BMJ Open

Predicting the risk of drug–drug interactions in psychiatric hospitals: a retrospective longitudinal pharmacovigilance study

Jan Wolff [ORCID] ,[1,2] Gudrun Hefner,[3] Claus Normann,[2] Klaus Kaier [ORCID] ,[4] Harald Binder,[4] Katharina Domschke,[2] Christoph Hiemke,[5] Michael Marschollek,[1] Ansgar Klimke[6,7]

Correspondence to
Dr Jan Wolff;
janwolff123@gmail.com

ABSTRACT

Objectives The aim was to use routine data available at a patient's admission to the hospital to predict polypharmacy and drug–drug interactions (DDI) and to evaluate the prediction performance with regard to its usefulness to support the efficient management of benefits and risks of drug prescriptions.

Design Retrospective, longitudinal study.

Setting We used data from a large multicentred pharmacovigilance project carried out in eight psychiatric hospitals in Hesse, Germany.

Participants Inpatient episodes consecutively discharged between 1 October 2017 and 30 September 2018 (year 1) or 1 January 2019 and 31 December 2019 (year 2).

Outcome measures The proportion of rightly classified hospital episodes.

Methods We used gradient boosting to predict respective outcomes. We tested the performance of our final models in unseen patients from another calendar year and separated the study sites used for training from the study sites used for performance testing.

Results A total of 53 909 episodes were included in the study. The models' performance, as measured by the area under the receiver operating characteristic, was 'excellent' (0.83) and 'acceptable' (0.72) compared with common benchmarks for the prediction of polypharmacy and DDI, respectively. Both models were substantially better than a naive prediction based solely on basic diagnostic grouping.

Conclusion This study has shown that polypharmacy and DDI can be predicted from routine data at patient admission. These predictions could support an efficient management of benefits and risks of hospital prescriptions, for instance by including pharmaceutical supervision early after admission for patients at risk before pharmacological treatment is established.

INTRODUCTION

The most common medical decision is the prescription of medicines.[1] Pharmacotherapy is also essential for the treatment of mental and behavioural disorders,[2–6] where more than 130 different drugs with proven efficacy are currently available.[7] The combination of multiple drugs is required in many

Strengths and limitations of this study

► Detailed and broad longitudinal prescription data from eight hospitals, allowing to delineate the daily prescription patterns for each hospital episode and identifying polypharmacy and drug–drug interactions.

► Restricted set of features in prediction models which should already be routinely available or easily possible to implement in many hospitals.

► Relatively low risk of information leakage and overfitting by testing the prediction performance in patients that were treated in another calendar year and by separating the study sites that were used for model training from the study sites that were used for testing the prediction performance.

► No delineation of patient-specific benefit-–risk balances of prescriptions, for instance by including drug serum levels, results of electrocardiograms or individual pharmacogenetic risk factors.

► Restriction to study sites from one large provider of inpatient psychiatric services in the region of Hesse, Germany.

clinical situations,[8] but this is associated with an increased risk of drug–drug interactions (DDI),[9] which enhances the risk of adverse drug reactions (ADR).[10]

The simultaneous use of five or more different pharmaceuticals is defined as polypharmacy.[11 12] Its prevalence is high in both outpatient and inpatient settings, especially in old aged patients.[13–17] Cost savings from reducing polypharmacy can be substantial.[18] Due to an ageing population and an increasing number of patients with multimorbidity, polypharmacy is likely to become more frequent in the future, thereby further complicating the efficient management of benefits and risks of prescriptions in hospital psychiatry.[19]

DDI can be pharmacokinetic or pharmacodynamic in nature. In psychiatry, most

pharmacokinetic DDI are related to the cytochrome P450 (CYP)-mediated drug metabolism.[20 21] Pharmacodynamic DDI arise when the pharmacological effect of one drug is affected directly by another one.[22] The latter is most relevant when combining drugs with anticholinergic or QT interval prolonging activity. Anticholinergic drugs decrease or block the actions of acetylcholine on central and peripheral acetylcholine receptors and may lead to multiple ADR like dry mouth, confusion, constipation, urinary retention, falls or delirium.[23 24] Combining anticholinergic drugs increases the risk of the occurrence and severity of the respective ADR.[24–26] QT interval prolonging properties of drugs can lead to ventricular arrhythmia (Torsade de Pointes, TdP) and cardiac arrest.[27–30] Not all but most DDI are related to negative clinical outcomes and this relationship has been studied thoroughly. Therefore, for most drug combinations avoiding of DDI is a way to improve drug safety.

The use of information technology and the inclusion of hospital pharmacists are frequently discussed ways to improve efficient management of benefits and risks of prescriptions in hospitals. While research has shown ambiguous results for the effectiveness and acceptance of information technology to support continuous drug management during the hospital stay,[31–35] systematic reviews have found improved patient outcomes for the inclusion of hospital pharmacists in adult[36] and paediatric inpatients.[37]

There is a lack of evidence considering the potential to identify patients at risk of DDI at their admission to the hospital. If potential DDI risks can be detected before a patient's pharmacological treatment is established, drug safety can improve by early focus on balancing patient-specific, combined benefit–risk ratios of all prescriptions. Furthermore, if polypharmacy can be predicted in advance, that is, detecting cases that will receive at least five drugs simultaneously, this allows early identification of patients that will have particular needs of drug safety management. The aim of the present study was to use routine data available at admission to the hospital in order to predict polypharmacy and DDI during the stay, respectively, and to evaluate the prediction performance with regard to its usefulness to support the efficient management of benefits and risks of drug prescriptions. A further aim was to compare the results of a machine learning approach with those achieved by means of logistic regression and by means of a naive baseline classifier.

METHODS
Design, setting and participants
We carried out a retrospective, longitudinal study in eight psychiatric hospitals in Hesse, Germany. Our study included all inpatient episodes consecutively discharged between 1 October 2017 and 30 September 2018 (year 1) or 1 January 2019 and 31 December 2019 (year 2). Episodes at departments for child and adolescent psychiatry were excluded. An inpatient episode was defined as an individual patient's stay at the hospital between admission and formal discharge with a planned duration of at least one complete day and night. Our study was part of a large clinical pharmacovigilance project sponsored by the Innovations Funds of the German Federal Joint Committee ('Optimization of pharmacological treatment in hospitalized psychiatric patients (OSA-PSY)', study number 01VSF16009).

Patient and public involvement
Patients and public were not directly involved in the design of this retrospective study.

Data
Our study used daily medication data from the electronic medical records at the study sites. These data contained each individual medication for each day of an inpatient episode. This allowed to investigate the medications given at each day separately and to include all modifications of pharmaceutical treatment during a stay. We defined polypharmacy as the concurrent prescription of at least five drugs per day.[11 12] DDI were defined as (i) pharmacokinetic cytochrome P450 CYP-mediated DDI (CYP450-Interaction), (ii) the prescription of more than one concurrent anticholinergic drug (Antichol. Combi.) and (iii) the prescription of more than one concurrent QT interval prolonging drug (QT-Combi.).

Daily prescription data per episode were matched with DDI-information obtained from guidelines and additional studies. CYP-mediated drugs were identified in accordance to the Consensus Guidelines for Therapeutic Drug Monitoring in Neuropsychopharmacology[38] restricted to clinically relevant drugs according to the Food and Drug Administration.[39] In addition, melperone,[40] levomepromazine[41] and perazine[42–44] were considered as CYP inhibitors. Additional non-psychotropic victim drugs were added based on CYP substrate properties defined by Hiemke and Eckermann.[45] In total, these sources resulted in covering the following isoforms for analyses of CYP-mediated DDI: CYP1A2, CYP2B6, CYP2C19, CYP2C9, CYP2D6, CYP2E1, CYP3A4.

QT interval prolonging drugs were identified based on the lists of the Arizona Center for Education and Research, which maintains lists of drugs that have either a known or a possible risk for TdP.[46 47] Anticholinergic activity of drugs was identified according to Hiemke and Eckermann, Chew *et al* and Lertxundi *et al*.[45 48 49]

The candidate feature variables that might potentially predict polypharmacy and DDI were obtained from routinely documented information in the electronic medical records and patient administration databases. A restricted set of feature variables was used that should be available in many hospitals at admission of patients. These were patients' age at admission, gender, somatic and psychiatric diagnoses according to the 10th revision of the International Statistical Classification of Diseases and Related Health Problems-German Modification (ICD-10-GM), the treatment setting (inpatient vs

day-clinic) and the scores at admission from the Global Assessment of Functioning Scale,[50] the Clinical Global Impression,[51] the Beck Depression Inventory[52] as well as the Positive and Negative Syndrome Scale.[53]

Analysis

The study compared the results obtained through a machine learning approach with logistic regression and a naive baseline classifier. The chosen machine learning approach was gradient boosting with trees,[54] as implemented in the CARET package in R.[55 56] For comparison, we carried out a logistic regression with the same feature variables that were used in the machine learning approach. Furthermore, a naive baseline classifier was obtained by using only basic diagnostic groups in a logistic regression. The basic diagnostic groups were F0/G3 organic mental disorders/other degenerative diseases of the nervous system, F1 substance-related mental disorders, F2 schizophrenia, schizotypal and delusional disorders, F3 affective disorders and others.

We trained, validated and tested our models on different patients in order to avoid overly optimistic results from the evaluation of their prediction performance. Therefore, we divided data into a training set, that is, a random 80% sample of patients discharged in the first year, a validation set, that is, the remaining 20% of patients of the first year and a test set, that is, patients discharged in the second year. After separating the datasets, missing data were addressed by mean imputation and categories representing missingness of each individual feature.

We tuned hyperparameters on the basis of the trained models' performance in the validation data set using the built-in grid search process in the CARET package, thereby modifying each of the four tuning parameters, that is, boosting iterations, max depth of trees, shrinkage and minimal terminal node size, until a maximum performance was reached in the validation sample.

The final models' performance was assessed in the test dataset of episodes discharged in year 2. For judging variability of performance and robustness across different study-sites, we thereby trained eight different models of episodes discharged in year 1, each holding-out one study-site, and used these models to predict the outcomes of episodes discharged in year 2 from the held-out study site to restrict assessment of performance to hospitals not involved in the training process.

Prediction performance was compared using receiver operating characteristic curves (ROC) and Precision and Recall plots (PR-plots). We calculated 95% DeLong CIs for the area under the ROC.[57] Furthermore, we defined different cut-off values for the operationalisation of the models that maximised sensitivity at a minimum precision of 0.6, 0.7 and 0.8. We chose a specificity threshold of 0.2 to be the minimum for a clinically meaningful application based on previous work of Tomašev *et al*.[58] Furthermore, we defined a sensitivity of 0.2 as the minimum threshold for clinically meaningful application.

The results of machine learning models may be difficult to understand since the processes between input and output are opaque.[59] This aspect is sometimes referred to the term 'black-box'. In response, several methods have been developed to make machine learning models more interpretable.[60] In our study, we calculated so called accumulated local effects of feature variables to show the average model prediction over the feature and to make their influence interpretable. Accumulated local effects describe the way features influence the prediction of a machine learning model on average.[61] They are computed as accumulated differences over the conditional distribution and are considered unbiased even when features are correlated.[62] The detailed methods for calculating accumulated local effects are described in detail in Molnar[61] and Apley and Zhu.[62]

RESULTS

A total of 53 909 episodes were included in the study. Table 1 shows the characteristics of the episodes in year 1 and year 2. The study periods used for training and validation (year 1) and for testing (year 2) were very similar in terms of both total number of inpatient episodes and their characteristics.

Figure 1 shows the association between the maximum number of different prescriptions an episode has received per day and the proportion of episodes with a DDI. As expected, the likelihood of all DDI increased with the number of prescriptions. It appears remarkable that the risk of QT-Combi. showed the steepest increase between one and four different medications and levelled off afterwards. In contrast, the risk of CYP450-Interaction and Antichol-Combi. increased at a relatively constant rate. Further detailed results describing the frequencies of DDI and polypharmacy are provided in online supplemental figure S1 (for each diagnosis group) and the online supplemental figure S2 (for each combination of DDI and polypharmacy). The overall incidence of at least one DDI or receiving polypharmacy was relatively high (63%) and therefore an import aspect of clinical management.

Figure 2 compares the possible combinations of sensitivity, that is, the proportion of correctly predicted actual positives, and specificity, that is, the proportion of correctly predicted actual negatives, that were reached by the different classifications. Furthermore, the operational points at the curves that maximise sensitivity at different minimum levels of precision are shown. Measured by the area under the ROC, the models for polypharmacy and DDI achieved a relatively good performance.

The area under the ROC can be a potentially misleading measure of model performance when observations are distributed very unbalanced between classes.[63] Therefore, figure 2 also compares the possible combinations of recall, a synonym for sensitivity, and precision, that is, the proportion of actual positives among all positive predictions. The analogous figures for the ROC and the PR-plots for each individual DDI are shown in online supplemental figure S3.

Table 1 Patient characteristics

		Year 1		Year 2	
Number of episodes (n)		26 949		26 960	
Study site (n and %)	1	1971	7	2039	8
	2	5977	22	6108	23
	3	2172	8	2157	8
	4	3432	13	3447	13
	5	5087	19	4874	18
	6	2710	10	2831	10
	7	3125	12	3073	11
	8	2475	9	2431	9
Day-clinic (n and %)		3444	13	3394	13
Female (n and %)		12 304	46	12 343	46
Age at admission (years, mean and SD)		47	18	47	18
Length of stay (days, median and IQR)		17	8–33	17	8–33
Diagnostic group (n and %)	F0/G3	2419	9	2330	9
	F1	8980	33	8791	33
	F2	4048	15	4086	15
	F3	8147	30	8259	31
	F4	1722	6	1650	6
	F6	1252	5	1347	5
	F7	185	1	265	1
	Others	196	1	228	1
	NA	0	0	4	0

IQR 25th–75th percentile.
Year 1: 1 October 2017 to 30 September 2018.
Year 2: 1 January 2019 to 31 December 2019.
F0, organic, including symptomatic, mental disorders; F1, mental and behavioural disorders due to psychoactive substance use; F2, schizophrenia, schizotypal and delusional disorders; F3, mood (affective) disorders; F4, neurotic, stress-related and somatoform disorders; F6, disorders of adult personality and behavior; F7, mental retardation; G3, other degenerative diseases of the nervous system.

Figure 3 provides additional measures of classification performance based on a set of different potentially meaningful operational points. These should help to facilitate a more intuitively understanding of actual clinical usability. For instance, the model for polypharmacy operated at a minimum precision of 0.70 generated three false alerts for seven true alerts. It triggered alerts for one half of all admissions, thereby rightly identifying slightly more than
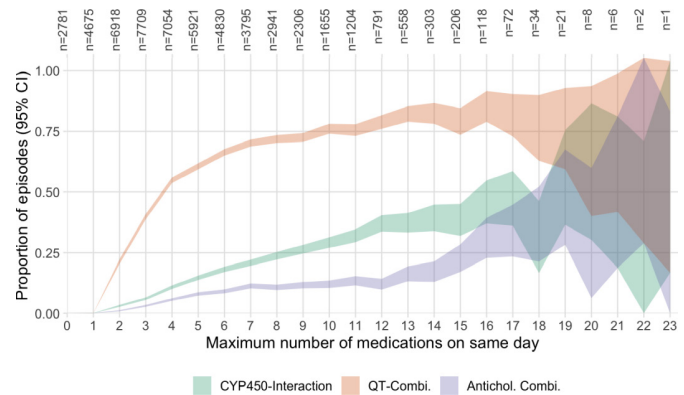

Figure 1 95% CI of proportion of hospital episodes with drug–drug interactions versus maximum number of medications per day. CYP450-Interaction: pharmacokinetic cytochrome P450 (CYP)-mediated drug–drug interaction. QT-Combi.: a combination of at least two drugs on the same day with known or possible risk of Torsade de Pointes according to the Arizona Center for Education and Research classification. Antichol. Combi.: a combination of at least two drugs on the same day with at least moderate anticholinergic activity.

three fourths of all actual positives and 72% of all actual negatives.

Two additional measures of model performance that are commonly used to evaluate the performance of clinical prediction models are calibration, that is, how consistent the predictions were with the observed rates, and the learning curve, that is, how the prediction performance increased with increasing number of episodes in the
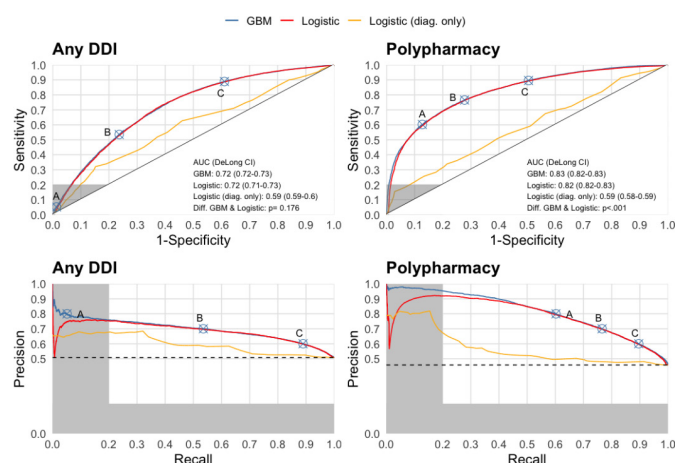

Figure 2 Receiver operating characteristic curves and precision and recall plots. Polypharmacy, receiving at least five different medications at the same day. A: precision at least 80%, B: precision at least 70%, C: precision at least 60%. Crossed circles show cut-off values that maximise sensitivity at different minimum thresholds of precision. Grey areas are not clinically meaningful because of a sensitivity or precision of less than 0.2. Dashed horizontal lines show the prevalence of the outcome. Diagonal lines show the random classifier bottom line. AUC, area under the curve; DDI, drug–drug interactions; GBM, gradient boosting machine.

	Preva-lence	Trig. Rate	True Pos. Rate	True Neg. Rate	False Pos. Rate	False Neg. Rate	Pos. Pred. Value	Neg. Pred. Value	
	0.51	0.75	0.89	0.39	0.61	0.11	0.6	0.77	Any DDI (Pr.: 0.6)
	0.51	0.39	0.54	0.76	0.24	0.46	0.7	0.61	Any DDI (Pr.: 0.7)
	0.51	0.03	0.05	0.99	0.01	0.95	0.8	0.5	Any DDI (Pr.: 0.8)
	0.46	0.68	0.9	0.5	0.5	0.1	0.6	0.85	Polyphar (Pr.: 0.6)
	0.46	0.5	0.77	0.72	0.28	0.23	0.7	0.78	Polyphar (Pr.: 0.7)
	0.46	0.34	0.6	0.87	0.13	0.4	0.8	0.72	Polyphar (Pr.: 0.8)

Figure 3 Measures of prediction performance. Polyphar.: receiving at least five different medications at the same day. Pr.: precision of at least. Prevalence: proportion of episodes with observed positive outcome. Trig. Rate: proportion of episodes that cause a positive prediction. True positive rate (a.k.a. sensitivity and recall): proportion of actual positives that are correctly identified as such. True negative rate (a.k.a. specificity): proportion of actual negatives that are correctly identified as such. False positive rate: proportion of actual negatives that are falsely predicted as positives. False negative rate,; proportion of actual positives that are falsely predicted as negatives. Pos. Pred. value (a.k.a. precision): proportion of actual positives in all positive predictions. Neg. Pred. value: proportion of actual negatives in all negative predictions. DDI, drug–drug interactions.

training data. These two measures are provided in online supplemental figures S4 and S5.

The overall prediction performance does not inform about the influence of individual feature variables on the predicted outcomes. These influences are not readily accessible in machine learning applications. Therefore, figures 4 and 5 show the accumulated local effects of the top eight features ranked by their variable importance in an effort to show the average model predictions across the features and make their influence interpretable. Variable importance is a dimensionless measure that represents the influence of each feature on the prediction performance relative to the other variables (the method is described in detail in Friedman[54]).

The plots describe how the features influenced the prediction on average, with negative values representing decreasing probabilities and positive values representing increasing probabilities. Furthermore, the plots show the type of associations between outcome and feature variable, that is, whether they were linear, monotonic or more complex. The influence of feature variables on outcomes was as clinically expected, supporting the assumption that the models provided meaningful predictions.

DISCUSSION
Key findings
The aim of the present study was to use routine data available at patient admission to the hospital to predict polypharmacy and DDI during the stay as well as to evaluate the prediction performance with regard to its usefulness
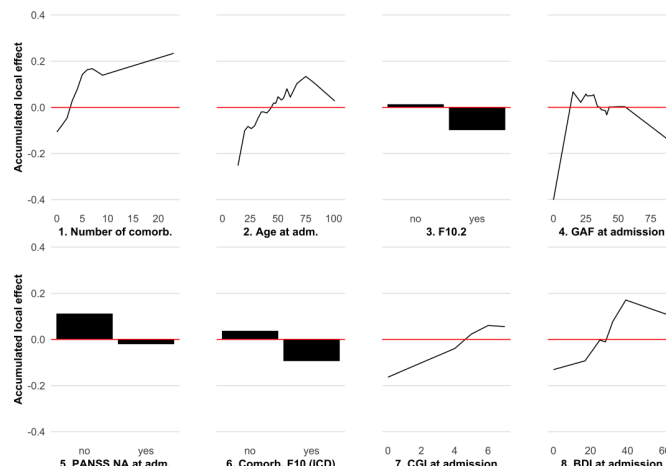

Figure 4 Accumulated local effects of top eight feature variables in predicting any drug–drug interaction. Comorb.: comorbidities. Age at adm.: age at admission. F10.2: dependence syndrome due to use of alcohol as main diagnosis. GAF at adm.: Global Assessment of Functioning at admission. PANSS NA at adm.: Positive and Negative Syndrome Scale missing at admission. Comorb. F10 (ICD): mental and behavioural disorders due to use of alcohol as secondary diagnosis. CGI at adm.: Clinical Global Impression at admission. BDI at adm.: Beck Depression Inventory at admission.

to support the efficient management of benefits and risks of drug prescriptions. A further aim was to compare the results of a machine learning approach with those achieved by means of logistic regression and by means of a naive baseline classifier. The models' performance, as measured by the area under the ROC, was 'excellent'
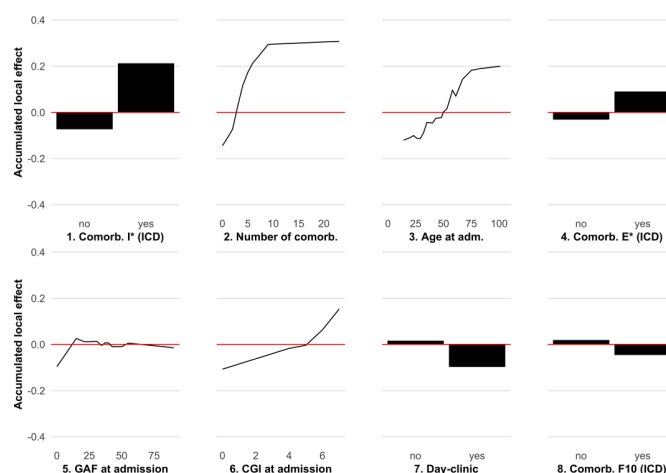

Figure 5 Accumulated local effects of top eight feature variables in predicting polypharmacy. Polypharmacy: receiving at least five different medications at the same day. Comorb. I* (ICD): diseases of the circulatory system as secondary diagnosis. Comorb.: comorbidities. Age at adm.: age at admission. Comorb. E* (ICD): endocrine, nutritional and metabolic diseases as secondary diagnosis. GAF at adm.: Global Assessment of Functioning at admission. CGI at admission: Clinical Global Impression at admission. Comorb. F10 (ICD): mental and behavioural disorders due to use of alcohol as secondary diagnosis.

and 'acceptable' compared with common benchmarks[64] for the prediction of polypharmacy and DDI, respectively. The performance of the machine learning approach was very similar to the performance of the logistic regression models. Both models were substantially better than a naive prediction based solely on basic diagnostic grouping.

A reliable identification of patients at admission that will likely receive polypharmacy during their stay or are at risk of DDI would be beneficial for an efficient management of pharmacotherapy and the promotion of patient safety.[65] These identifications would allow, for instance, the early involvement of pharmacists and specialised physicians that could supervise potentially complicated constellations and avoid DDI before a patient's pharmacological treatment is established.

However, it is difficult to objectively define the prediction performance required for a useful application in routine clinical practice and this was out of the scope of the present study (see for instance[66 67]). Different applications and clinical settings likely require their own trade-off decisions between reducing false alerts and increasing coverage of actual positives, sometimes requiring intensive focus on the avoidance of missing actual positive cases. Our model was trained, tested and validated in the context of inpatient psychiatry. It has the potential to be useful in any clinical setting where polypharmacy and DDI are relevant and where the feature variables are readily available.

Regardless of prediction performance, machine learning techniques must be used responsibly in clinical practice, otherwise unintended effects can have severe consequences.[68] The exact framework for a responsible use is currently under discussion.[69–71] For instance, caregivers have to be trained in using the provided results, patients' access to care has to remain equitable, real-world performance must be constantly scrutinised, and responsibilities in case of errors have to be clear.

Furthermore, the effective implementation in clinical practice requires more than high prediction performance. Improved patient outcomes depend on how predictions are translated into effective decision making.[72] This translation requires the predictions to be reasonably integrated in existing clinical processes in order to be accepted by medical staff and to create an actual benefit from better informed decisions.[73]

Present study in comparison to previous research
Previous studies found similar results considering the association between patient characteristics and the prediction of polypharmacy and DDI. Abolhassani et al studied 17 742 adult patients discharged between 2009 and 2015 from a department of internal medicine at s Swiss hospital.[74] They found age, number of comorbidities and a higher Charlson Comorbidity Index independently associated with polypharmacy. Pérez et al studied 38 299 patients in 44 general practices in Ireland between 2012 and 2015 and found that age and multimorbidity were associated with a higher risk of DDI.[75] Furthermore, they found that

hospital admissions themselves were independently associated with a higher risk of DDI.

These studies did not investigate psychiatric hospital care. To the best of our knowledge, there is currently no evidence comparable in scale and scope investigating the prediction of polypharmacy and the risk of DDI in hospital psychiatry. Predicting other outcomes in hospitalised patients has often been found to be more complex in psychiatry than in other medical disciplines.[76–78] Less distinct diagnostic concepts,[79–81] less standardisation of care[82] and a broader spectrum of acceptable therapeutic regimes[83] were reasons put forward for this.

Several previous studies aimed to use patient and service data to predict outcomes of psychiatric hospital stays. These studies have often used a broad range of feature variables. Moreover, studies were often restricted to specific settings and patients. Leigthon et al[84] attempted at predicting remission after 12 months in 79 patients with a first episode of psychosis applying a wide range of psychometric, demographic and socioeconomic feature variables and reached an area under the ROC of 0.65. Koutsouleris et al[85] reached a sensitivity of 71%, a specificity of 72% and a precision of 93% in 108 unseen patients with their top 10 demographic, socioeconomic and psychometric features to predict remission in first episode of psychosis. Lin et al[86] used single nucleotide polymorphisms from genetic analyses and clinical data and reached an area under the ROC of 0.82 in distinguishing responders from non-responders prior to antidepressant therapy in 455 patients with major depression.

Strengths and weaknesses of our study
A strength of this study is its detailed longitudinal prescription data. This allowed for delineating the daily prescription patterns for each episode and identifying polypharmacy and DDI. A further strength of this study is its broad coverage of psychiatric hospital care by including all inpatients admitted to eight psychiatric hospitals for 2 years. Furthermore, the study used a rather restricted set of features which should already be routinely available or easily possible to implement in many hospitals. Moreover, we were able to reduce the risk of information leakage and overfitting by testing the prediction performance of our models in patients that were treated in another calendar year and by separating the study sites that were used for model training from the study sites that were used for testing the prediction performance.

The present study did not include all potential types of pharmacokinetic and pharmacodynamic DDI that could occur under psychotropic medication. However, we have covered main types of DDI in hospital psychiatry, and our clinical prediction models will bring about a tool to support the efficient management of benefits and major risks of hospital prescriptions.

Our study did not delineate patient-specific benefit–risk balances of prescriptions, for instance by including drug serum levels, results of electrocardiograms or

individual pharmacogenetic risk factors. Therefore, it was not possible to differentiate between DDI and actually inadequate prescriptions. Neither did our study document actual ADR. DDI do not necessarily lead to ADR and negative patient outcomes. However, the association between undesired DDI and an increased risk of negative patient outcomes has been thoroughly established by previous studies.[10 24–30] Indeed, the association between DDI and actual negative patient outcomes might often be underestimated in psychiatry and therefore neglected in clinical practice.[87]

Another limitation of our study was the lack of time stamps on feature data. Patients were diagnosed and rated on clinical severity scales at admission and these values should mainly remained stable. However, we were not able to entirely rule out that these groupings might have been changed during the stay by clinical staff. A further general limitation of our study was its restriction to study sites from one large provider of inpatient psychiatric services in the region of Hesse, Germany. This obviously raises the question whether the prediction performance of the present models would remain stable if used in different healthcare systems or different clinical setting.

Conclusion

This study has shown that polypharmacy and DDI at a psychiatric hospital can be predicted from routine data at patient admission. These predictions could support an efficient management of benefits and risks of hospital prescriptions, for instance by including pharmaceutical supervision early after admission for patients at risk before pharmacological treatment is established. Future studies should investigate the clinical impact of such risk models on drug safety processes and patient outcomes in hospital psychiatry.

Author affiliations
[1]Peter L. Reichertz Institute for Medical Informatics, Hannover Medical School, Hannover, Germany
[2]Department of Psychiatry and Psychotherapy, Medical Center-University of Freiburg, Freiburg, Germany
[3]Vitos Clinic for Forensic Psychiatry, Vitos Rheingau, Eltville, Germany
[4]Institute for Medical Biometry and Statistics, Medical Center-University of Freiburg, Freiburg, Germany
[5]Department of Psychiatry and Psychotherapy, University Medical Center Mainz, Mainz, Germany
[6]Waldkrankenhaus Köppern, Vitos Hospital Hochtaunus, Friedrichsdorf, Germany
[7]Department of Psychiatry and Psychotherapy, Heinrich Heine University Düsseldorf, Düsseldorf, Germany

ORCID iDs
Jan Wolff http://orcid.org/0000-0003-2750-0606
Klaus Kaier http://orcid.org/0000-0003-0837-6945

REFERENCES

1 NICE. *Medicines optimisation: the safe and effective use of medicines to enable the best possible outcomes. NICE guideline NG5*, 2015.
2 Taylor M, Perera U. NICE CG178 psychosis and schizophrenia in adults: treatment and management - an evidence-based guideline? *Br J Psychiatry* 2015;206:357–9.
3 Bolea-Alamañac B, Nutt DJ, Adamou M, *et al*. Evidence-based guidelines for the pharmacological management of attention deficit hyperactivity disorder: update on recommendations from the British association for psychopharmacology. *J Psychopharmacol* 2014;28:179–203.
4 Goldberg SG, Wagner K. American psychological association practice guidelines for psychopharmacology: ethical practice considerations for psychologists involving psychotropic use with children and adolescents. *J Clin Psychol* 2019;75:344–63.
5 Goodwin GM, Young AH. The British association for psychopharmacology guidelines for treatment of bipolar disorder: a summary. *J Psychopharmacol* 2003;17:3–6.
6 Baldwin DS, Anderson IM, Nutt DJ, *et al*. Evidence-based guidelines for the pharmacological treatment of anxiety disorders: recommendations from the British association for psychopharmacology. *J Psychopharmacol* 2005;19:567–96.
7 Hiemke C, Shams M. Phenotyping and genotyping of drug metabolism to guide pharmacotherapy in psychiatry. *Curr Drug Deliv* 2013;10:46–53.
8 Preskorn SH, Lacey RL. Polypharmacy: when is it rational? *J Psychiatr Pract* 2007;13:97–105.
9 Kratz T, Diefenbacher A. Psychopharmacological treatment in older people. *Dtsch Arztebl Int* 2019;116:508–18.
10 Leone R, Magro L, Moretti U, *et al*. Identifying adverse drug reactions associated with drug-drug interactions: data mining of a spontaneous reporting database in Italy. *Drug Saf* 2010;33:667–75.
11 Masnoon N, Shakib S, Kalisch-Ellett L, *et al*. What is polypharmacy? A systematic review of definitions. *BMC Geriatr* 2017;17:230.

12 Gnjidic D, Hilmer SN, Blyth FM, *et al*. Polypharmacy cutoff and outcomes: five or more medicines were used to identify community-dwelling older men at risk of different adverse outcomes. *J Clin Epidemiol* 2012;65:989–95.

13 Mojtabai R, Olfson M. National trends in psychotropic medication polypharmacy in office-based psychiatry. *Arch Gen Psychiatry* 2010;67:26.

14 Glezer A, Byatt N, Cook R, *et al*. Polypharmacy prevalence rates in the treatment of unipolar depression in an outpatient clinic. *J Affect Disord* 2009;117:18–23.

15 McIntyre RS, Jerrell JM. Polypharmacy in children and adolescents treated for major depressive disorder: a claims database study. *J Clin Psychiatry* 2009;70:240–6.

16 Carmona-Huerta J, Castiello-de Obeso S, Ramírez-Palomino J, *et al*. Polypharmacy in a hospitalized psychiatric population: risk estimation and damage quantification. *BMC Psychiatry* 2019;19:78.

17 Möller H-J, Seemüller F, Schennach-Wolff R. History, background, concepts and current use of comedication and polypharmacy in psychiatry. *Int J Neuropsychopharm* 2014;17:983–96.

18 Valuck RJ, Morrato EH, Dodd S, *et al*. How expensive is antipsychotic polypharmacy? Experience from five US state Medicaid programs. *Curr Med Res Opin* 2007;23:2567–76.

19 Molokhia M, Majeed A. Current and future perspectives on the management of polypharmacy. *BMC Fam Pract* 2017;18:70.

20 Palleria C, Di Paolo A, Giofrè C, *et al*. Pharmacokinetic drug-drug interaction and their implication in clinical management. *J Res Med Sci* 2013;18:601–10.

21 Zanger UM, Schwab M. Cytochrome P450 enzymes in drug metabolism: regulation of gene expression, enzyme activities, and impact of genetic variation. *Pharmacol Ther* 2013;138:103–41.

22 Cascorbi I. Drug interactions—principles, examples and clinical consequences. *Dtsch Arztebl Int* 2012;109:546–55.

23 National Institute of Diabetes and Digestive and Kidney Diseases. Anticholinergic agents. In: *LiverTox: clinical and research information on drug-induced liver injury*. Bethesda, MD: National Institute of Diabetes and Digestive and Kidney Diseases, 2012.

24 Mintzer J, Burns A. Anticholinergic side-effects of drugs in elderly people. *J R Soc Med* 2000;93:457–62.

25 Kalisch Ellett LM, Pratt NL, Ramsay EN, *et al*. Multiple anticholinergic medication use and risk of hospital admission for confusion or dementia. *J Am Geriatr Soc* 2014;62:1916–22.

26 Lupu AM, Clinebell K, Gannon JM, *et al*. Reducing anticholinergic medication burden in patients with psychotic or bipolar disorders. *J Clin Psychiatry* 2017;78:e1270–5.

27 Pasquier M, Pantet O, Hugli O, *et al*. Prevalence and determinants of QT interval prolongation in medical inpatients. *Intern Med J* 2012;42:933–40.

28 Nosè M, Bighelli I, Castellazzi M, *et al*. Prevalence and correlates of QTc prolongation in Italian psychiatric care: cross-sectional multicentre study. *Epidemiol Psychiatr Sci* 2016;25:532–40.

29 Beach SR, Celano CM, Noseworthy PA, *et al*. QTc prolongation, torsades de pointes, and psychotropic medications. *Psychosomatics* 2013;54:1–13.

30 Hasnain M, Vieweg WVR. QTc interval prolongation and torsade de pointes associated with second-generation antipsychotics and antidepressants: a comprehensive review. *CNS Drugs* 2014;28:887–920.

31 Pirnejad H, Amiri P, Niazkhani Z, *et al*. Preventing potential drug-drug interactions through alerting decision support systems: a clinical context based methodology. *Int J Med Inform* 2019;127:18–26.

32 Bayati S, Bastani P, Sagheb ZM, *et al*. The performance implications of pharmacy information system at the University teaching hospitals of Shiraz, Iran: cluster approach. *J Adv Pharm Technol Res* 2017;8:125.

33 Alanazi A, Al Rabiah F, Gadi H, *et al*. Factors influencing pharmacists' intentions to use Pharmacy Information Systems. *Inform Med Unlocked* 2018;11:1–8.

34 Azaz-Livshits T, Levy M, Sadan B, *et al*. Computerized surveillance of adverse drug reactions in hospital: pilot study. *Br J Clin Pharmacol* 1998;45:309–14.

35 Bobb A, Gleason K, Husch M, *et al*. The epidemiology of prescribing errors: the potential impact of computerized prescriber order entry. *Arch Intern Med* 2004;164:785.

36 Kaboli PJ, Hoth AB, McClimon BJ, *et al*. Clinical pharmacists and inpatient medical care: a systematic review. *Arch Intern Med* 2006;166:955.

37 Drovandi A, Robertson K, Tucker M, *et al*. A systematic review of clinical pharmacist interventions in paediatric hospital patients. *Eur J Pediatr* 2018;177:1139–48.

38 Hiemke C, Bergemann N, Clement HW, *et al*. Consensus guidelines for therapeutic drug monitoring in Neuropsychopharmacology: update 2017. *Pharmacopsychiatry* 2018;51:9–62.

39 US Food and Drug Admistration. Drug Development and Drug Interactions. Table of substrates, inhibitors and inducers, 2020. Available: https://www.fda.gov/drugs/drug-interactions-labeling/drug-development-and-drug-interactions-table-substrates-inhibitors-and-inducers

40 Grözinger M, Dragicevic A, Hiemke C, *et al*. Melperone is an inhibitor of the CYP2D6 catalyzed O-demethylation of venlafaxine. *Pharmacopsychiatry* 2003;36:3–6.

41 Mannheimer B, von Bahr C, Pettersson H, *et al*. Impact of multiple inhibitors or substrates of cytochrome P450 2D6 on plasma risperidone levels in patients on polypharmacy. *Ther Drug Monit* 2008;30:565–9.

42 Wójcikowski J, Pichard-Garcia L, Maurel P, *et al*. Perazine as a potent inhibitor of human CYP1A2 but not CYP3A4. *Pol J Pharmacol* 2002;54:407–10.

43 Schaller G, Jacobi A, Rotter A, *et al*. Perazine elevates clozapine serum levels by inhibiting hepatic metabolism. *Prog Neuropsychopharmacol Biol Psychiatry* 2009;33:908–9.

44 Fischer M, Unterecker S, Deckert J, *et al*. Elevated clozapine serum levels in combination with perazine. *Psychopharmacology* 2013;226:623–5.

45 Hiemke CC, Eckermann G. Kombinationstherapie/Polypharmazie - Interaktionen von Psychopharmaka. *Arzneimitteltherapie* 2014;32:361–70.

46 Woosley R, Romero K. QT drugs list. AZCERT, Inc, 2020. Available: https://crediblemeds.org/

47 Romero RL, Woosley RL. Clarification to the www.qtdrugs.org updated lists. *Pharmacoepidemiol Drug Saf* 2009;18:423–4.

48 Chew ML, Mulsant BH, Pollock BG, *et al*. Anticholinergic activity of 107 medications commonly used by older adults. *J Am Geriatr Soc* 2008;56:1333–41.

49 Lertxundi U, Domingo-Echaburu S, Hernandez R. Expert-based drug lists to measure anticholinergic burden: similar names, different results: scores to measure anticholinergic burden. *Psychogeriatrics* 2013;13:17–24.

50 Jones SH, Thornicroft G, Coffey M, *et al*. A brief mental health outcome scale-reliability and validity of the global assessment of functioning (GAF). *Br J Psychiatry* 1995;166:654–9.

51 Busner J, Targum SD. The clinical global impressions scale. *Psychiatry* 2007;4:28–37.

52 Beck AT, Steer RA, Carbin MG. Psychometric properties of the beck depression inventory: twenty-five years of evaluation. *Clin Psychol Rev* 1988;8:77–100.

53 Lin C-H, Lin H-S, Lin S-C, *et al*. Early improvement in PANSS-30, PANSS-8, and PANSS-6 scores predicts ultimate response and remission during acute treatment of schizophrenia. *Acta Psychiatr Scand* 2018;137:98–108.

54 Friedman JH. Greedy function approximation: a gradient boosting machine. *Ann Stat* 2001;29:1189–232.

55 Kuhn M. Building predictive models in R using the CARET package. *J Stat Softw* 2008;28:1–26.

56 R Core Team. R: a language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria, 2018. Available: https://www.R-project.org/

57 DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988;44:837–45.

58 Tomašev N, Glorot X, Rae JW, *et al*. A clinically applicable approach to continuous prediction of future acute kidney injury. *Nature* 2019;572:116–9.

59 The Lancet Respiratory Medicine. Opening the black box of machine learning. *Lancet Respir Med* 2018;6:801.

60 Zhang Z, Beck MW, Winkler DA, *et al*. Opening the black box of neural networks: methods for interpreting neural network models in clinical applications. *Ann Transl Med* 2018;6:216.

61 Molnar C. Interpretable machine learning - a guide for making black box models explainable, 2020. Available: https://christophm.github.io/interpretable-ml-book/

62 Apley DW, Zhu J. Visualizing the effects of predictor variables in black box supervised learning models. arXiv:161208468 [stat]. Available: http://arxiv.org/abs/1612.08468

63 Saito T, Rehmsmeier M. The Precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS One* 2015;10.

64 Hosmer DW, Lemeshow S, Sturdivant RX. *Applied logistic regression*. Hoboken, NJ: Wiley InterScience, Wiley, 2013.

65 Forgerini M, Schiavo G, Lucchetta RC, *et al*. Drug interactions for elderly people with mental and behavioral disorders: a systematic scoping review. *Arch Gerontol Geriatr* 2021;93:104283.

66 Emanuel EJ, Wachter RM. Artificial intelligence in health care: will the value match the hype? *JAMA* 2019;321:2281–2.

67 Shah NH, Milstein A, Bagley PhD SC. Making machine learning models clinically useful. *JAMA* 2019;322:1351–2.

68 Obermeyer Z, Powers B, Vogeli C, *et al*. Dissecting racial bias in an algorithm used to manage the health of populations. *Science* 2019;366:447–53.

69 Char DS, Shah NH, Magnus D. Implementing machine learning in health care - addressing ethical challenges. *N Engl J Med* 2018;378:981–3.

70 Vayena E, Blasimme A, Cohen IG. Machine learning in medicine: addressing ethical challenges. *PLoS Med* 2018;15:e1002689.

71 Nebeker C, Torous J, Bartlett Ellis RJ. Building the case for actionable ethics in digital health research supported by artificial intelligence. *BMC Med* 2019;17:137.

72 Altman DG, Vergouwe Y, Royston P, *et al*. Prognosis and prognostic research: validating a prognostic model. *BMJ* 2009;338:b605.

73 Lindsell CJ, Stead WW, Johnson KB. Action-informed artificial intelligence-matching the algorithm to the problem. *JAMA* 2020;323:2141.

74 Abolhassani N, Castioni J, Santschi V. Trends and determinants of polypharmacy and potential drug-drug interactions at discharge from hospital between 2009–2015. *J Patient Saf* 2018;1.

75 Pérez T, Moriarty F, Wallace E, *et al*. Prevalence of potentially inappropriate prescribing in older people in primary care and its association with hospital admission: longitudinal study. *BMJ* 2018;363:k4524.

76 Wolff J, McCrone P, Koeser L, *et al*. Cost drivers of inpatient mental health care: a systematic review. *Epidemiol Psychiatr Sci* 2015;24:78–89.

77 Barry CL, Weiner JP, Lemke K, *et al*. Risk adjustment in health insurance exchanges for individuals with mental illness. *Am J Psychiatry* 2012;169:704–9.

78 Montz E, Layton T, Busch AB, *et al*. Risk-adjustment simulation: plans may have incentives to distort mental health and substance use coverage. *Health Aff* 2016;35:1022–8.

79 Wakefield JC. The concept of mental disorder: diagnostic implications of the harmful dysfunction analysis. *World Psychiatry* 2007;6:149–56.

80 Aboraya A, Rankin E, France C. The reliability of psychiatric diagnosis revisited. *Psychiatry* 2006;3:41–50.

81 Jablensky A. Psychiatric classifications: validity and utility. *World Psychiatry* 2016;15:26–31.

82 Evans-Lacko SE, Jarrett M, McCrone P, *et al*. Clinical pathways in psychiatry. *Br J Psychiatry* 2008;193:4–5.

83 Barbui C, Tansella M. Guideline implementation in mental health: current status and future goals. *Epidemiol Psychiatr Sci* 2012;21:227–9.

84 Leighton SP, Krishnadas R, Chung K, *et al*. Predicting one-year outcome in first episode psychosis using machine learning. *PLoS One* 2019;14:e0212846.

85 Koutsouleris N, Kahn RS, Chekroud AM, *et al*. Multisite prediction of 4-week and 52-week treatment outcomes in patients with first-episode psychosis: a machine learning approach. *Lancet Psychiatry* 2016;3:935–46.

86 Lin E, Kuo P-H, Liu Y-L, *et al*. A deep learning approach for predicting antidepressant response in major depression using clinical and genetic biomarkers. *Front Psychiatry* 2018;9:290.

87 de Leon J. Why do you keep telling me that drug-drug interactions are important in psychopharmacology when I do not see them in my clinical practice?: my failure to Convince clinicians. *J Clin Psychopharmacol* 2019;39:1–4.