HUMAN PSYCHOPHARMACOLOGY
Hum. Psychopharmacol Clin Exp 2011; **26**: 282–290.
Published online 2 June 2011 in Wiley Online Library
(wileyonlinelibrary.com) **DOI**: 10.1002/hup.1200

Early improvement on antipsychotic treatment as a predictor of subsequent response in schizophrenia: analyses from ziprasidone clinical studies

Cedric O'Gorman[1]*, Shitij Kapur[2], Sheela Kolluri[1] and John Kane[3]

[1]*Pfizer Inc., New York, New York, USA*
[2]*Institute of Psychiatry, King's College London, London, UK*
[3]*The Zucker Hillside Hospital, New York, New York, USA*

Objective We examine data from short-term placebo-controlled and comparator-controlled clinical trials of ziprasidone in schizophrenia to confirm the predictive capacity of early symptom changes for response. We pose the question of how early is too early to consider "stay or switch" and evaluate the predictive capability of a clinical measure in this regard.
Methods We presented two separate pooled analyses of (i) two placebo-controlled and (ii) two active comparator (risperidone and olanzapine) randomized trials of ziprasidone in schizophrenia. Relationship between early changes in Positive and Negative Syndrome Scale (PANSS) total, Brief Psychiatric Rating Scale (BPRS), and Clinical Global Impression-Improvement (CGI-I) scores and treatment outcome was evaluated.
Results Week 2 improvement was more reliably predictive of subsequent outcome than week 1 improvement using PANSS and BPRS scores with high sensitivity and specificity, whereas CGI-I had much lower specificity. Overall, non-improvement at week 1 or week 2 was highly predictive of non-response using BPRS scores and PANSS but not CGI-I.
Conclusions These data, independent of antipsychotic used, confirm prior research showing that early improvement in symptoms is predictive of response. There appears to be an important window of time, beyond week 1, during which important clinical decisions to stay or switch medication may be made. Copyright © 2011 John Wiley & Sons, Ltd.

KEY WORDS—schizophrenia; predictor; response; ziprasidone

INTRODUCTION

The extent to which early improvement in treated patients with schizophrenia is predictive of subsequent disease course, and outcome has been increasingly studied with a view to improving clinical management of the disorder as well as potentially optimizing clinical trial design. Early prediction of antipsychotic treatment response in schizophrenia, based on the presence or absence of initial improvement, may facilitate a timely decision to either continue treatment or switch to an alternative agent. Current guidelines for the treatment of schizophrenia do not address issues of response or methods of predicting response. Treatment guidelines recommend that physicians wait up to 4–8 weeks before switching antipsychotics because of lack of efficacy (Kinon *et al.*, 2008), and antipsychotic trial durations of approximately 6 weeks are the current recommendations. However, efficacy has been shown to occur as early as the first week of treatment (Leucht *et al.*, 2005; Agid *et al.*, 2006; Kinon *et al.*, 2008), so what is the significance of early symptomatic change for treatment decisions?

Studies have focused mainly on two predictive capacities: early improvement as a predictor of clinical response (sensitivity and positive predictive value) and lack of early improvement as a predictor of lack of response (specificity and negative predictive value). Sensitivity is the correct identification of subsequent responders, and specificity is the correct identification of subsequent non-responders. Positive predictive value (PPV) is the probability that early improvers show subsequent response, whereas negative predictive value (NPV) is the probability that early non-improvers are subsequent non-responders (Jäger *et al.*, 2009b). These are valid methods and are useful when assessing the overall predictive capacity of a symptom measure to evaluate the association between initial symptom change and subsequent outcome.

*Correspondence to: C. O'Gorman, Pfizer Inc., 235 East 42nd Street, New York, NY 10017, USA. Tel: 1-212-733-5293; Fax: 1-646-441-5605. E-mail: cedric.o'gorman@pfizer.com

Received 1 November 2010
Accepted 29 March 2011

Previous studies examined several thresholds for early improvement as a marker for later response using different methodologies. Some studies used an early improvement of ≥20% in Positive and Negative Syndrome Scale (PANSS) total or Brief Psychiatric Rating Scale (BPRS) total scores as a marker (Correll *et al.*, 2003; Chang *et al.*, 2006; Emsley *et al.*, 2006a, 2006b; Kinon *et al.*, 2008). A naturalistic study by Jäger *et al.* demonstrated that thresholds of a 20% and 30% reduction in PANSS total scores in the first two weeks had the highest total accuracy for prediction of remission (65%) and response (76%), respectively, at discharge (Jäger *et al.*, 2009b). In a recent prospective study, Kinon and colleagues found that higher thresholds for endpoint response (40% vs. 30% or 20% improvements in PANSS total score) led to higher sensitivity for the early improvement but lower specificity while the overall predictive value was similar (Kinon *et al.*, 2010). Although much work has been carried out using various definitions of clinical response at study endpoint (improvements of 20%–50% on either the PANSS or BPRS scales), a core clinical question is what do patients, physicians, and caregivers really view as an acceptable clinical outcome after 6, 8, and 12 weeks of treatment.

Accepting that improvement on antipsychotic therapy can occur in the first week (Leucht *et al.*, 2005; Agid *et al.*, 2006; Kinon *et al.*, 2008), and that early improvement/non-improvement has a valid and strong predictive capability for longer-term outcome, one important question is how early is too early to make the important "stay or switch" clinical decision. Although most retrospective analyses and even prospective analyses have studied week 2 improvement, some authors have evaluated week 1 as well (Correll *et al.*, 2003; Emsley *et al.*, 2006b). In *post hoc* analyses of a trial for the typical antipsychotic fluphenazine for patients with chronic schizophrenia, Correll and colleagues reported a high degree of specificity (100%) at week 1 for non-improvers using a 20% threshold for reduction in BPRS total score; although the sensitivity was very low with only 35% of week 4 responders showing improvement at week 1 (Correll *et al.*, 2003). In contrast, in first-episode patients with schizophrenia, Emsley *et al.* found specificity and sensitivity of 65% and 69%, respectively, for week 1 symptom changes in PANSS total scores (Emsley *et al.*, 2006b). More research would be helpful to elucidate the predictive value of week 1 findings in different patient populations.

What is the optimal threshold for early improvement for predicting later response and how does the predictive capacity of various early measures change over time? The data seems to suggest that the later "early" time points are associated with greater predictive accuracy (Derks *et al.*, 2010). However, in the interest of not forcing patients to endure treatment for unnecessarily long periods when they remain unresponsive, an important question to ask might be "how early is too early to make a stay or switch decision?"

Early prediction models have been developed for several first- and second-generation antipsychotics. Among the first studies using systematic prediction models assessing sensitivity and specificity was for the drug fluphenazine looking at a 20% threshold of improvement in the BPRS during week 1 (Correll *et al.*, 2003). Early symptom response (as a predictor of outcome) and overall response has also been assessed for the antipsychotics olanzapine and risperidone (Chang *et al.*, 2006; Kinon *et al.*, 2008). These findings have been validated in a recent prospective study that examined early improvement and further evaluated the benefit of a randomized switch for the early non-improvers from risperidone to olanzapine (Kinon *et al.*, 2010). Although these studies have detailed the relationship between early symptom changes and response, it is important to also verify this for individual drugs, such as ziprasidone. Our initial analysis of two short-term, fixed-dose, placebo-controlled studies of ziprasidone for the treatment of schizophrenia indicated that the predictive power (PP) of improvement/non-improvement at week 2 for the week 6 response was 71% (week 2 improvement and week 6 response defined as ≥10% and ≥30% reductions in PANSS total score, respectively) (Kane *et al.*, 2007).

The majority of work on early improvement/non-improvement for later outcomes in schizophrenia has focused on the PANSS and BPRS scales. These research tools are not readily used in clinical practice, yet the "stay or switch" conundrum is clearly a clinical one. For instance, Chang and colleagues used both PANSS and BPRS scores from the first two weeks to predict response (defined as a reduction in PANSS total score of ≥20%) at weeks 4 and 6 and found that PANSS scores were more accurate predictors than BPRS scores (Chang *et al.*, 2006). Recent publications have evaluated the correlation between early PANSS improvement and response and remission (Emsley *et al.*, 2007; Schennach-Wolff *et al.*, 2011). Masand and colleagues evaluated data from 10 schizophrenia studies for ziprasidone and found that a Clinical Global Impression of Improvement (CGI-I) score of 1 correlated with the remission criteria developed by the remission working group and that a CGI-I score of 1 or 2 at week 1 successfully predicted remission in schizophrenia (Masand *et al.*, 2011).

Here, we present analyses of placebo-controlled and comparator studies of ziprasidone in the treatment of

DOI: 10.1002/hup

schizophrenia. Using retrospective data from the placebo-controlled trials, we looked at this phenomenon specific to ziprasidone (Daniel *et al.*, 1999; Rappard *et al.*, 2006). Additionally, using data from the comparator studies (involving olanzapine and risperidone), we looked at this question generally without regard to which antipsychotic was used (Addington *et al.*, 2004; Simpson *et al.*, 2004). Our hypothesis was that these analyses would support the initial findings that demonstrated that early improvement/non-improvement predicts subsequent outcome independent of the antipsychotic chosen.

We examined endpoint responses of ≥30% in the placebo-controlled studies and ≥40% in the comparator studies (comprising ziprasidone, olanzapine, and risperidone). Our aim was to determine whether early improvement can predict response at 6 weeks. In particular, we queried whether week 2 scores could correctly identify the responders and non-responders (measures of sensitivity and specificity) at week 6 and assessed the PP of examining early scores. For the comparator studies, we further hypothesized that clinically meaningful decisions could not be made as early as week 1 because of several confounding reasons and compared the predictive parameters of week 1 and 2. Furthermore, we chose a relatively low response of 10% reduction in BPRS and PANSS scores to capture early signs of improvement at week 1. Our hypothesis was that there would be too many confounding factors in that early time frame to allow for meaningful clinical decisions to be made. Finally, we assessed the predictive capability of CGI-I at weeks 1 and 2 for later response; it would be important to know if clinical decisions about continuing or stopping treatment could be made based on CGI, which can be easily employed in clinical practice.

METHODS

These data reflect two separate, pooled, *post hoc* analyses. Data from two similarly designed placebo-controlled studies were pooled and analyzed. Separately, data from two active comparator trials were also pooled and analyzed.

Placebo-controlled studies

We pooled data from two similarly designed 6-week, fixed-dose, randomized, placebo-controlled trials (studies 114 and 115) of ziprasidone in hospitalized patients with schizophrenic illness (Daniel *et al.*, 1999; Rappard *et al.*, 2006). Following a 3-day to 7-day washout period, subjects were randomized to fixed-dose treatment with ziprasidone (40–160 mg) or placebo. Subjects in the 40-mg to 160-mg dose range ($N = 369$) were evaluated at weeks 2 and 6.

Inclusion criteria. Subjects had an acute exacerbation of a chronic or subchronic schizophrenic illness as per Diagnostic and Statistical Manual of Mental Disorders, 3rd Edition, Revised (DSM-III-R). All subjects were diagnosed at least 6 months prior to screening and had been hospitalized <4 weeks prior to screening. A baseline PANSS total score of ≥60, with a score of ≥4 on ≥2 core items of the PANSS (conceptual disorganization, hallucinatory behavior, suspiciousness, and unusual thought content) was also required, as was a score of ≥2 on the CGI-I scale.

Response criteria. Week 2 improvement was defined as a ≥10% reduction in PANSS total score, and lack of improvement was defined as a <10% reduction in PANSS total score. Week 6 response was defined as a ≥30% reduction in PANSS total score, and non-response was defined as a <30% reduction in PANSS total score.

Comparator studies

Data from two similarly designed, flexible-dose, randomized, double-blind, comparative trials (study 302 with ziprasidone versus risperidone for 8 weeks and study 548 with ziprasidone versus olanzapine for 6 weeks) in hospitalized patients experiencing an acute exacerbation of schizophrenia or schizoaffective disorder were pooled (Addington *et al.*, 2004; Simpson *et al.*, 2004).

In study 548, after screening and washout, subjects were randomized double-blind to either ziprasidone or olanzapine (Simpson *et al.*, 2004). During week 1, the fixed-dose regimen for ziprasidone was 40 mg b.i.d on days 1 and 2 and 80 mg b.i.d on days 3–7; for olanzapine, it was 5 mg/day on days 1 and 2 and 10 mg/day on days 3–7. From weeks 2 to 6, subjects received flexible doses of ziprasidone (40, 60, or 80 mg b.i.d) or olanzapine (5, 10, or 15 mg/day).

In study 302, after screening and washout, subjects were randomized double-blind to either ziprasidone or risperidone (Addington *et al.*, 2004). Ziprasidone subjects started on 40 mg b.i.d for the first week, which was then adjusted at weekly intervals in increments of 20 mg b.i.d within the range of 80 to 160 mg/day. Risperidone was titrated from 1 mg b.i.d (day 1) to 3 mg b.i.d (days 3 to 7) during the first week and was then adjusted at weekly intervals in 1 mg increments to a maximum of 5 mg b.i.d.

Inclusion criteria. Subjects for both comparator studies included men and women aged 18–64 years

DOI: 10.1002/hup

who were required to have a primary diagnosis of schizophrenia or schizoaffective disorder. For study 548, subjects with a baseline CGI score of ≥4 and a score of ≥4 on ≥1 of the PANSS positive symptom items were included. For study 302, subjects with a PANSS total score of ≥60 and a score of ≥4 on ≥2 core PANSS items were included.

Response criteria. Improvement at weeks 1 and 2 was defined as a ≥10% and a ≥20% reduction, respectively, from baseline on the BPRS and PANSS (data for week 1 only) scores or a measurement of 1, 2, or 3 for CGI-I. Response at week 6 was defined as a ≥40% reduction from baseline BPRS or PANSS score or a measurement of 1 or 2 for CGI-I. Subjects in all treatment arms were included in the analyzed data set.

Statistical analyses

We used measures of sensitivity, specificity, PPV, NPV, and PP to examine the early improvement and response to treatment. Sensitivity is the correct identification of subsequent responders, and specificity is the correct identification of subsequent non-responders. PPV is the probability that early improvers show subsequent response, whereas NPV is the probability that early non-improvers are subsequent non-responders (Jäger *et al.*, 2009b). A receiver operating characteristics (ROC) curve was generated to further assess the ability of week 1 or week 2 PANSS, BPRS, or CGI-I total scores to discriminate between week 6 responders and non-responders. Area under the ROC curve ranges from 0.5 (non-informative) to 1 (perfect test discrimination). Most researchers consider a value of 0.7–0.8 as reasonable, whereas >0.8 has good discriminative capacity (Weinstein and Fineberg, 1980).

Placebo-controlled studies. The baseline characteristics were continuous variables that were descriptively compared for improvers and non-improvers at week 2. At each post-baseline time point analyzed, a last observation carried forward approach was used to impute missing PANSS total scores.

Comparator studies. Baseline characteristics (demographics, PANSS, BPRS, and CGI-I scores) for week 1 responders and non-responders were compared descriptively. The analyses used only observed cases in both studies, such that subjects with missing observation at time point of interest were excluded from the analysis. As no symptom change scores were assessed in study 302 at week 2, data from this study were excluded for the week 2/week 6 comparison.

RESULTS

Baseline demographics and clinical characteristics of early improvers and early non-improvers in the placebo-controlled studies are shown in Table 1. Baseline demographics and clinical characteristics for subjects in the comparator studies are shown in Table 2 and are descriptively comparable across early responders (≥20% reduction in BPRS) and non-responders (<20% reduction in BPRS) at week 2.

Predictive capacity of week 2 improvement/non-improvement for week 6 response/non-response

Among a total of 369 subjects in the placebo-controlled trials, using PANSS total scores, 159 were week 2 improvers (≥10%) and 90 were week 6 responders (≥30%). Improvement (≥10% reduction in PANSS total score) at week 2 correctly predicted last visit (week 6) response (≥30% reduction in PANSS total score) in 71 of 90 (sensitivity, 78.9%) subjects. Likewise, week 2 non-improvers (<10% reduction in PANSS total score) were correctly identified as non-responders at week 6 (<30% reduction in PANSS total score) in 191 of 279 subjects (specificity, 68.5%). Of the 159 week 2 improvers, only 71 were week 6 responders (PPV, 44.6%). However, a higher proportion of week 2 non-improvers were week 6 non-responders (191 of 210, NPV 91%) (Table 3). The PP at week 2 using PANSS was 71.0%, and this was confirmed by the ROC curve for week 2 PANSS scores as a predictor of 30% response at week 6 (Figure 1a), which had an area under the curve (AUC) of 0.74.

Table 1. Baseline demographics and clinical characteristics of early improvers and early non-improvers in the placebo-controlled studies

	Early improvers (≥10% reduction in PANSS total at week 2)	Early non-improvers (<10% reduction in PANSS total at week 2)
Characteristic	*N* = 159	*N* = 210
Age, years	38.0	38.4
Male, *n* (%)	112 (70.4)	145 (69.1)
Race/ethnicity		
Caucasian, *n* (%)	109 (68.6)	146 (69.5)
African American, *n* (%)	35 (22.0)	43 (20.5)
Asian, *n* (%)	3 (1.9)	9 (4.3)
Other, *n* (%)	12 (7.6)	12 (5.7)
Weight, kg	75.9	78.5
Age at diagnosis, years	22.8	22.4
Diagnosis of schizophrenia (versus schizoaffective disorder), *n* (%)	113 (71.1)	160 (76.2)
PANSS total	95.6	93.6
CGI-S	4.8	4.9

PANSS, Positive and Negative Syndrome Scale; CGI-S, Clinical Global Impression-Severity.

DOI: 10.1002/hup

Table 2. Baseline demographics and clinical characteristics of early improvers and early non-improvers at week 1 in the comparator-controlled studies

Characteristic	Early improvers (≥10% reduction in PANSS total at week 1) $N=216$	Early non-improvers (<10% reduction in PANSS total at week 1) $N=300$
Age, years	36.1	36.2
Male, *n* (%)	147 (68.1)	211 (70.3)
Race/ethnicity		
Caucasian, *n* (%)	157 (72.7)	224 (74.7)
African American, *n* (%)	31 (14.4)	43 (14.3)
Asian, *n* (%)	7 (3.2)	6 (2.0)
Hispanic, *n* (%)	10 (4.6)	16 (5.3)
Other, *n* (%)	11 (5.1)	11 (3.6)
Weight, kg	81.3	80.2
Age at diagnosis, years	24.3	24.0
Diagnosis of schizophrenia (versus schizoaffective disorder), *n* (%)	172 (79.6)	231 (77.0)
BPRS total	53.1	52.9
PANSS total	93.7	92.7
CGI-S	4.9	4.8

PANSS, Positive and Negative Syndrome Scale; BPRS, Brief Psychiatric Rating Scale; CGI-I, Clinical Global Impression-Severity.

Brief Psychiatric Rating Scale and CGI-I scores for a total of 153 patients were available for analysis at week 2 (ziprasidone and olanzapine data only). Although the sensitivity for correctly identifying week 6 responders at week 2 was consistent between CGI-I (103 of 110, 93.6%) and BPRS (37 of 46, 80.4%), the specificity was much lower for CGI-I (11 of 43, 25.6%). The NPV for BPRS and CGI-I scores was 89.2% (74 of 83) and 61.1% (11 of 18), respectively. The PP for week 2 BPRS scores (72.5%) was similar to CGI-I (74.5%) (Table 3). For week 2, the ROC curves for the BPRS scores had an AUC of 0.678, and for the CGI-I scores, the AUC was 0.676 (Figure 1b, c).

Predictive capacity of week 1 improvement/non-improvement for week 6 response/non-response

Based on PANSS scores, among a total of 370 subjects, 162 were week 1 improvers (≥10% reduction), and 104 were week 6 responders (≥40% reduction). At week 1, 69 of 104 (66.3%, sensitivity) were correctly identified as week 6 responders; whereas 173 of 266 (65.0%, specificity) showing non-improvement at week 1 were week 6 non-responders. For the PANSS week 1 scores, the PP was 65.4% (Table 4).

Of a total of 369 subjects, 174 and 228 were week 1 improvers, and 107 and 238 were week 6 responders, using BPRS and CGI-I scores, respectively (Table 4). For BPRS, of the 195 week 1 non-improvers, 159 (81.5%, NPV) were week 6 non-responders. However, for CGI-I, the NPV was lower (47.5%) as only 67 week 1 non-improvers were amongst the 141 week 6 non-responders. The PP of the week 1 result was 62.3% for BPRS and 62.6% for CGI-I scores (Table 4). The previously mentioned observations were confirmed by ROC curves for week 1 PANSS, BPRS, and CGI-I scores. The AUC for week 1 was 0.576 for PANSS total response, 0.558 for BPRS response, and 0.643 for CGI-I response (Figure 1).

DISCUSSION

Week 2 early improvement/non-improvement predicts response/non-response

Our results confirm that early improvement/non-improvement in PANSS, BPRS, or CGI-I during week 2 can predict subsequent outcome. Across both

Table 3. Predictive capabilities of week 2 changes for study endpoint outcomes

Efficacy measure	Ziprasidone, placebo-controlled data PANSS	Ziprasidone/olanzapine[a] BPRS	Ziprasidone/olanzapine[a] CGI-I
N (Total subjects)	369	153	153
Week 2 improvers	159 (71 TP + 88 FP)	70 (37 TP + 33 FP)	135 (103 TP + 32 FP)
Week 2 non-improvers	210 (191 TN + 19 FN)	83 (74 TN + 9 FN)	18 (11 TN + 7 FN)
Week 6 responders	90	46	110
Week 6 non-responders	279	107	43
Sensitivity	78.9% (71/90)	80.4% (37/46)	93.6% (103/110)
Specificity	68.5% (191/279)	69.2% (74/107)	25.6% (11/43)
PPV	44.6% (71/159)	52.9% (37/70)	76.3% (103/135)
NPV	91.0% (191/210)	89.2% (74/83)	61.1% (11/18)
PP	71% (262/369)	72.5% (111/153)	74.5% (114/153)

PANSS, Positive and Negative Syndrome Scale; BPRS, Brief Psychiatric Rating Scale; CGI-I, Clinical Global Impression-Improvement; TP, true positives; FP, false positives; TN, true negatives; FN, false negatives; PPV, Positive Predictive Value; NPV, Negative Predictive Value; PP, Predictive Power. Sensitivity = TP/week 6 responders, Specificity = TN/week 6 non-responders; PPV = TP/(TP + FP); NPV = TN/(TN + FN); PP = (TP + TN)/N. For the purpose of our analyses, "TP + FP" refers to early improvers, and "TN + FN" refers to early non-improvers.
[a]Study 548 with olanzapine data only. Study 302 with risperidone did not have week 2 data.

DOI: 10.1002/hup

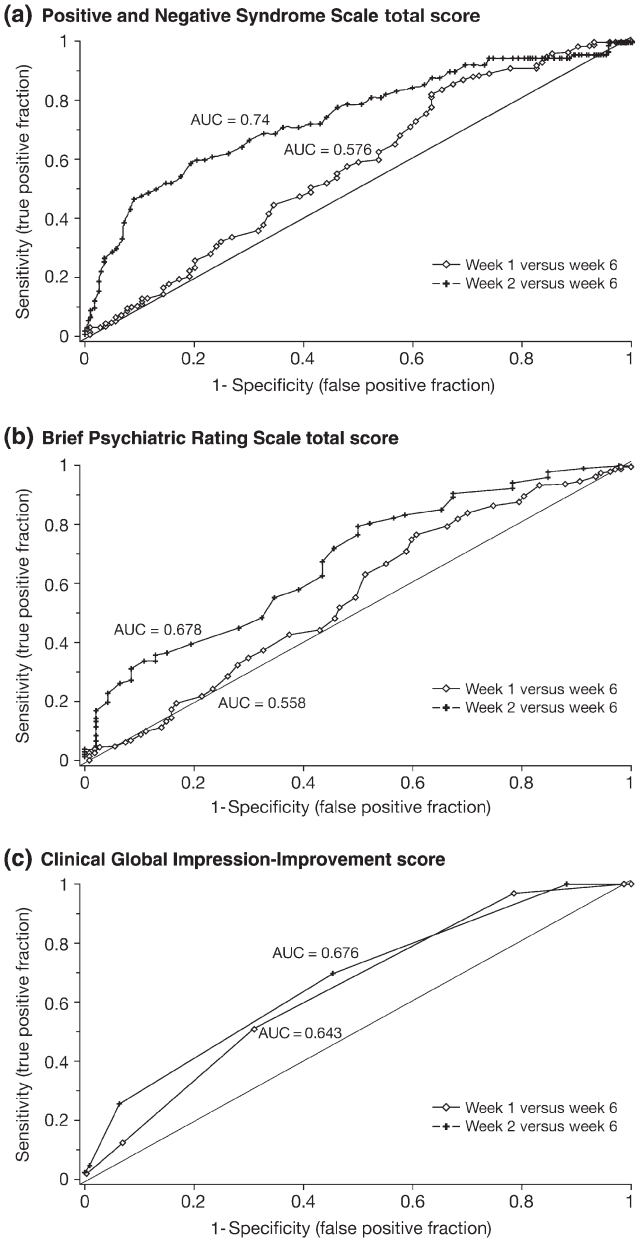

Figure 1. Receiver operating characteristics curves predicting response from early psychiatric testing in placebo-controlled and comparator studies (a) Positive and Negative Syndrome Scale total score week 1 versus week 6 (placebo data) and week 2 versus week 6 (ziprasidone, olanzapine data). (b) Brief Psychiatric Rating Scale total score week 1 versus week 6 (all comparator data) and week 2 versus week 6 (ziprasidone, olanzapine data). (c) Clinical Global Impression-Improvement total score week 1 versus week 6 (all comparator data) and week 2 versus week 6 (ziprasidone, olanzapine data). AUC, area under the curve

separate analyses (placebo and comparator data sets) with both BPRS and PANSS, early improvement with antipsychotics was associated with a sensitivity of approximately 80% and a specificity of just under 70%. These rates compare with studies reported by Chang *et al.* with high rates of both sensitivity (~80%) and specificity using a PANSS total score reduction threshold of ≥20% within the first two weeks (Chang *et al.*, 2006). However, a later study by Kinon and colleagues using the same threshold had a lower sensitivity (45%) but high specificity of 89% (Kinon *et al.*, 2008). These variations may be attributed to differences in baseline patient disease duration and severity.

Our analyses of the comparator data show that the CGI-I had a higher sensitivity at 93.6%, correctly identifying the majority of week 6 responders at week 2, whereas the specificity of CGI-I was just over 25%. Overall, the PP of CGI-I at week 2 was approximately equal to that of BPRS or PANSS. In particular, these analyses highlight that non-improvement is a particularly stronger predictor of non-response compared with early improvement predicting response.

An important aspect of this data set (PANSS and BPRS at week 2) was that approximately 50% of early improvers were false positives (they were not responders at week 6). As a result, the PPV of early improvement was low. Furthermore, in our analysis of the comparative dataset, we did not have week 2 data from study 302 with risperidone, thereby reducing the power of our analyses on BPRS and CGI-I at week 2. Although the number of subjects with week 2/6 PANSS data in the placebo-controlled analyses was higher than for the week 2/6 BPRS and CGI-I data in the comparative analyses (369 vs. 153), the overall PP at week 2 was comparable across efficacy tools, as were sensitivity, specificity, PPV, and NPV between BPRS and PANSS. CGI-I showed lower specificity and NPV than BPRS and PANSS at the week 2 time point.

Limitations of week 1 data to predict response/ non-response at week 6

The week 1 PANSS and BPRS data for early improvement on antipsychotic treatment (zipraisdone, olanzapine, or risperidone) showed sensitivity of approximately 66% and specificities in the range of 60%–65%. The CGI-I performed marginally better than the BPRS and PANSS scales in terms of sensitivity (68.9%) and outperformed both diagnostic scales on PPV (71.9% vs. 40.8% and 42.6%, respectively). Again, as with the week 2 data, CGI-I did not prove to be a particularly specific tool and underperformed the PANSS and BPRS in terms of NPV at week 1. Early non-improvement was highly predictive of later non-response for BPRS, but not for CGI-I, such that the absence of early improvement at

DOI: 10.1002/hup

Table 4. Predictive capabilities of week 1 changes for study endpoint outcomes

	Ziprasidone/olanzapine/risperidone (comparator studies)[a]		
Efficacy measure	PANSS	BPRS	CGI-I
N (Total subjects)	370	369	369
Week 1 improvers	162 (69 TP + 93 FP)	174 (71 TP + 103 FP)	228 (164 TP + 64 FP)
Week 1 non-improvers	208 (173 TN + 35 FN)	195 (159 TN + 36 FN)	141 (67 TN + 74 FN)
Week 6 responders	104	107	238
Week 6 non-responders	266	262	131
Sensitivity	66.3% (69/104)	66.4% (71/107)	68.9% (164/238)
Specificity	65.0% (173/266)	60.7% (159/262)	51.1% (67/131)
PPV	42.6% (69/162)	40.8% (71/174)	71.9% (164/228)
NPV	83.2% (173/208)	81.5% (159/195)	47.5% (67/141)
PP	65.4% (242/370)	62.3% (230/369)	62.6% (231/369)

PANSS, Positive and Negative Syndrome Scale; BPRS, Brief Psychiatric Rating Scale; CGI-I, Clinical Global Impression-Improvement; TP, true positives; FP, false positives; TN, true negatives; FN, false negatives; PPV, Positive Predictive Value; NPV, Negative Predictive Value; PP, Predictive Power. Sensitivity = TP/week 6 responders, Specificity = TN/week 6 non-responders; PPV = TP/(TP + FP); NPV = TN/(TN + FN); PP = (TP + TN)/N. For the purpose of our analyses, "TP + FP" refers to early improvers and "TN + FN" refers to early non-improvers.
[a]Data from comparator studies 548 (olanzapine) and 302 (risperidone) included.

week 1 or 2 was much more predictive (higher NPV) of non-response at week 6 using the BPRS scores. Week 1 PPs were again equivalent across the PANSS, BPRS, and CGI-I as they were at week 2, although overall predictive capacity increased from week 1 to week 2 by approximately 10%. These findings are reinforced by the ROCs with higher AUCs ranging from 0.67 to 0.74 for week 2/6 responses that showed acceptable discrimination compared with week 1/6 responses with AUCs ranging from 0.55 to 0.64.

There are some important limitations to interpreting the week 1 data. Although we have a greater number of subjects in the analyses of the comparative data for week 1/6 on PANSS, BPRS, and CGI-I, these studies were associated with titration schedules during the first week. As a result, it is questionable whether improvement can really be assessed before a stable dose is achieved. Additionally, the threshold of 10% improvement may have been too low to be meaningful on a predictive level. Although the intention was to be lenient to capture early signs of improvement, it had the effect of inflating the number of false positive and false negatives at week 1. As previous work has shown, although response can occur as early as the first 24 hours (Agid *et al.*, 2006), we find that the PP of early time points appears to increase from week 1 to 2 and onward. Previous studies have shown reasonable predictive validity of early improvement for a 46% PANSS total score improvement at week 2 and a 50% improvement for remission (AUC: response 0.707, remission 0.692) (Schennach-Wolff *et al.*, 2010). Although we made no distinction between first and multiple-episode subjects in our analyses, a higher threshold in the PANSS score within 2 weeks of treatment may well have been more discriminating. With the limitations associated with the frequent need for dose titration in clinical practice and because of the high degree of false positives at week 1, early predictive assessment at this time point for clinical decision making may be unrealistic. Indeed, this observation supports the recommendation for longer periods of assessment in hospital at a time when often speedy discharge of patients from inpatient care is advocated.

Potential value of Clinical Global Impression-Improvement to assess early improvement

As discussed previously, early non-improvement was highly predictive of later non-response for BPRS but not for CGI-I. These differences between BPRS and CGI-I in predicting later non-response suggests that further research is needed to compare the predictive capabilities of research versus clinical tools for outcomes in schizophrenia. CGI-I has been investigated as a proxy for remission in schizophrenia, and it would be worthwhile to further examine the utility of CGI-I as a tool in making "stay or switch" decisions, as it is considered more physician friendly and representative of global assessment. At present, the use of CGI-I in evaluating response and remission during clinical trials as a primary endpoint is limited; however, efforts are ongoing to determine if it is an appropriate tool for predicting both response and remission (Masand *et al.*, 2011). Nevertheless, what exactly these diagnostic scales measure is also an important consideration, as unlike PANSS and BPRS scales, the CGI ratings may correspond to relative change rather than absolute change (Leucht *et al.*, 2006).

DOI: 10.1002/hup

CONCLUSIONS

In light of several retrospective, prospective, and naturalistic studies, it appears that early prediction of response is possible and of value to clinicians. The complex matrix of demographic factors (gender, duration of episode, number of episodes, duration of untreated psychosis, impact of negative symptoms, and drugs selected) contributing to response variability needs to be further evaluated, and the interplay with improvement threshold by time point needs to be further elucidated (Mancama *et al.*, 2002; Gunduz-Bruce *et al.*, 2005; Gupta *et al.*, 2006; Emsley *et al.*, 2006a; Usall *et al.*, 2007; Jäger *et al.*, 2009a). Further work is needed to evaluate optimal early thresholds, how early is too early to switch, treatment trajectories, outcomes after switching to alternate medications, and how to better facilitate clinical decision making. In addition to improving patient care, advances in this area can optimize clinical trial design, thereby reducing study length and potentially reducing the costs associated with research and development.

CONFLICT OF INTEREST

This manuscript has not been published previously in a journal. Some of these analyses were presented by Dr Kane at the 47th Annual New Clinical Drug Evaluation Unit Meeting on 11–14 June 2007 in Boca Raton, FL, whereas further analyses were presented by Dr O'Gorman at the American College of Neuropharmacology Meeting on 7–11 December 2008, in Scottsdale, AZ.

Drs O'Gorman and Kolluri are employees of Pfizer Inc., New York, New York.

Dr Kane has served as consultant and/or speaker for Astra Zeneca, Boehringer Ingelheim, Bristol-Meyers Squibb, Cephalon Inc., Dainippon Sumitomo, Eisai Inc., Eli Lilly, H. Lundbeck A/S, Intracellular Therapeutics, Janssen Pharmaceutica, Johnson and Johnson, Merck, Myriad, Novartis, Otsuka Pharmaceutical, Pfizer, Proteus Biomedical, Rules Based Medicine, Takeda, Targacept, Vanda Pharmaceuticals, and Wyeth. Dr Kane is a shareholder of MedAvante Inc. Dr Kane has not received any compensation for development or preparation of this manuscript.

Dr Kapur has received grant support from AstraZeneca, Bristol-Myers Squibb, and GSK, and has served as consultant and/or speaker for AstraZeneca, Bioline, BMS, Eli Lilly, Janssen (J&J), Lundbeck, NeuroSearch, Otsuka, Pfizer, Roche, Servier, and Solvay Wyeth. Dr Kapur has not received any compensation for development or preparation of this manuscript.

ACKNOWLEDGEMENTS

This study was sponsored by Pfizer Inc., New York, New York. Editorial support was provided by Hajira Koeller, PhD, of PAREXEL and was funded by Pfizer Inc.

REFERENCES

Addington DE, Pantelis C, Dineen M, Benattia I, Romano SJ. 2004. Efficacy and tolerability of ziprasidone versus risperidone in patients with acute exacerbation of schizophrenia or schizoaffective disorder: an 8-week, double-blind, multicenter trial. *J Clin Psychiatry* **65**: 1624–1633.

Agid O, Seeman P, Kapur S. 2006. The "delayed onset" of antipsychotic action — an idea whose time has come and gone. *J Psychiatry Neurosci* **31**: 93–100.

Chang YC, Lane HY, Yang KH, Huang CL. 2006. Optimizing early prediction for antipsychotic response in schizophrenia. *J Clin Psychopharmacol* **26**: 554–559.

Correll CU, Malhotra AK, Kaushik S, McMeniman M, Kane JM. 2003. Early prediction of antipsychotic response in schizophrenia. *Am J Psychiatry* **160**: 2063–2065.

Daniel DG, Zimbroff DL, Potkin SG, Reeves KR, Harrigan EP, Lakshminarayanan M. 1999. Ziprasidone 80 mg/day and 160 mg/day in the acute exacerbation of schizophrenia and schizoaffective disorder: a 6-week placebo-controlled trial. Ziprasidone Study Group. *Neuropsychopharmacology* **20**: 491–505.

Derks EM, Fleischhacker WW, Boter H, Peuskens J, Kahn RS. 2010. Antipsychotic drug treatment in first-episode psychosis: should patients be switched to a different antipsychotic drug after 2, 4, or 6 weeks of nonresponse? *J Clin Psychopharmacol* **30**: 176–180.

Emsley R, Oosthuizen PP, Kidd M, Koen L, Niehaus DJ, Turner HJ. 2006a. Remission in first-episode psychosis: predictor variables and symptom improvement patterns. *J Clin Psychiatry* **67**: 1707–1712.

Emsley R, Rabinowitz J, Medori R. 2006b. Time course for antipsychotic treatment response in first-episode schizophrenia. *Am J Psychiatry* **163**: 743–745.

Emsley R, Rabinowitz J, Medori R. 2007. Remission in early psychosis: Rates, predictors, and clinical and functional outcome correlates. *Schizophr Res* **89**: 129–139.

Gunduz-Bruce H, McMeniman M, Robinson DG, *et al.* 2005. Duration of untreated psychosis and time to treatment response for delusions and hallucinations. *Am J Psychiatry* **162**: 1966–1969.

Gupta S, Jain S, Brahmachari SK, Kukreti R. 2006. Pharmacogenomics: a path to predictive medicine for schizophrenia. *Pharmacogenomics* **7**: 31–47.

Jäger M, Riedel M, Schmauss M, *et al.* 2009a. Prediction of symptom remission in schizophrenia during inpatient treatment. *World J Biol Psychiatry* **10**: 426–434.

Jäger M, Schmauss M, Laux G, *et al.* 2009b. Early improvement as a predictor of remission and response in schizophrenia: results from a naturalistic study. *Eur Psychiatry* **24**: 501–506.

Kane J, Vanderburg D, Loebel A, Roberts H, Batzar E. Early improvement with ziprasidone as a predictor of later treatment outcome in hospitalized patients with schizophrenia. *47th Annual New Clinical Drug Evaluation Unit Meeting*. June 11–14, 2007. Boca Raton, FL.

Kinon BJ, Chen L, Ascher-Svanum H, *et al.* 2008. Predicting response to atypical antipsychotics based on early response in the treatment of schizophrenia. *Schizophr Res* **102**: 230–240.

Kinon BJ, Chen L, Ascher-Svanum H, *et al.* 2010. Early response to antipsychotic drug therapy as a clinical marker of subsequent response in the treatment of schizophrenia. *Neuropsychopharmacology* **35**: 581–590.

Leucht S, Busch R, Hamann J, Kissling W, Kane JM. 2005. Early-onset hypothesis of antipsychotic drug action: a hypothesis tested, confirmed and extended. *Biol Psychiatry* **57**: 1543–1549.

Leucht S, Kane JM, Etschel E, Kissling W, Hamann J, Engel RR. 2006. Linking the PANSS, BPRS, and CGI: clinical implications. *Neuropsychopharmacology* **31**: 2318–2325.

Mancama D, Arranz MJ, Kerwin RW. 2002. Genetic predictors of therapeutic response to clozapine: current status of research. *CNS Drugs* **16**: 317–324.

Masand P, O'Gorman C, Mandel FS. 2011. Clinical Global Impression of Improvement (CGI-I) as a valid proxy measure for remission in schizophrenia: analyses of ziprasidone clinical study data. *Schizophr Res* **126**: 174–183.

Rappard F, Warrington LE, Lombardo I, Loebel I, Yang R. Efficacy of ziprasidone in the treatment of schizoaffective disorder: an analysis of 2 fixed-dose, placebo-controlled trials. *19th Congress of the European College of Neuropsychopharmacology*, September 16, 2006. Paris, France.

Schennach-Wolff R, Seemuller FH, Mayr A, *et al.* 2010. An early improvement threshold to predict response and remission in first-episode schizophrenia. *Br J Psychiatry* **196**: 460–466.

Schennach-Wolff R, Jäger M, Mayr A, *et al.* 2011. Predictors of response and remission in the acute treatment of first-episode schizophrenia patients - Is it all about early response? *Eur Neuropsychopharmacol* **21**: 370–378.

Simpson GM, Glick ID, Weiden PJ, Romano SJ, Siu CO. 2004. Randomized, controlled, double-blind multicenter comparison of the efficacy and tolerability of ziprasidone and olanzapine in acutely ill inpatients with schizophrenia or schizoaffective disorder. *Am J Psychiatry* **161**: 1837–1847.

Usall J, Suarez D, Haro JM. 2007. Gender differences in response to antipsychotic treatment in outpatients with schizophrenia. *Psychiatry Res* **153**: 225–231.

Weinstein MC, Fineberg HV. 1980. *Clinical Decision Analysis* W.B. Saunders: Philadelphia, PA.

DOI: 10.1002/hup