Published in final edited form as:

Am J Psychiatry. 2011 November ; 168(11): 1186–1194. doi:10.1176/appi.ajp.2011.11010048.

Diagnostic Stability Revisited: Shifts during the Decade Following First-Admission for Psychosis

Evelyn J. Bromet, PhD¹, Roman Kotov, PhD¹, Laura J. Fochtmann, MD¹, Gabrielle A. Carlson, MD¹, Marsha Tanenberg-Karant, MD¹, Camilo Ruggero, PhD², and Su-wei Chang, PhD³

¹Department of Psychiatry and Behavioral Science, Stony Brook University School of Medicine, Stony Brook, NY

²Department of Psychology, University of North Texas, Denton, TX

³Institute of Biomedical Sciences, Academia Sinica, Nankang, Taipei, Taiwan

Abstract

OBJECTIVE—Diagnostic shifts have been prospectively examined in the short-term, but the long-term stability of initial and follow-up diagnoses have rarely been evaluated.

METHOD—A cohort of 470 first-admission patients with psychotic disorders was systematically assessed at baseline, 6-month, 2-year, and 10-year follow-up. Longitudinal best-estimate consensus diagnoses were formulated after each assessment.

RESULTS—At baseline, the diagnostic distribution was: schizophrenia spectrum disorders 29.6%, bipolar disorder with psychotic features 21.1%, major depression with psychotic features 17.0%, substance-induced psychosis 2.4%, and other psychoses 27.9%. At year 10, the distribution changed to 49.8%, 24.0%, 11.1%, 7.0%, and 8.1%, respectively. Overall, 50.7% changed diagnoses at some point during the study. Most participants who were initially diagnosed with schizophrenia or bipolar disorder retained the diagnosis at year 10 (89.2% and 77.8%, respectively). However, 32.0% of participants (N=98) originally given a non-schizophrenia diagnosis gradually shifted into schizophrenia at year 10. The second biggest shift was to bipolar disorder (10.7% of those not given this diagnosis at baseline). Changes in the clinical picture explained many diagnostic shifts. In particular, poorer functioning and greater negative and psychotic symptoms predicted a subsequent shift to schizophrenia. Better functioning and lower negative and depressive symptoms predicted the shift to bipolar disorder.

CONCLUSIONS—First-admission patients run the risk of being misclassified at early stages in the illness course, including more than 2 years after first hospitalization. Diagnosis should be reassessed at all follow-up points.

Keywords

diagnosis; psychosis; schizophrenia; depression; bipolar disorder

In their classic paper on diagnostic validity, Robins and Guze (1) commented that a change in diagnosis in follow-up studies provides compelling evidence about diagnostic heterogeneity. Shifts in diagnosis mean that reliance on the initial categorization could lead to biased estimates of risk factors, familial aggregation, and prognosis (1–4), as well as

misjudgments about optimal treatment (5). Prospective studies of systematically diagnosed patients with first-episode psychosis have found that by 5-year follow-up, initial diagnoses of schizophrenia or bipolar disorder were retained by 80–90% of patients (3, 6–7), but other diagnoses, such as major depression with psychosis, drug-induced psychosis, and schizophreniform disorder, were frequently revised, suggesting substantial misclassification (7, 8–17). Little is known about diagnostic stability of psychotic disorders beyond five years or the temporal stability of the follow-up diagnoses. The World Health Organization (WHO) Determinants of Outcome of Severe Mental Disorders (DOSMeD) first-contact study suggested that agreement continues to erode with increasing time from initial diagnosis. In that cohort, 12.8% shifted out of the ICD-10 schizophrenia spectrum 13 years later, and 7.5% shifted in, for an overall kappa (k) of 0.43 (18). The Nottingham DOSMeD site found better agreement for DSM-III-R diagnoses ($k=0.60$), with all subjects initially diagnosed with schizophrenia retaining the diagnosis (19). We found a similar level of agreement between baseline and 10-year follow-up diagnosis of schizophrenia in preliminary analyses of the Suffolk County Mental Health Project ($k=0.52$) (20). Long-term studies of mood disorders also reported substantial shifting from major depression to bipolar disorder and to schizophrenia (21–25). Since these studies compared diagnoses at two points in time, it is unknown how often the diagnosis may have shifted before becoming stable.

Information on the determinants of short- and long-term diagnostic shifts is limited. The key determinant involves the evolution of the disorder since diagnosis is based on the presence or duration of specific symptoms and/or decline in functioning (7, 18, 26).

This paper uses a heterogeneous, first-admission sample with psychotic disorders reassessed over a period of 10 years to examine the stability of five broad diagnostic categories: schizophrenia spectrum disorders, bipolar disorder and major depression (MDD) with psychotic features, substance induced psychosis, and other psychotic conditions (primarily Psychosis NOS). Diagnoses were formulated by consensus at baseline and at the 6-month, 2-year and 10-year follow-ups (20). We specifically evaluated (a) the distributions, stability, and trajectories of these diagnostic categories; (b) the associations of changes in symptom severity and treatment with changes in diagnosis; and (c) the ability of early clinical features to forecast the diagnostic changes.

Method

Design

The research reported here is from the Suffolk County Mental Health Project, a naturalistic study of the course and outcome of psychotic disorders. The sampling frame consisted of consecutive first admissions with psychosis to the 12 psychiatric inpatient facilities in Suffolk County, New York, from 1989 to 1995. Inclusion criteria were first admission current or within six months, clinical evidence of psychosis (any positive symptoms or antipsychotic medication), ages 15–60, IQ >70, proficiency with English, and no apparent general medical etiology.

The study was approved annually by the Committees on Research Involving Human Subjects at Stony Brook University and the institutional review boards of participating hospitals. Treating physicians determined capacity to provide consent. The head nurse or social worker referred potentially eligible patients to the study. Written consent was obtained from adult participants and from parents of patients <18 years old.

Face-to-face assessments by master's level mental health professionals were conducted at baseline, 6-month, 2-year, and 10-year follow-up. Medical records and interviews with informants, usually a family member, were obtained at each assessment.

Sample

We initially interviewed 675 participants (72% of referrals); N=628 met the eligibility criteria. Forty-two participants died during the 10 years. Of the remaining 586 participants, N=470 (80.2%) were successfully contacted at 10-year follow-up and comprise the analysis sample. Reasons for non-participation were refusal (N=61), untraceable (N=36, including 9 who left the country), uncooperative relatives (N=10), and lacking capacity to provide consent (N=9).

Diagnosis

At baseline, month 6, and year 2, we administered the Structured Clinical Interview for DSM-III-R (SCID; 27); at year 10, we administered sections of the SCID for DSM-IV. Follow-up SCIDs covered the interval from last assessment. The interviewers were aware of previous SCID information. The depression module was administered without skip-outs. We inserted items about severity of suicide attempts and aggression. SCID symptom ratings integrated interview, medical record, and significant other sources. The SCID trainer observed 5–10% of interviews. Agreement (intraclass correlation) for the baseline to 2-year waves was 0.75 for psychotic symptoms, 0.78 for negative symptoms, and $k=0.73$ for depressive symptoms; at 10 years, the corresponding statistics were 0.81, 0.87, and 0.79 (28–30).

The primary study diagnosis was determined by consensus. At baseline, two psychiatrists independently completed the SCID diagnosis module; inconsistent diagnoses, occurring for <10% of participants, were reviewed by a third psychiatrist (13). At follow-up, at least four psychiatrists formulated best-estimate longitudinal consensus diagnoses from information accumulated over time (except prior research diagnoses), including the interviewers' narratives (7, 20). If consensus was not reached or the diagnosis did not fit a DSM category, the diagnosis was coded as 'unknown.' At baseline, 12.8% of diagnoses were unknown (60/470). At follow-up, the figures were: 4.0% at 6 months (19/438); 1.7% at 2 years (8/459); and 0.9% at 10 years (4/470). 'Unknown' was included in the Other category.

Baseline diagnoses used DSM-III-R; follow-up diagnoses used DSM-IV. Although DSM-III-R and IV criteria varied somewhat, a review of 6-month diagnoses using both criteria sets indicated that for the broad categories considered here, the differences were negligible (i.e., four 6-month DSM-III-R cases were revised under DSM-IV).

Clinical and Treatment Variables

Eight clinical ratings were obtained at each assessment: (1) negative symptoms (sum of 18 items from the Scale for the Assessment of Negative Symptoms [SANS] (excludes inattentiveness during mental status testing) (31); (2) psychotic symptoms (16 items on delusions and hallucinations from the Scale for the Assessment of Positive Symptoms [SAPS] (30, 32); (3) disorganized symptoms (13 SAPS items on bizarre behavior and thought disorder); (4) depressive symptoms (sum of 9 SCID past-month depression symptoms; 30); (5) mania severity (excitement item of the Brief Psychiatric Rating Scale [BPRS]) (33); (6) suicide attempts (lifetime at baseline; past interval at follow-up); (7) aggression, rated 1=never to 5=frequent violence toward people or property; and (8) Global Assessment of Functioning (GAF) for the best month in the year before baseline and year 10 and in the interval between assessments at month 6 and year 2.

Treatment variables included rehospitalization during follow-up intervals; antipsychotic, antidepressant, and anti-manic medication use at each contact; and substance abuse treatment in the prior 6 months. There was good agreement between self-report and medication information in outpatient records (34).

Statistical Analyses

Agreement of earlier diagnoses with 10-year diagnosis was examined using kappa, positive predictive value, negative predictive value, sensitivity and specificity.

Symptom and treatment determinants of shifts in diagnosis were examined using mixed-effects logistic regression (35) estimated in SAS version 9.1 with PROC NLMIXED. The time-varying symptom composites and treatment variables were entered simultaneously into separate regression models examining changes in each diagnostic category (coded 1=present, 0=absent). Continuous variables were standardized with respect to their grand means and standard deviations (across all subjects and follow-up points) to facilitate interpretation. Slopes of independent variables and the intercepts were random terms in order to model associations for each participant. Time was modeled as a categorical variable to control for average changes in the dependent and independent variables across assessments. The random effects covariance structure was specified as an unstructured covariance matrix.

We then tested whether the variables that were significant in the mixed-effect logistic regression models predicted subsequent shifts in diagnosis. Using structural equation modeling (Mplus version 5.1), we specified cross-lagged models in which the follow-up diagnostic status was predicted jointly by diagnostic status and participant characteristic from the preceding assessment point (online data supplement, Figure 1). In evaluating the models, we examined the Comparative Fit Index, the Tucker Lewis Index, and the Root Mean Square Error of Approximation (36).

Missing data were addressed in structural equation modeling using the full information maximum likelihood method (37), which estimates models from all available data, thus minimizing attrition-related biases. An analogous approach was employed in mixed-effects logistic regression so that data from each participant were included in the analysis. The longitudinal analyses were based on 1837 observations from 470 participants.

Results

Sample characteristics

About half of the 10-year follow-up sample was male (57.2%), under age 28 at baseline (50.4%), and from blue collar households (47.4%) (Table 1). Three-quarters (74.3%) were Caucasian. At baseline, about half (46.4%) had lifetime episodes of MDD (46.4%). One-fifth (21.3%) had a history of frequent/serious aggression.

Compared to non-participants, the 10-year sample had poorer baseline SANS and GAF ratings, and more participants came from blue collar households (Table 1). No other significant differences were found, including the baseline research diagnosis.

Distribution and Stability of Diagnosis

The proportion diagnosed with schizophrenia spectrum disorders increased progressively from 29.6% of the sample at baseline to 49.8% at year 10 (Figure 1). Schizophrenia increased from 20.9% at baseline to 38.1% at year 10, and schizoaffective disorder from 3.4% at baseline to 11.5% at year 10. In contrast, schizophreniform disorder decreased from 5.3% of the sample at baseline to 0.2% at year 10. Eighty percent of participants with baseline schizophreniform disorder were re-diagnosed with schizophrenia or schizoaffective disorder. Similar proportions were diagnosed each time with bipolar (21.1% at baseline; 24.0% at year 10) and substance-induced (4.5% at baseline; 7.0% at year 10) disorders,

while MDD fell from 17.0% at baseline to 11.1% at year 10, and Other disorders decreased from 27.9% at baseline to 8.1% at year 10.

Agreement of baseline with 10-year diagnoses was low (with kappa ranging 0.13 – 0.65), but reasons for inconsistency differed. Schizophrenia showed relatively low negative predictive value and sensitivity but high positive predictive value, indicating low false positive and high false negative rates. In contrast, bipolar, MDD and substance disorders showed relatively weak sensitivity and positive predictive values (prospective consistency), indicating high false negative and false positive rates. The same was true for Other psychoses except that the false negative and false positive rates were higher. Agreement improved over time, with kappa for 2 to 10 year comparisons ranging from 0.69 to 0.76, except for Other psychoses ($k=0.45$).

Patterns of Diagnostic Shifts

To examine the patterns of diagnostic shifts, we focused on the 432 participants given a research diagnosis at all waves (Figure 2). For each baseline category, we traced the number receiving the same diagnosis each time (straight line), the number with the same baseline and 10-year diagnosis but different at 6-months and/or 2-years (squiggly line), and the number with a different diagnosis at year 10 (broken line). Only 49.3% (213/432) retained their original diagnosis each time. Participants initially diagnosed with schizophrenia were most likely to retain the diagnosis throughout (78.6%, 99/126), followed by bipolar disorder (69.4%, 66/95), substance-induced psychosis (56.3%, 9/16), and MDD (42.9%, 33/77). Only a small proportion (8.5%) stayed in the Other category.

The largest proportion of diagnostic shifts was to schizophrenia. Among 306 participants with a non-schizophrenia diagnosis at baseline, 98 (32.0%) were eventually diagnosed with schizophrenia, with one-third of these shifts (36/98) occurring after year 2. Shifts from mood disorders were primarily to schizoaffective disorder (15/23 from MDD, 8/14 from bipolar disorder). The second largest shift was to psychotic bipolar disorder, involving 10.7% of participants with a non-bipolar diagnosis at baseline (36/337); one-third of them (12/36) occurred after year 2. Eleven participants with baseline MDD (14.3%) switched to bipolar disorder, half (5/11) after year 2.

The right half of Figure 2 shows the composition of the 10-year diagnostic groups relative to these trajectories. Looking retrospectively, 68.8% with MDD at 10-year follow-up received the same diagnosis since baseline, followed by bipolar disorder (60.0%), schizophrenia (47.1%), substance-induced psychosis (31.0%), and other disorders (28.6%).

Determinants of Diagnostic Changes

Mixed-effects logistic regression was used to examine the changes in the clinical picture and treatment exposures that contributed to changes in diagnosis. Given the number of comparisons, we focus on findings with $p<0.01$ (Table 3).

The shift to schizophrenia was more likely to occur when there was a decrease in GAF and depression symptoms, an increase in negative and psychotic symptoms, and initiation/reinstatement of antipsychotic medications. Shifts to psychotic mood disorders were associated with improvement on the GAF, increased depressive symptoms, and decreased negative and psychotic symptoms. Improvement on the GAF was particularly pronounced for a shift to bipolar disorder, while the increase in depressive symptoms was especially important for a shift to MDD. The change to bipolar disorder was also associated with an increase in excitement ratings and initiation/reinstatement of mood stabilizers. A change to MDD was preceded by initiation/reinstatement of antidepressants and discontinuation of

mood stabilizers. Rediagnosis to substance-induced psychosis followed initiation/reinstatement of substance abuse treatment.

Antecedents of Diagnostic Shifts

To determine whether we could forecast changes in diagnosis, we selected the significant variables from the mixed-effects logistic regression models and constructed 18 models using structural equation modeling. These models showed reasonably good fit (Table 4). For schizophrenia, poorer GAF, SANS, and SAPS predicted shifts into this category from baseline to month 6 and from year 2 to year 10, but none predicted a shift from month 6 to year 2. For bipolar disorder, better GAF, lower SANS, lower depression, greater excitement, and anti-manic medication antedated shifts from baseline to month 6. The first three also predicted the shift from year 2 to year 10, but only treatment with anti-manic medication forecasted the shift from month 6 to year 2. For MDD, increased depressive symptoms, lower SAPS, use of antidepressants, and not using anti-manic medication predicted a shift from baseline to month 6. None of the selected variables predicted later shifts. For substance-induced psychosis, substance abuse treatment predicted a shift from baseline to month 6, but not at later intervals.

Discussion

This report examined diagnostic stability in a first-admission cohort across 4 assessments over a 10 year period. We had earlier found considerable shifting between baseline, 6-month, and 2-year follow-up, most notably from MDD and Other psychoses to schizophrenia (7, 13). In this paper, we again found a substantial number of revisions at the 10-year follow-up, including 20.7% whose diagnosis changed from year 2 to year 10. Only half of the cohort retained the same diagnosis throughout the study. Changes in clinical symptoms and treatment were important determinants of shifts in diagnosis. The observed effects were fully consistent with expectations, except that disorganized symptoms and rehospitalization were largely unrelated to change in diagnosis. Furthermore, some diagnostic changes could be anticipated in advance. Participants who did not meet criteria for schizophrenia but exhibited poor functioning and greater negative and psychotic symptoms were likely to shift into that category, whereas better functioning, and lower negative and depressive symptoms predicted a later shift to bipolar disorder.

Our findings must be viewed within the context of the limitations of the study. First, the sample included patients hospitalized with psychotic symptoms, and the results may not generalize to patients who were never hospitalized or did not have co-occurring psychosis. Second, the substance-induced and MDD groups were small, and thus the modeling analyses were able to detect antecedents only for the baseline to 6-month shifts, when changes were more common. Third, our measure of mania severity (BPRS excitement) was crude. Our study began with a focus on schizophrenia and was designed to minimize exclusion of false negatives. We later realized that many participants had a primary mood disorder. Starting at 2-year follow-up, we added a mania rating scale. Fourth, DSM-IV was published as we were completing the 6-month diagnoses. We updated all of the 6-month diagnoses but were unable to recheck the baseline diagnoses. However, only four 6-month diagnoses warranted a change. The fact that the shifts occurred across the 10 year period and were not limited to the baseline to 6-month period also suggests that this is not the primary explanation for our findings. Fifth, our diagnoses were formulated by consensus, which precluded examining inter-rater (psychiatrist) agreement. Sixth, we do not know precisely when, during the 2- to 10-year follow-up, the diagnostic team would have concluded that a change in diagnosis was warranted. Lastly, the interviewers and psychiatrists had access to multiple longitudinal sources of information and were blind only to prior research diagnoses. Diagnoses established in this fashion are not comparable to diagnoses determined by clinical judgment

or cross-sectional SCID ratings. However, longitudinal information is essential to most diagnoses, and we wished to improve the accuracy of the consensus diagnosis with the best possible chronological record of the evolution of the disorder. If anything, having access to longitudinal information should have led the research psychiatrists to maintain the same diagnosis without clear evidence to the contrary. All in all, the diagnostic instability reported here should be regarded as a “best case scenario.”

We could not locate prior studies that considered serial research diagnoses in samples with psychosis. However, studies have examined serial clinical diagnoses in treatment samples (e.g., 39–40). These studies also report temporal variability in diagnosis, with schizophrenia having the best and personality disorders the worst agreement.

Changes in diagnosis may have a number of explanations. By definition, some diagnoses require specific temporal patterns of symptoms (e.g., schizoaffective disorders). Other diagnoses, such as bipolar disorder, include episodes with different polarities that take time to unfold. Many symptoms, such as social withdrawal or agitation, may be present in more than one disorder. In terms of psychotic symptoms, none is pathognomonic of a specific diagnosis. At the time of initial presentation, there are often gaps or ambiguities in the information available to establish a diagnosis. In addition, we included participants with significant substance use histories to construct a generalizable sample, and this too may have confounded the clinical presentation and ultimately the diagnoses. Events occurring just before an acute decompensation may be “red herrings” when viewed in the context of the overall illness course and take on undue weight in the initial diagnosis. Thus, it is not surprising that first-episode patients run the greatest risk of a shift in diagnosis with longitudinal assessment.

Our results demonstrate that diagnostic reconsideration is often linked to changes in functioning or symptoms. In addition, we found that changes in medication regimens administered by community physicians forecasted shifts in diagnosis. Although at first glance, this may seem tautological, the treatments administered by community physicians could also reflect their awareness of diagnostically important symptoms, even when they are subsyndromal or at low to moderate levels of severity. This raises the possibility that strict adherence to the diagnostic criteria may have led us to miss clues utilized by practitioners and misjudge the illness initially. There is obviously a tension between strict implementation of diagnostic criteria, in which equal weight is given to all of the components, and community diagnoses that consider some symptoms and behaviors as more salient than others. Nevertheless, if we regard our 10-year diagnoses as the gold standard, then half of the study population was misclassified based on our initial rigorous application of DSM criteria. This is a very concerning finding given that patients have treatments (with their associated side effects) recommended on a long-term basis based upon presumptive diagnoses that our data suggest have a 50–50 chance of being revised. Misclassification also has serious implications for research by promoting non-reproducible results and potentially erroneous conclusions across a broad range of studies (e.g., therapeutic indications and outcome predictors; biomarkers; genetics; etiologic factors of specific disorders).

The DSM-III and subsequent revisions have been celebrated for introducing a reliable system of classification. Yet most assessments of diagnostic reliability have focused on inter-rater reliability at a given point in time rather than temporal reliability of initial and subsequent diagnoses determined from prospective research. Our results make clear that a reliable cross-sectional diagnosis may still have poor reliability over time. Conceivably, representative samples, as opposed to participants in clinical trials, include many patients who do not have ‘classic’ clinical presentations. Development of future criteria for

psychiatric diagnosis will need to give greater consideration to temporal reliability and predictive validity, rather than cross-sectional reliability of diagnosis.

Robins and Guze (1), writing before the publication of DSM-III, were prescient in drawing attention to the fundamental importance of longitudinal diagnosis for both research and clinical care. Our results, along with a recent study of mood disorder (26), reinforce the importance of reassessing diagnosis over the long-term. As a naturalistic study, our findings highlight the complexity of formulating a diagnosis in the face of multiple comorbidities (e.g., mood symptoms, psychotic symptoms, substance use). They also emphasize the clinical significance of judiciously integrating longitudinal information from multiple sources. Finally, these findings underscore the need to periodically re-evaluate clinical diagnoses to assure that patients are receiving appropriate interventions.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We gratefully acknowledge the support of the participants and mental health community of Suffolk County for graciously contributing their time and energy to this project. We also are indebted to the interviewers for their careful assessments, and to the psychiatrists who contributed to the consensus diagnoses: Alan Brown, Eduardo Constantino, Thomas Craig, Frank Dowling, Shmuel Fennig, Silvana Fennig, Beatrice Kovaszny, Alan Miller, Ramin Mojtabai, Bushra Naz, Joan Rubinstein, Carlos Pato, Michele Pato, Ranganathan Ram, Charles Rich, and Ezra Susser. Special thanks to Janet Lavelle and Al Hamdy for coordinating the study.

FUNDING: National Institute of Health (MH-44801 to EJB)

REFERENCES

1. Robins E, Guze SB. Establishment of diagnostic validity in psychiatric illness: its application to schizophrenia. *Am J Psychiatry*. 1970; 126:983–987. [PubMed: 5409569]
2. Kendler KS. The impact of diagnostic misclassification on the pattern of familial aggregation and coaggregation of psychiatric illness. *J Psychiatr Res*. 1987; 21:55–91. [PubMed: 3560007]
3. Chang WC, Chan SSM, Chung DWS. Diagnostic stability of functional psychosis: a systematic review. *Hong Kong J Psychiatry*. 2009; 19:30–41.
4. Höfler M. The effect of misclassification on the estimation of association: a review. *Int J Methods Psychiatr Res*. 2005; 14:92–101. [PubMed: 16175878]
5. Craig TJ, Fennig S, Tanenberg-Karant M, Bromet EJ. Rapid vs delayed readmission in first-admission psychosis: quality indicators for managed care? *Ann Clin Psychiatry*. 2000; 12:233–238. [PubMed: 11140925]
6. Salvatore P, Baldessarini RJ, Tohen M, Khalsa H-M, Sanchez-Toledo JP, Zarate CA, Vieta E, Maggini C. McLean-Harvard international first-episode project: two-year stability of ICD-10 diagnoses in 500 first-episode psychotic disorder patients. *J Clin Psychiatry*. 2010 Jul 13. e-pub.
7. Schwartz JE, Fennig S, Tanenberg-Karant M, Carlson G, Craig T, Galambos N, Lavelle J, Bromet EJ. Congruence of diagnoses 2 years after a first-admission diagnosis of psychosis. *Arch Gen Psychiatry*. 2000; 57:593–600. [PubMed: 10839338]
8. Amini H, Alaghband-rad J, Omid A, Sharifi V, Davari-Ashtiani R, Momeni F, Aminipour Z. Diagnostic stability in patients with first-episode psychosis. *Australas Psychiatry*. 2005; 13:388–392. [PubMed: 16403137]
9. Baldwin P, Browne D, Scully PJ, Quinn JF, Morgan MG, Kinsella A, Owens JM, Russell V, O'Callaghan E, Waddington JL. Epidemiology of first-episode psychosis: illustrating the challenges across diagnostic boundaries through the Cavan-Monaghan study at 8 years. *Schizophr Bull*. 2005; 31:624–638. [PubMed: 15944446]

10. Castagnini A, Bertelsen A, Berrios GE. Incidence and diagnostic stability of ICD-10 acute and transient psychotic disorders. *Compr Psychiatry*. 2008; 49:255–261. [PubMed: 18396184]
11. Caton CL, Hasin DS, Shrout PE, Drake RE, Dominguez B, First MB, Samet S, Schanzer B. Stability of early-phase primary psychotic disorders with concurrent substance use and substance-induced psychosis. *Br J Psychiatry*. 2007; 190:105–111. [PubMed: 17267925]
12. Crebbin K, Mitford E, Paxton R, Turkington D. First-episode drug-induced psychosis: a medium term follow up study reveals a high-risk group. *Soc Psychiatry Psychiatr Epidemiol*. 2009; 44:710–715. [PubMed: 19183816]
13. Fennig S, Kovasznay B, Rich C, Ram R, Pato C, Miller A, Rubinstein J, Carlson G, Schwartz JE, Phelan J, Lavelle J, Craig T, Bromet E. Six-month stability of psychiatric diagnoses in first-admission patients with psychosis. 1994; 151:1200–1208.
14. Haahr U, Friis S, Larsen TK, Melle I, Johannessen JO, Opjordsmoen S, Simonsen E, Rund BR, Vaglum P, McGlashan T. First-episode psychosis: diagnostic stability over one and two years. *Psychopathology*. 2008; 41:322–329. [PubMed: 18635935]
15. Naz B, Bromet EJ, Mojtabai R. Distinguishing between first-admission schizophreniform disorder and schizophrenia. *Schizophr Res*. 2003; 62:51–58. [PubMed: 12765743]
16. Strakowski SM. Diagnostic validity of schizophreniform disorder. *Am J Psychiatry*. 1994; 151:815–824. [PubMed: 8184991]
17. Whitty P, Clarke M, McTigue O, Browne S, Kamali M, Larkin C, O'Callaghan E. Diagnostic stability four years after a first episode of psychosis. *Psychiatr Serv*. 2005; 56:1084–1088. [PubMed: 16148321]
18. Craig, TJ.; Wanderling, JA.; Janca, A. Long-term diagnostic stability in international cohorts of persons with schizophrenia and related psychoses, in *Recovery from Schizophrenia: An International Perspective*. Hopper, K.; Harrison, G.; Janca, A.; Sartorius, N., editors. New York: Oxford University Press; 2007. p. 50-60.
19. Mason P, Harrison G, Croudace T, Glazebrook C, Medley I. The predictive validity of a diagnosis of schizophrenia. *Br J Psychiatry*. 1997; 170:321–327. [PubMed: 9246249]
20. Bromet EJ, Naz B, Fochtmann LJ, Carlson GA, Tanenberg-Karant M. Long-term diagnostic stability and outcome in recent first-episode cohort studies of schizophrenia. *Schizophr Bull*. 2005; 31:639–649. [PubMed: 15976012]
21. Clayton PJ, Guze SB, Cloninger CR, Martin RL. Unipolar depression: diagnostic inconsistency and its implications. *J Affect Disord*. 1992; 26:111–116. [PubMed: 1447428]
22. Goldberg JF, Harrow M, Whiteside JE. Risk for bipolar illness in patients initially hospitalized for unipolar depression. *Am J Psychiatry*. 2001; 158:1265–1270. [PubMed: 11481161]
23. Kessing LV. Diagnostic stability in depressive disorder as according to ICD-10 in clinical practice. *Psychopathology*. 2005; 38:32–37. [PubMed: 15714010]
24. Ruggero CJ, Carlson GA, Kotov R, Bromet EJ. Ten-year diagnostic consistency of bipolar disorder in a first-admission sample. *Bipolar Disord*. 2010; 12:21–31. [PubMed: 20148864]
25. Vetter P, Köller O. Stability of diagnoses in various psychiatric disorders: a study of long-term course. *Psychopathology*. 1993; 26:173–180. [PubMed: 8234632]
26. Fiedorowicz JG, Endicott J, Leon AC, Solomon DA, Keller MB, Coryell WH. Subthreshold hypomanic symptoms in progression from unipolar major depression to bipolar disorder. *Am J Psychiatry*. 2011; 168:40–48. [PubMed: 21078709]
27. First MB, Spitzer RL, Gibbon M, Williams JBM. Structured clinical interview diagnostic (SCID) for DSM-IV axis I disorders-clinician version (SCID-CV). 1997
28. Bromet EJ, Schwartz JE, Fennig S, Geller L, Jandorf L, Kovasznay B, Lavelle J, Miller A, Pato C, Ram R, Rich C. The epidemiology of psychosis: the Suffolk County Mental Health Project. *Schizophr Bull*. 1992; 18:243–255. [PubMed: 1621071]
29. Brown AS, Susser ES, Jandorf L, Bromet EJ. Social class of origin and cardinal symptoms of schizophrenic disorders over the early illness course. *Soc Psychiatry Psychiatr Epidemiol*. 2000; 35:53–60. [PubMed: 10784367]
30. Kotov R, Guey LT, Bromet EJ, Schwartz JE. Smoking in schizophrenia: diagnostic specificity, symptom correlates, and illness severity. *Schizophr Bull*. 2010; 36:173–181. [PubMed: 18562340]

31. Andreasen, NC. The Scale for the Assessment of Negative Symptoms (SANS). Iowa City: The University of Iowa; 1983.
32. Andreasen, NC. The Scale for the Assessment of Positive Symptoms (SAPS). Iowa City: The University of Iowa; 1984.
33. Woerner M, Manuzza S, Kane J. Anchoring the BPRS: an aid to improved reliability. *Psychopharmacol Bull.* 1988; 24:112–124. [PubMed: 3387514]
34. Mojtabai R, Lavelle J, Gibson PJ, Bromet EJ. Atypical anti-psychotics in first-admission schizophrenia: medication continuation and outcomes. *Schizophr Bull.* 2003; 29:519–530. [PubMed: 14609245]
35. Verbeke, G.; Molenberghs, G. Linear mixed models for longitudinal data. New York: Springer-Verlag; 2000.
36. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling.* 1999; 6:1–55.
37. Schafer JL, Graham JW. Missing data: our view of the state of the art. *Psychol Methods.* 2002; 7:147–177. [PubMed: 12090408]
38. McArdle JJ, Hamagami F. Modeling incomplete longitudinal and cross-sectional data using latent growth structural models. *Exp Aging Res.* 1992; 18:145–166. [PubMed: 1459161]
39. Baca-Garcia E, Perez-Rodriguez MM, Basurte-Villamor I, Lopez-Castroman J, Fernandez del Moral AL, Jimenez-Arriero MA, Gonzalez de Rivera JL, Saiz-Ruiz J, Murillo-Leiva JM, de Prado-Cumplido M, Santiago-Mozos R, Artes-Rodriguez A, Oquendo MA, de Leon J. Diagnostic stability and evolution of bipolar disorder in clinical practice: a prospective cohort study. *Acta Psychiatr Scand.* 2007; 115:473–480. [PubMed: 17498159]
40. Forrester A, Owens DG, Johnstone EC. Diagnostic stability in subjects with multiple admissions for psychotic illness. *Psychol Med.* 2001; 31:151–158. [PubMed: 11200954]

Figure 1. Cross-sectional Distributions of Research Diagnoses at Each Assessment Point
 Baseline and 10-year N=470; 6-month N=438; 24-month N=459

			10-Year Study Diagnoses				
			Schizo- phrenia 210 %	Bipolar 110 %	MDD 48 %	Substance induced 29 %	Other 35 %
Baseline (N)		10 Year (N)					
Schizo- phrenia (126)		Schizophrenia (13) Schizophrenia (99) Bipolar (5) MDD (3) Substance (1) Other-Psy (5)	6.2 <u>47.1</u>	4.5	6.3	3.4	14.3
Bipolar Disorder (95)		Bipolar (8) Bipolar (66) Schizophrenia (14) MDD (1) Substance (3) Other (3)	6.7	7.3 <u>60.0</u>	2.1	10.3	8.6
MDD (77)		MDD (1) MDD (33) Schizophrenia (23) Bipolar (11) Substance (4) Other (5)	11.0	10.0	2.1 <u>68.8</u>	13.8	14.3
Sub- stance (16)		Substance (0) Substance (9) Schizophrenia (2) Bipolar (1) MDD (1) Other (3)	1.0	0.9	2.1	0 <u>31.0</u>	8.6
Other (118)		Other (9) Other (10) Schizophrenia (59) Bipolar (19) MDD (9) Substance (12)	28.1	17.3	18.8	41.4	25.7 <u>28.6</u>

Figure 2. Pattern of Shifts for 432 Study Participants Diagnosed at All 4 Assessment Points^a

^aStraight line: same diagnosis at baseline (BL), 6-months, year 2, and year 10. Squiggly line: same diagnosis at BL and year 10, but different diagnosis at 6 months or year 2. Broken line: final study diagnosis different from baseline.

Table 1

Baseline Characteristics of Follow-up Cohort (N=470) and Surviving Non-Participants (N=116)

	10-Year Cohort		Non-Participants	
<i>Demographic characteristics</i>				
Male	57.2		62.9	
Age less than 28	50.4		44.8	
Blue collar household ^{**}	47.4		33.6	
Race, Caucasian	74.3		75.0	
<i>Clinical history</i>				
Lifetime suicide attempt	28.3		19.8	
Lifetime major depressive disorder	46.4		38.8	
History of frequent aggression ^a	21.3		18.1	
<i>Research diagnosis</i>				
Schizophrenia spectrum disorders	29.6		30.2	
Bipolar I with psychosis	21.1		24.1	
Major depression with psychosis	17.0		15.5	
Substance induced psychosis	4.5		4.3	
Other/undetermined psychosis	27.9		25.9	
<i>Ratings</i>	<i>Mean</i>	<i>(SD)</i>	<i>Mean</i>	<i>(SD)</i>
GAF (best month in year before hospitalization) ^{**}	57.6	(14.1)	62.0	(14.8)
Negative symptoms [*]	17.6	(13.6)	14.7	(12.5)
Positive symptoms	11.3	(9.3)	12.3	(9.8)
Disorganized symptoms	6.8	(6.5)	6.6	(6.0)
Excitement	1.6	(1.1)	1.7	(1.2)

^a Coded 4–5 on a scale of 1=never to 5=frequently^{*} p<0.05^{**} p<0.01

Table 2

Agreement of Baseline, Month 6 and Year 2 Research Diagnosis with 10-year Diagnosis

Diagnosis (N ^a)	Kappa	Positive Predictive Value (%)	Negative Predictive Value (%)	Sensitivity (%)	Specificity (%)
Schizophrenia (10-year N=234)					
Baseline to year 10	0.47	89.2	66.8	53.0	93.6
Month 6 to year 10	0.57	88.5	73.0	64.7	91.9
Year 2 to year 10	0.75	95.2	82.4	78.8	96.1
Bipolar disorder (10-year N=113)					
Baseline to year 10	0.65	77.8	90.3	68.1	93.8
Month 6 to year 10	0.66	72.6	91.9	76.6	90.2
Year 2 to year 10	0.76	79.3	95.3	85.7	92.8
Major depression (10-year N=52)					
Baseline to year 10	0.48	45.0	95.9	69.2	89.5
Month 6 to year 10	0.53	47.0	97.5	81.2	88.7
Year 2 to year 10	0.70	62.3	99.0	92.3	92.9
Substance-induced (10-year N=33)					
Baseline to year 10	0.37	52.4	95.1	33.3	97.7
Month 6 to year 10	0.49	50.0	96.8	55.2	96.1
Year 2 to year 10	0.69	71.0	97.9	71.0	97.9
Other (10-year N=38)					
Baseline to year 10	0.13	15.3	94.7	52.6	74.3
Month 6 to year 10	0.24	26.5	94.3	37.1	91.1
Year 2 to year 10	0.45	46.5	95.7	52.6	94.5

NOTE: positive predictive value=prospective consistency; sensitivity=retrospective consistency

^aN = 470 at baseline and 10 years, 438 at 6 months, and 464 at 2 years

Table 3

Determinants of Diagnostic Shifts: Mixed-Effects Logistic Regression

	Schizophrenia			Bipolar Disorder			MDD			Substance-Induced		
	Adjusted Odds Ratio	95% Confidence Interval		Adjusted Odds Ratio	95% Confidence Interval		Adjusted Odds Ratio	95% Confidence Interval		Adjusted Odds Ratio	95% Confidence Interval	
Changes in symptoms												
GAF	0.28 ***	0.19–0.41		4.61 ***	2.80–7.59		2.12 **	1.32–3.41		1.31	0.73–2.35	
Negative symptoms	3.02 ***	2.11–4.32		0.42 ***	0.25–0.69		0.51 **	0.31–0.85		0.89	0.49–1.60	
Psychotic symptoms	3.49 ***	2.43–5.01		0.45 ***	0.28–0.71		0.31 ***	0.17–0.57		0.65	0.36–1.19	
Disorganized symptoms	0.70 *	0.50–1.00		1.45	0.98–2.14		0.75	0.43–1.32		0.85	0.47–1.54	
Depression symptoms	0.45 ***	0.33–0.63		1.76 **	1.19–2.61		4.23 ***	2.77–6.46		1.21	0.76–1.91	
BPRS excitement	0.70 *	0.51–0.97		1.66 **	1.19–2.30		0.49 *	0.27–0.88		0.91	0.56–1.47	
Suicide attempts	0.63	0.26–1.52		0.57	0.17–1.95		3.25 *	1.18–8.90		0.96	0.28–3.25	
Aggression	0.79	0.43–1.44		1.00	0.49–2.04		0.90	0.42–1.94		2.83 *	1.14–7.05	
Changes in treatment												
Rehospitalization	1.27	0.66–2.43		1.33	0.58–3.03		0.50	0.21–1.21		2.23	0.77–6.47	
Antipsychotics	6.92 ***	3.62–13.25		0.52	0.25–1.06		0.46 *	0.22–0.99		0.46	0.19–1.12	
Antidepressants	0.61	0.33–1.11		0.39	0.18–0.82		3.85 ***	1.93–7.69		0.62	0.24–1.58	
Anti-manic	0.44 *	0.22–0.86		13.92 ***	6.81–28.45		0.22 ***	0.09–0.55		0.35	0.11–1.17	
Substance abuse	0.42 *	0.19–0.94		0.91	0.36–2.30		0.35 *	0.12–0.99		5.91 ***	2.38–14.65	

NOTE: Odds ratios adjusted for effects of all other predictors

* p<0.05

**

*** p<0.01

*** p<0.001

Table 4
Antecedents of Subsequent Diagnostic Shifts in Psychotic Disorders: Structural Equation Modeling

	Fit Indices ^a			Path coefficients ^b		
	CFI	TLI	RMSEA	Baseline to month 6	6 months to year 2	Year 2 to year 10
Schizophrenia spectrum disorders						
GAF	0.994	0.996	0.055	-0.25***	-0.06	-0.13*
Negative symptoms	0.991	0.994	0.064	0.29***	0.10	0.16**
Psychotic symptoms	0.996	0.996	0.054	0.09*	-0.01	0.11*
Depression symptoms	0.986	0.982	0.106	-0.08	-0.05	0.08
Antipsychotics	0.999	0.999	0.021	0.04	0.00	0.07
Bipolar disorder						
GAF	0.994	0.996	0.049	0.32***	0.04	0.13*
Negative symptoms	0.993	0.996	0.055	-0.34***	0.00	-0.16*
Psychotic symptoms	0.986	0.983	0.108	0.01	0.00	-0.10
Depression symptoms	0.987	0.984	0.098	-0.14**	0.03	-0.12*
BPRS excitement	0.984	0.979	0.116	0.13**	0.07	-0.08
Anti-manic medication	0.996	0.997	0.051	0.18***	0.15*	-0.03
Major depression						
GAF best	0.987	0.986	0.077	0.03	0.00	0.09
Negative symptoms	0.980	0.978	0.096	-0.04	0.02	-0.06
Psychotic symptoms	0.991	0.987	0.072	-0.28***	0.03	-0.12
Depression symptoms	0.984	0.982	0.086	0.36***	0.09	-0.08
Anti-depressants	0.994	0.993	0.056	0.27***	-0.12	0.01
Anti-manic medication	0.994	0.992	0.061	-0.15*	-0.15	0.08
Substance-induced disorder						
Substance treatment	0.996	0.995	0.026	0.25***	0.20	-0.03

NOTE. Variables were selected from significant associations found in the mixed-effects logistic regression analyses.

^aCFI = Comparative Fit Index; TLI = Tucker Lewis Index; RMSEA = Root Mean Square Error of Approximation.

Standardized path coefficient for the symptom or treatment characteristic at one wave to the diagnostic status at the next wave, controlling for diagnosis at the former wave.

b

* $p < 0.05$
** $p < 0.01$
*** $p < 0.001$