Internet Interventions 21 (2020) 100337

Contents lists available at ScienceDirect

Internet Interventions

journal homepage: www.elsevier.com/locate/invent

Applying machine learning on health record data from general practitioners to predict suicidality

Kasper van Mens[a,d,*], Elke Elzinga[b], Mark Nielen[c], Joran Lokkerbol[d], Rune Poortvliet[c], Gé Donker[c], Marianne Heins[c], Joke Korevaar[c], Michel Dückers[c], Claire Aussems[c], Marco Helbich[e], Bea Tiemens[f], Renske Gilissen[b], Aartjan Beekman[g,h], Derek de Beurs[d,i]

[a] Altrecht Mental Healthcare, Utrecht, the Netherlands
[b] 113 Suicide Prevention, Amsterdam, the Netherlands
[c] Nivel, Netherlands Institute for Health Services Research, Utrecht, the Netherlands
[d] Trimbos Institute (Netherlands Institute of Mental Health), Utrecht, the Netherlands
[e] Human Geography and Spatial Planning, Utrecht University, Utrecht, the Netherlands
[f] Behavioural Science Institute, Radboud University, Nijmegen, the Netherlands
[g] Psychiatry, Amsterdam Public Health (research institute), Amsterdam UMC, Vrije Universiteit Amsterdam, the Netherlands
[h] GGZ inGeest Specialized Mental Health Care, Amsterdam, the Netherlands
[i] Clinical Psychology, Amsterdam Public Health, Vrije Universiteit Amsterdam, the Netherlands

A R T I C L E I N F O

Keywords:
Suicide
General practice
Electronic health records
Machine learning

A B S T R A C T

Background: Suicidal behaviour is difficult to detect in the general practice. Machine learning (ML) algorithms using routinely collected data might support General Practitioners (GPs) in the detection of suicidal behaviour. In this paper, we applied machine learning techniques to support GPs recognizing suicidal behaviour in primary care patients using routinely collected general practice data.
Methods: This case-control study used data from a national representative primary care database including over 1.5 million patients (Nivel Primary Care Database). Patients with a suicide (attempt) in 2017 were selected as cases (N = 574) and an at risk control group (N = 207,308) was selected from patients with psychological vulnerability but without a suicide attempt in 2017. RandomForest was trained on a small subsample of the data (training set), and evaluated on unseen data (test set).
Results: Almost two-third (65%) of the cases visited their GP within the last 30 days before the suicide (attempt). RandomForest showed a positive predictive value (PPV) of 0.05 (0.04–0.06), with a sensitivity of 0.39 (0.32–0.47) and area under the curve (AUC) of 0.85 (0.81–0.88). Almost all controls were accurately labeled as controls (specificity = 0.98 (0.97–0.98)). Among a sample of 650 at-risk primary care patients, the algorithm would label 20 patients as high-risk. Of those, one would be an actual case and additionally, one case would be missed.
Conclusion: In this study, we applied machine learning to predict suicidal behaviour using general practice data. Our results showed that these techniques can be used as a complementary step in the identification and stratification of patients at risk of suicidal behaviour. The results are encouraging and provide a first step to use automated screening directly in clinical practice. Additional data from different social domains, such as employment and education, might improve accuracy.

1. Introduction

Suicide is a major public health problem, with an estimated 800,000 deaths a year worldwide (World Health Organization, 2014). Suicide attempts are about 20 times more frequent (World Health Organization, 2014). The WHO deemed suicide prevention a global imperative and urges countries to develop and implement a national strategy on suicide prevention.

In countries with a health system including a general practitioner (GP), patients frequently contact their GP prior to engaging in suicidal behaviour. Studies found that around 50% was in contact within the month before the suicide (attempt) (de Beurs et al., 2016; Luoma et al.,

* Corresponding author at: Trimbos Institute (Netherlands Institute of Mental Health), Utrecht, the Netherlands.
E-mail address: ka.van.mens@altrecht.nl (K. van Mens).

https://doi.org/10.1016/j.invent.2020.100337
Received 24 January 2020; Received in revised form 29 June 2020; Accepted 20 July 2020
Available online 27 August 2020

2002; Stene-Larsen and Reneflot, 2019). GPs indicated, in hindsight, that suicidal behaviour was a serious risk in 30% of these final contacts (Marquet et al., 2005; de Beurs et al., 2016). Suicidal ideation, associated with increased risk for future suicidal behaviour (Franklin et al., 2017), is generally hard to detect. Few patients express their suicidal thoughts to GPs unsolicited. Pearson et al. (2009) reported that only 15% of the patients had expressed suicidal thoughts during the final consultation before their suicide. This stresses the importance of proactive suicide exploration by GPs. However, a recent Dutch study showed that among depressed patients, a well-established risk group for suicidal behaviour (Hawton et al., 2013), GPs explored suicidality in only 44% of these patients (Elzinga et al., 2019). Therefore, this group may still be too large and unspecified for GPs to systematically explore suicidal feelings. An algorithm based automated pre-screening tool might help the GP identify which patients are most likely at risk for suicidal behaviour.

Machine Learning (ML) techniques can be used to detect patterns in large sets of patient data containing many different predictors (Kuhn and Johnson, 2013; Walsh et al., 2017; Iniesta et al., 2016; Franklin et al., 2017). Earlier studies showed encouraging results in predicting suicide (attempts) among patients using only data from electronic health records (Kessler et al., 2015; Barak-Corren et al., 2017; Simon et al., 2018; Walsh et al., 2017). Overall, the studies produced accurate classification models. However, since suicide is such a rare event, even the slightest prediction errors result in high false positive rates, meaning that patients will be incorrectly classified as suicidal. Therefore, predictive models should not be seen as substitution but as complementary step in the identification and stratification of patients at risk for future suicidal behaviour. A possible method for identifying primary care patients at risk for suicidal behaviour, would include, first, selecting all patients with psychological vulnerability, and second, narrowing this group down using a predictive algorithm. The last step would include GPs actively assessing suicidal ideation among the marked patients.

This study aims at supporting GPs in the second step, the identification of suicidal behaviour, by applying ML techniques to a large and representative primary care database (the Nivel database). Ultimately, the goal would be to create an automatic signalling system that warns GPs during consultations for future suicidal behaviour of patients. Although this technology has the potential to substantially strengthen the timely detection and treatment of patients at risks, currently it is still in its infancy. This study seeks to contribute to the uptake of ML algorithms in everyday clinical practice by testing its predictive accuracy on real-life consultation data from medical records kept by GPs.

2. Methods

2.1. Dataset

For this case-control study, we used the Nivel Primary Care Database containing a representative sample of about 500 general practices including over 1.5 million registered patients (approximately 10% of the Dutch population) in 2017. Nivel collects routinely recorded information from electronic health records systems. In Dutch healthcare, insurance is mandatory, and everybody is assigned to a GP (Kroneman et al., 2016). This means that the database is representative of the total population. It is important to note that the registrations do not always reflect face-to-face consultations, but may also include administrative acts or consultatations over the phone. We compared patients with a registration of suicide (attempt) with an at-risk control group without a recorded suicide (attempt) in 2017.

2.2. Ethical statement

Dutch law allows the use of electronic health records for research purposes under certain conditions. According to this legislation, neither obtaining informed consent from patients nor approval by a medical ethics committee is obligatory for this type of observational studies containing no directly identifiable data (Dutch Civil Law, Article 7: 458). This study has been approved according to the governance code of Nivel Primary Care Database, under number NZR-00318.009.

2.3. Cases

Consultations and other registrations are labeled in the data with a diagnosis according to the International Classification of Primary Care (ICPC) system (Lambert and Wood, 1987). ICPC codes, 685 in total, are clustered in 17 different chapters; such as the psychological (P) or the social (Z) chapter. Since both suicide and suicide attempt are registered within the P-chapter as P77, they cannot be discriminated from each other. Patients with a registration of P77 in 2017 were selected as cases. Patients may have had multiple registrations of P77, sometimes even within a week, likely referring to one suicide (attempt). Therefore, we operationalized our cases by selecting the first recording of P77 in 2017 as time of the suicide (attempt). Patients who also had a recording of P77 in 2016 or 2015 were excluded, to ensure that the first registration in 2017 most likely referred to a new suicidal event. This yielded a total of 574 new suicide (attempt) cases, which implies a prevalence of 0.12% (in a population of 486,488 primary care patients with a least one registration in 2017). We then added information about primary care service use in the 12 months before the event. Forty of these patients were excluded from the model building process, since they did not have any registration in the previous 12 months and cannot be identified by automatic signalling technology implemented in a GP's system.

2.4. Controls

Our goal is identifying suicidal behaviour among patients from an at-risk group. We selected patients with psychological vulnerability as at risk, which are patients with at least one psychological related registration in their patient file (N = 207,308). For each of the controls, one year of registration data prior to the first recording in 2017 was included as well. Just as the cases, control patients without any registrations in 2017 were excluded further on in the model building process.

2.5. Time to event

Within the Nivel database, patients could have a registration for any of the ICPC codes on any day of the year. In order to create computationally feasible and meaningful features, months were used as time units. All registrations in 1 to 30 days before the suicide (attempt) or control event are grouped in 'month before', registrations in 31 days up to 61 days in '2nd month before', registrations in 62 up to 92 days in '3rd month before' and each registration in 93 up to 365 days in '4th–12th months before'. The latter group is seen as a baseline period of 9 months, which is used to calculate the relative increase in registration per month in the first, second and third month before.

2.6. Feature engineering

Different from data collected for research purposes, registration data is not collected to answer a specific research question. The raw data in the Nivel database contained many details about service use that most likely would not be relevant to predict suicidal behaviour. Therefore, feature engineering from the raw data is necessary to identify meaningful patterns of service use (James et al., 2015). Several feature types were created based on expert knowledge and literature (Kämpfer et al., 2016; Leavey et al., 2016; Pedersen et al., 2019; Wiborg et al., 2013; Windfuhr et al., 2016; de Beurs et al., 2016; Marquet et al., 2005).

1. The number of registrations for each ICPC cluster in the first, second and third month before the suicide (attempt).
2. The average number of registrations per month for each cluster in the baseline-period (4–12 months before the suicide (attempt)).
3. The relative increase in total registrations in the first, second and third month compared to the baseline number of registrations.
4. The relative increase in psychological (P), social (Z) and Medically Unspecified Physical Symptoms (MUPS; see Appendix) registrations in the first, second, and third month compared to their respective baselines.
5. The number of registrations for each of the individual ICPC codes from the P-cluster in the first, second and third month before the suicide (attempt).

2.7. Descriptive statistics

First, we present basic statistics to describe the at-risk control group and the health care uptake of suicide cases. We show how many patients visited the GP in the week and months before the registration of a suicide (attempt) and we describe the reason(s) for the last consultation before the suicide (attempt). This description includes the patients without any consultation in the year prior to the event, who were excluded in the model building process.

2.8. Modelling strategy

In this stage, our data included 534 cases. Based on ML principles, we created a training sample (70%) to build our models and a test sample (30%) to evaluate our final model on unseen test data. We created a random subset of the cases (N = 373) and added 1865 random controls to create the training set (N = 2238). To improve classifier performance, we under-sampled the controls to artificially increase the prevalence of suicide (attempt) to 17% in this hypothetical training situation. As a rule of thumb, around 17% can be seen as the threshold for a balanced dataset (He and Garcia, 2009). We used 10 times 10-fold cross validation to estimate the performance of the models and chose the final model. The final model was evaluated once in a real-world setting on an unseen test sample consisting of the remaining 161 cases. To represent the real-world prevalence of suicide (attempts) in the at-risk group (0.28%) of this dataset, 53,666 controls were added to the test set.

Guided by model competition studies (Hagenauer et al., 2019; Fernández-Delgado et al., 2014), a *randomForest* algorithm was used to create the predictive model. A *randomForest* is an example of ensemble learning, which is an algorithm that combines multiple predictors to make a single prediction. In the case of a *randomForest*, multiple decision trees are combined to create a 'forest'. Each single decision tree recursively partitions the data into smaller subsets until all patients in a subset have similar outcomes (case or control) or until further partitioning does not add value to the predictions. A decision tree has the advantage of being able to model complex interactions and non-linear relationships. The downside of a random forest model is that the combination of multiple decision trees can create a black box in which the relation between input and prediction can be hard to interpret. The package *randomForest* as implemented in the statistical software R was used (Breiman et al., 2015).

2.9. Performance metrics

Model performance was evaluated using sensitivity, specificity, Area Under the Curve (AUC), Positive Predictive Value (PPV) and Balanced Accuracy. Sensitivity describes the proportion of cases that is correctly classified as such among all cases. Specificity indicates the proportion of controls that is correctly classified as controls among all controls. AUC summarizes the trade-off between sensitivity and specificity. The PPV is the percentage of correctly predicted cases among all cases predicted as such (Kuhn and Johnson, 2013). The balanced accuracy describes the average proportion of correct classifications. The R packages *caret* and *pROC* were used for this (Kuhn et al., 2017; Robin et al., 2011).

2.10. Variable importance

Although random forests do not yield predictor coefficients to describe the impact of variables like traditional models do, it is possible to order variables according to their importance (Kuhn and Johnson, 2013). The importance of a variable describes how much a variable contributes to the improvement of the model. We use the permutation importance option as implemented in the *randomForest* package to determine variable importance. It indicates the mean decrease in classification performance after permutating a specific variable over all trees of the random forest (Breiman et al., 2015).

2.11. Role of the funding source

The funders had no role in study design, data collection, data analysis, data interpretation or writing of this article. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication. Opinions, interpretations, conclusions and recommendations are those of the authors and are not necessarily endorsed by the funders.

3. Results

Table 1 shows the demographic features of the 574 suicide (attempt) cases and the 207,308 controls. Cases are, compared to controls, more often male and younger aged. They also have a higher mean number of registrations in the previous year and a higher percentage consulted their GP at least once in the previous year as opposed to controls. Further, cases consulted about twice as much for psychological and social problems, and they more often had a recording of a Medically Unexplained Physical Symptoms (MUPS).

In Fig. 1, the number of cases presenting themselves at the GP in a given time period is given. About a quarter of the cases had a

Table 1
Demographic features of the cases and the at-risk control group.

	Cases	At-risk control group[a]	p-Value
Number	574	207,308	
Percentage male	46%	42%	0.024
Mean age (SD)	46 (17)	52 (19)	< 0.001
Mean number of registrations in one year (SD)	18 (19)	10 (12)	< 0.001
Percentage with at least one registration	93%	86%	< 0.001
Percentage of total registrations for psychological reasons	52%	23%	< 0.001
Percentage of total registrations for social reasons	12%	6%	< 0.001
Percentage with at least one MUPS registration	60%	51%	< 0.001

[a] Patients with at least one P-registration in registration data in the period 2011–2017. MUPS = Medically Unexplained Psychological Symptoms.

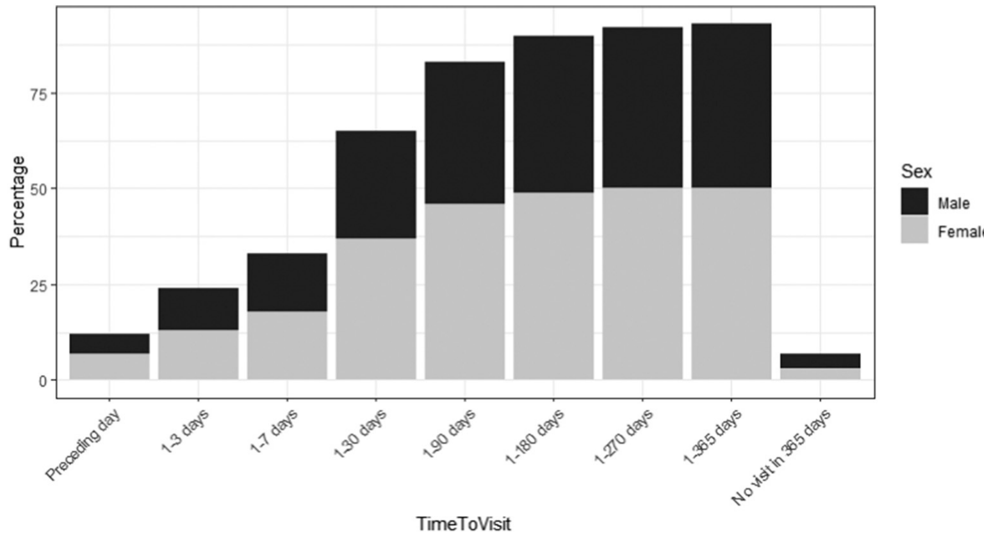

Fig. 1. Number of cases with a registration in their GP file prior to a suicide (attempt) (N = 574).

registration in the 3 days before and one-third in the week before their suicide (attempt). About two thirds (65%) of the patients visited their GP one month and 83% visited their GP in the three months prior to their suicide (attempt). Only 40 patients (7%) did not visit the GP in the year before the suicide (attempt).

Table 2 shows the 10 most often mentioned topics of the last registration before the suicide (attempt). It shows large variation in topics: in 10% of the cases depression was the topic of the last recorded registration, chronic alcohol abuse, and diabetes both in 3%. In 2% of the cases, the topic of final registration was a personality disorder, no disease, essential hypertension, crisis/stress reaction, other psychological symptoms, or anxiety.

3.1. Performance metrics

The confusion matrix and performance metrics of the random forest are shown in Table 3. Of the 161 cases in the test set, the algorithm identified 63 correctly as cases, resulting in a sensitivity of 0.39 (0.32–0.47). Almost all controls were identified correctly as controls, giving a specificity of 0.98 (0.97–0.98). The PPV is 0.05 (0.04–0.06) and the balanced accuracy is 0.68.

3.2. Variable importance

Table 4 shows the variable importance as identified by the *randomForest*. The relative healthcare uptake in one month before the suicide (attempt) is the best predictor. Second-best is the number of registrations in the P-cluster in the month before the suicide (attempt).

Table 2
Reason for last registration (ICPC-code) prior to suicide (attempt) of cases who consulted a GP (N = 534).

Topic of last registration (chapter)	Cases
Depression (P)	53 (10%)
Chronic alcohol abuse (P)	16 (3%)
Diabetes (other)	14 (3%)
Affective psychosis (P)	13 (2%)
Personality disorder (P)	13 (2%)
No disease (other)	11 (2%)
Essential hypertension (other)	11 (2%)
Crisis/stress reaction (P)	10 (2%)
Other psychological symptoms (P)	10 (2%)
Anxiety (P)	10 (2%)

P = psychological.

Table 3
Prediction metrics of the random forest.

	Random forest
Area under the curve (95% CI)	0.82 (0.78–0.86)
Sensitivity	0.39 (0.32–0.47)
Specificity	0.98 (0.97–0.98)
PPV	0.05 (0.04–0.06)
Balanced accuracy	0.68

	Actual case	Actual control
Predicted case	63	1298
Predicted control	98	52,368

Third, the age of the patient, followed by the relative increase in MUPS registrations in the month before the suicide (attempt) compared to baseline.

4. Discussion

This is one of the first studies to apply ML techniques to predict a suicide (attempt) using routinely collected GP data. We used the ICPC registrations to describe healthcare uptake patterns prior to a suicide (attempt). Subsequently we used these patterns to predict suicide (attempts). Descriptive statistics showed that almost all cases (93%) consulted their GP in the year before their suicide (attempt). More than half of the cases had a registration in the month before the suicide (attempt) and about a third in the week before. This indicates that for the majority of suicidal patients, there is an opportunity for the GP to signal suicidal behaviour.

The random forest model resulted in a PPV of 0.05 (0.04–0.06), sensitivity of 0.39 (0.32–0.47) and AUC of 0.82 (0.78–0.86). Specificity was 0.98 (0.97–0.98), meaning that almost all controls were accurately labeled as controls (52,368 out of 53,666). These results are in line with a recent meta review (Belsher et al., 2019). The PPV of an algorithm will depend on the prevalence of suicidal behaviour in the population in which the algorithm will be used. The lower the prevalence, the lower the PPV. Some studies in the review of Belsher et al. (Belsher et al., 2019) achieved a higher PPVs, however these studies used an ultra-high-risk sample in which the prevelance was up to 50%. In our real-world at-risk sample, with a prevalence of 0.28%, a high PPV is very hard to achieve. Predictive models can be build and tuned to lay more emphasis on either the positive or negative cases. By focusing more on

Table 4
Ranking of variable importance as identified by the random forest model.

Rank	Variable
1	Relative healthcare uptake (all registrations) 1 month before compared to baseline
2	Number of P-registrations 1 month before
3	Age
4	Relative healthcare uptake MUPS-registrations 1 month before compared to baseline
5	Number of MUPS-registrations 1 month before
6	Relative healthcare uptake P-registrations 1 month before compared to baseline
7	Number of depression registrations 1 month before
8	Relative healthcare uptake (all registrations) 3 months before compared to baseline
9	Number of P-registrations 2 months before
10	Relative healthcare uptake MUPS-registrations 3 months before compared to baseline

the positive case and thereby increasing the sensitivity, the false negative rate will be reduced. Contrary, by focusing on the negative cases, specificity will increase and false positives will be reduced. The optimal balance between sensitivity and specificity will depend on the context of the research. The impact of a low PPV depends on how the algorithm is used in clinical practice. In our design, the algorithm is used as a complementary step in suicide risk assessment. Although our algorithms were evaluated in a simplified case-control design, we translated the performance of the *randomForest* model to the following real-world example: Among a sample of 650 at-risk primary care patients, 20 patients would be labeled as high-risk by the algorithm (1 in 33 patients). Of those, 1 would be an actual case and, additionally, 1 case would be missed by the algorithm. These identified high-risk patients can be screened for suicide ideation, which is deemed good clinical practice (Van Hemert et al., 2012). This example shows that even with a high specificity (0.98), the identification of suicidal behaviour is correct in only 5% of the cases. However, the GP can now assess suicide ideation in a group of 20 patients instead of 650 patients, in which the prevalence is 18 times higher (5% compared to 0.28%). More worrisome is that the algorithm also missed one in two cases. There seems to be a group of patients that show no discriminative pattern in health care uptake at the GP compared to controls. GPs identified about one-third of their patients correctly as at risk for suicide (de Beurs et al., 2016; Pearson et al., 2009). Future studies should reveal whether GPs and the algorithm identify the same patients as at risk.

The major strength of the paper is that we applied a novel algorithm to a set of routinely available data. This research is a first step towards an automated screening device which could be implemented in the registration system of the GP. The growth of routinely available data, and data from other sources such as wearables could be integrated to further improve performance. The development of stronger algorithms provides opportunities for exploration, hypothesis generation and prediction. Another strength of this study is that we included the frequency of registration in the run-up to the suicide (attempt). Similar to Walsh et al. (Walsh et al., 2017), we compared registration rate in the final month with the registration rate at baseline (month 4–12 prior to the suicide (attempt)), this enabled us to distinguish changes in personal registration pattern of patients. As expected, this factor turned out to be (one of) the most important predictor(s) for suicide (attempts). Additionally, this is among the few studies that offer insights into the reason of consultation prior to a suicide (attempt).

There were some limitations in this study that need to be addressed. First, in this study we applied a predictive algorithm in a case-control design, in which we selected cases and compared them with random patients from an at-risk control sample. The algorithm was evaluated on patient level in which they predicted whether the next registration would be a suicide (attempt). However, identifying suicidal behaviour at an earlier consultation would also be clinically relevant. Further analysis should investigate whether algorithms can be evaluated on consultation level, meaning that patients can be flagged for risk on future suicidal behaviour during multiple consultations. This would also increase the intervention options for GPs. Second, we applied the algorithms to data from an at-risk control population based on psychological vulnerability (recorded P-consultation). The idea is that the algorithms serve as a complementary step in further stratification of patients at risk of suicidal behaviour. It should be determined whether this sub population is the most appropriate population to stratify by an algorithm or whether there is more added value in a different (at-risk) population. A different population would require a new model building process. Third, our data includes an underrepresentation of suicide attempts. Many registrations include severe attempts requiring health care or are preceded by preceded by a letter from a formal authority, such as the police or the emergency department of a hospital. Less severe suicide attempts are less likely to get registered in the GP system. Another important limitation is that we were not able to distinguish suicides from attempts. Since these are both registered within one code (P77), it is not possible to further differentiate our outcome measure. To disentangle the suicides form attempts, the primary care database should be linked to the National Statistics' cause-of-death register. This would not only enable future researchers to differentiate in type of suicidal behaviour, but also enrich the dataset with other relevant variables such as income level, employment status, ethnicity and living situation, which may add to the accuracy of the algorithm.

5. Conclusion

In this study, we applied machine learning to predict suicidal behaviour using general practice data. Our results showed that these techniques can be used as a complementary step in the identification and stratification of patients at risk of suicidal behaviour. The results are encouraging and provide a first step to use automated screening directly in clinical practice. Additional data from different social domains, such as employment and education, might improve accuracy.

Funding

Netherlands Organisation for Health Research and Development (ZONMW), Dutch ministry of Health.

Contributors

DPdB designed and prepared the study. KvM performed the data analysis. KvM, DPdB and EE drafted the manuscript. All authors made critical revisions, contributed to, and approved the final manuscript.

Declaration of competing interest

All authors declare no competing interests.

Acknowledgements

The Nivel database is funded by a long term grant from the Dutch

Ministry of Health. Part of the writing and analysis of this study was funded by the fellowship mental health of DdB, awarded by the Netherlands Organisation for Health Research and Development (ZONMW). Project number is 636320002. EE was funded by the Dutch Ministry of Health, Welfare and Sport grant number 326961, 2017.

Appendix 1. Overview of MUPS-related ICPC-codes (Yzermans et al., 2016)

General and Unspecified

A04 Weakness/tiredness general

Digestive

D01 Abdominal pain/cramps general
D02 Abdominal pain epigastric
D06 Abdominal pain localized other
D09 Nausea
D11 Diarrhoea
D12 Constipation

Eye

F01 Eye pain
F02 Red eye

Ear

H01 Ear pain/earache
H02 Hearing complaint
H03 Tinnitus, ringing/buzzing ear
H13 Plugged feeling ear

Cardiovascular

K03 Cardiovascular pain NOS
K04 Palpitations/awareness of heart

Musculoskeletal

L01 Neck symptom/complain
L02 Back symptom/complaint
L03 Low back symptom/complaint
L04 Chest symptom/complaint
L05 Flank/axilla symptom/complai
L08 Shoulder symptom/complaint
L09 Arm symptom/complaint
L10 Elbow symptom/complaint
L11 Wrist symptom/complaint
L12 Hand/finger symptom/complaint
L13 Hip symptom/complaint
L14 Leg/thigh symptom/complaint
L15 Knee symptom/complaint
L17 Foot/toe symptom/complaint
L18 Muscle pain

Neurological

N01 Headache
N02 Tension headache
N05 Tingling fingers/feet/toes
N17 Vertigo/dizziness

Psychological

P01 Feeling anxious/nervous/tense
P02 Acute stress reaction
P03 Feeling depressed
P04 Feeling/behaving irritable/angry
P06 Sleep disturbance
P20 Memory disturbance

Respiratory

R02 Shortness of breath/dyspnoea
R03 Wheezing
R04 Breathing problem, other
R05 Cough
R07 Sneezing/nasal congestion
R29 Respiratory symptom/complaint other

Skin

S01 Pain/tenderness of skin
S06 Rash localized
S07 Rash generalized

Endocrine/Metabolic and Nutritional

T07 Weight gain
T08 Weight loss

References

Barak-Corren, Y., et al., 2017. Predicting suicidal behavior from longitudinal electronic health records. Am. J. Psychiatr. 174 (2), 154–162.
Belsher, B.E., et al., 2019. Prediction models for suicide attempts and deaths: a systematic review and simulation. JAMA Psychiatry 76 (6), 642–651.
de Beurs, D.P., et al., 2016. Trends in suicidal behaviour in Dutch general practice 1983–2013: a retrospective observational study. BMJ Open 6 (5), e010868.
Breiman, L., et al., 2015. The randomForest Package.
Elzinga, E., et al., 2019. Discussing suicidality with depressed patients: an observational study in Dutch sentinel general practices. BMJ Open 9 (4).
Fernández-Delgado, M., et al., 2014. Do We Need Hundreds of Classifiers to Solve Real World Classification Problems?.
Franklin, J.C., et al., 2017. Risk factors for suicidal thoughts and behaviors: a meta-analysis of 50 years of research. Psychol. Bull. 143 (2), 187–232.
Hagenauer, J., Omrani, H., Helbich, M., 2019. Assessing the Performance of 38 Machine Learning Models: The Case of Land Consumption Rates in Bavaria. International Journal of Geographical Information Science, Germany.
Hawton, K., et al., 2013. Risk factors for suicide in individuals with depression: a systematic review. J. Affect. Disord. 147 (1–3), 17–28.
He, H., Garcia, E.A., 2009. Learning from imbalanced data. IEEE Trans. Knowl. Data Eng. 21 (9), 1263–1284.
Iniesta, R., Stahl, D., McGuffin, P., 2016. Machine learning, statistical learning and the future of biological research in psychiatry. Psychol. Med. 46 (12), 2455–2465.
James, G., et al., 2015. An Introduction to Statistical Learning With Applications in R. Springer, New York, NY.
Kämpfer, N., et al., 2016. Suicidality in patients with somatoform disorder – the speechless expression of anger? Psychiatry Res. 246, 485–491.
Kessler, R.C., et al., 2015. Predicting suicides after psychiatric hospitalization in US army soldiers: the Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS). JAMA Psychiatry 72 (1), 49–57.
Kroneman, M., et al., 2016. Netherlands Health System Review.
Kuhn, M., Johnson, K., 2013. Applied Predictive Modeling. Springer New York, New York, NY.
Kuhn, M., Wing, Jed, Weston, Steve, Williams, Andre, Keefer, Chris, Engelhardt, Allan, Cooper, Tony, Mayer, Zachary, Kenkel, Brenton, the R Core Team, Benesty, Michael, Lescarbeau, Reynald, Ziem, Andrew, Scrucca, Luca, Tang, Yuan, Candan, Can, Hunt, T., 2017. Contributions from. caret: Classification and Regression Training.
Lambert, H., Wood, M. (Eds.), 1987. ICPC: International Classification of Primary Care. In Oxford: Oxford University Press. Oxford University Press, USA.
Leavey, G., et al., 2016. Patterns and predictors of help-seeking contacts with health services and general practitioner detection of suicidality prior to suicide: a cohort analysis of suicides occurring over a two-year period. BMC Psychiatry 16 (1), 120.
Luoma, J.B., Martin, C.E., Pearson, J.L., 2002. Contact with mental health and primary care providers before suicide: a review of the evidence. Am. J. Psychiatr. 159 (6), 909–916.
Marquet, R.L., et al., 2005. The epidemiology of suicide and attempted suicide in Dutch general practice 1983–2003. BMC Family Practice 6 (1), 45.
Pearson, A., et al., 2009. Primary care contact prior to suicide in individuals with mental illness. Br. J. Gen. Pract. 59 (568), 825–832.

Pedersen, H.S., et al., 2019. Frequency of health care utilization in the year prior to completed suicide: a Danish nationwide matched comparative study V. De Luca, ed. PLoS ONE 14 (3), e0214605.
Robin, X., et al., 2011. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics 12 (1), 77.
Simon, G.E., et al., 2018. Predicting suicide attempts and suicide deaths following outpatient visits using electronic health records. Am. J. Psychiatr. 175 (10), 951–960.
Stene-Larsen, K., Reneflot, A., 2019. Contact with primary and mental health care prior to suicide: a systematic review of the literature from 2000 to 2017. Scandinavian Journal of Public Health 47 (1), 9–17.
Van Hemert, A.M., et al., 2012. Multidisciplinaire richtlijn diagnostiek en behandeling van suïcidaal gedrag. De Tijdstroom, Utrecht.
Walsh, C.G., Ribeiro, J.D., Franklin, J.C., 2017. Predicting risk of suicide attempts over time through machine learning. Clin. Psychol. Sci. 5 (3), 457–469.
Wiborg, J.F., et al., 2013. Suicidality in primary care patients with somatoform disorders. Psychosom. Med. 75 (9), 800–806.
Windfuhr, K., et al., 2016. Suicide risk linked with clinical consultation frequency, psychiatric diagnoses and psychotropic medication prescribing in a national study of primary-care patients. Psychol. Med. 46 (16), 3407–3417.
World Health Organization, 2014. Preventing Suicide. (A global imperative A global imperative).
Yzermans, J., et al., 2016. Assessing non-specific symptoms in epidemiological studies: development and validation of the Symptoms and Perceptions (SaP) questionnaire. Int. J. Hyg. Environ. Health 219 (1), 53–65.