Diagnostic Stability in Bipolar Disorder:
A Follow-up Study in 130,000 Patient-Years

Fanny Cegla-Schvartzman, MD[a]; Santiago Ovejero, MD, PhD[a]; Jorge López-Castroman, MD, PhD[b,c]; Nora Palomar-Ciria, MD[d]; Marta Migoya-Borja, MD[a]; Hugo Bello, PhD[e]; Gonzalo Martínez-Alés, MD[f]; and Enrique Baca-García, MD, PhD[a,b,g,h,i,j,k,l,*]

ABSTRACT

Objective: Diagnostic stability is the degree to which a diagnosis remains unchanged during time. Our main objective was to evaluate the diagnostic stability of bipolar disorder (BD) in psychiatric outpatient consultations and determine the socio-demographic variables influencing its stability.

Methods: The Cumulative Register of Cases of the Community of Madrid provided data on all outpatient visits conducted at Madrid's Community Mental Healthcare Centers between 1980–2009. Diagnoses were made according to *ICD-9*/*ICD-10* criteria. Two indices were measured: temporal consistency (maintenance of the diagnosis over time) and diagnostic constancy (presence of BD diagnosis in at least 75% of visits). κ coefficient measured the agreement between diagnoses in the first and last evaluations (prospective and retrospective consistency).

Results: 14,557 patients were diagnosed with BD for at least 1 evaluation and had at least 10 visits and 1 year of follow-up. At first evaluation, 3,988 patients were diagnosed with BD (prospective consistency 50.8%), and at last evaluation 5,396 patients were diagnosed with BD (retrospective consistency 37.5%). A total of 2,026 patients were diagnosed with BD at their first and last evaluations (prospective consistency 18.3%).

Conclusions: This longitudinal study conducted in community mental health centers reflects common diagnostic practices in outpatient settings over a 30-year period (130,000 patient-years). Delay of > 10 years was found to achieve diagnostic stability. Frequent diagnostic shifts were found in relation to BD, the most common being with other affective disorders. Anxiety was also a common misdiagnosis. Greater stability was associated with having been diagnosed after hospitalization, having an age at onset > 25 years, and having an age at diagnosis < 24 years.

J Clin Psychiatry 2021;82(6):20m13764

To cite: Cegla-Schvartzman M, Ovejero S, López-Castroman J, et al. Diagnostic stability in bipolar disorder: a follow-up study in 130,000 patient-years. *J Clin Psychiatry*. 2021;82(6):20m13764.

To share: https://doi.org/10.4088/JCP.20m13764

[a]Department of Psychiatry, Hospital Universitario Fundación Jiménez Díaz, Madrid, Spain
[b]Department of Psychiatry, Centre Hospitalier Universitaire de Nîmes
[c]INSERM Unit 1061 and University of Montpellier, Montpellier, France
[d]Psychiatry Service, Complejo Hospitalario de Soria, Soria, Spain
[e]Department of Applied Mathematics, Universidad de Valladolid, Duques de Soria Campus, Soria, Spain
[f]Department of Epidemiology, Columbia University Mailman School of Public Health, New York, New York
[g]Department of Psychiatry, Universidad Autónoma, Madrid, Spain
[h]Department of Psychiatry, Hospital Universitario Rey Juan Carlos, Móstoles, Spain
[i]Department of Psychiatry, Hospital General de Villalba, Madrid, Spain
[j]Department of Psychiatry, Hospital Universitario Infanta Elena, Valdemoro, Spain
[k]CIBERSAM (Centro de Investigación en Salud Mental), Carlos III Institute of Health, Madrid, Spain
[l]Universidad Católica del Maule, Talca, Chile
**Corresponding author:* Enrique Baca-García, MD, PhD, Department of Psychiatry, Hospital Universitario Fundación Jiménez Díaz, Avda. Reyes Católicos, 2. 28040. Madrid, Spain (ebaca@fjd.es).

Recent estimates suggest a global prevalence of bipolar disorder (BD) of around 45 million cases.[1] Because BD entails substantial functional disability,[2] it is considered a major contributor to the number of years lived with disability worldwide. In addition, BD is associated to an increased risk of all-cause mortality and, in particular, of suicide mortality.[3] As a result, previous research has highlighted the high cost and burden driven by BD in a variety of settings.[4,5]

A common limitation of psychiatric epidemiology studies is that psychiatric diagnoses are based on clinical assessments rather than biological measurements.[6] In the absence of an objective measurement that can serve as a gold standard, interobserver reliability and diagnostic stability over time are key components of the validity of psychiatric diagnoses. Diagnostic stability over time is presumed to be characteristic of psychiatric conditions with a tendency to chronicity and relapses over time, such as BD. However, stability varies markedly across chronic psychiatric disorders. For instance, schizophrenia has been found to be one of the most stable diagnoses.[7–10]

There is a paucity of studies focusing on the diagnostic stability of BD, despite the interest in mental health planning. Most recent studies suggest moderate to high levels of diagnostic stability for BD.[11–20] However, these studies are limited by technical difficulties. For instance, many studies have used only a few evaluation points, 2 or 3 at most, over limited follow-up periods,[21–24] raising concerns about the generalization to wider time periods and suggesting the need for studies including more evaluation points over longer follow-up periods.

The importance of diagnostic stability lies in the fact that those who are misdiagnosed or unstable are inadequately treated, which leads to more hospital admissions and also more suicides.[25,26]

This study estimated the real-world long-term clinical stability of BD diagnoses, using data from repeated outpatient visits, and explored which

Clinical Points

- This longitudinal study was conducted in community mental health centers, in a real-world scenario and in the general population, and reflects conditions in a daily practice over a 30-year period, including both outpatient consultations and hospitalization.
- Frequent diagnostic shifts were found in relation to BD, the most common being with other affective disorders. Anxiety was also a common misdiagnosis.
- Greater stability was observed if age at onset was > 25 years, BD diagnosis was made at age < 24 years, or diagnosis was made after hospitalization.

mental health diagnoses were the most common before and after receiving a BD diagnosis among individuals with no diagnostic stability.

METHODS

Study Setting, Sample, and Measurements

Using the Cumulative Register of Cases of the Community of Madrid, an electronic health care record that includes sociodemographic data and *International Classification of Diseases* (*ICD*) diagnostic codes for all outpatient psychiatric visits held between January 1980–December 2009 at Madrid's Community Mental Healthcare Centers (roughly 14% of Madrid's total population in 1996—5,022,289),[27] we selected all records of adults aged ≥ 18 years who (1) received a BD diagnosis in at least 1 visit and (2) undertook at least 10 visits over the study period (minimal adequacy of care, defined as having ≥ 4 outpatient visits in the last year and use of psychotropic medication, or ≥ 8 outpatient visits with or without a medication, a definition used in prior studies).[28,29] Out of a total population of 691,526 patients that were evaluated during 30 years, 14,557 met these inclusion criteria (Figure 1).

Madrid's Community Mental Healthcare Centers are part of the national health service of Spain, which has universal coverage, is financed by taxes, and has no direct cost for patients. The database includes the entire community of Madrid.

In this register, anonymity was ensured by a numerical coding system based on the assignment of a relational registration number. Accordingly, this study did not require participants' informed consent, in agreement with the Spanish law.[30] This study was overseen by the Institutional Review Board at Instituto de Investigación Sanitaria—Fundación Jiménez Díaz. The RECORD guidelines were followed to report findings.[31]

Diagnoses were made by board-certified psychiatrists in a variety of settings including both outpatient consultations and hospitalization, following guidelines in accordance with the 9th or 10th edition of the *International Classification of Diseases* (*ICD*), taking into account equivalence tables bridging both editions.[32] In addition, records included sociodemographic variables (Table 1).

Data Analysis

We used 2 indices of diagnostic stability:

1. Temporal consistency: the presence or absence of a particular disorder at first and last evaluations. Two indices were considered: prospective consistency and retrospective consistency. Of note, some recent papers use the term *diagnostic stability coefficient* as a synonymous of prospective consistency.[33] Using the broad *ICD-10* F1–F9 categories as diagnoses, we computed prospective consistency comparing diagnoses made at the initial evaluation with those made at the final one, and retrospective consistency comparing diagnoses made at the final evaluation with those made at the initial one.
2. Diagnostic constancy: Since prospective and retrospective consistency were based on only 2 evaluations, they often do not reflect diagnostic processes based on multiple evaluations characteristic of routine clinical practice.[34] We thus included a criterion according to which subjects who received diagnoses of BD in at least 75% of the evaluations were categorized as having a "stable BD diagnosis," since it is a common consensus measure that has been used in previous studies.[35]

We used χ^2 and Fisher exact test to test sociodemographic differences between people with stable and non-stable BD diagnosis. We then applied a multivariable logistic model to examine predictors of diagnostic shift including the significant variables of univariate analysis as covariates (gender, marital status, educational level, employment status, occupation, and type of cohabitation and background; described in Table 1), selecting the final model with a progressive elimination method (the likelihood ratio was used as criteria for model fit).

Survival analyses were used to estimate the time from the beginning of the follow-up to the first diagnosis of BD. Since the follow-up time contributes to the stability of the

Figure 1. Sample Selection

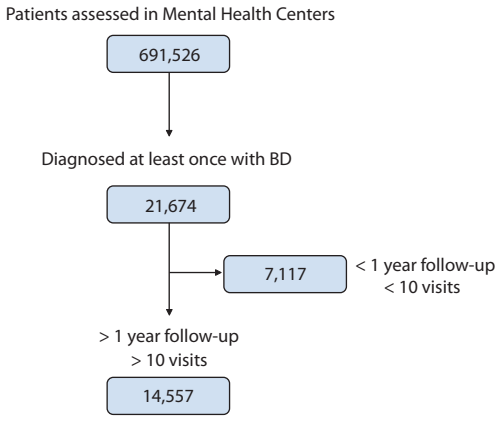


Table 1. Sociodemographic Data

	n	%
Sex		
Female	9,134	63.9
Male	5,161	36.1
Total	14,295	100
Data missing	2	0
Marital status		
Married	7,297	51
Divorced	320	2.2
Single	4,873	34.1
Widower	834	5.8
Separated	568	4
Data missing	405	2.8
Level of education		
Illiterate	365	2.6
No studies	1,458	10.2
Elementary	4,646	32.5
Middle school	2,505	17.5
High school	2,922	20.5
College degree	1,314	9.2
Other	161	1.1
Data missing	926	6.5
Employment status		
Military	18	0.1
Temporal incapacity to work	1,091	7.6
Permanent incapacity to work	271	1.9
Active	4,228	29.6
Looking for first job	197	1.4
Subsidized unemployment	411	2.9
Unsubsidized unemployment	820	5.7
Retirement	1,676	11.7
Rentier	40	0.3
Studying	769	5.4
Work at home	3,427	24
Data missing	1,349	9.4
Employment		
No job	4,549	31.8
Professionals and technicians	1,258	8.8
Management	120	0.8
Administrative	1,116	7.8
Commercial	456	3.2
Hotels and security services	1,385	9.7
Agriculture	124	0.9
Construction industry	763	5.3
Other	4,462	31.2
Armed forces	63	0.4
Data missing	1	0
Residential situation		
Other	586	4.1
Alone	1,216	8.5
Spouse	6,980	48.8
Couple	450	3.1
Family	2,518	17.6
Father only	125	0.9
Mother only	673	4.7
Children	927	6.5
Other family members	506	3.5
Institutionalized	186	1.3
Data missing	130	0.9

diagnosis, the survival predictors were analyzed with the Mantel-Cox model taking into account other covariates and the follow-up time.

RESULTS

Sample Description

A total of 14,557 patients were diagnosed with BD. These patients received 848,147 psychiatric and/or psychological consultations. The mean follow-up time for these patients was 3,295.9 days (standard deviation [SD] = 1,967.6 days), the mean number of visits was 58.3 (SD = 66.7), and the median was 38 visits. Sociodemographic data are shown in Table 1.

Prospective Consistency of Psychiatric Diagnoses

Our consistency comparisons included 15,082 diagnoses made at the initial evaluation and 15,507 at the final one. Figure 2 depicts retrospective diagnostic shifts.

The greatest prospective consistency was found among subjects diagnosed with mood/affective disorders (F3 category): 77.7% of patients diagnosed with F3 in the initial evaluation received a diagnosis under the same category at the final evaluation.

We also found a high prospective consistency in patients diagnosed with schizophrenia, schizotypal disorders, and delusional disorders (F2): 60% of these patients also received the same diagnosis. In contrast, patients diagnosed with mental and behavioral disorders due to the use of psychoactive substances (F1) had a low prospective consistency of 30.5%.

Prospective Consistency of BD Diagnoses

A total of 3,988 patients received a BD diagnosis in their first visit, 5,396 received it in their final visit, and 2,026 received it in both visits. Prospective and retrospective consistencies were, respectively, 50.8% and 37.5%. Cohen κ between first and last BD diagnoses was found to be low (k = 0.17).

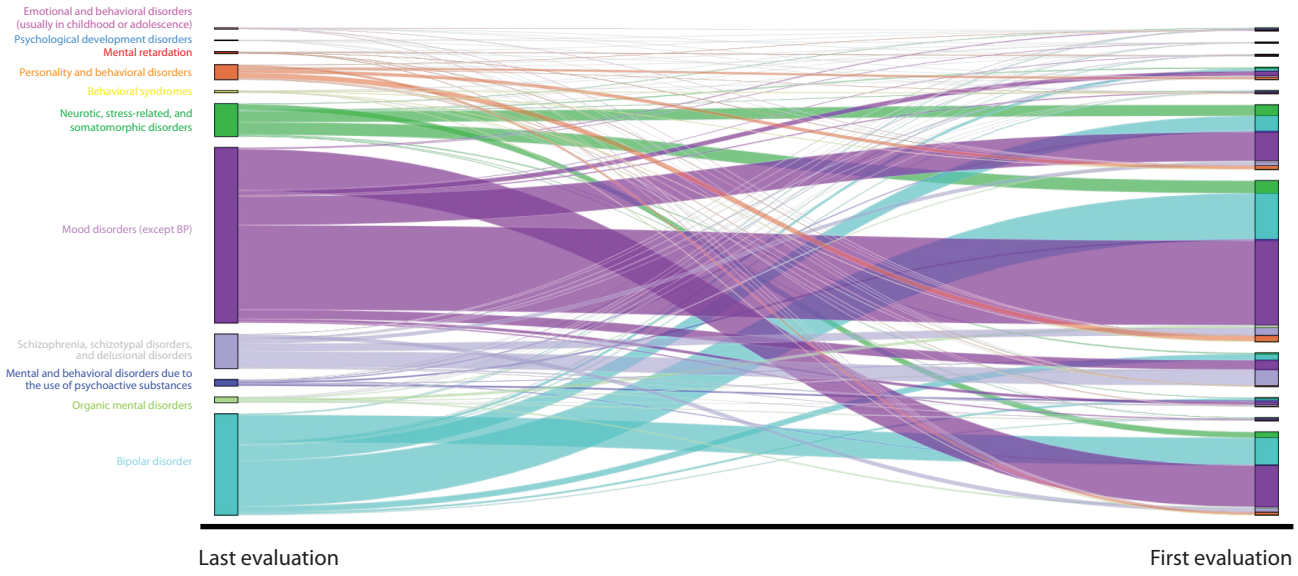
A category that led to diagnostic shift was schizophrenia, schizotypal disorders, and delusional disorders (F2): 8.3% (n = 449) of patients finally diagnosed with BD received an initial F2 diagnosis, and 7.5% (n = 301) of cases who were diagnosed with BD at the beginning had an F2 diagnosis as the final diagnosis.

With regard to non-bipolar affective disorders, out of 8,141 patients initially diagnosed with mood/affective disorders (F3), 51% (n = 4,153) had a non-BD diagnosis. In the final evaluation, 9,717 patients were assigned a mood/affective disorders diagnosis (F3), of which 5,396 were bipolar (F31).

One in 5 patients (21%, n = 1,135) initially diagnosed with neurotic disorders, stress-related disorders, and somato-morphic disorders (F4) were diagnosed with BD in the last visit. Conversely, 10.7% (n = 429) of patients diagnosed with BD at the first evaluation ended up with a diagnosis of neurotic, stress-related, and somato-morphic disorders.

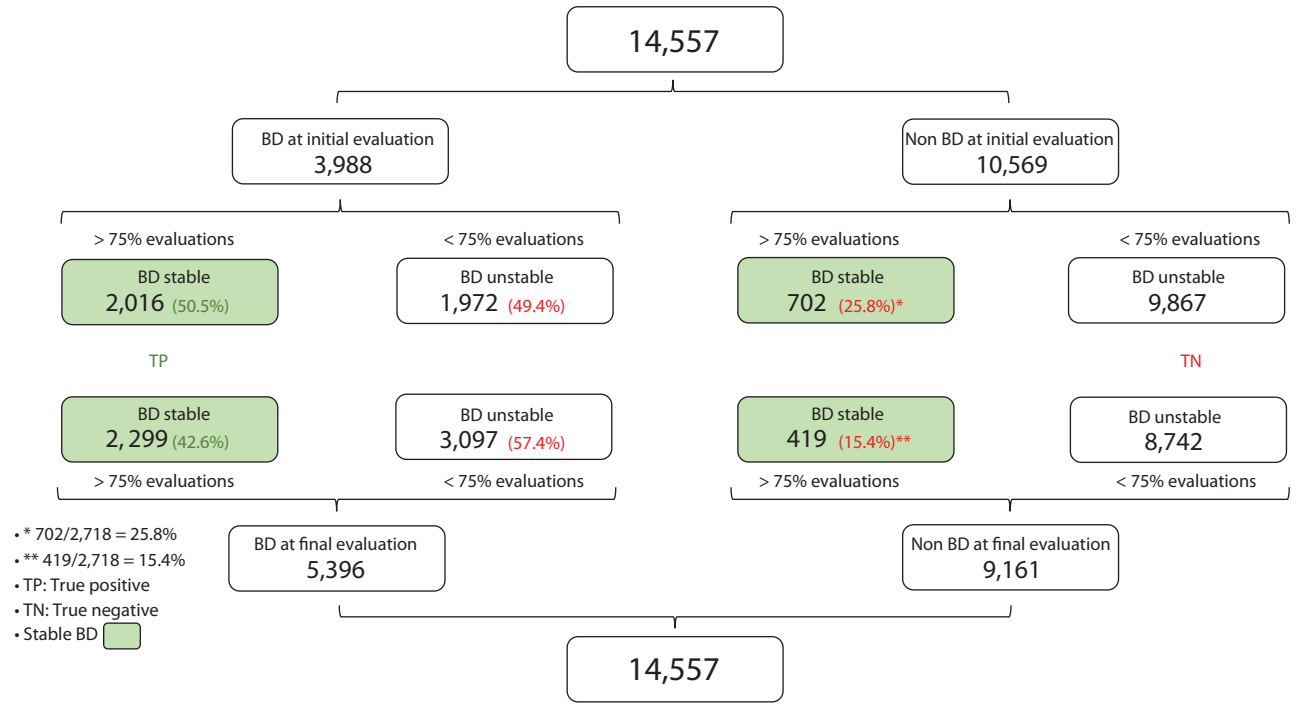
Diagnoses of personality and behavioral disorders in adults (F6) were initially assigned to 3.6% (n = 195) of those with a final diagnosis of BD. Conversely, in the last evaluation, personality and behavioral disorders in adults' diagnoses amounted to 4.2% (n = 167) of those initially categorized as having BD. The remaining diagnostic categories appeared in less than 3% of initial or final BD diagnoses.

Figure 2. Retrospective Consistency in Psychiatric Diagnoses[a]



[a]This graph is an "alluvial diagram." On the left side, the final diagnoses are shown; the width of each bar represents the number of patients with that diagnosis. On the right side, the initial diagnoses are shown, also in proportion. If the reader focuses on how a particular color of bar on the left side (final diagnosis) splits into several other bars on the right side, the proportions of the different initial diagnoses that converge to the same final diagnosis can be traced.

Figure 3. Bipolar Disorder Diagnoses Related to Follow-up Time



Diagnostic Stability of BD Diagnoses

Out of the total sample of 14,557 patients, only 18.6% (n = 2,718) were categorized as having a stable BD diagnosis (eg, retained the BD diagnosis in > 75% of clinical encounters). We summarize the findings in Figure 3.

Among these 2,718 "stable" patients diagnosed with BD, the mean time from the first therapeutic contact with the Mental Healthcare Center to the first time the patient was diagnosed with BD was 318.1 days (95% CI, 188.5–347.8). The average time from the first therapeutic contact within the Mental Healthcare Center to the last time the patient was diagnosed with BD was 7,386.7 days. The median was 7,429 days (95% CI, 7,068.2–7,705.2). There was a difference between the time needed to make the first diagnosis of BD

between those who kept a stable diagnosis (median = 0 days) and those who did not (median = 966 days) (Mantel-Cox test, $\chi^2 = 2{,}852.10$; $P < .0001$). If only patients with a stable BD diagnosis are taken into account, they represent 0.4% of the sample. Taking into account the total number of patients evaluated, about 3.1% had a diagnosis of BD at some point during follow-up.

More than 50% of the sample has been evaluated by the same psychiatrist at least 66% of the time. Among stable BD patients, 77% of the time the patients were seen by the same psychiatrist, and among BD unstable patients, 64% of the time the patients were seen by the same psychiatrist (Student $t_{14.083} = 27.081$, $P < .001$). Among stable BD patients, 14.2% were not evaluated by the same psychiatrist at the first and last evaluation, and among stable BD patients, 28.1% were seen by the same psychiatrist at the first and the last evaluation, with an OR of 2.375 (95% CI, 2.174–2.594).

The agreement found between the first diagnosis with BD stable group compared to the last diagnosis with BD stable group was as follows: BD stable—first diagnosis: k = 0.492, $P \leq .000$; BD stable—last diagnosis: k = 0.420, $P \leq .000$, respectively.

Factors Related to the Diagnostic Stability of BD

The United Nations, for statistical purposes, defines persons between the ages of 15 and 24 years as youth. Following this definition when performing the analysis on the stable BD group (when retained the BD diagnosis in > 75% of clinical encounters), greater stability was found if the diagnosis was made after hospitalization—OR = 1.932 (95% CI, 1.682–2.219), if the age at onset was > 25 years—OR = 4.318 (95% CI, 2.527–7.377), if the diagnosis of BD was made at age < 24 years—OR = 6.133 (95% CI, 3.477–10.817), if > 65% of the visits were held by the same psychiatrist—OR = 2.246 (95% CI, 1.978–2.550), and if the patient had been assessed by the same psychiatrist in the first and last assessments—OR = 1.667 (95% CI, 1.475–1.883) (Hosmer and Lemeshow test: $\chi^2_7 = 10{,}620$, $P = .156$).

DISCUSSION

The present study addresses the issue of diagnostic stability of BD in outpatient settings and contributes to the knowledge about the temporal consistency of BD and the usual diagnostic changes that occur during its evolution.

The results showed a limited number of stable BD diagnoses, notably lower than previous studies. Some methodological reasons could explain the differences with previous studies, especially the low number of evaluations and the shorter follow-up period used in previous studies. To be sure of the diagnosis of stable BD, at least 6–12 months is needed. The administrative prevalence (the proportion of the population in a defined area—the community of Madrid in this study—who are receiving services) of BD in this psychiatric sample is 0.4%, lower than usually reported in clinical and nonclinical samples.[4,36,37] However, this could be related to the fact that diagnoses were made in outpatient settings, and, as reported in previous studies,[16,17] a higher diagnostic stability is observed when the diagnosis is made after a discharge from hospital. A more detailed study of the factors that influence the stability of BD and a better knowledge of the course of diagnoses throughout its evolution are proposed as future lines of research.

The natural evolution of BD is prone to a high variability; however, the central symptoms of affective episodes are not present as frequently, and the presence of comorbid disorders, which is quite common, leads to misdiagnoses during daily clinical practice.[38,39] The present study found that the stability of BD was low, and even lower than in previous studies,[13,16,17,19,40] with the 3 different indices used.

In the first evaluation, 27.4% of the patients received a diagnosis of BD, and only 18.3% of the total sample was considered stable according to the criteria established in this study. Additionally, confusion surrounding the usual differential diagnoses of BD was found. These conclusions are detailed and discussed in the following sections.

It is important to emphasize that the increased specificity of the diagnostic criteria for BD in *ICD-10* versus *ICD-9* may have somewhat influenced our results.

It is essential to clarify that there has been no deinstitutionalization in Spain and that the registry was conceived as a tool during the psychiatric reform. Of note, this is one of the few epidemiologic studies on this issue conducted outside Scandinavia in which multiple types of stability measures were used, and this database has been used in previous works.[35,41]

BD Diagnosis at the First Evaluation

Diagnostic shifts in BD are especially frequent at first contact with the physician, with misleading initial symptoms due to substance abuse, depressive, or psychotic symptoms. The greatest prospective consistency was found in the mood/affective disorders category (F3), since the sample was selected among patients with at least 1 BD diagnosis. While only 27.4% of subjects were diagnosed with BD at the first evaluation, the rest were diagnosed at least once during subsequent evaluations. Similar results were presented in a previous study,[40] where it was noted that these figures were consistent with the high prevalence of misdiagnosis (48% and 69%) found in naturalistic research using self-administered questionnaires in general practitioner consultations[42,43] and also in studies in which diagnoses were based on the application of *DSM-IV* criteria.[44]

However, in our sample, 50.5% of patients who were diagnosed with BD at the first evaluation remained stable in three-quarters of the evaluations. This fact is not consistent with the figure reported by Chen et al,[13] who noted that 70% of the subjects with an initial diagnosis of BD did not change to a different category over time. On the other hand, the percentage of patients with a stable diagnosis of BD (n = 2,718) who were correctly diagnosed in the first evaluation (n = 2,016) increases in our sample to 74.2%. These results support the hypothesis of the diagnostic difficulty of BD in the first evaluations.

BD Diagnosis at the Last Evaluation

The latest evaluation showed an increase in the number of diagnoses of BD (37.1% of the sample), and, of those, 42.6% had been stable throughout the study. On the other hand, 84.6% of patients with stable diagnoses (n = 2,718) were accurately diagnosed in their last visit (n = 2,299).

This result may reflect a progressive increase in diagnostic stability throughout the evaluations (in our case a minimum of 10), which is congruent with the idea that routine reassessment could improve the chances of a successful diagnostic process. However, Schwartz et al in 2000[17] reported that the retrospective consistency of BD was 85% when comparing 6-month and 24-month diagnoses but was reduced to 73% when comparing initial and 24-month diagnoses. This would mean that consistency rates for some diagnoses decreased as the follow-up period increased. In any case, the retrospective consistency of BD in our study (37.5%) is low compared to other studies that measured it (58.4%–94.4%), similar to that presented by Baca-García et al in 2007 (38%),[35] and higher than Weeke in 1984 (20%).[45] The low retrospective consistency might be explained by the fact that this is a longitudinal study based on data retrieved from community mental health centers and hence conducted in a real-world scenario with the general population, and not in a BD-specific unit where patients start off already correctly diagnosed.

Diagnostic Stability of BD

To our knowledge, this is the largest longitudinal study, with 14,557 patients over 30 years of study, that has evaluated the diagnostic stability of BD under ecological conditions. In 2005, Kessing[40] mentioned that no study had investigated the diagnostic stability of the most common *ICD-10* psychiatric diagnoses administered under ecological clinical conditions. This is the case for our study, which has shown a low stability of the *ICD-10* BD categories, measured by temporal consistency and diagnostic constancy, with findings considerably lower than in previous studies. The reasons for these differences in diagnostic temporal stability are not clear but may be due to the large sample size, the extensive duration of follow-up, the high number of evaluations, diagnostic criteria, or sociodemographic variables.

Time consistency showed low results with a prospective consistency of 50.8% and a retrospective consistency of 37.5%. It should be noted that the κ value was low (κ = 0.17) between the first and last diagnosis. However, since κ values take into account stable positive cases and stable negative cases, but also cases that remit and new cases, low κ values can be observed if a high number of new or remitting cases occur[46] and therefore do not necessarily reflect a lack of diagnostic stability.

The results of our study showed that only 18.3% of patients were diagnosed with BD in 75% of the evaluations. In these patients with stable diagnosis, the mean number of evaluations until the first diagnosis of BD was 3.7, with an average time of 318.1 days. These values were increased to 21.2 evaluations and 1,511.2 days within the group with a non-stable diagnosis. Thus, patients with a stable diagnosis of BD were diagnosed earlier (less than 1 year) and needed fewer evaluations than those without a stable diagnosis (somewhat more than 4 years until the BD diagnosis was made).

In our study, patients with a stable BD diagnosis achieved diagnostic stability at 7,386.7 days (slightly more than 20 years) and 279 follow-up visits. Patients with a non-stable diagnosis had their diagnosis withdrawn at 2,929 days (approximately 8 years) and after 55 evaluations. These results could be in line with previous reports by Hirschfeld et al[42] in 2003 and Baldessarini et al[47] in 1999, who reported a delay in correct diagnosis of about 8–10 years from the onset of the disease. In our study, the data suggest that both consolidating and withdrawing the diagnosis of BD are tasks that require many years of follow-up and numerous evaluations; while consolidating required about 14 visits/year, withdrawing entailed fewer than 7 visits/year. This fact may reflect that patients with a stable diagnosis of BD are more complex and require more health care than those for whom this diagnosis is withdrawn.

Although it was not the main objective of our study, 4 variables not only were included but were predictive of diagnostic stability of BD: marital status, educational level, work situation, and personal history of psychiatric care. In another preliminary study,[6] 4 variables related to the stability of bipolar diagnosis were found: sex, age ≥ 40 years, number of psychiatric consultations, and outpatient Mental Health Centers. In any case, more studies focusing on these variables are needed.

The higher consistency rates found by other authors[13,16,17,19,40] may have been influenced by a number of drawbacks that diminish the generalizability of these studies.

Diagnostic Shifts in BD

Patients with a stable BD diagnosis had some diagnostic fluctuation that included the typical differential diagnoses of BD. Our study found high rates of misdiagnosis of BD with other affective disorders: 44.5% of patients who were diagnosed at the first evaluation of a non-bipolar affective disorder were eventually diagnosed at the last follow-up visit with BD, and 51% of patients who were initially diagnosed with BD were no longer diagnosed at the last evaluation. Previous studies concur that the high rates of misdiagnosis derive from confusion with unipolar depression,[42,43] especially in cases in which the BD debuts with 1 or more depressive episodes. As for neurotic and anxiety disorders, the percentage of these diagnoses at the beginning is high (21.03%) in patients who are ultimately diagnosed with BD. Other less frequent diagnostic shifts occurred with the spectrum of schizophrenia (7.5% at first evaluation and 8.3% at the last evaluation) and with personality disorders (4.2% baseline and 3.6% final).

Many factors may be involved in the unstable progression of a psychiatric diagnosis. Schwartz et al[17] mentioned that diagnostic changes over time may reflect the evolution of a disease, the emergence of new information, or the

unreliability of measurements. The relative lack of stability in diagnoses over time in this study may be due to disease progression or reflect weaknesses inherent in clinical evaluations. There could be other ways to determine the validity of stable diagnoses such as the use of prescriptions for mood stabilizers, this being a limitation of the study.

The results of this study raise concerns about psychiatric research findings, especially in studies with short follow-up periods for chronic conditions that may not allow enough time to reach an accurate diagnosis or in studies that do not take into account the context.

Our study has limitations. Our study has limitations. First, we did not consider differences between type I and II bipolar disorder, despite clinical and prognostic implications, because codes were mostly recorded following *ICD-9*, where type II bipolar disorder could not be specified. Second, diagnoses in our database are recorded following the independent judgment of clinicians rather than alternative assessments such as research scales. This, however, enhances external validity of our results, as they likely reflect the course of illness from a real clinical practice perspective. While diagnostic scales can reduce measurement error, they (1) require specific training and are too time-consuming to be used routinely and (2) are mostly validated in the context of highly selected samples of patients, for research purposes. In conclusion, as noted previously in the literature,[48] our results should be considered an externally valid representation of patterns of real clinical diagnostic change, which has important implications for treatment planning, rather than patterns in the prevalence of the disorder. Also, a limitation was that the form filled out at each visit consisted of sociodemographic data and psychiatric diagnosis, leaving out other relevant data.

A point to keep in mind is that the question of whether diagnostic changes in our data (eg, patients whose diagnoses changed from or to bipolar disorder) reflect misdiagnosis, the natural history of the phenotypical presentation of these patients' disease, or a mix of both cannot be clarified using this data source. Accordingly, conclusions regarding over- or underdiagnosis of BD based on our results should be made with caution. The study is also limited by the possible existence of uncontrolled pathways of psychiatric care but may more accurately reflect real clinical practice, perhaps revealing the poor accuracy of clinical evaluation systems in usual practice.

CONCLUSIONS

1. This work reflects real conditions in a daily practice over a 30-year period of observation, including both outpatient consultations and hospitalization.
2. In our sample, the administrative prevalence of stable BD is 0.4%; however, the diagnosed prevalence is 3.1% when all patients diagnosed with BD are included.
3. A delay of > 10 years to achieve diagnostic stability was found.
4. Frequent diagnostic shifts were found in relation to BD, the most common being with other affective disorders. Anxiety was also a common misdiagnosis. There is a 50% diagnostic error rate when BD is diagnosed in the first evaluation.
5. The most diagnostically stable patients are diagnosed at the first visit.
6. Greater stability was observed if age at onset was > 25 years, BD diagnosis was made at age < 24 years, or diagnosis was made after hospitalization (which may explain the low prevalence we found in this study in comparison to other studies in which only hospitalized patients were considered).
7. The low stability detected in this study should be taken into account when evaluating results compared to clinical and epidemiologic trials, in which samples were smaller and followed for less time.

Submitted: November 10, 2020; accepted April 12, 2021.

Published online: September 14, 2021.

Potential conflicts of interest: None.

Funding/support: This study was partly funded by Carlos III (ISCIII PI16/01852), American Foundation for Suicide Prevention (LSRG-1-005-16), the Madrid Regional Government (B2017/BMD-3740 AGES-CM 2CM; Y2018/TCS-4705 PRACTICO-CM), MINECO/FEDER (ADVENTURE, id. TEC2015-69868-C2-1-R), MCIU Explora Grant aMBITION (id. TEC2017-92552-EXP), and MICINN (CLARA, id. RTI2018-099655-B-I00).

Role of the sponsor: The sponsor was not involved in any of the following: design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, or approval of the manuscript.

Acknowledgments: The authors thank the health and administrative staff who participated in the data collection for the study. They also thank the reviewers for their comments and observations, which have contributed to the improvement of this work.

REFERENCES

1. Institute for Health Metrics and Evaluation [Internet]. Institute for Health Metrics and Evaluation. Cited September 28, 2020. http://www.healthdata.org/institute-health-metrics-and-evaluation
2. Ferrari AJ, Stockings E, Khoo J-P, et al. The prevalence and burden of bipolar disorder: findings from the Global Burden of Disease Study 2013. *Bipolar Disord*. 2016;18(5):440–450.
3. Hayes JF, Marston L, Walters K, et al. Mortality gap for people with bipolar disorder and schizophrenia: UK-based cohort study 2000–2014. *Br J Psychiatry*. 2017;211(3):175–181.
4. Hidalgo-Mazzei D, Undurraga J, Reinares M, et al. The real world cost and health resource utilization associated to manic episodes: the MANACOR study. *Rev Psiquiatr Salud Ment*. 2015;8(2):55–64.
5. Pini S, de Queiroz V, Pagnin D, et al. Prevalence and burden of bipolar disorders in European countries. *Eur Neuropsychopharmacol*. 2005;15(4):425–434.
6. Baca-Garcia E, Perez-Rodriguez MM, Basurte-Villamor I, et al. Diagnostic stability of psychiatric disorders in clinical practice. *Br J Psychiatry*. 2007;190(3):210–216.
7. Palomar-Ciria N, Cegla-Schvartzman F, Lopez-Morinigo J-D, et al. Diagnostic stability of schizophrenia: a systematic review. *Psychiatry Res*. 2019;279:306–314.
8. Woo BKP, Sevilla CC, Obrocea GV. Factors influencing the stability of psychiatric diagnoses in the emergency setting: review of 934 consecutively inpatient admissions. *Gen Hosp Psychiatry*. 2006;28(5):434–436.
9. Daradkeh TK. Stability of psychiatric diagnoses in clinical practice. *Int J Soc Psychiatry*. 1996;42(3):207–212.
10. Daradkeh T, El-Rufaie O, Younis Y, et al. The diagnostic stability of *ICD-10* psychiatric diagnoses in clinical practice. *Eur Psychiatry*. 1997;12(3):136–139.
11. Addington J, Chaves A, Addington D. Diagnostic stability over one year in first-episode psychosis. *Schizophr Res*. 2006;86(1–3):71–75.
12. Amin S, Singh SP, Brewin J, et al. Diagnostic stability of first-episode psychosis. Comparison of *ICD-10* and *DSM-III-R* systems. *Br J Psychiatry*. 1999;175(6):537–543.

13. Chen YR, Swann AC, Johnson BA. Stability of diagnosis in bipolar disorder. *J Nerv Ment Dis*. 1998;186(1):17–23.
14. Fennig S, Kovasznay B, Rich C, et al. Six-month stability of psychiatric diagnoses in first-admission patients with psychosis. *Am J Psychiatry*. 1994;151(8):1200–1208.
15. Marneros A, Deister A, Rohde A. Stability of diagnoses in affective, schizoaffective and schizophrenic disorders cross-sectional versus longitudinal diagnosis. *Eur Arch Psychiatry Clin Neurosci*. 1991;241(3):187–192.
16. Schimmelmann BG, Conus P, Edwards J, et al. Diagnostic stability 18 months after treatment initiation for first-episode psychosis. *J Clin Psychiatry*. 2005;66(10):1239–1246.
17. Schwartz JE, Fennig S, Tanenberg-Karant M, et al. Congruence of diagnoses 2 years after a first-admission diagnosis of psychosis. *Arch Gen Psychiatry*. 2000;57(6):593–600.
18. Tsuang MT, Woolson RF, Winokur G, et al. Stability of psychiatric diagnosis. Schizophrenia and affective disorders followed up over a 30- to 40-year period. *Arch Gen Psychiatry*. 1981;38(5):535–539.
19. Veen ND, Selten J-P, Schols D, et al. Diagnostic stability in a Dutch psychosis incidence cohort. *Br J Psychiatry*. 2004;185(6):460–464.
20. Cegla-Schvartzman FB, Ovejero S, López-Castroman J, et al. Diagnostic stability in bipolar disorder: a narrative review. *Harv Rev Psychiatry*. 2019;27(1):3–14.
21. Ratheesh A, Cotton SM, Betts JK, et al. Prospective progression from high-prevalence disorders to bipolar disorder: exploring characteristics of pre-illness stages. *J Affect Disord*. 2015;183:45–48.
22. Amini H, Alaghband-rad J, Omid A, et al. Diagnostic stability in patients with first-episode psychosis. *Australas Psychiatry*. 2005;13(4):388–392.
23. Baldwin P, Browne D, Scully PJ, et al. Epidemiology of first-episode psychosis: illustrating the challenges across diagnostic boundaries through the Cavan-Monaghan study at 8 years. *Schizophr Bull*. 2005;31(3):624–638.
24. Rufino ACTBF, Uchida RR, Vilela JAA, et al. Stability of the diagnosis of first-episode psychosis made in an emergency setting. *Gen Hosp Psychiatry*. 2005;27(3):189–193.
25. Parker G. Borderline personality disorder and bipolar disorder: commentary on Paris and Black. *J Nerv Ment Dis*. 2015;203(1):13–14.
26. Moore PJ, Little MA, McSharry PE, et al. Forecasting depression in bipolar disorder. *IEEE Trans Biomed Eng*. 2012;59(10):2801–2807.
27. Padrón continuo 1996. Total Comunidad de Madrid [Internet]. Available from: http://www.madrid.org/iestadis//fijas/estructu/demograficas/padron/pc96i1.htm
28. Carmona R, Cook BL, Baca-García E, et al. Comparison of mental health treatment adequacy and costs in public hospitals in Boston and Madrid. *J Behav Health Serv Res*. 2019;46(3):464–474.
29. Wang PS, Demler O, Kessler RC. Adequacy of treatment for serious mental illness in the United States. *Am J Public Health*. 2002;92(1):92–98.
30. BOE. BOE - Ley 14/2007 [Internet]. 2007. Available from: https://www.boe.es/boe/dias/2007/07/04/
31. Benchimol EI, Smeeth L, Guttmann A, et al; RECORD Working Committee. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. *PLoS Med*. 2015;12(10):e1001885.
32. eCIE-Maps - CIE-9-MC [Internet]. [cited 2020 Sep 28]. Available from: https://eciemaps.mscbs.gob.es/ecieMaps/browser/index_9_mc.html
33. Pierce K, Gazestani VH, Bacon E, et al. Evaluation of the diagnostic stability of the early autism spectrum disorder phenotype in the general population starting at 12 months. *JAMA Pediatr*. 2019;173(6):578–587.
34. Pettit JW, Morgan S, Paukert AL. The stability of axis I diagnoses in youth across multiple psychiatric hospitalizations. *Child Psychiatry Hum Dev*. 2005;36(1):53–71.
35. Baca-Garcia E, Perez-Rodriguez MM, Basurte-Villamor I, et al. Diagnostic stability and evolution of bipolar disorder in clinical practice: a prospective cohort study. *Acta Psychiatr Scand*. 2007;115(6):473–480.
36. Alonso J, Angermeyer MC, Lépine JP; European Study of the Epidemiology of Mental Disorders (ESEMeD) Project. The European Study of the Epidemiology of Mental Disorders (ESEMeD) project: an epidemiological basis for informing mental health policies in Europe. *Acta Psychiatr Scand suppl*. 2004;109(420):5–7.
37. Bijl RV, Ravelli A, van Zessen G. Prevalence of psychiatric disorder in the general population: results of The Netherlands Mental Health Survey and Incidence Study (NEMESIS). *Soc Psychiatry Psychiatr Epidemiol*. 1998;33(12):587–595.
38. Blacker D, Tsuang MT. Contested boundaries of bipolar disorder and the limits of categorical diagnosis in psychiatry. *Am J Psychiatry*. 1992;149(11):1473–1483.
39. Ghaemi N, Sachs GS, Goodwin FK. What is to be done? controversies in the diagnosis and treatment of manic-depressive illness. *World J Biol Psychiatry*. 2000;1(2):65–74.
40. Kessing LV. Diagnostic stability in bipolar disorder in clinical practise as according to ICD-10. *J Affect Disord*. 2005;85(3):293–299.
41. Lopez-Castroman J, Leiva-Murillo JM, Cegla-Schvartzman F, et al. Onset of schizophrenia diagnoses in a large clinical cohort. *Sci Rep*. 2019;9(1):9865.
42. Hirschfeld RMA, Lewis L, Vornik LA. Perceptions and impact of bipolar disorder: how far have we really come? results of the National Depressive and Manic-Depressive Association 2000 Survey of Individuals With Bipolar Disorder. *J Clin Psychiatry*. 2003;64(2):161–174.
43. Lish JD, Dime-Meenan S, Whybrow PC, et al. The National Depressive and Manic-depressive Association (DMDA) survey of bipolar members. *J Affect Disord*. 1994;31(4):281–294.
44. Ghaemi SN, Sachs GS, Chiou AM, et al. Is bipolar disorder still underdiagnosed? are antidepressants overutilized? *J Affect Disord*. 1999;52(1-3):135–144.
45. Weeke A. Admission pattern and diagnostic stability among unipolar and bipolar manic-depressive patients. *Acta Psychiatr Scand*. 1984;70(6):603–613.
46. Mattanah JJ, Becker DF, Levy KN, et al. Diagnostic stability in adolescents followed up 2 years after hospitalization. *Am J Psychiatry*. 1995;152(6):889–894.
47. Baldessarini RJ, Tondo L, Hennen J. Treatment delays in bipolar disorders. *Am J Psychiatry*. 1999;156(5):811–812.
48. Moreno C, Laje G, Blanco C, et al. National trends in the outpatient diagnosis and treatment of bipolar disorder in youth. *Arch Gen Psychiatry*. 2007;64(9):1032–1039.