Psychiatry Research 170 (2009) 161–167

Contents lists available at ScienceDirect

Psychiatry Research

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

Defining "good" and "poor" outcomes in patients with schizophrenia or schizoaffective disorder: A multidimensional data-driven approach ☆

Ilya A. Lipkovich [a,*], Walter Deberdt [a], John G. Csernansky [b,e], Peter Buckley [c], Joseph Peuskens [d], Sara Kollack-Walker [a,*], Matthew Rotelli [a], John P. Houston [a]

[a] Lilly Research Laboratories, Eli Lilly and Company, Indianapolis, IN 46285, USA
[b] Department of Psychiatry, Washington University School of Medicine, 660 South Euclid Ave, St Louis, MO 63110, USA
[c] Department of Psychiatry and Health Behavior, Medical College of Georgia, 1515 Pope Avenue, Augusta, GA 30912, USA
[d] The Catholic University of Leuven, University Centre St Jozef, Kortenberg, Belgium
[e] Department of Psychiatry and Behavioral Sciences, Northwestern University Feinberg School of Medicine, Chicago, IL, USA

ARTICLE INFO

Article history:
Received 28 September 2007
Received in revised form 13 February 2008
Accepted 19 September 2008

Keywords:
Psychiatric symptoms
Functioning
Cluster analysis
Antipsychotic

ABSTRACT

The study's goal was to characterize the typology of patient outcomes based on social and occupational functioning and psychiatric symptoms following antipsychotic drug treatment, and to explore predictors of group membership representing the best/worst outcomes. A hierarchical cluster analysis was used to define groups of patients ($n = 1449$) based on endpoint values for psychiatric symptoms, social functioning, and useful work measured up to 30 weeks of treatment. Stepwise logistic regression was used to construct predictive models of cluster membership for baseline predictors, and with 2/4/8 weeks of treatment. Five distinct clusters of patients were identified at endpoint (Clusters A–E). Patients in Cluster A (25.6%, best outcome) had minimal psychiatric symptoms and mild functional impairment, while patients in Cluster D (14.3%) and E (14.8%) (worst outcome) had moderate-to-severe symptoms and severe functional impairment. Occupational functioning, disorganized thinking, and positive symptoms were sufficient to describe the clusters. Membership in the best/worst clusters was predicted by baseline scores for functioning and symptom severity, and by early changes in symptoms with treatment. Psychiatric symptoms and functioning provided complementary information to describe treatment outcomes. Early symptom response significantly improved the prediction of outcome, suggesting that early monitoring of treatment response may be useful in clinical practice.

1. Introduction

The issue of what constitutes "success" in treating patients with schizophrenia has been the focus of a number of studies. While many studies report significant reductions in symptomatology (i.e., mean change in score for specific psychiatric scales) as evidence of successful treatment, some groups have suggested the need to meet specific categorical criteria for symptom remission (Andreasen et al., 2005; van Os et al., 2006). Andreasen et al. (2005) define 'remission' in schizophrenia as an absolute threshold across a range of positive and negative symptoms, wherein patients have to maintain a score of mild or less (i.e., Positive and Negative Syndrome Scale item scores ≤ 3) on all items for a minimum of 6 months. However, while these criteria for remission in schizophrenia focus only on sustained improvement in symptoms, remission criteria for other mental disorders have included cognitive and psychosocial measures (for review, see Nasrallah and Lasser, 2006). In turn, other groups have focused on 'recovery' in schizophrenia (Rund, 1990; Torgalsboen and Rund, 1998; Torgalsboen, 1999; Liberman and Kopelowicz, 2005; Corrigan, 2006), a concept that includes improvements in occupational and psychosocial functioning as well as symptomatology.

As an alternative to using pre-specified cut-offs to define remission or recovery, one can employ a *data-driven* approach to identify groups or clusters of patients with differing treatment outcomes. Cluster analysis has been used to assess a variety of outcomes in schizophrenia including course of schizophrenia (Rabinowitz et al., 2007), outcomes for psychiatric symptoms and functioning (Di Michele and Bolino, 2004), alignment of resources with patients' clinical and psychosocial needs (Lora et al., 2001), and symptom subtypes (Dollfus et al., 1996). Cluster analysis is well suited to explore multivariate data that may facilitate identifying subpopulations of patients with distinct profiles in terms of different treatment outcomes (i.e., best vs. worst outcomes). In addition, using a broad range of outcome variables allows one to determine based

☆ This work has been previously presented at scientific meetings including the American Society for Experimental NeuroTherapeutics Annual Conference in Washington, DC, March 3–5, 2005; Society of Biological Psychiatry Annual Conference in Atlanta, GA, May 19–21, 2005; and New Clinical Drug Evaluation Unit Annual Conference in Boca Raton, FL, June 6–9, 2005.

* Corresponding authors. Lipkovich is to be contacted at Tel.: +1 317 651 6095; fax: +1 317 276 7100. Kollack-Walker, Tel.: +1 317 433 4654; fax: +1 317 276 7100.
E-mail addresses: lipkovichia@lilly.com (I.A. Lipkovich), Kollack-Walker_Sara@lilly.com (S. Kollack-Walker).

0165-1781/$ – see front matter © 2008 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.psychres.2008.09.004

on the observed data which outcome measures are most meaningful in discriminating cluster membership.

In this study, we evaluated the importance of including direct measures of functioning in the assessment of treatment outcomes. Our analysis involved two stages. First, we (i) identified clusters of patients having distinct treatment outcomes within a large diverse clinical trial patient population using both disease state symptoms and social and occupational functioning at study endpoint, and (ii) constructed parsimonious clinical descriptors (or rules) to identify cluster membership. Second, we identified variables at baseline and during early treatment that we hypothesized would predict cluster membership.

2. Methods

2.1. Datasets

The present analysis focused on patients diagnosed with schizophrenia, schizophreniform disorder, or schizoaffective disorder. Data were obtained from a subset of 2161 patients randomized to a variety of drug therapies in six double-blind, active-control studies involving olanzapine as one of the treatment arms. As an integral part of each clinical trial, patients were given a complete description of the study, and each patient gave written informed consent prior to participation. These six studies were selected because of similarities in the schedules and scales for assessment of psychiatric symptoms and functioning over roughly 30 weeks. Patients ($n=1449$) from these six studies who had received an antipsychotic drug for at least 10 weeks, and who had a post-baseline measure after 10 weeks of treatment for all eight variables, were included in the analysis. A summary of each study is provided in Table 1.

2.2. Outcome measures

Psychiatric symptoms were assessed using the Positive and Negative Syndrome Scale (PANSS) (Kay et al., 1986). Previous research by Lindenmayer et al. (1995) has demonstrated that the symptoms of schizophrenia best fit a five-factor model. The exact items comprising each factor of the PANSS can differ; and in the present study, we followed the factor analysis by Marder and colleagues (Marder et al., 1997). The five main factor scores of the PANSS included the following: 1) *PANSS Negative factor* (PANSS NEG; seven items: 8–11,13,21,30); 2) *PANSS Positive factor* (PANSS POS; 8 items: 1,3,5,6,14,15,23,26); 3) *PANSS Disorganized factor* (PANSS DIS; 7 items: 2,12,19,24,25, 27,29); 4) *PANSS Hostility factor* (PANSS HOS; 4 items: 4,7,22,28); and 5) PANSS *Anxiety/Depression factor* (PANSS DEP; 4 items: 16–18,20). Scores on individual items of the PANSS ranged from 1 (no symptoms) to 7 (severe symptomatology). To aid in comparisons among the different factor scores, the mean value of the items included in each factor was computed (range: 1 to 7).

Functioning was assessed using two domains of the Quality of Life Scale (QLS) (Heinrichs et al., 1984): (i) *QLS Interpersonal Relations domain* (QLS Intprs) — measuring the qualitative aspects of interpersonal relationships; mean value of eight items, and (ii) *QLS Instrumental Role domain* (QLS Instru) — measuring the level of, and the satisfaction with, occupational role functioning; mean value of four items. The QLS items range from 0 (worst) to 6 (best). These two subscales of the QLS represent the part of the QLS that reflects the functional aspects of a patient's quality of life, including the patient's ability to work and to maintain relationships with others.

As a measure of patient's engagement in "useful activities," we used an item (Amount of Useful Activity/Work, *Useful Work*) from the Resource Utilization questionnaire which was available in all studies. This item measures the amount of time a patient effectively engaged in activities, including school, housework, volunteer activities, or paid employment on a five-point scale: 1 (0% time), 2 (<25% time), 3 (25–50% time), 4 (51–75% time), and 5 (>75% time). Since this instrument has not been previously validated as an effective measure of functioning in patients with schizophrenia, we evaluated its correlations with the QLS items that are most relevant for assessing "useful activities" from data pooled from seven studies at screening evaluation ($N=3171$), specifically, QLS Item 9—Extent of Occupational Role Functioning ($r=0.51$, $P<0.0001$), and QLS Item 11—Degree of Underemployment ($r=0.37$, $P<0.0001$).

2.3. Data analytic procedure

A hierarchical cluster analysis (based on Ward's minimum-variance method) (Everitt, 1993) was used to define groups of patients based on endpoint values for eight continuous measures including the the five PANSS factors, two QLS domains, and "useful work". The endpoints for analyses were based on the last observation carried forward (LOCF) principle. The clustering procedure was applied to the data converted into z scores (having zero means and unit standard deviations) by subtracting the observed mean and by dividing by the standard deviation for each measure. Data from patients who had discontinued prior to Week 10 were not included in this analysis.

The number of clusters was chosen based on the proportion of variation in the data (R^2) captured by the clusters. Specifically, the number of clusters was chosen so that clustering with a smaller number of groups led to a substantial deterioration of R^2. While there are sophisticated methods developed recently for selection of the optimal number of clusters [(e.g. see Yan and Ye, 2007) and references therein], in this presentation, we relied on informal and visual criteria. While anticipating high degree of uncertainty about the "true" number of clusters due to the presence of intermediate clusters (inevitable in data on treatment outcomes of patients who started as a relatively homogeneous group), we therefore were more concerned about capturing the extreme clusters and less so on choosing the optimal way of splitting data among intermediate clusters.

A classification tree algorithm (CART) (Breiman et al., 1983) was used to generate parsimonious descriptors for each cluster in terms of the same outcome variables used for data clustering.

Stepwise logistic regression was used to construct predictive models of class membership using data available at baseline: patient demographics, disease history, symptom severity, pre-existing extrapyramidal symptoms, assigned treatment, and effect of study. Selected significant baseline predictors were then forced in the subsequent modeling step where additional predictors reflecting early change in symptoms and tolerability at 2 and 4 weeks of treatment were identified using forward selection. Tolerability was measured as a maximum severity on a scale of 0 to 3 of any adverse event occurring during Weeks 2 and 4. Finally, predictive models combining predictors at Week 0 (baseline) and Weeks 2 and 4 were constructed. Weeks 2 and 4 were common time points available across the studies and therefore were selected as time points reflecting early change.

SAS® version 8.02 (SAS Institute, Cary, N.C.) was used for construction of clusters, logistic regression analyses, and multiple imputation. Rpart library for S-plus version 6.0 (Insightful Corp.) was used for construction of classification tree (Therneau and Atkinson, 1997).

3. Results

3.1. Identifying distinct outcome profiles

Five distinct clusters of patients were identified based on endpoint values for residual psychiatric symptoms and social and occupational functioning (Fig. 1A). These clusters are illustrated on a biplot display which provides a two-dimensional joint projection of subjects (indicated by symbols) and outcomes (indicated by rays) in a multivariate dataset (Lipkovich and Smith, 2002). To facilitate visual assessment and interpretation of the clusters, a unique symbol and color was used to represent all patients comprising a given cluster. The biplot revealed that while there was a substantial overlap in intermediate clusters, the extreme clusters (labeled A and E on the biplot) represented patients with distinct profiles in terms of both psychiatric symptoms and functional outcomes. Also, measures of functioning and psychiatric symptoms formed two distinct dimensions underlying the observed data so that measures within either group of outcomes were correlated better than measures across

Table 1
Design features and baseline severity of symptoms and functioning in individual studies

Treatment	Study 1 HGBG OLZ-RIS		Study 2 HGDH OLZ-HAL		Study 3 HGGN OLZ-HAL-RIS			Study 4 HGHJ OLZ-ZIP		Study 5 HGJU OLZ-ZIP		Study 6[f] HGJB OLZ-QUE	
Patients randomized	172	167	131	132	159	97	15	277	271	202	192	171	175
PANSS POS at baseline; mean (S.D.)	3.5 (0.8)	3.5 (0.8)	3.3 (0.7)	3.4 (0.7)	3.2 (0.7)	3.2 (0.7)	3.2 (0.6)	3.6 (0.8)	3.7 (0.8)	2.7 (0.8)	2.7 (0.8)	2.8 (0.7)	2.9 (0.7)
PANSS NEG at baseline; mean (S.D.)	3.6 (1.0)	3.5 (0.8)	2.6 (0.9)	2.8 (1.1)	2.9 (0.8)	2.8 (0.8)	2.9 (0.8)	3.7 (0.9)	3.8 (1.1)	2.9 (0.9)	2.8 (0.8)	3.5 (0.6)	3.5 (0.6)
QLS Instru at baseline; mean (S.D.)	1.7 (1.5)	1.6 (1.6)	2.3 (1.7)	2.3 (1.9)	3.3 (0.9)	3.4 (0.9)	3.3 (0.9)	1.6 (1.4)	1.5 (1.4)	2.1 (1.5)	1.9 (1.5)	1.7 (1.6)	1.8 (1.5)
QLS Intprs at baseline; mean (S.D.)	2.1 (1.5)	2.3 (1.2)	2.7 (1.3)	2.7 (1.4)	2.5 (1.2)	2.6 (1.1)	2.5 (1.1)	2.0 (1.0)	1.9 (1.1)	2.5 (1.2)	2.4 (1.2)	2.2 (1.1)	2.3 (1.1)
Dose range (mg/day)	10–20	4–12	5–20	5–20	5–20	2–20	2–10	10–20	80–160	10–20	80–160	10–20	300–700
Relevant entry criteria	Acutely psychotic		First episode		Impaired cognition			Acutely psychotic		Depressed		Prominent negative symptoms	
Study duration	28 weeks		52 weeks		32 weeks			28 weeks		24 weeks		24 weeks	

Abbreviations: HAL, haloperidol; OLZ, olanzapine; RIS, risperidone; QUE, quetiapine; ZIP, ziprasidone.
Study 1 (HGBG): Tran et al. (1997) *J. Clin. Psychopharmacol.*, 17: 407–18. *Study 2* (HGDH): Lieberman et al. (2005) *Arch. Gen. Psychiatry*, 62: 361–70. *Study 3* (HGGN): Keefe et al. (2006) *Schizophr. Res.*, 81: 1–15. *Study 4* (HGHJ): Breier et al. (2005) *Am. J. Psychiatry*, 162: 1879–87. *Study 5* (HGJU): Kinon et al. (2006a) *J. Clin. Psychopharmacol.*, 26: 157–62. *Study 6* (HGJB): Kinon et al. (2006b) *J. Clin. Psychopharmacol.*, 26: 453–61.