pone.0074262
42. Anderson HD, Pace WD, Brandt E, Nielsen RD, Allen RR, Libby AM, et al. Monitoring suicidal patients in primary care using electronic health records. *J Am Board Fam Med.* (2015) 28:65–71. doi: 10.3122/jabfm.2015.01.140181
43. Downs J, Velupillai S, Gkotsis G, Holden R, Kikoler M, Dean H, et al. Detection of suicidality in adolescents with autism spectrum disorders: developing a natural language processing approach for use in electronic health records. In: *AMIA Annual Symposium. AMIA*. Washington, DC (2017).
44. Gkotsis G, Velupillai S, Oellrich A, Dean H, Liakata M, Dutta R. Don't let notes be misunderstood: a negation detection method for assessing risk of suicide in mental health records. In: *Proceedings of the Third Workshop on Computational Linguistics and Clinical Psychology*. San Diego, CA: Association for Computational Linguistics (2016). p. 95–105.
45. Haerian K, Salmasian H, Friedman C. Methods for identifying suicide or suicidal ideation in EHRs. In: *AMIA Annual Symposium Proceedings*. Chicago, IL: American Medical Informatics Association (2012). p. 1244–53.
46. Leonard Westgate C, Shiner B, Thompson P, Watts BV. Evaluation of veterans' suicide risk with the use of linguistic detection methods. *Psychiatr Serv*. (2015) 66:1051–6. doi: 10.1176/appi.ps.201400283
47. Schofield P. Big data in mental health research – do the ns justify the means? Using large data-sets of electronic health records for mental health research. *Psychiatrist* (2017) 41:129–32. doi: 10.1192/pb.bp.116.055053
48. Tissot H, Roberts A, Derczynski L, Gorrell G, Del Fabro MD. Analysis of temporal expressions annotated in clinical notes. In: *Proceedings of 11th Joint ACL-ISO Workshop on Interoperable Semantic Annotation*. London (2015). p. 93–102.
49. Caruana R, Lou Y, Gehrke J, Koch P, Sturm M, Elhadad N. Intelligible models for healthcare: predicting pneumonia risk and hospital 30-day readmission. In: *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '15*. Sydney, NSW: ACM (2015). p. 1721–30.
50. Ford RA, Price WN. Privacy and Accountability in Black-Box Medicine. *Mich Telecomm Tech L Rev*. (2016) 23:1–43.
51. Freitas AA. Comprehensible classification models: a position paper. *SIGKDD Explor Newsl.* (2014) 15:1–10. doi: 10.1145/2594473.2594475
52. Craven MW. *Extracting Comprehensible Models From Trained Neural Networks*. University of Wisconsin–Madison. (1996).
53. Ribeiro MT, Singh S, Guestrin C. "Why should i trust you?": Explaining the predictions of any classifier. In: *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16*. San Francisco, CA: ACM (2016). p. 1135–44.
54. Aubakirova M, Bansal M. Interpreting neural networks to improve politeness comprehension. In: *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*. Austin, TX: Association for Computational Linguistics (2016). p. 2035–41. doi: 10.18653/v1/D16-1216
55. Lei T, Barzilay R, Jaakkola T. Rationalizing neural predictions. In: *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*. Austin, TX: Association for Computational Linguistics (2016). p. 107–17. doi: 10.18653/v1/D16-1011
56. Li J, Monroe W, Jurafsky D. *Understanding Neural Networks through Representation Erasure*. *arXiv Prepr. arXiv*:161208220 (2016).
57. Brown S-A. Patient similarity: emerging concepts in systems and precision medicine. *Front Physiol*. (2016) 7:561. doi: 10.3389/fphys.2016.00561
58. Lipton ZC. *The Mythos of Model Interpretability*. CoRR abs/1606.03490. (2016).
59. Nguyen D. Comparing automatic and human evaluation of local explanations for text classification. In: *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Vol. 1 (Long Papers). New Orleans, LA: Association for Computational Linguistics (2018). p. 1069–78. doi: 10.18653/v1/N18-1097
60. Krause J, Perer A, Ng K. Interacting with predictions: visual inspection of black-box machine learning models. In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, CHI '16*. Santa Clara, CA: ACM (2016). p. 5686–97.
61. Perlis RH, Iosifescu DV, Castro VM, Murphy SN, Gainer VS, Minnier J, et al. Using electronic medical records to enable large-scale studies in psychiatry: treatment resistant depression as a model. *Psychol. Med*. (2012) 42:41–50. doi: 10.1017/S0033291711000997
62. Hansen MM, Miron-Shatz T, Lau AYS, Paton C. Big data in science and healthcare: a review of recent literature and perspectives. contribution of the IMIA Social Media Working Group. *IMIA Yearb*. (2014) 23:21–6. doi: 10.15265/IY-2014-0004
63. Vahabzadeh A, Sahin N, Kalali A. Digital suicide prevention: can technology become a game-changer? *Innov Clin Neurosci*. (2016) 13:16–20.
64. Appleby L, Hunt IM, Kapur N. New policy and evidence on suicide prevention. *Lancet Psychiatry* (2017) 4:658–60. doi: 10.1016/S2215-0366(17)30238-9
65. Brodsky BS, Spruch-Feiner A, Stanley B. The zero suicide model: applying evidence-based suicide prevention practices to clinical care. *Front Psychiatry* (2018) 9:33. doi: 10.3389/fpsyt.2018.00033
66. De Choudhury M, De S. Mental health discourse on reddit: self-disclosure, social support, and anonymity. In: *Proc. International AAAI Conference on Web and Social Media*. Ann Arbor, MI (2014).
67. Gkotsis G, Oellrich A, Velupillai S, Liakata M, Hubbard TJP, Dobson RJB, et al. Characterisation of mental health conditions in social media using Informed Deep Learning. *Sci Rep.* (2017) 7:45141. doi: 10.1038/srep45141

68. Pavalanathan U, De Choudhury M. Identity management and mental health discourse in social media. *Proc Int World-Wide Web Conf Int WWW Conf*. (2015) 2015:315–21. doi: 10.1145/2740908.2743049
69. Barrett JR, Shetty H, Broadbent M, Cross S, Hotopf M, Stewart R, et al. 'He left me a message on Facebook': comparing the risk profiles of self-harming patients who leave paper suicide notes with those who leave messages on new media. *Br J Psychiatry Open* (2016) 2:217–20. doi: 10.1192/bjpo.bp.116.002832
70. Luxton DD, June JD, Chalker SA. Mobile health technologies for suicide prevention: feature review and recommendations for use in clinical care. *Curr Treat Options Psychiatry* (2015) 2:349–62. doi: 10.1007/s40501-015-0057-2
71. Roski J, Bo-Linn GW, Andrews TA. Creating value in health care through big data: opportunities and policy implications. *Health Aff Millwood* (2014) 33:1115–22. doi: 10.1377/hlthaff.2014.0147
72. Bruffaerts R, Demyttenaere K, Hwang I, Chiu W-T, Sampson N, Kessler RC, et al. Treatment of suicidal people around the world. *Br J Psychiatry J Ment Sci*. (2011) 199:64–70. doi: 10.1192/bjp.bp.110.084129
73. Alvarez-Jimenez M, Bendall S, Lederman R, Wadley G, Chinnery G, Vargas S, et al. On the HORYZON: Moderated online social therapy for long-term recovery in first episode psychosis. *Schizophr Res*. (2013) 143:143–9. doi: 10.1016/j.schres.2012.10.009
74. Benton A, Coppersmith G, Dredze M. Ethical research protocols for social media health research. In: *Proceedings of the First ACL Workshop on Ethics in Natural Language Processing*. Valencia: Association for Computational Linguistics (2017). p. 94–102. doi: 10.18653/v1/W17-1612
75. Flores M, Glusman G, Brogaard K, Price ND, Hood L. P4 medicine: how systems medicine will transform the healthcare sector and society. *Pers Med*. (2013) 10:565–76. doi: 10.2217/pme.13.57
76. Stokes B. (2015). *Who are Europe's Millennials?* Pew Research Center. Available online at: http://www.pewresearch.org/fact-tank/2015/02/09/who-are-europes-millennials/
77. Carter P, Laurie GT, Dixon-Woods M. The social licence for research: why care.data ran into trouble. *J. Med. Ethics* (2015) 41:404–9. doi: 10.1136/medethics-2014-102374

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Velupillai, Hadlaczky, Baca-Garcia, Gorrell, Werbeloff, Nguyen, Patel, Leightley, Downs, Hotopf and Dutta. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.