Predicting mechanical restraint of psychiatric inpatients by applying machine learning on electronic health data

Danielsen, A. A.; Fenger, M. H.J.; Østergaard, S. D.; Nielbo, K. L.; Mors, O.

Published in:
Acta Psychiatrica Scandinavica

DOI:
10.1111/acps.13061

Publication date:
2019

Document version
Accepted manuscript

Citation for published version (APA):

Terms of use
This work is brought to you by the University of Southern Denmark through the SDU Research Portal. Unless otherwise specified it has been shared according to the terms for self-archiving. If no other license is stated, these terms apply:

- You may download this work for personal use only.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying this open access version

If you believe that this document breaches copyright please contact us providing details and we will investigate your claim. Please direct all enquiries to puresupport@bib.sdu.dk
MR ANDREAS AALKJÆR DANIELSEN (Orcid ID : 0000-0002-6585-3616)
DR SØREN DINESEN ØSTERGAARD (Orcid ID : 0000-0002-8032-6208)

Article type: Original Article

Abstract word count: 200
Main text word count: 5,064
Total word count (abstract, text, references): 6,542
Number of tables: 3
Number of figures: 2

Predicting Mechanical Restraint of Psychiatric Inpatients by Applying Machine Learning on Electronic Health Data

Running Title: Predicting Mechanical Restraint

Andreas A. Danielsen MD¹,²,³, Morten H.J. Fenger MSc PhD⁴, Søren D. Østergaard MD PhD²,³,⁵,⁶, Kristoffer L. Nielbo MA PhD⁷, Ole Mors MD PhD¹,²,³

¹Psychosis Research Unit, Department for Psychosis, Aarhus University Hospital - Psychiatry, Denmark. ²The Lundbeck Foundation Initiative for Integrative Psychiatric Research, iPSYCH, Denmark. ³Department of Clinical Medicine, Aarhus University, Denmark. ⁴Department of Economics and Business Economics, Aarhus University, Denmark. ⁵Department of Affective Disorders, Aarhus University Hospital – Psychiatry, Denmark, ⁶Aarhus Institute of Advanced Studies, Aarhus University, Denmark, ⁷Department of History, University of Southern Denmark.
Corresponding author
Andreas A. Danielsen, MD
Psychosis Research Unit
Aarhus University Hospital - Psychiatry
Palle Juul-Jensens Boulevard 175
8200 Aarhus N
Denmark
E-mail: andrdani@rm.dk
Phone: +45 30268627

Acknowledgements

AAD is supported by grants from Aarhus University, the Lundbeck Foundation Initiative for Integrative Psychiatric Research (iPSYCH) and Ledelsesforum for Universitet-Regionssamarbejdet Region Midtjylland (LUR)

SDØ is supported by a grant from Independent Research Fund Denmark (7016-00048).

This article is protected by copyright. All rights reserved
Declaration of interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available from Central Denmark Region (Permission is granted by the Danish Patient Safety Authority) and the Danish Health Data Authority.

Abstract
Objective

Mechanical restraint (MR) is used to prevent patients from harming themselves or others during inpatient treatment. The objective of this study was to investigate whether incident MR occurring in the first three days following admission could be predicted based on analysis of electronic health data available after the first hour of admission.

Methods

The dataset consisted of clinical notes from Electronic Health Records from the Central Denmark Region and data from the Danish Health Registers from patients admitted to a psychiatric department in the period from 2011 to 2015. Supervised machine learning algorithms were trained on a randomly selected subset of the data and validated using an independent test dataset.

Results

A total of 5,050 patients with 8,869 admissions were included in the study. One-hundred patients were mechanically restrained in the period between one hour and three days after the admission. A Random Forest algorithm predicted MR with an area under the curve of 0.87 (95% CI 0.79-0.93). At 94% specificity, the sensitivity was 56%. Among the ten strongest predictors, nine were derived from the clinical notes.

Conclusions

These findings open for the development of an early warning system that may guide interventions to reduce the use of MR.

Key Words

Mental Disorders; Coercion; Electronic Medical Records; Supervised Machine Learning; Natural Language Processing

Significant outcome
• Using machine learning techniques on electronic health data we identified patients at high risk of being subjected to mechanical restraint.
• A Random Forest algorithm predicted mechanical restraint with an area under the curve of 0.87 (95% CI 0.79-0.93).
• Among the ten strongest predictors, nine were derived from the clinical notes.

Limitations
• Since we only analyzed data from the electronic health record system of the Central Denmark Region, the prediction model may not be applicable to data from other sources.
• The risk of recurrence was not considered in this study as it was restricted to incident episodes of mechanical restraint.

Introduction
Mechanical restraint (MR) - i.e. restraining a patient to a bed using belts or straps - is a coercive measure used in many countries to avoid that patients suffering from mental disorders harm themselves or others during inpatient treatment at psychiatric hospitals\textsuperscript{1-3}. MR has been associated with many adverse outcomes for the patients subjected to this intervention, such as
significant psychological distress, shoulder injuries, venous thromboembolism and even death\textsuperscript{4}. For these reasons, the ethics of MR regarding patient autonomy and nonmaleficence is often debated\textsuperscript{5}. Furthermore, the adverse outcomes associated with MR are not restricted to the patients being subjected to MR as the hospital staff involved in the procedure are also at risk of developing psychological- as well as physical injuries\textsuperscript{4}. For these reasons, reducing the use of MR and other coercive measures as much as possible is a priority for mental health services across the globe\textsuperscript{6,7}.

Reducing the use of MR in clinical practice relies on being able to identify patients at risk of being subjected to this procedure. Prior studies have identified several risk factors for MR including diagnosis (e.g. schizophrenia or organic mental disorder), male sex, living alone and involuntary admission\textsuperscript{8,9}. Prior studies have also suggested that early identification of patients at high risk for being exposed to coercive measures in psychiatric hospitals may reduce the use of these measures\textsuperscript{10,11}. However, only little is known on how identified risk factors interact - and prediction and prevention of MR therefore remains a challenge to clinical practice.

While a few prior studies have aimed to investigate how risk factors for MR interact, they have all had important limitations: i) No studies have focused specifically on patients at high risk for incident MR (MR for the first time ever). This is a very important focus since timely interventions could potentially prevent this group of patients to be subjected to MR at all. ii) Some studies have created models aimed at predicting coercive measures using historical data, but have not validated their findings in an independent test set\textsuperscript{12-15}, which is recommended to assess the extent of potential overfitting\textsuperscript{16,17}. iii) One study specifically aimed at predicting MR using tools developed to assess the risk of violent incidents, but only had a relatively small sample available\textsuperscript{18}. In order to overcome these limitations, studies aimed at predicting MR should i) focus on incident MR episodes, ii) validate the prediction algorithm in an independent test set, and iii) be based on large datasets. A methodological approach, which can handle large datasets and allows for consideration of interactions between predictors and subsequent validation in independent test sets, is machine learning\textsuperscript{19}.  

This article is protected by copyright. All rights reserved
Machine learning is a set of heterogeneous algorithms that can identify patterns in large datasets and has been shown to be a powerful tool for developing prediction models\(^1\). In psychiatry, a number of studies have used machine learning to create prediction models for several important patient outcomes such as suicide, domestic abuse and treatment response\(^{20-22}\). In these studies, various information was extracted from databases and then used to create potential predictors. Based on these potential predictors, a machine learning algorithm was trained to create a prediction model that could identify patients at high risk of experiencing the outcome of interest. Due to the promising results of prior studies using machine learning to predict outcomes of great relevance to clinical psychiatry\(^{20-22}\), we sought to use machine learning to predict MR.

**Aims of the Study**
The aim of this study was to develop and validate a prediction model that, based on electronic data available after the first hour of admission, could predict incident MR within the following three days. If such a model proves sufficiently accurate, it may open for the implementation of an automatic early warning system that can guide interventions to reduce the use of MR in clinical practice – to the benefit of both patients and staff at psychiatric hospitals.

**Method**

**Setting**
The setting of this study is the Central Denmark Region (CDR) – a Danish region with approximately 1.3 million inhabitants. The Danish regions (five in total) are administrative units with health care as their main responsibility. CDR was chosen for this study because the electronic
health record system “MidtEPJ” has been used consistently since 2012 across the seven psychiatric departments covering this region.

**Data sources**

Data was extracted from three databases: i) The MidtEPJ electronic health record system ii) The Registry of Coercive Measures in Psychiatric Treatment\(^\text{23}\), and iii) The Danish Psychiatric Central Research Register\(^\text{24}\). These three data sources and the extracted information used for this study are described in further detail below. Permission to use and store these data for research purposes was granted by the Danish Data Protection Agency (File no. 1-16-02-527-15) and the Danish Patient Safety Authority (File no. 3-3013-1244/1/). Ethical review board approval is not required for this study according to Danish legislation.

**MidtEPJ:** MidtEPJ is an electronic health record system, which contains individual-level information collected as part of standard clinical practice in relation to hospital contacts in CDR. One of the main types of data regarding psychiatric inpatients in MidtEPJ is the clinical notes, which consist predominantly of natural language, but also contain binary data (e.g. whether a specific symptom is present or not (yes/no)) and numerical data such as the sum score on the Brøset Violence Checklist (BVC)\(^\text{25}\) – an instrument measuring acute risk of violence. For each clinical note, MidtEPJ contains metadata such as a timestamp, a unique personal identification number and information regarding the hospital and department where the note was written. The notes are written by health care professionals, e.g. doctors, nurses, psychologists etc. as part of standard clinical practice at the psychiatric departments. All notes have a predefined heading describing their thematic content (‘current social functioning’, ‘previous social functioning’, ‘suicide risk assessment’). When completing a note, adherence to the predefined heading/theme is expected. For this study, all clinical notes from the psychiatric departments from the period between the implementation of MidtEPJ (2011/2012) and 2015 (inclusive) were extracted.

**The Registry of Coercive Measures in Psychiatry:** This register holds information regarding all coercive measures employed in relation to inpatient treatment at psychiatric hospitals in Denmark since 1999\(^\text{23}\). Coercive measures include involuntary admissions, involuntary treatment
(predominantly pharmacological) and MR. According to Danish law, the use of coercive measures at psychiatric hospitals must immediately be recorded in the clinical notes. Subsequently, information regarding the type of coercion, the reason for coercion and the precise timing of initiation and termination of the coercive measure is forwarded to the Registry of Coercive Measures in Psychiatry. For this study, information on all coercive measures occurring from 1999 and onwards was extracted for all psychiatric inpatients registered with at least one clinical note in MidtEPJ.

The Danish Psychiatric Central Research Register: This register contains information concerning all psychiatric admissions (e.g. diagnoses, hospital, department, date/hour/minute of admission and date of discharge) since April 1st, 1969\(^4\). In 1995 this registry was expanded to also include information regarding outpatient psychiatric treatment and contacts to psychiatric emergency rooms. For this study, administrative data regarding all psychiatric treatments/contacts (from 1995 and onwards) was extracted for all psychiatric inpatients registered with at least one clinical note in MidtEPJ.

Data linkage
Information from the three data sources described above were linked at the level of the individual by means of the unique 10-digit personal identification number, which is assigned to all inhabitants in Denmark at the time of birth or when obtaining legal residency\(^2\).

Study cohort
Patients were included in the study cohort if they had at least one psychiatric admission that met the following requirements:

i) The admission was recorded in MidtEPJ

ii) All prior contacts to psychiatric departments in the CDR were recorded in MidtEPJ

iii) No prior contact to child- and adolescent psychiatric services in CDR

iv) There were no prior MR episodes recorded in the Registry of Coercive Measures in Psychiatry.

v) At the time of admission, the patient was \(\geq 18\) years old.
For each included patient, all psychiatric admissions in the CDR were identified. Admissions were excluded if one of the following conditions were present:

i) The patient had been subjected to MR in relation to a prior admission (e.g. if a patient had three admissions and was subjected to MR during the second admission, the third admission was not included in the dataset).

ii) The patient was mechanically restrained within one hour of the admission.

iii) The patient was admitted to a somatic hospital and then simultaneously involuntarily admitted to a psychiatric hospital (a so-called double admission).

Double admissions were identified by the first author by reading clinical notes from involuntary admissions. Double admission takes place when patients are in need of treatment for a life-threatening somatic condition, but refuses to receive treatment due to the mental disorder (most often due to delusions).

The selection of patients and admissions eligible for the study is illustrated in Figure 1.

Figure 1 approximately here

Case definition
An admission was considered a “case-admission” if the patient was mechanically restrained in the 71-hour period between one hour after admission and three days after admission. All other included admissions were used as controls-admissions. Thus, one patient could have several admissions that served as control-admissions but only one admission that served as a case-admission. This approach has been used before and enables modeling of changes in risk over time (new predicted risk for each included admission) within patients.

Data preparation and predictor construction
The included patients were randomly assigned to a training (70%) and a test set (30%) stratified on MR episodes, such that the training and test datasets contained an equal proportion of MR
episodes. In the training set, only the last admission for each patient was included. This was done to avoid that several consecutive admissions of a patient would be attributed to much weight in the analysis. This ensured that all case-admissions in the training set were included for model development. SAS software, version 9.4 (SAS Institute) was used for data management.

For each admission used in the training set, data from the three databases were collected including clinical information up to the first hour of the admission. Predictors known or hypothesized to be associated with risk of MR were extracted from the registers, such as age, sex, diagnosed mental disorders, number of prior involuntary psychiatric admissions and outpatient treatment. An explorative approach was adopted for inclusion/exclusion of predictors in order to preserve important interactions. From MidtEPJ clinical notes describing five different themes were chosen. Three themes were unstructured and written in natural language: ‘Subjective Mental State’, ‘Current Objective Mental State’, and ‘Current Social Functioning’. Two themes were structured and numerical: ‘Brøset Violence Checklist’ (range: 0-6. 0 = little risk of violence, 1-2 = moderate risk of violence, and >2 = high risk of violence[25]) and ‘Current Risk of Suicide’ (range: 1-3. 1 = no risk of suicide, 2 = moderate risk of suicide, and 3 = high risk of suicide). For the numerical notes only the most recent note available for each theme was used to construct categorical predictors. Since all themes describe a current status, only notes written within the last two days before admission were included – except for ‘Current Social Functioning’ where notes written up to one month prior to admission were included. Notes in natural language were concatenated into one string within each theme, effectively constructing three note-types in natural language for each admission.

In order to identify a small set of semantic topics in each note-type, a three-step text mining pipeline was created, consisting of i) text preprocessing, ii) word feature extraction, and iii) topic feature extraction. The first author manually labeled semantic topics. This was done based on i) a clinically oriented assessment of the individual terms included in the identified semantic topics, and ii) reading of the individual clinical notes that contained the topic being labeled, looking for a common theme. SAS Enterprise Miner
14.1 including the Text Miner Node add-on with a Danish parser was used to implement the pipeline. See the Supplementary Material for a full description of the pipeline.

**Model Development**

Based on the constructed predictors, five different supervised machine learning algorithms were trained on the training set: I) Neural Network, II) Support Vector Machine, III) Random Forest, IV) Stepwise Forward Logistic Regression, and V) Least Absolute Shrinkage and Selection Operator (LASSO). The Random Forest model was trained in a two-step procedure: First, a Random Forest was trained with predictor selection; second, only the selected predictors were then used to train a new final Random Forest model. See the Supplementary Material for a full description of the model development. Confidence intervals for the prediction accuracy in the test set were calculated post hoc in Python (custom script, 1000 bootstrapped samples).

**Assessment of model prediction accuracy**

Based on the area under the receiver operating characteristic curve (ROC AUC), the best performing model (of the five described above) was chosen for validation using the test set. All results pertaining predictive performance are based on the test set and include ROC AUC, specificity, sensitivity and positive predictive value (PPV).

**Results**

**Study cohort**

Five-thousand-and-fifty patients with a total of 8,869 admissions met the inclusion criteria and were included in the model development and validation. The case definition (MR in the 71-hour period between one hour after admission and three days after admission) was met by 100 (1.1%) of the included admissions and the remaining 8,769 (98.9%) admissions were used as control-
admissions. The training set contained 3,509 patients (of which 73 patients were subjected to incident MR) with a total of 6191 admissions. The test set contained 1541 patients (of which 27 patients were subjected to incident MR) with a total of 2678 admissions. The distribution of time intervals from admission to MR in hours was: 1.15 hours (minimum), 5.50 hours (25% quantile), 19.9 hours (50% quantile), 37.4 hours (75% quantile) and 71.6 hours (maximum).

Characteristics for the study cohort (at the level of admissions) are presented in Table 1. If a patient had several known mental disorders, diagnosed according to the International Classification of Diseases 10th edition (ICD-10)\(^3\), at the time of admission, the most severe diagnosis (e.g., lowest ICD-10 F-code) was chosen. The only exception to this rule was when a patient had a psychotic disorder (ICD-10: F20-F29) and a substance use disorder (F10-F19). In such cases the psychotic disorder prevailed.

\textit{Table 1 approximately here}

All predictors (such as previous psychiatric admissions or coercion) extracted from the two register databases (the Registry of Coercive Measures in Psychiatry and the Danish Psychiatric Central Research Register) were available for all admissions. In contrast, not all predictors extracted from MidtEPJ were available for all admissions because we only considered notes written up to the first hour after admission. Admissions with at least one specific note-type from MidtEPJ present were: 3,135 (35.5%) for ‘Brøset Violence Checklist’; 5,322 (60.0%) for ‘Current Suicide Risk’; 5,945 (67.0%) for ‘Current Social Functioning’; 6,855 (77.3%) for ‘Current Objective Mental State’; 6,993 (78.8%) for ‘Subjective Mental State’. Between group differences are presented in Table 2. Patients were more likely to be assessed for violent behavior in relation to case-admissions compared to control-admissions. The opposite was the case for suicide risk and social functioning, where patients were less likely to be assessed in relation to case-admissions compared to control-admissions.

\textit{Table 2 approximately here}
Models were trained based on 86 predictors of which 8 were categorical. A total of 78 predictors were derived from electronic patient record clinical notes in natural language; 60 from ‘Subjective Mental State’; 10 from ‘Current Objective Mental State’; 8 from ‘Current Social Functioning’.

**Model performance and evaluation**

The Random Forest model was chosen as the optimal model (best performance in the training set) achieving a ROC AUC measure of 0.87 (95% CI: 0.79-0.93) on the test set (Figure 2). At 94% specificity the sensitivity was 56% and PPV was 8.1%, at 88% specificity the sensitivity was 74% and PPV was 6.0%. In other words, the 6% of admissions with the highest risk score covers 56% of the case-admissions (i.e. incident MR).

*Figure 2 approximately here*

Of the 86 predictors the final Random Forest model used a total of 45 predictors (see supplementary table S1 for a description of all used predictors). The top ten predictors are presented in Table 3 - ranked by importance according to the Out-Of-Bag Margin Reduction (a Random Forrest measure for the variables relative importance\(^{16}\)). The two most important predictors were categorical: admission type (voluntary; involuntary because of danger; involuntary because of urgent need for treatment) and Brøset Violence Checklist (total sum of zero; total sum greater than zero; missing value). The remaining eight were derived from notes in natural language of which seven belonged to ‘Subjective Mental State’ and one belonged to ‘Current Objective Mental State’.

*Table 3 approximately here*

**Discussion**

We developed a Random Forest machine learning model incorporating data from electronic health records to predict incident MR episodes occurring between one hour after admission and three days after admission. The Random Forest model achieved a ROC AUC of 0.87 (95% CI: 0.79-0.93) and a sensitivity of 56% at 94% specificity when validated using an independent test set. This
precision is exceeding- or comparable to that of other screening tools already in use in psychiatry (for detection of suicide risk)\textsuperscript{31}, but also compared to screening tools used in other medical specialties such as cardiology (for detection of cardiovascular disease)\textsuperscript{32,33}.

Out of the ten most important predictors, nine were derived from the clinical notes, and eight of these were based on clinical notes in natural language. Below, we discuss how the characteristics/behavior covered by the ten most important predictors ordered by predictive ability plausibly related to the risk of MR. Interestingly, all of these predictors relate to the psychopathology of the patient as documented by the hospital staff:

**Involuntary admission**: The most important predictor in our model was involuntary admission, which occurs when patients poses an acute danger to themselves or others or if there is an urgent need for treatment\textsuperscript{34}. Patients admitted involuntarily typically display psychotic symptoms, lack of insight and aggression. Such behavior has previously been associated with increased risk of MR\textsuperscript{35-37}.

**BVC score**: Irrespective of the type of admission, patients are screened for increased risk of violent acts using the BVC, which is based on staff assessment of confusion, irritability, boisterousness, physical threats, verbal threats and attack on objects. That the BVC score is predictive with regard to MR is therefore an expected finding, since the BVC was developed to predict violence for the next 24 hour period\textsuperscript{38} – and violent acts are an indication for MR\textsuperscript{34}.

**Somatic comorbidity**: This semantic topic identifies patients with somatic diseases, who are admitted involuntarily because of a comorbid mental disorder. This may imply that these patients suffer from organic mental disorders (e.g. delirium and/or dementia), which are known risk factors for MR\textsuperscript{8}. Sparse/non-coherent verbal response and non-informative verbal response: Although the terms defining these topics were inconclusive, the notes loading on these topics clearly identified patients that typically were either unable or unwilling to answer questions asked by the staff. Thus, because the answers from the patients were missing/non-informative, the health professional had to document that despite they ("I") had asked the right questions the answer had been inadequate. This lack of coherent response likely reflected underlying severe psychopathology such as marked thought disorder (illogical or incoherent speech) or prominent
persecutory delusions. That psychopathology of this nature is associated with increased risk of MR is not surprising from a clinical perspective. We are however somewhat surprised that such a complex construct can seemingly be identified using unsupervised machine learning.

**Abnormal behavior:** This semantic topic was derived from the ‘Current objective mental state’ notes and therefore reflects the health care professionals’ assessments of the behavior of the patients. The terms defining this topic were rather vague and did not reveal a definite label. However, many of the clinical notes loading on this topic described patients (predominantly males) who were behaving unexpectedly and/or abnormally although not to a degree where the patient was clearly psychotic. Therefore, this topic likely represents sub-clinical psychotic states – and psychosis is a known risk factor for MR\(^{39}\).

**Threatening behavior:** That patients with threatening behavior, who – according to the words included in this topic – may even threaten to kill other people or themselves are at elevated risk of MR is expected given that such behavior is an indication for MR according to Danish law\(^{34}\).

**Good social status:** This finding is in line with that of a prior study showing that social factors such as high socioeconomic status and stable family relations are associated with reduced risk for MR\(^{8}\).

**Suicidal ideation - car crash:** It has been shown that suicidal ideation is associated with the risk of MR \(^{39}\). Interestingly, our results seem to indicate that suicidal ideation or suicide attempts involving a car (crash), may have a particularly strong association to MR.

**Persecutory ideation:** This semantic topic is defined by words such as “apartment”, “police”, “neighbor” and “surveillance”, and therefore clearly seems to tap into persecutory ideation. Patients in this state may be convinced that the hospital staff or fellow patients are plotting against them or trying to hurt them and may respond with hostility or overt aggression, which ultimately leads to MR.

As outlined above, all of the ten most important predictors seem to plausibly relate to the risk of MR. In other words – the predictors are meaningful from a clinical perspective, something that might prove useful when trying to understand and explain the model’s decisions to patients. However, care should – as always with statistical measures – be taken not to infer a direct causal relationship between the identified predictors and MR\(^{40,41}\). Conversely, we were initially somewhat surprised to see that none of the patients subjected to MR appeared to have a
diagnosis of a personality disorder (F60-F69), mental retardation (F70-F79) or a psychological
development disorder such as autism (F80-F89) in our dataset. There are probably four main
explanations for this finding, namely i) that patients who had a contact to a psychiatric
department in CDR before implementation of MidtEPJ were excluded from our analysis, ii) that
more than 50% of the case-admissions represent the very first admission to a psychiatric hospital
in the CDR, iii) that patients with a personality disorder will often suffer from a comorbid
depressive- or anxiety disorder, which will “overrule” the personality disorder in the diagnostic
hierarchy employed in this study, and iv) that patients with mental retardation or autism will
almost exclusively have been diagnosed within child- and adolescent psychiatric services.

This study has demonstrated that it is possible to identify patients at high risk for incident MR
under a ‘treatment as usual’ management. The immediate clinical benefit of that knowledge is
that it will allow the staff to make more timely and/or more intensive interventions and thereby
reduce the risk of the first MR episode. By avoiding MR, patients will experience a less traumatic
admission\textsuperscript{42}, the effect of which is likely to carry over to potential subsequent admissions and thus
reduce the risk of MR in the future as well. In other words, by avoiding the first MR episode, the
patient’s entire psychiatric treatment trajectory may take a more positive direction. Avoiding
recurrent episodes of MR would be a tremendous relief for patients, staff members and hospital
systems.

Another benefit of having an algorithm assisting risk assessment in relation to MR is that it may,
only to some extent of course, compensate for lack of clinical experience with this type of
assessment among staff members. We do not believe that the model should or could replace
clinical assessment but given the mere rarity of incident MR episodes, it is difficult for health care
professionals to gain sufficient clinical experience in identifying patients at high risk for incident
MR. Even when predicting a non-rare event such as any restraint episodes within the next 24
hours health care professionals, being assisted by a risk assessment tool, classified MR with a ROC
AUC of only 0.73 (95% CI: 0.69-0.77)\textsuperscript{18}. The timely early warning from the model developed in our
study would thus represent a substantial improvement. Considering, that one reason for
mechanical restraint is the perceived risk of future violence, another approach to identify patients
at elevated risk of MR would be to employ assessment tools aimed at identifying the risk of
violence. A systematic review has identified over 150 risk-assessment tools for violence. Another systematic review investigated the nine most commonly used of these tools and found that a majority of the identified studies obtained ROC AUC measures of less than 0.80. When comparing our results with many of these risk-assessment tools, we note that these tools often predict violence in the long term (often months) and/or in specific populations (e.g. forensic patients, outpatients, sexual offenders), whereas our model predicts into the immediate future for all admissions, making it a much more practical assessment tool. A recent review describes previous assessment tools evaluating the acute risk of violence among psychiatric inpatients, and these tools obtain ROC AUC values ranging between 0.85-0.93 (Brøset Violence Checklist), 0.71-0.86 (Dynamic Appraisal of Situational Aggression – Inpatient Version), 0.82-0.84 (Violence Risk Screening–10) and 0.61-0.84 (Short-term Assessment of Risk and Treatability). In a direct comparison, our model (AUC ROC=0.87) thus performs equally well or better than these risk-assessment tools. However, this comparison is obviously suboptimal since i) these tools were not designed nor tested to identify patients at risk for incident violence and ii) in several of these studies the health care professionals were allowed to make interventions to reduce the risk of violence. Importantly, it is therefore not the natural course that is predicted in these studies and thus the precision depends on the effect of the de-escalating intervention being used. Ideally, the predictive power of our algorithm versus that of classical risk-assessment tools should be investigated head to head in a prospective randomized controlled trial.

The results of this study have implications for both clinical practice and for future research. From the clinical perspective, the predictive ability of our model has spurred the administrators of the psychiatric hospitals in the CDR to initiate a process in which an automatic MR early warning system developed based on our results will be feasibility tested in clinical practice. If this feasibility test is successful, the system may subsequently be implemented throughout the region. Furthermore, if the results of this study are replicated within psychiatric services elsewhere – implementation of early warning systems could follow in these settings as well. From the research perspective, our results serve as a proof of concept that clinical notes written as part of standard psychiatric assessments contain important information that can readily be extracted by unsupervised machine learning algorithms. This suggests that machine learning may be a very
useful tool for clinical psychiatry in the future – a perspective, which definitely merits further study.

This study has several strengths. First, the prediction was based on data, which was already collected as part of standard clinical care. Hence, when using our model, no additional data collection from health professionals is needed to identify patients at high risk of MR. We believe that this is important since it would allow the sparse human resources at psychiatric wards to be allocated to actual prevention of MR rather than identification of those at risk. Second, our model identifies high-risk patients already one hour after admission. This is important since many patients have no or few prior admissions and are therefore often not known by the staff responsible for welcoming them to the ward. Third, the prediction is valid for the next 71 hours following the first hour of an admission and therefore secure the staff time to launch measures that could potentially prevent MR.

The findings reported here should be interpreted in the context of the limitations of the study. First and foremost, since our predictors from natural language lean heavily on the thematic structure of the clinical notes in MidtEPJ, the Random Forest prediction model developed in this study may not be applicable to data from other electronic health record systems without contextual adaptation. This is a known limitation within predictive machine learning. However, since the use of coercive measures appears to be quite dependent on local conditions, we do not believe that a universally functioning model can be developed. Rather, the purpose of this study was to investigate if prediction of MR was possible using the wealth of information, which is registered in electronic health records at psychiatric hospitals. However, since electronic health care data are available in many other psychiatric hospital settings and machine learning algorithms are increasingly used, prediction models tailored to these specific settings could readily be developed – for instance by following the approach used in the study presented here.

Second, 8,038 patients with an admission in CDR were excluded from the analysis because they had a contact to a psychiatric department in CDR, which was not documented in MidtEPJ. This approach was chosen because the predictive value of information from clinical notes written before an admission were unknown a priori to this study. However, since the prediction model
developed in this study only base its predictions on clinical notes written within the past month before an admission, insufficient prior patient records appear to be less of a concern. Therefore, future studies need not be as restrictive in their sampling criterions as we were.

Third, when identifying which admissions to exclude in this study (those with double admission and those subjected to MR within an hour from admission), some clinical notes were read manually. This manual handling cannot be carried out by an automatic decision support system. Therefore, such patients would be scored by the model if used in the clinic. However, this limitation has little practical relevance for prediction accuracy and the reported AUC of 0.87 is the predictive precision for admissions where an action would be appropriate and feasible.

In conclusion, using a machine learning approach, we were able to develop and validate a model that could accurately predict incident MR. These findings are forming the basis of the ongoing development of an automatic early warning system aimed at preventing MR at the psychiatric hospitals where the data from this study stem from. A feasibility test of this system is currently being planned. Future studies should aim at replicating the results of this study based on data from psychiatric services in other countries and settings.

References


Table 1. Characteristics for patients (recorded at the time of each included admission).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Control-admissions</th>
<th>Case-admissions</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Column %</td>
<td>N</td>
</tr>
</tbody>
</table>

Age Group
<table>
<thead>
<tr>
<th>Age Group</th>
<th>Total</th>
<th>Percentage</th>
<th>Male</th>
<th>Female</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;30 years old</td>
<td>3,042</td>
<td>34.69 %</td>
<td>48</td>
<td>48.00 %</td>
<td>3,090</td>
</tr>
<tr>
<td>30-45 years old</td>
<td>2,237</td>
<td>25.51 %</td>
<td>17</td>
<td>17.00 %</td>
<td>2,254</td>
</tr>
<tr>
<td>45-60 years old</td>
<td>1,810</td>
<td>20.64 %</td>
<td>17</td>
<td>17.00 %</td>
<td>1,827</td>
</tr>
<tr>
<td>&gt;60 years old</td>
<td>1,680</td>
<td>19.16 %</td>
<td>18</td>
<td>18.00 %</td>
<td>1,689</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>4,495</td>
<td>51.26 %</td>
<td>69</td>
<td>69.00 %</td>
<td>4,564</td>
</tr>
<tr>
<td>Female</td>
<td>4,274</td>
<td>48.74 %</td>
<td>31</td>
<td>31.00 %</td>
<td>4,305</td>
</tr>
<tr>
<td><strong>Prior mental disorders</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organic mental disorders (F00-F09)</td>
<td>223</td>
<td>2.54 %</td>
<td>6</td>
<td>6.00 %</td>
<td>229</td>
</tr>
<tr>
<td>Substance abuse disorder (F10-F19)</td>
<td>821</td>
<td>9.36 %</td>
<td>9</td>
<td>9.00 %</td>
<td>830</td>
</tr>
<tr>
<td>Psychotic disorders (F20-F29)</td>
<td>956</td>
<td>10.90 %</td>
<td>15</td>
<td>15.00 %</td>
<td>971</td>
</tr>
<tr>
<td>Mood disorders (F30-F39)</td>
<td>2,137</td>
<td>24.37 %</td>
<td>8</td>
<td>8.00 %</td>
<td>2,145</td>
</tr>
<tr>
<td>Anxiety disorders (F40-F48)</td>
<td>737</td>
<td>8.40 %</td>
<td>8</td>
<td>8.00 %</td>
<td>745</td>
</tr>
<tr>
<td>Sleep and eating disorders (F50-F59)</td>
<td>11</td>
<td>0.13 %</td>
<td>0</td>
<td>0.00 %</td>
<td>11</td>
</tr>
<tr>
<td>Personality disorders (F60-F69)</td>
<td>47</td>
<td>0.54 %</td>
<td>0</td>
<td>0.00 %</td>
<td>47</td>
</tr>
<tr>
<td>Mental retardation (F70-F79)</td>
<td>8</td>
<td>0.09 %</td>
<td>0</td>
<td>0.00 %</td>
<td>8</td>
</tr>
<tr>
<td>Disorders of psychological development (F80-F89)</td>
<td>8</td>
<td>0.09 %</td>
<td>0</td>
<td>0.00 %</td>
<td>8</td>
</tr>
<tr>
<td>Other behavioral and emotional disorders in</td>
<td>106</td>
<td>1.21 %</td>
<td>0</td>
<td>0.00 %</td>
<td>106</td>
</tr>
<tr>
<td>Predictor name</td>
<td>Number of control-admissions containing note-type (% of all control-admissions)</td>
<td>Number of case-admissions containing note-type (% of all case-admissions)</td>
<td>Chi-Square</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------------------------</td>
<td>--------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------</td>
<td>------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brøset Violence Checklist</td>
<td>3,090 (35.2%)</td>
<td>45 (45.0%)</td>
<td>0.042</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Current Risk of Suicide</td>
<td>5,277 (60.2%)</td>
<td>45 (45.0%)</td>
<td>0.002</td>
<td></td>
<td></td>
</tr>
<tr>
<td>‘Current Social Functioning’</td>
<td>5,891 (67.2%)</td>
<td>54 (54%)</td>
<td>0.006</td>
<td></td>
<td></td>
</tr>
<tr>
<td>‘Current Objective Mental State’</td>
<td>6,775 (77.3%)</td>
<td>80 (80.0%)</td>
<td>0.516</td>
<td></td>
<td></td>
</tr>
<tr>
<td>‘Subjective Mental State’</td>
<td>6,913 (78.8%)</td>
<td>80 (80.0%)</td>
<td>0.777</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Group differences of number of admissions with at least one specific note-type present.
Table 3. Top ten predictors ranked by importance according to the Out-Of-Bag Margin Reduction.

<table>
<thead>
<tr>
<th>Data Source</th>
<th>Out-Of-Bag: Margin Reduction</th>
<th>Terms (Freely translated from Danish)</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>Register (type of admission)</td>
<td>7.9x10^{-4}</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>MidtEPJ (BVC score)</td>
<td>5.1x10^{-4}</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>MidtEPJ (Subjective Mental State)</td>
<td>3.0x10^{-4}</td>
<td>Department, paper, somatic, red*, admission</td>
<td>Somatic comorbidity</td>
</tr>
<tr>
<td>MidtEPJ (Subjective Mental State)</td>
<td>1.9x10^{-4}</td>
<td>I, I, ask, we, say</td>
<td>Sparse/non-coherent verbal response</td>
</tr>
<tr>
<td>MidtEPJ (Subjective Mental State)</td>
<td>1.7x10^{-4}</td>
<td>Answer, question, describe, asked, answered</td>
<td>Non-informative verbal response</td>
</tr>
<tr>
<td>MidtEPJ (Current Objective Mental State)</td>
<td>8x10^{-5}</td>
<td>He, he, they, like, at</td>
<td>Abnormal behavior (males)</td>
</tr>
<tr>
<td>MidtEPJ (Subjective Mental State)</td>
<td>7x10^{-5}</td>
<td>Kill, hit, threaten, increase, man</td>
<td>Threatening behavior</td>
</tr>
<tr>
<td>MidtEPJ (Subjective Mental State)</td>
<td>6x10^{-5}</td>
<td>Good, may, social, current, child</td>
<td>Good social status</td>
</tr>
<tr>
<td>MidtEPJ (Subjective Mental State)</td>
<td>6x10^{-5}</td>
<td>Car, drive, driver, crash, wife</td>
<td>Suicidal ideation - car crash</td>
</tr>
<tr>
<td>MidtEPJ (Subjective Mental State)</td>
<td>5x10^{-5}</td>
<td>Apartment, police, neighbor, surveillance, him</td>
<td>Persecutory ideation</td>
</tr>
</tbody>
</table>
Our clinical interpretations of these predictors are put into perspective in the discussion section. BVC: Brøset Violence Checklist, N/A: Not applicable. *Refers to so-called “red papers”, which are used (compulsory) when patients are admitted involuntarily.
Figure 1. Flowchart illustrating the selection of the sample used in the study

Individuals admitted to a psychiatric department in CDR between implementation of MidtEPJ and 01.01.2016 (N = 13,458)

All psychiatric admissions in CDR documented in MidtEPJ (N = 5,420)

Patients with contact to a psychiatric department in CDR not documented in MidtEPJ (N = 8,038)

Patients with an MR episode before their first admission in CDR (N = 102)

At least one admission with no prior MR (N = 5,318)

< 18 years old (N = 18)

≥ 18 years old (N = 5,300)

Patients with i) MR within the first hour during their first admission or ii) only double admissions (N = 250)

Excluding double-admissions and admissions with MR within the first hour (N = 5,050)

MR: Mechanical restraint.
Figure 2: Receiver Operating Characteristic for the random forest prediction model in the test set.