Novel Quality Control Metric for the Pharmacotherapy of Major Depressive Disorder:

Measuring Guideline Concordance and Its Impact on Symptom Severity

Mason T. Breitzig, MPH; Fan He, PhD; Lan Kong, PhD; Guodong Liu, PhD; Daniel A. Waschbusch, PhD; Jeff D. Yanosky, ScD; Erika F. H. Saunders, MD; and Duanping Liao, MD, PhD

Abstract

Objective: Studies suggest that people with major depressive disorder (MDD) often receive treatment that is not concordant with practice guidelines. To evaluate this, we (1) developed a guideline concordance algorithm for MDD pharmacotherapy (GCA-8), (2) scored it using clinical data, and (3) compared its explanation of patient-reported symptom severity to a traditional concordance measure.

Methods: This study evaluated 1,403 adults (67% female, 85% non-Hispanic/Latino White, mean age 43 years) with non-psychotic MDD (per *ICD-10* codes), from the Penn State Psychiatry Clinical Assessment and Rating Evaluation System (PCARES) registry (visits from February 1, 2015, to April 13, 2021). We (1) scored 1-year concordance using the Canadian Network for Mood and Anxiety Treatments (CANMAT) guidelines and deviation from 8 pharmacotherapy-related criteria and (2) examined associations between concordance and Patient Health Questionnaire depression module (PHQ-9) scores.

Results: The mean GCA-8 score was 6.37 (standard deviation [SD]=1.30; 8.00=perfect concordance). Among those who switched drugs (n=671), 81% (n=542) did not have their dose increased to the recommended maximum before switching. In our adjusted analyses, we found that a 1 SD increase in the GCA-8 was associated with a 0.78 improvement in the mean PHQ-9 score ($P<.001$). The comparison concordance measure was not associated with the mean PHQ-9 score ($\beta=-0.20$; $P=.20$; $R^2=0.53$), and adding the GCA-8 score significantly improved the model ($R^2=0.54$; Vuong test $P=.008$).

Conclusions: By measuring naturalistic MDD pharmacotherapy guideline concordance with the GCA-8, we revealed potential treatment gaps and an inverse association between guideline concordance and MDD symptom severity.

J Clin Psychiatry 2024;85(1):23m14916

Author affiliations are listed at the end of this article.

According to 2019 data provided by the Institute for Health Metrics and Evaluation, depression represents the fifth largest percentage (5%) of years lived with disability in the United States.[1] As a result of health care utilization, loss of productivity, and suicide, major depressive disorder (MDD) is associated with burdensome personal and societal costs, both functionally and financially.[2–4] Still, suboptimal MDD treatment response remains a tremendous challenge for modern medicine.

Robust psychopharmacologic studies (eg, randomized efficacy trials) establish the superiority of myriad MDD pharmacotherapeutics versus placebo,[5–10] and one study even showed that effect sizes in psychiatry were comparable to those in general medicine.[11] Despite this, a trial of sequenced, measurement-driven treatment in the US indicated that two-thirds of patients required more than a first-line antidepressant,[6,12,13] and one-third tried at least 4 antidepressants.[6,13] Furthermore, efficacy trials demonstrate MDD remission rates of 30%–50% within 6–8 weeks,[6,14,15] while effectiveness trials suggest 11%–43% in a year.[6,13,16] Estimated relapse rates range from 20% to 83% within 6–12 months.[4,6,13,17]

Patients with MDD who do not respond to adequate trial (ie, dose and duration) of 2 distinct antidepressants are often considered to have difficult-to-treat depression,[13,18,19] which is associated with a 23% higher risk of all-cause mortality.[20] The prevalence of difficult-to-treat depression

Clinical Points

- Several quality clinical guidelines exist for treating major depressive disorder, but research on concordance with their recommendations is limited, and their link with real-world patient outcomes remains understudied.
- Avoiding 1 guideline discordant pharmacotherapeutic event, during 1 year of treatment, is associated with nearly a 1-point lower mean score on the Patient Health Questionnaire depression module.

(approximately 30%)[13,20] highlights the need for better MDD treatment optimization strategies.[13,19] Indeed, treatment choice is crucial for reducing therapeutic steps, improving outcomes, and avoiding relapse,[13] but is dependent on the needs of each patient.[21–23]

Necessarily, the relative similarity in psychotropic drug-to-drug efficacy[24] and effectiveness,[12,13] the syndromic nature of MDD, and the nascence of predictive strategies[21,22,25,26] lead clinicians to artfully prescribe using observed response, acceptability, and tolerability.[22,24] Various clinical guidelines for the treatment of MDD seek to lend objectivity to this art, but this strategy has its flaws. Studies indicate that clinical guidelines (1) can be derived from a body of trials with a moderate-to-high degree of bias[24,27–31] and potential generalizability issues (eg, via inclusion/exclusion criteria or the use of structured protocols)[32,33] with varying impact on guidelines,[34,35] (2) are not equally applied across practice,[36] and (3) are seldom studied in connection with real-world patient outcomes.[31,36,37]

We propose that some of the aforementioned differences in pharmacotherapy efficacy and effectiveness stem from issues with the generalizability of guideline evidence and guideline concordance in practice. If we take the guidelines to be well constructed, a core assumption of this study, measuring guideline concordance serves two key purposes: to determine (1) whether real-world practice reflects these guidelines (and vice versa) and (2) whether practicing guideline-concordant pharmacotherapy translates to better patient outcomes.

Studies on MDD guideline concordance in the late 1990s to early 2000s showed that, although adherence to clinical guidelines was associated with lower MDD severity[37–39] and hospitalization frequency,[40] researchers operationalized guideline concordance differently.[36,37] Studies often tested whether a patient received any drug, rather than which drug, and whether dosing was minimally adequate, rather than how many times it was changed.[36] They also rarely considered the cross-product of multiple criteria.[36] Consequently, many aspects of guideline concordance remain understudied, especially the link with patient outcomes.[37–40]

In 2011, Duhoux et al[36] conducted a systematic review of MDD guideline concordance between 1990 and 2010 and identified 65 articles, 8 (12%) of which used medical records. Of the reviewed articles, 17 (26%) examined initial prescription rates, 16 (24%) dosing and duration, and 8 (12%) modification, among others.[36,41,42] The observed concordance rates were vastly different.[36] More recent studies, and those investigating the acute phase, reported higher concordance, while those measuring dose and duration, visit frequency, maintenance, and minimal adequacy tended to report lower concordance.[38–40]

To improve and attempt to standardize the study of guideline concordance, as has been called for,[37] as well as further determine whether adherence to guidelines influences patient outcomes, we developed a multidimensional guideline concordance algorithm (GCA-8) based on the Canadian Network for Mood and Anxiety Treatments (CANMAT) guidelines.[43] We leveraged the Penn State Psychiatry Clinical Assessment and Rating Evaluation System (PCARES) registry to test the hypotheses that (1) concordance with CANMAT-derived criteria can be precisely and systematically measured in clinical data, (2a) higher GCA-8 scores are significantly associated with lower patient-reported mean and final MDD symptom severity (via patient health questionnaire depression module [PHQ-9]) scores during a 1-year treatment window, and (2b) the GCA-8 is more strongly associated with patient-reported symptom severity than traditional measurement methods. Through this work, we aim to standardize the measure of MDD pharmacotherapy guideline concordance and, most importantly, explore its role in improving real-world clinical outcomes.

METHODS

PCARES Registry Description

The PCARES registry includes a systematic clinical sample of 3,556 individuals with mental illness.[44] Participants sought mental health care services at a psychiatry and behavioral health outpatient clinic and partial hospitalization program in a central Pennsylvania academic medical center between February 17, 2015, and May 30, 2020.[44] New admissions and returning patients were evaluated by a board-certified psychiatrist or licensed clinical psychologist at each visit and administered patient-reported outcome (PRO) assessments. The registry also includes electronic medical record (EMR) data. Clinicians were instructed in measurement-based care, but no specific clinical guidelines were mandated.[45] As a clinical quality improvement project, the PCARES registry was granted exemption from review by the Penn State College of Medicine Institutional Review Board (IRB; #00019704). This retrospective study was conducted according to the ethical principles of the Declaration of Helsinki and approved by the PCARES Steering Committee and Penn State University College of Medicine IRB (#00020184).

Study Participants

Eligible participants (1) were in the PCARES registry (n = 3,556; individuals with significant cognitive impairment were excluded); (2) were 18 years or older at baseline; (3) had an *ICD-10* diagnosis for MDD within 1 year before and 1 year after their PCARES encounter (February 1, 2015, to April 13, 2021); (4) had at least 1 PHQ-9 and 1 World Health Organization Disability Assessment Schedule (WHODAS 2.0) measure (no minimum score required); and (5) had no diagnosis of bipolar disorder or psychosis within 1 year before baseline.

These criteria resulted in 1,403 eligible participants, for which we calculated the GCA-8 score. However, some individuals had missing data: average MDD episode duration (n = 19), body mass index (BMI; n = 84), insurance info (n = 26; those classified as "other type," eg, self-pay, were treated as missing), marital status (n = 7), municipal rural/urban classification (n = 20), and race and ethnicity (n = 27). Because non-random missingness precluded imputation, we conducted a sensitivity analysis comparing the 162 participants with missing data to the 1,241 with complete data. We only identified a significant difference for insurance type: a greater proportion of excluded individuals had commercial insurance. In light of this, a complete case analysis was used.

Selection of MDD Clinical Guidelines

Several high-quality clinical guidelines exist for the treatment of MDD. We refer readers to the work by Gabriel et al,[30] Zafra-Tanaka et al,[31] and others,[46,47] who summarized, analyzed, and compared MDD treatment guidelines. The GCA-8 is based on evidence that underpins several clinical guidelines, but defers to the 2016 CANMAT guidelines for criteria that some guidelines may not cover (eg, specific duration thresholds).[30,43,48] Because the CANMAT guidelines themselves are congruent with other guidelines, the GCA-8 may still be relevant for clinics where other pharmacotherapy guidelines are applied.[30]

With this being said, we selected the CANMAT guidelines as our primary source for 3 specific reasons. First, CANMAT's evaluation of research quality aligns closely with the international Appraisal of Guidelines, Research, and Evaluation (AGREE II) criteria and the Grading of Recommendations, Assessment, Development, and Evaluation (GRADE) method for the evaluation of source evidence, but favors expert opinion.[43,49,50] We believed CANMAT's emphasis on expert opinion would result in a more accurate real-world assessment of pharmacotherapy guideline concordance.[43] Second, the CANMAT guidelines are more specific than other clinical guidelines,[30,43] as highlighted by Yang et al[51] in a similar study. For example, the CANMAT guidelines made specific drug recommendations rather than referring to drug classes.[43] Lastly,[31,32] because the GCA-8 was developed in a North American setting and focused on pharmacotherapy, North American guidelines that recommended treatment with pharmacotherapy for moderate depression were given preference.[30,31,43]

Guideline Concordance Algorithm (GCA-8)

The novel pharmacotherapy guideline concordance algorithm (GCA-8) is fully described in the supplementary flowchart (Supplementary Appendix 1). Briefly, the ordinal score is generated by 8 criteria: 3 focused on prescription sequence; 3 on dosing, duration, and modifications; 1 on drug-drug interactions (including cytochrome P450 interactions and, conservatively, those that can lead to serotonin syndrome); and 1 on visit frequency. For this study, 1 point was deducted for each failed criterion, during a 1-year treatment window, from the baseline score of 8. Repeatable events were evaluated against the median number of discordant events in the sample, counting only excess occurrences as discordant for the final score (thus avoiding conflation of individual discordance with population/practice level discordance). This study did not evaluate clinical notes and elected to focus on psychopharmacology.

Comparison (Traditional) Concordance Score

A "traditional" concordance score was assembled based on existing literature reviewed by Duhoux et al.[36] Two key metrics were used by several studies: whether an antidepressant was prescribed (1) at baseline and (2) at the minimally adequate dose.[36] Studies also examined (3) treatment duration (90 days, 180 days total, etc) and (4) visit frequency, including 1 visit per month for 3 months and 3 visits every 6 months.[36] Such thresholds were generally selected according to clinical guidelines and the time needed to see a therapeutic response.[43] To be more comparable to the GCA-8, we computed binary responses to these 4 criteria (1 = concordant) and then took the sum to get an ordinal score with a range of 0 to 4. Individuals with fewer than 180 days in the registry (n = 170/1,403) were held to more appropriate thresholds for criteria 3 and 4 to minimize censoring bias: 90 days versus 180 days for treatment duration and 3 visits per 6 months rather than 1 visit per month for 3 months.

PCARES PRO Data

From the PCARES PRO battery,[45] we studied the PHQ-9 and adjusted for the WHODAS.[52] We calculated (1) an arithmetic mean of all PHQ-9 scores in 1 year (except the baseline score; the median number of scores was 4); (2) the final PHQ-9 score in 1 year (median occurrence at 265 days); and (3) the standard deviation (SD) of PHQ-9 scores (including baseline) for those with at least 3 scores. The mean PHQ-9 score was used to minimize the impact of spurious measures and provide a steady indicator of patient status in the 1-year window.

Table 1.

Characteristics of People With Major Depressive Disorder in the PCARES Registry and Unadjusted Differences in GCA-8 (n = 1,403)

	n (%)	Mean (SD)	β Coefficient (SE) or mean GCA-8 score (SD)	*P* value[a]
Age (y)	1,403 (100)	43.06 (17.30)	1.72×10^{-4} (2.01×10^{-3})	.93
Baseline PHQ-9 score (0–27)	1,403 (100)	12.48 (6.83)	−0.03 (0.01)	< .001
Baseline WHODAS score (0–48)	1,403 (100)	15.7 (10.32)	−0.02 (0.003)	< .001
BMI[b]	1,319 (94)	30.32 (7.98)	1.58×10^{-3} (0.005)	.73
Insurance[c]				.11
Commercial	776 (56)		6.42 (1.29)	
Medicaid	252 (18)		6.22 (1.36)	
Medicare	349 (25)		6.38 (1.26)	
Marital status				.29
Divorced	155 (11)		6.21 (1.22)	
Married	577 (41)		6.36 (1.35)	
Separated/widowed	86 (6)		6.35 (1.21)	
Single	578 (41)		6.44 (1.29)	
Municipal rurality				.51
Rural	193 (14)		6.31 (1.29)	
Urban	1,190 (86)		6.38 (1.31)	
Patient-reported gender				.01
Female	945 (67)		6.31 (1.32)	
Male	458 (33)		6.50 (1.26)	
Race and ethnicity				.17
All others[d]	56 (4)		6.68 (1.21)	
Hispanic or Latino	72 (5)		6.17 (1.37)	
Non-Hispanic/Latino Black	76 (6)		6.32 (1.48)	
Non-Hispanic/Latino White	1,172 (85)		6.37 (1.29)	

[a]Test for unadjusted linear association with the continuous score, such that the β estimate is significantly different from zero for continuous variables or the group means demonstrate significant inequality.
[b]Continuous BMI values lying beyond the 99th percentile at baseline were trimmed.
[c]All miscellaneous types (eg, self-pay) that did not fall under the categories of commercial, Medicaid, or Medicare were excluded (n = 21).
[d]Where "all others" represents those identifying as Alaska Native, Asian, Native American, Native Hawaiian or Pacific Islander, 2 or more races, or another group, or where multiple ethnicities were identified.
Abbreviations: BMI = body mass index, GCA-8 = guideline concordance algorithm, PCARES registry = Penn State Psychiatry Clinical Assessment and Rating Evaluation System, PHQ-9 = Patient Health Questionnaire depression module, SD = standard deviation, SE = standard error, WHODAS = World Health Organization Disability Assessment Schedule 2.0.

The final score in the treatment window represented symptom status temporally after most clinical decisions. Lastly, the SD of the PHQ-9 scores was used to examine measure-to-measure variability. PHQ-9 and WHODAS questionnaires missing more than 1 answer were excluded (n = 235, 1.23% of all PCARES PHQ-9 measures and n = 397, 2.10% of all PCARES WHODAS measures). Mean imputation was used to address surveys with 1 missing value (n = 672, 3.51% of all PCARES PHQ-9 measures and n = 1,197, 6.33% of all PCARES WHODAS measures).[53]

PCARES EMR Data

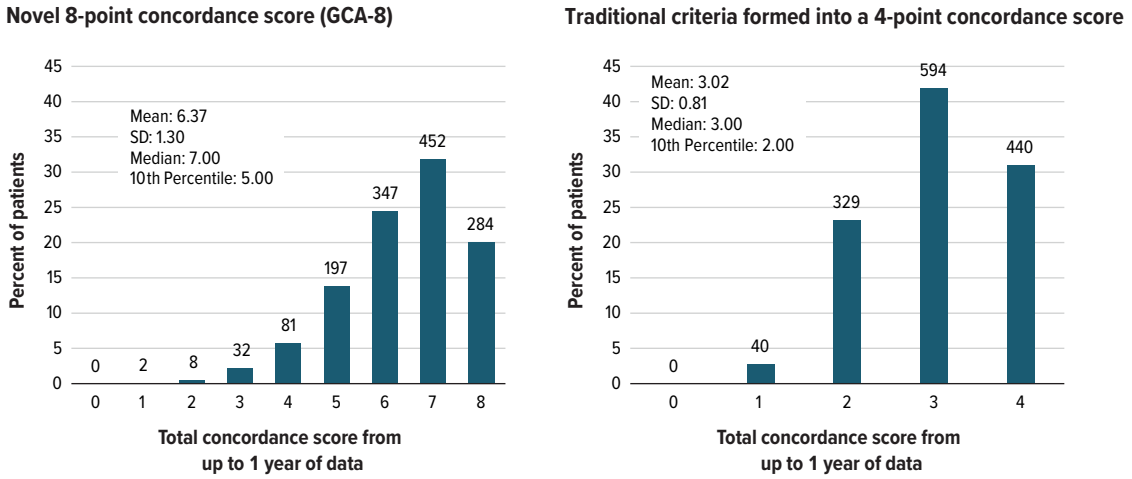
PCARES EMR data consist of self-reported demographic information, anthropometrics, diagnoses (*ICD-10* codes), prescription drug data (eg, generic name, dose, frequency, etc), and blood laboratory results. Patients' municipal-level rural/urban status was determined for the PCARES registry based on the population being less than the mean state density of 284 people per square mile, or less than 2,500 with < 50% residing in an urbanized area.[54]

Statistical Analyses

To facilitate regression coefficient comparison, the GCA-8 score and traditional score were divided by their respective sample SD (n = 1,403). Multivariable linear regression was used to determine the added explanatory value of the GCA-8 over models that included the traditional score, sociodemographic characteristics, and the baseline measures (eg, PHQ-9). Model differences were tested using the Vuong closeness test. SAS 9.4 was used to perform statistical analyses with an α of 0.05 for statistical significance.

Figure 1.

Distribution of Guideline Concordance Scores in Patients With MDD From the PCARES Registry in 1 Year (n = 1,403)



Abbreviations: GCA-8 = guideline concordance algorithm, MDD = major depressive disorder, PCARES registry = Penn State Psychiatry Clinical Assessment and Rating Evaluation System, SD = standard deviation.

RESULTS

Baseline Characteristics of Study Participants

The study sample of 1,403 individuals with MDD had a mean age of 43.06 ± 17.30 years, was 67% female, and was 85% non-Hispanic/Latino White at baseline (Table 1). Participants tended to be married (41%), live in an urban municipality (86%), and have commercial insurance (56%). The average BMI was 30.32 ± 7.98; the mean PHQ-9 score at baseline was 12.48 ± 6.83, indicating moderate depression; and the mean WHODAS score at baseline was 15.7 ± 10.32, indicating some functional disability. In the 1-year window, participants contributed a median of 7 months of data or 5 visits with MDD prescription events. Before baseline, 87% were prescribed MDD-related prescriptions, while 53% received a prescription for MDD at baseline, 61% within 2 weeks of baseline, and 82% within 90 days of baseline. Of the full sample, 40 (2.85%) patients were enrolled in PCARES as new admissions (no records before encounter). For sociodemographic characteristics, mean GCA-8 scores were significantly higher for males at 6.50 (SD = 1.26; $P = .01$) and decreased by 0.03 points for every 1-point increase in baseline PHQ-9 score (SD = 0.01; $P < .001$).

Guideline Concordance Score Distribution

The mean score for the GCA-8 was 6.37/8.00 (SD = 1.30; Figure 1). In order of highest to lowest discordance proportion (Table 2), criterion 4 showed that 542 (81%) of the 671 (48%) who had a drug switch did not reach the maximum recommended dose for the previous treatment. Only 42% were prescribed medications/doses for the recommended durations with limited interruption (defined, using order/stop dates, as a gap < 31 days; criterion 5). Criterion 7 indicated that 23% were concurrently prescribed drugs susceptible to drug-drug interactions, and criterion 6 showed that 20% of those who switched had no modification. Finally, 18% did not have a sufficient visit frequency (criterion 8), while criteria 1–3 demonstrated discordance proportions of less than 9%. Overall, 284 (20%) PCARES patients received a score of 8/8 (Figure 1).

Traditional Score Distribution

The mean for the traditional score was 3.02 (SD = 0.81; Figure 1). In order of highest to lowest discordance proportion (Table 2), criterion 1 showed that 47% of patients did not have an antidepressant prescribed at the baseline visit. Criterion 3 indicated that 27% did not have their treatment continued (per start/stop dates) for a sufficient duration. Of those who were prescribed an antidepressant, and had dose information (n = 740), 81% were prescribed a drug at or above the minimum adequate dose (criterion 2). Finally, criterion 4 showed that 14% did not meet the visit frequency guidelines. Overall, using the traditional concordance criteria, 440 (31%) PCARES patients received a concordance score of 4/4 (Figure 1).

Comparison of the Two Scores and the Added Contribution of the Novel GCA-8

Unadjusted analysis showed significant associations between the GCA-8 and PHQ-9 metrics (Table 3). In

Table 2.

Comparative Frequencies of Criteria Failures in Patients With MDD From the PCARES Registry (n = 1,403)

	GCA-8			Reconstructed 4-point score of traditional criteria		
Domain	Criterion	Concordance, n (%)	Discordance, n (%)	Criterion	Concordance, n (%)	Discordance, n (%)
Initial treatment and drug sequence	1. The initial Rx is a first-line (eg, SSRI), or a second-line if the patient has recurrent MDD.	1,377 (98.15)	26 (1.85)	1. The patient is prescribed any antidepressant at the baseline visit (including augmenting medications).	744 (53.03)	659 (46.97)
	2. The initial second-line Rx is after no more than 2 first-line drugs.	1,283 (91.45)	120 (8.55)			
	3. The initial third-line Rx (eg, MAOI) is after a second-line drug.	1,366 (97.36)	37 (2.64)			
Dose and switching	4. The Rx dose is increased to the recommended maximum before switching to a new drug (no. of discordant events ≤ PCARES median of 0) *and* the total number of switches is not more than the median (no. of switches ≤ PCARES median of 1).	820 (58.45)	583 (41.55)	2. At least 1 of the medications prescribed at baseline is listed at or above the minimum adequate dose for treating MDD (n = 740 had dose info).	600 (81.08)	140 (18.92)
	The Rx dose is increased to the recommended maximum before switching, among those who switched (n = 671)	129 (19.23)	542 (80.77)			
Duration and consistency	5. Treatment duration should be at least 4 weeks per Rx, 2 weeks per dose (no. of discordant events ≤ PCARES median of 0; tapering excluded), with an average gap between prescriptions of less than 31 days and a limited number of gaps (no. of gaps ≤ PCARES median of 1).	593 (42.27)	810 (57.73)	3. The treatment duration should be at least 180 days for any given treatment, or 90 days for those who had fewer than 180 days in the PCARES registry.	1,025 (73.06)	378 (26.94)
Modification	6. Treatment should be modified by changing the dose (tapering excluded) or adding an adjunctive/augmenting medication, before switching to a new drug (no. of discordant events ≤ PCARES median of 0).	1,270 (90.52)	133 (9.48)			
	Among those who switched drugs (n = 671).	538 (80.18)	133 (19.82)			
Interactions	7. The Rx combo is without notable drug-drug interactions (no. of interactions ≤ PCARES median of 0; eg, cytochrome P450).	1,080 (76.98)	323 (23.02)			
Visit frequency	8. The psychiatric visit frequency should be at least 3 visits every 6 months.	1,152 (82.11)	251 (17.89)	4. Patients should visit once per month for 3 months, or 3 visits in 6 months for those who had fewer than 180 days in the system.	1,208 (86.10)	195 (13.90)
Spearman rank correlation between the 2 concordance scores (n = 1,403)					−0.17 (*P* < .001)	
Spearman rank correlation between the 2 concordance scores (n = 1,241)					−0.15 (*P* < .001)	

Abbreviations: GCA-8 = guideline concordance algorithm, MAOIs = monoamine oxidase inhibitors, MDD = major depressive disorder, PCARES registry = Penn State Psychiatry Clinical Assessment and Rating Evaluation System, SSRIs = selective serotonin reuptake inhibitors, Rx = prescription drug (Rx).

contrast, the traditional score was not significantly associated with any of the 3 PHQ-9 metrics in unadjusted analyses. After adjusting for other covariates, including baseline symptom severity, the multivariable models (Table 4, Model 1) explained substantially more variability for all 3 outcome measures. However, the traditional concordance score remained a nonsignificant contributor. Adding the GCA-8 (Table 4, Model 2) not only explained more outcome variability but also demonstrated that the GCA-8 was significantly associated with all 3 outcomes. For example, 1 SD increase in the GCA-8 was associated with a 0.78 decrease in the mean PHQ-9 score (standard error [SE] = 0.15; $P < .001$); the traditional score was marginally nonsignificant (β = −0.28; SE = 0.15; $P = .06$).

The final model for mean PHQ-9 score explained 54% of the response variability and was estimated to be significantly closer (an increase of 0.01 in the R^2) to a true model of the underlying process than the model without the GCA-8 (Vuong test $P = .008$). A similar GCA-8 finding was reported for the final PHQ-9 score (β = −0.80; SE = 0.18; $P < .001$) with the traditional score remaining nonsignificant (β = −0.16; SE = 0.19; $P = .40$). This model explained 41% of the variability in the final PHQ-9 score and was also significantly favored (Vuong test $P = .04$). Lastly, 1 SD increase in the GCA-8 score was associated with a 0.27 decrease in PHQ-9 SD (SE = 0.08; $P < .001$). The final model explained only 9% of the variability and was a nonsignificant improvement on the model without the GCA-8 (Vuong test $P = .11$).

Table 3.
Unadjusted Associations Between Standardized Scores of Guideline Concordance and PHQ-9 Scores[a]

			Unadjusted model			
Outcomes	Predictor of interest	Sample size	Regression coefficient	SE	Coefficient *P* value	R^2
Mean PHQ-9 score in 1 year (excluding baseline measure)	GCA-8	958	−1.68	0.19	<.001	0.07
	4-point		0.33	0.22	.12	0.002
Final PHQ-9 score in 1 year	GCA-8	958	−1.65	0.21	<.001	0.06
	4-point		0.42	0.23	.07	0.003
SD of 3+ PHQ-9 scores in 1 year[b]	GCA-8	741	−0.25	0.07	<.001	0.01
	4-point		0.05	0.08	.54	<0.001

[a]The regression coefficient represents slope; mean change in outcome variable per one standard deviation (SD) increase in the given guideline concordance score; scores were standardized before modeling by dividing the variable of interest by the SD of the given predictor as calculated using the entire cohort (n = 1,403).
[b]Including the baseline measure.
Abbreviations: GCA-8 = guideline concordance algorithm, PHQ-9 = Patient Health Questionnaire depression module.

DISCUSSION

In this study, we evaluated a novel scoring framework for MDD pharmacotherapy guideline concordance (GCA-8) using 1 year of clinical data (Figure 1). Similar to previous studies, which exhibited concordance with first-line recommendations to be 62%–79%,[36] or closer to 96% in a more recent study in China,[51] we noted no major discordance for drug sequence (concordance ≥ 91%), drug-drug interactions (77%), or visit frequency (82%) in PCARES.[36] On the other hand, we observed high discordance for treatment optimization.

According to the STAR*D trial, identifying an optimal therapeutic effect necessitated more aggressive dosing.[13,55,56] However, a 2017 German observational study reported that 86% (n = 132/153) of continuously monitored participants did not have their dose increased and 15% (n = 52/344) received doses that were lower than recommended.[57] In our study, only 20% of participants who switched did not have their drug regimen modified, but 81% of participants did not reach a maximum recommended dose before switching.

Although we only observed switching for 48% of PCARES participants with MDD, this proportion is corroborated by other studies: Hepner et al[39] reported that only 38% of 1,131 US primary care patients with MDD received a treatment adjustment after 6 months of nonresponse, and Yang et al[51] reported a change of medication for only 16% of the 19,955 Chinese patients with MDD in their study.

Insights into treatment duration were another key takeaway from the STAR*D trial, finding that although peak response and remission rates were achieved after 6 weeks of treatment with citalopram, treatment for as long as 12 weeks may be necessary.[13,42,56] CANMAT guidelines suggest switching no sooner than 2–4 weeks after initiation.[58] In PCARES, although 73% had a sufficient overall treatment duration (traditional criterion 3), 58% had at least 1 independent drug or dose step where the duration was inadequate (< 2–4 weeks) or inconsistent (gap ≥ 31 days).

All criteria considered, our results show that dose and treatment duration, as well as modification, are key components of guideline-derived clinical decision-making. Both dose and treatment duration are easily impacted by access to medications (eg, cost), physical tolerance, and contraindications.[58] Acknowledging that clinical situations necessitate unique approaches, and stepped care systems may be challenging to implement, our findings corroborate that pharmacotherapy optimization strategies may need to be more aggressive with dosing and persistent with duration and should attempt modification before switching.[13,42,56]

Aside from its distribution in real-world practice, a crucial question about guideline concordance is its link with patient outcomes. If adherence to guidelines does not influence patient outcomes, then it may be argued that the current guidelines are of less value to clinicians. Fortney et al showed that of 106 US participants with depression, 29% received guideline-concordant care (defined as having been taking an antidepressant for 75% or more of an 8-week period [patient reported] or 8 or more visits over 12 weeks) and were likely to have a significant improvement in depression severity ($P = .02$).[38] Hepner et al demonstrated that higher quality care, as measured by their depression quality index (DQI), was associated with lower odds of scoring higher for symptom severity at 18 and 24 months (OR [95% CI]: 0.64 [0.41, 0.98] and 0.59 [0.38, 0.91], respectively).[39] Lastly, a meta-analysis of randomized and non-randomized experimental designs found that guideline-concordant treatment for psychiatric conditions, including depression, was superior to treatment as usual.[37]

In our adjusted analyses, avoiding 1 discordant event (increasing the concordance score by 1.3) was associated with nearly a 1-point lower mean PHQ-9 score (0.78; Table 4). This finding was independent of the baseline PHQ-9 score. We attribute the significant improvement of the GCA-8 over the traditional measure to the (1) direct measurement of treatment events (ie, repeatable events) from EMR data, (2) inclusion of multiple treatment steps (ie, initial pharmacotherapy versus second step), and (3) measurement of less frequently analyzed constructs, such as modification and drug-drug interactions.[36]

Overall, our results supported our hypotheses, demonstrating (1) concordance with CANMAT-derived guidelines could be measured in real-world clinical data with

Table 4.

Adjusted Associations Between Standardized Scores of Guideline Concordance and Patient Outcomes Measures[a]

Outcomes (excluding baseline measure)	Predictor of interest	Sample size	Model 1 Covariables + traditional score: Regression coefficient	SE	Coefficient *P* value	R^2	Model 2 Model 1 + GCA-8: Regression coefficient	SE	Coefficient *P* value	R^2	*P* value for model fit[b]
Mean PHQ-9 score in 1 year	GCA-8	958	...	...	...	0.53	−0.78	0.15	<.001	0.54	.008 Favoring model 2
	4-point		−0.20	0.15	.20		−0.28	0.15	.06		
	Severity		3.03	0.19	<.001		2.95	0.19	<.001		
Final PHQ-9 score in 1 year	GCA-8	958	...	...	...	0.39	−0.80	0.18	<.001	0.41	.04 Favoring model 2
	4-point		−0.07	0.19	.71		−0.16	0.19	.40		
	Severity		2.76	0.23	<.001		2.69	0.23	<.001		
The SD of 3+ PHQ-9 scores in 1 year[c]	GCA-8	741	...	...	...	0.07	−0.27	0.08	<.001	0.09	.11
	4-point		0.04	0.08	.65		0.004	0.08	.96		

[a]The regression coefficient represents slope; mean change in outcome variable per 1 standard deviation increase in the predictor; scores were standardized before modeling by dividing the variable of interest by the SD of the given predictor as calculated using the entire cohort (n = 1,403); model 1: traditional concordance score and age, average MDD episode duration (defined as the average duration between MDD episodes, detailed further in the supplementary methods [Supplementary Appendix 1]), BMI, insurance type, marital status, municipal rural/urban classification, patient-reported gender, race and ethnicity, standardized baseline PHQ-9 scores (as appropriate), standardized baseline WHODAS scores, whether MDD was recurrent or non-recurrent at baseline (per *ICD-10* codes), and cohort. The cohort variable represents the wave of recruitment into the PCARES registry based on the forms that were collected during that period, effectively making it a temporal adjustment; model 2: all variables in Model 1 + GCA-8.

[b]Vuong's closeness test, Akaike adjusted, assessing whether 2 models are equally close to the true data generating process. Quadratic and cubic terms for the 8-point guideline concordance score were examined and found to be nonsignificant for all outcomes.

[c]Also adjusted for 1-year score difference.

Abbreviations: BMI = body mass index, GCA-8 = guideline concordance algorithm, *ICD-10* = International Statistical Classification of Diseases and Related Health Problems Tenth Revision, MDD = major depressive disorder, PCARES registry = Penn State Psychiatry Clinical Assessment and Rating Evaluation System, PHQ-9 = Patient Health Questionnaire depression module, SD = standard deviation, SE = standard error, 4-point score = traditional guideline concordance score, WHODAS = World Health Organization Disability Assessment Schedule 2.0.

a more systematic and nuanced approach; (2a) the GCA-8 was significantly, and inversely, associated with PHQ-9 scores, and (2b) statistically significant after adjusting for confounders and the traditional concordance criteria. These findings are supported by a large and clinically representative sample (a wide range of MDD severity and pharmacotherapies).

This study has several limitations. Temporality and causality cannot be established. The PCARES registry only includes those who seek treatment in a specific geographic region, and, as such, has limited generalizability. Researchers or clinicians wishing to apply the GCA-8 outside the US, or in a clinical environment where other guidelines are likely to be used, should expect a different GCA-8 score distribution. Due to registry limitations and design choices, there is no information on patient adherence, psychotherapy, providers, smoking, out-of-network treatment, or treatment received before 2015. These, and other sources of residual confounding, may impact the score distributions and further limit generalizability. Finally, misclassification may occur due to diagnostic errors or lack of treatment context.

Herein we demonstrated that the GCA-8 affords a multidimensional approach to measuring pharmacotherapy guideline concordance that is significantly associated with patients' symptom severity. Our findings also corroborate the evidence that, despite advances, dosing and treatment duration could be further strengthened. Future work should focus on longitudinal discordance patterns and the association between guideline concordance and patient functioning. We encourage researchers to apply the GCA-8 to other populations with MDD. Studying the distribution of the GCA-8 in a variety of populations with MDD will improve the interpretation and generalizability of its score and help elucidate the relationship between real-world guideline application and patient outcomes. Tools like the GCA-8 will encourage standardized measurement of clinical processes and help systematically inform both guidelines and practice.

Article Information

Published Online: January 3, 2024. https://doi.org/10.4088/JCP.23m14916

Submitted: May 2, 2023; accepted October 16, 2023.

To Cite: Breitzig MT, He F, Kong L, et al. Novel quality control metric for the pharmacotherapy of major depressive disorder: measuring guideline concordance and its impact on symptom severity. *J Clin Psychiatry.* 2024;85(1):23m14916

Author Affiliations: Department of Public Health Sciences, Penn State College of Medicine, Hershey, Pennsylvania (Breitzig, He, Kong, Liu, Yanosky, Liao); Department of Psychiatry and Behavioral Health, Penn State College of Medicine and Penn State Milton S. Hershey Medical Center, Pennsylvania (Liu, Waschbusch, Saunders).

Corresponding Author: Duanping Liao, MD, Penn State College of Medicine and Penn State Milton S Hershey Medical Center, 700 HMC Cres Rd, Hershey, PA 17033 (dliao@pennstatehealth.psu.edu).

Author Contributions: All listed authors made substantial contributions to the work within this manuscript, including contributions to its conception and design or analysis and interpretation, as well as its written form. All authors participated in the revision of the manuscript and approved its final version.

Relevant Financial Relationships: All authors declare that there are no conflicts of interest.

Funding/Support: None.

Previous Presentation: The study was presented at the American Public Health Association annual conference under the same title on November 14, 2023, in Atlanta, Georgia.

Acknowledgments: The authors thank the PCARES participants, without whom this study would not be possible.

Supplementary Material: Available at Psychiatrist.com.

References

1. Institute for Health Metrics and Evaluation (IHME). GBD Compare Data Visualization. IHME, University of Washington. Published 2020. Accessed March 26, 2023. https://vizhub.healthdata.org/gbd-compare
2. Hammer-Helmich L, Haro JM, Jönsson B, et al. Functional impairment in patients with major depressive disorder: the 2-year PERFORM study. *Neuropsychiatr Dis Treat*. 2018;14:239–249.
3. Greenberg PE, Fournier AA, Sisitsky T, et al. The economic burden of adults with major depressive disorder in the United States (2010 and 2018). *PharmacoEconomics*. 2021;39(6):653–665.
4. Severe J, Greden JF, Reddy P. Consequences of recurrence of major depressive disorder: is stopping effective antidepressant medications ever safe? *Focus Am Psychiatr Publ*. 2020;18(2):120–128.
5. Thase ME, Mahableshwarkar AR, Dragheim M, et al. A meta-analysis of randomized, placebo-controlled trials of vortioxetine for the treatment of major depressive disorder in adults. *Eur Neuropsychopharmacol*. 2016;26(6):979–993.
6. Saragoussi D, Touya M, Haro JM, et al. Factors associated with failure to achieve remission and with relapse after remission in patients with major depressive disorder in the PERFORM study. *Neuropsychiatr Dis Treat*. 2017;13:2151–2165.
7. Carrasco JL, Kornstein SG, McIntyre RS, et al. An integrated analysis of the efficacy and safety of desvenlafaxine in the treatment of major depressive disorder. *Int Clin Psychopharmacol*. 2016;31(3):134–146.
8. Liebowitz MR, Yeung PP, Entsuah R. A randomized, double-blind, placebo-controlled trial of desvenlafaxine succinate in adult outpatients with major depressive disorder. *J Clin Psychiatry*. 2007;68(11):1663–1672.
9. DeMartinis NA, Yeung PP, Entsuah R, et al. A double-blind, placebo-controlled study of the efficacy and safety of desvenlafaxine succinate in the treatment of major depressive disorder. *J Clin Psychiatry*. 2007;68(5):677–688.
10. Fournier JC, DeRubeis RJ, Hollon SD, et al. Antidepressant drug effects and depression severity: a patient-level meta-analysis. *JAMA*. 2010;303(1):47–53.
11. Leucht S, Hierl S, Kissling W, et al. Putting the efficacy of psychiatric and general medicine medication into perspective: review of meta-analyses. *Br J Psychiatry*. 2012;200(2):97–106.
12. Gaynes BN, Rush AJ, Trivedi MH, et al. The STAR*D study: treating depression in the real world. *Cleve Clin J Med*. 2008;75(1):57–66.
13. Rush AJ, Trivedi MH, Wisniewski SR, et al. Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR*D report. *Am J Psychiatry*. 2006;163(11):1905–1917.
14. Kennedy SH, Andersen HF, Thase ME. Escitalopram in the treatment of major depressive disorder: a meta-analysis. *Curr Med Res Opin*. 2009;25(1):161–175.
15. Thase ME, Nierenberg AA, Vrijland P, et al. Remission with mirtazapine and selective serotonin reuptake inhibitors: a meta-analysis of individual patient data from 15 controlled trials of acute phase treatment of major depression. *Int Clin Psychopharmacol*. 2010;25(4):189–198.
16. Rush AJ, Trivedi M, Carmody TJ, et al. One-year clinical outcomes of depressed public sector outpatients: a benchmark for subsequent studies. *Biol Psychiatry*. 2004;56(1):46–53.
17. Hengartner MP. How effective are antidepressants for depression over the long term? A critical review of relapse prevention trials and the issue of withdrawal confounding. *Ther Adv Psychopharmacol*. 2020;10:2045125320921694.
18. Singh JB, Fedgchin M, Daly EJ, et al. A double-blind, randomized, placebo-controlled, dose-frequency study of intravenous ketamine in patients with treatment-resistant depression. *Am J Psychiatry*. 2016;173(8):816–826.
19. Fava M, Davidson KG. Definition and epidemiology of treatment-resistant depression. *Psychiatr Clin North Am*. 1996;19(2):179–200.
20. Lundberg J, Cars T, Lööv SÅ, et al. Association of treatment-resistant depression with patient outcomes and health care resource utilization in a population-wide study. *JAMA Psychiatry*. 2023;80(2):167–175.
21. Perlis RH. Abandoning personalization to get to precision in the pharmacotherapy of depression. *World Psychiatry*. 2016;15(3):228–235.
22. Simon GE, Perlis RH. Personalized medicine for depression: can we match patients with treatments? *Am J Psychiatry*. 2010;167(12):1445–1455.
23. Simon G. Choosing a first-line antidepressant: equal on average does not mean equal for everyone. *JAMA*. 2001;286(23):3003–3004.
24. Cipriani A, Furukawa TA, Salanti G, et al. Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. *Lancet*. 2018;391(10128):1357–1366.
25. van Bronswijk SC, DeRubeis RJ, Lemmens LHJM, et al. Precision medicine for long-term depression outcomes using the Personalized Advantage Index approach: cognitive therapy or interpersonal psychotherapy? *Psychol Med*. 2021;51(2):279–289.
26. Cipriani A, Geddes J. Predicting treatment outcome in depression: so far, so good. *Lancet Psychiatry*. 2016;3(3):192–194.
27. Deshauer D, Moher D, Fergusson D, et al. Selective serotonin reuptake inhibitors for unipolar depression: a systematic review of classic long-term randomized controlled trials. *CMAJ*. 2008;178(10):1293–1301.
28. Turner EH, Matthews AM, Linardatos E, et al. Selective publication of antidepressant trials and its influence on apparent efficacy. *N Engl J Med*. 2008;358(3):252–260.
29. Driessen E, Hollon SD, Bockting CLH, et al. Does publication bias inflate the apparent efficacy of psychological treatment for major depressive disorder? a systematic review and meta-analysis of US National Institutes of Health-funded trials. *PLoS One*. 2015;10(9):e0137864.
30. Gabriel FC, de Melo DO, Fráguas R, et al. Pharmacological treatment of depression: a systematic review comparing clinical practice guideline recommendations. *PLoS One*. 2020;15(4):e0231700. doi:10.1371/journal.pone.0231700
31. Zafra-Tanaka JH, Goicochea-Lugo S, Villarreal-Zegarra D, et al. Characteristics and quality of clinical practice guidelines for depression in adults: a scoping review. *BMC Psychiatry*. 2019;19(1):76.
32. Malmivaara A. Generalizability of findings from randomized controlled trials is limited in the leading general medical journals. *J Clin Epidemiol*. 2019;107:36–41.
33. Blanco C, Olfson M, Goodwin RD, et al. Generalizability of clinical trial results for major depression to community samples: results from the National Epidemiologic Survey on Alcohol and Related Conditions. *J Clin Psychiatry*. 2008;69(8):1276–1280.
34. van der Lem R, de Wever WW, van der Wee NJ, et al. The generalizability of psychotherapy efficacy trials in major depressive disorder: an analysis of the influence of patient selection in efficacy trials on symptom outcome in daily practice. *BMC Psychiatry*. 2012;12(1):192.
35. von Wolff A, Jansen M, Hölzel LP, et al. Generalizability of findings from efficacy trials for chronic depression: an analysis of eligibility criteria. *Psychiatr Serv*. 2014;65(7):897–904.
36. Duhoux A, Fournier L, Menear M. Quality indicators for depression treatment in primary care: a systematic literature review. *Curr Psychiatry Rev*. 2011;7(2):104–137.
37. Setkowski K, Boogert K, Hoogendoorn AW, et al. Guidelines improve patient outcomes in specialised mental health care: a systematic review and meta-analysis. *Acta Psychiatr Scand*. 2021;144(3):246–258.
38. Fortney J, Rost K, Zhang M, et al. The relationship between quality and outcomes in routine depression care. *Psychiatr Serv*. 2001;52(1):56–62.
39. Hepner KA, Rowe M, Rost K, et al. The effect of adherence to practice guidelines on depression outcomes. *Ann Intern Med*. 2007;147(5):320–329. https://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med6&NEWS=N&AN=17785487
40. Sewitch MJ, Blais R, Rahme E, et al. Receiving guideline-concordant pharmacotherapy for major depression: impact on ambulatory and inpatient health service use. *Can J Psychiatry*. 2007;52(3):191–200.
41. Simon GE, Von Korff M, Rutter CM, et al. Treatment process and outcomes for managed care patients receiving new antidepressant prescriptions from psychiatrists and primary care physicians. *Arch Gen Psychiatry*. 2001;58(4):395–401.
42. Katon W, Von Korff M, Lin E, et al. Collaborative management to achieve treatment guidelines. impact on depression in primary care. *JAMA*. 1995;273(13):1026–1031.
43. Kennedy SH, Lam RW, McIntyre RS, et al; CANMAT Depression Work Group. Canadian Network for Mood and Anxiety Treatments (CANMAT) 2016 clinical guidelines for the management of adults with major depressive disorder: section 3. pharmacological treatments. *Can J Psychiatry*. 2016;61(9):540–560.
44. Saunders EFH, Brady M, Mukherjee D, et al. Gender differences in transdiagnostic domains and function of adults measured by *DSM-5* assessment scales at the first clinical visit: a cohort study. *BMC Psychiatry*. 2023;23(1):709.
45. Gomaa H, Baweja R, Mukherjee D, et al. Transdiagnostic and functional predictors of depression severity and trajectory in the Penn State Psychiatry Clinical Assessment and Rating Evaluation System (PCARES) registry. *J Affect Disord*. 2022;298(Pt A):86–94.
46. Davidson JRT. Major depressive disorder treatment guidelines in America and Europe. *J Clin Psychiatry*. 2010;71(suppl E1):e04.

47. Wang HR, Bahk WM, Seo JS, et al. Korean medication algorithm for depressive disorder: comparisons with other treatment guidelines. *Clin Psychopharmacol Neurosci*. 2017;15(3):199–209.
48. American Psychiatric Association. Practice Guideline For the Treatment of Patients With Major Depressive Disorder. Published November 2010. Accessed August 24, 2023. https://psychiatryonline.org/guidelines
49. Brouwers MC, Kerkvliet K, Spithoff K; AGREE Next Steps Consortium. The AGREE Reporting Checklist: a tool to improve reporting of clinical practice guidelines. *BMJ*. 2016;352:i1152.
50. Malmivaara A. Methodological considerations of the GRADE method. *Ann Med*. 2015;47(1):1–5.
51. Yang L, Su Y, Dong S, et al. Concordance of the treatment patterns for major depressive disorders between the Canadian Network for Mood and Anxiety Treatments (CANMAT) algorithm and real-world practice in China. *Front Pharmacol*. 2022;13:954973.
52. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med*. 2001;16(9):606–613.
53. Kroenke K, Spitzer RL, Williams JBW, et al. The Patient Health Questionnaire Somatic, Anxiety, and Depressive Symptom Scales: a systematic review. *Gen Hosp Psychiatry*. 2010;32(4):345–359.
54. The Center for Rural Pennsylvania. Rural-Urban Definitions. Published 2020. https://www.rural.pa.gov/data/rural-urban-definitions
55. Trivedi MH, Rush AJ, Wisniewski SR, et al; STAR*D Study Team. Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: implications for clinical practice. *Am J Psychiatry*. 2006;163(1):28–40.
56. Trivedi MH, Rush AJ, Gaynes BN, et al. Maximizing the adequacy of medication treatment in controlled trials and clinical practice: STAR(*)D measurement-based care. *Neuropsychopharmacology*. 2007;32(12):2479–2489.
57. Herzog DP, Wagner S, Ruckes C, et al. Guideline adherence of antidepressant treatment in outpatients with major depressive disorder: a naturalistic study. *Eur Arch Psychiatry Clin Neurosci*. 2017;267(8):711–721.
58. Kennedy SH, Lam RW, McIntyre RS, et al; CANMAT Depression Work Group. Canadian Network for Mood and Anxiety Treatments (CANMAT) 2016 clinical guidelines for the management of adults with major depressive disorder: section 3. pharmacological treatments. *Can J Psychiatry*. 2016;61(9):540–560.

The Journal of Clinical Psychiatry

Supplementary Material

Article Title: Novel Quality Control Metric for the Pharmacotherapy of Major Depressive Disorder: Measuring Guideline Concordance and Its Impact on Symptom Severity

Authors: Mason T. Breitzig, MPH; Fan He, PhD; Lan Kong, PhD; Guodong Liu, PhD; Daniel A. Waschbusch, PhD; Jeff D. Yanosky, ScD; Erika F. H. Saunders, MD; and Duanping Liao, MD, PhD

DOI Number: 10.4088/JCP.23m14916

LIST OF SUPPLEMENTARY MATERIAL FOR THE ARTICLE

1. Appendix 1 Guideline Concordance Algorithm (GCA-8)

DISCLAIMER

This Supplementary Material has been provided by the author(s) as an enhancement to the published article. It has been approved by peer review; however, it has undergone neither editing nor formatting by in-house editorial staff. The material is presented in the manner supplied by the author.

Appendix 1 - Guideline Concordance Algorithm (GCA-8)

The eight criteria that form the GCA-8 are first derived from general practice guidelines common to many major depressive disorder (MDD) treatment guidelines then, where specifics are needed, from the 2016 Canadian Network for Mood and Anxiety Treatments (CANMAT) guidelines (*e.g.*, which drugs are first-line) or the work of Stahl* (*e.g.*, specific dose ranges); Additional thresholds (*e.g.*, gaps between treatments) were decided upon by our research team using distributional information from Penn State Psychiatry Clinical Assessment and Rating Evaluation System (PCARES) registry and clinical experience. The algorithm is focused on MDD pharmacotherapy but does not address all use scenarios for every available drug. Future adjustments may be needed. Several data pre-processing steps may be necessary for the application of this algorithm to electronic medical record data, some steps are described below.

Drug (Rx) classification with respect to MDD:

- First-line: agomelatine, bupropion, citalopram, desvenlafaxine, duloxetine, escitalopram, fluoxetine, fluvoxamine, mianserin, milnacipran, mirtazapine, paroxetine, sertraline, venlafaxine, and vortioxetine.
- Second-line: amitriptyline, amoxapine, clomipramine, desipramine, doxepin, imipramine, levomilnacipran, moclobemide, nefazodone, protriptyline, quetiapine, selegiline, trimipramine, and vilazodone.
- Third-line: isocarboxazid, lurasidone, maprotiline, nortriptyline, paliperidone, phenelzine, reboxetine, and tranylcypromine.
- Adjunctive/augmentation therapies: aripiprazole, brexpiprazole, buspirone, lithium, lamotrigine, methylphenidate, olanzapine, pindolol, and quetiapine.
- Bupropion and quetiapine were excluded due to being monotherapies. Cariprazine and ketamine were considered experimental/under study and were not included for PCARES.

Multipurpose drug classifications:

- Doxepin: if prescribed at a dose ≤ 10 mg (threshold based on insurance coverage; the typical dose is 6 mg daily for sleep disorders) and the patient has an ICD-10 code for a sleeping disorder, the drug is considered to be used for the sleep disorder. Doxepin is included as a second-line with other tricyclic antidepressants.
- Duloxetine: if prescribed with an ICD-10 code for fibromyalgia, the drug is considered to be used for the fibromyalgia. Duloxetine can be prescribed for generalized anxiety disorder but is assumed to be targeting MDD.
- Milnacipran: if prescribed and an ICD-10 code for fibromyalgia is present, the drug is not prescribed for MDD.
- Paroxetine: if prescribed and the patient has an ICD-10 code for an anxiety disorder, it is considered to be prescribed for anxiety and if the dose is > 62.5 mg, it is prescribed for other conditions.
- Risperidone: if prescribed at a dose > 3 mg, the drug is not considered to be related to the treatment of psychiatric disorders; if the dose is ≤ 3, the drug is considered adjunctive.
- Trazodone: if prescribed at a dose ≥ 150 mg, the drug is considered a second-line treatment for MDD; if it is prescribed at a dose < 150 mg and the patient has an ICD-10 code for a sleep disorder, the drug is considered to be prescribed for the sleep disorder; if it is prescribed at a dose of < 150 mg and there is no ICD-10 code for a sleep disorder, the drug is assumed to be prescribed as an augmenting therapy.

Given the above codes, if a drug is a first-line, second-line, third-line, adjunctive, or augmenting therapy, and not likely to be prescribed for a comorbid condition, it is considered "related" to the treatment of MDD for this algorithm. Classifications may overlap and not all classifications are represented. All MDD-related drugs in a combination treatment were evaluated where applicable; the more complex the regimen the greater the risk of misclassification for some criteria. Because drug priority could not be assigned (*i.e.*, which drug is primary), drugs prescribed within a given visit were sorted alphabetically. Drug classification was completed using the American Hospital Formulary Service Index.

Classification of MDD and selection of the baseline visit date:

- The "baseline" visit date was set to the MDD diagnosis date nearest to, and within one year prior and one year after, the registry encounter visit. The indicator for recurrent MDD is intended to help capture treatments that would be logical for those who have had MDD in the past and may try treatments not recommended for those with first-time episodes.
- Non-recurrent MDD was classified using the following ICD-10 codes: F32.0, F32.1, F32.2, F32.9, F33.0, F33.1, F33.2, or F33.9
- Recurrent MDD was classified using F33.0, F33.1, F33.2, F33.9, F33.40, F33.41, or F33.42

Identifying switching and continuances:

- Medications were compiled into visit-by-visit entries such that one row represented the related drugs prescribed on that visit. Subsequently, "switching" was identified as events in which the treatment at a given visit was characteristically different from the most recent prior combination regardless of doses. A "modification" also included dose changes. Dropping a drug from a combination treatment was not considered a switch in the treatment approach. Because of the structure of electronic medical record data, it was helpful to collapse consecutive treatments with no pause between them, termed a "continuance" of treatment; identical drugs that were prescribed consecutively at different doses were considered separate entries. This step was done by taking only the earliest start date for a drug at a given dose and the latest stop date. The continuance indicator was coded as 0 at the baseline date and at the date of the first prescription (if different).

Classifying treatment gaps:

- Treatment gaps were assessed as a gap of > 30 days and < 180 days between the stop date of a given drug and the order date of the next drug in the dataset. A gap of > 30 days precluded a "continuance." Gaps > 180 days were considered indicative of relapse and therefore not treated as a gap in treatment for one given episode. Drug entries where the order date was more than 270 days after the visit were considered errors and were dropped.

Median thresholds:

- Researchers can decide whether to set the threshold to 1) zero to evaluate all discordance, 2) the median value for the full study sample (assuming it is not 0) to account for some population or provider level discordance that may be a product of separate processes (*e.g.*, local resources), or 3) the median for those who experience the relevant events (*i.e.*, among those who do switch at least once).

Average episode duration:

- Participants who are in treatment longer are not only more likely to have more complex clinical status but also more likely to try multiple treatment approaches and thus have greater potential to fail additional treatment guidelines. As a result, careful adjustment for MDD episode duration and visit frequency is important. The difference in days between visit dates with consecutive ICD-10 codes for active depression was calculated to obtain the sum of days with an active episode. The summed durations were then divided by the number of episodes (per codes for remitted depression: F32.4, F32.5, F33.40, F33.41, or F33.42) to obtain the average episode duration per patient in days. Those without remission events had their average episode duration set to the full time spent in the registry. Those with an average duration of less than 30 days were set to missing.

Starting GCA-8 score
All patients start with a maximum score of 8.

+8

Criterion 1
The first medication in the treatment window

Using the baseline date, assess the first drug prescribed in the given study window. If two or more medications are prescribed first, as long as a first-line (or second-line in the case of recurrent MDD) is included the event is considered concordant. Adjunctive and augmenting therapies should not be considered.

- The patient had non-recurrent depression and initially received a received a second-line or third-line medication. → -1
- The patient had recurrent depression and their first medication was a third-line medication. → -1
- The patient's first treatment was concordant or there was no eligible event. → -0

Criterion 2
The initial second-line medication is not after more than two first-line medications.

Tally the number of unique first-line drugs that are prescribed before the initial second-line drug.

- The patient's number of unique first-line drugs was greater than two. → -1
- The patient's initial second-line treatment was concordant or there was no eligible event. → -0

Criterion 3
The initial third-line medication is after a second-line medication.

If a patient receives a third-line medication during the treatment window, that medication should be tried after a second-line medication was tried. For this criterion, as long as a second-line drug was tried, it does not matter whether it directly precedes the third-line drug.

- The patient's initial third-line medication is not after a second-line medication. → -1
- The initial third-line treatment was concordant or there was no eligible event. → -0

Criterion 4
The Rx dose is increased to the recommended maximum before switching (# of discordant events ≤ median) and the total number of switches is not more than the median (# of switches ≤ median).

This criterion considers only drugs related to MDD treatment and excludes tapering from the evaluation of the maximum dose tried. The highest dose tried for a given combination should be evaluated against the recommended maximum (per CANMAT and Stahl) for the same drug. To simplify the algorithm, if any drug in a multi-drug approach reaches its maximum dose before a switch event, the criterion is considered satisfied.

Evaluate the median number of switch events and the median number of times failing to reach the maximum dose before switching.

- The patient had a switch event count that exceeded the median or the median number of events where the maximum dose was not reached before a switch was exceeded. → -1
- The patient's regimen did not exceed the median number of switch events or the median number of events where the maximum dose was not reached before a switch. → -0

Criterion 5
Treatment duration should be at least four weeks per Rx, two weeks per dose (# of discordant events ≤ median; excluding tapering), with an average gap of fewer than 31 days and a limited number of gaps (# of gaps ≤ median).

The duration of treatments was set using prescription start and stop dates. Tapering periods shorter than two weeks were not treated as discordant. There is a risk of misclassification for complex treatment approaches that involved multiple drugs.

Evaluate the median number of events where a drug or dose was prematurely changed and the median number of treatment gaps (discontinuities in treatment greater than 0 days and less than 180 days) in the sample.

- The patient had a drug or dose switched prematurely more times than the median, an average treatment gap of 31 days or more, or a number of treatment gaps greater than the median. → -1
- The patient had no more premature drug and dose switches than the median, an average treatment gap of 30 days or less, and a number of treatment gaps less than or equal to the median. → -0

Criterion 6
Treatment should be modified by changing the dose (tapering excluded) or adding an adjunctive/augmenting drug, before switching to a new med. (# of discordant events ≤ median).

The two treatment steps before a switch were evaluated to determine whether the same drug was prescribed but a dose was modified. Dropping drugs, tapering doses, or ultimately switching back to a drug was not considered discordant.

Evaluate the median number of events where the drug or dose was not modified prior to switching.

- The patient's regimen had more failures to modify than the median threshold. → -1
- The patient's regimen was modified appropriately. → -0

Criterion 7
The Rx combo is without notable drug-drug interactions (# of interactions ≤ median).

Important drug-drug interactions were identified using CANMAT and literature,[*,**] primarily those occurring via the cytochrome P450 pathway. Those that may induce serotonin syndrome were also included. A variant of the "continuance" indicator was used for this step to ignore different doses of the same medication (as opposed to treating them as separate entries).

Evaluate the median number of potential drug-drug interactions.

- The number of potential drug-drug interactions exceeded the median threshold. → -1
- The number of potential drug-drug interactions is less than or equal to the median threshold. → -0

Criterion 8
The visit frequency should be at least three visits every six months.

A simple tally of visits within six months was used to establish whether a patient met the concordance criteria in one year.

- The number of visits was less than three in six months. → -1
- The number of visits was at least three in six months. → -0

Final GCA-8 score

[*]Stahl SM. Prescriber's Guide. Cambridge University Press; 2020. doi:10.1017/9781108921275

[#]Lynch T, Price A. The effect of cytochrome P450 metabolism on drug response, interactions, and adverse effects. Am Fam Physician. 2007;76(3):391-396. http://www.ncbi.nlm.nih.gov/pubmed/17708140

[##]Preissner S, Kroll K, Dunkel M, et al. SuperCYP: a comprehensive database on Cytochrome P450 enzymes including a tool for analysis of CYP-drug interactions. Nucleic Acids Res. 2010;38(suppl_1):D237-D243. doi:10.1093/nar/gkp970