Psychiatry Research 326 (2023) 115328

Contents lists available at ScienceDirect

Psychiatry Research

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

Outcome prediction of electroconvulsive therapy for depression

Yuri van der Does [a,*], Rosanne J. Turner [a,b], Miel J.H. Bartels [a], Karin Hagoort [a], Aäron Metselaar [a], Floortje Scheepers [a], Peter D. Grünwald [b,c], Metten Somers [a], Edwin van Dellen [a,d]

[a] Department of Psychiatry, UMC Utrecht Brain Center, University Medical Center Utrecht, Utrecht University, 3584 CX Utrecht, the Netherlands
[b] Machine Learning Group, CWI (national research institute for mathematics and computer science), Amsterdam, the Netherlands
[c] Department of Mathematics, Leiden University, Leiden, Netherlands
[d] Department of Neurology, UZ Brussel and Vrije Universiteit Brussel, Brussels, Belgium

ARTICLE INFO

Keywords:
Affective disorders
Treatment
Shared decision-making
Bayesian statistics
Systematic literature search
Hospital psychiatry
Remission

ABSTRACT

Introduction: We developed and tested a Bayesian network(BN) model to predict ECT remission for depression, with non-response as a secondary outcome.
Methods: We performed a systematic literature search on clinically available predictors. We combined these predictors with variables from a dataset of clinical ECT trajectories (performed in the University Medical Center Utrecht) to create priors and train the BN. Temporal validation was performed in an independent sample.
Results: The systematic literature search yielded three meta-analyses, which provided prior knowledge on outcome predictors. The clinical dataset consisted of 248 treatment trajectories in the training set and 44 trajectories in the test set at the same medical center. The AUC for the primary outcome remission estimated on an independent validation set was 0.686 (95%CI 0.513–0.859) (AUC values of 0.505 – 0.763 observed in 5-fold cross validation of the model within the train set). Accuracy 0.73 (balanced accuracy 0.67), sensitivity 0.55, specificity 0.79, after temporal validation in the independent sample. Prior literature information marginally reduced CI width.
Discussion: A BN model comprised of prior knowledge and clinical data can predict remission of depression after ECT with reasonable performance. This approach can be used to make outcome predictions in psychiatry, and offers a methodological framework to weigh additional information, such as patient characteristics, symptoms and biomarkers. In time, it may be used to improve shared decision-making in clinical practice.

1. Introduction

Depression is a leading cause of disability according to the World Health Organization, affecting one in six people during their lifetime (Kessler et al., 2005; WHO, 2022). Electroconvulsive therapy (ECT) is the most effective available treatment for severe depression (Lisanby, 2007). In practice, ECT is usually reserved for patients who show insufficient response to antidepressant medications and psychotherapy, in part because of stigma and anticipated cognitive side effects (Leiknes et al., 2012). Although highly effective on a group level, a substantial number of patients show no or insufficient response to ECT. There are several factors associated with response to ECT, including age and presence of psychotic symptoms (van Diermen et al., 2018). However, in current psychiatric practice, neither systematic assessment of these independent predictors, nor assessment of cumulative predictive value of multiple predictors are routinely used in the decision to initiate ECT for individual patients. As a result, treatment outcome on the individual level remains largely unpredictable.

Clinical decision support systems (CDSSs) are computerized tools which provide clinicians individualized information based on various sources of information, for instance demographic characteristics and information from electronic health records (EHRs). CDSSs make use of prediction models or algorithms for systematic assessment of information. CDSSs are used in several clinical specialties, such as in cardiovascular medicine (collaboration SwgaECr 2021; Hageman et al., 2022). In psychiatry, the availability of CDSSs is modest at best, as was illustrated by Koposov and colleagues (Koposov et al., 2017). Bright and colleagues give an overview of clinically implemented CDSSs across all

* Corresponding author at: Postal address: UMC Utrecht, Afdeling psychiatrie, polikliniek A1, Postbus 85500, 3508 GA Utrecht, the Netherlands.
E-mail address: y.vanderdoes@umcutrecht.nl (Y. van der Does).

https://doi.org/10.1016/j.psychres.2023.115328
Received 15 February 2023; Received in revised form 26 June 2023; Accepted 30 June 2023
Available online 2 July 2023

medical specialties in a systematic review and meta-analysis of randomized controlled trials. They found that clinicians are more likely to appoint the appropriate treatment when informed by CDSSs compared to clinicians who did not use these systems, based on 46 studies across diverse venues and systems (OR 1.57, 95%CI 1.35 – 1.82) (Bright et al., 2012). A recent Cochrane review by Stacey and colleagues, which assessed the effect of decision aids, reported that patients who used CDSSs were better informed on treatment options, felt more knowledgeable, and were likely to have more accurate risk perceptions (Stacey et al., 2017). A CDSS which can predict the effect of ECT for individual patients could be useful to inform patients and facilitate shared decision making before treatment is initiated. In order to realize this, a prediction model for ECT outcome is required.

In this study, we developed a personalized effect prediction model for the prediction of remission after ECT, and, secondary, ECT non-response, using a Bayesian network (BN) model. BNs are a combination of intuitive graphical representations of causal or predictive dependencies between variables, and the corresponding underlying quantitative model (for an insightful tutorial aimed at psychopathology researchers see (Briganti et al., 2022)). The presence and underlying quantitative model of these dependencies can be derived from data, can be obtained from expert opinion, or both (Arora et al., 2019). The aim was to predict the effect of an ECT trajectory using data which was clinically available before ECT was initiated. We used a systematic review to identify predictors of ECT outcome to inform a BN with prior knowledge from literature. Subsequently, we used expert knowledge to further design the BN, and tested its performance in clinical data. Finally, we validated the performance of this prediction model in a validation dataset.

2. Methods

We used a stepwise approach to create the BN model for ECT outcome prediction: 1) acquiring prior knowledge by performing a literature search for clinically obtainable predictors; 2) Creation and training of a BN using prior knowledge from literature and a clinical dataset; and 3) validation of the trained model in a validation cohort.

2.1. acquiring prior knowledge

2.1.1. Systematic review

To acquire high quality prior knowledge from literature, we performed a systematic review on predictors of ECT outcome, in which we only searched for high quality meta-analyses. We performed a literature search in the online libraries MEDLINE and EMBASE according to the PRISMA guidelines for systematic reviews (Page et al., 2021). The protocol of this review was not registered in advance and is not available for review as such. The question this systematic review addressed was formulated as: “for adult patients undergoing ECT, what are clinical predictors for outcome (response or remission) of ECT”. Search terms used were: (ect OR electroconv*) AND predict* AND (remission OR respons* OR outcome*); all published articles available before 16–11–2022 were reviewed. Articles were excluded in screening of title and abstract if they were: non-human, non-English language, when treatment was not ECT and if study design was not a meta-analysis. Systematic reviews without meta-analysis were excluded. When the full-text studies were not available, the authors were contacted. Eligibility criteria for inclusion were studies on predictors of ECT outcomes of which data were readily available at baseline in most patients. These were defined as demographic predictors, clinical assessment predictors, comorbidity predictors, pharmacological or technical ECT aspects predictors. The definition did not include MRI findings predictors, because MRI scans are not performed as standard practice at the start of ECT.

The screening and quality assessment of articles were performed by two independent reviewers (YD and AM), without automation tools. Discrepancies in results were resolved by consensus, or by a third reviewer in case of disagreement (ED). We performed a ROBIS quality assessment to assess the risk of bias in the identified meta-analyses (Whiting et al., 2016). Only studies with an overall low risk of bias were included in model development.

Data were collected from each individual predictor for both remission and response. Standardized mean differences with 95% confidence intervals (95%CI) were collected for continuous predictors, odds ratios (ORs) with 95%CI were collected for dichotomous predictors. Outcome was defined as “remission” or “response”, without further specification, in order to include all relevant studies. When two or more meta-analyses provided data for a single predictor and outcome, the authors would, after consensus, only extract data from the most relevant meta-analysis available, based on date of publication and quality assessment. Results of the data extraction were used as priors for the BN model.

2.2. Bayesian network model development

2.2.1. Study population

For the BN model, we used individual patient data from patients who were treated with ECT in the University Medical Center Utrecht (UMCU) in the Netherlands between 1 January 2008 and 27 September 2019. We included all patients receiving ECT for a depressive episode, including patients with bipolar and schizoaffective disorders, who had a discharge letter with conclusion of ECT trajectory available. When patients had multiple ECT trajectories, a subsequent trajectory was only included after a time interval of at least 90 days. We retrospectively acquired as many patient characteristics as are routinely collected in standard clinical practice and used these as predictors for the outcomes. These were age, sex, somatic comorbidity, age of onset of symptoms, duration of depressive episode, ECT naivety, co-morbid personality disorder, severity of depression, psychotic features, catatonic features, and diagnostic context of ECT indication (e.g. depressive disorder, bipolar disorder, or schizoaffective disorder). Clinical patient characteristics were extracted from Electronic Health Records (EHRs). Patient data were anonymized using DEDUCE and therefore the institutional medical ethics review board waived informed consent (Menger et al., 2018). Validation of the model was performed in a cohort which consisted of patients who received ECT in the UMCU between 17 July 2018 and 22 October 2021.

2.2.2. Outcomes

The primary outcome was remission after ECT. Remission outcome was assessed by deriving the conclusion from the psychiatrist’s discharge letter, stating “remission”, indicating an absence of depressive symptoms. This dichotomous outcome has been used in meta-analytic research and has clinical usefulness, because it is informative and understandable for both clinicians and patients (Pagnin et al., 2004).

Non-response was assessed as secondary outcome, and was defined as absence of any amelioration of symptoms of depression. This was also assessed by deriving the conclusion from the discharge letter.

2.2.3. Statistical analysis

To explore the data used for training the model, group means or proportions for the predictor variables were compared between remission and non-remission using the appropriate hypothesis tests, where we used Bonferroni correction to correct for multiple testing. In case of missing data we used multiple imputation, using IterativeImputer (Python) for the training set and MICE for the validation set (R).

To gain insight into the associations and/or causal relations between predictors on multiple levels, a BN was fitted on the UMCU data with the “bnlearn” package in R (Scutari, 2010). Prior to learning the structure of the network, black- and whitelists were created based on the data derived from the meta-analyses combined with expert knowledge from authors YD, MS and ED. Associations on these lists were either by default included (for the whitelist)) in, or excluded (for the blacklist) from the network. Adding this sort of prior knowledge can vastly improve the

stability of overparameterized networks (Briganti et al., 2022). On the blacklist, response or remission was excluded as a predictor of other variables in the network, and age and gender were excluded as being dependent on other variables in the network. Somatic comorbidity and cognitive disorder were also excluded as possible direct predictors of non-response or remission. On the whitelist, all predictors except catatonic symptoms, forced care and first ECT trajectory were included as direct predictors of response or remission. Personality disorder was included as a direct predictor of somatic comorbidity, age of onset, relapse, episode duration and psychotic and catatonic symptoms. The structure of the BN was determined through applying the score-based (i.e., aimed at optimizing the predictive performance of the network) "Hill-Climbing" algorithm on the data 100 times through bootstrapping: to improve stability, only dependencies occurring in at least 85% of the bootstrapped networks were included in the final structure (Briganti et al., 2022).

Based on the dependencies found in the BN a hierarchical Bayesian logistical regression model was fitted specifically for predicting response to ECT with the "arm" package in R, as the bnlearn package did not offer fitting such models with prior information. As predictors, all variables found to be associated with response to ECT from the meta-analyses and all variables included as predictors of the outcome variable in the BN were included. Priors on coefficients were chosen to be normal, with mean and standard deviation either estimated through the meta-analysis, or set to 0 and 1 in the absence of prior information; all prior settings can be found in Table 2.

To generate insights into model performance, 5-fold cross-validation was performed and mean accuracy, ROC-AUC (receiver operator characteristics area under the curve) and corresponding 95% confidence intervals were calculated. ROC-AUC can be interpreted as the ability of the predictor to distinguish between true positive and negative cases (or probability that it will do so correctly). The model was subsequently validated in the independent dataset, where ROC-AUC curves, sensitivity, specificity and calibration curves were calculated to assess the external validity of the model. The model creation and validation were in accordance with the TRIPOD statement (Collins et al., 2015). For a comparison between BN and other statistical approaches, we used the R package "caret" to train and evaluate other machine learning techniques: regression (penalized and unpenalized), random forest and boosting (glm, glmnet,rf, xgbTree). We used 10-fold cross validation with an oversampling within the cross validation, according to ROSE sampling. Results were reported in the material.

3. Results

3.1. Acquiring prior knowledge

3.1.1. Systematic review

We found a total of 1638 articles after removing duplicates. Of these, 1614 were excluded after screening of title and abstract. A total of 24 articles were sought for retrieval, of which six were eventually not available. Full text screening for eligibility was performed in 18 articles, of which five were included (van Diermen et al., 2018; Baldwin and Oxlad, 1996; Haq et al., 2015; Havaki-Kontaxaki et al., 2006; Kho et al., 2003). One additional article was found while screening manually for relevant references in the included articles (Heijnen et al., 2010). Flowchart and quality assessment summary are reported in the supplementary material. We included three meta-analyses with an overall low risk of bias (van Diermen et al., 2018; Haq et al., 2015; Heijnen et al., 2010). All meta-analyses showed overlap of included studies. For the predictors psychotic symptoms, age, melancholic symptoms and depression severity, two studies reported data on response outcome (van Diermen et al., 2018; Haq et al., 2015). For these predictors, we only extracted data from the meta-analysis of Van Diermen and colleagues (van Diermen et al., 2018), because this meta-analysis was more recent and was assessed as having an overall lower risk of bias in the ROBIS quality assessment. For the predictor medication failure, data for remission outcome from Heijnen and others (Baldwin and Oxlad, 1996) and data for response outcome was extracted from the meta-analysis from Haq and others (Pagnin et al., 2004). Extracted data of predictors of ECT outcomes were reported in supplementary material.

3.2. Bayesian network model development

3.2.1. Training and validation datasets

We included a total of 248 treatment trajectories of patients receiving ECT at the UMCU between 2009 and 2019 in the training dataset, of which 90 (36%) were classified as remission and 63 (25%) as non-response. The validation set consisted of 44 independent treatment trajectories, of which 11 (25%) were classified as remission and 5 (9%) as non-response. Summary statistics (mean values or proportions for patients with and without remission) of both datasets can be found in Table 1. In the training data, nine treatment trajectories had an unknown episode duration. In the independent validation cohort, 23 cases had missing data, for the relapse, episode duration, catatonic symptoms and age of onset predictors. All missing variables were imputed. A total of 32 treatments had missing values for their number of medication switches in the current episode and nine treatments had an unknown episode duration. If any other clinical parameter was missing, the treatment trajectory was excluded from analysis. Age was the only predictor with statistically significant differences after Bonferroni correction between non-responders and patients with response/remission, with higher age being associated with higher remission rate, in both the training set and in the validation set ($p = 0.0005$ and $p = 0.0034$ respectively).

3.2.2. Bayesian network model and hierarchical model for predicting remission

The Bayesian network found with the Hill-Climbing algorithms revealed no new direct dependencies between predictor variables and outcome variable remission that were not already present on the whitelist provided by the experts and meta-analysis (supplementary figure 1.)

The model containing solely priors from literature had an AUC of 0.63 (95% CI 0.56 – 0.70) and an accuracy of 0.63 (balanced accuracy: 0.60) for predicting remission of UMCU patients in the training set. After updating the model coefficients using the data of UMCU patients, the AUC was 0.629 (values 0.505 – 0.763 observed in 5-fold cross validation) and the classification accuracy estimated through 5-fold cross-validation was 0.637 (balanced accuracy: 0.569). The trained hierarchical Bayesian logistic regression model and an overview of priors can be found in Table 2. For completeness, a model containing only patient derived data with no prior information, showed a mean AUC of 0.59 (values 0.53 – 0.82 observed in 5-fold cross-validation) and a mean accuracy for remission of 0.66 (values 0.60 – 0.81 observed in 5-fold cross-validation, balanced accuracy: 0.602, values 0.52 – 0.72).

Validation of the updated model predicting remission on the 44 patients in the validation set resulted in an AUC of 0.686 (95%CI 0.513–0.859), with an accuracy of 0.73 (balanced accuracy: 0.67). Remission occurred in 11 patients: there were 26 true negatives (59.1% of the validation set), 6 true positives (13.6% of the validation set), 5 false negatives (11.4% of the validation set) and 7 false positives (15.9% of the validation set). The corresponding sensitivity of the model assessed on the validation set was 0.55 and the specificity 0.79. A model without prior information resulted in an AUC of 0.686 (95%CI 0.491–0.881) and an accuracy of 0.75 (balanced accuracy: 0.66).

An overview of data of misclassified cases of remission is given below in Table 3. False negative cases (patients predicted as not achieving remission after ECT while in reality they did), were generally younger, without psychotic symptoms. False positive cases were generally older, with psychotic symptoms, and did not have personality disorders, which were strong predictors in the final model for remission (see Table 2).

Table 1
Summary statistics. Mean values (continuous variables) or proportions (categorical variables) of predictors of patients with the corresponding property (dichotomous variables) and p-values based on a *t*-test for continuous data or Fisher's exact test for dichotomous data of patients with or without remission in the train and test data. Missing variables were excluded column-wise. For dichotomous variables, counts of patients with the corresponding property are included between brackets. For the test set, non-imputed data are shown.

Summary statistics						
	Training set ($n = 248$)			Validation set ($n = 44$)		
	Mean No remission ($n = 158$)	p-value Remission ($n = 90$)		Mean p-value No Remission ($n = 33$)	Remission ($n = 11$)	
Relapse (y/n)	0.80 (126)	0.84 (76)	0.40	0.867 (Gross et al., 2018)	1 (Stacey et al., 2017)	0.556
Episode duration (months)	24.5	18.0	0.039	27.9	17.0	0.214
Age (years)	48.5	55.9	0.000527	49.5	68.5	0.0034
First ECT (y/n)	0.772 (122)	0.733 (66)	0.538	0.733 (Kho et al., 2003)	0.778 (Hageman et al., 2022)	1
Psychotic Symptoms(y/n)	0.241 (Montgomery and Asberg, 1979)	0.333 (Yip et al., 2021)	0.139	0.355 (Briganti et al., 2022)	0.400 (Leiknes et al., 2012)	1
Catatonic symptoms (y/n)	0.070 (Briganti et al., 2022)	0.100 (Bright et al., 2012)	0.469	0.097 (Lisanby, 2007)	0.000 (0)	0.564
Severe depressive Episode (y/n)	0.791 (125)	0.888 (79)	0.119	0.364 (Arora et al., 2019)	0.273 (Lisanby, 2007)	0.722
Forced care (y/n)	0.045 (Hageman et al., 2022)	0.033 (Lisanby, 2007)	0.751	0.152 (van Diermen et al., 2018)	0 (0)	0.309
ECT trajectory number (count)	1.20	1.18	0.792	1.061	1.091	1
Somatic comorbidity (y/n)	0.430 (68)	0.400 (Andrade et al., 2016)	0.689	0.424 (Whiting et al., 2016)	0.364 (Leiknes et al., 2012)	1
Female (y/n)	0.639 (101)	0.733 (66)	0.159	0.727 (Berlin and Golub, 2014)	0.636 (Hageman et al., 2022)	0.706
Age of onset (years)	33.8	36.6	0.166	38.46	47.10	0.317
Medication Failure (y/n)	0.082 (Page et al., 2021)	0.133 (Arora et al., 2019)	0.272	0.182 (collaboration SwgaECr 2021)	0.091 (Kessler et al., 2005)	0.659
Personality Disorder (y/n)	0.310 (49)	0.178 (Pagnin et al., 2004)	0.0246	0.303 (Stacey et al., 2017)	0.364 (Leiknes et al., 2012)	0.722
Bipolar Disorder (y/n)	0.095 (Menger et al., 2018)	0.089 (Koposov et al., 2017)	1	0.061 (WHO, 2022)	0.091 (Kessler et al., 2005)	1
Cognitive Impairment (y/n)	0 (0)	0.011 (Kessler et al., 2005)	0.363	0.091 (Lisanby, 2007)	0.091 (Kessler et al., 2005)	1
Major depressive Disorder (y/n)	0.867 (137)	0.856 (77)	0.849	0.909 (Yip et al., 2021)	0.818 (Bright et al., 2012)	0.586
Schizoaffective Disorder (y/n)	0.0380 (collaboration SwgaECr 2021)	0.0444 (Leiknes et al., 2012)	1	0 (0)	0.091 (Kessler et al., 2005)	0.250

Table 2
the final logistic regression model for predicting remission, and the priors used for fitting the model. NA indicates "not available": prior estimates of mean and sd were available for four out of 18 predictors. 13 predictors were selected to be included in the final model through the Bayesian network analysis.

Final logistic regression model with predicting remission				
Predictor	Mean estimate	Sd estimate	Coefficient	Coefficient SE
Medication failure	−0,65,393	0,143,841	−0,55,991	0,13,831
Severe depressive Episode	−0,097	0,05	−0,08,764	0,049,603
Age	0,258	0,063	0,052,953	0,013,273
Psychotic symptoms	0,383,901	0,116,449	0,388,671	0,109,663
Personality disorder	NA	NA	−0,50,342	0,34,052
Bipolar disorder	NA	NA	−0,46,309	0,657,538
Relapse	NA	NA	−0,37,544	0,41,086
ECT trajectory number	NA	NA	−0,2943	0,270,555
Major depressive disorder	NA	NA	−0,18,883	0,618,723
Schizoaffective disorder	NA	NA	−0,10,911	0,700,235
Age of onset	NA	NA	−0,02,874	0,013,801
Episode duration	NA	NA	−0,01,756	0,007,355
Female	NA	NA	0,489,148	0,295,852
First ECT	NA	NA	NA	NA
Catatonic symptoms	NA	NA	NA	NA
Forced care	NA	NA	NA	NA
Somatic comorbidity	NA	NA	NA	NA
Cognitive impairment	NA	NA	NA	NA

Table 3
Group means (for continuous data) or proportions with the corresponding property (for dichotomous data) for misclassified cases in the validation set, split based on false negative or false positive misclassification.

Misclassification analysis		
	False negative ($n = 5$)	False positive ($n = 7$)
Relapse	1	1
Episode duration	24	18.8
Age	58.0	76.0
First ECT	0.750	0.571
Psychotic symptoms	0.250	0.714
Catatonic features	0	0.143
Severe depressive episode	0.400	0
Forced care	0	0.286
ECT trajectory number	1.00	1.14
Somatic comorbidity	0.400	0.571
Female	0.400	0.857
Age of onset	42.3	65.5
Medication failure	0	0
Personality disorder	0.600	0.143
Bipolar disorder	0.200	0
Cognitive impairment	0	0.143
Major depressive disorder	0.800	1.00
Schizoaffective disorder	0	0

These false positives could possibly explain the decreasing trend in the calibration plot in the bins with the highest predicted probabilities of remission, where the model overestimates the success probabilities (see Fig. 1).

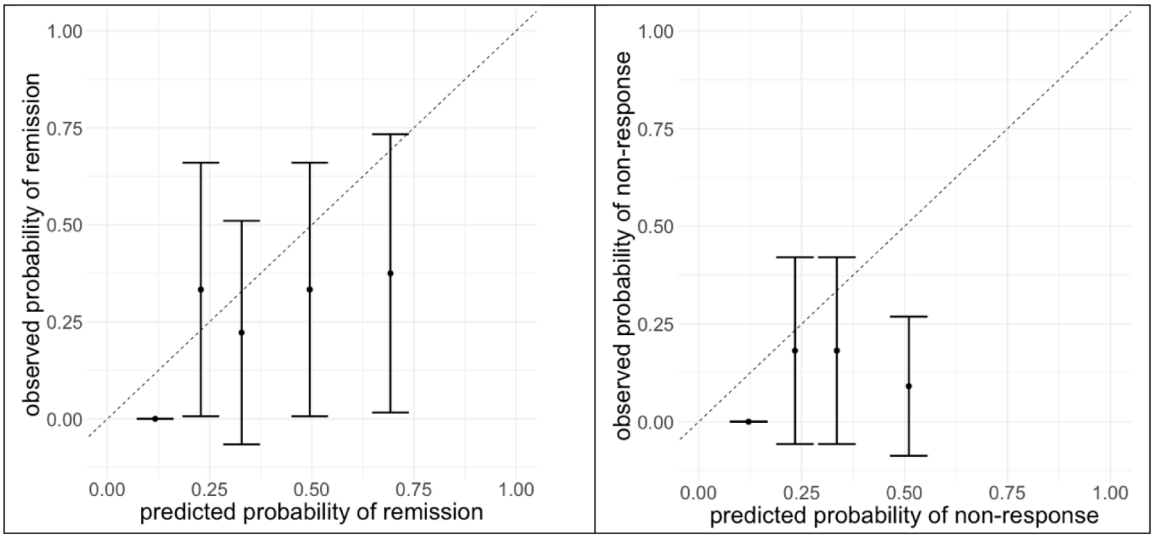

Fig. 1. Calibration plots of the model for prediction remission (left) and non-response (right) on the validation set. Patients in the validation set were divided into 5 or 4 equal bins, depending on the probability of remission or non-response as predicted by the model. For those bins, the observed probability of remission or non-response and corresponding upper- and lower confidence bounds were estimated based on the patient data, resulting in the figures depicted above.

3.2.3. Bayesian network model and hierarchical model for predicting secondary outcome non-response

In the training set, 63 (25%) of trajectories were classified as non-responders. The AUC for the model for predicting the secondary outcome, ECT non-response, was 0.644 (values 0.603–0.675 observed in 5-fold cross-validation), with a classification accuracy estimated through 5-fold cross-validation of 0.746 (balanced accuracy: 0.542). In the validation set, non-response occurred in 5 out of 44 patients. Validation of the updated model resulted in an AUC of 0.528 (95%CI 0.250–0.807) for predicting (non-)response, with an accuracy of 0.84 (balanced accuracy: 0.562), a sensitivity of 0.20 and a specificity of 0.92. The trained hierarchical Bayesian logistic regression model and an overview of priors for non-response can be found in supplementary files. The four alternative machine learning techniques showed comparable AUCs: 0.545 for unpenalized regression (glm), 0.628 for penalized regression (glmnet), 0.591 for random forest (rf) and 0.614 for boosting (xgbTree). Other metrics were reported in the supplementary material.

4. Discussion

In this study, we created and temporally validated BN model to predict outcome after ECT for depression, using prior knowledge from literature combined with single center clinical patient data. We found a mean AUC of 0.629 (values 0.505 – 0.763 observed in 5-fold cross-validation) for the training set and an AUC for the validation set of 0.686 (95%CI 0.513–0.859) for predicting remission to ECT. These findings suggest that probability of remission of a depressive episode using ECT can be reasonably well estimated with readily available clinical predictors for individual patients. For non-response, we found a mean AUC of 0.644 and an AUC for the validation set of 0.528 (95%CI 0.250–0.807)

High-quality meta-analyses are considered as the highest level of evidence in evidence-based medicine. However, one of the downsides of meta-analyses is that the aggregated data have no direct clinical value to individual patients (Berlin and Golub, 2014). In this study we used the knowledge from the best meta-analyses available in a BN model to create a clinical decision support system which calculates personalized outcome predictions for ECT. Although these methods have been studied before, this study is, to our knowledge, the first to investigate the outcome of ECT using a BN. Previous studies of BNs in psychiatry focused on dementia and cognitive impairment (Jin et al., 2016; Gross et al., 2018; Moreira and Namen, 2018). BNs are mostly used in the fields of cardiology and oncology, but have not yet been adopted as a standard technique in medical decision making. One explanation is that previous publications on BNs mostly emphasized technical aspects instead of clinical usefulness (McLachlan et al., 2020; Kyrimi et al., 2021). We found that the addition of prior information to our model did not influence the AUC for remission, but marginally reduced CI width, from an AUC of 0.686 (95%CI 0.491–0.881) in the no priors model, to an AUC of 0.686 (95%CI 0.513–0.859) in the final model. Based on these findings, including prior information hypothetically decreases the sampling variability in a model, by increasing the number of samples of which data is derived. An additional value of priors is that they can be used as an extra validation of findings in a study cohort. If significant discrepancies are observed, further investigation on bias is warranted. There was no improvement in model performance between BN and other machine learning approaches. This shows that our model is comparable with more standard approaches. Further optimization of performance may require more predictors.

Our findings showed that the presence of psychotic symptoms was a strong predictor for remission, as well as the absence of a personality disorder and the absence of medication failure. These findings were expected because previous studies which identified these variables were used as prior knowledge in our study (van Diermen et al., 2018; Heijnen et al., 2010). Several studies found reduced effectiveness of ECT in patients with personality disorders (Yip et al., 2021; Prudic et al., 2004). Higher age was a statistically significant predictor for remission in our study, which is in line with previous research (van Diermen et al., 2018). The secondary outcome of non-response did not yield significant results.

In the misclassification analyses and calibration plots for both remission and non-response, we found a decreasing trend in the plots in the higher predicted probabilities, resulting in an overestimation of success observed probabilities (Fig. 1). Specificity was relatively high, but several cases were falsely positive, resulting in low sensitivity, as reflected in the low balanced accuracy We infer that the dataset may be confounded. However, because of the small sample size of the validation cohort, we cannot assess to what extent. Exploratory analyses of potential confounders, preferably in a larger validation cohort, may yield additional clinical predictors. Next to clinical and demographical parameters, several previous studies reported on biomarkers as predictors of ECT outcome, including MRI, EEG and genetic findings (Luykx et al., 2022; Levy et al., 2019; Simon et al., 2021). Hypothetically, the

accuracy of our model could be increased by including these predictors. However, the problem with these data is that these are not routinely obtained in clinical practice, and therefore often unavailable for the treatment decision about ECT. Therefore, although we were unable to include biomarkers in the model due to unavailability in our data, the clinical model presented here may be easier to implement in clinical practice than a model based on biomarker data.

Although outcome prediction of ECT may benefit shared decision-making, prospective studies are necessary before this model can be implemented as CDSS in standard practice. For example, the subjective experience and needs of individual patients are essential for treatment decisions but were not included here. Also, all factors that may be of influence should be reported in a standardized manner, covering items from all clusters of the biopsychosocial model. Moreover, in our sample, the decision to initiate ECT was already made. This resulted in a selected population of patients who were willing to undergo ECT. To assess clinical usefulness, it is necessary to also analyze the patients who decide not to start ECT, and why this decision is made. Misclassification bias may arise after implementation if treatment decisions are made differently because they are informed by a CDSS, and this adaptive change in decision making is not accounted for. One solution for this potential bias is a stepped-wedge cluster randomized controlled trial, in which the CDSS intervention (and its impact on outcomes) is gradually introduced and evaluated at sites (Hemming et al., 2018). Another factor is the unknown generalizability of findings from our single center study at a university hospital to other treatment settings. We speculate that this could have resulted in an increased severity of depression in our sample, and maybe in other unknown selection biases. An (inter)national, multicenter trial could increase generalizability of our current findings. When looking ahead, a CDSS based on the studies mentioned above could fundamentally alter the care for patients with affective disorders. Currently, ECT is an end-of-the-line treatment, and is initiated when other non-pharmacological and pharmacological treatments have failed. A thoroughly validated CDSS may advise to either start ECT as a first line treatment, or advise to not start ECT at all, given the low chance of success and/or side effects. In the end, a CDSS will provide an advice, it is up to the patient and psychiatrist to make the decision.

Our model did not include adverse effects of ECT. This was due to the fact that adverse effects were not recorded systematically, which may have led to a reporting bias. Adverse effects of ECT consist of amnesia, headache and nausea and occur in most patients during treatment (Andrade et al., 2016). Adverse events may be mild, but can also be a reason to halt ECT prematurely, for example in the case of severe amnesia or delirium. Halting treatment may consequently influence the outcome. We hypothesize that there may also be dependencies between these predictors, and that these could be incorporated to the BN model. Additionally, the inclusion of data generated during each session of ECT, such as seizure duration could be used to predict outcomes more accurately during the treatment. However, this would require a model with repeated measurements, with updated probabilities after each session. This approach could guide psychiatrists and patients in their decision to continue, stop or alter frequency of ECT. We aim to expand our model to include these factors and to further test for generalizability in future work.

We used a systematic review of meta-analyses for the collection of prior knowledge. A downside of this method is missing data of recent studies which are not yet included in a meta-analysis. Another problem was that several studies were included in more than one meta-analysis, and that meta-analyses on the same subject reported different outcomes. We considered risk of bias smallest if we analyzed the searches of multiple research groups and selected the one meta-analysis with the highest quality, with the potential risk of sacrificing some recency of data, and loss of prior information based on single studies. An admission for ECT treatment includes more than only the performing of ECT. Other aspects of the treatment, such as changes in medication, social factors, and psychotherapy may be of influence in the outcome of the treatment. In our study, detailed information specific medication switches, including dosing and timing was not available. Therefore, we de facto studied an admission for ECT in our hospital, including all standard care on pharmacological optimization, social and psychotherapy interventions. A prospective study design, with inclusion of standardized reporting on these factors may provide additional insights. We used clinical discharge letters with the final outcome of ECT to define the outcomes remission and response. Quantitative assessment of depression, for instance using the Hamilton Rating Scale for depression (HRSD) or Montgomery–Åsberg Depression Rating Scale (MADRS), is often used in clinical trials (van Diermen et al., 2018; Hamilton, 1967; Montgomery and Asberg, 1979). Outcomes remission and (partial) response are defined using a reduction of the score by a certain percentage, or below an arbitrary threshold. The potential upside of this approach is that, in theory, treatment can be evaluated objectively. However, there is an ongoing debate about the use of the reliability and validity of depression instruments (Fried et al., 2022). One of the hypothetical downsides of depression instruments is that the score is comprised of several symptom clusters. An equal reduction in scores of two patients after ECT may not resemble the same effect. Additionally, in clinical practice, standardized application of quantitative assessments requires additional time and training of staff. Therefore, we chose to use the most clinically relevant outcome assessment available, the conclusion of the discharge letter. This outcome included both clinician assessment and subjective patient experience. There are downsides to this approach, for there is no standardization, and deriving the outcome requires interpretation. As a result, there may have been both false positives and false negatives. Although discharge letters are far from a gold standard, this outcome does offer an integrated and personalized conclusion. Further conceptual research in validity of depression severity may use these integrative concepts to formulate a standardized and clinically useful outcome measurement. In 23 cases, we had missing data on clinical variables. We used multiple imputation to make optimal use of data. Although multiple imputation is superior to complete case analysis regarding potential bias, it may influence model performance (Steyerberg, 2009). In this study, we used a validation sample from our own hospital setting. In general, external validation usually reduces performance. In our study, we assume that the greatest effect on the performance will be due to a selection bias of the population of our center, a tertiary care university clinic. This selection bias may be based on residential area of patients and reasons for referral from other mental health institutions, which may be correlated with comorbidity, educational level and support from a community. To assess generalizability of our findings, external validation is necessary to account for setting-specific confounders and selection bias. Possibilities for retrospective external validation could be (multicenter) ECT registries. Prospective validation in a multicenter cohort study, as stated above, can yield even more insights.

5. Conclusion

In this study, we found that a BN model comprised of prior knowledge and clinical data can predict remission of depression after ECT with reasonable performance. This approach can be used to make outcome predictions in psychiatry, and offers a methodological framework to weigh additional information, such as patient characteristics, symptoms and biomarkers. In time, it may be used improve shared decision-making in clinical practice.

CRediT authorship contribution statement

Yuri van der Does: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Writing – original draft, Writing – review & editing. **Rosanne J. Turner:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Miel J.H. Bartels:**

Data curation, Formal analysis, Software, Writing – review & editing. **Karin Hagoort:** Data curation, Formal analysis, Writing – review & editing. **Aäron Metselaar:** Data curation, Formal analysis, Methodology, Writing – review & editing. **Floortje Scheepers:** Supervision, Writing – review & editing. **Peter D. Grünwald:** Formal analysis, Funding acquisition, Software, Supervision, Writing – review & editing. **Metten Somers:** Conceptualization, Methodology, Writing – review & editing. **Edwin van Dellen:** Conceptualization, Data curation, Investigation, Methodology, Supervision, Validation, Visualization, Writing – review & editing.

Declaration of Competing Interest

Yuri van der Does, Rosanne J. Turner, Miel J.H. Bartels, Karin Hagoort, Aäron Metselaar, Floortje Scheepers, Peter D. Grünwald, Metten Somers, Edwin van Dellen None of the authors declare any conflict of interest.

Acknowledgements

Part of this work was performed as part of the Enabling Personalized Interventions (EPI) project, which is supported by the Dutch Research Council (NWO) in the Commit2 Data –Data2Person program under contract 628.011.028.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.psychres.2023.115328.

References

Kessler, R.C., Chiu, W.T., Demler, O., Merikangas, K.R., Walters, E.E., 2005. Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. Arch. Gen. Psychiatry 62 (6), 617–627.

WHO. Fact sheet depression 2022 [Available from: https://www.who.int/news-room/fact-sheets/detail/depression.

Lisanby, S.H., 2007. Electroconvulsive therapy for depression. N. Engl. J. Med. 357 (19), 1939–1945.

Leiknes, K.A., Jarosh-von Schweder, L., Høie, B., 2012. Contemporary use and practice of electroconvulsive therapy worldwide. Brain Behav. 2 (3), 283–344.

van Diermen, L., van den Ameele, S., Kamperman, A.M., Sabbe, B.C.G., Vermeulen, T., Schrijvers, D., et al., 2018. Prediction of electroconvulsive therapy response and remission in major depression: meta-analysis. Br. J. Psychiatry 212 (2), 71–80.

collaboration SwgaECr, 2021. SCORE2 risk prediction algorithms: new models to estimate 10-year risk of cardiovascular disease in Europe. Eur. Heart J. 42 (25), 2439–2454.

Hageman, S.H.J., McKay, A.J., Ueda, P., Gunn, L.H., Jernberg, T., Hagström, E., et al., 2022. Estimation of recurrent atherosclerotic cardiovascular event risk in patients with established cardiovascular disease: the updated SMART2 algorithm. Eur. Heart J. 43 (18), 1715–1727.

Koposov, R., Fossum, S., Frodl, T., Nytrø, Ø., Leventhal, B., Sourander, A., et al., 2017. Clinical decision support systems in child and adolescent psychiatry: a systematic review. Eur. Child Adolesc. Psychiatry 26 (11), 1309–1317.

Bright, T.J., Wong, A., Dhurjati, R., Bristow, E., Bastian, L., Coeytaux, R.R., et al., 2012. Effect of clinical decision-support systems: a systematic review. Ann. Intern. Med. 157 (1), 29–43.

Stacey, D., Légaré, F., Lewis, K., Barry, M.J., Bennett, C.L., Eden, K.B., et al., 2017. Decision aids for people facing health treatment or screening decisions. Cochrane Database Syst. Rev. 4 (4), Cd001431.

Briganti, G., Scutari, M., McNally, R.J., 2022. A tutorial on bayesian networks for psychopathology researchers. Psychol. Methods.

Arora, P., Boyne, D., Slater, J.J., Gupta, A., Brenner, D.R., Druzdzel, M.J., 2019. Bayesian networks for risk prediction using real-world data: a tool for precision medicine. Value Health 22 (4), 439–445.

Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., et al., 2021. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 372, n71.

Whiting, P., Savović, J., Higgins, J.P., Caldwell, D.M., Reeves, B.C., Shea, B., et al., 2016. ROBIS: a new tool to assess risk of bias in systematic reviews was developed. J. Clin. Epidemiol. 69, 225–234.

Menger, V., Scheepers, F., van Wijk, L.M., Spruit, M.R., 2018. DEDUCE : a pattern matching method for automatic de-identification of Dutch medical text. Telematics Inf. 35 (4), 727 null.

Pagnin, D., de Queiroz, V., Pini, S., Cassano, G.B., 2004. Efficacy of ECT in depression: a meta-analytic review. J. ECT 20 (1), 13–20.

Scutari, M., 2010. Learning Bayesian Networks with the bnlearn R Package. J. Stat. Softw. 35 (3), 1–22.

Collins, G.S., Reitsma, J.B., Altman, D.G., Moons, K.G., 2015. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. Ann. Intern. Med. 162 (1), 55–63.

Baldwin, S., Oxlad, M., 1996. Multiple case sampling of ECT administration to 217 minors: review and meta-analysis. J. Mental Health 5 (5), 451–464.

Haq, A.U., Sitzmann, A.F., Goldman, M.L., Maixner, D.F., Mickey, B.J., 2015. Response of depression to electroconvulsive therapy: a meta-analysis of clinical predictors. J. Clin. Psychiatry 76 (10), 1374–1384.

Havaki-Kontaxaki, B.J., Ferentinos, P.P., Kontaxakis, V.P., Paplos, K.G., Soldatos, C.R., 2006. Concurrent administration of clozapine and electroconvulsive therapy in clozapine-resistant schizophrenia. Clin. Neuropharmacol. 29 (1), 52–56.

Kho, K.H., Van Vreeswijk, M.F., Simpson, S., Zwinderman, A.H., 2003. A meta-analysis of electroconvulsive therapy efficacy in depression. J. ECT 19 (3), 139–147.

Heijnen, W.T., Birkenhäger, T.K., Wierdsma, A.I., van den Broek, W.W., 2010. Antidepressant pharmacotherapy failure and response to subsequent electroconvulsive therapy: a meta-analysis. J. Clin. Psychopharmacol. 30 (5), 616–619.

Berlin, J.A., Golub, R.M., 2014. Meta-analysis as evidence: building a better pyramid. JAMA 312 (6), 603–605.

Jin, Y., Su, Y., Zhou, X.H., Huang, S., 2016. Heterogeneous multimodal biomarkers analysis for Alzheimer's disease via Bayesian network. EURASIP J. Bioinform. Syst. Biol. 2016 (1), 12.

Gross, T.J., Araújo, R.B., Vale, F.A.C., Bessani, M., Maciel, C.D., 2018. Dependence between cognitive impairment and metabolic syndrome applied to a Brazilian elderly dataset. Artif. Intell. Med. 90, 53–60.

Moreira, L.B., Namen, A.A., 2018. A hybrid data mining model for diagnosis of patients with clinical suspicion of dementia. Comput. Methods Programs Biomed. 165, 139–149.

McLachlan, S., Dube, K., Hitman, G.A., Fenton, N.E., Kyrimi, E., 2020. Bayesian networks in healthcare: distribution by medical condition. Artif. Intell. Med. 107, 101912.

Kyrimi, E., Dube, K., Fenton, N., Fahmi, A., Neves, M.R., Marsh, W., et al., 2021. Bayesian networks in healthcare: what is preventing their adoption? Artif. Intell. Med. 116, 102079.

Yip, A.G., Ressler, K.J., Rodriguez-Villa, F., Siddiqi, S.H., Seiner, S.J., 2021. Treatment outcomes of electroconvulsive therapy for depressed patients with and without borderline personality disorder: a retrospective cohort study. J. Clin. Psychiatry 82 (2).

Prudic, J., Olfson, M., Marcus, S.C., Fuller, R.B., Sackeim, H.A., 2004. Effectiveness of electroconvulsive therapy in community settings. Biol. Psychiatry 55 (3), 301–312.

Luykx, J.J., Loef, D., Lin, B., van Diermen, L., Nuninga, J.O., van Exel, E., et al., 2022. Interrogating Associations Between Polygenic Liabilities and Electroconvulsive Therapy Effectiveness. Biol. Psychiatry 91 (6), 531–539.

Levy, A., Taib, S., Arbus, C., Péran, P., Sauvaget, A., Schmitt, L., et al., 2019. Neuroimaging biomarkers at baseline predict electroconvulsive therapy overall clinical response in depression: a systematic review. J. ECT 35 (2), 77–83.

Simon, L., Blay, M., Galvao, F., Brunelin, J., 2021. Using EEG to predict clinical response to electroconvulsive therapy in patients with major depression: a comprehensive review. Front. Psychiatry 12, 643710.

Hemming, K., Taljaard, M., McKenzie, J.E., Hooper, R., Copas, A., Thompson, J.A., et al., 2018. Reporting of stepped wedge cluster randomised trials: extension of the CONSORT 2010 statement with explanation and elaboration. BMJ 363, k1614.

Andrade, C., Arumugham, S.S., Thirthalli, J., 2016. Adverse effects of electroconvulsive therapy. Psychiatr. Clin. North Am. 39 (3), 513–530.

Hamilton, M., 1967. Development of a rating scale for primary depressive illness. Br. J. Soc. Clin. Psychol. 6 (4), 278–296.

Montgomery, S.A., Asberg, M., 1979. A new depression scale designed to be sensitive to change. Br. J. Psychiatry 134, 382–389.

Fried, E.I., Flake, J.K., Robinaugh, D.J., 2022. Revisiting the theoretical and methodological foundations of depression measurement. Nat. Rev. Psychol. 1 (6), 358–368.

Steyerberg, E.W., 2009. Clinical Prediction Models. Springer New York, New York.