Meta-analysis of the strength of exploratory suicide prediction models; from clinicians to computers

Michelle Corke, Katherine Mullin, Helena Angel-Scott, Shelley Xia and Matthew Large

**Background**

Suicide prediction models have been formulated in a variety of ways and are heterogeneous in the strength of their predictions. Machine learning has been proposed as a way of improving suicide predictions by incorporating more suicide risk factors.

**Aims**

To determine whether machine learning and the number of suicide risk factors included in suicide prediction models are associated with the strength of the resulting predictions.

**Method**

Random-effect meta-analysis of exploratory suicide prediction models constructed by combining two or more suicide risk factors or using clinical judgement (Prospero Registration CRD42017059665). Studies were located by searching for papers indexed in PubMed before 15 August 2020 with the term suicid* in the title.

**Results**

In total, 86 papers reported 102 suicide prediction models and included 20,210,411 people and 106,902 suicides. The pooled odds ratio was 7.7 (95% CI 6.7–8.8) with high between-study heterogeneity ($I^2 = 99.5$). Machine learning was associated with a non-significantly higher odds ratio of 11.6 (95% CI 6.0–22.3) and instead found that the number of included suicide risk factors did not explain between-study heterogeneity in ORs.

**Conclusions**

Machine learning might have the potential to improve the performance of suicide prediction models by increasing the number of included suicide risk factors but its superiority over other methods is unproven.

**Keywords**

Suicide; risk assessment; self-harm; suicide attempt.

**Copyright and usage**

© The Author(s), 2021. Published by Cambridge University Press on behalf of the Royal College of Psychiatrists. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.
suicide risk factors included in the prediction models. We also examined other possible moderators of the ORs according to different diagnostic groups, different study settings and measures of study reporting strength. Finally, we identified the suicide risk factors that were most commonly included in the suicide prediction models. In accordance with the unanswered questions posed by the 2016 meta-analysis we hypothesised that the use of machine learning and a larger number of included suicide risk factors would both be associated with statically stronger suicide predictions.

**Method**

We conducted a registered meta-analysis (PROSPERO; CRD42017059665) according to Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) and Meta-analysis Of Observational Studies in Epidemiology (MOOSE) guidelines.\(^1\)\(^2\)\(^3\)

**Search strategy**

Preliminary searches of Medline, Embase and PsyCINFO from inception to January 2019 using the word ‘suicide’ resulted in an impractical number of hits,\(^3\)\(^4\) while searches with limits using the terms ‘prediction’, ‘model’, or ‘stratification’, missed papers that were identified in earlier reviews.\(^5\)\(^6\)\(^7\) In contrast, searches of PubMed using variants of the word ‘suicide’ in the title reliably located papers identified in searches of multiple databases. Therefore, in order to obtain a large and representative sample of relevant studies, English language papers with accompanying abstracts listed in PubMed from inception to 15 August 2020 were identified using title searches with the term suicid* (expanded to ‘suicida[ti] OR suicide[ti] OR suicides[ti] OR suicidal[ti] OR suicidal/death[ti] OR suicidality[ti] OR suicide[ti] OR suicide/attempted[ti] OR suicide/self[ti] OR suicide/suicide[ti] OR suicide’s[ti] OR suicidetagger[ti] OR suicides[ti] OR suicides[ti] OR suicide’s[ti] OR suicidality[ti] OR suicidal[ti] OR suicidio[ti] OR suicidios[ti] OR suicidogenic[ti] OR suicidology[ti]’) (Fig. 1). Two authors (M.C. and M.L.) screened full-text publications for inclusion and exclusion criteria. Electronic searches were supplemented by hand searches of the reference lists of included studies and earlier relevant meta-analyses.\(^8\)\(^9\)\(^10\)

**Inclusion and exclusion criteria**

We included data from studies that:

(a) examined groups of people who died by suicide or survived;
(b) used cohort or case–control methods;
(c) reported on a type of exploratory suicide prediction model using two or more clinical suicide risk factors (other than age and gender) or clinical judgement (assuming that clinicians consider more than one risk factor in making a clinical suicide risk assessment); and
(d) reported sufficient data to calculate the OR (i.e. the number of higher- and lower-risk patients that survived or died by suicide, the sensitivity and specificity, or other effect-size data).

We excluded data from studies that:

(a) examined validated existing suicide risk scales (including validation stages of machine-learning studies);
(b) reported on suicide attempts or a combined outcome of suicide and suicide attempts;
(c) used psychological autopsy methods;
(d) reported a suicide prediction model based on a single suicide risk factor combined with age and/or gender;
(e) only examined biological markers; or
(f) reported insufficient effect-size data to include in the meta-analysis.

Validation studies of suicide risk scales were excluded because the number and nature of items in suicide risk scales are not empirically derived from the validation data-set.

**Data extraction**

Two authors (M.L. and M.C.) independently extracted the effect-size data that was then reconciled by a third author (S.X.). Differences in effect-size data were resolved further by examination and consensus by M.L. and M.C. Where possible the effect-size data recorded in the form of counts of true positives and total positives, and false negatives and total negatives to allow a meta-analysis of sensitivity and specificity of the suicide prediction models. Most studies dichotomised the patients into higher and lower-risk groups but, when three or more risk strata were reported, the cut-off point associated with the highest OR was used. Two authors (M.L. and H.A.-S.) independently extracted the moderator data that was then reconciled by a third author (K.M.), with differences resolved by further examination by M.L., M.C. and K.M. and consensus. The list of suicide risk factors used in each suicide prediction model was recorded by M.L. and crosschecked by K.M. Suicide risk factors that were included in more than five suicide prediction models were considered to be replicated independent predictors of suicide.

**Moderator variables**

We classified the type of the suicide prediction models according to:

(a) clinical judgement (including both heuristic assessments of suicide risk recorded in medical records and risk categorisation by researchers who were masked to suicide outcomes);
(b) multi-item experimental scales with items selected after bivariate testing of suicide risk factors;
(c) multi-item experimental scales with items selected after multivariate testing of suicide risk factors;
(d) multivariate modelling other than machine learning (using statistically optimised risk models, for example classification plots generated by multiple regressions techniques); and
(e) machine learning (defined as any advanced experimental technique utilising computerised learning).

Where more than one type of suicide prediction model was described in a publication (as defined above) both suicide prediction models were included. When more than one suicide prediction model was reported using a single type of model (as defined above) only the model with the highest OR was used. When more than one population was reported in a single publication (such as men and women) both suicide prediction models were included.

Four continuous moderator variables were collected to investigate between-study heterogeneity in ORs:

(a) the number of included suicide risk factors in the suicide prediction models (because risk models that rely on more detailed information might be more accurate);
(b) the number of potential suicide risk factors (because studies considering more factors initially might then include more factors in the suicide prediction model);
(c) the mean length of follow-up (because studies with longer follow-up are less likely to misclassify eventual suicides); and
(d) publication year (because prediction might have improved over time with the introduction of more advanced types of suicide prediction model such as machine learning).

Downloaded from https://www.cambridge.org/core. 19 Jan 2021 at 06:04:28, subject to the Cambridge Core terms of use.
The number of potential suicide risk factors examined was often reported in the primary research paper or in an associated publication. Otherwise the number of potential risk factors was ascertained by counting all the independent variables listed in the paper. The number and nature of suicide risk factors included in the suicide prediction models was clearly documented in almost every study. The study diagnostic group (schizophrenia spectrum, affective disorder, mixed diagnosis or other) and the research setting (general population/non-psychiatric care, after any form of self-harm or suicide attempt, specialist mental healthcare, discharged psychiatric in-patients, current psychiatric in-patients and in correctional settings) were collected as potential moderators of the strength of suicide prediction.

Assessment of strength of reporting in primary studies

The strength of reporting in the primary studies was assessed using a 0–6 point scale derived from the Newcastle-Ottawa scale for the assessment of reporting strength of non-randomised studies\(^{15}\) and adapted from a scale used in two earlier meta-analyses of suicide prediction models.\(^7,^{14}\) One point was assigned for each of the following items:

(a) cohort study (rather than a controlled study);
(b) gender-matched cohorts or controls;
(c) individuals recruited from a defined geographic area;
(d) suicide risk factors recorded prospectively (rather than by examination of medical records);
(e) suicides ascertained using an external mortality database (rather than suicide as determined by the researchers); and
(f) used ‘persons’ rather than clinical contacts as the denominator.

Studies with a score of 4 or more were considered to have stronger reporting and to be at lower risk of bias.

Data synthesis

A random-effects model was used to calculate the pooled effect size in the form of ORs using Comprehensive Meta-Analysis.\(^{16}\) Secondary outcomes included the sensitivity, specificity and positive predictive value. Between-study heterogeneity was examined using the \(I^2\) and Q-value statistics. A meta-analytic estimate of the receiver operator curve and area under the curve was calculated using MetaDiSc.\(^{17}\) The possibility of publication bias was assessed using Egger’s regression\(^{18}\) and was quantified using Duval and Tweedie’s trim and fill method.\(^{19}\)

The analysis was conducted at the level of suicide prediction model but a sensitivity analysis was conducted using publication as the unit of analysis. Between-group heterogeneity was examined using a mixed-effects model without assuming a common within-study variance and the significance of between-group heterogeneity was determined with Q-value statistics. Between-group analyses

---

Fig. 1 Flow chart of searches for studies reporting exploratory suicide prediction models (SPM).

Database: PubMed
Date: Inception to 15 August 2020
Search term: “suicid* [ti]”
Limits: English language & abstract
\((n = 28,810)\)

Hand searching of included articles and reviews
\((n = 2416)\)

Titles reviewed
\((n = 31,226)\)

Abstracts screened
\((n = 859)\)

Excluded at abstract examination
\((n = 372)\)

Full-text articles assessed for eligibility
\((n = 487)\)

Excluded at full-text examination
\((n = 401)\)

Multivariate analysis but no SPM = 117
No multivariate statistics = 107
Psychological autopsy methods = 46
Outcome included suicide attempts = 31
Suicide risk scale = 28
Data unsuitable for meta-analysis = 6
Duplicated data = 6
Non-English language = 5
Other/lacked relevant data = 55

Studies included
\((n = 86)\)

Models included in the meta-analysis
\((n = 102)\)
were performed to examine the moderation of OR according to: (a) the type of the suicide prediction model; (b) cohort versus control design; (c) diagnostic group; (d) research setting; and (e) strength of reporting items.

Random-effects meta-regression (method of moments) was used to examine whether continuous moderators were associated with between-study heterogeneity in ORs. Multiple meta-regression was used to assess the independence of the associations between the moderator of machine learning, the number of included suicide risk factors and the OR.

**Ethical approval**

No ethical approval was required.

---

**Results**

**Searches and data extraction**

The searches identified 86 studies\(^{20–105}\) reporting 102 samples of people who were categorised as being at high suicide risk using an exploratory suicide prediction model (see Supplementary Material Table S1: table of included studies reporting exploratory suicide prediction models; Supplementary Material Table S2: data used in the meta-analysis and ratings of strength of reporting in primary research; available at https://doi.org/10.1192/bjo.2020.162). There were four disagreements about the selection of studies that were resolved by consensus. An initial data extraction from a subset of studies resulted in disagreements in almost one-third of the effect-size data because of the selection of different cut-off points and different suicide prediction models. A second independent data extraction of the full data-set resulted in disagreements about effect-size data in 10 of the 102 (10%) suicide prediction models, all of which were resolved by further examination. Re-examination resolved 198 differences in 1428 (14%) moderator or reporting strength data points.

The included papers examined 20 210 411 people of whom 106 902 died by suicide. The earliest paper was published in 1966 and the median publication year was 2007. The total number of potential suicide risk factors examined was 12 242 (mean per model = 135, s. d. = 387) after excluding the samples that used clinical judgement (in which the number of potential and included suicide risk factors could not be ascertained) and a single machine-learning study that used 8071 predictor variables.\(^{62}\) The total number of included suicide risk factors in the suicide prediction models (excluding ten clinical judgement samples and three models in which the number of included suicide risk variables was not reported) was significantly associated with the ORs, but this moderator was not significantly associated with the ORs among the 79 non-machine-learning, non-clinical judgement studies (Table 2). Machine-learning studies included a mean of 26 risk suicide factors whereas non-machine-learning studies (excluding clinical judgement studies that were not included in any analyses of the number of risk factors) incorporated a mean of 6.5 suicide risk factors. A multiple meta-regression found that neither machine learning (coefficient 0.21, 95% CI −0.50 to 0.91, Z-value = 0.57, P = 0.56) or the number of included suicide risk factors (coefficient 0.01, 95% CI −0.01 to 0.04, Z-value = 1.42, P = 0.15) made an independent contribution to heterogeneity in ORs and together explained about one tenth of the between-study heterogeneity (Q = 5.56, d.f. = 2, P = 0.06, R-square analogue 10.1%).

Cohort studies, those conducted in geographically defined catchment areas and studies that used external suicide mortality data had a lower pooled OR, respectively, than controlled studies, those that were based on a hospital or health service sample and those that assessed suicide mortality using local data (Table 1). Groups of studies that were matched for gender and those that collected research data prospectively had similar pooled ORs to those that were, respectively, not gender matched or that extracted the research data from medical records (Table 1).

The group of studies that examined suicides based on clinic contacts rather than individuals reported a higher pooled OR (Table 1). Suicide prediction models performed better among cohorts of people with schizophrenia, in primary health/general population samples and among psychiatric in-patients. Suicide prediction models performed less well among cohorts of patients who had presented with suicide attempts or self-harm (Table 3).

Meta-regression found no evidence that the year of publication or length of follow-up contributed to between-study heterogeneity (Table 2). The number of potential suicide risk factors did not contribute to between-sample heterogeneity (Table 2). Further, machine learning made no independent contribution to between-sample heterogeneity in various post hoc multiple meta-regression that included the methodological variable of whether the study used people or clinical contacts in the denominator and other indicators of reporting strength. (For the full data-set used in the study including effect-size data, moderators and strength of reporting data see Supplementary Material Table S2.)

**Suicide risk factors in the suicide prediction models**

A previous suicide attempt or any form of self-harm was the most common included suicide risk factor that was present in 68 of the 102 suicide prediction models. This was followed by depressive diagnosis or symptoms (n = 50), alcohol or substance use (n = 34), suicidal ideation (n = 31), previous psychiatric hospital admission (n = 24), male gender (n = 20), psychotrophic medication (n = 20), physical comorbidity (n = 19), psychiatric diagnosis (n = 19),
Study name | Odds ratio and 95% CI
--- | ---
Ahmedani et al (2017) MV
Allgulander & Fisher (1990) F, MV
Allgulander & Fisher (1990) M, MV
Appleby et al (1999) CJ
Bickley et al (2018) MVES
Blau et al (2005) MV
Blau et al (2018) MV
Buglass & Homburgh (1974) ES
Buglass & McCulloch (1970) F, ES
Buglass & McCulloch (1970) M, ES
Chang et al (2020) ES
Choi et al (2018) ML
Choi et al (2018) MV
Conlon et al (2007) CJ
DeHert et al (2001) ES
Dean et al (1969) F
Del Pozo-Barrionuevo et al (2018) ML
Dexam et al (2006) ES
Dexam et al (2006) MVES
Dong et al (2005) MVES
Drake et al (1988) MV
Dutta et al (2019) MV
Epstein et al (1979) CJ
Farberow & MacKinnon (1974) ES
Farberow et al (1966) CJ
Fazet al et al (2019) MV
Fosse et al (2017) MV
Funahashi et al (2000) MVES
Goldberger et al (2015) MV
Goldstein et al (1991) MV
Gradus et al (2019) F, MV
Gradus et al (2019) M, MV
Harriss & Hawton (2003) F, MV
Harriss & Hawton (2003) MVES
Hastie et al (1998) MV
Hawton et al (1999) MVES
Hofmunn et al et al (2015) ES
Holtstrnad et al (2015) MVES
Humber et al (2017) CJ
Humber et al (2013) MVES
Hung et al (2010) MV
Hurt et al (2007) MVES
Hurt et al (2008) MVES
Hurt et al (2014) MVES
Igen et al (2009) ML
Kim et al (2007) MVES
Kraepelin et al (2002) ES
Kessler et al (2016) ML
Kessler et al (2017) ML
Kessler et al (2017) ML
Kessler et al (2018) ML
Kien et al (2017) MV
Kien et al (2017) MVES
Klim et al (2001) MVES
Klim et al (2000) MVES
Klare et al (1993) ES
Kelsberg et al (1994) MVES
Krupinski et al (1998) MV
Krupinski et al (2000) MV
Levi et al (2016) MV
Lin et al (2014) MVES
Lind et al (2014) CJ
Lopez-Morrinigo et al (2016) MV
Lopez-Morrinigo et al (2018) MVES
Lukaschek et al (2014) MV
Madsen et al (2012) MVES
McCarty et al (2013) MV
McCoy et al et al (2016) MV
Modestin & Kopp (1988) MV
Modestin & Kopp (1988) ES
Modestin et al (1992) MV
Motto & Boissom (1995) MVES
Motto et al (1989) ES
Motto et al (1996) ES
Powell et al (2004) MV
Pratt et al (2013) CJ
Randall et al (2014) ML
Randall et al (2016) MVES
Rosen (1967) ES
Roy (1992) ES
Roy (1992) SCZ ES
Sahdev et al (2017) ML
Sanderson et al (2019) ML
Sanderson et al (2019) MV
Shaffer et al (1974) CJ
Shaffer et al (1974) M
Sharma et al (1988) CJ
Simon et al (2019) HC, ML
Simon et al (2019) SHC, ML
Spies et al (2005) MV
Stein et al (1999) AD, MV
Stein et al (1999) SCZ, MV
Stein et al (2012) MVES
Stenseth et al (1999) ES
Stenseth et al (2006) ES
Wabuy et al (2006) ES
Wabuy et al (2006) MV
Weiss et al (2016) ES
Wendt et al (2016) ES
Winkler et al (2015) MV

Fig. 2 Forest plot of suicide prediction models.

CJ, clinical judgement; MV, multivariate model; MVES, experimental scale based on multivariate analysis; ES, experimental scale based no bivariate analysis; ML, machine learning; M, male; F, female; AD, affective disorder; SCZ, schizophrenia spectrum; PHC, primary health care; SHC, secondary health care; OR, odds ratio.
psychotic symptoms \((n = 19)\), older age \((n = 17)\), previous psychiatric care \((n = 17)\), adverse life events \((n = 12)\), family history of mental disorder or suicide \((n = 12)\), childhood or developmental adversity \((n = 11)\), ethnicity \((n = 9)\), anxiety disorder \((n = 8)\), forensic history \((n = 7)\), housing problems \((n = 7)\), bipolar disorder \((n = 7)\), personality disorder \((n = 7)\), schizophrenia spectrum disorder \((n = 7)\), problems with psychiatric discharge \((n = 6)\), longer psychiatric hospitalisation \((n = 6)\), problems with rapport \((n = 6)\), current financial stress \((n = 5)\) and involuntary psychiatric treatment \((n = 5)\). Some risk factors were included with opposing effect sizes, for example both employment and unemployment were included in \((n = 4)\) suicide prediction models, and being divorced or not married \((n = 6)\) and being married \((n = 3)\) appeared in suicide prediction models.

**Discussion**

This synthesis of a large and representative sample of exploratory suicide prediction models found a strong statistical association between being allocated to the highest suicide risk category and suicide. The meta-analyses of sensitivity and specificity suggests that under half of all suicides might be anticipated by suicide prediction models, with non-suicide being incorrectly predicted in more than one in ten cases. Meta-analyses of a subsample of cohort studies suggested that 34 in 35 suicide predictions are likely to be false positive predictions.

The first aim of our study was to assess whether machine-learning suicide prediction models had stronger prediction than studies using...
other types of suicide prediction. The second aim was to determine whether suicide prediction models with a larger number of included suicide risk factors had stronger prediction than suicide prediction models with fewer factors. While we had conceptualised these as separate questions, machine-learning studies included an average of four times the number of suicide risk factors than other types of suicide prediction model. When all the studies (except clinical judgement studies) were included, the number of included suicide risk factors was positively associated with the OR in line with our hypothesis. However, when machine-learning studies were excluded or when
machine learning was used as a covariable in a multiple meta-regression, there was no association between the number of included suicide risk variables and the OR. This is in line with our findings in longitudinal suicide studies\(^7\) and among samples of psychiatric in-patients\(^{14}\) that models with more risk factors do not necessarily produce greater predictive strength.

This result is of interest because it challenges the traditional notion that considering more risk factors will necessarily improve suicide risk assessment. Further, the result is consistent with a key difference between machine learning and other types of suicide prediction model in that machine learning can potentially recognise the context of factors, such that factors with no overall association with suicide might be included if they confer risk in some contexts but are protective in others. Further research might show that machine learning can improve suicide risk prediction because it can recognise the context and non-linear relationships between risk factors.

However, in this data-set even though the effect size derived from machine learning was moderately stronger than other suicide prediction models, the \(P\)-value was below statistical significance. A future meta-analysis with more machine-learning studies may have a different result. Although this is encouraging for future machine-learning research, we urge some caution. In particular, the machine-learning studies we included had other study characteristics that might have inflated their results. In addition to examining a large number of potential suicide risk factors increasing the possibility of chance associations,\(^{46,39,97}\) some machine-learning studies examined clinical contacts rather than people, resulting in possible oversampling of suicides if they were more frequent attenders.\(^{46,60,77,97}\) Other machine-learning studies oversimplified the spectrum of non-suicide presentations by selective sampling and overweighting of a small proportion of controls.\(^{46,60,61}\) Still other machine-learning studies developed a large number of synthetic or compound potential suicide risk factors (for example alcohol use by various age and gender groups) increasing the risk of chance associations.\(^{46,97}\) We conclude that the case for machine learning as a statistically superior method of suicide prediction model is not yet conclusive.

Our findings also present an interesting conundrum for the clinician: although increasing the number of included risk factors will improve the strength of machine-learning programs, clinicians who use fewer risk factors to identify suicide risk are likely to do as well as those who use more complex risk models.

**Limitations to the generalisability of the pooled estimates**

The very high between-study heterogeneity in effect size means that our pooled estimates may not be generalisable. Moreover, it is likely that the pooled effect size we report would not be as strong in replication/validation studies, given our results are based on exploratory studies in which the suicide prediction models were retrofitted to suicide and survivor data.\(^{106}\) The potential for chance findings in the primary research and therefore our pooled ORs is illustrated by the threshold \(P\)-value of 0.05 for the inclusion of a suicide risk factor and the large number of potential suicide risk factors examined. Of the over 12 000 potential suicide risk factors examined in the primary research, 1 in 20, or over 600 (77% of all the 777 included risk factors) might have been included by chance with a threshold of \(P = 0.05\). This may not have had a great effect on the ultimate strength of the pooled effect size because the number of potential risk factors was not associated with the ultimate effect size and the number of included suicide risk factors (possibly including some chance associations) did not increase the strength of the predictions in non-machine-learning models. Notably, studies that used two or three included suicide risk factors performed just as well as studies that used more variables. Other cautions about the strength of the effect size reported flow from the statistical evidence of publication bias in favour of studies with a larger OR as well as the lower OR in cohort studies and those with generally stronger methods.

**Other findings**

We found evidence that suicide prediction models work better in the general community/non-mental health settings and in current

---

**Table 2** Meta-regression of continuous moderator variables and the strength of prediction models

| Samples, \(n\) | Coefficient | s.e. | Lower limit 95% CI | Upper limit 95% CI | \(Z\)-value | \(P\) |
|---|---|---|---|---|---|---|
| Number included suicide risk factors | 89 | 0.02 | 0.01 | 0.003 | 0.04 | 2.3 | 0.02 |
| Number of included suicide risk factors excluding machine-learning studies | 79 | 0.01 | 0.02 | -0.03 | 0.05 | 0.52 | 0.60 |
| Number of potential suicide risk factors | 92 | 0.0001 | 0.0001 | -0.0001 | 0.0002 | 0.45 | 0.65 |
| Mean length of follow-up | 84 | 0.003 | 0.0009 | -0.0015 | 0.002 | 0.36 | 0.70 |
| Publication year | 102 | -0.002 | 0.007 | 0.36 | 0.70 |

---

**Table 3** Meta-analysis of diagnostic groups and research settings and the strength of prediction models

| Samples, \(n\) | Odds ratio | Lower limit 95% CI | Upper limit 95% CI | \(I^2\) | Between-group heterogeneity | Q-value | d.f. (\(Q\)) | \(P\) |
|---|---|---|---|---|---|---|---|---|
| Diagnostic group | | | | | | | | | |
| Affective disorders | 8 | 10.4 | 7.1 | 15.1 | 31.2 | 11.4 | 2 | 0.003 |
| Schizophrenia spectrum | 12 | 17.2 | 9.5 | 31.2 | 66.6 | 71.4 | 5 | <0.0001 |
| Mixed diagnosis and other diagnoses\(^{a}\) | 82 | 7.0 | 6.0 | 8.1 | 99.4 | |
| Study setting | | | | | | | | | |
| Primary health/general population | 24 | 10.9 | 8.7 | 13.8 | 99.8 | 71.4 | 5 | <0.0001 |
| Specialist mental healthcare | 15 | 5.5 | 2.7 | 11.2 | 99.6 |
| Post self-harm | 12 | 2.9 | 2.3 | 3.7 | 32.3 |
| Post psychiatric discharge | 25 | 8.6 | 5.7 | 13.0 | 99.2 |
| Current psychiatric in-patients | 22 | 10.5 | 6.8 | 16.2 | 84.9 |
| Correctional settings | 4 | 8.0 | 3.5 | 18.5 | 78.2 |

\(a\) Includes one sample of patients with borderline personality disorder and one sample of people with epilepsy.
in-patient settings. Although these finding were incidental and might not be replicated, it is possible that suicide risk factors are more meaningful in settings where the risk factors are less prevalent (such as a psychiatric diagnosis in the general community) and when more accurate and detailed risk profiles are available (such as in a hospital). Suicide prediction models seemed to perform less well after suicide attempts or self-harm. The most obvious reason for this is that suicide attempts or self-harm, which were the single most commonly included suicide risk factor, cannot be used in the suicide prediction model if both the suicides and survivors have this risk factor.

**Implications**

Machine learning makes it possible for vast amounts of information to be modelled to predict suicide risk, but to do this, much more information, including those for economic, social and network factors not traditionally examined in a clinical setting, might be needed. It is possible that machine learning will ultimately produce better suicide prediction models that are more clinically applicable, as well as better reflecting our intuitive knowledge about the complex social, cultural and personal factors that contribute to an individual’s suicide. Alternatively, it might be that suicide is fundamentally unpredictable, with most of the uncertainty about suicide being aleatory rather than epistemic and therefore not very amenable to prediction. In the meantime, we suggest that future machine-learning studies should focus on clinically relevant input variables, use people rather than clinical contacts as the denominator, and should examine ways of combining computational results with broader clinical considerations.

Although the potential of machine learning in suicide prediction remains uncertain, it should not be forgotten that the utility of any suicide prediction model depends on its clinical application and not its effect size or P-value. Even a suicide prediction model producing a strong statistical association may not be useful if there are no rational interventions that can be provided for people who are predicted to die by suicide, remembering the vast majority of predictions are false positives, or if there are no rational interventions that should be withheld from patients classified as lower risk, among whom many suicides might occur. Ultimately the utility of a suicide prediction model should not be evaluated by its statistical strength or perceived suitability to guide interventions, but should be judged by its contribution to a reduction in suicide mortality. Until such time as the use of any suicide prediction model has been shown to reliably reduce suicide, our clinical advice is to focus on understanding and caring for the person in front of us, including by treating any modifiable risk factors, irrespective of estimates of any overall suicide risk category.

---

**Supplementary material**

Supplementary material is available online at https://doi.org/10.1119/bjo.2020.162

**Funding**

None.

**Acknowledgements**

We would like to thank Roar Fosse, Navneet Kapur, Chian-Jue Kuo, and Fredrick Walby for correspondence about their studies and Dr Amy Conder for her assistance with the manuscript.

**Data availability**

All the data upon which the study is based has been submitted in the supplementary material.

**Author contributions**

M.C.: study conception, data acquisition and manuscript preparation. K.M.: data acquisition, data analysis and manuscript preparation. H.A.-S.: data acquisition, data analysis and manuscript preparation. S.X.: data acquisition and manuscript preparation. M.L.: study conception, overall supervision, data acquisition, data analysis and manuscript preparation.

**Declaration of interest**

M.L. provides expert opinion to courts on matters related to suicide. All other authors have no conflicts of interest.

**References**

1. 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.
2. Ribeiro JD, Franklin JC, Fox KR, Bentley KH, Kleiman EM, Chang BP, et al. Self-injurious thoughts and behaviors as risk factors for future suicide ideation, attempts, and death: a meta-analysis of longitudinal studies. Psychol Med 2016; 46: 225–36.
3. Chan MK, Bhatti H, Meader N, Stockton S, Evans J, O'Connor RC, et al. Predicting suicide following self-harm: systematic review of risk factors and risk scales. Br J Psychiatry 2016; 209: 277–83.
4. Carter G, Miller 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.
5. Runeson B, Odeberg J, Pettersson A, Edborn T, Ildelidk Adamsson I, Waem M. Instruments for the assessment of suicide risk: a systematic review evaluating the certainty of the evidence. PLoS One 2017; 12: e0180292.
6. Belsher BE, Smolenski DJ, Pruett LD, Bush NE, Beech EH, Workman DE, et al. Prediction models for suicide attempts and deaths: a systematic review and simulation. JAMA Psychiatry 2019; 76: 642–51.
7. Large M, Kaneson M, Nyles 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.
8. Large M, Gallyard C, Myles N, Ryan CJ, Myles H. Known unknowns and unknown unknowns in suicide risk assessment: evidence from meta-analyses of aleatory and epistemic uncertainty. B.Psych Bull 2017; 41: 160–3.
9. Wilkinson J, Arnold K, Murray E, van Smeden M, Carr K, Sippky R, et al. Time to reality check the promises of machine learning-powered precision medicine. Lancet Digital Health 2020; 2: e677–80.
10. Mchugh CM, Large MM. Can machine-learning methods really help predict suicide? Curr Opinion Psychiatry 2020; 33: 369–74.
11. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. J Clin Epidemiol 2009; 62: e1–34.
12. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis of Observational Studies in Epidemiology (MOOSE) group. JAMA 2000; 283: 2088–10.
13. Mchugh CM, Corderoy A, Ryan CJ, Hickie IB, Large MM. Association between suicidal ideation and suicide: meta-analyses of odds ratios, sensitivity, specificity and positive predictive value. B.Psych Open 2019; 5: e1.
14. Large M, Nyles 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.
15. Wells GA, Shea B, O'Connell D, Peterson J, Welch V, Losos M, et al. The Newcastle-Ottawa Scale (NOS) for Assessing the Quality of Nonrandomised Studies in Meta-Analyses. The Ottawa Hospital Research Institute, 2013 (http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp).
16. Borenstein M, Hedges L, Higgins J, Rothstein H. Comprehensive Meta-Analysis Version 3. Biostat, 2013.
Corke et al

17. Zamora J, Abbarai V, Muriel A, Khan K, Coomarasamy A. Meta-DisC: a software for meta-analysis of test accuracy data. BMC Med Res Methodol 2006; 6: 31.

18. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997; 315: 629–34.

19. Duval S, Tweedie R. Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. Biometrics 2000; 56: 455–63.

20. Ahmedani BK, Peterson EL, Hu Y, Lynch F, Lu CY, Waitzfelder BE, et al. Major physical health conditions and risk of suicide. Am J Prev Med 2017; 53: 308–15.

21. Alligulander C, Fisher LD. Clinical predictors of completed suicide and repeated self-poisoning in 8895 self-poisoning patients. Eur Arch Psychiatry Neurol Sci 1990; 239: 270–6.

22. Appleby L, Dennehy JH, Thomas CS, Faragher EB, Lewis G. Aftercare and clinician characteristics of people with mental illness who commit suicide: a case-control study. Br J Psychiatry 1999; 175: 1306–10.

23. Bickley H, Hunt BM, Windfuhr K, Shaw J, Appleby L, Kapur N. Suicide within two weeks of discharge from psychiatric inpatient care: a case-control study. Psychiatr Serv 2013; 64: 653–9.

24. Blauw E, Kerkhof AJ, Hayes LM. Demographic, criminal, and psychiatric factors related to inmate suicide. Suicide Life Threat Behav 2005; 35: 63–75.

25. Bruer RA, Rodway-Norman M, Large M. Closer to the truth: admission to multifacility psychiatric facilities and an inaccurate history of hospitalization are strongly associated with inmate suicide. Can J Psychiatry 2018; 63: 748–56.

26. Baglass D, Horton J. A scale for predicting subsequent suicidal behaviour. Br J Psychiatry 1974; 124: 573–8.

27. Chang CF, Yeh MK, Chien WC, Chung CH, Li TT, Lai EC. Interactions between psychiatric and physical disorders and their effects on the risks of suicide: a nested case-control study. Ann N Y Acad Sci 2020; 1462: 79–91.

28. Choi SB, Lee W, Yoon JH, Won JU, Kim DW. Ten-year prediction of suicide death among 61,707 men in the Korean male military conscription cohort study in South Korea. J Affect Disord 2018; 231: 8–14.

29. Conlon L, Garland M, Prescott P, Mannion L, Leonard M, Fahy TJ. Psychiatric aftercare and suicide risk: a case-control study using blind rating. Arch Suicide Res 2007; 11: 291–5.

30. Dean RA, Farberow NL, MacKinnon D. A suicide prediction schedule for neuropsychiatric patients. J Nerv Ment Dis 1966; 142: 49–57.

31. De Hert M, McKenize K, Peusken J. Risk factors for suicide in young people suffering from schizophrenia: a long-term follow-up study. Schizophren Res 2001; 47: 127–34.

32. DelPizzo-Banos M, John A, Petkov N, Mark Berndrie D, Southern K, Lloyd K, et al. Using neural networks with routine health records to identify suicide risk: feasibility study. JMR Ment Health 2018; 8: e10144.

33. Didham R, Dovey S, Reith D. Characteristics of general practitioner prior to suicide: a nested case-control study in New Zealand. N Z Med J 2006; 119: U2358.

34. Dong JY, Ho TP, Kan CK. A case-control study of 92 cases of in-patient suicides. J Nerv Ment Dis 1974; 158: 25–9.

35. Drake RE, Gates C, Cotton PG, Whitaker A. Suicide among schizophrenics. Br J Psychiatry 1966; 117: 223–5.

36. Dutta R, Murray RM, Allardyce J, Jones PB, Boydell J. Developing a practical suicide risk prediction model for targeting high-risk patients in the Veterans Health Administration. Int J Methods Psychiatr Res 2017; 26: e1575.

37. Edberg S, Rohde P, Carlsson C. The predictive power of the Suicide Intent Scale. Br J Psychiatry 1973; 123: 296–301.

38. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997; 315: 629–34.

39. EGGER M, DAVEY SMITH G, SCHNEIDER M, MINDER C. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997; 315: 629–34.

40. Ferguson SW, Upton RS, Willoughby J, Amiel D, Calloway D, Mathews A, et al. Prediction of suicide in veterans based on medical student data. J Nerv Ment Dis 1973; 156: 19–29.

41. Ferguson SW, Upton RS, Willoughby J, Amiel D, Calloway D, Mathews A, et al. Prediction of suicide in veterans based on medical student data. J Nerv Ment Dis 1973; 156: 19–29.

42. Foerster S, Koenig HG, Schenker H, Schurch K, Laux E, Ploner M, et al. The predictive power of the Suicide Intent Scale. J Affect Disord 2005; 87: 91–5.

43. Drake RE, Gates C, Cotton PG, Whitaker A. Suicide among schizophrenics. Who is at risk? J Nerv Ment Dis 1984; 172: 613–7.

44. Dutta R, Murray RM, Allardyce J, Jones PB, Boydell J. Early risk factors for suicide in an epidemiological first episode psychosis cohort. Schizophren Res 2011; 126: 11–9.

45. Epstein LC, Thomas CB, Shaffer JV, Perlin S. Clinical prediction of physician used medical student data. J Nerv Ment Dis 1973; 156: 19–29.

46. Farberow NL, MacKinnon D. A suicide prediction schedule for neuropsychiatric hospital patients. J Nerv Ment Dis 1966; 142: 32–44.

47. Fanelli D, Pescosolido B, Zagli S, Roncalli E, Guelfi D, Cecchi L, et al. Prediction of suicide in severe mental illness: development and validation of a clinical prediction rule (OxMIS). Transl Psychiatry 2019; 9: 98.

48. Fosse R, Ryberg W, Carlsson MK, Hammer J. Predictors of suicide in the patient population admitted to a locked-door psychiatric acute ward. PLoS One 2017; 12: e0173958.

49. Funahashi T, Ukiyama Y, Iino S, Nakamura T, Ikeda F, Motoyama T, et al. Prediction of suicide in severe mental illness: development and validation of a clinical prediction rule (OxMIS). Transl Psychiatry 2019; 9: 98.

50. Fosse R, Ryberg W, Carlsson MK, Hammer J. Predictors of suicide in the patient population admitted to a locked-door psychiatric acute ward. PLoS One 2017; 12: e0173958.

51. Funahashi T, Ukiyama Y, Iino S, Nakamura T, Ikeda F, Motoyama T, et al. Prediction of suicide in severe mental illness: development and validation of a clinical prediction rule (OxMIS). Transl Psychiatry 2019; 9: 98.
