

RESEARCH

Open Access



# Evaluating the clinical utility of an easily applicable prediction model of suicide attempts, newly developed and validated with a general community sample of adults

Marcel Miché<sup>1\*</sup>, Marie-Pierre F. Strippoli<sup>2</sup>, Martin Preisig<sup>2</sup> and Roselind Lieb<sup>1</sup>

## Abstract

**Background** A suicide attempt (SA) is a clinically serious action. Researchers have argued that reducing long-term SA risk may be possible, provided that at-risk individuals are identified and receive adequate treatment. Algorithms may accurately identify at-risk individuals. However, the clinical utility of algorithmically estimated long-term SA risk has never been the predominant focus of any study.

**Methods** The data of this report stem from CoLaus|PsyCoLaus, a prospective longitudinal study of general community adults from Lausanne, Switzerland. Participants ( $N=4,097$ ;  $M_{\text{age}} = 54$  years, range: 36–86; 54% female) were assessed up to four times, starting in 2003, approximately every 4–5 years. Long-term individual SA risk was prospectively predicted, using logistic regression. This algorithm's clinical utility was assessed by net benefit (NB). Clinical utility expresses a tool's benefit after having taken this tool's potential harm into account. Net benefit is obtained, first, by weighing the false positives, e.g., 400 individuals, at the risk threshold, e.g., 1%, using its odds (odds of 1% yields  $1/(100-1) = 1/99$ ), then by subtracting the result ( $400 * 1/99 = 4.04$ ) from the true positives, e.g., 5 individuals ( $5 - 4.04$ ), and by dividing the result (0.96) by the sample size, e.g., 800 ( $0.96/800$ ). All results are based on 100 internal cross-validations. The predictors used in this study were: lifetime SA, any lifetime mental disorder, sex, and age.

**Results** SA at any of the three follow-up study assessments was reported by 1.2%. For a range of seven a priori selected threshold probabilities, ranging between 0.5% and 2%, logistic regression showed highest overall NB in 97.4% of all 700 internal cross-validations (100 for each selected threshold probability).

**Conclusion** Despite the strong class imbalance of the outcome (98.8% no, 1.2% yes) and only four predictors, clinical utility was observed. That is, using the logistic regression model for clinical decision making provided the most true positives, without an increase of false positives, compared to all competing decision strategies. Clinical utility is one among several important prerequisites of implementing an algorithm in routine practice, and may possibly guide a clinicians' treatment decision making to reduce long-term individual SA risk. The novel metric NB may become a

\*Correspondence:  
Marcel Miché  
marcel.miche@unibas.ch

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

standard performance measure, because the a priori invested clinical considerations enable clinicians to interpret the results directly.

**Keywords** Suicide attempt, Clinical utility, Adult, Algorithm, Decision support, Net benefit

## Background

Whether a patient with mental health issues may be at risk of attempting suicide is among the questions that await the therapist's decision in routine clinical practice. Aside from acute risk of a suicide attempt [SA; 1], there is also a long-term risk, for example, within the next 12 months [2] or beyond. The American Psychological Association defines a SA "as a deliberate but unsuccessful attempt to take one's own life" [3].

The reduction of long-term SA risk requires the identification of individuals who are at increased SA risk, and who may benefit from interventions that aim to reduce SA risk [e.g., 4–7]. The identification method may be questions from a clinical routine suicide risk assessment. These usually focus on acute risk, which we understand as: the current presence of an individual's suicidal intentions. It may also be an algorithmically supported risk assessment [8], which mainly focuses on long-term risk of a SA, or of suicide [9].

To the best of our knowledge, and although many (56+) published research reports on suicidality (suicide ideation, SA, and suicide) have used so-called supervised machine learning methodologies [e.g., 10], no report has provided empirical evidence that directly relates to questions of the clinical utility of algorithmically estimated SA risk (for a glossary, see Additional File 1). Clinical utility is understood as a benefit after potential harm has been taken into account (e.g., accepting the benefit of a drug despite its side effects). This concept of clinical utility can be applied to individuals and populations [11].

Among the 56 studies included in the meta-analysis by Kusuma et al. [10], three used an adult community sample to prospectively predict SA. One study used Danish national registries as their data source [12] ( $N=22,974$  SA cases, 265,183 controls, SA rate=8.7%, 1458 candidate predictors); the other two [13, 14] used the National Epidemiologic Survey on Alcohol and Related Conditions ( $N=34,653$  and 32,700, candidate predictors 643 (minimum) and 55, respectively, SA rate in both studies=0.6%). We found an additional study, where a Korean adult community sample was used to predict the combined outcome of suicide plan/attempt [15] ( $N=488$ , SA rate=50%, 57 candidate predictors). The only result, which has been consistently reported across these studies is, how well, on average, the algorithm can correctly rank an individual who reported the outcome (case) versus an individual who did not report the outcome (non-case). Correct ranking means that the algorithm assigns a higher risk to a case than to a non-case. The

closer this average ranking success approaches the value one, the better the algorithm can separate a case from a non-case. Results varied between 0.82 and 0.9 across the four studies. It is important to emphasize that this average ranking success of an algorithm is of very little relevance regarding clinical usefulness [16]. Only Lee and Pak [15] included logistic regression among their selected algorithms, whereas Machado et al. [14] included elastic net, which is an extended way of using logistic regression. However, none of these studies has evaluated the clinical utility of an algorithm that prospectively predicts individual long-term SA risk.

We aimed to evaluate the clinical utility of the logistic regression algorithm, using an adult sample of the general population, and prospectively predict individual long-term SA risk. Based on a meta-analysis of risk factors for suicidal thoughts and behaviors [17], including 365 studies, published between 1965 and 2015, we employed, as predictors, the four SA risk factors, namely, lifetime SA, any lifetime diagnosis of a mental disorder, sex, and age.

## Method

### Study participants

The research data stems from the prospective cohort study CoLaus|PsyCoLaus [18, 19] designed to assess (1) cardiovascular risk factors (CVRFs) and mental disorders in the community, and (2) the associations between CVRFs and mental disorders. CoLaus|PsyCoLaus includes a random sample of 6,734 participants (age range: 35–75 years) selected from the general population according to the civil register of the city of Lausanne (Switzerland) between 2003 and 2006. After a first physical and psychiatric investigation, which took place between 2003 and 2008, the cohort was followed up for approximately 5 (first follow-up, FU1), 9 (second follow-up, FU2), and 13 (third follow-up, FU3) years. At baseline, the psychiatric evaluation was restricted to the 35- to 66-year-old participants in the physical exam, resulting in a 67% participation within this age range ( $N=3,719$ ). From FU1 on, all individuals from the initial cohort were eligible for psychiatric evaluation. A total of 5,120 participants agreed to at least one psychiatric evaluation. The present study uses data from the 4,097 (54.5% women, age range 35.8–86.6 years) participants who additionally completed at least a second follow-up psychiatric evaluation. Forty-seven participants were excluded because of incomplete data (Sect. 1 in Additional File 2), resulting in a final sample of 4,050 participants.

## Ethics

The institutional Ethics Committee of the University of Lausanne, which afterward became the Ethics Commission of the Canton of Vaud ([www.cer-vd.ch](http://www.cer-vd.ch)), approved the baseline CoLaus|PsyCoLaus study (reference 16/03; 134-03,134-05bis, 134-05-2to5 addenda 1to4). The approval was renewed for the first (reference 33/09; 239/09), second (reference 26/14; 239/09 addendum 2), and third (PB\_2018-00040; 239/09 addenda 3to4) follow-ups. The study was performed in agreement with the Helsinki declaration and its former amendments, and in accordance with the applicable Swiss legislation. All participants gave written informed consent.

## Measurements

Sex and age were obtained via participants' self-report. Mental health information was gathered with the French version [20] of the Diagnostic Interview for Genetic Studies (DIGS) [21], a semi-structured clinical interview assessing symptoms of *DSM-IV-TR* mental disorders [22]. Interrater agreement of the French DIGS was excellent, albeit with a slightly lower 6-week test–retest reliability for psychotic mood disorders [23] and substance use disorders [24]. The DIGS was completed with the anxiety disorders sections of the French version of the Schedule for Affective Disorders and Schizophrenia–Lifetime and Anxiety disorder version [25, 26]. SA history was assessed in a separate interview module, asking whether participants had ever (first evaluation) or since the last interview (follow-up assessments) attempted to end their life. All four risk factors were assessed at the first psychiatric evaluation. The prospective outcome SA was assessed across the three follow-ups. The interviewers were master-level psychologists who were trained over a 1- to 2-month period. Each interview and diagnostic assessment was examined by an experienced psychologist. Participants who confirmed a SA, were asked additional questions related to the SA. Of the total of 48 participants with a SA during follow-up, 32 confirmed the question whether they really wanted to die, while 12 answered no, four participants were not sure. Medical treatment due to the SA was required in 27 of the 48 SAs, 20 negated the question, one participant was not sure. Of 18 recorded suicides in the overall research sample of 6,734 participants, there were three suicide victims among the 4,050 participants analyzed in this study.

## Selection of predictors

The predictors that were a priori selected are lifetime SA, any lifetime mental disorder (major depressive disorder, any anxiety disorder [generalized anxiety disorder, panic disorder, agoraphobia, social phobia], alcohol abuse or dependence, or illicit drug abuse or dependence), sex, and age. The following reasons guided the predictor selection:

First, all four predictors are long-known risk factors for SA [17]. Second, these risk factors are already assessed or can easily be assessed in routine clinical practice. Third, we assumed that measurement error in these four predictors is very low [27]. Fourth, the prevalence rates of the binary predictors (lifetime SA, any lifetime mental disorder, sex) are not all low, which is why the heuristic of 10 outcome events per variable (EPV) may yield sufficiently robust prediction model coefficients [28].

## Logistic regression

The logistic regression algorithm was our primary choice for three reasons: firstly, various review articles have concluded that logistic regression is not outperformed by modern machine learning algorithms [e.g., 29–32]; secondly, it is transparent regarding its inner mechanisms (linear algebra); thirdly, its output can be regarded as predicted probability that the outcome was observed, owing to logistic regression being rooted in probability theory.

## Competing algorithm: CART

We selected the classification and regression tree (CART) model [33, 34] to compete against logistic regression. Good statistical practice suggests providing empirical evidence that the analyst's preferred data model is better, or at least not inferior to, alternative models. CART is a strong competitor, because unlike logistic regression, it automatically makes use of possible interactions between predictors. Furthermore, like logistic regression, CART's inner mechanisms can be made fully transparent, even to lay users [35]. Because of our limited effective sample size of 48 outcome cases, instead of properly optimizing CART, we conducted a sensitivity analysis, by a posteriori pruning of the decision tree (Sect. 2 in Additional File 2). We used CART's option of case weights, which corresponded to the harm-to-benefit ratio (explained below, see Clinical utility measure: Net benefit).

## Repeated internal cross-validation

The estimation of real-world clinical utility is based on resampling procedures [36], generally referred to as cross-validation. We used 100 repetitions of holdout resampling, each containing 3,240 individuals (outcome no=3,202, yes=38) for training (80%), and 810 individuals (outcome no=800, yes=10) for testing (20%). These testing subsets formed the basis for evaluating the model's clinical utility.

## Software and prediction modeling guidelines

For all analyses and their reporting, including visualization of results, we used the R statistical software environment [37], specifically, the R software packages *rpart* [38], *rms* [39], *precrec* [40], and *ggplot2* [41]. Furthermore, we used R code that was provided as an appendix

in Austin and Steyerberg [42]. We followed the guidelines for transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) [43, 44], which we have documented in Sect. 3 of Additional File 2.

#### **Clinical utility measure: net benefit**

Net benefit (NB) is a decision analytic measure, which expresses clinical utility [45, 46]. NB of a decision strategy is defined as providing more than zero true positive individuals, without an increase of false positive individuals. A true positive is defined as a correct prediction, whereas a false positive is defined as a mistaken prediction (i.e., a false alarm). NB is the key result derived from a so-called decision curve analysis (DCA), which asks whether the clinical benefits exceed the expected costs. NB is our main measure of interest in this study. That is, clinical prediction models are developed to help improve clinical treatment decisions. The first important decision, however, is whether a prediction model is of clinical value, which NB can answer directly, as opposed to other commonly reported measures [47, 48].

The clinical utility of using the prediction model to guide treatment decision making competes against two other decision strategies in a DCA, termed “treat none” and “treat all”. Treating all can theoretically prevent the outcome in the entire treated population, at the cost of a possibly very large part of the population being needlessly treated. Conversely, deciding to treat nobody makes any treatment-related benefits or harms impossible. The decision strategy that is of highest clinical value is the one that displays the highest NB, compared to all competitors, e.g., two or more competing prediction models or one prediction model with two or more different predictor sets. Importantly, there are no further criteria, regarding what qualifies as the highest NB, i.e., should two NB results differ by, say, 0.0001, the higher of the two NB results counts as higher.

Before NB can be computed, one must consider how much more important it would be in a clinical setting for a prediction model to correctly predict the outcome (true positive), such as future SA, compared to falsely predicting it (false positive). For instance, clinicians might come to a consensus that for every true positive prediction, there may be 99 false positive predictions. This consensus should be based on the expected benefits of preventing a SA, in contrast to the expected prevention costs (e.g., the resources to treat 99 false positive individuals for each treated true positive individual). Benefits include the noninterrupted participation in life and avoiding costly treatments arising from an attempted suicide. Costs include the intervention’s implications, such as each individual’s involved efforts, the clinic’s or health system’s

available resources, and the expected side effects of the intervention.

NB ranges between 0 and 1 and is obtained at a given threshold probability  $p_t$  (e.g., 0.01, 1%), which always corresponds to a specific harm-to-benefit ratio [e.g., 1:99, or  $0.01/(1-0.01)$ ]. For example, an NB of 0.2 means there will be 20 more true positives among 100 tested individuals, without an increase in false positives, compared to treating nobody.

Overall, when assessing the clinical utility of a prediction model against other decision strategies (e.g., treat all) across the a priori selected range of reasonable thresholds, the decision strategy with the highest NB qualifies as the primary source to guide treatment decision making (for a more detailed description of DCA and NB, see Sect. 4 in Additional File 2).

#### **Delta NB**

Delta NB was defined as the additional net increase of true positives at a given  $p_t$ , when using logistic regression and CART, respectively. That is, for  $p_t$  less than the outcome incidence of 1.23%, each prediction model was compared to the treat all decision strategy, whereas for  $p_t$  equal to or greater than the outcome incidence, each prediction model was compared to the treat none decision strategy. A negative delta NB indicates that the model may cause more harmful clinical decisions, compared to an alternative decision strategy, e.g., treat all [49], and therefore cannot be recommended for clinical use.

#### **Reasonable range of threshold probabilities**

DCA requires researchers to set a range of reasonable threshold probabilities, accommodating varying threshold preferences across individuals for deciding whether to take outcome-preventing actions [45, 50, 51]. We selected the following threshold probabilities: 0.5%, 0.75%, 1%, 1.25%, 1.5%, 1.75%, and 2%.

Clinically, these low thresholds emphasize the benefit of capturing true positive individuals over the cost of capturing false positive individuals. This trade-off eventually raises legal, ethical, and economic concerns [52], which are beyond the scope of this report. Methodologically, thresholds that are close to, as opposed to being distant from, the outcome rate in the study sample, render the prediction model less sensitive to model miscalibration [49]. This seems reasonable for an initial proof-of-concept report.

#### **Prediction performance measures**

We report the area under the precision-recall curve (PR AUC), the area under the receiver operating characteristic curve (ROC AUC), [40], and the Integrated Calibration Index (ICI; Austin and Steyerberg [42]). All three measures summarize performance across all threshold

**Table 1** Distribution of the variables used in this report ( $N=4,050$ )

Variable	0		1	
	n	%	n	%
Sex (0 = male, 1 = female)	1,841	45.46	2,209	54.54
Lifetime mental disorder <sup>a</sup>	1,853	45.75	2,197	54.25
Major depressive disorder	2,373	58.59	1,677	41.41
Any anxiety disorder	3,405	84.07	645	15.93
Alcohol abuse or dependence	3,642	89.93	408	10.07
Illicit drug abuse or dependence	3,830	94.57	220	5.43
Lifetime SA	3,845	94.94	205	5.06
Follow-up SA	4,002	98.81	48	1.19

Note. Except for sex, 0=no and 1=yes;  $n$ =number of participants; SA=suicide attempt. <sup>a</sup>Major depressive disorder, any anxiety disorder (generalized anxiety disorder, panic disorder, agoraphobia, social phobia), alcohol abuse or dependence, illicit drug abuse or dependence

**Table 2** Logistic regression model coefficients

	Coefficient
Intercept	-4.0296
Lifetime SA	2.5446
Lifetime mental disorder	1.1711
Sex	-0.0631
Age	-0.0423

Note. SA=suicide attempt

probabilities. The chance level for the PR AUC is the outcome rate in the validation data (in our study, 0.0123) and 0.5 in the ROC AUC. Perfect discrimination is represented by the value 1 in both the PR AUC and ROC AUC. The ICI shows better calibration the closer it is to 0.

Additional performance results and visualizations are presented in Sect. 5 of Additional File 2. The prediction performance measures are reported only for logistic regression (as CART was selected as a competitor only regarding clinical utility). Performance results for CART are available upon request.

## Results

Table 1 provides the distribution of the predictors used in this report. Major depressive disorder, any anxiety disorder, abuse or dependence of alcohol and illicit drugs, respectively, were merged into lifetime mental disorder, which was then used as one of four predictors. The participants' ages ranged from 35.8 to 85.6 years ( $M=53.9$  years,  $SD=11.1$ ). For further information, see Sect. 1 of Additional File 2.

### Logistic regression model coefficients

The estimated model coefficients of the logistic regression model for the full sample ( $N=4,050$ ) are presented in Table 2. For more detailed results, see Sect. 6 of Additional File 2.

### NB and delta NB

The 100 repetitions of the cross-validated NB for each of the selected threshold probabilities are summarized, using the median NB, in Fig. 1; Table 3 (for detailed results and visualization, see supplementary R package). Logistic regression showed a somewhat higher median NB across the entire range of thresholds, compared to treat all and treat none. CART showed a lower median NB at the 0.5% threshold, compared to treat all. Across 700 cross-validations, logistic regression indicated 18 times that it was a potentially harmful decision strategy (including 14 times at  $p_t$  0.5% and two times at  $p_t$  0.75%), compared to 119 times for CART (including 75 times at  $p_t$  0.5% and 38 times at  $p_t$  0.75%). For more detailed NB results and visualization of all 100 cross-validated decision curves, see Sect. 7 of Additional File 2, especially the instruction and example therein, under the headline Detailed results.

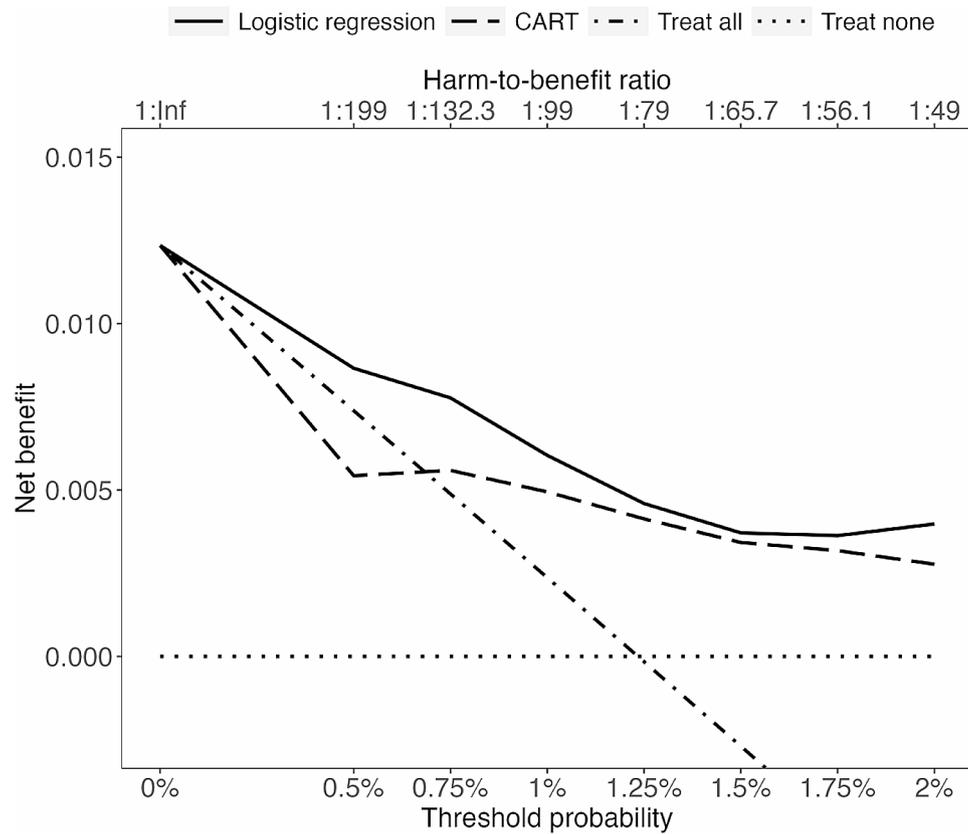
### Logistic regression prediction performance

The PR AUC ranged between 0.028 and 0.436, with 75% of results being below 0.163. Subjects who attempted suicide received a higher predicted probability in 80% of all pairwise comparisons between subjects with vs. subjects without a reported SA (see ROC AUC median in Table 4). Further results, in full detail, such as true positives, false positives, sensitivity, and positive predictive value, can be found in the supplementary R package (see also Sect. 7 of Additional File 2, headline Detailed results).

## Discussion

We presented the evaluation of the clinical utility of using logistic regression to prospectively predict long-term SA risk in adult individuals of a general community sample. Clinical utility was observed across the entire range of selected threshold probabilities, qualifying logistic regression as the best source to guide clinical decision making, compared to the alternatives of CART, treating everybody, and treating nobody. However, the CART sensitivity analysis indicated that using optimization, CART may qualify as the best decision guide for two of the seven threshold probabilities (1.5% and 1.75%); see Sect. 2 of Additional File 2.

In our investigation we used an extremely parsimonious approach, which is in line with clinically realistic demands. Indeed, Jaccobucci et al. [53] indicated that it may be premature to believe that only lots of predictors can produce clinically useful predictions of a suicidal outcome. It may be that logistic regression with four predictors suffices to predict long-term SA risk, as we have demonstrated for, we believe, the first time. Notably, our approach was parsimonious; for instance, the four predictors are commonly obtained in routine clinical practice, which may greatly facilitate the implementation of



**Fig. 1** Median net benefit among 100 test subsamples, based on the resampling procedure, which was the same for each of the seven selected threshold probabilities. The solid line shows the logistic regression model, and the dot-dashed line shows the classification and regression tree (CART) model. The harm-to-benefit ratios are rounded to one decimal place (Inf=infinity, expressing an infinite harm to missing any true positive subject in the entire population)

**Table 3** Net benefit and delta net benefit results, as shown in Fig. 1

$p_t$ %	Net benefit		Median net benefit across 100 cross-validations			
	Treat all	Treat none	Logistic regression		CART	
			Total	Delta	Total	Delta
0.00	0.0123	0	0.0123	0	0.0123	0
0.50	0.0074	0	0.0087	0.0013	0.0054	-0.0020
0.75	0.0049	0	0.0078	0.0029	0.0056	0.0007
1.00	0.0024	0	0.0060	0.0037	0.0049	0.0026
1.25	<0	0	0.0046	-	0.0041	-
1.50	<0	0	0.0037	-	0.0034	-
1.75	<0	0	0.0036	-	0.0032	-
2.00	<0	0	0.0040	-	0.0028	-

Note. CART=Classification and regression tree; the column  $p_t$  % contains the seven selected threshold probabilities; 0% was added to satisfy reporting guidelines for decision curve analysis. The  $p_t$  of 1.25% is larger than the outcome rate in the test data (0.0123=1.23%); yielding a negative net benefit (<0) for treat all, as do all  $p_t > 1.25\%$ . Empty cells are empty because delta net benefit is equal to net benefit, according to our delta net benefit definition. Lower net benefit results were observed 18 times (logistic regression) and 119 times (CART) across 700 cross-validations, compared to either treat all or treat none

**Table 4** Logistic regression prediction performance results across 100 cross-validations

Predictor	Min	1st Qu.	Mean	Median	3rd Qu.	Max
PR AUC	0.028	0.074	0.132	0.106	0.163	0.436
ROC AUC	0.644	0.771	0.805	0.800	0.846	0.916
ICI	0.003	0.006	0.008	0.008	0.010	0.018

Note. Min=minimum; 1st Qu. = 25% quantile; 3rd Qu. = 75% quantile; Max=maximum; PR AUC=area under the precision-recall curve; ROC AUC=area under the receiver operating characteristic curve; for both, a value of 1 represents perfect discrimination; ICI=Integrated Calibration Index; a value of 0 for the ICI represents perfect calibration

such an algorithm in that setting, provided that prediction success can be replicated and that it is continuously supervised once implemented [54].

The parsimonious approach we took may have another advantage. It may maximize the ease of conducting external, as opposed to internal, model validation [55], which is among the most important tasks one should complete prior to model deployment in the real world [56]. External validation could thus be conducted in a large number of independent datasets and then be meta-analyzed, providing empirical arguments for or against large-scale implementation of a SA prediction model in routine clinical care of adults.

The primary reason to consider using a prediction model in clinical practice, despite knowing that false positive predictions will be inevitable, is that it may still improve clinical decision making. This report is the first, in the realm of SAs, to have focused on NB. One of the most important prerequisites of NB is to present an idea of what the clinical decision may entail, because only then can a reasonable range of threshold probabilities be agreed upon across researchers, clinicians, and other stakeholders.

As research of clinical prediction models of SAs in adults progresses [8], it will be important that other researchers also present clinically feasible models, e.g., open box, as opposed to black box, prediction models, containing few candidate predictors which are easy to assess in clinical practice. Should the NB, if presented in a sufficient number of studies, indicate clinical utility across a reasonable range of risk thresholds, discussing clinical implementation is warranted. This recommendation statement [57] and the criticism of it [58] are current examples of discussing implementation of screening instruments, in a population wide effort of preventing depression and suicide in US adults. The range of reasonable risk thresholds depend on the evaluation of attempted suicide, which we evaluate as a very serious outcome, and the clinical action to be provided to individuals whose estimated risk exceeds the risk threshold. For instance, more extensive diagnostic procedures may justify a wide dissemination, due to relatively low costs, as opposed to a therapeutic intervention, due to relatively high costs. Wide dissemination means a high number of false positive individuals, e.g., 99, for each true positive individual may clinically be acceptable, i.e., 1:99 (risk threshold of 1%). However, such matters must be discussed and agreed upon by all involved stakeholders, who belong to diverse groups, such as therapists, patients, public health politicians, and lawyers.

Eventually, should official public health institutions approve the use of a SA prediction model in clinical practice, its use must be very simple. That is, both the therapist and the patient must be able to use the risk algorithm

and interpret the result effortlessly. For example, a man, 51 years old, having reported a lifetime suicide attempt, but having no lifetime mental disorder diagnosis, receives an estimated risk of a future SA of 2.5%, using our logistic regression model, which is presented in Table 2 (the model is published as part of the supplementary R package). Since 2.5% exceeds the maximum value of our suggested reasonable range of risk thresholds (0.5–2%), the therapist would offer this man the preventive action, e.g., more extensive SA diagnostics.

Overall, visual presentation, a clear interpretation, and a transparent explanation of the individual's estimated SA risk is warranted, which includes communicating to the individual the uncertainty of his or her estimated risk [59]. Some authors suggest the use of so-called nomograms [60, 61], which may facilitate the use of risk algorithms in clinical practice. A nomogram is a graphical representation of a mathematical formula, which is what a prediction model is. Such future possibilities require substantial amounts of clinically relevant research.

### Strengths

First, we followed recommendations against using any of the up or downsampling methods for logistic regression [62]. Second, we used only four candidate predictors, which are already assessed in routine clinical care. Third, we present our complete analysis code, as well as the logistic regression prediction model (supplementary R package), which can be downloaded from this GitHub repository <https://github.com/mmiche/predictSuiatt-PsyCoLaus>. Although we cannot publish the original CoLaus|PsyCoLaus study data, we provide code to simulate data, which is superficially similar to the original study data.

### Limitations

First, we had a very small effective sample size ( $N=48$  outcome cases). However, by using four predictors, we met the EPV heuristic of approximately 10 outcome cases per predictor in the training subsample. Second, we used a single question to measure the outcome SA, instead of employing additional qualifiers, such as the intention to die of the person attempting suicide. In our view, it appears justified to regard all participants who affirm this question as constituting a clinically homogeneous and relevant group. Third, we used internal validation, which is the minimum requirement for prediction modeling research [63]. However, this is the first proof-of-concept paper describing such a parsimonious approach to predict SA in a general community adult sample, which is why we think internal validation is a justifiable limitation. Note that Sect. 8 of Additional File 2 permits readers to judge for themselves whether overfitting was strong.

## Conclusion

Despite the strong class imbalance of the outcome (98.8% no, 1.2% yes) and only four predictors, clinical utility (number of true positives greater than the weighted number of false positives) was observed across the full range of reasonable risk thresholds. If comparable future research also indicated clinical benefit of an algorithm's estimates of adult SA risk, clinician's routine use of a risk algorithm would improve their decision making overall, by using the tool that was empirically better than all other tested clinical decision strategies. This may eventually help save lives or prevent individuals from attempting suicide.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12888-024-05647-w>.

Additional File 1: Glossary of terms, used in the main document.

Additional File 2: Additional information, such as a detailed explanation of decision curve analysis, results in more detail, and sensitivity analyses.

## Acknowledgements

Not applicable.

## Author contributions

M.M. is responsible for the conception and design of the documents, the statistical analyses and the interpretation of results, preparation of tables and figures, and for writing the main manuscript text, Additional File 2, and the supplementary R package. M.P.F. wrote the Methods' subsections Participants, Ethics, and Measurements, and provided critical feedback. M.P. provided critical feedback and is the principal investigator of the PsyCoLaus study arm. R.L. provided critical feedback and suggestions regarding the visual presentation of the net benefit results. All authors read and approved the final manuscript.

## Funding

Not applicable.

Open access funding provided by University of Basel

## Data availability

Yes, Additional File 2 and the supplementary R package. The supplementary R package can be downloaded from this GitHub repository <https://github.com/mmiche/predictSuiattPsyCoLaus>. Although we cannot publish the original CoLaus|PsyCoLaus study data, we provide code to simulate data, which is superficially similar to the original study data.

## Declarations

### Ethics approval and consent to participate

The institutional Ethics Committee of the University of Lausanne, which afterward became the Ethics Commission of the Canton of Vaud ([www.cer-vd.ch](http://www.cer-vd.ch)), approved the baseline CoLaus|PsyCoLaus study (reference 16/03; 134-03, 134-05bis, 134-05-2to5 addenda 1to4). The approval was renewed for the first (reference 33/09; 239/09), second (reference 26/14; 239/09 addendum 2), and third (PB\_2018-00040; 239/09 addenda 3to4) follow-ups. The study was performed in agreement with the Helsinki declaration and its former amendments, and in accordance with the applicable Swiss legislation. All participants gave written informed consent.

### Consent for publication

Not applicable.

## Competing interests

The authors declare no competing interests.

## Author details

<sup>1</sup>Department of Psychology, Division of Clinical Psychology and Epidemiology, University of Basel, Missionsstrasse 60-62, 4055 Basel, Switzerland

<sup>2</sup>Psychiatric Epidemiology and Psychopathology Research Center, Lausanne University Hospital, University of Lausanne, Prilly, Switzerland

Received: 19 September 2023 / Accepted: 28 February 2024

Published online: 20 March 2024

## References

1. Bagge CL, Littlefield AK, Wiegand TJ, Hawkins E, Trim RS, Schumacher JA et al. A controlled examination of acute warning signs for suicide attempts among hospitalized patients. *Psychol Med*. 2023 [cited 2023 Apr 2];53(7):2768–76. Available from: [https://www.cambridge.org/core/product/identifier/S0033291721004712/type/journal\\_article](https://www.cambridge.org/core/product/identifier/S0033291721004712/type/journal_article).
2. Bommersbach TJ, Rosenheck RA, Rhee TG. National trends of mental health care among US adults who attempted suicide in the past 12 months. *JAMA Psychiatry*. 2022 Mar 1 [cited 2023 Apr 2];79(3):219. Available from: <https://jamanetwork.com/journals/jamapsychiatry/fullarticle/2787969>.
3. American Psychological Association. APA Dictionary of Psychology. attempted suicide. 2013 [cited 2024 Jan 16]. Available from: <https://dictionary.apa.org/attempted-suicide>.
4. Inagaki M, Kawashima Y, Yonemoto N, Yamada M. Active contact and follow-up interventions to prevent repeat suicide attempts during high-risk periods among patients admitted to emergency departments for suicidal behavior: a systematic review and meta-analysis. *BMC Psychiatry*. 2019 Dec [cited 2023 May 26];19(1):44. Available from: <https://bmcp psychiatry.biomedcentral.com/articles/https://doi.org/10.1186/s12888-019-2017-7>.
5. Meerwijk EL, Parekh A, Oquendo MA, Allen IE, Franck LS, Lee KA. Direct versus indirect psychosocial and behavioural interventions to prevent suicide and suicide attempts: a systematic review and meta-analysis. *Lancet Psychiatry*. 2016 Jun [cited 2023 May 26];3(6):544–54. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S221503661600064X>.
6. Sheehan L, Oexle N, Bushman M, Glover L, Lewy S, Armas SA et al. To share or not to share? Evaluation of a strategic disclosure program for suicide attempt survivors. *Death Stud*. 2023 Apr 21 [cited 2023 Apr 2];47(4):392–9. Available from: <https://www.tandfonline.com/doi/full/https://doi.org/10.1080/07481187.2022.2076266>.
7. Torok M, Han J, Baker S, Werner-Seidler A, Wong I, Larsen ME et al. Suicide prevention using self-guided digital interventions: a systematic review and meta-analysis of randomised controlled trials. *Lancet Digit Health*. 2020 Jan [cited 2023 May 26];2(1):e25–36. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S2589750019301992>.
8. Kirtley OJ, van Mens K, Hoogendoorn M, Kapur N, de Beurs D. Translating promise into practice: a review of machine learning in suicide research and prevention. *Lancet Psychiatry*. 2022 Mar [cited 2023 Apr 2];9(3):243–52. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S2215036621002546>.
9. Kessler RC, Bauer MS, Bishop TM, Bossarte RM, Castro VM, Demler OV et al. Evaluation of a model to target high-risk psychiatric inpatients for an intensive postdischarge suicide prevention intervention. *JAMA Psychiatry*. 2023 Mar 1 [cited 2023 Apr 2];80(3):230–40. Available from: <https://jamanetwork.com/journals/jamapsychiatry/fullarticle/2800171>.
10. Kusuma K, Larsen M, Quiroz JC, Gillies M, Burnett A, Qian J et al. The performance of machine learning models in predicting suicidal ideation, attempts, and deaths: A meta-analysis and systematic review. *J Psychiatr Res*. 2022 Nov [cited 2023 Apr 2];155:579–88. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0022395622005416>.
11. Pauker SG, Kassirer JP. Therapeutic decision making: A cost-benefit analysis. *N Engl J Med*. 1975 Jul 31 [cited 2023 Apr 2];293(5):229–34. Available from: <http://www.nejm.org/doi/abs/https://doi.org/10.1056/NEJM197507312930505>.
12. Gradus JL, Rosellini AJ, Horváth-Puhó E, Jiang T, Street AE, Galatzer-Levy I et al. Predicting sex-specific nonfatal suicide attempt risk using machine learning and data from Danish national registries. *Am J Epidemiol*. 2021 Dec 1 [cited 2023 Apr 2];190(12):2517–27. Available from: <https://academic.oup.com/aje/article/190/12/2517/6239818>.

13. de la García Á, Blanco C, Olsson M, Wall MM. Identification of suicide attempt risk factors in a national US survey using machine learning. *JAMA Psychiatry*. 2021 Apr 1 [cited 2023 Apr 2];78(4):398. Available from: <https://jamanetwork.com/journals/jamapsychiatry/fullarticle/2774348>.
14. Machado Cdos, Ballester S, Cao PL, Mwangi B, Caldieraro B, Kapczynski MA, F. Prediction of suicide attempts in a prospective cohort study with a nationally representative sample of the US population. *Psychol Med*. 2022 Oct [cited 2023 Apr 2];52(14):2985–96. Available from: [https://www.cambridge.org/core/product/identifier/S0033291720004997/type/journal\\_article](https://www.cambridge.org/core/product/identifier/S0033291720004997/type/journal_article).
15. Lee J, Pak TY. Machine learning prediction of suicidal ideation, planning, and attempt among Korean adults: A population-based study. *SSM - Popul Health*. 2022 Sep [cited 2023 Apr 14];19:110231. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S2352827322002105>.
16. Janssens ACJW, Martens FK. Reflection on modern methods: Revisiting the area under the ROC Curve. *Int J Epidemiol*. 2020 Aug 1 [cited 2024 Jan 11];49(4):1397–403. Available from: <https://academic.oup.com/ije/article/49/4/1397/5714095>.
17. 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 [cited 2023 Apr 2];143(2):187–232. Available from: <http://doi.apa.org/getdoi.cfm?doi=10.1037/bul0000084>.
18. Firmann M, Mayor V, Vidal PM, Bochud M, Pécoud A, Hayoz D et al. The COLaus study: a population-based study to investigate the epidemiology and genetic determinants of cardiovascular risk factors and metabolic syndrome. *BMC Cardiovasc Disord*. 2008 Dec [cited 2023 Jul 5];8(1):6. Available from: <https://bmccardiovascdisord.biomedcentral.com/articles/https://doi.org/10.1186/1471-2261-8-6>.
19. Preisig M, Waeber G, Vollenweider P, Bovet P, Rothen S, Vandelur C et al. The PsyCoLaus study: methodology and characteristics of the sample of a population-based survey on psychiatric disorders and their association with genetic and cardiovascular risk factors. *BMC Psychiatry*. 2009 Dec [cited 2023 Apr 2];9(1):9. Available from: <http://bmcp psychiatry.biomedcentral.com/articles/https://doi.org/10.1186/1471-244X-9-9>.
20. Leboyer M, Barbe B, Gorwood P, Teheriani M, Allilaire JF, Preisig M, et al. Interview diagnostique pour les études génétiques. Paris: INSERM; 1995.
21. Nurnberger JI. Diagnostic Interview for Genetic Studies: Rationale, Unique Features, and Training. *Arch Gen Psychiatry*. 1994 Nov 1 [cited 2023 Apr 2];51(11):849–59. Available from: <http://archpsyc.jamanetwork.com/article.aspx?doi=10.1001/archpsyc.1994.03950110009002>.
22. American Psychiatric Association. DSM-IV-TR: Diagnostic and statistical manual of mental disorders. Author: Washington, DC; 2000.
23. Preisig M, Fenton BT, Matthey ML, Berney A, Ferrero F. Diagnostic Interview for Genetic Studies (DIGS): Inter-rater and test-retest reliability of the French version. *Eur Arch Psychiatry Clin Neurosci*. 1999 Aug 13 [cited 2023 Jun 2];249(4):174–9. Available from: <http://link.springer.com/https://doi.org/10.1007/s004060050084>.
24. Berney A, Preisig M, Matthey ML, Ferrero F, Fenton BT. Diagnostic interview for genetic studies (DIGS): inter-rater and test-retest reliability of alcohol and drug diagnoses. *Drug Alcohol Depend*. 2002 Jan [cited 2023 Apr 2];65(2):149–58. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0376871601001569>.
25. Endicott JA. Diagnostic Interview. The Schedule for Affective Disorders and Schizophrenia. *Arch Gen Psychiatry*. 1978 Jul 1 [cited 2023 Jun 2];35(7):837–44. Available from: <http://archpsyc.jamanetwork.com/article.aspx?doi=10.1001/archpsyc.1978.01770310043002>.
26. Leboyer M, Maier W, Teherani M, Lichtermann D, D'Amato T, Franke P et al. The reliability of the SADS-LA in a family study setting. *Eur Arch Psychiatry Clin Neurosci*. 1991 Dec [cited 2023 Jun 2];241(3):165–9. Available from: <http://link.springer.com/https://doi.org/10.1007/BF02219716>.
27. Jacobucci R, Grimm KJ. Machine learning and psychological research: The unexplored effect of measurement. *Perspect Psychol Sci*. 2020 May [cited 2023 Apr 2];15(3):809–16. Available from: <http://journals.sagepub.com/doi/10.1177/1745691620902467>.
28. Ogundimu EO, Altman DG, Collins GS. Adequate sample size for developing prediction models is not simply related to events per variable. *J Clin Epidemiol*. 2016 Aug [cited 2023 Apr 2];76:175–82. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0895435616300117>.
29. Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J Clin Epidemiol*. 2019 Jun [cited 2023 Apr 2];110:12–22. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0895435618310813>.
30. Lynam AL, Dennis JM, Owen KR, Oram RA, Jones AG, Shields BM et al. Logistic regression has similar performance to optimised machine learning algorithms in a clinical setting: application to the discrimination between type 1 and type 2 diabetes in young adults. *Diagn Progn Res*. 2020 Dec [cited 2023 Apr 2];4(1):6. Available from: <https://diagnprognres.biomedcentral.com/articles/https://doi.org/10.1186/s41512-020-00075-2>.
31. Nusinovi S, Tham YC, Chak Yan MY, Wei Ting DS, Li J, Sabanayagam C et al. Logistic regression was as good as machine learning for predicting major chronic diseases. *J Clin Epidemiol*. 2020 Jun [cited 2023 Apr 2];122:56–69. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0895435619310194>.
32. Song X, Liu X, Liu F, Wang C. Comparison of machine learning and logistic regression models in predicting acute kidney injury: A systematic review and meta-analysis. *Int J Med Inf*. 2021 Jul [cited 2023 Apr 2];151:104484. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S1386505621001106>.
33. Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification and regression trees. 1st ed. Routledge; 1984 [cited 2022 Mar 25]. Available from: <https://www.taylorfrancis.com/books/9781351460491>.
34. Hill RM, Oosterhoff B, Do C. Using machine learning to identify suicide risk: A classification tree approach to prospectively identify adolescent suicide attempters. *Arch Suicide Res*. 2020 Apr 2 [cited 2023 Apr 2];24(2):218–35. Available from: <https://www.tandfonline.com/doi/full/https://doi.org/10.1080/01381118.2019.1615018>.
35. Strobl C, Malley J, Tutz G. An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychol Methods*. 2009 Dec [cited 2023 Jun 7];14(4):323–48. Available from: <http://doi.apa.org/getdoi.cfm?doi=10.1037/a0016973>.
36. Torgo L, Branco P, Ribeiro RP, Pfahringer B. Resampling strategies for regression. *Expert Syst*. 2015 Jun [cited 2023 May 4];32(3):465–76. Available from: <https://onlinelibrary.wiley.com/doi/https://doi.org/10.1111/exsy.12081>.
37. R Core Team. R: a language and environment for statistical computing. 2020. Available from: <https://www.R-project.org/>.
38. Therneau T, Atkinson B, Ripley B. rpart: Recursive partitioning and regression trees. 2022. Available from: <https://cran.r-project.org/web/packages/rpart/index.html>.
39. Harrell FE Jr. rms: Regression modeling strategies. 2020. Available from: <https://cran.r-project.org/web/packages/rms/index.html>.
40. Saito T, Rehmsmeier M. Precrec: calculate accurate precision-recall and ROC (receiver operator characteristics) curves. 2023. Available from: <https://cran.r-project.org/web/packages/precrec/index.html>.
41. Wickham H. ggplot2: elegant graphics for data analysis. 2nd ed. New York: Springer International Publishing; 2016. p. 260. (Use R!).
42. Austin PC, Steyerberg EW. The Integrated Calibration Index (ICI) and related metrics for quantifying the calibration of logistic regression models. *Stat Med*. 2019 Sep 20 [cited 2023 Apr 3];38(21):4051–65. Available from: <https://onlinelibrary.wiley.com/doi/https://doi.org/10.1002/sim.8281>.
43. Andaur Navarro CL, Damen JAA, Takada T, Nijman SWJ, Dhiman P, Ma J et al. Completeness of reporting of clinical prediction models developed using supervised machine learning: a systematic review. *BMC Med Res Methodol*. 2022 Dec [cited 2023 May 26];22(1):12. Available from: <https://bmcmredresmethodol.biomedcentral.com/articles/https://doi.org/10.1186/s12874-021-01469-6>.
44. Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): Explanation and elaboration. *Ann Intern Med*. 2015 Jan 6 [cited 2023 May 26];162(1):W1–73. Available from: <https://www.acpjournals.org/doi/https://doi.org/10.7326/M14-0698>.
45. Vickers AJ, van Calster B, Steyerberg EW. A simple, step-by-step guide to interpreting decision curve analysis. *Diagn Progn Res*. 2019 Dec [cited 2023 Apr 2];3(1):18. Available from: <https://diagnprognres.biomedcentral.com/articles/https://doi.org/10.1186/s41512-019-0064-7>.
46. Vickers AJ, Elkin EB. Decision curve analysis: A novel method for evaluating prediction models. *Med Decis Making*. 2006 Nov [cited 2023 Apr 2];26(6):565–74. Available from: <http://journals.sagepub.com/doi/https://doi.org/10.1177/0272989X06295361>.
47. Vickers AJ, Cronin AM. Traditional Statistical Methods for Evaluating Prediction Models Are Uninformative as to Clinical Value: Towards a Decision Analytic Framework. *Semin Oncol*. 2010 Feb [cited 2024 Jan 16];37(1):31–8. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0093775409002292>.
48. Vickers AJ, Van Cluster B, Wynants L, Steyerberg EW. Decision curve analysis: confidence intervals and hypothesis testing for net benefit. *Diagn Progn Res*.

- 2023 Jun 6 [cited 2024 Jan 16];7(1):11. Available from: <https://diagnprognres.biomedcentral.com/articles/https://doi.org/10.1186/s41512-023-00148-y>.
49. Van Calster B, Vickers AJ. Calibration of risk prediction models: Impact on decision-analytic performance. *Med Decis Making*. 2015 Feb [cited 2023 Jun 7];35(2):162–9. Available from: <http://journals.sagepub.com/doi/https://doi.org/10.1177/0272989X14547233>.
  50. Van Calster B, Wynants L, Verbeek JFM, Verbakel JY, Christodoulou E, Vickers AJ et al. Reporting and interpreting decision curve analysis: A guide for investigators. *Eur Urol*. 2018 Dec [cited 2023 Apr 2];74(6):796–804. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0302283818306407>.
  51. Vickers AJ, Holland F. Decision curve analysis to evaluate the clinical benefit of prediction models. *Spine J*. 2021 Oct [cited 2023 Apr 2];21(10):1643–8. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S1529943021001121>.
  52. Luk JW, Pruitt LD, Smolenski DJ, Tucker J, Workman DE, Belsher BE. From everyday life predictions to suicide prevention: Clinical and ethical considerations in suicide predictive analytic tools. *J Clin Psychol*. 2022 Feb [cited 2023 May 2];78(2):137–48. Available from: <https://onlinelibrary.wiley.com/doi/https://doi.org/10.1002/jclp.23202>.
  53. Jacobucci R, Littlefield AK, Millner AJ, Kleiman EM, Steinley D. Evidence of inflated prediction performance: A commentary on machine learning and suicide research. *Clin Psychol Sci*. 2021 Jan [cited 2023 Apr 2];9(1):129–34. Available from: <http://journals.sagepub.com/doi/https://doi.org/10.1177/2167702620954216>.
  54. Davis SE, Greevy RA, Lasko TA, Walsh CG, Matheny ME. Detection of calibration drift in clinical prediction models to inform model updating. *J Biomed Inform*. 2020 Dec [cited 2023 Apr 3];112:103611. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S1532046420302392>.
  55. Steyerberg EW, Harrell FE. Prediction models need appropriate internal, internal–external, and external validation. *J Clin Epidemiol*. 2016 Jan [cited 2023 May 26];69:245–7. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0895435615001754>.
  56. Andaur Navarro CL, Damen JAA, van Smeden M, Takada T, Nijman SWJ, Dhiman P et al. Systematic review identifies the design and methodological conduct of studies on machine learning-based prediction models. *J Clin Epidemiol*. 2023 Feb [cited 2023 Apr 2];154:8–22. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0895435622003006>.
  57. Barry MJ, Nicholson WK, Silverstein M, Chelmos D, Coker TR, Davidson KW et al. Screening for Depression and Suicide Risk in Adults: US Preventive Services Task Force Recommendation Statement. *JAMA*. 2023 Jun 20 [cited 2024 Jan 15];329(23):2057. Available from: <https://jamanetwork.com/journals/jama/fullarticle/2806144>.
  58. Thoms BD, Markham S, Rice DB, Ziegelstein RC. Screening for depression and anxiety in general practice. *BMJ*. 2023 Jul 17 [cited 2024 Jan 15];p1615. Available from: <https://www.bmj.com/lookup/doi/10.1136/bmj.p1615>.
  59. Kompa B, Snoek J, Beam AL. Second opinion needed: communicating uncertainty in medical machine learning. *Npj Digit Med*. 2021 Jan 5 [cited 2024 Jan 15];4(1):4. Available from: <https://www.nature.com/articles/s41746-020-00367-3>.
  60. Shariat SF, Capitano U, Jeldres C, Karakiewicz PI. Can nomograms be superior to other prediction tools? *BJU Int*. 2009 Feb [cited 2024 Jan 16];103(4):492–7. Available from: <https://bjui-journals.onlinelibrary.wiley.com/doi/https://doi.org/10.1111/j.1464-410X.2008.08073.x>.
  61. Joyce DW, Kormilitzin A, Smith KA, Cipriani A. Explainable artificial intelligence for mental health through transparency and interpretability for understandability. *Npj Digit Med*. 2023 Jan 18 [cited 2024 Jan 16];6(1):6. Available from: <https://www.nature.com/articles/s41746-023-00751-9>.
  62. van den Goorbergh R, van Smeden M, Timmerman D, Van Calster B. The harm of class imbalance corrections for risk prediction models: illustration and simulation using logistic regression. *J Am Med Inform Assoc*. 2022 Aug 16 [cited 2023 Apr 2];29(9):1525–34. Available from: <https://academic.oup.com/jamia/article/29/9/1525/6605096>.
  63. Binuya MAE, Engelhardt EG, Schats W, Schmidt MK, Steyerberg EW. Methodological guidance for the evaluation and updating of clinical prediction models: a systematic review. *BMC Med Res Methodol*. 2022 Dec 12 [cited 2023 Apr 2];22(1):316. Available from: <https://bmcmmedresmethodol.biomedcentral.com/articles/https://doi.org/10.1186/s12874-022-01801-8>.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.