Using an explicit guideline-based criterion and implicit review to assess antipsychotic dosing performance for schizophrenia

RICHARD R. OWEN1,2, CAROL R. THRUSH1,2, TERESA J. HUDSON1,2, SUSAN R. MALLORY3, ELLEN P. FISCHER1,2, JAMES A. CLARDY2 AND D. KEITH WILLIAMS2,4

1Veterans Affairs Health Services Research and Development Service, Center for Mental Healthcare and Outcomes Research (CeMHOR), Central Arkansas Veterans Healthcare System, 2Centers for Mental Healthcare Research, Department of Psychiatry and Behavioral Sciences, University of Arkansas for Medical Sciences, Little Rock, AR, 3Valley Mental Health, Salt Lake City, UT and 4Division of Biometry, University of Arkansas for Medical Sciences, Little Rock, AR, USA

Abstract

Objective. Using structured implicit review as the gold standard, this study assessed the sensitivity and specificity of an explicit antipsychotic dose criterion derived from schizophrenia guidelines.

Design. Two psychiatrists reviewed medical records and made consensus-structured implicit review ratings of the appropriateness of discharge antipsychotic dosages for hospitalized patients who participated in a schizophrenia outcomes study. Structured implicit review ratings were compared with the explicit criterion: whether antipsychotic dose was within the guideline-recommended range of 300–1000 chlorpromazine milligram equivalents (CPZE). In addition, reasons for deviation from guideline dose recommendations were examined.

Setting and study participants. A total of 66 patients hospitalized for acute schizophrenia at a Veterans Affairs medical center or state hospital in the southeastern US.

Main outcome measures. The sensitivity and specificity of the explicit dose criterion at hospital discharge were determined in comparison with the gold standard of structured implicit review.

Results. At hospital discharge, 61% of patients (n=40) were receiving doses within the guideline-recommended range. According to structured implicit review ratings, antipsychotic dose management was appropriate for 80% (n=53) of patients. When the 300–1000 CPZE dose criterion (dosage within or outside the recommended range) was compared with structured implicit review, it demonstrated 84.6% sensitivity and 71.7% specificity for detecting inappropriate antipsychotic dose.

Conclusions. The explicit antipsychotic dose criterion may provide a useful and efficient screen to identify patients at significant risk for quality of care problems; however, the relatively low specificity suggests that the measure may not be appropriate for quality measurement programs that compare performance among health plans.

Keywords: antipsychotic agents, guidelines, quality indicators, schizophrenia, sensitivity and specificity
phrenia Patient Outcomes Research Team recommendations and the American Psychiatric Association guidelines, for example, recommend treatment of acute exacerbations of schizophrenia with 300–1000 chlorpromazine milligram equivalents (CPZE) per day [10,12]. However, numerous studies have shown that routine prescribing practices often do not conform to evidence-based recommendations [14–18]. Therefore, guideline-based performance monitoring and improvement efforts for schizophrenia could initially focus on whether antipsychotic medications are prescribed within recommended dose ranges.

We recently described the development and initial application of a clinical performance measure that assesses adherence to an explicit guideline-based antipsychotic dose criterion for treatment of acute schizophrenia (300–1000 CPZE), and have reported on the measure’s accuracy, feasibility, and predictive validity [18,19]. It is also important to assess the concurrent validity of any new clinical performance measure [20,21]. In this case, concurrent validity can be assessed by examining how well the explicit dose criterion identifies inappropriate medication management as compared with direct assessment of medication management by a qualified psychiatrist who has reviewed comprehensive information on the patient’s clinical status, functioning, and past response to medications (implicit review). Implicit review is commonly the reference standard or ‘gold standard’ for validation of explicit quality indicators [20]. Implicit review is costly, time consuming, and subject to inter-observer bias [22,23]; however, it takes into account clinical details and subtleties of care that explicit review cannot consider [24]. In structured implicit review (SIR), reviewers are instructed to base their quality judgments on specific information on the chart [25].

In this report, we examine the sensitivity and specificity of a guideline-derived explicit antipsychotic dose criterion in comparison with SIR by trained psychiatrists. The explicit criterion for antipsychotic dose can be viewed as a new ‘laboratory test’ for inappropriate care, and as with any new test, the clinician must be aware of its sensitivity and specificity for detecting the condition in comparison to a gold standard test. The ‘condition’ to be addressed in this case is inappropriate medication management and the ‘gold standard’ is the psychiatrists’ consensus SIR rating of appropriateness of medication dose. In this context, a ‘positive test’ refers to an inappropriate antipsychotic dose, either <300 CPZE or >1000 CPZE. Thus, a patient who was prescribed a dose outside the guideline-recommended dose range and whose care was rated ‘inappropriate’ by SIR would be considered a true test positive. In contrast, a patient prescribed an out-of-range dose whose care was rated ‘appropriate’ by SIR would be considered a false positive.

In order to understand better how this performance measure works we also conducted a descriptive analysis of the frequency and reasons for deviation from guideline recommendations. To our knowledge, there is no published literature on the nature and frequency of common and legitimate reasons (also called ‘acceptable alternatives’ [25]), which may account for observed deviations from medication management guidelines for acute schizophrenia (e.g. the false positives). This report attempts to address that gap by examining the documented reasons for deviation from guideline-recommended antipsychotic doses (300–1000 CPZE) at hospital discharge for 66 patients with schizophrenia. Our analyses have implications for both quality assessment and clinical quality improvement efforts.

**Methods**

**Sample**

**SCHIZOM I dataset**

The database for the first validation study of the Schizophrenia Outcomes Module [SCHIZOM I] was the primary dataset used in this study. The study and database have been described in detail elsewhere [26,27]. Briefly, the database includes detailed longitudinal information for 160 patients with schizophrenia who were recruited into the study during an index hospitalization in 1992 or 1993. Patients were recruited from the Central Arkansas Veterans Healthcare System Medical Center (47%) and from the Arkansas State Hospital (53%), a state-funded facility run by the Arkansas Division of Mental Health Services. All patients met DSM-III-R criteria for schizophrenia, confirmed using the Structured Clinical Interview (SCID) for DSM-III-R [28], and were between 18 and 55 years of age.

**Exclusion criteria**

For the present study, we excluded SCHIZOM I patients who were participating in double-blind medication studies (n = 5), were admitted primarily for treatment of substance abuse rather than schizophrenia (n = 16), had been hospitalized during the entire 6-month follow-up period (n = 2), or had incomplete data (n = 24). In addition, because there is controversy over methods for converting long-acting injectable (depot) medication doses to CPZE, and because the predictive validity of explicit dose criteria has not yet been established for depot prescribing, we excluded all patients who were receiving long-acting injectable medications alone or in combination with oral medications (n = 47). Data for the remaining 66 patients were utilized in analyses reported here.

**Additional data collection**

To facilitate SIR ratings, detailed medication management and treatment data were abstracted from medical records by a trained, experienced psychiatric nurse practitioner, using a structured chart-abstraction instrument. The instrument, developed specifically to assess the quality of medication management for schizophrenia, examines documentation in multiple domains, including admission evaluation data, substance abuse history, treatment history, medical history, medication compliance, side effects, and hospital discharge data (instrument available from corresponding author). The nurse practitioner also compiled narrative summaries of the treatment course during hospitalization, including data on medication prescribing, medication response, adverse events,
Table 1  Structured implicit review form—item assessing neuroleptic dose

<table>
<thead>
<tr>
<th>Review criteria</th>
<th>Rating scale¹</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rate the quality of medication management with respect to neuroleptic dose.</td>
<td>Overall, this aspect of medication management is:</td>
</tr>
<tr>
<td>Consider the following questions to develop this rating:</td>
<td>1. Appropriate</td>
</tr>
<tr>
<td>Is dose within the chlorpromazine-equivalent accepted ranges (300–400 mg/day to 1000 mg/day)?</td>
<td>2. Appears to be appropriate</td>
</tr>
<tr>
<td>If dose is outside the range, is it documented appropriately or inappropriate</td>
<td>3. Appears to be inappropriate</td>
</tr>
</tbody>
</table>
| given the clinical circumstances or given the documentation aspects of the process of care. Additionally, the implicit review team adopted a four-item response scale that eliminated the equivocal rating [30] and included explicit anchors for the ratings of appropriateness (Table 1). Ratings of ‘1’ and ‘4’ were used when the quality of medication management with respect to neuroleptic dose was assessed to be clearly appropriate or inappropriate, respectively, based on the available information and the clinical situation, but for which there was limited documentation concerning the clinical rationale for the antipsychotic dose prescribed. The revised form was piloted on 20 practice records. Although the resulting SIR form includes 12 items rating the appropriateness of various aspects of treatment, including medication dosage, medication choice, compliance, and aftercare arrangements, in this report we discuss only the item assessing appropriateness of antipsychotic medication dose (Table 1).

SIR ratings

Two implicit review-team psychiatrists independently reviewed each patient’s chart abstraction and narrative summary and completed the corresponding SIR rating form. Afterwards, the psychiatrists met to resolve discrepancies in SIR ratings. Raters discussed and reached a consensus rating for patients whose medication dosage was rated appropriate (a 1 or 2) by one rater but inappropriate (a 3 or 4) by the other. The process of resolving disagreements during these meetings involved reviewing the chart abstraction forms and narrative case summaries for each subject with discrepant SIR ratings and then carefully applying the explicit dose criterion recommended by guidelines. In cases in which antipsychotic dose varied from the recommended range, reviewers considered whether that variation was nevertheless acceptable, given the clinical circumstances or given the documentation of a reason for the unusual dose. To facilitate calculation of sensitivity, specificity, and agreement between raters, SIR ratings were dichotomized as appropriate (ratings of 1 or 2) or inappropriate (ratings of 3 or 4).

Agreement between raters

The psychiatrists’ initial independent SIR ratings were in agreement for 76% (n = 50) of the sample and discordant for 24% of the total sample (n = 16). Agreement between raters
Table 2  Sensitivity and specificity equations \((n=66)\)

<table>
<thead>
<tr>
<th>Explicit criterion—antipsychotic dosage at hospital discharge</th>
<th>Structured implicit review rating</th>
<th>Inappropriate ((n=13))</th>
<th>Appropriate ((n=53))</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;300 CPZE ((n=12))</td>
<td>True positive (TP) (n=3)</td>
<td>False positive (FP) (n=9)</td>
<td></td>
</tr>
<tr>
<td>300–1000 CPZE ((n=40))</td>
<td>False negative (FN) (n=2)</td>
<td>True negative (TN) (n=38)</td>
<td></td>
</tr>
<tr>
<td>&gt;1000 CPZE ((n=14))</td>
<td>True positive (TP) (n=8)</td>
<td>False positive (FP) (n=6)</td>
<td></td>
</tr>
</tbody>
</table>

Sensitivity = \((TP/TP + FN) \times 100\) = \((11/11 + 2)\) \times 100 = 84.6%

Specificity = \((TN/TN + FP) \times 100\) = \((38/38 + 15)\) \times 100 = 71.7%

was statistically significant according to kappa, but low in magnitude \((k=0.25, \ P<0.03)\). The percentile agreement is high and the kappa is low due to the low prevalence of inappropriate care \([34,35]\). The final consensus rating for the 16 initially discordant ratings agreed with one reviewer’s initial (pre-consensus) rating in nine cases and the other reviewer’s initial rating in the remaining seven cases, which suggests that there was not a systematic bias among the raters in reaching a consensus—the final rating was equally likely to agree with the initial ratings of the individual rater.

**Data analysis**

All daily antipsychotic doses were converted to CPZE \([12, 36,37]\). We calculated the sensitivity and specificity of the explicit antipsychotic dose criterion in comparison with the consensus SIR ratings of dose appropriateness, as shown in Table 2. Also, we conducted a descriptive analysis of the false positive and true positive cases. Before examining the data, we developed a list of reasons why discharge antipsychotic doses might legitimately fall outside the guideline-recommended range (e.g. prior response to high/low antipsychotic dosages, patient preference). We then reviewed all available data for the 15 false positive cases [dosage <300 \((n=9)\) or >1000 CPZE \((n=6)\) when SIR = appropriate] and the 11 true positive cases [dosage <300 \((n=3)\) or >1000 \((n=8)\) when SIR = inappropriate], and then classified each in terms of the apparent reasons for out-of-range doses.

**Results**

**Criterion validity**

At hospital discharge, the majority of the 66 patients included in this study \((61%; n=40)\) were receiving doses within the recommended 300–1000 CPZE range; 18% \((n=12)\) were receiving antipsychotic doses below the guideline-recommended range (<300 CPZE), and 21% \((n=14)\) were receiving doses >1000 CPZE. According to the SIR ratings, antipsychotic dose management was appropriate for 80% \((n=53)\) of patients and inappropriate for 20% \((n=13)\). As shown in Table 2, the explicit criterion accurately identified 84.6% (95% CI = 55–98%) of inappropriately managed cases and 71.7% (95% CI = 58–83%) of appropriately managed cases. In this sample, with a 20% ‘gold standard’ prevalence of inappropriate care, positive predictive value (PPV) for the explicit criterion was 42.3% (95% CI = 23–63%), and the negative predictive value (NPV) was 95% (95% CI = 83–99%).

Figure 1 is a scatter-plot of antipsychotic doses at hospital discharge by structured implicit review (SIR) ratings.

**Deviation from guideline-recommended doses**

In 11 of the 15 false positive cases (appropriate by SIR despite out-of-range doses) there was adequate documentation in the medical record of a previous history or current good response to out-of-range doses to justify deviating from guideline-recommended antipsychotic doses. For the remaining four false positive cases, it was clear that the dosage was being titrated in the direction of the guideline-recommended range, but the patient had requested hospital discharge before titration was complete.
Discussion
This study assessed the concurrent validity of an explicit indicator of poor quality care for schizophrenia: prescription of a daily antipsychotic dose outside the guideline-recommended range of 300–1000 CPZE at the time of hospital discharge. Using this guideline-based explicit criterion, medication management was classified as inappropriate for 39% of the subjects, a level of non-adherence to dosage guidelines comparable to that reported in other recent studies of the quality of routine care for schizophrenia [14,17,38,39]. The explicit criterion demonstrated high sensitivity, somewhat lower specificity, modest positive predictive value, and high negative predictive value using SIR as the gold standard.

In order for a quality indicator to be useful it must be sensitive, that is, it must correctly identify a high proportion of cases of poor quality. It must also be specific, correctly identifying a high proportion of appropriate care cases. The finding that the explicit dose criterion is sensitive in comparison with SIR suggests that it can be used as an indicator of potentially inappropriate antipsychotic dosing for patients with schizophrenia. We have previously demonstrated that this indicator correlates with subsequent symptom outcomes [18], and other work has suggested that an explicit antipsychotic dose criterion be used as a measure of guideline conformance [16,18,40]. The indicator’s somewhat lower specificity, on the other hand, means that reliance on the explicit criterion alone would result in a relatively high rate of false positives, overestimating the extent of inappropriate antipsychotic prescribing.

Based on these findings, we suggest that the explicit dose criterion would be useful as part of a two-stage clinical quality assessment or improvement effort. The explicit criterion could be used as a screen to identify patients at greatest potential risk of poor quality medication management, in order to substantially reduce the number of cases that would need to be more intensively evaluated using structured implicit review or other methods. A two-step sample reduction method that uses explicit criteria followed by SIR to assess quality and direct clinical quality improvement has been advocated by others [5], and has been shown to be effective in a variety of clinical situations [41]. To our knowledge, this is the first report suggesting its utility in evaluating the quality of medication management in schizophrenia. It is also important to note that because the explicit dose criterion may result in a systematic overestimate of the extent of inappropriate dosing, performance reports that presented only these explicit findings to compare health plans could be potentially misleading. When quality indicators are used to compare health plans, it is better to err on the side of specificity [42].

One way to improve the specificity of the explicit antipsychotic dose indicator would be to combine it with data that reflect justified reasons for deviating from guidelines (i.e. accepting cases outside the dose range where an appropriate reason is documented). Although others have noted that there are a number of potentially legitimate reasons to deviate from clinical guideline recommendations in the routine care for particular disorders [25], we are not aware of any other studies that have examined the type or frequency of acceptable alternatives to guideline-recommended antipsychotic doses for acute schizophrenia treatment. The clinical literature suggests that antipsychotic treatment for schizophrenia using lower-than-recommended doses may be appropriate for patients with a previous history of good treatment response to such doses, or for those who suffer from intolerable side effects at higher doses [10]. On the other hand, higher-than-recommended antipsychotic doses may be appropriate when oral medication is used to supplement depot medication, or when multiple trials demonstrate that a patient does not respond at recommended doses [10,43]. Our descriptive analysis of the reasons for deviating from guideline recommendations suggests that individual variation in response to antipsychotic medication may account for a substantial proportion of appropriate deviations from guideline-recommended antipsychotic doses (false positives). Further work is needed to specify acceptable alternatives to guideline-recommended doses as well as how often such variation occurs.

The explicit guideline-derived dose criterion that we have tested here could be applied on a large scale by extracting data from administrative pharmacy databases [19] or by chart review. This approach could be enhanced by incorporating additional electronic or chart review data that reflect acceptable alternatives to guideline-concordant care [16]. As noted by Chen et al. [16], however, such data are not currently available in administrative databases. To make such an approach viable in routine quality improvement efforts in the future, a field representing reasons for deviation would be needed in administrative databases. Health systems could also focus on improving provider documentation of reasons for variation from guideline recommendations so that this type of data could readily be abstracted via chart review.

Another quality improvement approach to apply the explicit dose criterion could involve the use of patient outcomes data. Efforts to improve the quality of health care could be enhanced significantly by having such information. For instance, we could focus quality improvement efforts for antipsychotic dosing on patient cases where dose was not guideline concordant and where outcomes were suboptimal. In addition, some patients might be receiving moderate doses yet still have poor outcomes. For these patients, other factors
could be examined (e.g. medication choice, medication compliance, side effects, substance use), and interventions to improve the quality of care could be targeted appropriately. However, unless a system of care has an outcomes management system or other approach in place for routinely collecting patient outcomes data, this approach would be unlikely to be feasible for routine quality assessment and improvement efforts. As noted above, the best use of the explicit criterion, based on current evidence, would be to use two stages: screening with the explicit dose criterion, followed by SIR of cases that do not meet the criterion.

The modest initial agreement between SIR raters is a limitation of the present study. Given that the overall initial agreement between the two raters in this study was 76%, we recommend that additional structuring and testing of the SIR review process is carried out before implementing our two-stage quality review approach. Fifteen of the 16 cases of disagreement involved ratings of ‘appears to be appropriate’ and ‘appears to be inappropriate’. Because these intermediate ratings represented uncertainty due to poor documentation, agreement could be improved by providing further guidance to reviewers about clinical circumstances that would justify variation from guideline dose recommendations. In this validation study, we used two raters to conduct the structured quality reviews, a method that would probably not be feasible in routine quality of care work. With better inter-rater reliability, we could be more confident recommending the use of a single rater with the SIR approach we used. The present study is also limited by its relatively small sample, and the fact that the dataset was collected in 1992–93, before introduction of most newer second-generation antipsychotic medications. The relatively wide confidence intervals around our estimates of sensitivity and specificity reflect the size of the sample. Although our data do not necessarily reflect current prescribing practices, the level of guideline adherence for antipsychotic dose we observed is quite similar to more recent studies [14,17,38,39]. Nevertheless, this study needs to be replicated in a larger sample with more representative prescribing of second-generation antipsychotics. More research is needed to develop an explicit dose criterion for patients who are receiving long-acting injectable antipsychotics as well as those receiving the second-generation agents. Future research should also focus on developing and testing performance measures that would apply to patients with dual diagnoses (e.g. schizophrenia and substance abuse) [44].

**Conclusion**

Two major challenges in quality assessment and improvement are the development and testing of quality indicators. To be useful, an indicator must reliably identify subjects at risk for poor quality care, taking into account clinically appropriate deviations from the indicator. Furthermore, if it is to be used in routine quality measurement, the indicator must be simple and relatively inexpensive to apply, and provide meaningful information. As noted by Hofer, few quality indicators meet these standards [20]. Our focus on antipsychotic prescribing for schizophrenia is important because high-quality medication management can contribute to positive clinical and functional status outcomes [10,45]. Study results suggest that an explicit guideline-derived antipsychotic dose criterion is sufficiently sensitive to be useful for internal clinical quality improvement efforts in this area. However, given that the explicit dose criterion was not highly specific, it may not be appropriate to use as a stand-alone external quality indicator for reporting and accreditation purposes. Thus, until we can identify a more specific way to identify the patients at greatest risk of poor quality care, our results suggest that a combination of explicit and implicit review would be most useful for improving the appropriateness of antipsychotic dosing for schizophrenia.

**Acknowledgements**

This research was supported by grants from the Department of Veterans Affairs HSR&D Service (HSR&D SDR 91–005 and IIR 95–020) and the National Institute of Mental Health (R03 MH49123). Dr Owen’s work on this project was supported by a Research Career Development Award from the VA Health Services Research and Development Service. The views expressed in this article are those of the authors and do not necessarily represent the views of the Department of Veterans Affairs or the National Institute of Mental Health.

**References**


Accepted for publication 29 January 2002