Predictive model of the treatment effect for patients with major depressive disorder

Igor Akushevich
Center for Population Health and Aging,
Duke University
Durham, NC, 27708
Phone: 1(919)668-2715
ia6@duke.edu

Julia Kravchenko
Department of Surgery,
Duke University
Medical Center
Durham, NC, 27705
Phone: 1(919)668-6809
yk27@duke.edu

Ken Gersing
Department of Psychiatry,
Duke University
Medical Center
Durham, NC 27705
Phone: 1 (919)684-0103
kenneth.gersing@duke.edu

Ketan K. Mane
RENCI, UNC-Chapel Hill
100 Europa Dr.,
Suite 540,
Chapel Hill, NC
Phone:1(919)445-9703
kmane@renci.org

ABSTRACT
The model to evaluate and predict the effectiveness of treatment of the Major Depressive Disorder (MDD) was developed and estimated using MindLinc data. The clinical global impression (CGI) scale with seven categories was used to measure the patient's state. The proportional odds model was selected because of ordinal nature of the outcome. The set of predictors included i) CGI score measured at preceded visit, ii) three groups of medications (antidepressants, atypical medicine, and augmentation medicine), all categorized for appropriate number of strata (from six to nine) and their daily doses, iii) psychiatric comorbidities, iv) type of the therapy used (talk vs. medications), v) demographic variables (e.g., age group, sex), and vi) the history of the efficiency of prior treatment. More than a half of a million records with measured CGI scores and their predictors were identified in the MindLinc database and used for model estimation. The predicted model of future CGI scales was developed and evaluated for single and recurrent episodes of MDD. Significant estimates were obtained for demographic factors, history of previous SGI scales, and for comorbidity and treatment indices. The methods of causal inferences based on the inverse probability weighting approach were applied to evaluate the treatment effects. The model extensions allowing for addressing the limitations of the proportional odds model are discussed.

Categories and Subject Descriptors
D.3.3 [**Programming Languages**]: SAS

General Terms
Algorithms, Design.

Keywords
CGI, MindLinc, Major Depressive Disorder, causal effects, predictive modeling

BCB '13, September 22 - 25, 2013, Washington, DC, USA

1. INTRODUCTION
The major depressive disorder (MDD) is one of the most prevalent mental disorder in developed countries that affect approximately 1 in 6 people during their lifetime and causes both occupational and social impairments [1,2]. MDD is expected to become one of the three leading causes of burden of disease in 2030 [3]. MDD represents the severe end of a spectrum of depressive disorders. It could be associated with other psychiatric disorders or/and with non-psychiatric diseases such as cardio- and cerebrovascular diseases, ulcer, asthma, diabetes, obesity and others [4]. Despite its high prevalence and substantial impairments it causes, very little is known about the factors that impact effectiveness of MDD treatment, its interactions with other mental disorders and somatic diseases [5]. Regardless the introduction of new generations of antidepressants, more than a half of the patients do not respond to the treatment (e.g., for the therapy with the first-line antidepressants) [6]. Lack of symptomatic remission affects economic burden of MDD due to the loss of productivity caused by disability and pharmacological therapies [7,8]. Because there is a lot of variability in individual outcomes from MDD treatments, there is a need in "increasing efforts to identify the factors that may predict the optimal treatment modality for a given patient" [9]. However, there is still a gap in developing such approaches that can be applied to the large existing datasets and will be able to make prognoses of therapy effectiveness.

Medical records are undergoing a major transition from paper-to-electronic storage in electronic medical record systems (EMRs). That presents a huge opportunity to better understand the data to improve diagnosis and effectiveness of treatment and to make a step toward a personalized medicine approach. Traditionally, practitioners rely on observed causal correlations in determining treatment plan; however, in the era of expanding medical data and limited patient-physician (e.g., visit) time, there is a growing need for the tools that can facilitate physicians with their decision making process. Use of EMR data, and a data-driven approach offers promise to bridge the evidence gap, and to leverage collective evidence in clinical care. Predictive modeling is one of these approaches that can use EMR data, derive insight, and guide physician's selection of treatment choice that has the more likelihood of working for the target patient.

Predictive modeling is focused on creating a statistical tool capable of projecting future patient's characteristics (e.g., their

health status). In modeling process, data are collected for the relevant predictors, a statistical model is formulated, predictions are made and the model is validated (or revised) as additional data becomes available. With relevant data available, modeling effort offers an opportunity to build different models to learn more about how to prevent complications; control readmissions; generate more precise diagnoses and treatments; predict risk; and control costs for a more diverse array of population segments than previously attempted. In this paper, we focus on developing and using the predictive modeling approach to determine the treatment outcome in patients with MDD as a primary diagnosis.

2. DATA AND METHODS

The individual health state in respect of MDD is represented by the Clinical Global Impression (CGI) scale that is commonly used to measure the symptoms severity, response to the treatment, and treatment effectiveness for the patients with mental disorders [10,11]. There are two versions of the scale: severity (CGI-S) and improvement (CGI-I). Both of them are the 7-point scales that require the clinician to evaluate the severity of patient's condition at the time of the visit (CGI-S scale) and change of patient's status (i.e., improvement or worseness) (CGI-I scale) comparing to the baseline. The seven categories for severity are i) no disease, healthy; ii) borderline mentally ill; iii) mildly ill; iv) moderately ill; v) markedly ill; vi) severely ill; or vii) extremely ill. The seven categories for improvements are: i) very much improved; ii) much improved; iii) minimally improved; iv) no change; v) minimally worse; vi) much worse; or vii) very much worse. The CGI score was developed and validated for use by the National Institute of Mental Health [10,11] and is used in virtually all FDA-regulated and most CNS trials. The CGI provides a brief, stand-alone score of the clinician's assessment of the patient's global functioning prior to and after initiating a treatment. The purpose of the developed predictive model is to predict CGI (severity or improvement) for the next patient's visit based on the information available at the time of a current visit.

2.1 MindLinc dataset

The MindLinc EMR system is the largest de-identified psychiatry outcome data warehouse in the United States, and it is a clinically representative sample of the data collected in psychiatric practice [12]. The MindLinc data warehouse represents 110,000 patients or 2,400,000 clinical encounters collected during the past ten years. This data are drawn from the various types of mental health facilities—academic medical centers (25%), community mental health centers (50%) and other practices (25%)—from geographically different areas of the country (North, East, South and West). The patients' data from each psychiatric practice site are pooled every 6 months into a de-identified, HIPAA compliant data warehouse. The MindLinc EMR system stores data on patients' demographics, psychiatric diagnoses current and past medications, their side-effects, psychiatric and non-psychiatric comorbidities, and other related clinical data, e.g., therapeutic outcomes.

Within the MindLinc, all treatment outcomes are recorded in the form of CGI improvement and severity scores for each patient. There are about 32,000 individuals who had at least one diagnosis of MDD identified in the data set. The all MDD cases are classified as single (ICD-9-CM 296.2x) or recurrent (296.3x).

Individual follow-up period is separated on time periods between measured CGI. Thus, the dataset reconstructed to be appropriate for data analysis and modeling (including estimation of parameters of the predictive model) consists of these time periods that are considered as statistical units of our data analysis. The effect of a current treatment is expected to occur between the fifth and the eighth week. Therefore, the dependent variable is estimated as a minimal CGI measured in that period. Thus, the time period is characterized by the dependent variable and the set of independent variables (or predictors) available at the time of the visit including information about treatment, comorbidities, history of previously measured CGIs.

The time period of the individual follow-up starts when the (initial) CGI is measured. Note, that different CGIs can be assigned to an individual in the same date, because several exams of different medical specialists can be performed at the same date. Therefore, not all CGIs have the same quality. In our study, only CGIs assigned during an outpatient visit were considered. Specifically, a unique CGI for a day is defined as the latest CGI assigned at an outpatient visit. If there are no outpatient visits, the day-specific CGI is the latest CGI among all types of visits.

Several groups of medications (such as antidepressants, atypical medicine, and augmentation medicine) were available for each patient at the baseline. Individual histories of prescribed treatment were also available for analysis. Each medication was considered to be effective before information on the next prescription would be available. Therefore, the prescribed (and effective) medication was known during the patient's medical history. We used these assumptions to fill missing data when a medicine is prescribed in the date different than the date of CGI assignment. Each of the three main groups were further divided for sub-groups, e.g., anti-depressant categories included atypical dopamine, atypical, monoamine oxidase (MAO) inhibitors, tricyclic antidepressants (TCAs), serotonin-norepinephrine reuptake inhibitors (SNRIs), and selective serotonin reuptake inhibitors (SSRIs). The categories of augmentation included anti-mania, anxiolytics, CNS stimulant, lithium, mood stabilizer, thyroid, atypical, and other.

Importantly, daily doses of antidepressants were reconstructed for each category of antidepressants using information available in prescriptions. Therefore, each type of antidepressants is represented by the continuous variable. Binary indicators for other groups of medicine are currently used.

Eighteen psychiatric comorbidities coded according to the Diagnostic and Statistical Manual of Mental Disorders (DSM) code were considered: 1) adjustment disorders, 2) anxiety disorders, 3) delirium, dementia, amnestic, cognitive, 4) dissociative disorders, 5) eating disorders, 6) factitious disorders, 7) impulse control, 8) infancy, childhood, adolescence, 9) mental disorders due to GMC, 10) mood disorders, 11) other conditions, 12) personality disorders, 13) schizophrenia, 14) sexual, gender identity, 15) sleep disorders, 16) somatoform disorders, 17) substance-related disorders, and 18) additional codes. Patient can have several comorbidities at a time of a visit; therefore, we used eighteen binary indicators for all types of comorbidities.

Information about the type of medical personnel (such as allied, counselor, physician) was used as a cofactor in the predictive model Again, since several types of the treatments can be available for the same day, we used the respective binary indicators for all types of medical personnel. Note, these variables also reflect the information about the type of treatment, i.e., talk (allied or counselor) vs. medicine (physician).

Demographic variables included age group, sex (male, female, and unknown), and race (White, Asian, Black, Hispanic, Multiracial, Native American, other, and undetermined). Age-at-visit groups were defined as 0-30, 31-40, 41-50, 51-65, and 66+ years old.

The history of previous patient state and the effect of last treatment are represented by the last change of CGI (denoted as $\Delta\mathrm{CGI}$). A separate category is used for the first visit, i.e., when there is no history. The categories are: 'non', -2- (i.e., $\Delta\mathrm{CGI\_I} \le -2$), -1, 0, 1, and 2+ (i.e., $\Delta\mathrm{CGI\_I} \ge 2$).

3. The MODEL

CGI are ordinal measures, therefore, the proportional odds model was chosen as a base modeling approach. The model fits a cumulative model with the common slopes. The cumulative model has the form

$$g(\Pr(Y \le i) \mid x) = \alpha_i + \sum\nolimits_j \beta_j x_j \qquad (1)$$

where Y is CGI (severity or improvement) categorized in seven groups, i runs from 1 to 7, x is the vector of predictors, j runs from 1 to the number of predictors, $\alpha_1,\ldots,\alpha_7$ are seven intercept parameters, and β is the vector of slope parameters. The function $g()$ is so-called link function allowing us to consider a function of the cumulative probability as a new dependent variable. SAS Proc LOGISTIC provided parameter estimation.

This model is the base predictive model. There are two basic aims that can be addressed using such models. The first is the predictions of future CGI based on the individual characteristics available at the time of visit. The second is the evaluation of the causal effects of counterfactual treatments on future values of the CGI.

We found that predictions using this model are quite precise. We considered three types of patient inclusion criteria: i) all patients if they have severe MDD at least once during a life, ii) only single episodes, and iii) only recurrent episodes. The list of ICD-9 codes included codes 296.23 and 296.24 (single episode without and with psychotic features) and 296.33 and 296.34 (recurrent episode without and with psychotic features). In total, there were 31,654 patients who have been diagnosed with MDD at least once. Based on these selection criteria, we identified 479,055, 17,605, and 76,452 visits for parameter estimations, respectively. Variables of all types (as presented in the Data and Methods Section) were tested to be predictors of the CGI (both severity and improvement) at the visits occurred between the fifth and the eight week. The most strong (i.e., the most statistically significant) predictors were initial CGI score and the history of previous CGI measurements. This means that the trajectories of individual CGIs during an episode were quite stable and can explain majority of CGI variability. Among other predictors with the significant effects were demographic factors: for example, being a male or being Black or Hispanic was associated with better prognosis (i.e., decreased future CGI). The prognosis was better for the age group of 31-40; while being younger than 31 years or older than 40 years were associated with worse prognosis. The effects of psychiatric comorbidities and treatment were controversial; that

Table 1. *Odds ratios (OR), confidence intervals (CI), and p-values estimated using the model (1) for single and recurrent MDD episodes*

Variable	Category	Single	Recurrent
		OR (CI), p-value	
Initial CGI (vs. 4)	1	104 (80.38,133.3), <.0001	101 (88.87,114.3), <.0001
	2	7.15 (6.25, 8.18), <.0001	7.54 (7.03, 8.08), <.0001
	3	1.39 (1.25, 1.56), <.0001	2.27 (2.16, 2.40), <.0001
	5	0.06 (0.06, 0.07), <.0001	0.20 (0.19, 0.21), <.0001
	6	0.05 (0.04, 0.05), <.0001	0.08 (0.08, 0.09), <.0001
	7	0.96 (0.58, 1.61), 0.8903	0.06 (0.04, 0.07), <.0001
Sex (vs. Female)	Male	0.39 (0.33, 0.47), <.0001	0.99 (0.97, 1.00), 0.0715
Race (vs. White)	Black	0.90 (0.84, 0.97), 0.0045	0.85 (0.83, 0.88), <.0001
	Hispanic	0.99 (0.92, 1.07), 0.7511	0.97 (0.94, 1.01), 0.1379
	Other	0.86 (0.82, 0.91), <.0001	0.97 (0.94, 0.99), 0.0115
Age group (vs. 0-30)	31-40	0.83 (0.77, 0.89), <.0001	0.96 (0.93, 0.99), 0.0070
	41-50	1.10 (1.04, 1.16), 0.0015	1.16 (1.13, 1.19), <.0001
	51-65	1.09 (1.04, 1.15), 0.0009	1.08 (1.06, 1.11), <.0001
	66+	0.96 (0.89, 1.04), 0.3635	1.10 (1.06, 1.14), <.0001
History of previous patient state (vs. no previous state measured)	0	0.63 (0.59, 0.67), <.0001	0.70 (0.68, 0.72), <.0001
	1	2.27 (2.06, 2.51), <.0001	1.82 (1.73, 1.91), <.0001
	2 or more	11.2 (9.62,13.00), <.0001	9.45 (8.78,10.16), <.0001
	-1	0.51 (0.46, 0.56), <.0001	0.56 (0.53, 0.59), <.0001
	-2 or less	0.43 (0.37, 0.51), <.0001	0.52 (0.48, 0.56), <.0001
Comorbidity		1.24 (1.20, 1.27), <.0001	1.21 (1.19, 1.22), <.0001
Current therapy (vs. 1 antidepres-sant [ant] and no aug-mentation medicine [augment])	No therapy	1.28 (1.19, 1.38), <.0001	0.82 (0.61, 1.10), 0.1859
	1 ant. + augment.	0.91 (0.83, 0.99), 0.0232	0.96 (0.71, 1.29), 0.7787
	2 ant.	0.82 (0.76, 0.88), <.0001	0.83 (0.62, 1.11), 0.2140
	2 ant. + 1 augment.	1.21 (1.09, 1.35), 0.0005	1.36 (1.01, 1.83), 0.0441
	3 ant.	0.99 (0.87, 1.12), 0.8477	1.02 (0.76, 1.38), 0.8981
	ECT	n/a	1.21 (0.21, 7.10), 0.8294

suggested testing their aggregated characteristics. Considering aggregated indices could help to obtain significant and interpretable estimates of comorbidity and treatment effects and to avoid possible model overfitting. Therefore, we created and incorporated comorbidity index (as the sum of the presence of comorbid diseases) and aggregated therapy indices as a monotherapy (just one antidepressant independent of dosage), therapy that included two and three antidepressants, therapies with augmentation medicine, and electroconvulsive therapy (ECT). Predictive power represented by the c-index (represented the

Table 2. *Frequencies (in %) of predictors for specific treatments: no treatment (T0), treatment by one antidepressant (T1), treatment by two or more antidepressants (T2), treatment with antidepressants and augmentation medicine (T3) without and with the corrections for selection bias (i.e., with unit weights and inverse probability weights).*

		With unit weights					With IP weights				
Variable	Category	T0	T1	T2	T3	p-value	T0	T1	T2	T3	p-value
Initial CGI	3	23.8	30.0	31.3	26.0	<.0001	33.3	24.0	28.4	29.3	0.6479
	4	53.1	48.1	47.4	53.0		43.5	50.0	47.6	51.0	
	5	23.2	21.9	21.3	21.0		23.2	26.1	24.0	19.8	
Sex	Female	72.5	67.4	76.0	81.8	<.0001	59.3	66.3	72.1	67.9	0.0916
	Male	27.5	32.6	24.0	18.2		40.7	33.7	27.9	32.1	
Race	Black	7.8	11.7	11.8	9.3	<.0001	14.8	10.0	11.0	11.2	0.5944
	Hispanic	6.5	11.4	7.2	11.8		7.5	7.6	7.1	7.5	
	Other	24.7	33.7	24.9	24.5		24.8	24.2	22.8	17.3	
	White	61.0	43.3	56.1	54.3		52.9	58.2	59.0	64.0	
Age group	0-30	32.1	19.7	14.9	12.6	<.0001	19.5	20.2	17.0	23.1	0.6137
	66+	5.2	11.5	8.4	9.9		14.8	12.2	11.5	11.6	
	31-40	14.5	14.4	14.2	12.3		13.2	15.8	23.9	15.5	
	41-50	20.8	21.2	27.6	30.4		24.5	20.8	22.0	20.6	
	51-65	27.5	33.3	34.9	34.8		28.0	31.0	25.6	29.2	
History of previous patient state	0	67.4	77.1	75.7	74.4	<.0001	60.5	69.2	65.0	65.2	0.1428
	1	9.6	8.0	8.9	9.9		15.1	9.5	9.4	8.8	
	2 or more	4.6	2.7	3.1	2.6		9.7	5.3	3.2	8.2	
	-1	8.9	9.0	9.8	11.0		9.8	9.8	18.0	13.7	
	-2 or less	2.0	1.4	1.6	1.0		2.4	3.8	1.8	1.8	
	non	7.5	1.8	0.9	1.1		2.4	2.3	2.4	2.2	
Time since last visit (in weeks)	0	22.5	26.8	31.4	27.4	<.0001	21.6	27.6	23.0	20.8	0.2341
	1	26.3	20.5	20.8	21.0		19.3	18.6	17.1	24.1	
	2	17.5	13.9	14.2	16.1		17.7	13.3	13.1	10.6	
	3	6.9	8.0	7.3	8.0		6.3	6.6	5.6	8.8	
	4	5.6	8.7	8.4	7.6		7.2	9.9	9.0	7.6	
	5	3.5	4.1	4.0	3.4		8.0	2.9	3.6	3.3	
	6 or more	17.7	18.0	13.8	16.6		19.8	21.1	28.6	24.8	
Previous therapy	0	86.2	3.0	1.3	1.9	<.0001	15.1	14.1	13.8	16.4	1.0000
	1	3.5	92.3	5.5	2.9		48.3	45.7	46.8	45.4	
	1 ant.+ aug.	0.9	0.4	0.0	57.9		9.4	10.9	9.2	10.0	
	2 ant.	0.5	1.7	69.2	0.9		16.9	15.6	15.5	14.9	
	2 ant.+aug.	0.2	0.1	1.1	34.1		4.2	6.0	7.0	6.0	
	3 ant.	0.0	0.1	21.5	0.6		3.1	4.6	4.6	4.4	
	no	8.6	2.4	1.5	1.6		3.1	3.0	3.1	2.8	
Comorbidity Index	0 or 1	43.0	33.9	26.8	26.2	<.0001	35.3	32.2	44.4	36.5	0.2261
	2	20.8	37.4	44.9	36.5		33.2	32.2	31.2	30.0	
	3	19.2	17.8	16.2	16.0		20.2	19.3	13.9	16.4	
	4 or more	17.0	10.9	12.1	21.4		11.3	16.3	10.5	17.1	

fraction of correctly predicted events) for the original model was 0.88-0.89 for CGI-S and 0.75-0.77 for CGI-I scale. For the model with aggregated indices for comorbidity and treatment the predictive power is similar: c-index is 0.88 for CGI-S and 0.74-0.76 for CGI-I scale for all three selection criteria.

Validation of the model was performed by separating the complete MindLinc dataset into equal estimation and validation dataset and testing how the model estimated using the estimation dataset predict CGIs in validation dataset. We found that characteristics of predictive power were similar in evaluation and validation parts of the dataset for all considered cases. For example, the Brier score [13] was BS=0.42 for both estimation and validation datasets for recurrent episodes. Note, the prediction of equal probabilities for all seven outcomes results in the Brier score equaling 0.86 and the prediction of probability 1/3 (or ⅓) for observed outcome and equal for remaining outcomes results in BS=0.52 (or 0.29).

The Table 1 presents the results of parameter estimations of the model (1) for CGI severity. The most responsive to the therapy patients were: i) males, ii) non-Whites, iii) patients aged 31-40, or/and iv) patients without (or with minor) comorbidities. Changes of individual states represented by the variable (history of previous patient state) have a tendency to conserve. The results shows that the most efficient treatment strategy is the treatment with two different antidepressants (or one antidepressant and one augmentation medicine), however, as we discuss below such interpretation is naïve and additional statistical analysis and modeling is required.

There are several aspects that make a direct application of the developed model to evaluation of the treatment effect challenging. The most important of them is the selection bias that does not allow us to separate the effects of a treatment itself from the effect of underlying patient state for that this treatment was administrated. Another obstacle is incompleteness of measured information about the patient state used by a physician to make a decision about specific treatment. One approach allowing to address the selection bias or, in the other words, to evaluate the causal effects of treatments involves the propensity-score-based approach with the inverse-probability weighting (IPW). The underlying idea of the approach is to assign the weights for each patient such that the patients groups become randomized in respect of probabilities of treatment assignment. The estimation of treatment effects for 'weighted' patients results in estimates adjusted for treatment selection. The first step is the calculation of individual weights estimated as reciprocal of the probability to have actually observed treatments. The probabilities of all treatment options (or propensity scores) were evaluated using the logistic (generalized logistic in the case of more than two alternative treatments) model. Predictors of treatment included demographic characteristics (like age, sex, and race), current CGI score and history of its previous measurements, aggregated comorbidity score, previous treatment and time since the previous visit. To make sure that the propensity score model correctly adjusts for treatment selection, we compared treatment-specific baseline characteristics with and without using the weights. Tables 2 gives an example of the distributions of the baseline characteristics in treatment groups for CGI severity and single MDD episodes. Four treatment strategies were considered: i) no treatment (T0), ii) treatment by one antidepressant (T1), iii) treatment by two or more antidepressants (T2), and iv) treatment with antidepressants and augmentation medicine (T3). The p-values provide the probabilities that the frequencies of specific variables are not equal for these treatment groups. The differences are found to be strongly significant for the original variables (i.e., with using unit weights or not corrected for the selection bias) and the difference becomes minor after applying the correction to the selection bias (i.e., when using inverse-probability weights). This indicates that inverse-probability weights properly correct for the selection bias. Then, the estimates of the treatment effects (T0 vs. T1, T2 vs. T1, and T3 vs. T1) were obtained using the proportional odds model for individuals with unit weights and weights estimated within inverse-probability weighting approach. The results obtained with using the IPW (presented in Table 3) were interpreted as causal effects of a specific treatment on improvement of patient state represented by CGI. The results showed, that if the treatment effects were controversial before the correction for the selection bias, then the estimates obtained with the inverse-probability weights were quite stable. The 'no treatment' strategy was the less beneficial for all considered case, while the treatment with two types of antidepressants and treatment involving augmentation medicine are beneficial for certain cases presented in Table 3.

Limitations of this approach includes i) remaining (and unobserved in the used subset of variables) heterogeneity in patient groups, ii) insufficient information for describing baseline patient state, and iii) the absence of the detailed treatment considerations (e.g., daily doses of antidepressants were not taken into account in this approach). Moreover, the assumptions made in the Section 2 when the data were prepared to the analyses, also

Table 3. *Effects of treatments (no treatment (T0), treatment by two or more antidepressants (T2), treatment with antidepressants and augmentation medicine (T3) vs. treatment by one antidepressant (T1) considered as referent) before and after the correction for the selection bias (i.e., with unit weights and inverse probability weights).*

	CGI Improvement		CGI Severity	
	Single	Recurrent	Single	Recurrent
No correction for the selection bias (i.e., unit weights)				
T0 vs. T1	1.12 (1.04, 1.22), 0.0033	0.67 (0.65, 0.70), <.0001	1.28 (1.20, 1.37), <.0001	0.82 (0.79, 0.85), <.0001
T2 vs. T2	0.99 (0.92, 1.06), 0.7400	1.17 (1.14, 1.21), <.0001	0.87 (0.82, 0.93), <.0001	0.99 (0.96, 1.02), 0.4647
T3 vs. T1	1.19 (1.10, 1.28), <.0001	1.22 (1.18, 1.26), <.0001	1.09 (1.02, 1.16), 0.0139	1.27 (1.23, 1.30), <.0001
After the correction for the selection bias (i.e., IP weights)				
T0 vs. T1	1.16 (1.09, 1.24), <.0001	1.13 (1.09, 1.16), <.0001	1.57 (1.48, 1.66), <.0001	1.09 (1.06, 1.11), <.0001
T2 vs. T2	0.76 (0.71, 0.81), <.0001	0.98 (0.95, 1.01), 0.2640	0.72 (0.69, 0.77), <.0001	0.88 (0.86, 0.91), <.0001
T3 vs. T1	1.19 (1.12, 1.27), <.0001	0.89 (0.87, 0.92), <.0001	0.99 (0.94, 1.05), 0.7089	0.97 (0.95, 1.00), 0.0330

have to be kept in mind when interpreting the results. Therefore, the obtained results are illustrative: they demonstrate the application of the techniques of statistical modeling to the 'big data' collected in MDD patients based on the clinical observations. Although, currently these results should not be clinically interpreted, they allowed us to uncover limitations and obstacles on the way to finally develop such interpretations. For example, a baseline patient state is represented by CGI (severity or improvement) that is subjectively estimated by a physician during the patient visit. This estimate is performed on the basis of multiple measures (e.g., measures represented patient symptoms) that are also available in the dataset. Actually, practicing physician makes his/her decision about specific treatment on the basis of these multiple symptoms in addition to (or even instead of) CGI. Therefore, the using of these measures in the model of treatment decision would be beneficial when detailed stratification of treatment options needs to be used in analysis.

3.1 Further Model Extension

The proportional odds model allows the researcher to address multiple research and practical questions, but this model has limitations because of specific assumptions required for the validity of this model. Two extensions of the model are important because they allow us to soften some assumptions of the model.

One assumption in the base proportional odds model is that the prediction does not depend on the time period between visits. The first alternative model deals with the probability to change CGI (se verity or improvement) during a specific time period. The idea of this modeling is to incorporate unmeasured CGI in certain time periods between visits. Two variants of time period are considered for CGI in intermediate time periods: time periods of one day and one week. Justification for using one week is that the distribution of time between visits has clear maxima in 7 days, 14 days, 21 days, and etc. (i.e., individuals have appointments in the same day of a week). Parameter estimation is based on likelihood maximization. The likelihood function is constructed as follows. First, all time-paths between two measured points are identified and then total path-specific probabilities are calculated. For example, if initially measured CGI is 4 and final CGI measured in three weeks is 5 and if we assume that CGI can change by 1 during one week, then there are 5 unique time-paths: 4445, 4455, 4555, 4565, and 4345. In the simplest model with assumption that improvement and deterioration probabilities (P_- and P_+) does not depend on current CGI, probability of each path is

$$P_{4445} = P_{4455} = P_{4555} = (1 - P_- - P_+)^2 P_+,$$
$$P_{4565} = P_{4345} = P_+^{\,2} P_-$$

The contribution of this measurement to likelihood function is the probability to have exactly 5 in three weeks:

$$L_i = P_{4445} + P_{4455} + P_{4555} + P_{4345} + P_{4563}$$
$$= 3(1 - P_- - P_+)^2 P_+ + 2P_+^{\,2} P_-$$

The total likelihood function is $L = \prod_i L_i$, where i runs over all our statistical units, i.e., over all time periods with measured CGI in the beginning and in the end of the period. The transitional probabilities P_- and P_+ depends on characteristics measured in the beginning of time periods (i.e., on all variables discussed in Section Data). Respective slopes are model parameters, that are estimated by maximizing likelihood function (or, that is more appropriate for computation, logarithm of likelihood function). It is more reasonable to assume that improvement and deterioration probabilities depend also on the initial CGI, i.e., we consider

$$P_{\pm c} = \Pr(\Delta\mathrm{CGI} = \pm 1 \mid \mathrm{CGI}{=}c)\,.$$

Five probabilities in the above example become:

$$P_{4445} = (1 - P_{-4} - P_{+4})^2 P_{+4},$$
$$P_{4455} = (1 - P_{-4} - P_{+4}) P_{+4} (1 - P_{-5} - P_{+5})$$
$$P_{4555} = P_{+4} (1 - P_{-5} - P_{+5})^2,$$
$$P_{4565} = P_{+4} P_{+5} P_{6-},$$
$$P_{4345} = P_{-4} P_{+3} P_{+4.}$$

Similar model involving paths and transitional probabilities was developed by our group and applied to analysis of HPV clearance [14].

The second alternative approach is based on the boosted regression approach (a data mining approach) allowing for avoiding the strict assumptions about the logistic regression shape of the relation of the predicted CGI with predictors. The R package 'gbm' was used for analysis.

4. CONCLUSION

At our best knowledge, the proposed approach based on the MindLinc data is the first attempt to develop and validate the prognostic model based on CGI-I and CGI-S scales for patients with MDD and different comorbidities. Since its developing, CGI-S and CGI-I scales have been used predominantly for evaluation of effectiveness of treatment of MDD in clinical trials. These trials have studied short-term (within several weeks) and long-term (from 6 to 24 months) effects of the prescribed treatments for MDD [15,16] and for combination of MDD with other disorders including insomnia, fibromyalgia, and others [17,18]. However, the numbers of patients in these trials was usually small and the period of observation not long enough for making conclusions about the found associations on a large-scale projection. For developing the tool allowing for making projections on both treatment effectiveness and changes of patient health (both mental and somatic) status large dataset is required with multiple measurements and long follow-up. The one used in this study fits perfect for that purpose. Such rich dataset requires advanced methods of analysis to be applied that will take into account multiple potential confoundings.

5. REFERENCES

[1] Wittchen, H. U., Jacobi, F., Rehm, J., Gustavsson, A., Svensson, M., Jönsson, B., ... & Steinhausen, H. C. (2011). The size and burden of mental disorders and other disorders of the brain in Europe 2010. *European Neuropsychopharmacology*, *21*(9), 655-679.

[2] Kessler, R. C., Berglund, P., Demler, O., Jin, R., Koretz, D., Merikangas, K. R., ... & Wang, P. S. (2003). The

epidemiology of major depressive disorder. *JAMA: the journal of the American Medical Association*, *289*(23), 3095-3105.

[3] Mathers, C. D., & Loncar, D. (2006). Projections of global mortality and burden of disease from 2002 to 2030. *PLoS medicine*, *3*(11), e442.

[4] Merikangas, A. K., Mendola, P., Pastor, P. N., Reuben, C. A., & Cleary, S. D. (2012). The association between major depressive disorder and obesity in US adolescents: results from the 2001–2004 National Health and Nutrition Examination Survey. *Journal of behavioral medicine*, *35*(2), 149-154.

[5] Woodward, A. T., Taylor, R. J., Abelson, J. M., & Matusko, N. (2013). Major depressive disorder among older african americans, caribbean blacks, and non-hispanic whites: secondary analysis of the national survey of american life. *Depression and anxiety*. In press.

[6] Connolly, K. R., & Thase, M. E. (2011). If at First You Don't Succeed. *Drugs*, *71*(1), 43-64.

[7] Sobocki, P., Ekman, M., Ågren, H., Runeson, B., & Jönsson, B. (2006). The mission is remission: health economic consequences of achieving full remission with antidepressant treatment for depression. *International journal of clinical practice*, *60*(7), 791-798.

[8] Garcia-Toro, M., Medina, E., Galan, J., Gonzalez, M., & Maurino, J. (2012). Treatment patterns in major depressive disorder after an inadequate response to first-line antidepressant treatment. *BMC psychiatry*, *12*(1), 143.

[9] Dunlop, B. W., Kelley, M. E., Mletzko, T. C., Velasquez, C. M., Craighead, W. E., & Mayberg, H. S. (2012). Depression beliefs, treatment preference, and outcomes in a randomized trial for major depressive disorder. *Journal of psychiatric research*, *46*(3), 375-381.

[10] Guy W: ECDEU Assessment Manual for Psychopharmacology —Revised (DHEW Publ No ADM 76-338). Rockville, MD, U.S. Department of Health, Education, and Welfare, Public Health Service, Alcohol, Drug Abuse, and Mental Health Administration, NIMH Psychopharmacology Research Branch, Division of Extramural Research Programs, 1976, pp 218–222

[11] Guy W: Clinical Global Impressions (CGI) Scale. Modified From: Rush J, *et al.*: Psychiatric Measures, APA, Washington DC, 2000.

[12] Gersing, K. and Krishnan, R. (2003). Clinical Computing: Clinical Management Research Information System (CRIS). *Psychiatric Services (Washington, D.C.)*. 54, 9 (Sep. 2003), 1199–1200.

[13] Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly weather review*, *78*(1), 1-3.

[14] Kravchenko, J., Akushevich, I., Sudenga, S. L., Wilson, C. M., Levitan, E. B., & Shrestha, S. (2012). Transitional Probability-Based Model for HPV Clearance in HIV-1-Positive Adolescent Females. *PloS one*, *7*(1), e30736.

[15] Mallinckrodt, C. H., Prakash, A., Andorn, A. C., Watkin, J. G., & Wohlreich, M. M. (2006). Duloxetine for the treatment of major depressive disorder: a closer look at efficacy and safety data across the approved dose range. *Journal of psychiatric research*, *40*(4), 337-348;

[16] Robinson, M., Oakes, T. M., Raskin, J., Liu, P., Shoemaker, S., & Nelson, J. C. (2013). Acute and long-term treatment of late-life major depressive disorder: duloxetine versus placebo. *The American Journal of Geriatric Psychiatry*. In press.

[17] Fava, M., McCall, W. V., Krystal, A., Wessel, T., Rubens, R., Caron, J., ... & Roth, T. (2006). Eszopiclone co-administered with fluoxetine in patients with insomnia coexisting with major depressive disorder. *Biological Psychiatry, 59*, 1052-1060

[18] Arnold, L. M., Rosen, A., Pritchett, Y. L., D'Souza, D. N., Goldstein, D. J., Iyengar, S., & Wernicke, J. F. (2005). A randomized, double-blind, placebo-controlled trial of duloxetine in the treatment of women with fibromyalgia with or without major depressive disorder. *Pain*, *119*(1), 5-15.