Psychiatry Research 350 (2025) 116537

Contents lists available at ScienceDirect

Psychiatry Research

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

RIPTOSO: The development of a screening tool for adverse events during forensic-psychiatric inpatient treatments of offenders with schizophrenia spectrum disorders

Andreas B. Hofmann [a,*], Marc Dörner [b,c], Philip E. Klassen [d], Lena Machetanz [a,e], Johannes Kirchebner [e]

[a] Adult Psychiatry and Psychotherapy, University Hospital of Psychiatry Zurich, Faculty of Medicine, University of Zurich, Zurich, Switzerland
[b] Department of Consultation-Liaison-Psychiatry and Psychosomatic Medicine, University Hospital Zurich, University of Zurich, 8091 Zurich, Switzerland
[c] German Center for Neurodegenerative Diseases (DZNE) within the Helmholtz Association, 39120 Magdeburg, Germany
[d] Ontario Shores Centre for Mental Health Sciences, Canada
[e] Forensic Psychiatry and Psychotherapy, University Hospital of Psychiatry Zurich, Faculty of Medicine, University of Zurich, Zurich, Switzerland

A R T I C L E I N F O

Keywords:
Machine learning
Artificial intelligence
Forensic psychiatry
Schizophrenia spectrum disorders
Offenders
Psychotic disorders
Screening tool

A B S T R A C T

Adverse events such as compulsory measures, absconding, illicit substance use, self-harm, aggressive behavior, and prolonged hospitalization pose significant challenges in forensic psychiatric inpatient care. This study introduces a machine learning-based tool to predict these events in patients with schizophrenia spectrum disorders (SSD) upon admission. Data from 370 court-mandated forensic inpatients treated at an academic center in Zurich, Switzerland, were retrospectively analyzed. Twenty-seven variables, available upon admission in clinical settings, were tested using six machine learning algorithms (support vector machines (SVM), logistic regression, naive Bayes, gradient boosting, fine trees, and neural networks). Predictive performance was assessed using metrics such as area under the curve (AUC) and balanced accuracy. SVM demonstrated the highest performance, achieving an AUC of 0.79 and a balanced accuracy of 69.8 %. These results suggest that the tool can identify patients at higher risk for problematic treatment courses, enabling earlier interventions and more efficient resource allocation. The simplicity of the model, based on routinely collected data, enhances its clinical applicability. However, validation studies in multi-center and international settings are essential to confirm its robustness and generalizability. This tool represents a promising step toward integrating machine learning into forensic psychiatry to improve treatment outcomes and patient safety.

1. Introduction

Prognostic tools play an important role in both forensic and general psychiatric inpatient settings. Such measures are helpful in daily routine in order to assess the probability of certain clinical outcomes. Regarding forensic psychiatric treatment, which is – depending on national legislation – often imposed by court decision, severe incidents in the treatment process may lead to harm of patients and staff as well as potentially avoidable prolongation of custodial measures. Thus, identification of risk factors for adverse events as well as accurate risk appraisal are needed to guide the therapeutic process. Over the last decades, a myriad of psychometric and prognostic tools have been developed and brought into clinical practice, and the importance of precise prediction is stressed throughout the literature (Fusar-Poli et al., 2018; Moller, 2009). However, existing albeit limited data suggest that psychiatrists often do not use existing assessment tools for a variety of reasons, including the perception that these scales are time-consuming, not helpful, not regularly required, or beyond the clinicians' (perceived) lack of training (Gilbody et al., 2002; Zimmerman and McGlinchey, 2008).

Violence and aggressive behavior against staff, patients or other individuals are common phenomena seen not only in forensic but also in general psychiatric treatment, derived from patient, staff, and organizational characteristics (Hvidhjelm et al., 2014), and often resulting in prolonged hospitalization (Broderick et al., 2015). In order to recognize and thereby prevent aggressive behavior, several scales have been developed. A tool widely used in daily clinical practice is the Brøset

* Corresponding author at: University Hospital of Psychiatry Zurich, Lenggstrasse 31, 8032 Zurich, Switzerland.
E-mail address: andreas.hofmann@pukzh.ch (A.B. Hofmann).

https://doi.org/10.1016/j.psychres.2025.116537
Received 7 January 2025; Received in revised form 30 April 2025; Accepted 8 May 2025
Available online 9 May 2025

Violence Checklist (BVC). It consists of six items and features moderate sensitivity and inter-rater reliability, combined with high specificity. Its goal is to assess the risk of aggression in the upcoming 24 hours (Woods and Almvik, 2002). In a prospective study, the BVC was able to correctly predict violence in a forensic inpatient setting from Denmark, achieving an adequate sensitivity and specificity using the common threshold of 3. In order to reach satisfactory results, the staff had to apply the checklist three times daily, (Hvidhjelm et al., 2014), thus much effort was needed to predict violence, requiring many resources in times of severe staff shortage (Wainberg et al., 2017).

Another scale commonly used to measure the risk of violence is the Modified Overt Aggression Scale (MOAS) (Kay et al., 1988). It uses a five-point scale to retrospectively assess four aggression dimensions on a weekly basis, namely verbal and physical aggression, property damage, and self-harm. Several studies used MOAS as an outcome parameter in patients with Schizophrenia spectrum disorders (SSD), corroborating previous findings that aggression is a frequent phenomenon seen in these patients (Li et al., 2020). Nevertheless, while the MOAS features appropriate statistical characteristics like good inter-rater reliability and sensitivity in detecting incidents of low severity, clinically important parameters like origin of the aggressive behavior are not taken into account (Harris et al., 2013).

Absconding can also affect the forensic patient's care, i.e., through prolongation of institutionalization, but also people being in contact with absconding inpatients, for example through victimization, although the risk appears to be rather low (Campagnolo et al., 2019; Wilkie et al., 2014). Existing data suggests a wide range of absconding rates, up to 54 %, depending on the setting (Olagunju et al., 2022). Recently, assessment tools for the risk of absconding in forensic patients, e.g., the BEAT, have been developed, but the underlying population includes a variety of diagnoses other than psychotic disorders (Booth et al., 2021). Moreover, external validation has yet to be performed (Neumann et al., 2023).

In summation, there are some tools assessing particular risks, but few comprehensive approaches encompassing several ones combined. The Short-Term Assessment of Risk and Treatability (START) was designed as a structured professional judgement tool developed on various diagnoses in forensic mental health care regarding partially similar adverse outcomes as the ones described below (Webster et al., 2006). While its predictive validity is strong, the influence following its clinical use on reduced incidence of violent behavior is low. Both risk and protective factors included are subject to clinicians` appraisal, associated with perceived difficulties in finer distinction between the items` scales. Further, a weekly application is often recommended, which is time-consuming in the tight schedule of clinical practice (O'Shea and Dickens, 2014).

Modern medicine now promotes individualized diagnostic and therapeutic measures, and machine learning (ML) – an umbrella term for multiple methods used in a variety of disciplines – plays an important role in facilitating precision medicine and psychiatry (Bzdok and Meyer-Lindenberg, 2018; Johnson et al., 2021). In forensic psychiatric research, especially that involving inpatients, clinical decision making still is predominantly based on data not derived from ML. Our previous work focused on the development of ML algorithms identifying characteristics associated with adverse outcomes in forensic inpatients diagnosed with SSD, as these are the most common primary diagnoses in forensic settings (de Girolamo et al., 2023). Some key findings were:

1) Prolonged forensic hospitalizations occur in cases in which the index offence is homicide, or the victims` injuries were severe (Kirchebner et al., 2020).
2) Coercive measures are more often applied to patients exhibiting threatening or violent behavior and distinct psychopathological peculiarities (Gunther et al., 2020).
3) Younger age as well as more symptoms of depression and anxiety are associated with self-harm during inpatient treatment (Kappes et al., 2023).
4) Amongst others, forbidden substance use during hospitalization, higher doses of antipsychotics, and disobedient behavior were associated with absconding (Kirchebner et al., 2021).
5) Substance use while hospitalized was rare, but younger age and early onset of schizophrenia symptoms are associated with this (Sonnweber et al., 2022).
6) Aggressive behavior is linked to variables such as breaking ward rules, PANSS score upon admission, and higher impulsivity (Machetanz et al., 2022b).

All ML models yielded satisfactory statistical properties, including sensitivity, specificity, and goodness of fit. In the presence of elevating number of patients and scarce financial and staff resources (Schanda et al., 2009; Wainberg et al., 2017), an easily applicable prognostic tool specifically for SSD capable of assessing the risk of future adverse events and potentially predicting treatment outcome already upon admission is needed. We assume that an ideal instrument would be easy to use and have robust statistical properties. As described above, previously developed ML tools have already been proven to be useful in screening for variables associated with complex treatment (Machetanz et al., 2022a).

The current work focuses on a revised version of the existing models, called *RIPTOSO (Rheinau Inventory for the Prediction of Treatment Outcome in Schizophrenic Offenders)*. A number of prior variables – 27 in total – were incorporated into this model. It was our hope that *RIPTOSO* would correctly predict the following adverse events at the outset of hospitalization of forensic inpatients diagnosed with SSD (for a comprehensive definition of all variables please refer to the appendix):

1) Use of compulsory measures (i.e., seclusion, restraint, forced medication)
2) Absconding (threatened or carried out)
3) Use of forbidden substances (suspected or proven)
4) Self-harm (with both suicidal and non-suicidal intention)
5) Aggressive behavior (verbally or physically) against others and objects
6) Prolonged hospitalization (i.e., >75th percentile regarding length of stay)

2. Methods

2.1. Study population

Our study used the files of 370 court-mandated forensic patients of male and female gender, all diagnosed with SSD (ICD-10 F2X, ICD-9 295.X (Slee, 1978; World Health Organization, 2004). All of them received treatment at the Center for Inpatient Forensic Therapy, a facility within the University Hospital of Psychiatry Zurich, located in the municipality of Rheinau, Switzerland. Treatment took place between 1982 and 2016, with the majority, i.e., 296 cases, after the year 2000. The patients were referred to this facility due to both violent (e.g., (attempted) homicide, assault, violent offenses against sexual integrity, robbery, and arson) and non-violent (e.g., threat and coercion, property crime, drug offenses) crimes committed in the context of their underlying psychiatric disorders. The legal basis for the referral is Article 59 of the Swiss Criminal Code (Fedlex, 2023).

2.2. Source of data

All data were retrospectively extracted from the patients` files using qualitative content analysis independently performed by two experienced psychiatrists according to a rating protocol based on a set of criteria described by Seifert and Nedopil (Assarroudi et al., 2018; Seifert, 1997). The case files are comprehensive and feature broad information about sociodemographic, clinical, and crime-related data. The whole data base has been used for various other analyses before as part of an overarching research project.

2.3. Predictor and outcome variables

In order to create a tool capable of predicting adverse events during the upcoming forensic inpatient hospitalization, only variables available upon admission were included. Thus, not all variables that had been previously found to be influential in the models were used for the analysis as predictor variables (Gunther et al., 2020; Kappes et al., 2023; Kirchebner et al., 2020, 2021; Machetanz et al., 2022b; Sonnweber et al., 2022). Altogether, 27 predictor variables were incorporated after having omitted variables depicting the current course of treatment. Regarding outcome variables, six dichotomous variables were defined in total: use of compulsory measures (i.e., seclusion, restraint, forced medication), absconding, use of forbidden substances, self-harming behavior in both suicidal and non-suicidal intention, any aggressive behavior against others and objects, and prolonged hospitalization. If at least one of these outcome variables was present, the respective case was considered as having had adverse outcome. For a comprehensive description of all predictor and outcome variables, please refer to the appendix.

2.4. Data analysis

We used supervised machine learning to test the ability of the variables to predict adverse events in the course of forensic care. In a first step, categorical variables were translated to binary code; ordinal and continuous variables remained unmodified. The appearance of any of the outcome variables was then dichotomized. Since all predictor variables were already identified in previous publications (for a comprehensive overview of each item's occurrence in all six source studies described in the introduction, please refer to the Supplementary Table S1), the dataset had not to be split in training and validation subsets. Thus, the whole set was used for the analysis. Missing values in the dataset were then imputed via 10 iterations of Multiple Imputation by Chained Equations (MICE) (Azur et al., 2011). The proportion of missing values of each predictor variable is mentioned in Supplementary Table S2.

Subsequently, several algorithms were tested in order to identify the one with the most robust predictive properties, including sensitivity (i.e., true positive rate), specificity (i.e., true negative rate), positive predictive value, negative predictive value, balanced accuracy (arithmetic mean of sensitivity and specificity), and the area under the receiver operating characteristic curve (AUC). The tested algorithms comprise support vector machines (SVM), naive Bayes, logistic regression, fine trees, gradient boosting, and neural networks. To assess the internal validity of the model and to mitigate the risk of overfitting, a 10-fold cross-validation was conducted. In this procedure, the data set was randomly partitioned into ten equal-sized folds; in each iteration, nine folds were used for training and one fold for validation. The reported performance metrics are averaged across the ten folds.

Last, all predictor variables were ranked according to their relative influence upon the model. All developmental steps are illustrated in Fig. 1.

Machine learning calculations were performed using MATLAB 2024b (MATLAB and Statistics Toolbox Release 2012, The MathWorks, Inc., Natick, Massachusetts, United States). Potential multicollinearity of all 27 predictor variables was tested with the variance inflation factor (VIF) based on the "car" package in R version 3.6.3. (R Project, Vienna, Austria). The coding protocol can be accessed (Kirchebner, 2025).

2.5. Ethics statement

This research project was reviewed and approved by the Ethics Committee Zurich (reference number KEK-ZH—NR 2014–0480, date of approval: 05/19/2015. As the study design was retrospective, patient consent was waived since it is not necessarily required in the Canton of Zurich, Switzerland.

3. Results

Table 1 depicts the sociodemographic characteristics of the final study sample. The inpatients were predominantly of male gender and in their early thirties upon admission. Most patients had a history of migration. The majority of the patients were not in an intimate relationship at the time of the offence that led to the forensic hospitalization. Schizophrenia was the most frequent diagnosis (nearly 80 %), followed by schizoaffective disorder and other psychotic disorders. Co-morbid substance abuse or dependence was common.

Of the 370 patients included, outcome data were available for 369 cases. One case was excluded from model evaluation due to missing outcome information. Of the remaining 369 cases, 239 individuals (64.8 %, i.e., outcome prevalence) experienced at least one adverse event,

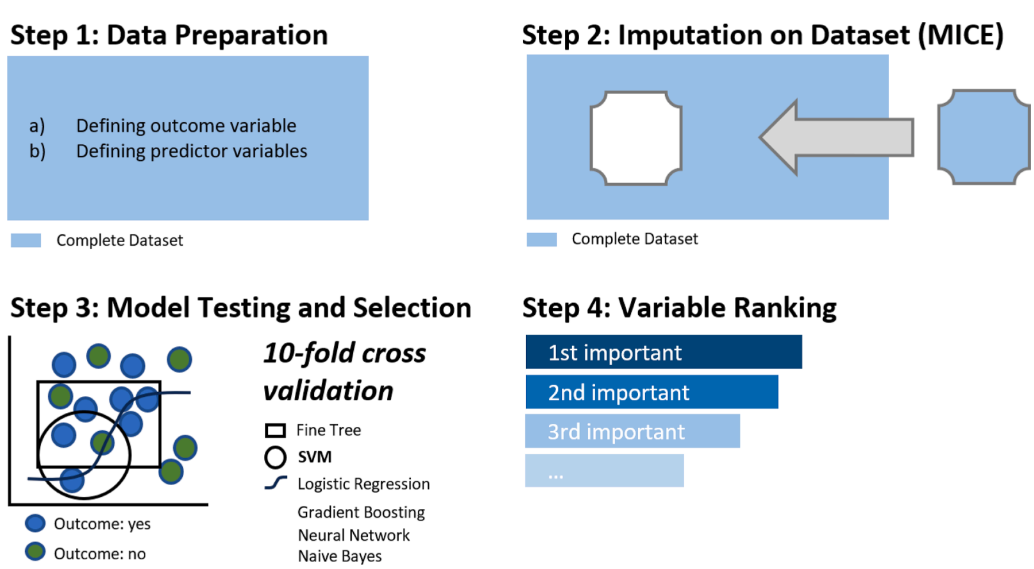

Fig. 1. Overview of statistical procedures.

Table 1
Sociodemographic characteristics.

Characteristics	Total
	n/N (%)
Male sex *	339/370 (91.6)
Age at admission (mean, SD)	34.15 (10.23)
Native Country Switzerland	167/370 (45.1)
Single (at offence)	297/364 (81.6)
Diagnosis: Schizophrenia	294/370 (79.5)
Diagnosis: Schizoaffective disorder	26/370 (7)
Diagnosis: Other psychotic disorder	50/370 (13.5)
Co-Diagnosis substance abuse/dependence	270/370 (73)

[*] according to patient files

Note. SD = standard deviation; N = total study population; n = subgroup with characteristic.

thereby fulfilling the criteria of the composite outcome. While the use of compulsory measures (i.e., seclusion, restraint, forced medication) was necessary in 131/358 (36.6 %) cases, further adverse events like absconding (34/274, 12.4 %), use of forbidden substances (51/364, 14.0 %), self-harming behavior in both suicidal and non-suicidal intention (87/356, 24.4 %), any aggressive behavior against others and objects (113/352, 32.1 %), and prolonged hospitalizations (42/358, 11.7 %) were less frequent (multiple events per observed case were possible).

Multiple algorithms were run on the whole dataset, as described in the methods section. Out of these six algorithms, support vector machines demonstrated the strongest performance parameters, as indicated by a receiver operating characteristic area under the curve (AUC) – a value depicting the level of discrimination – of 0.79 (95 % confidence intervals ranging from 0.73 to 0.85). The balanced accuracy, i.e., the arithmetic mean of sensitivity and specificity, reached 69.8 % (95 % confidence intervals 64.5 % - 74.8 %). Other algorithms performed similarly; logistic regression generated better balanced accuracy, while the AUC was not as strong as in SVM. For an overview of AUC and the balanced accuracy of all tested algorithms, as well as extensive performance parameters of SVM, please refer to Tables 2 and 3, respectively.

The variables differed in their importance, with some contributing more to the model than others. For RIPTOSO, the item “PA_A” (modified PANSS sum score upon admission (Kay et al., 1987)), was most important, while item “PA12” (PANSS item regarding difficulties in abstract thinking) contributed least. The full variable ranking and the variables’ meaning is displayed in Fig. 2. Multicollinearity was considered generally acceptable. Most predictors showed VIFs <3, indicating rather low collinearity. Three variables (SD1, D6, R9a) exhibited slightly elevated VIFs above 3. Missing values in the predictor variables ranged from 0 % to 31.7 %. For a comprehensive overview of all VIF scores and proportion of missing values, please refer to Supplementary Table S2.

4. Discussion

The present work focuses on the development of a tool capable of predicting adverse events during a forensic hospitalization, upon admission. For this purpose, we used 27 predictor variables, all of which were identified as associated with adverse outcomes in previous studies, but not within a combined model. The tool, called *RIPTOSO*, shows satisfactory statistical performance parameters, as indicated by an AUC of 0.79 and a balanced accuracy of 69.8 %.

Table 2
Machine learning models and balanced accuracy/AUC.

Statistical Procedure	Accuracy (%)	AUC
Logistic regression	73.4	0.76
Fine tree	65.9	0.65
Neural network	67.5	0.67
Gradient boosting	65.6	0.72
Naive Bayes	71	0.73

Note. AUC = area under the curve (level of discrimination).

Table 3
SVM model performance measures.

Performance measures	% (95 % CI)
Balanced accuracy	69.8 (64.5 - 74.8)
AUC	79 (73 - 85)
Sensitivity	76 (70.3 – 81)
Specificity	63.6 (53.6 – 72.5)
PPV	83.7 (78.2 – 88)
NPV	51.9 (43 – 60.7)

Note. AUC = area under the curve; CI = Confidence interval; PPV = Positive predictive value; NPV = Negative predictive value.

While several instruments reportedly predict inpatient outcomes, such as aggression and violence (Ogonah et al., 2023; Woods and Almvik, 2002), to our knowledge, none of these scoring systems were developed for multiple unfavorable events simultaneously specifically in SSD. Given that offenders with SSD account for a large proportion of patients in forensic psychiatric facilities, *RIPTOSO* focuses on this particular population, helping to identify individuals at high risk of a problematic treatment course. Thus, resources can be allocated to patients at higher risk in order to avert negative outcomes for themselves and staff.

In nearly all approaches, PANSS variables, either as a sum score or particular items, played an important role, and were as such incorporated into *RIPTOSO* (Gunther et al., 2020; Kappes et al., 2023; Kirchebner et al., 2020, 2021; Machetanz et al., 2022b). The modified three-tier sum score was shown to be most influential in our model when ranking the predictors. The high relative influence of variables related to psychopathology in our model goes to show that adverse behavior during the hospitalization could significantly be targeted through sufficient and needs-oriented interventions as well (Hodgins, 2001; Mullen, 2006). Other variables like age upon admission and first diagnosis, negative experiences during childhood and adolescence (e.g., childhood poverty (Galloway and Skardhamar, 2010; Page et al., 2014)), as well as type and number of index offences (i.e., those leading to the forensic referral) are static in nature, but still might be targeted as part of the therapy. Several items, e.g., PANSS or olanzapine-equivalent dose, could also serve as potential follow-up parameters, but this has yet to be proven in further analyses.

As already mentioned, our model showed an AUC of 0.79, which indicates a high level of discrimination. This performance quality is particularly noteworthy when looking at other well-established and frequently used prognostic tools in forensic psychiatry, namely instruments assessing reoffending such as the VRAG and HCR-20, although they do not necessarily measure inpatient aggressive behavior in particular and problematic courses of treatment in general. A recent meta-analysis found AUC values for violent recidivism, as indicated by HCR-20v2 (also applies for general recidivism) and VRAG, of 0.69 (95 % confidence interval 0.65 – 0.72) and 0.69 (0.63 – 0.75), respectively. The authors conclude that current tools still have limited evidence, partly because of weak methodology. Other parameters like specificity and sensitivity were rarely reported, and, if available, ranged from 55 %–85 % and 33 %–80 %, respectively (Ogonah et al., 2023). Another meta-analysis investigating the statistical properties of START detected an AUC of approximately 0.74 regarding any aggressive behavior and self-harm risk (O’Shea and Dickens, 2014). *RIPTOSO* performed at least slightly better regarding AUC and similarly regarding balanced accuracy, but we consider it helpful to boost its power by combining it with other tools in order to substantially influence comprehensive clinical decision making. However, it has to be stressed that the current version is based on a relatively small sample (n = 370); following the planned multi-center validation, statistical parameters might alter. In general, the predictive power of multiple assessment tools used in criminology has been shown to be only poor to moderate, and

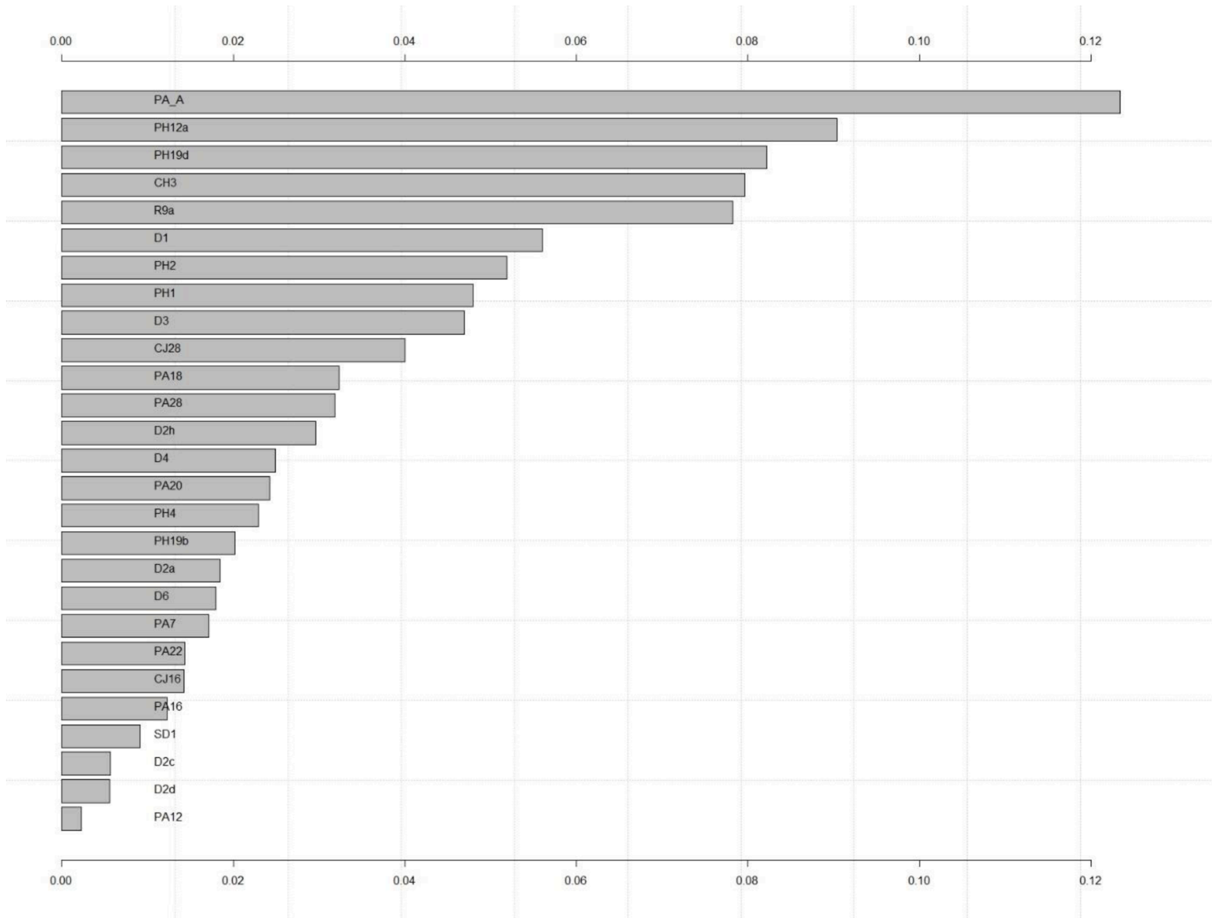

Fig. 2. Variable importance ranking and variable description.
PA_A: modified PANSS upon admission: sum score, **PH12a:** use of coercive measures during previous hospitalization, **PH19d:** time between discharge from the last inpatient treatment and index offense, **CH3:** age at the time of first entry in the Swiss criminal record, **R9a:** Olanzapine equivalent dose at the time of admission, **D1:** number of index offenses, **PH2:** age at the time of first presentation of SSD symptoms, **PH1:** age at the time of diagnosis (of SSD), **D3:** victim with serious/life-threatening injuries, **CJ28:** substance use disorder during childhood, **PA18:** modified PANSS upon admission: tension, **PA28:** modified PANSS upon admission: poor impulse control, **D2h:** violent property crime, **D4:** victim killed, **PA20:** modified PANSS upon admission: depression, **PH4:** history of hallucinations, **PH19b:** age at the time of first psychiatric inpatient treatment, **D2a:** (attempted) homicide, **D6:** age at the time of the index offense, **PA7:** modified PANSS upon admission: hostility, **PA22:** modified PANSS upon admission: uncooperativeness, **CJ16:** poverty in patient's family, **PA16:** modified PANSS upon admission: anxiety, **SD1:** age upon admission, **D2c:** threat, coercion, **D2d:** sexual abuse of children, **PA12:** modified PANSS upon admission: difficulty in abstract thinking.

underlying studies often lack important parameters other than AUC like false positive and negative rate, impairing comparability (Fazel et al., 2022).

Before introducing any assessment algorithm into all-day clinical practice, ethical implications must be considered. We refrain from notions that treatment decisions should be merely based on computations since this approach would so far not be covered by current research (Fazel et al., 2012). A scoring system should always be interpreted along multiple other factors, especially in forensic psychiatry, due to the inherent legally mandated restrictions (MacIntyre et al., 2023; Starke et al., 2021). In other words, a patient *RIPTOSO* might identify as at high risk of having a problematic course of treatment, must not receive more restrictions only because of a calculated score. So far, *RIPTOSO* is especially useful to raise awareness for patients at high risk and thus allow the allocation of specific supportive resources in order to avert future problems. With ML becoming more and more part of all medical disciplines, we need further ethical discussions and frameworks on what to use for clinical work (Wiese and Friston, 2022). Current research also stresses the importance of using unbiased data for any artificial intelligence tool (MacIntyre et al., 2023). With items mainly based on robust data, like age upon certain events or clinical scale like PANSS, we consider this prerequisite as given.

4.1. Strengths and limitations

With *RIPTOSO,* we were able to employ variables that are usually readily available upon forensic psychiatric admission. This would allow for an easily applicable assessment instrument for the risk of upcoming problematic courses of treatment on the basis of a few accessible variables, enabling early countermeasures. In times of staff shortages, covering risk assessment for several adverse treatment events in one tool is efficient and economic. Compared to other risk assessment tools, *RIPTOSO* performs equally well. *RIPTOSO* is more or less on the very threshold between fair/acceptable ($0.7 \leq$ AUC < 0.8) and good/excellent ($0.8 \leq$ AUC < 0.9) (Mandrekar, 2010; Nahm, 2022). We therefore consider the present preliminary tool as psychometrically acceptable. Also, it is designed for use in SSD offenders and features satisfying explanatory power after being used only once right after admission.

There are some limitations that have to be taken into account. First, our study population comprised patients treated over a long span (back to 1982). Even though the majority of cases stemmed from treatment periods after 2000, clinical practices and procedures have changed over time, thus the results could therefore be biased. The long observation period is vital though to collect enough data for ML analyses – and while 370 patient files seem to be sufficient for many calculations, ML depends on large numbers. In this case, performance parameters like AUC and balanced accuracy might even be higher if there had been more data (Kokol et al., 2022; Koppe et al., 2021). In contrast, it has been shown that smaller studies with less than 500 individuals report rather higher AUC (Fazel et al., 2022). Subsequent replication projects, performed independently if possible, could address this issue. Furthermore, multicollinearity among predictor variables was assessed using the variance inflation factor (VIF). Most predictors exhibited VIF scores below 3, indicating acceptable multicollinearity following common interpretation guidelines (VIF < 5 considered acceptable) (James et al., 2013; Kim, 2019). Only three variables (SD1, D6, and R9a) showed slightly elevated VIF values above 3, suggesting minor interdependencies. Although multicollinearity was not a major issue in our model, this should be considered when interpreting the variable influences. Another potential limitation relates to missing data in some predictor variables. While most variables exhibited low rates of missingness, a few had higher proportions, up to approximately 30 %. However, to minimize potential bias and retain the full sample size, we applied MICE, a well-established and effective imputation method for handling incomplete data. Due to the retrospective nature of data collection, some predictor variables were occasionally unavailable at the time of admission. Nevertheless, this does not interfere with the general goal of the model, as all 27 variables—based on a clinical perspective—can be readily collected within the first days of forensic hospitalization. Prospective validation studies are needed to further address this limitation.

So far, our results are limited to forensic inpatients diagnosed with SSD, and, as can be expected in an offender population, the investigated sample is mainly male. Moreover, the transferability of *RIPTOSO* to other legal systems – which play a vital role in forensic therapy – has yet to be proven in upcoming international analyses, which are also important to address another limitation: currently, *RIPTOSO*'*s* perspective is directed to the *upcoming* hospitalization, an unspecified period, fluctuating from patient to patient. Future projects could include follow-up assessments subsequent to the time of admission, for example at a yearly frequency. This would also allow for measuring the performance depending on the duration of stay, since any tool might be most informative and meaningful at certain moments. Other limitations include the use of modified three-tier PANSS item (as described above), a common but not validated method. The current version of *RIPTOSO* aims at identifying inpatients at high risk of *any* adverse event, but currently does not allow a specification of the nature of the adverse event in each individual case. Again, this is due to a paucity of data, but also facilitates an efficient assessment despite the potential loss of information. Last, all variables identified in previous calculations derived from the same dataset that was now used for the development of *RIPTOSO*. As a consequence, overfitting – a common problem in ML (Dietterich, 1995) – might have indirectly influenced the present work, since all models that led to the development of *RIPTOSO* are based on this dataset. This has to be considered before implementing the tool into clinical practice. Nevertheless, this bias could easily be relieved with the upcoming multi-center, international dataset validations.

5. Conclusion and outlook

In forensic psychiatry, a number of prognostic tools exists for predicting criminal recidivism, but few tools for predicting difficult treatment courses. *RIPTOSO*, a novel tool, is particularly designed for forensic inpatients diagnosed with any schizophrenia spectrum disorder. If patients at high risk for expressing difficult behavior can be identified when admitted to a forensic-psychiatric care facility, a corresponding reaction and intervention tailored to the individual's risk- and need-based requirements can be established. In other words, *RIPTOSO* allows for more effective resource allocation. Additionally, we expect a lower prevalence of critical incidents during treatment, thus a shorter duration of treatment and confinement. This would not only result in a reduction of overall costs, but more importantly could directly benefit patients and forensic health care staff. It is particularly worth mentioning that *RIPTOSO* can be completed quickly and easily with 27 data points as soon as the patient arrives. In the near future, both international and Swiss data derived from multiple treatment sites across legal systems shall be used to further investigate its validity retrospectively. Prospective studies are needed. The tool could also be adapted to diagnoses other than SSD and to general psychiatric patients, given sufficient data.

CRediT authorship contribution statement

Andreas B. Hofmann: Writing – review & editing, Writing – original draft, Visualization, Methodology, Data curation, Software, Validation. **Marc Dörner:** Writing – review & editing, Writing – original draft, Methodology. **Philip E. Klassen:** Writing – review & editing, Writing – original draft, Conceptualization, Validation. **Lena Machetanz:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Formal analysis, Data curation, Conceptualization, Investigation, Project administration. **Johannes Kirchebner:** Writing – review & editing, Writing – original draft, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement and funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.psychres.2025.116537.

References

Assarroudi, A., Heshmati Nabavi, F., Armat, M.R., Ebadi, A., Vaismoradi, M., 2018. Directed qualitative content analysis: the description and elaboration of its underpinning methods and data analysis process. J. Res. Nurs. 23 (1), 42–55.

Azur, M.J., Stuart, E.A., Frangakis, C., Leaf, P.J., 2011. Multiple imputation by chained equations: what is it and how does it work? Int. J. Methods Psychiatr. Res. 20 (1), 40–49.

Booth, B.D., Michel, S.F., Baglole, J.S., Healey, L.V., Robertson, H.V., 2021. Validation of the booth evaluation of absconding tool for assessment of absconding risk. J. Am. Acad. Psychiatry Law 49 (3), 338–349.

Broderick, C., Azizian, A., Kornbluh, R., Warburton, K., 2015. Prevalence of physical violence in a forensic psychiatric hospital system during 2011-2013: patient assaults, staff assaults, and repeatedly violent patients. CNS Spectr. 20 (3), 319–330.

Bzdok, D., Meyer-Lindenberg, A., 2018. Machine learning for precision psychiatry: opportunities and challenges. Biol. Psychiatry Cogn. Neurosci. NeuroimAging 3 (3), 223–230.

Campagnolo, D., Furimsky, I., Chaimowitz, G., 2019. Absconsion from forensic psychiatric institutions: a review of the literature. Int. J. Risk Recovery 2.

de Girolamo, G., Iozzino, L., Ferrari, C., Gosek, P., Heitzman, J., Salize, H.J., Wancata, J., Picchioni, M., Macis, A., 2023. A multinational case-control study comparing forensic and non-forensic patients with schizophrenia spectrum disorders: the EU-VIORMED project. Psychol. Med. 53 (5), 1814–1824.

Dietterich, T., 1995. Overfitting and undercomputing in machine learning. ACM Comput. Surv. (CSUR) 27 (3), 326–327.

Fazel, S., Burghart, M., Fanshawe, T., Gil, S.D., Monahan, J., Yu, R., 2022. The predictive performance of criminal risk assessment tools used at sentencing: systematic review of validation studies. J. Crim. Justice 81, 101902.

Fazel, S., Singh, J.P., Doll, H., Grann, M., 2012. Use of risk assessment instruments to predict violence and antisocial behaviour in 73 samples involving 24 827 people: systematic review and meta-analysis. BMJ 345, e4692.

Fedlex, 2023. Swiss Criminal Code.

Fusar-Poli, P., Hijazi, Z., Stahl, D., Steyerberg, E.W., 2018. The science of prognosis in psychiatry: a review. JAMA Psychiatry 75 (12), 1289–1297.

Galloway, T.A., Skardhamar, T., 2010. Does parental income matter for onset of offending? Eur. J. Criminol. 7 (6), 424–441.
Gilbody, S.M., House, A.O., Sheldon, T.A., 2002. Psychiatrists in the UK do not use outcomes measures. National survey. Br. J. Psychiatry 180, 101–103.
Gunther, M.P., Kirchebner, J., Lau, S., 2020. Identifying direct coercion in a high risk subgroup of offender patients with schizophrenia via machine learning algorithms. Front. Psychiatry 11, 415.
Harris, S.T., Oakley, C., Picchioni, M., 2013. Quantifying violence in mental health research. Aggress. Violent Behav. 18 (6), 695–701.
Hodgins, S., 2001. The major mental disorders and crime: stop debating and start treating and preventing. Int. J. Law Psychiatry 24 (4–5), 427–446.
Hvidhjelm, J., Sestoft, D., Skovgaard, L.T., Bue Bjorner, J., 2014. Sensitivity and specificity of the broset violence checklist as predictor of violence in forensic psychiatry. Nord. J. Psychiatry 68 (8), 536–542.
James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. Linear Regression, An Introduction to Statistical Learning: with Applications in R. Springer, New York, New York, NY, pp. 59–126.
Johnson, K.B., Wei, W.Q., Weeraratne, D., Frisse, M.E., Misulis, K., Rhee, K., Zhao, J., Snowdon, J.L., 2021. Precision medicine, AI, and the future of personalized health care. Clin. Transl. Sci. 14 (1), 86–93.
Kappes, J.R., Huber, D.A., Kirchebner, J., Sonnweber, M., Gunther, M.P., Lau, S., 2023. Self-harm among forensic psychiatric inpatients with schizophrenia spectrum disorders: an explorative analysis. Int. J. Offender. Ther. Comp. Criminol. 67 (4), 352–372.
Kay, S.R., Fiszbein, A., Opler, L.A., 1987. The positive and negative syndrome scale (PANSS) for schizophrenia. Schizophr. Bull. 13 (2), 261–276.
Kay, S.R., Wolkenfeld, F., Murrill, L.M., 1988. Profiles of aggression among psychiatric patients. I. nature and prevalence. J. Nerv. Ment. Dis. 176 (9), 539–546.
Kim, J.H., 2019. Multicollinearity and misleading statistical results. Korean J. Anesth. 72 (6), 558–569.
Kirchebner, J., 2025. RIPTOSO - the development of a screening tool for adverse events during forensic-psychiatric inpatient treatments of offenders with schizophrenia spectrum disorders. Zenodo. https://doi.org/10.5281/zenodo.15297352.
Kirchebner, J., Gunther, M.P., Sonnweber, M., King, A., Lau, S., 2020. Factors and predictors of length of stay in offenders diagnosed with schizophrenia - a machine-learning-based approach. BMC. Psychiatry 20 (1), 201.
Kirchebner, J., Lau, S., Sonnweber, M., 2021. Escape and absconding among offenders with schizophrenia spectrum disorder - an explorative analysis of characteristics. BMC Psychiatry 21 (1), 122.
Kokol, P., Kokol, M., Zagoranski, S., 2022. Machine learning on small size samples: a synthetic knowledge synthesis. Sci. Prog. 105 (1), 368504211029777.
Koppe, G., Meyer-Lindenberg, A., Durstewitz, D., 2021. Deep learning for small and big data in psychiatry. Neuropsychopharmacology 46 (1), 176–190.
Li, W., Yang, Y., Hong, L., An, F.R., Ungvari, G.S., Ng, C.H., Xiang, Y.T., 2020. Prevalence of aggression in patients with schizophrenia: a systematic review and meta-analysis of observational studies. Asian J. Psychiatr. 47, 101846.
Machetanz, L., Huber, D., Lau, S., Kirchebner, J., 2022a. Model building in forensic psychiatry: a machine learning approach to screening offender patients with SSD. Diagn. (Basel) 12 (10).
Machetanz, L., Lau, S., Kirchebner, J., 2022b. Advantages of machine learning in forensic psychiatric research—uncovering the complexities of aggressive behavior in schizophrenia. Appl. Sci. 12, 819.
MacIntyre, M.R., Cockerill, R.G., Mirza, O.F., Appel, J.M., 2023. Ethical considerations for the use of artificial intelligence in medical decision-making capacity assessments. Psychiatry Res. 328, 115466.
Mandrekar, J.N., 2010. Receiver operating characteristic curve in diagnostic test assessment. J. Thorac. Oncol. 5 (9), 1315–1316.
Moller, H.J., 2009. Standardised rating scales in psychiatry: methodological basis, their possibilities and limitations and descriptions of important rating scales. World J. Biol. Psychiatry 10 (1), 6–26.
Mullen, P.E., 2006. Schizophrenia and violence: from correlations to preventive strategies. Adv. Psychiatr. Treat. 12 (4), 239–248.
Nahm, F.S., 2022. Receiver operating characteristic curve: overview and practical use for clinicians. Korean J. Anesth. 75 (1), 25–36.
Neumann, M., Wittland, I., Schmitz, S.L., 2023. Ein Überblick über ausgewählte standardisierte Instrumente zur Erfassung des Risikos von Entweichungen im Kontext von Lockerungsprognosen in forensisch-psychiatrischen Einrichtungen. Forens. Psychiatr. Psychol. Kriminol. 17 (2), 239–250.
O'Shea, L.E., Dickens, G.L., 2014. Short-Term assessment of risk and treatability (START): systematic review and meta-analysis. Psychol. Assess. 26 (3), 990–1002.
Ogonah, M.G.T., Seyedsalehi, A., Whiting, D., Fazel, S., 2023. Violence risk assessment instruments in forensic psychiatric populations: a systematic review and meta-analysis. Lancet Psychiatry 10 (10), 780–789.
Olagunju, A.T., Bouskill, S.L., Olagunju, T.O., Prat, S.S., Mamak, M., Chaimowitz, G.A., 2022. Absconsion in forensic psychiatric services: a systematic review of literature. CNS Spectr. 27 (1), 46–57.
Page, A., Lewis, G., Kidger, J., Heron, J., Chittleborough, C., Evans, J., Gunnell, D., 2014. Parental socio-economic position during childhood as a determinant of self-harm in adolescence. Soc. Psychiatry Psychiatr. Epidemiol. 49 (2), 193–203.
Schanda, H., Stompe, T., Ortwein-Swoboda, G., 2009. Dangerous or merely difficult? the new population of forensic mental hospitals. Eur. Psychiatry 24 (6), 365–372.
Seifert, D., 1997. Die entwicklung des psychiatrischen Massregelvollzugs (§ 63StGB) in Nordrhein-Wesfalen. Psychiat Prax 24, 237–244.
Slee, V.N., 1978. The International Classification of Diseases: Ninth Revision (ICD-9). American College of Physicians, pp. 424–426.
Sonnweber, M., Kirchebner, J., Gunther, M.P., Kappes, J.R., Lau, S., 2022. Exploring substance use as rule-violating behaviour during inpatient treatment of offender patients with schizophrenia. Crim. Behav. Ment. Health 32 (4), 255–266.
Starke, G., De Clercq, E., Borgwardt, S., Elger, B.S., 2021. Computing schizophrenia: ethical challenges for machine learning in psychiatry. Psychol. Med. 51 (15), 2515–2521.
Wainberg, M.L., Scorza, P., Shultz, J.M., Helpman, L., Mootz, J.J., Johnson, K.A., Neria, Y., Bradford, J.E., Oquendo, M.A., Arbuckle, M.R., 2017. Challenges and opportunities in global mental health: a research-to-practice perspective. Curr. Psychiatry Rep. 19 (5), 28.
Webster, C.D., Nicholls, T.L., Martin, M.L., Desmarais, S.L., Brink, J., 2006. Short-term assessment of risk and treatability (START): the case for a new structured professional judgment scheme. Behav. Sci. Law 24 (6), 747–766.
Wiese, W., Friston, K.J., 2022. AI ethics in computational psychiatry: from the neuroscience of consciousness to the ethics of consciousness. Behav. Brain Res. 420, 113704.
Wilkie, T., Penney, S.R., Fernane, S., Simpson, A.I., 2014. Characteristics and motivations of absconders from forensic mental health services: a case-control study. BMC. Psychiatry 14, 91.
Woods, P., Almvik, R., 2002. The Broset violence checklist (BVC). Acta Psychiatr. Scand. Suppl. (412), 103–105.
World Health Organization, 2004. ICD-10: International Statistical Classification of Diseases and Related Health Problems: Tenth Revision. World Health Organization.
Zimmerman, M., McGlinchey, J.B., 2008. Why don't psychiatrists use scales to measure outcome when treating depressed patients? J. Clin. Psychiatry 69 (12), 1916–1919.