Title

Methods for Extracting and Validating Psychiatric Phenotypes: Advancements Using Electronic Health Records in Colombia

Author

De la Hoz Gomez, Juan Fernando

Publication Date

2023

UNIVERSITY OF CALIFORNIA

Los Angeles

Methods for Extracting and Validating Psychiatric Phenotypes: Advancements Using Electronic Health Records in Colombia

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy
in Bioinformatics

by

Juan Fernando De la Hoz Gomez

2023

ABSTRACT OF THE DISSERTATION

Methods for Extracting and Validating Psychiatric Phenotypes: Advancements Using Electronic Health Records in Colombia

by

Juan Fernando De la Hoz Gomez

Doctor of Philosophy in Bioinformatics

University of California, Los Angeles, 2023

Professor Nelson B. Freimer, Co-Chair

Professor Loes Marlein Olde Loohuis, Co-Chair

Biobanks linked to Electronic Health Records (EHRs) herald a new era of opportunities for etiological research of Severe Mental Illness (SMI). However, because EHRs are not primarily designed for research, translating these opportunities into actionable insights demands innovative frameworks and accurate phenotyping tools. This dissertation harnesses the potential of EHRs from psychiatric hospitals for in-depth studies of SMI. I set the stage by contextualizing the relevance of EHRs in psychiatric genetic research. Then, I describe the organizational makeup and data types within the EHR of the Clinica San Juan de Dios in Manizales — a regional psychiatric hospital in Colombia.

The subsequent chapters explore transdiagnostic phenotypes by combining clinical notes and diagnostic codes, leading to the delineation of disease trajectories in SMI. Then, I explore the extraction and validation of psychiatric diagnoses through both rule-based and machine learning strategies. And finally, I conclude with the design and validation of a Clinical Natural Language Processing (cNLP) tool for extracting highly detailed psychiatric phenotypes from unstructured text.

Three strengths of EHRs are emphasized throughout this work: the integration of multi-dimensional data, enabling a comprehensive perspective of patient phenotypes; the innovative application of cNLP for symptom extraction from clinical narratives in Spanish; and the capacity of EHRs to provide longitudinal insights into patients' course of illness. Taken together, this dissertation not only highlights the potential of EHRs but also navigates the intricacies of employing them for psychiatric genetic research.

The dissertation of Juan Fernando De la Hoz Gomez is approved.

Alex A.T. Bui

Bogdan Pasaniuc

Chiara Sabatti

Loes Marlein Olde Loohuis, Committee Co-Chair

Nelson B. Freimer, Committee Co-Chair

University of California, Los Angeles

2023

To my aunt and uncle, Clara and Miguel Gomez del Corral

TABLE OF CONTENTS

LIST OF FIGURES

damage and dysfunction and to physical disease (F06), SUD: Mental and behavioral disorders due to multiple drug use and use of other psychoactive substances (F19), BPE: Acute and transient psychotic disorders (F23), MDE: Major Depressive Episode (F32), PMD: Persistent mood disorders (F34), UMD: Unspecified mood disorder (F39), ANX: Other anxiety disorders (F41), PTSD: Reaction to severe stress, and adjustment disorders (F43), ADHD: Hyperkinetic disorders (F90), CON: Conduct disorders (F91)

Figure 3.4 *Diagnostic stability over time. At each visit k, the proportion of patients that will switch primary diagnosis code on their next visit k+1. A) Stratified by age groups: age at 1st visit before and after 30 years. B) Stratified by having previously switched diagnoses (from visit k-1). n=12,962 patients (Supplementary Figure 1).*

Figure 4.1 *Random Forest model for diagnosis of SMI based on 162 EHR-extracted features. A and B) Precision recall and ROC curves. The performance of the rule-based model (most recent SMI ICD-10 code) is shown by the dots. AUPRC: MDD: 0.86; BD: 0.891; SCZ: 0.95. AUROC: MDD: 0.913; BD: 0.90; SCZ: 0.991. C and D) probabilities for diagnosis of MDD, BD and SCZ. Dots represent 3-way probabilities for each patient. Individuals are colored in Figure 1C according to their SCID diagnosis and in Figure 1D to their combination of EHR-SCID diagnoses from Table 4. E and F) SHAP value plots of the 15 most important predictors for diagnoses of BD and SCZ, respectively.*

Figure 4.2 *Ternary plots interpreting diagnostic probabilities. A) shows the frequency of nine features across the three-way probability space when ICD-10 codes are not part of the prediction model. In order, the features are delusions, hallucinations, grandiosity, suicidal ideation, suicide attempt, ICD-10 codes of neurotic, stress-related and somatoform disorders (F4), use of antipsychotics, use of mood stabilizers, and use of antidepressants. B) positions of "worsening" individuals based on their EHR features in their first (red) and last (green) five visits. They had no hospitalization in the first five visits and two or more in the last five visits (N=107). Small ternary plot shows centroids for the first and last five visits.*

Figure 4.3 *Density plot of the probability distribution of the binary classifier for BD type I and MDD, stratified by SCID diagnosis: BD type II (N=350), other specified BD (N=168), and unspecified BD (N=31).*

Figure 5.1 *A) Density of concepts per document in the five sections of origin. B) Distribution of concept occurrences (examples) in the 2000 documents annotated.*

F.5 Supplementary Figure 5. *Diagnostic stability over time. Using time since the first encounter instead of visit number. For every year, the observed proportion of visits that will have a diagnostic switch on the next visit is plotted as a dot with 95% confidence intervals. The solid line is the average probability of switching at any given visit during that year, as estimated by the model. The corresponding shaded area is the 95% confidence interval. A) Patients are stratified by the age of their first visit: before and after 30 years. B) Visits are stratified by having switched diagnoses from visit k-1. N=12,962 patients.*

LIST OF TABLES

T.4 Supplementary Table 4*. List of diagnostic pairs from the F chapter of ICD-10 that are, by definition, incompatible with each other and, therefore, represent diagnostic switches. All other combinations of diagnoses are considered comorbidities. There are two exceptions to this rule: the pairs F30-F31 and F32-F33. These are neither switches nor comorbidities.*

T.5 Supplementary Table 5. *Demographic and clinical characteristics of the study cohort. Patients are classified by their most recent SMI diagnosis. Medians with interquartile range (IQR) are presented for: visits per patient, age at the most recent visit, length of stay, and length of the medical record. Tests comparing these values across the three main diagnoses (MDD, BD and SCZ) are provided in the bottom part of the table. Test*: across the three diagnoses, differences in percentages are tested with a chi-squared test and differences in distributions with a Kruskal-Wallis test. Test**: between pairs of diagnoses, differences in percentages are tested with z-tests and differences in distributions with Mann-Whitney tests. Asterisks mark significant results at the Bonferroni-corrected alpha threshold of 0.05/8=0.006.*

T.6 Supplementary Table 6. *Estimates of kappa and PPV from comparisons of ICD-10 diagnoses extracted from the EHR and clinician diagnoses obtained from chart review, considering all visits and considering inpatient visits only. The narrow definition refers to the SMI codes: F20X (SCZ), F301, F302, F310, F311, F312, F313, F314, F315, F316, F317 (BD), F322, F323, F331, F332, F333, F334 (Severe/Recurrent MDD). The broad definition encompasses, additionally, all F31X (including F318 and F319), F32X (including F320, F321, F328 and F329) and F33X (including F330, F338 and F339). Kappa values are estimated both for individual diagnoses and across all diagnoses. 95% confidence intervals for kappa values are shown in parentheses. Kappa values between 0.6-0.8 are considered "very good", while those > 0.8 are considered "excellent".*

T.7 Supplementary Table 7. *Annotation results for four clinical features. Two clinicians independently reviewed and annotated 3,600 sentences. The columns show the number of sentences containing the clinical features identified by Clinician A (Clin. A), Clinician B (Clin. B) or either clinician (Union), and their level of agreement as estimated by Cohen's kappa (Kappa). Kappa values between 0.4-0.6 are considered "good", those between 0.6-0.8 are considered "very good", and those > 0.8 are considered "excellent".*

T.8 Supplementary Table 8. *Performance of the NLP algorithm for extraction of clinical features. A) Sentence-level performance on the annotated gold standard. B) Patient-level performance on patient records* *manually reviewed by a clinician (n=104, as one patient was removed for having only one clinical note). C) Patient-level performance after post-hoc review of true and false positives. The average affirmative and negative instances of each feature per patient are, respectively: 1 and 0 for Suicide Attempt, 4 and 12 for Suicidal Ideation, 17 and 19 for Delusions, 10 and 25 for Hallucinations.*

T.9 Supplementary Table 9*. ICD-10 code severity and psychosis qualifiers, recorded at each visit for individuals receiving an ICD-10 mood disorder diagnosis demonstrate strong association with clinical feature profiles extracted from the notes during the same visit. The top section shows the association of clinical features with codes representing mood disorder diagnoses designated as severe compared to those designated as either mild or moderate. The binary variable "severe" is defined as 1 when the visit code is one of F301, F302, F311, F312, F314, F315, F322, F323, F332, or F333, and is 0 when the visit code is one of F300, F310, F313, F320, F321, F330, or F331. For this section, N is 47,186 visits in 9,203 people. The bottom section shows the association of psychotic features (Delusions and Hallucinations) with codes designating the presence of psychotic symptoms during visits in which a code designating a mood disorder episode as severe has been recorded. The binary variable "psychosis" is defined as 1 when the visit code is one of F302, F312, F315, F323, or F333 and is 0 when the visit code is one of F301, F311, F314, F322, or F332. For this section, N is 15,120 visits in 4,075 people. Analyses are described in Supplementary Note 6.*

T.10 Supplementary Table 10. *Frequency of each clinical feature (in percentages), for patients with SMI diagnoses, stratified by gender and inpatient history (yes: patients with a history of at least one inpatient hospitalization; no: individuals without any history of inpatient hospitalization). The first two columns show the total number of individuals included in this analysis, while the other columns show the frequencies of each clinical feature. The rows display the total numbers and frequencies considering all SMI diagnoses ("All") and then considering each of the three diagnoses separately. All patients included in this table had at least two clinical notes in their EHR.*

T.11 Supplementary Table 11. *Odds ratios for patient-level associations of each clinical feature with gender, diagnosis, and the other clinical features. Bonferroni-corrected alpha is 0.05/12=0.0041 for table A and 0.05/16=0.0031 for table B. Analyses are described in Supplementary Note 2.*

T.12 Supplementary Table 12. *Percentage of individuals with comorbidities within each SMI diagnosis, as observed in patients with at least 3 encounters (n=12,962). The ICD codes for the 20 most frequent diagnoses are shown.*

T.13 Supplementary Table 13. *Stability of diagnoses. A) Counts and comparative statistics for long-term stability of SMI diagnoses. First: number of individuals with that diagnosis on their first visit. Last: number of individuals with that diagnosis on their last visit. Both: number of individuals with that diagnosis on both visits. Prospective stability is calculated as 100*both/first, and retrospective as 100*both/last. B) comparison of stability values between two groups: those whose first visit was before age 30 and those whose first visit was after age 30. C) comparison between prospective and retrospective stability for all patients and stratified by age group*

Acknowledgments

As I reflect on the path that brought me here, I am filled with gratitude for all the people whose efforts and support have gotten me to this place.

First, I want to thank my advisors, Loes and Nelson. None of this would have been possible without their vision, advice, and support. Loes, thank you for trusting me from the beginning. Witnessing the lab grow under your leadership has been truly inspirational. Nelson, I am profoundly grateful for your unwavering support throughout all these years. I feel privileged to have had many opportunities to share ideas with you and learn how the best science is done.

Thanks to my doctoral committee, Alex Bui, Bogdan Pasaniuc, and Chiara Sabatti, for their encouragement and valuable feedback.

My success really is the fruit of the work of many people over many years. In particular, PIs who conceived ambitious projects and created a fertile ground for other scientists to grow. I am deeply grateful to Carlos López-Jaramillo, Carrie Bearden, Javier Escobar, Victor Reus, Nelson Freimer, and Loes Olde Loohuis.

To the Colombian team, Mauricio Castaño, Maria Cecilia López, Luis Guillermo Agudelo, Sergio Sánchez, Juanita Melo, and especially to Maria Pérez, Daniel Londoño, and Alejandro Arias, for always going above and beyond at every step of the way.

Thank you to the Gonda people, Margaret Chu, Susan Service, Terri Teshiba, Binh Vuong, and Adara Lui, for your generosity in lending me a hand whenever I needed it.

I want to thank the Fulbright Commission in Colombia and the Neurobehavioral T32 training grant for the financial and logistical support. Thank you to Silvia Restrepo, Jorge Duitama, and Bodo Raatz for preparing me for the PhD.

I feel lucky to have found so many great communities at UCLA.

It has been a huge pleasure to work side by side with the beautiful people of the Freimer, Loes and Ophoff Labs. Kerneau Seok, Aditya Pimplaskar, Greta Gerdes, Nora Liu, Ana Maria Ramirez, Laura Mena, Janet Song, Marfred Muñoz, Alex Chubick, Kevin Wojta, Lingyu Zhan, Lianne Reus, Carolinne Alvarado, Toni Boltz, Merel Bot, Anil Ori, and Marcelo Francia. A special thank you to Ana María Diaz for bringing so much joy to the office and always believing in me.

To the undergraduates I had the privilege of working with: Julia Bowers, Ben Simon, Sandy Assaf, Ryan Foundoulis, Ray Guo, and Dorothy Duan, thank you for your trust and passion.

Thank you to my bioinformatics cohort, Jesse, Tommer, Leah, Mike, Alec, Brandon, Ha, Megan, Kofi, Sarah, Yang, Christa, and Ruthie, and to the broader bioinformatics/MPG community, Chris Robles, Arun Durvasula, Kiku Koyano, Rob Brown, Maria Palafox, Michal Sadowski, Jonatan Hervoso, Lena Krockenberger, Sandy Kim, Terrence Li, Jon Mah, Kim Insigne, Mudra Choudhury, Adriana Arneson, Artur Jaroszewicz.

To the Life Sciences Core and CIRTL, particularly Diana Azurdia, Katie Dixie, Kristin McCully, and K. Supriya, thank you for fostering my joy of teaching and mentoring.

These years in graduate school have given me an amazing community of friends. I am infinitely grateful for your friendship and the happy days – the beach volleyball afternoons, the art Sundays, the morning surfing sessions, the wine tastings, the mountain biking, the salsa dancing, and more. Allison Daly, Mike Lauria, Camila Marrero, Amara Thind, Eric Heinrichs, Ana Sias, Matt Miller, Sebastian Solarte, Emilie Tarouilly, Laura Correa, Juan Camilo del Rio, Alejandra Abril, Carlos Osorio, Miya Shaffer, Peter Polack, Zep Kalb, Sara Vallejo, Juan Camilo Mora.

Thank you, Sebastián Ramirez, for walking this path with me.

Last and most important, I want to thank my parents, Luz Alba and Fernando, my brother Alejandro, and my fiancée Stephanie, for their unconditional support. I would not be here without you.

Specific contributions to the chapters are as follows:

Chapter 2 describes the foundational work with the database of the Clínica San Juan de Dios in Manizales. Some of the people without which this work wouldn´t have been possible are Cristian Gallego, who did the first data export and deidentification of the database. Alejandro Arias with whom I spent countless hours navigating data tables. He oversaw all the EHR data freezes. Janet Song did some of the data cleaning, documentation, and integration of external data sources. Marfred Muñoz, Laura Mena and Sandy Assaf coded much of the demographic data. Greta Gerdes and Jiahao Tian wrote documentation for the database. Mauricio Castaño and Ana Maria Diaz contributed invaluable clinical expertise to all analyses. Clara Frydman and Selina Wu incorporated external data.

Chapter 3 is a manuscript currently being submitted under the name "..." by Juan F. De la Hoz, Alejandro Arias, Susan K. Service, Mauricio Castaño, Ana M. Diaz-Zuluaga, Janet Song, Cristian Gallego, Sergio Ruiz-Sánchez, Javier I Escobar, Alex A. T. Bui, Carrie E. Bearden, Victor Reus, Carlos Lopez-Jaramillo, Nelson B. Freimer, Loes M. Olde Loohuis. NF, LOL, and JDLH designed the study and wrote the manuscript. JDLH and SKS performed the analyses. MC did the chart annotations. All authors read and reviewed the manuscript and provided insightful feedback.

Chapter 4 is currently being prepared as a manuscript for submission with Nora Liu and JDLH as co-first authors. LOL, NL and JDLH designed the study. NL and JDLH performed the analyses and drafted the manuscript.

Chapter 5 contains materials from an article currently in preparation. LOL and JDLH designed the study with input from Alex Bui, NF, Victor Reus, and Carlos Lopez-Jaramillo. Alejandro Arias arranged and managed all the information systems. María Pérez Vallejo and Daniel Londoño did the annotation.

VITA

2009-2014	B.A. Microbiology Universidad de los Andes. Bogotá, Colombia
2014-2017	Research Assistant International Center for Tropical Agriculture. Cali, Colombia
2015-2016	Specialization in Applied Statistics Universidad del Valle. Cali, Colombia
2017-2023	Graduate Student Researcher, Bioinformatics Interdepartmental Program University of California, Los Angeles
2022	Teaching Assistant, LS 30A: Mathematics for Life Sciences University of California, Los Angeles
2023	CIRTL Practitioner

Awards

2016 - 2021	Fulbright - Minciencias
2017 - 2021	Extramural funding award
2018 - 2020	Predoctoral fellowship, associated with: 5T32MH073526
2021	Outstanding Mentorship and Outstanding Teaching Assistant

Presentations

09/2022	Invited Speaker Diversity and Science Lecture Series (DASL) Symposium.
09/2022	(Talk) Using longitudinal EHR data to delineate diverse trajectories of SMI World Congress of Psychiatric Genetics. Florence, Italy
09/2019	(Poster) NLP Strategies for phenotyping Severe Mental Illness from EHR World Congress of Psychiatric Genetics. Los Angeles, CA

Publications

De la Hoz, J. F.*, Y. Liu*, N. B. Freimer, L. M. Olde Loohuis, et al. (2023). "*Transcending Diagnostic Categories: Unveiling Phenotypic Overlaps in SMI through EHR Data and Machine Learning*". In: Prep.
* Shared co-first authors.

De la Hoz, J. F., A. Arias, M. Pérez-Vallejo, J. D. Londoño, V. I. Reus, C. Lopez-Jaramillo, N. B. Freimer, L. M. Olde Loohuis, et al. (2023). "*Extraction of Symptom-Level Phenotypes from Clinical Notes*". In: Prep.

De la Hoz, J. F., A. Arias, S. Service, M. Castaño Ramírez, A. M. Díaz-Zuluaga, J. Song,C. Gallego, S. Ruiz Sánchez, J. I. Escobar, A. A. Bui, C. E. Bearden, V. I. Reus, C. Lopez-Jaramillo, N. B. Freimer, and L. M. Olde Loohuis (2022). "*Electronic health records reveal transdiagnostic clinical features and diverse trajectories of serious mental illness*". In: medRxiv. url: doi.org/10.1101/2022.08.20.22279007v2

Service, S.K., **De la Hoz, J. F.,** Diaz-Zuluaga, A.M., Arias, A., Pimplaskar, A., Luu, C., Mena, L., Valencia, J., Ramírez, M.C., Bearden, C.E. and Sabbati, C., (2023). "*Predicting diagnostic conversion from major depressive disorder to bipolar disorder: an EHR based study from Colombia*". In: medRxiv. url: doi.org/10.1101/2023.09.28.23296092v1

Song, J., M. Castaño Ramírez, J. Okano, S. Service**, J. F. De la Hoz,** A. M. Diaz-Zuluaga, C. Vargas Upegui, C. Gallego, A. Arias, A. Valderrama Sánchez, T. Teshiba, C. Sabatti, R. Gur, C. E. Bearden, J. I. Escobar, V. I. Reus, C. Lopez-Jaramillo, N. B. Freimer, L. M. Olde Loohuis, and S. Blower (2022). "*Geospatial analysis reveals distinct hotspots of severe mental illness*". In: medRxiv. url: doi.org/10.1101/2022.03.23.22272776

Service, S., C. V. Upegui, M. Castaño Ramírez, A. M. Port, T. M. Moore, M. Munoz Umanes, L. G. Agudelo Arango, A. M. Díaz-Zuluaga, J. Melo Espejo, M. C. López, J. D. Palacio, S. Ruiz Sánchez, J. Valencia, T. M. Teshiba, A. Espinoza, L. M. Olde Loohuis**, J. F. De la Hoz,** B. B. Brodey, C. Sabatti, J. I. Escobar, V. I. Reus, C. Lopez-Jaramillo, R. C. Gur, C. E. Bearden, and N. B. Freimer (2020). "*Distinct and shared contributions of diagnosis and symptom domains to cognitive performance in severe mental illness in the Paisa population: a case-control study*". In: The Lancet Psychiatry 7.5, pp. 411–419.

Perea, C., **J. F. De la Hoz,** D. F. Cruz, J. D. Lobaton, P. Izquierdo, J. C. Quintero, B. Raatz, and J. Duitama (2016). "*Bioinformatic analysis of genotype by sequencing (GBS) data with NGSEP*". In: BMC genomics 17.5, pp. 539–551.

Brito, J. J., T. Mosqueiro, J. Rotman, V. Xue, D. J. Chapski, **J. F. De la Hoz**, P. Matias, L. S. Martin, A. Zelikovsky, M. Pellegrini, and S. Mangul (2020). "*Telescope: an interactive tool for managing large-scale analysis from mobile devices*". In: GigaScience 9.1, giz163.

Fuentes, R. R., D. Chebotarov, J. Duitama, S. Smith, **J. F. De la Hoz**, M. Mohiyuddin, R. A. Wing, K. L. McNally, T. Tatarinova, A. Grigoriev, R. Mauleon, and N. Alexandrov (2019). "*Structural variants in 3000 rice genomes*". In: Genome research 29.5, pp. 870–880.

Keller, B., D. Ariza-Suarez, J. F**. De la Hoz, J.** S. Aparicio, A. E. Portilla-Benavides, H. F. Buendia, V. M. Mayor, B. Studer, and B. Raatz (2020). "*Genomic prediction of agronomic traits in common bean (Phaseolus vulgaris L.) under environmental stress*". In: Frontiers in plant science 11, p. 1001.

Worthington, M., M. Ebina, N. Yamanaka, C. Heffelfinger, C. Quintero, Y. P. Zapata, J. G. Perez, M. Selvaraj, M. Ishitani, J. Duitama, **J. F. De la Hoz**, I. Rao, S. Dellaporta, J. Tohme, and J. Arango (2019). "*Translocation of a parthenogenesis gene candidate to an alternate carrier chromosome in apomictic Brachiaria humidicola*". In: BMC genomics 20, pp. 1–18.

Diaz, S., D. Ariza-Suarez, P. Izquierdo, J. D. Lobaton, **J. F. De la Hoz**, F. Acevedo, J. Duitama, A. F. Guerrero, C. Cajiao, V. Mayor, S. E. Beebe, and B. Raatz (2020). "*Genetic mapping for agronomic traits in a MAGIC population of common bean (Phaseolus vulgaris L.) under drought conditions*". In: BMC genomics 21.1, pp. 1–20.

Lobaton, J. D., T. Miller, J. Gil, D. Ariza, **J. F. De la Hoz**, A. Soler, S. E. Beebe, J. Duitama, P. Gepts, and B. Raatz (2018). "*Resequencing of common bean identifies regions of inter–gene pool introgression and provides comprehensive resources for molecular breeding*". In: The plant genome 11.2, p. 170068.

De la Hoz-Restrepo, F., N. J. Alvis-Zakzuk, **J. F. De la Hoz**, A. De la Hoz, L. G. Del Corral, and N. Alvis-Guzmán (2020). "*Is Colombia an example of successful containment of the 2020 COVID-19 pandemic? A critical analysis of the epidemiological data, March to July 2020*". In: International Journal of Infectious Diseases 99, pp. 522–529.

CHAPTER 1

Leveraging Electronic Health Records in the Advancement of Precision Psychiatry

1.1 Introduction

Severe Mental Illness (SMI) encompasses a range of mental, behavioral, and emotional disorders that significantly impair daily functioning [1]. Despite society's growing awareness of these disorders [2–4], and after decades of extensive biomedical research [5,6], the biological mechanisms underlying SMI remain elusive. However, recent advances in genetic technologies and bioinformatics and increased access to extensive research cohorts [7–9] have paved the way to increasingly frequent breakthroughs that are gradually improving our understanding of SMI genetics.

Genome-wide association studies (GWAS) have successfully identified common genetic variants associated with psychiatric disorders. For example, they have highlighted the role of neurons, especially synapses, in the etiology of schizophrenia [10]. GWAS for Bipolar disorder has reaffirmed the significance of calcium signaling pathways in its biology [11]. Similarly, GWAS for Major Depression has revealed an enrichment of associated loci in brain regions like the frontal cortex and specific molecular pathways tied to neurogenesis [12]. Moreover, robust evidence has accumulated in recent years demonstrating the shared genetic components across psychiatric disorders [13]. Cross-disorder GWAS, in particular, has uncovered over a hundred pleiotropic loci that deepen our understanding of the common biology connecting distinct disorders [14]. These studies are creating new opportunities to disentangle the web of shared risk factors and overlapping clinical features across psychiatric disorders.

Two critical insights have emerged from genetic studies of complex traits. The first is the need for expansive sample sizes, ranging from thousands to hundreds of thousands of participants, to have enough statistical power to detect small genetic associations [15]. The realization that “larger is better” has fueled the move away from traditional small cohorts, recruited from specialized research centers, towards increasingly large cohorts recruited through non-conventional methods that prioritize reducing costs per participant. Often, the rapid explosion of sample size in genetic studies has come at the expense of high-quality phenotyping, prompting calls for better protocols that preserve the quality of phenotypic data [16]. The second insight is the pronounced bias in the ancestral composition of GWAS cohorts. Most human genetic research in recent decades has focused on European ancestries. As Martin et al. [17], such a narrow focus risks making research outcomes less universally applicable, thereby limiting therapeutic advancements for non-European populations. To maximize the impact of future genetic studies of SMI, research cohorts need to be both large and representative of global genetic diversity while, in parallel, prioritizing high-throughput, high-quality phenotyping protocols.

Over the past decade, the declining costs of genotyping have driven the proliferation of GWAS. However, phenotyping, particularly for SMI, remains resource-intensive and methodologically challenging. Unlike somatic conditions, psychiatric disorders lack biomarkers to delineate disease states objectively. This situation results in a diagnostic landscape with nebulous boundaries between disorders, with symptoms that overlap multiple diagnoses and often lead to conflicting clinical interpretations [18–21]. Without a universally endorsed gold standard for diagnosis, rigorous, research-quality phenotypes largely depend on costly clinical interviews [22,23], which don’t scale easily. In the place of interviews, prominent biobank projects such as the UK Biobank and companies like 23andMe employ self-report questionnaires [7,9]. These enable the capture of a broad spectrum of phenotypes at a significantly lower cost per individual, creating some of the largest datasets available today. While questionnaires offer a pragmatic approach to examining the complex dimensions of serious mental illness (SMI), including medical history, family predispositions, and other health determinants, they are often critiqued for yielding phenotypes of lower quality, especially in terms of precision [24].

An alternative strategy for high-throughput phenotyping that overcomes the limitations of self-report questionnaires is using Electronic Health Records (EHRs). These provide comprehensive patient data, encompassing symptomatic narratives, diagnoses, familial medical histories, prescription and treatment regimens, and other types of healthcare usage data. Originating primarily to facilitate clinical care and streamline administrative processes [25], EHRs meticulously document a patient's trajectory within the healthcare system, thereby facilitating rich, longitudinal studies of chronic conditions [26]. In fields ranging from cystic fibrosis to depression, researchers have demonstrated the potential of these databases to contribute research-quality phenotypes to genetic research [27,28]. Consequently, EHRs expand the number and diversity of traits available for genetic research, potentially enhancing the quality of phenotypes in biobanks linked to healthcare settings.

Multiple hospitals and academic institutions have recently begun building extensive biobanks from their patient populations [29,30]. They recognize the value of EHRs in advancing precision medicine and are harnessing them to extract clinical phenotypes for guiding genomic research [31]. Leading healthcare institutions, like UCLA Health, Vanderbilt University Medical Center, and Mount Sinai Health System, have established dedicated programs such as Atlas, BioVU, and BioMe, which link DNA samples with de-identified EHRs. Similarly, the Million

Veteran Program (MVP), one of the largest biobanks globally, combines genetic, clinical, lifestyle, and military exposure data from over a million participants [8], facilitating studies into the genetics of PTSD and substance use disorders. Along this same line, the PsychEMERGE (Electronic MEdical Records and GEnomics) Network combines EHR-linked biobanks across multiple institutions to increase sample sizes for genetic studies of neuropsychiatric illness to develop preventive and therapeutic interventions [32].

The work of leveraging EHRs for large-scale research requires the concerted effort of multiple parties, such as engineers, doctors, researchers, and even ethicists, to ensure data quality and contiguity, harmonize heterogeneous data sources [33], and navigate the ethical landscape associated with using Protected Health Information [34]. EHR research often relies on structured data—diagnostic codes, medications, lab values—which constitute the primary elements of most phenotype definitions. However, multiple factors such as hospital protocols or insurance dynamics can have a major effect on how and when they are recorded; taking these data at face value might, therefore, introduce biases and other sources of error that limit our ability to draw conclusions from them. Consequently, validating extracted phenotypes is a critical step in utilizing EHRs for research.

In lower and middle-income countries (LMICs), the increasing availability of Electronic Health Records (EHRs) presents a promising opportunity for creating large-scale biobanks tailored for GWAS. Such projects can contribute to democratizing access to genetic research capacities around the globe while at the same time addressing the current imbalance in ancestral diversity in genetic studies. However, achieving this requires international and cross-cultural collaborative efforts and equitable resource-sharing.

This dissertation explores the potential of EHRs from regional psychiatric hospitals in Colombia in facilitating large-scale studies into severe mental illness. This work develops three strengths of EHRs that distinguish them from other phenotyping methods in the research of psychiatric genetics:

1. *Longitudinal healthcare information.* EHRs chronicle patients' healthcare interactions, offering valuable insights into disease progression and treatment outcomes. This research leverages the longitudinal perspective to reveal patterns and trends often overlooked in cross-sectional data.
2. *High-dimensional data integration.* EHRs are repositories of multifaceted data, providing a more holistic and nuanced representation of patients than traditional categorical diagnoses. This research leverages these rich, high-dimensional data to scrutinize and augment conventional diagnostic classifications, aiming to characterize the overlaps and distinctions between diagnoses in a real-world clinical context.
3. *Symptom and behavior phenotyping.* EHRs document detailed descriptions of a patient's state, including the symptoms and behaviors that shape their clinical presentation. This information is crucial for prognosis and treatment decisions, offering a potentially truer reflection of the biological underpinnings of SMI than diagnosis alone. Despite their value, these detailed phenotypes and social, environmental, and developmental factors often reside in free-text clinical narratives, making them hard to access on a large scale. Nevertheless, advances in Clinical Natural Language Processing (cNLP) enable the extraction of these phenotypes from clinical notes [35–37] This research presents the development of a rule-based cNLP algorithm designed for symptom-level phenotype extraction from Spanish-language EHRs.

In summary, this dissertation examines how EHRs enhance genetic studies of SMI and expand biobank access globally.

1.2 Overview of Chapters

Misión Origen (MO) is a multi-institutional research project focused on advancing the study of the genetics of SMI and enhancing the diversity of human genetic research cohorts. To achieve this, MO is recruiting 100,000 participants from the culturally and genetically distinct Paisa region in Colombia [38]. The project includes establishing a biobank for storing and analyzing the genetic data of these participants. The work presented here aims to facilitate the collection of such a large sample size by extracting high-quality phenotypes from the EHRs of participating institutions. The hospitals featured in this study are the Clínica San Juan de Dios, in Manizales (CSJDM) and Hospital Mental the Antioquia (HOMO), in Bello.

Chapter 2 details the EHR database of the CSJDM. Here, I outline the data types and sources incorporated into the research database and the steps taken for data cleaning, deidentification, and integration of external data like ATC codes and geographical coordinates.

Chapter 3 presents the research paper titled "Investigations of Electronic Health Records Databases Enable Scalable Analyses of Transdiagnostic Clinical Features and Reveal Highly Diverse Trajectories of Serious Mental Illness." In this paper, we validate phenotypic data from the EHR of the CSJDM using manual chart reviews and explore the longitudinal trends of psychiatric diagnoses, emphasizing the factors contributing to diagnostic instability.

Chapter 4 showcases our diagnostic definitions, both rule-based and machine learning-driven, from the EHR of patients at CSJDM. The chapter underscores the potential of machine learning in deriving diagnostic probabilities and their potential use in genetic research in psychiatry.

Chapter 5 moves the focus to HOMO. Here, I document the development and validation of a clinical NLP algorithm for extracting phenotypes from clinical narratives. A key highlight will be ensuring the completeness and robustness of this extraction.

Through these chapters, I offer a holistic view of the opportunities and limitations of using EHRs to conduct research on psychiatric genetics in the Colombian context.

1.3 Conclusion

This dissertation explores the intersection of psychiatry, and medical informatics, aiming to advance the field of psychiatric genetics using EHRs from Colombian regional hospitals. The core thread of this work is the development of methods to extract clinical data from complex database architectures, enabling thorough validation of psychiatric phenotypes. This sets the groundwork for an EHR phenotyping framework for future research into this population.

1.4 References

1. NIMH.Mental Illness. https://www.nimh.nih.gov/health/statistics/mental-illness.

2. Pescosolido, B. A., Halpern-Manners, A., Luo, L. & Perry, B. Trends in Public Stigma of Mental Illness in the US, 1996-2018. JAMA Netw Open 4, e2140202 (2021).

3. Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet Psychiatry 9, 137–150 (2022).

4. SAMHSA. As Part of President Biden's Mental Health Strategy, HHS Awards Nearly $105 Million to States and Territories to Strengthen Crisis Call Center Services in Advance of July Transition to 988. https://www.samhsa.gov/newsroom/press-announcements/20220419/hhs-awards-105-million-states-territories-strengthen-crisis-call-center-services.

5. Carvalho, A. F. et al. Evidence-based umbrella review of 162 peripheral biomarkers for major mental disorders. Transl Psychiatry 10, (2020).

6. Liston, C. et al. Understanding the biological basis of psychiatric disease: What's next? Cell 185, 1–3 (2022).

7. Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. Nature 562, 203–209 (2018).

8. Gaziano, J. M. et al. Million Veteran Program: A mega-biobank to study genetic influences on health and disease. J Clin Epidemiol 70, 214–223 (2016).

9. Tung, J. Y. et al. Efficient replication of over 180 genetic associations with self-reported medical data. PLoS One 6, (2011).

10. Trubetskoy, V. et al. Mapping genomic loci implicates genes and synaptic biology in schizophrenia. Nature 604, 502–508 (2022).

11. Mullins, N. et al. Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. Nat Genet 53, 817–829 (2021).

12. Howard, D. M. et al. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. Nat Neurosci 22, 343–352 (2019).

13. Anttila, V. et al. Analysis of shared heritability in common disorders of the brain. Science (1979) 360, (2018).

14. Lee, P. H. et al. Genomic Relationships, Novel Loci, and Pleiotropic Mechanisms across Eight Psychiatric Disorders. Cell 179, 1469-1482.e11 (2019).

15. Abdellaoui, A., Yengo, L., Verweij, K. J. H. & Visscher, P. M. 15 years of GWAS discovery: Realizing the promise. Am J Hum Genet 110, 179–194 (2023).

16. Cai, N. et al. Minimal phenotyping yields genome-wide association signals of low specificity for major depression. Nat Genet 52, 437–447 (2020).

17. Martin, A. R. et al. Clinical use of current polygenic risk scores may exacerbate health disparities. Nat Genet 51, 584–591 (2019).

18. Cuthbert, B. N. & Insel, T. R. Toward the future of psychiatric diagnosis: The seven pillars of RDoC. BMC Med 11, (2013).

19. Regier, D. A. et al. DSM-5 field trials in the United States and Canada, part II: Test-retest reliability of selected categorical diagnoses. American Journal of Psychiatry 170, 59–70 (2013).

20. Bearden, C. E., Reus, V. I. & Freimer, N. B. Why genetic investigation of psychiatric disorders is so difficult. Curr Opin Genet Dev 14, 280–286 (2004).

21. Kendell, R. & Jablensky, A. Distinguishing Between the Validity and Utility of Psychiatric Diagnoses. American Journal of Psychiatry 160, 4–12 (2003).

22. Kendler, K. S. & Neale, M. C. Endophenotype: A conceptual analysis. Mol Psychiatry 15, 789–797 (2010).

23. Hyman, S. E. The diagnosis of mental disorders: The problem of reification. Annu Rev Clin Psychol 6, 155–179 (2010).

24. Cai, N. et al. Minimal phenotyping yields GWAS hits of low specificity for major depression. bioRxiv 440735 (2019).

25. Steve Alder. What is the HITECH Act? The HIPAA Journal https://www.hipaajournal.com/what-is-the-hitech-act/.

26. Crawford, D. C. et al. EMERGEing progress in genomics-the first seven years. Front Genet 5, 1–11 (2014).

27. Denny, J. C. et al. PheWAS: Demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. Bioinformatics 26, 1205–1210 (2010).

28. Bastarache, L. Using Phecodes for Research with the Electronic Health Record: From PheWAS to PheRS. Annu Rev Biomed Data Sci 4, 1–19 (2021).

29. Roden, D. et al. Development of a Large-Scale De-Identified DNA Biobank to Enable Personalized Medicine. Clin Pharmacol Ther 84, 362–369 (2008).

30. Bielinski, S. J. et al. Preemptive Genotyping for Personalized Medicine: Design of the Right Drug, Right Dose, Right Time—Using Genomic Data to Individualize Treatment Protocol. Mayo Clin Proc 89, 25–33 (2014).

31. Wei, W. Q. & Denny, J. C. Extracting research-quality phenotypes from electronic health records to support precision medicine. Genome Med 7, 1–14 (2015).

32. PsychEMERGE. http://psychemerge.com.

33. Weiskopf, N. G. & Weng, C. Methods and dimensions of electronic health record data quality assessment: Enabling reuse for clinical research. Journal of the American Medical Informatics Association 20, 144–151 (2013).

34. Office for Civil Rights. Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule. https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html (2012).

35. Jackson, R. G. et al. Natural language processing to extract symptoms of severe mental illness from clinical text: The Clinical Record Interactive Search Comprehensive Data Extraction (CRIS-CODE) project. BMJ Open 7, 1–10 (2017).

36. Liu, Q. et al. Symptom-based patient stratification in mental illness using clinical notes. J Biomed Inform 98, 103274 (2019).

37. Wu, C. S., Kuo, C. J., Su, C. H., Wang, S. H. & Dai, H. J. Using text mining to extract depressive symptoms and to validate the diagnosis of major depressive disorder from electronic health records. J Affect Disord 260, 617–623 (2020).

38. Service, S. K. et al. Distinct and shared contributions of diagnosis and symptom domains to cognitive performance in severe mental illness in the Paisa population: a case-control study. Lancet Psychiatry 7, 411–430 (2020).

CHAPTER 2

Overview of the Structure and Development of a Deidentified Research Database from Electronic Health Records in a Psychiatric Hospital in Colombia.

2.1 Introduction

Hospital databases serve multiple primary functions. These include supporting clinical decision-making and record-keeping, enforcing adherence to clinical guidelines and legal rules, and billing and remuneration. Additionally, they support secondary applications such as biomedical research, quality improvement, and public health surveillance by allowing access to aggregated and often de-identified data from patient's Electronic Health Records (EHRs). Each of these applications has its unique database requirements[1], and these requirements can change over time. In this way, databases evolve as a response to changing protocols, the introduction of new treatments, or even events like pandemics. These transformations, then, introduce inconsistencies in data collection, potentially compromising the integrity and quality of longitudinal datasets and limiting researchers' access to consistent, long-term data. Repurposing EHR databases for research, thus, presents significant challenges.[2,3] This chapter has two aims: first, to describe the repurposing of the EHR database from Clínica San Juan de Dios in Manizales (CSJDM), and second, to provide an overview of the database's content.

2.2 Hospital Setting

We conducted this work at the Clínica San Juan de Dios, located in Manizales, the capital municipality of the Caldas department in Colombia. The CSJDM provides comprehensive mental healthcare for the one million inhabitants of Caldas [4]. Of this population, Manizales accounts for approximately 44% of the total. The clinic does not discriminate based on health insurance or socioeconomic status. CSJDM is one of the primary recruitment sites for the genetics projects linked to the Misión Origen project that aims to establish a Latin American Biobank for Severe Mental Illness focused on the Paisa population.[5]

2.2.1 Two EHR Databases at the CSJDM

CSJDM in Colombia was ahead of its time by adopting an EHR system in 2005 – before most US hospitals. Their database has evolved over the years. From 2005 to 2015, a highly flexible system called Mentor, explicitly designed for the hospital, was used and maintained internally. Between 2016 and 2023, the hospital transitioned to another system called Compuconta, which was provided by a company specializing in EHR solutions for various hospitals. However, clinic personnel retained the ability to customize many of the data formats in this system to suit their requirements. In mid-2023, the hospital began using an EHR system from a larger company named Dinámica. This section will focus solely on data from the first two databases, Mentor and Compuconta, spanning the years 2005-2023. Figure 2.1 demonstrates the continuity of visit data during the transition between these databases (2015-2016).

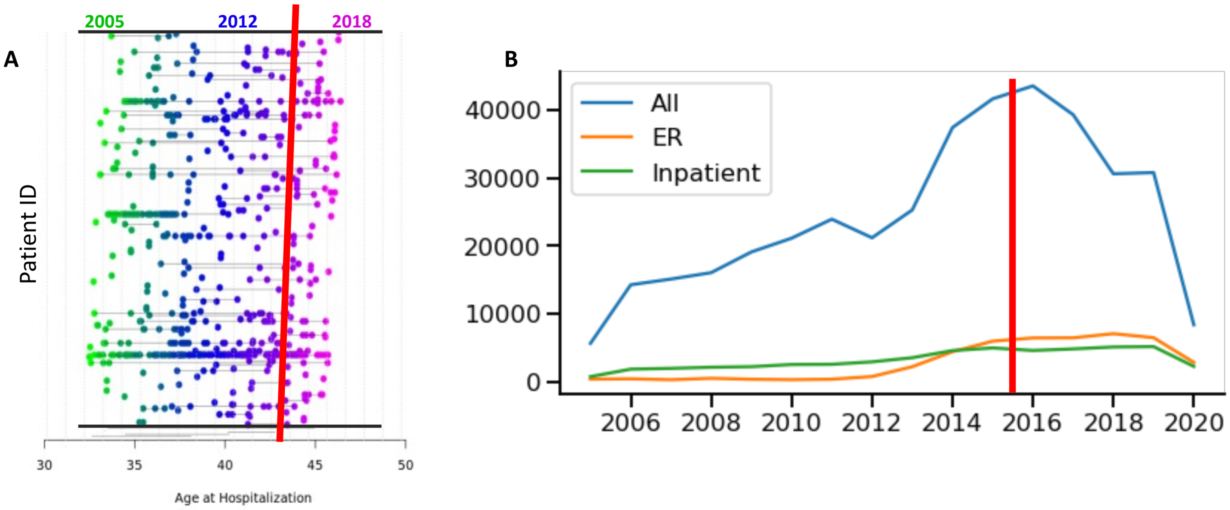


Figure 2.1 *A) sequence of hospitalizations of individual patients (aged 32-34 in 2005) across the two EHR databases at CSJDM. B) Frequency of overall visits and the number of inpatient or Emergency Room (ER) visits by year. The red lines indicate the time of the change from the first to the second EHR database. The smooth trend between both sides of the red line shows that the change in the database system didn't lead to significant disruptions in the recording of data.*

2.3 Privacy Considerations and Data Indexing

The IRBs of UCLA, Universidad de Antioquia (UdeA), and CJSDM approved the analysis and deidentification of the EHR database used for these studies. In particular, we removed all the protected health information (PHI) elements from medical records, as designated by HIPAA [7] (short for Health Insurance Portability and Accountability Act), before transitioning them to a HIPAA-compliant cloud environment.

While many PHI elements, like telephone numbers or emails, can be swiftly identified and omitted, names and, more crucially, patient IDs are essential for linking various events to a single individual – especially when linking across the two EHR databases at CSJDM. While each database used its own indexing, developing a unified approach was essential for assimilating data from both systems. For instance, one of the databases linked medications with prescription orders and prescription orders to patient encounters, which were then indexed chronologically. These, however, were different schemes for inpatient and outpatient prescriptions and had variations over time corresponding with changes in medication sourcing. This granularity, while beneficial for administrative tasks, complicates exploratory analyses and phenotype extraction.

Our solution was a patient-centered indexing method that prioritized chronological information, essentially transforming the raw tables to focus on patient ID and age. This allowed us to map and compare all records for a patient using just these two data points. The patient ID (EHR_ID) is a hashed version of each patient's national ID, enhanced with a unique "salt" held by an authorized hospital representative. The patient's age is stored as days since birth at the time of each record. Additionally, we include the calendar year to facilitate a comprehensive overview of trends over time, such as visit frequency, hospitalizations, and patterns of medication usage.

Lastly, we adopted a string-matching strategy to de-identify free text in medical records, targeting names and ID numbers. Specifically, we organized individual names alphabetically from those recorded in the central EHR registry. We used this list to do a regular expression search, replacing any matches with the placeholder *nombre (name). We applied a similar approach to numbers exceeding five digits, replacing them with the placeholder *cedula (national ID).

2.4 Data Tables and Types

2.4.1 Structured Data

2.4.1.1 Main Data The database's core table consists of a single entry for every patient, encompassing personal-level details. Each patient is identified uniquely by an ID (EHR_ID) derived from a hash of the cédula combined with a confidential "salt." This table captures demographic attributes such as gender, age at the first hospital visit, and the recruitment date (for patients participating in the Paisa or Misión Origen projects). To reflect project enrollment criteria, we incorporated information on the number of surnames an individual has that are associated with the Paisa population; this variable ranged from 0 to 2. Additional fields summarize the patient’s socio-demographic profile, including marital status, educational background, insurance type (private or public), religion, current occupation, and living arrangement (e.g., residing with parents, living independently, or homeless).

2.4.1.2 Visits to the Hospital A separate ‘visits’ table logs every clinical encounter at CSJDM for all patients. Key columns in this table include the unique patient ID, the period of the encounter (expressed in age days, as outlined previously), and the distinct calendar year of the visit. This table is a foundation for indexing associated events during a patient’s visit, be it medication administration or risk assessment scales. Additional clinical details in this table include the primary diagnosis for the visit, represented with ICD-10 codes, and the type of visit, denoted by two specific indicators – emergency visit and inpatient admission. Complementary fields provide a broader view of each visit, entailing details such as the treating physician’s ID, the type of insurance coverage, optional secondary diagnoses, the hospital unit that was the site of the encounter, and the reasons leading to the discharge of the patient.

2.4.1.3 Medications Medication data are crucial for clinical research, particularly in studying treatment regimens and outcomes. This approach can provide biological insights beyond mere diagnostic labels, such as identifying subgroups of patients who are responders or non-responders to treatments. It is essential to note that this data type has two settings: one encompasses inpatient stays, and another pertains to take-home prescriptions. Parsing these data requires considerable effort, since the values are in free text format or inputted via a diverse combination of drop-down menus. Consequently, we adopted an approach for text standardization that used regular expressions for extracting pertinent details.

Core fields in these tables are the patient’s ID, the date of prescription (in age-days), and the calendar year of the prescription. Other relevant fields include the medication name, the associated Anatomical Therapeutic Chemical (ATC) code (only available for some medications), the dose, the prescribed frequency, and the total number of doses. The method of administration is also denoted, with categories spanning tablets to injections. Supplementary fields include the diagnosis associated with the prescription, the ID for the prescribing physician, the insurance coverage, confirmation of medication collection from the hospital’s pharmacy, and the class of the medication —specifically, antidepressants, antipsychotics, mood stabilizers, and benzodiazepines.

2.4.1.4 Laboratory Tests and Other Procedures The ‘labs’ table records orders for laboratory tests and other procedures. Key fields in this table include the patient’s ID, the order date (expressed in age days), and the calendar year. Each test is identified by its name and a unique Test ID. Since the lab that does the testing is a third party, the test results are not recorded in the hospital’s database and are not part of the research database.

2.4.1.5 Routine Risk Scales developed in-house are typically completed by nurses and recorded across multiple tables. These include specific scales for risk of escaping from the hospital, risk of falling, risk of pressure ulcers, and risk of attempting suicide. Recently, the Global Assessment of Functioning (GAF) was also added to the system. The tables include item-level responses as well as the aggregate score for each scale. Common fields for each table are the patient’s ID, evaluation date (expressed in age-days), and calendar year.

2.4.2 Unstructured Data: Clinical Notes

Clinical notes, ranging from intake to discharge, provide detailed narratives that supplement structured data, offering a more rounded view of patient care and outcomes as observed by healthcare professionals, including physicians and nurses.

2.4.2.1 Intake Notes detail a patient's health status upon entry into the healthcare system and at the beginning of each subsequent hospitalization. These notes indicate the primary reasons for a patient's visit and often provide a broader health overview. Each note is tagged with a unique patient ID, date (in age-days), and calendar year. Key sections include Chief Complaint, Physical Exam, Mental Status Exam (MSE), Subjective and Objective sections, Analysis, and Treatment Plan.

2.4.2.2 Discharge Notes Discharge notes summarize a patient's hospital stay, detailing treatment and providing alerts for symptom surveillance. Each note is uniquely identified by a patient ID, discharge date (in age-days), and calendar year. It features the discharge diagnosis (ICD code), a synthesized analysis based on in-hospital observations, the reason for discharge, and a post-hospital care and monitoring plan. Patients receive a copy of this note upon leaving the hospital.

2.4.2.3 Follow-up and Progress Notes Outpatient and progress notes, while contextually distinct, share a standard format. Each note has a unique patient ID, the date (in age-days), and the calendar year. They encompass the primary diagnosis (ICD codes), patient-reported symptoms, objective clinical observations, an interpretative analysis, and an evolving treatment strategy.

2.4.2.4 Nurse Notes Nurses' notes, recorded twice daily, concisely record a patient's status, symptoms, and treatment responses.

2.5 Incorporating External Data

External data contained within the EHR database include ATC codes via the RxNorm API. We have used this utility to match medication names in Spanish from the CSJDM database with their corresponding ATC codes. Additionally, we have utilized for research the home addresses of CSJDM patients, which are included in the EHRs; combining this information with data from geocoding services, such as OpenStreetMap, makes it possible to evaluate geospatial aspects of serious mental illnesses.[4]

2.6 Descriptive Statistics of the Hospital Population

As of data freeze v11 (from August 9, 2023), the EHR of the CSJDM had records for 92,505 patients (52% female, N=47,646), with a total of 465,593 visits (15% of them resulting in an inpatient stay). Overall, 38% of all patients (N=35,551) have had an inpatient stay, and 40% (N=36,641) have been to the hospital only once. 32,935 patients have been to the hospital before their 18th birthday and 14,029 before their 12th. Figure 2.2 shows the distribution of common ICD-10 diagnoses, and the frequency of visits to the hospital, stratified by age and sex.

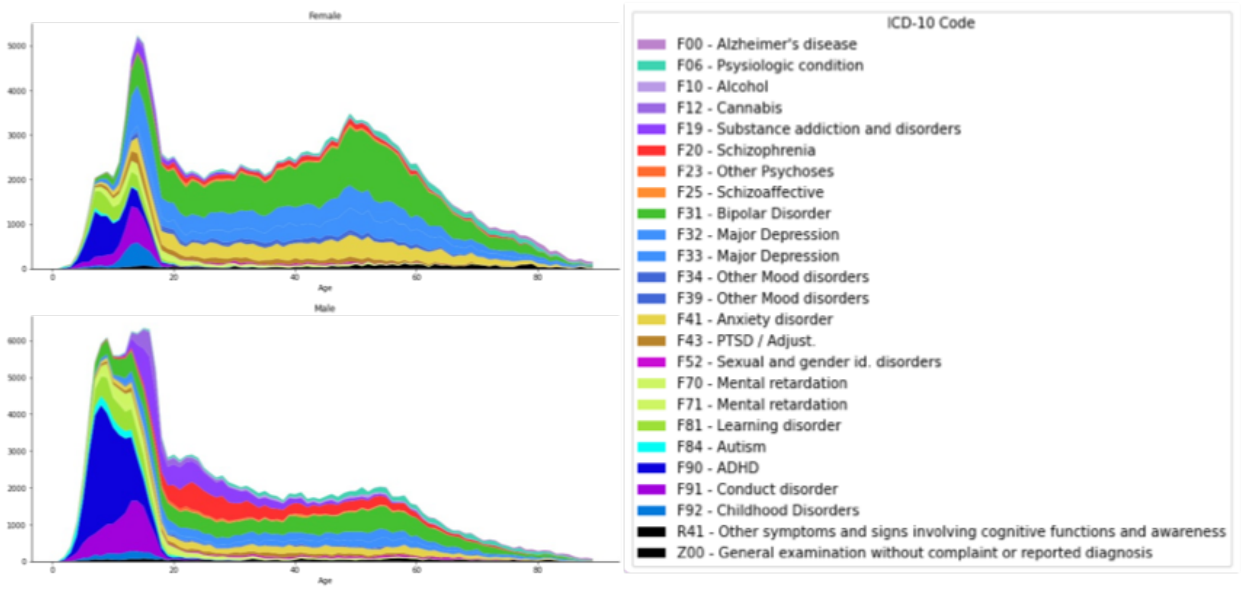


Figure 2.2 *Area plot of the distribution of common ICD-10 diagnoses by age and sex. A) Female (N= 47,646), B) Male (N= 44,870).*

2.7. Discussion

The EHR database at CSJDM provides an important tool for research in serious mental illness [4,5,8], as has been performed through collaborations between investigators at UdeA, CSJDM, and UCLA. The approach used for structuring the data from this EHR for research purposes may be extended to enable further research that harmonizes these data with those from other Colombian hospitals. Such data integration requires rigorous data quality assessments on topics such as data extraction and deidentification.

2.8. Conclusion

Effectively leveraging EHRs is critical to advancing precision medicine. Developing the CSJDM de-identified EHR database illuminates complexities and opportunities in this field. As EHRs continue to gain prominence in genomic research, it is imperative to ensure data precision, reliability, and inclusivity.

2.9 References

1. Meystre, S. M. et al. Clinical Data Reuse or Secondary Use: Current Status and Potential Future Progress. Yearbook of medical informatics vol. 26 38–52 Preprint at https://doi.org/10.15265/IY-2017-007 (2017).

2. Gagalova, K. K., Angelica Leon Elizalde, M., Portales-Casamar, E. & Görges, M. What you need to know before implementing a clinical research data warehouse: Comparative review of integrated data repositories in health care institutions. JMIR Formative Research vol. 4 Preprint at https://doi.org/10.2196/17687 (2020).

3. Bastarache, L. et al. Developing real-world evidence from real-world data: Transforming raw data into analytical datasets. Learn Health Syst 6, (2022).

4. Song, J. et al. Geospatial analysis reveals distinct hotspots of severe mental illness. medRxiv (2022) doi:10.1101/2022.03.23.22272776.

5. Service, S. K. et al. Distinct and shared contributions of diagnosis and symptom domains to cognitive performance in severe mental illness in the Paisa population: a case-control study. Articles Lancet Psychiatry vol. 7 www.thelancet.com/psychiatry (2020).

6. Steve Alder. What is the HITECH Act? The HIPAA Journal https://www.hipaajournal.com/what-is-the-hitech-act/.

7. Office for Civil Rights. Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule. (2012).

8. De la Hoz, J. F. et al. Analysis of diagnosis instability in electronic health records reveals diverse disease trajectories of severe mental illness. medRxiv (2022) doi:10.1101/2022.08.20.22279007.

CHAPTER 3

Electronic health records reveal transdiagnostic clinical features and diverse trajectories of serious mental illness.

Abstract

Objective: Electronic health record (EHR) databases enable scalable investigations of serious mental illness (SMI), including bipolar disorder (BD), severe or recurrent major depressive disorder (MDD), schizophrenia (SCZ), and other chronic psychoses. The authors analyzed structured and unstructured EHR data from a large mental health facility to characterize SMI clinical features and trajectories.

Methods: Diagnostic codes, information from clinical notes, and healthcare use data, were extracted from the EHR database of Clínica San Juan de Dios in Manizales, Colombia for the years 2005-2022, including 22,447 individuals (ages 4-90, 60% female) treated for SMI. The reliability of diagnostic codes was assessed in relation to diagnoses obtained from manual chart review (n=105). A Natural Language Processing (NLP) pipeline was developed to extract features from clinical notes. Diagnostic stability was quantified in patients with ≥ 3 visits (n=12,962). Finally, mixed-effect logistic regression models were used to identify factors associated with diagnostic stability.

Results: Assigned EHR diagnoses showed very good agreement with those obtained from manual chart review (Cohen's kappa 0.78). The NLP algorithm (which demonstrated excellent balance between precision and recall with average F1=0.88) identified high frequencies of suicidality and psychosis, transdiagnostically. Most SMI patients (64%) displayed multiple EHR diagnoses, including switches between primary diagnoses (19%), comorbidities (30%), and combinations of both (15%). Predictors of changes in EHR diagnoses include Delusions in clinical notes (OR=1.50, $p=2e^{-18}$) and a history of previous diagnostic changes (OR=4.02, $p=3e^{-250}$).

Conclusions: Longitudinal EHR databases enable scalable investigation of transdiagnostic clinical features and delineation of granular SMI trajectories through the integration of information from clinical notes and diagnostic codes.

3.1 Introduction

Examination of disease trajectories through longitudinal observation of symptoms led to the development of modern classification systems for mental disorders; such data formed the main basis for differentiating key categories of serious mental illness (SMI), such as schizophrenia (SCZ), bipolar disorder (BD), and major depressive disorder (MDD). While these classification systems advocate a parsimonious, longitudinal perspective, current research on SMI relies primarily on cross-sectional assessments, usually of patients with a unique diagnosis, in which the only available trajectory information is supplied by patient recall [1–3]. This lack of detailed longitudinal data may be a factor contributing to heterogeneity within our current SMI categories [4], as evidenced in cross-disorder genetic studies [3] Furthermore, concentrating on individual diagnoses ignores the fact that many features of psychiatric illness (such as suicidality or psychosis) are important transdiagnostically.

Recent studies using longitudinal data collected from participants in national registries [5,6], precision health initiatives [7], and birth cohorts [8], have begun to identify risk factors for specific diagnoses and to describe patterns of variation in these diagnoses over time [5–8]. These resources, which are mainly limited to upper-income countries (UIC), typically contain only sparse data for individual clinical features, including symptoms and behaviors. In contrast, electronic health record (EHR) data which are available in both UIC and in many low- or middle–income countries (LMIC), may contain extensive, descriptions of such clinical features during the periods when individuals actually experience them. EHR databases thus facilitate investigations of features that are important both transdiagnostically and longitudinally, and that may predict clinically important outcomes, such as the onset of psychosis or suicidal behaviors [9,10].

We demonstrate here that the EHR database from a psychiatric hospital located in a middle-income country, enables longitudinal population-scale investigations of SMI diagnoses, individual clinical features, and trajectories [11], through the combination of novel NLP methodologies applied to detailed clinical notes with analyses of structured data. The Clínica San Juan de Dios in Manizales (CSJDM) [12] implemented an EHR in 2005, which provided us both structured and unstructured data from all visits of more than 20,000 SMI patients, from that date until June 2022. The CSJDM provides comprehensive mental healthcare to the one million inhabitants of Caldas, and its EHR captures data concerning SMI covering the entire region[11]. We characterized trajectories of SCZ, BD, and MDD through longitudinal analyses of diagnoses (from assigned diagnostic codes) and of four features (Suicidal Ideation, Suicide Attempt, Delusions, and Hallucinations) described in clinical notes recorded in the EHR, each of which may be important in categorizing SMI and its trajectories, transdiagnostically.

To conduct these analyses, we extracted diagnoses and developed a Natural Language Processing (NLP) pipeline for extraction of transdiagnostic features from the free-text, in Spanish. We first established the reliability and completeness of the EHR and our phenotype extraction pipelines, and showed that features recorded in the notes at individual visits align with

ICD-10 diagnostic severity qualifiers at those visits. We then characterized trajectories of each SMI diagnosis, distinguishing between diagnostic *switches* and the accumulation of *comorbidities*, and identifying both global and diagnosis-specific patterns for each. We quantify the probability of patients changing diagnoses across admissions or visits and identify factors contributing to such changes. In particular, we evaluate the utility of transdiagnostic features from narrative notes in delineating SMI trajectories.

3.2 Methods

3.2.1 EHR Database

To investigate SMI trajectories we extracted from the EHR both structured data (demographic information; duration, type and site of visits [inpatient, outpatient, or emergency department]; diagnostic codes [ICD-10]), and unstructured data, consisting of free-text from clinical notes. These notes include psychiatrists' intake, progress, and discharge notes and nursing notes (from inpatient hospitalizations); psychiatrists' outpatient notes, and psychiatrists' triage notes (from emergency department visits).

Prior to performing the analyses reported here, we removed from the EHR any fields considered Protected Health Information [13], using procedures approved by the institutional review boards from UCLA, UdeA, and CSJDM. We used regular expression matching to strip from the text names and numbers exceeding five digits (potential ID numbers), to further reduce the possibility of including identifying information in our dataset.

For our analyses, we first identified all patients in CSJDM with at least one clinical note (n=77,538) and excluded patients with missing gender information (n=626). We then excluded visits outside the age range of 4-90, without a valid diagnostic code or with primary diagnostic codes outside of the Mental, Behavioral, and Neurodevelopmental Disorders categories (excluded n=20,982 visits from 5,056 patients, Supplementary Figure 1).

3.2.2 ICD-10 Codes Extraction and Cohort Definition

Following each visit to the hospital, a patient is assigned a single primary ICD-10 diagnosis by their treating psychiatrist, generating a time-stamped sequence of diagnoses. We extracted this sequence for every patient and selected for analyses patients who had at least one primary diagnosis of SMI, defined here as BD (F301, F302, F310, F311, F312, F313, F314, F315, F316, F317), Severe/Recurrent MDD (F322, F323, F331, F332, F333, F334), SCZ (F20X), and other chronic psychoses (Delusional Disorder; F22X. Schizoaffective Disorder; F25X). (Supplementary Table 1). In total, this cohort includes 22,447 patients with 157,003 visits (Supplementary Figure 1).

3.2.3 Primary Diagnosis Classification and Reliability Estimation

We assessed, in a subsample of the 22,447 patients described above, the reliability of the ICD-10 diagnoses recorded in the EHR in comparison with those made by an expert research clinician (MC) performing a complete manual chart review. To enable a sufficiently precise estimation of the degree of agreement between these two sets of diagnoses, we selected 120 patients for this record review, chosen at random from among participants whom we had previously recruited at CSJDM in an ongoing study of BD, MDD, and SCZ [12]. Of these individuals (40 from each of the three diagnostic groups), we excluded 15 whose most recently recorded ICD-10 diagnoses (F318, F319, F321, F328, or F339) were not among the codes that met our criteria for SMI, as defined above. The clinician review of the remaining 105 records yielded a checklist of symptoms and other clinical features of SMI (see below) and assignment of a *current* primary diagnosis based on DSM-5 criteria [14]. We then, in these 105 patients, evaluated (using Cohen's kappa statistic [15]) the agreement between the diagnosis assigned through this review and the *most recently* recorded ICD-10 SMI code.

3.2.4 NLP Algorithm to Extract Clinical Features

To enable the identification of specific clinical features in the clinical notes of each of the 22,447 patients with an SMI primary diagnosis, in the CSJDM database, we developed a Spanish-language NLP algorithm; the procedures used to develop, train, and validate this algorithm are detailed in Supplementary Note 1 (also see Supplementary Tables 2 and 3). For this study, we used the algorithm to identify the presence of four transdiagnostic features that are routinely assessed in clinical encounters: Suicide Attempts, Suicidal Ideation, Delusions, and Hallucinations.

Briefly, two clinicians independently reviewed a randomly selected sample consisting of 3,600 passages of free text (which we term "sentences") from the inpatient notes, progress notes, and outpatient notes of 2,788 unique patients with ICD-10 SMI codes, flagging those sentences in which any of the four features were present. We stopped sentence annotation at this point, as we had identified a sufficient set of positive instances of each feature to obtain accurate estimates of the algorithm's performance.

We evaluated the algorithm's performance using a held-out, gold standard set of 290 sentences, in which each putative feature was also annotated as being affirmed or negated. Metrics used to assess performance are precision (positive predictive value), recall (sensitivity), and their harmonic mean (F1). Then, we designated features as “present” for a given patient, if they were identified by the algorithm as having been mentioned affirmatively in a least two notes over the entire course of their EHR. The requirement of two notes was selected as it yielded optimal performance of the NLP algorithm (Supplementary Figure 2). To determine the accuracy with which the algorithm designated the presence or absence of each feature, we compared its output (for the patients in whom we had conducted a manual chart review, as described in the section above), to the checklist of these features compiled by the clinician from the chart review (using the same metrics as for the sentence level evaluation, described above). Finally, as an evaluation of the attributions made by the clinician conducting in the manual review, we conducted an additional set of manual reviews of selected records; two additional clinicians conducted independent reviews of each of the charts with false positive instances (and an equal number of randomly chosen charts with true positives, Supplementary Note 1).

3.2.5 Characterization of Extracted Clinical Features in Relation to ICD-10 Codes

We evaluated, in the entire study cohort, the correspondence between the clinical features extracted from notes using the NLP algorithm and the current-state severity qualifiers in the ICD-10 diagnoses recorded in the EHR at each individual visit. For these comparisons we used a mixed-effect logistic regression, which accounts for multiple visits per person. Additionally, we include covariates to correct for potential confounding factors such as inpatient status.

3.2.6 Patient-level Associations Between Clinical Features and ICD-10 Diagnoses

We assessed the relationship, at the individual level, between the presence of clinical features at any timepoint in the EHR, the most recently recorded diagnoses (considering only codes for

MDD, BD, and SCZ) and gender. For this analysis, we evaluated all patients with an SMI diagnosis and at least two separately recorded clinical notes (n=20,658 out of 22,447 SMI patients, Supplementary Figure 1). Association tests were performed using logistic regression including both diagnosis and gender while adjusting for the length of patients' records and hospitalization history (Supplementary Note 2). We tested four models, one for each feature. To test for interactions between gender and diagnosis, we expanded the model to include an interaction term.

To evaluate the relationship between clinical features that co-occur in patients (considering their entire longitudinal EHR) we used the same logistic modelling framework, but added, for each feature, the presence, recorded at any point, of the three remaining features, in the same model. To test for interactions between gender and the second co-occurring feature, we expanded the model to include an interaction term.

3.2.7 Diagnostic Switches, Comorbidities, and Trajectories

We defined two types of diagnostic changes: diagnostic switches and comorbidities. We use the term *diagnostic switches* to refer to changes between two psychiatric diagnoses that cannot, by definition, be held at the same time, specifically, the diagnoses in the ICD-10 F3 and F2 chapters (mood and psychotic disorders, respectively; see Supplementary Note 3 and Supplementary Table 4 for details). By contrast, we use the term *comorbidities* to refer to all other combinations of ICD-10 codes; comorbid diagnoses can accumulate over time, without limit. Using these two definitions, an individual patient's diagnostic trajectory may include both switches and comorbidities.

3.2.8 Diagnostic Stability

We assessed the stability of diagnostic categories over time, considering a diagnosis unstable only if a patient switched to another diagnosis (Supplementary Table 4). For individual SMI diagnoses, we estimated long-term prospective and retrospective stability in those individuals with > 10 visits, (n=12,962, Supplementary Figure 1); we chose this number of visits to represent trajectories of sufficient length for analysis of stability to be meaningful [16,17]. Prospective stability is the probability of a patient's first diagnosis being the same as their last diagnosis, and is analogous to the precision of the initial diagnosis in predicting the final diagnosis. Retrospective stability is the probability of a patient's final diagnosis being the same as their first one and is analogous to recall of the first diagnosis relative to the final. Differences in stability across diagnoses and age groups (before or after age 30) were evaluated using z-tests (Supplementary Note 4).

3.2.9 Factors Affecting Diagnostic Stability

Next, we explored factors contributing to visit-to-visit diagnostic stability. Specifically, we evaluated the effects of patient sex and age, primary diagnosis, inpatient status, previous switch, clinical features, and receiving a Not Otherwise Specified (NOS) code at the previous visit. For these analyses we first used a mixed-effect logistic regression to estimate the probability of switching diagnoses over time (using number of visits as a proxy), accounting for repeated patient observations. We then expanded this model to evaluate the effects of the demographic and clinical factors listed above (Supplementary Note 5). An NOS code indicates diagnostic uncertainty in cases of atypical or confusing patient presentations, or when temporal criteria are not yet met [18], and therefore serves as a positive control; we expect to see increased diagnostic instability associated with these codes. In a sensitivity analysis, we explored the impact of measuring time by years since the first encounter rather than by visit number.

Finally, to evaluate the possibility that clinical features extracted from the notes at a given visit *anticipate* specific diagnostic changes recorded in future visits, we tested whether psychosis features (Delusions and Hallucinations) predict the application of psychosis current-state qualifiers in ICD-10 diagnoses of BD or MDD (Supplementary Note 6).

3.2.10 Significance Thresholds

We applied Bonferroni correction for multiple testing in all our analyses. Model details with corresponding significance thresholds are described in Supplementary Notes 2-6.

3.3 Results

3.3.1 Study Sample

As of June 2022, the CSJDM EHR included 157,003 visits from 22,447 patients who were assigned an SMI diagnosis at any point from their first visit onwards (Supplementary Figure 1). The demographic and clinical characteristics of this sample are described in Supplementary Table 5.

3.3.2 Reliability of Diagnoses and Clinical Features Extracted from EHR Compared to those Identified through Manual Chart Review by Expert Clinicians

For the 105 randomly selected patients in whom we conducted a complete manual chart review, the diagnoses extracted from their EHR (most recent assigned ICD-10 code) demonstrated agreement with their diagnoses from manual review of their entire EHR at a kappa level typically considered "very good" to "excellent" for such comparisons [19,20]. The kappa estimates for specific diagnoses considering both inpatient visits and outpatient visits (Supplementary Table 6) were: 0.74 (95% CI: 0.60-0.89) for MDD, 0.74 (95% CI: 0.60-0.87) for BD, 0.90 (95% CI: 0.81-0.99) for SCZ; overall kappa=0.78 (95% CI: 0.69-0.88). Positive predictive values (PPVs) were 0.84 for MDD, 0.80 for BD, and 0.92 for SCZ. Estimates for kappas and PPVs were similarly high when considering inpatient visits only (Supplementary Table 6). These levels of agreement are also similar to those from previous studies comparing diagnoses from ICD codes recorded in the EHR with diagnoses from manual chart reviews [21].

In training our NLP algorithm for extracting clinical features from the EHR notes, the kappas indicated "good" to "excellent" agreement between two independent annotators for all four features (Supplementary Table 7). Then, application of the algorithm to extract features from the gold standard set of sentences, demonstrated that it performed with a high rate of precision (range: 0.88-1.0) and recall (range: 0.62-1.0), resulting in a satisfactory F1 for all features (Suicide Attempt: 0.82, Suicidal Ideation: 0.73, Delusions: 1.0, Hallucinations: 0.95), (Supplementary Table 8A, see Supplementary Note 7 for a description of errors). We evaluated different thresholds for the number of affirmative mentions of a feature in the output of the NLP algorithm that we would require to consider that feature "present", for a given patient, over the lifetime of their EHR (Supplementary Figure 2); requiring two such mentions provided an optimal balance (as measured by F1) between precision and recall. At this threshold the algorithm output and the designation of features from manual chart review were highly concordant for all four features; Suicide Attempt (92/104, F1 = 0.68), Suicidal Ideation (89/104, F1 = 0.79), Delusions (84/104, F1 = 0.82), and Hallucinations (87/104, F1 = 0.84), Supplementary Table 8B. Further investigation suggested that the above comparison actually underestimated the performance of the algorithm (Supplementary Note 1).

3.3.3 Comparison Between ICD-10 Diagnoses of MDD and BD Assigned at Each Visit with Clinical Features Identified from the Notes at the Same Visit

The ICD-10 codes for MDD and BD include qualifiers indicating the severity of episodes (unipolar depressive, bipolar depressive, and manic) and the presence or absence of psychotic features at a given visit; this information provides the opportunity to evaluate the relationship between these qualifiers and the clinical features extracted from the notes using the NLP algorithm. As would be expected, we observe strong positive associations for all four of the extracted clinical features with both episode severity and presence of psychosis, as recorded in the ICD-10 codes (Supplementary Note 6, Supplementary Table 9).

In contrast to the ICD-10 codes, which simply report the presence or absence of psychotic symptoms at a given visit, the application of the NLP algorithm to the clinical notes for these visits provides rich information on such symptoms. The notes reveal that, in depressive episodes (both unipolar and bipolar) Delusions and Hallucinations are observed at relatively similar frequencies, while in manic episodes Delusions represent by far the predominant psychotic feature (Figure 1). This finding is consistent with observations made in a manual review of 1,715 case notes from a North American tertiary care hospital [22], while a systematic review article of types of psychotic symptoms in bipolar depression and mania reported less consistent pattern across multiple, smaller studies [23].

The notes contain corroborating examples for most of the instances in which an ICD-10 diagnosis of depression or mania with psychotic features was assigned; for visits at which such diagnoses were recorded, the algorithm identified either Hallucinations or Delusions at a frequency ranging from 72% (unipolar depression) to 84% (mania, see Figure 1A). Examples of such psychotic features were also found, however, in a substantial proportion of the notes of visits for which the recorded ICD-10 diagnosis indicated an absence of psychotic features (ranging from 21% [unipolar depression] to 56% [mania]), Figure 1B). It is unclear what accounts for this apparent discrepancy, but one contributing factor could be differences between the clinicians recording ICD-10 diagnoses and the NLP algorithm in how they consider psychosis. Such differences could at least partially explain the high frequency of Delusions in the notes from visits for which the diagnosis of "mania without psychotic features" was assigned (47%). Grandiose delusional beliefs comprise more than 30% of the Delusions for such visits identified by the NLP algorithm (Supplementary Figure 3, but, in practice, the point at which grandiosity (a cardinal feature of mania) reaches psychotic proportions may be difficult to define [24]. Additionally, persistence of delusionary beliefs may have been a consideration for clinicians recording ICD-10 diagnoses, while in some instances the information extracted from the notes may reflect transient beliefs that resolved relatively quickly.

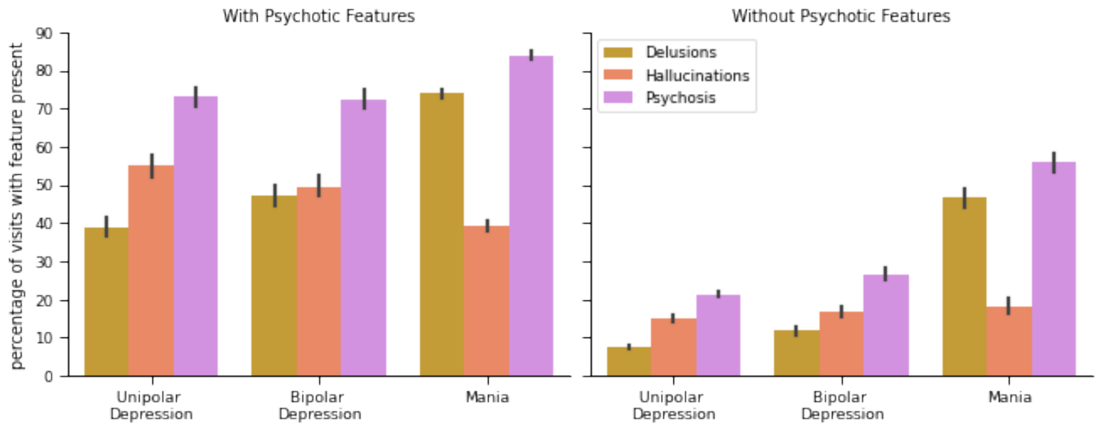


Figure 3.1 *Psychotic features extracted from the clinical notes of visits for severe unipolar and bipolar depression and mania. The ICD-10 codes for these disorders (F32, F33, F31) include qualifiers for the clinician to specify the presence or absence of psychosis for each visit. (1A) The percentage of visits assigned a diagnosis of "with psychotic features" for which the NLP algorithm identified the features Delusions and Hallucinations, considered together ("Psychosis") and separately. (1B) The percentage of visits assigned a diagnosis of "without psychotic features" for which the NLP algorithm identified the features Delusions and Hallucinations, considered together ("Psychosis") and separately. Error bars indicate 95% confidence intervals obtained through bootstrapping, n=5,961 patients and 13,928 visits.*

3.3.4 Transdiagnostic characterization of features extracted from EHR notes

The above comparisons indicated that the NLP algorithm identifies the presence of psychosis and suicidality clinical features in the EHR database with high sensitivity. Suicide Attempt, Suicidal Ideation, Delusions, and Hallucinations each occur in all of the SMI diagnoses at a frequency of >5%, stratified by gender, demonstrating their transdiagnostic quality (Figure 2A). Several patterns of these frequencies are, however, noteworthy. In contrast to most reports in the literature [25,26], Suicidal Ideation in the CSJDM database is less frequently observed in females compared to males, after correcting for diagnoses, inpatient history and number of visits (OR=0.84, p=8.42e^{-7}, Supplementary Tables 10 and 11). This difference is driven largely by a lower rate of Suicidal Ideation in females with MDD specifically (34% versus 45% in males, interaction OR=0.65, p=4.2e^{-9},). A similar reduced frequency of psychotic features was observed in females (OR=0.67, p=1.99e^{-22} Delusions; and OR=0.88, p=5.9e^{-4} Hallucinations,), while rates of Suicide Attempt are similar in both genders. The four features subdivide the patient population according to the combination of comorbidities (Figure 2B). Aside from the expected co-occurence of suicide-related features and psychotic features, we found, unexpectedly, that the mention of Delusions in the notes decreases the likelihood of notes mentioning Suicidal Ideation or Suicide Attempts in the same patient and vice versa (OR between 0.59-0.62, p < 1.78e^{-17}, accounting for gender, diagnosis, inpatient history and number of visits [Supplementary Table 11); the reverse is true for Hallucinations (OR between 1.29-2.05, p < 2.89e^{-7}).

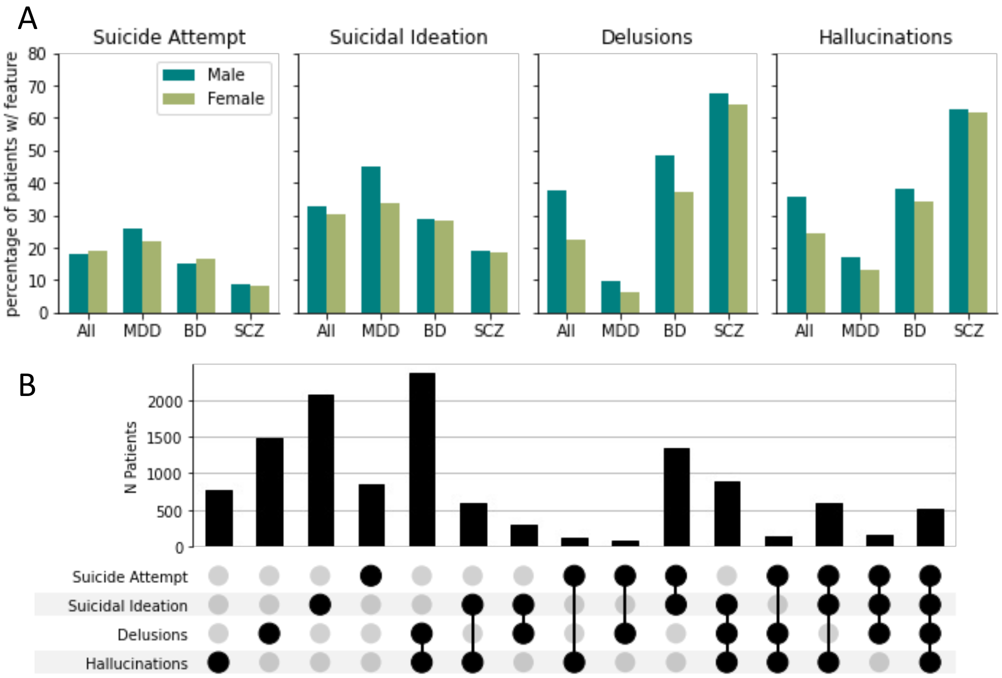


Figure 3.2 Transdiagnostic characterization and co-occurrence of clinical features extracted from EHR notes. A) Proportion of patients with each of the four features stratified by primary diagnosis. B) Number of patients with co-occurrence of 2, 3, or 4 clinical features. All data in these plots are limited to patients with at least two EHR notes.

3.3.5 Diverse Diagnostic Trajectories in SMI Patients

We evaluated diagnostic trajectories among all SMI patients with at least three visits (n=12,962, Supplementary Figure 4). The majority (64%, Figure 3A) had multiple diagnoses recorded in their EHR, broken down as follows: 30% displayed comorbidities (orange bars; Supplementary Table 12, 19% displayed diagnostic switches (teal bars), and 15% displayed both switches and comorbidities (purple bars).

While some pairs of diagnoses in the trajectories are common, for example, the switch from MDD to BD (observed in 24% of current BD patients, figure 3B) and the comorbidity between MDD and Other Anxiety Disorders (observed in 28% of current MDD patients), the majority of patients (58%) follow rare trajectories (occurring in fewer than 1% of patients). Altogether, we observed 3,149 unique trajectories.

We estimated prospective and retrospective stability for each diagnosis, evaluating trajectories with 10 or more visits (n=5,016). Prospective stability was lower for MDD compared to BD or SCZ (56% vs. 88% and 83%, respectively; 2-df chi-square=383; $p=5e^{-84}$). Retrospective stability, by contrast, while lower than prospective stability for all diagnoses, was highest in MDD (53% vs. 48% and 40% in BD and SCZ respectively; 2-df chi-square=34.5; $p=3e^{-8}$). (Supplementary Note 4 and Supplementary Table 13).

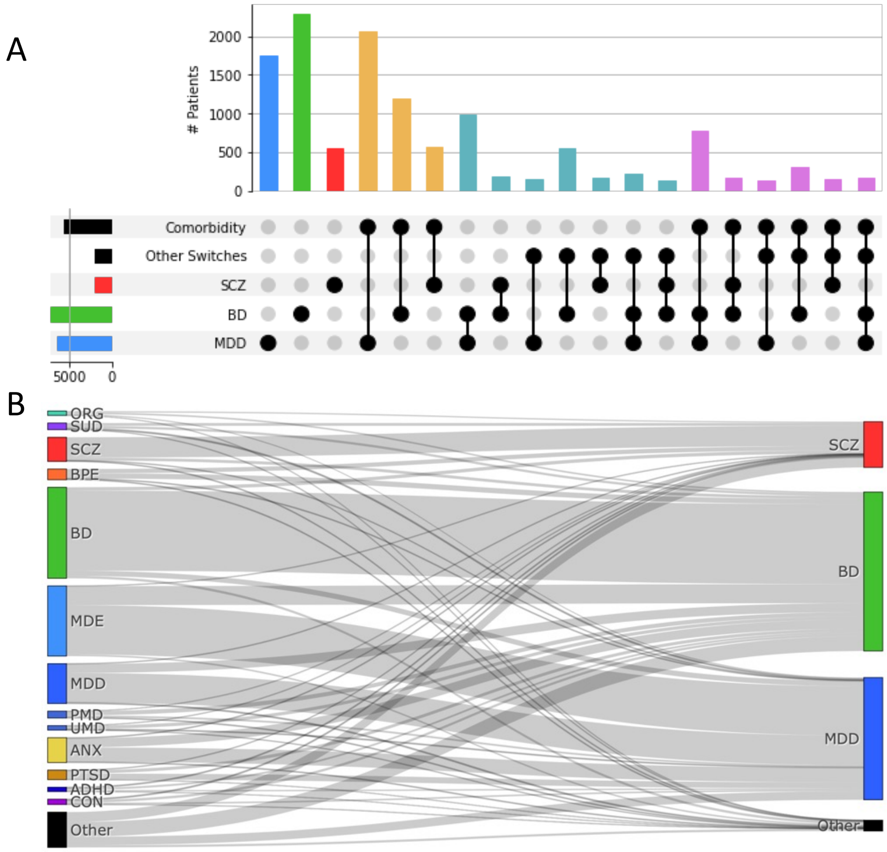


Figure 3.3 Disease trajectories of SMI in patients with at least three visits. A) UpSet plot presenting diagnostic switches (between SMI categories) and comorbidities (SMI and non-SMI categories). Patients with a single SMI diagnosis (blue, green, red, total n=4,620); a single SMI diagnosis and other comorbidities (orange n=3,955); multiple SMI diagnoses and no other comorbidities (teal n=2,468); multiple SMI diagnoses and other comorbidities (purple, n=1,919). Bars with n<100 are not shown. B) Sankey diagram of ICD-10 code trajectories. Left nodes represent the diagnosis given at the initial visit and right nodes represent the most recent SMI code. (Diagnostic switches within SMI are shown in Supplementary Figure 4). ORG: Other mental disorders due to brain damage and dysfunction and to physical disease (F06), SUD: Mental and behavioral disorders due to multiple drug use and use of other psychoactive

substances (F19), BPE: Acute and transient psychotic disorders (F23), MDE: Major Depressive Episode (F32), PMD: Persistent mood disorders (F34), UMD: Unspecified mood disorder (F39), ANX: Other anxiety disorders (F41), PTSD: Reaction to severe stress, and adjustment disorders (F43), ADHD: Hyperkinetic disorders (F90), CON: Conduct disorders (F91)

3.3.6 Clinical Features, Time, and Other Factors Affecting Diagnostic Stability and Specific Trajectories

We identified multiple factors that influenced diagnostic stability. Diagnostic switching was most frequent during the early stages of treatment. While 11.3% of the patients changed diagnosis on their second visit, this percentage decreased over the patient's course of illness (Figure 4A; log10(k) OR=0.56, p-value $5e^{-66}$) and stabilized at around 4% after the tenth visit. Additional predictors of future diagnostic instability included the following observations at the current visit: a diagnostic switch from the previous visit (Figure 4B; OR=4.02, p-value $3e^{-250}$), an inpatient visit (OR=1.7, p-value $5e^{-35}$), an NOS diagnosis (OR=1.61, p-value $2e^{-47}$), and the presence of the clinical features Delusions or Hallucinations (OR=1.50 and 1.17, p-values $2e^{-18}$ and $3e^{-4}$, respectively). Predictors of future diagnostic stability included: diagnoses of SCZ or BD compared to MDD (ORs=0.31 and 0.32; p-values $<3e^{-70}$), male gender (OR=0.71, p-value $2e^{-16}$), and increasing age (OR per decade=0.96, p-value $8e^{-4}$). Sensitivity analyses confirmed these findings; the same pattern was observed when modeling switching by time rather than visit number (Supplementary Figure 5).

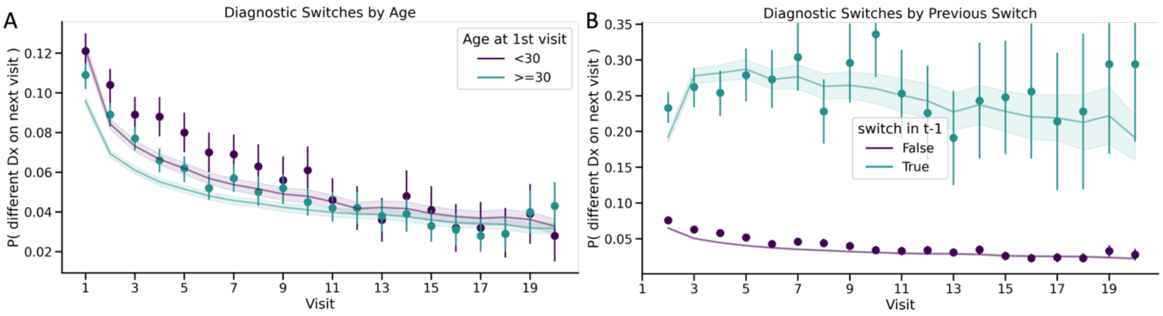


Figure 3.4 *Diagnostic stability over time. At each visit k, the proportion of patients that will switch primary diagnosis code on their next visit k+1. A) Stratified by age groups: age at 1*st *visit before and after 30 years. B) Stratified by having previously switched diagnoses (from visit k-1). n=12,962 patients (Supplementary Figure 1).*

3.4 Discussion

By analyzing EHR data spanning 17 years and encompassing over 20,000 patients from a single large mental health facility, we characterize SMI, its relation to transdiagnostic clinical features, and its longitudinal trajectories. We show that both the diagnostic codes recorded in the EHR and our custom, rule-based NLP pipeline for extracting clinical features from narrative notes, reliably reflect the diagnostic impressions of expert clinicians conducting manual chart reviews. By applying this pipeline to our data we were able to perform more granular analyses of four key clinical features of SMI – Suicide Attempts, Suicidal Ideation, Delusions, and Hallucinations – than has been possible in most previous EHR studies of SMI which have relied only on the information in the diagnostic codes. Additionally, analysis of the NLP-algorithm output in relation to the recorded diagnostic codes reveals the high degree of association between the information contained in these two types of data. The four clinical features each occur at a high frequency transdiagnostically, and they co-occur in both expected and unexpected patterns. Finally, we use the information on clinical features in the notes together with the ICD-10 diagnostic codes to characterize SMI trajectories, including prediction of future diagnostic changes, differentiation of diagnostic switches from the accumulation of comorbidities, and factors contributing to the stability of diagnoses.

Our NLP pipeline overcomes the performance and usability issues of "off-the-shelf" NLP pipelines [27,28] and is, to our knowledge, the first of its kind for extracting information from Spanish language psychiatric notes. The algorithm identified the clinical features in the notes from multiple individuals where the initial chart review failed to do so, and its results were mostly confirmed by a further round of manual reviews. These observations suggest, therefore, that this automated approach to retrieving clinical data is more accurate than manual review, as well as being vastly more scalable.

In considering the data on psychotic features that we obtained from the clinical notes in relation to what we gleaned from the ICD-10 codes, we illustrate the potential utility of the NLP pipeline for analyzing a vast array of information that is often unavailable in large-scale studies of SMI, for which the phenotypes may be based on brief interviews or self-report scales (e.g., [29]). Notably, several previous studies have reported conflicting results with respect to the relative frequencies of Delusions and Hallucinations in psychotic mania compared to psychotic depression, which may reflect their differing designs and mostly small sample sizes of the studies that they have evaluated [23]. In contrast, by applying a uniform methodology to analyze information extracted from the notes of nearly 6,000 patients, in almost 14,000 clinical encounters, we were able to show that Delusions are far more frequent than Hallucinations in mania, while Hallucinations have the same or even greater frequency than Delusions in both unipolar and bipolar depression. This observation only partially reflects grandiose delusional beliefs characteristic of mania. Further development of the NLP algorithm will enable even finer-grained characterization of psychosis in these disorders, e.g., specifying whether psychotic symptoms in the notes are congruent or incongruent with the predominant mood states of an episode [23].

Our comparison of the information from the notes and from the diagnostic codes also highlights the complementarity of these data sources, as, in most instances, the former provide specific examples to corroborate the assignment of the “with psychotic features” codes. On the other hand, the apparent discrepancies between the notes and the codes – the identification of psychotic features in the free text at visits where such features were not recorded in the diagnoses – provide a reminder that the EHR is primarily a clinical record and the contents of its various components reflect their differing purposes. The assigned diagnostic codes represent a clinician’s overall impression of a patient’s predominant clinical state at a given visit, rather than a comprehensive representation of all of the clinical information obtained at that visit. Even though such information may not be incorporated within the formal diagnosis, it is, however, of great value for addressing a number of important questions.

Widespread recognition of the inadequacies of diagnosis-based taxonomies for classifying mental illness has generated growing interest in investigation of transdiagnostic features of SMI [30]. Evidence has accumulated indicating a high degree of shared genetic risk across SMI diagnoses, but assembling adequately-scaled datasets that are suitable for genetic analysis of systematically assessed transdiagnostic phenotypes has proven challenging. Our finding that all four of the clinical features that we extracted from the EHRs are present in substantial numbers in each of the SMI diagnostic categories suggests that extending this approach to additional features, and to the EHR databases of additional facilities, could enable well-powered genetic association studies of a number of phenotypes that are relevant to SMI. Further transdiagnostic explorations will also be important to follow up unexpected findings from our current analyses for which we have no obvious explanation, including the lower rates [lifetime EHR] of Suicidal Ideation, Delusions, and Hallucinations identified in females compared to males, and the contrasting patterns of association between the suicidal features and the psychotic features (negative for Delusions, and positive for Hallucinations).

Our findings regarding longitudinal diagnostic trajectories of SMI rest on an EHR database that is, to our knowledge, unique. Because the CSJDM provides comprehensive care to all individuals living with SMI in a geographically defined catchment area, the EHR provides an essentially continuous record of treatment encounters, over this period in the more than 20,000 individuals included in our analyses. These data enabled us to model simultaneously the dynamics of switching between incompatible SMI diagnoses and the accumulation of comorbidities; by doing so, we found that approximately half of the SMI population had one or more psychiatric comorbidities, and over one-third have switched diagnosis at least once. The most frequent comorbities are between anxiety disorders and MDD, while the most frequent diagnostic switch is between MDD and BD. The combination of comorbidities and diagnostic switches describes a broad variety of disease trajectories; a major challenge for future studies will be to determine which of these patterns are most meaningful, either from the standpoint of our understanding of disease causation or in terms of clinical utility. Additionally, as have others we found that diagnostic instability is characteristic mainly of the early stages of SMI [5,31].

Most previous studies of SMI trajectories at a population level have utilized national registries available mainly in a few Northern European countries [5,6] In particular, studies of diagnostic progression in patients with an index psychiatric diagnosis have observed similar degrees of instability of initial SMI diagnoses as we report here [6]. While the concordance of our results with those of these registry studies provides an additional validation for our approach, our integration of features from clinical notes together with diagnostic codes and their qualifiers, has enabled us to delineate trajectories at a level of granularity not available in registry data, and to identify patterns that could have important clinical implications. As an example, we observed that mentions of psychosis in the notes from a clinical visit significantly preceded diagnostic switches at future visits. More specifically, we noted that examples of Delusions in the notes from a clinical visit at which the ICD-10 qualifier "without psychotic features" was assigned, anticipated the application of the qualifier "with psychotic features" at the subsequent visit. This observation will stimulate further research with the aim of determining how such information could be used for clinical prediction.

Our finding that diagnostic switches increase the likelihood of future switches is driven in large part by a small number of patients who rapidly accumulate diagnoses. While we have not yet identified any specific features that characterize these patients, we hypothesize that they constitute a group for whom research classifications that assign all case-participants a single primary lifetime diagnosis are particularly imprecise descriptions of their phenotype. This group is difficult to recognize in cross-sectional studies, and we hypothesize that they may constitute a source of variance in the many large-scale psychiatric genetics investigations that rely on such classification systems. By contrast, utilizing longitudinal EHR data to selecti samples for genetic studies may not only facilitate the early identification of individuals with extreme diagnostic instability and reduce heterogeneity of research datasets, but alsoenable discovery of distinct features that characterize this group.

Limitations of the study are its focus on a single mental health care facility, the CSJDM, and the early stage of development of the NLP algorithm that we use here to automate the extraction of clinical information from its EHR. As we have already noted, the CSJDM was ideal for implementing our approach due to the extensiveness of its EHR database and given the fact that it continues to provide most of the care for SMI to the ~ 1,000,000 inhabitants of the Department of Caldas. We are now extending the approaches described here, including further testing of the NLP algorithm, to enable longitudinal studies of SMI in other facilities, including the Hospital Mental de Antioquia, one of the largest psychiatric hospitals in Colombia. These implementations will enable estimations of factors such as institutional biases in reporting styles.

Further development of the NLP algorithm to incorporate a larger number of symptoms and behaviors will enable us to place greater confidence in the validity of the assigned ICD-10 diagnoses, e.g., for analyses of trajectories or for future genetic association studies. For example, our current definition of diagnostic switches include individuals who have transitioned from a diagnosis of BD to one of MDD. Such a transition implies either that the clinician assigning the former diagnosis erroneously included an episode of mania or hypomania, or that the clinician assigning the latter diagnosis erroneously overlooked such an episode; our aim is to use the algorithm to help resolve such uncertainties. Additionally, while we currently only differentiate between affirmations and negations of particular terms, in future work we will incorporate a larger range of contexts that are relevant for delineating trajectories, e.g., distinguishing between symptoms that are improving and those that are worsening over the course of a hospitalization.

Our results support the notion that research classifications that incorporate past and future trajectory data will likely be less heterogeneous and more realistic than current systems that assign patients a single ‘lifetime’ diagnosis. Evidence from prior studies suggests that distinctive genetic risk profiles may partially underlie trajectory features, such as polarity at onset in BD [32] or conversion from non-psychotic to psychotic illness [33–35]. Efforts to replicate and extend such findings, however, have been limited by variation in ascertainment strategies, reliance on patient recall [32], and small sample sizes [33,35] . Future research should evaluate the relationship between genetic risk and diagnostic or clinical stability, with the aim of establishing more genetically homogeneous subgroups. As analyzing datasets with thousands of uncommon trajectories will be impractical, developing improved methods for reducing dimensionality by clustering patients with similar trajectories should be an important focus of future work[36]. EHR databases usually contain information on interventions, such as pharmacological treatments, that likely influence disease trajectories. Modeling this impact will be an important direction of future research.

3.5 Acknowledgments

Chapter 3 is based on a draft of a manuscript in preparation, author citation and affiliations are as follows:

Juan F. De la Hoz[1], Alejandro Arias[2], Susan K. Service[1], Mauricio Castaño[2], Ana M. Diaz-Zuluaga[1], Janet Song[1], Cristian Gallego[2], Sergio Ruiz-Sánchez[3], Javier I Escobar[4], Alex A. T. Bui[5], Carrie E. Bearden[1], Victor Reus[6], Carlos Lopez-Jaramillo[3], Nelson B. Freimer[1], Loes M. Olde Loohuis[1].

1. Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, USA
2. Department of Mental Health and Human Behavior, University of Caldas, Manizales, Colombia
3. Department of Psychiatry, University of Antioquia, Medellín, Colombia
4. Global Health, Robert Stempel School of Public Health and Social Work, Florida International University, Miami, USA
5. Department of Radiological Sciences, University of California Los Angeles, California, USA

6. Department of Psychiatry and Biobehavioral Sciences, University of California San Francisco, San Francisco, USA

Research reported here was supported by R01MH123157 (to LMOL, CLJ, and NBF), R01 MH113078 (to CEB, CLJ, and NBF), R00 MH116115 (to LMOL), T32 MH073526 (to JFDLH) and the Fulbright Commission in Colombia under the Fulbright-Colciencias grant (to JFDLH). The content is solely the responsibility of the authors and does not represent the official views of the Fulbright Program or the National Institutes of Health.

3.6 References

1. Kendler, K. S. The nature of psychiatric disorders. World Psychiatry 15, 5–12 (2016).

2. Hyman, S. E. The diagnosis of mental disorders: The problem of reification. Annual Review of Clinical Psychology vol. 6 155–179 Preprint at https://doi.org/10.1146/annurev.clinpsy.3.022806.091532 (2010).

3. Anttila, V. et al. Analysis of shared heritability in common disorders of the brain. Science (1979) 360, (2018).

4. Regier, D. A. et al. DSM-5 field trials in the United States and Canada, part II: Test-retest reliability of selected categorical diagnoses. American Journal of Psychiatry 170, 59–70 (2013).

5. Plana-Ripoll, O. et al. Exploring Comorbidity Within Mental Disorders among a Danish National Population. JAMA Psychiatry 76, 259–270 (2019).

6. Høj Jørgensen, T. S., Osler, M., Jorgensen, M. B. & Jorgensen, A. Mapping diagnostic trajectories from the first hospital diagnosis of a psychiatric disorder: a Danish nationwide cohort study using sequence analysis. Lancet Psychiatry 10, 12–20 (2023).

7. Barr, P. B., Bigdeli, T. B. & Meyers, J. L. Prevalence, Comorbidity, and Sociodemographic Correlates of Psychiatric Disorders Reported in the All of Us Research Program. JAMA Psychiatry 79, 622–628 (2022).

8. Caspi, A. et al. Longitudinal Assessment of Mental Health Disorders and Comorbidities Across 4 Decades Among Participants in the Dunedin Birth Cohort Study. JAMA Netw Open 3, e203221 (2020).

9. Barak-Corren, Y. et al. Predicting suicidal behavior from longitudinal electronic health records. American Journal of Psychiatry 174, 154–162 (2017).

10. Raket, L. L. et al. Dynamic ElecTronic hEalth reCord deTection (DETECT) of individuals at risk of a first episode of psychosis: a case-control development and validation study. Lancet Digit Health 2, e229–e239 (2020).

11. Song, J. et al. Geospatial analysis reveals distinct hotspots of severe mental illness. medRxiv (2022) doi:10.1101/2022.03.23.22272776.

12. Service, S. K. et al. Distinct and shared contributions of diagnosis and symptom domains to cognitive performance in severe mental illness in the Paisa population: a case-control study. Articles Lancet Psychiatry vol. 7 www.thelancet.com/psychiatry (2020).

13. Office for Civil Rights. Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule. (2012).

14. American Psychiatric Association. Diagnostic and statistical manual of mental disorders: DSM-5. vol. 5 (American psychiatric association Washington, DC, 2013).

15. Cohen, J. A Coefficient of Agreement for Nominal Scales. Educ Psychol Meas 20, 37–46 (1960).

16. Schwartz, J. E. et al. Congruence of Diagnoses 2 Years After a First-Admission Diagnosis of Psychosis. Arch Gen Psychiatry 57, 593–600 (2000).

17. Baca-Garcia, E. et al. Diagnostic stability of psychiatric disorders in clinical practice. British Journal of Psychiatry 190, 210–216 (2007).

18. First, M. B. et al. Do mental health professionals use diagnostic classifications the way we think they do? A global survey. World Psychiatry 17, 187–195 (2018).

19. Clarke, D. E. et al. DSM-5 Field Trials in the United States and Canada, Part I: Study Design, Sampling Strategy, Implementation, and Analytic Approaches. American Journal of Psychiatry 170, 43–58 (2013).

20. Kraemer, H. C., Kupfer, D. J., Clarke, D. E., Narrow, W. E. & Regier, D. DSM-5: How reliable is reliable enough? American Journal of Psychiatry vol. 169 13–15 Preprint at https://doi.org/10.1176/appi.ajp.2011.11010050 (2012).

21. Davis, K. A. S., Sudlow, C. L. M. & Hotopf, M. Can mental health diagnoses in administrative data be used for research? A systematic review of the accuracy of routinely collected diagnoses. BMC Psychiatry 16, (2016).

22. Black, D. W. & Nasrallah, A. Hallucinations and delusions in 1,715 patients with unipolar and bipolar affective disorders. Psychopathology 22, 28–34 (1989).

23. Chakrabarti, S. & Singh, N. Psychotic symptoms in bipolar disorder and their impact on the illness: A systematic review. World J Psychiatry 12, 1204–1232 (2022).

24. Canuso, C. M., Bossie, C. A., Zhu, Y., Youssef, E. & Dunner, D. L. Psychotic symptoms in patients with bipolar mania. J Affect Disord 111, 164–9 (2008).

25. Canetto, S. S. & Sakinofsky, I. The gender paradox in suicide. Suicide Life Threat Behav 28, 1–23 (1998).

26. Murphy, G. E. Why women are less likely than men to commit suicide. Compr Psychiatry 39, 165–75 (1998).

27. Akhtyamova, L. Named Entity Recognition in Spanish Biomedical Literature: Short Review and Bert Model. in Conference of Open Innovation Association, FRUCT vols 2020-April 3–9 (IEEE Computer Society, 2020).

28. Cotik Viviana and Rodríguez, H. and V. J. Spanish Named Entity Recognition in the Biomedical Domain. in Information Management and Big Data (ed. Lossio-Ventura Juan Antonio and Muñante, D. and A.-S. H.) 233–248 (Springer International Publishing, 2019).

29. Cai, N. et al. Minimal phenotyping yields genome-wide association signals of low specificity for major depression. Nat Genet 52, 437–447 (2020).

30. Insel, T. et al. Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. Am J Psychiatry 167, 748–51 (2010).

31. Bromet, E. J. et al. Diagnostic Shifts during the decade Following First Admission for Psychosis. American Journal of Psychiatry 168, 1186–1194 (2011).

32. Kalman, J. L. et al. Characterisation of age and polarity at onset in bipolar disorder. British Journal of Psychiatry 219, 659–669 (2021).

33. Perkins, D. O. et al. Polygenic risk score contribution to psychosis prediction in a target population of persons at clinical high risk. American Journal of Psychiatry 177, 155–163 (2020).

34. Musliner, K. L. et al. Polygenic risk and progression to bipolar or psychotic disorders among individuals diagnosed with unipolar depression in early life. American Journal of Psychiatry 177, 936–943 (2020).

35. Jonas, K. G. et al. Schizophrenia polygenic risk score and 20-year course of illness in psychotic disorders. Transl Psychiatry 9, (2019).

36. Krebs, M. D. et al. Associations between patterns in comorbid diagnostic trajectories of individuals with schizophrenia and etiological factors. Nat Commun 12, (2021).

CHAPTER 4

Transcending Diagnostic Categories: Unveiling Phenotypic Overlaps in SMI through EHR Data and Machine Learning

4.1 Introduction

Serious Mental Illness (SMI) consists, by definition, of disorders characterized by symptoms that significantly impair an individual's function. Such disorders are commonly classified according to distinct diagnostic categories. However, the overlap in symptoms between such categories and growing evidence that neither their causation nor their trajectories are fully distinct from one another (as described in Chapter Three, Figure 3.3B) suggest the importance of developing methods that can more accurately classify SMI.

Electronic Health Records (EHRs) offer a tool for implementing novel approaches for phenotyping SMI. As we and others have demonstrated, they provide large-scale, detailed, and longitudinal information, which can be used to generate detailed research phenotypes. Here, we extend our previous work, demonstrating the potential utility of EHR data for a dimensional phenotyping approach, considering disease risk as following a liability threshold model, occurring along a spectrum rather than as dichotomous.

In Chapter Three, we showed that diagnoses assigned based on ICD-codes recorded in EHRs were highly accurate compared with those assigned based on manual clinical review of these records. As an extension, we compare EHR-based diagnoses with those from structured interviews of a larger sample of individuals and identify factors that could lead to diagnostic discrepancies. Then, we investigate the potential of multidimensional EHR data, including symptoms, family history, substance use, and more, to reduce discrepancies between EHR and interview-based diagnoses. Lastly, we begin to explore the use of such dimensional information to generate probability-based diagnoses, raising the possibility that such an approach might be useful for classifying cases that do not fit simply into our current categorical diagnostic system. In this chapter, we present preliminary findings for this ongoing project.

4.2 Methods

4.2.1 Study Setting

For this study, we leveraged the EHR from the Clínica San Juan de Dios in Manizales (CSJDM). Demographic information, diagnoses, medications, healthcare usage, and clinical notes were extracted from the EHR database following approved protocols by IRBs at UCLA, UdeA, and CSJDM, as described in Chapter Two.

4.2.2 SCID Interviews

As part of the Paisa Project [1], potential study participants were identified by screening the CSJDM EHR for ICD codes, indicating one of the SMI diagnostic categories. Additional inclusion criteria and recruitment and assessment procedures are specified in [1]. Briefly, Participants were assessed using the online version of the Structured Clinical Interview for DSM-5 (NetSCID-5) [2], hereafter referred to as SCID, a semi-structured interview designed to elicit information from a patient about their symptoms and history. Between October 2017 and June 2022, 4,592 participants were interviewed at the CSJDM with the SCID and received one of the following diagnoses: major depressive disorder (MDD), bipolar disorder (BD; including type I, type II, unspecified bipolar disorder, and other specified bipolar disorder), schizophrenia (SCZ;

including schizophreniform disorder), and other psychoses (including schizoaffective disorder (SZA) and delusional disorder (DD)).

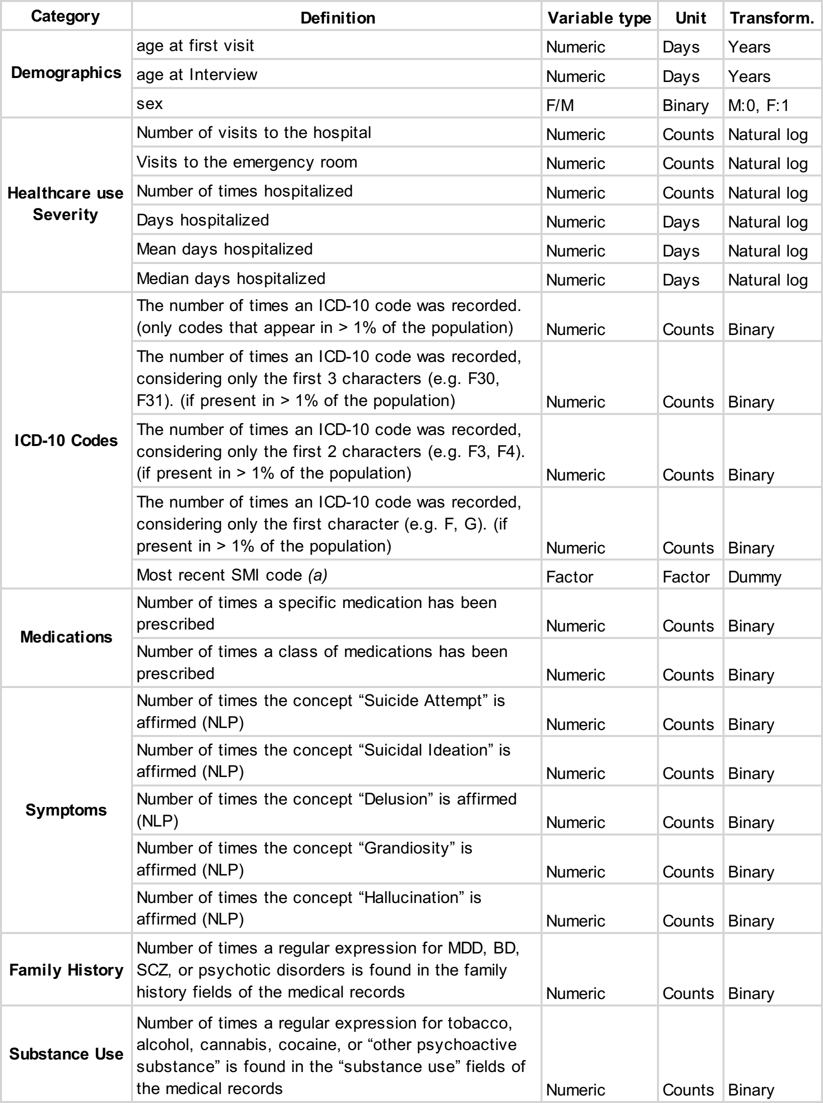
4.2.3 Definition of SMI Based on ICD-10 Codes

The tenth revision of the International Classification of Diseases[3] (ICD-10) is a system of codes used to classify and communicate medical diagnoses. Mental disorders are cataloged in the F chapter and grouped by categories that include mood disorders, anxiety disorders, psychotic disorders, substance use disorders, personality disorders, etc. We define Serious Mental Illness (SMI) using ICD-10 codes as in Chapter Three (section 3.2.2). 3,430 individuals had at least one of these SMI diagnoses on record before the day of their interview. For these patients, we extracted their longitudinal EHRs dating back to 2005 and up to the day of their assessment (n=37,038 visits).

4.2.4 Demographic and Clinical Features Extracted from the EHR

We extracted seven different types of information from the EHR database described in Chapter Two: demographics (sex, age at the first visits, and age at recruitment), diagnostic codes (ICD-10), clinical features extracted using the NLP pipeline described in Chapter Three (suicidal ideation or attempts, hallucinations, and delusions) [4], medications prescribed (name and class of medication), usage of hospital services (number and severity of visits.), family history of mental illness, and substance use history. The last two were extracted using regular expression matching. (Table 4.1).

Patient-level phenotypes were defined from the above list of variables in two different ways. First, counts for categorical variables were collected from each patient’s entire EHR history. Categorical variables included ICD-10 codes, medication class names, and NLP features. The presence of these features was summarized with binary variables. Second, quantitative variables, such as the length of stay during inpatient visits, were summarized using the mean, median, and total sum and were log-transformed. (Table 4.1).



Category	Definition	Variable type	Unit	Transform.
Demographics	age at first visit	Numeric	Days	Years
	age at Interview	Numeric	Days	Years
	sex	F/M	Binary	M:0, F:1
Healthcare use Severity	Number of visits to the hospital	Numeric	Counts	Natural log
	Visits to the emergency room	Numeric	Counts	Natural log
	Number of times hospitalized	Numeric	Counts	Natural log
	Days hospitalized	Numeric	Days	Natural log
	Mean days hospitalized	Numeric	Days	Natural log
	Median days hospitalized	Numeric	Days	Natural log
ICD-10 Codes	The number of times an ICD-10 code was recorded. (only codes that appear in > 1% of the population)	Numeric	Counts	Binary
	The number of times an ICD-10 code was recorded, considering only the first 3 characters (e.g. F30, F31). (if present in > 1% of the population)	Numeric	Counts	Binary
	The number of times an ICD-10 code was recorded, considering only the first 2 characters (e.g. F3, F4). (if present in > 1% of the population)	Numeric	Counts	Binary
	The number of times an ICD-10 code was recorded, considering only the first character (e.g. F, G). (if present in > 1% of the population)	Numeric	Counts	Binary
	Most recent SMI code *(a)*	Factor	Factor	Dummy
Medications	Number of times a specific medication has been prescribed	Numeric	Counts	Binary
	Number of times a class of medications has been prescribed	Numeric	Counts	Binary
Symptoms	Number of times the concept “Suicide Attempt” is affirmed (NLP)	Numeric	Counts	Binary
	Number of times the concept “Suicidal Ideation” is affirmed (NLP)	Numeric	Counts	Binary
	Number of times the concept “Delusion” is affirmed (NLP)	Numeric	Counts	Binary
	Number of times the concept “Grandiosity” is affirmed (NLP)	Numeric	Counts	Binary
	Number of times the concept “Hallucination” is affirmed (NLP)	Numeric	Counts	Binary
Family History	Number of times a regular expression for MDD, BD, SCZ, or psychotic disorders is found in the family history fields of the medical records	Numeric	Counts	Binary
Substance Use	Number of times a regular expression for tobacco, alcohol, cannabis, cocaine, or “other psychoactive substance” is found in the “substance use” fields of the medical records	Numeric	Counts	Binary

***Table 4.1** Demographic and clinical features extracted from the EHR. (a) variable used in the rule-based model*

4.2.5 A Rule-based Definition of EHR Diagnosis

The initial phase of this study involved evaluating the congruence between diagnoses from the EHR and those from the SCID. For this purpose, we employed two rule-based criteria for determining an EHR diagnosis [5,6]: A) the most recent SMI code in a patient's trajectory and B) the most common SMI code in the trajectory. We selected the most recent of these codes in the case of a tie. For each diagnosis definition, we evaluated the agreement between EHR and SCID using the following metrics: Cohen's kappa [7], Positive Predictive Value (PPV), Sensitivity, and F1. To identify features in the EHR that might account for discrepancies in diagnosis, we juxtaposed patients whose EHR and SCID diagnoses were identical with those where the EHR diagnosis matched, but the SCID diagnosis varied. This contrast was executed based on demographic and clinical parameters, utilizing z and t-tests.

4.2.6 A Machine Learning Definition of EHR Diagnosis

Next, we tested whether harnessing additional data from the EHR in a supervised Machine Learning (ML) framework could enhance the alignment between EHR and SCID diagnoses. For this, we restricted our analyses to the three more common SMI diagnoses (MDD, BD, and SCZ) and trained our model using 162 demographic and clinical features extracted from the EHR (Table 4.1). We evaluated four different machine learning models: random forest, XGBoost, elastic-net, and LASSO logistic regression. To fine-tune model parameters and ensure predictions for each participant, we used a 5-fold nested cross-validation. To gauge the effectiveness of the ML models, we used the same metrics employed for rule-based models, supplemented with the area under the precision-recall curve (AUPRC) and receiver operating characteristic (ROC) curve (AUROC). The estimated diagnosis probability for each patient was visually represented using ternary plots. To determine which features had the greatest impact on model predictions, especially in decision tree models, we used the SHAP package [8].

4.2.7 Interpretation of Diagnostic Probabilities

To evaluate the interpretability of diagnostic probabilities from our model, we proposed two experiments. First, we fit a new random forest without using any ICD-10 codes, obtained individual's predicted probabilities, and plotted them on the ternary plot. Then, we divided the ternary plot into regions of equal size and estimated the average value of the features within each region. This visual representation allowed us to investigate the spatial distribution of features based on their relation to diagnostic probabilities when ICD-10 codes were not a major contributor to the model.

Second, we focused on individuals who had more than ten hospital visits (N=1,233). From this subset, we categorized patients into two distinct groups: those showing an "improving" trajectory and those on a "worsening" path. This classification was made based on the number of inpatient visits during their first and last five visits. Specifically, those classified as "improving" had a minimum of two hospital admissions in their first five visits but none in their last five (N=58), whereas the "worsening" group had the inverse pattern – no inpatient visits in the first five visits and at least two in the last five (N=107). With these classifications in hand, we analyzed the shifts in the position of patients on the ternary plots, comparing their early visits to their more recent ones. This procedure allowed us to evaluate changes in diagnostic probability over time.

4.2.8 A Mood-Disorder Spectrum Defined from EHR Features

To investigate the spectrum of mood disorders. We aimed at developing a measure that could distinguish between the two ends of the spectrum: MDD and BD type I. Utilizing a total of 2,473 cases, we developed a binary classifier for these two SCID diagnoses. The methodology employed mirrored our prior machine-learning protocol. Specifically, we used a 5,5-cross-validation strategy to train our model on the selected cohort and the same number of features.

4.3 Results

4.3.1 SMI Patient Population

The cohort analyzed in this study consists of 3,430 patients. The distribution of SCID diagnoses was: 45% MDD, 43% BD, 9% SCZ, 3% SZA, and 0.1% (2 individuals) DD. As a description of this cohort, we summarize in Table 4.2 the demographic and clinical information extracted from the EHR (Table 4.1) and stratified by diagnosis and sex.

	SCID Diagnoses	MDD		BD		SCZ	
	Sex	Male	Female	Male	Female	Male	Female
	N	494	1041	484	1003	251	59
Demographics (mean (SD))	Age at interview	4.2(1.5)	4.3(1.5)	4.6(1.6)	4.7(1.5)	3.8(1.4)	4.7(1.4)
	Age at first visit	3.9(1.5)	3.9(1.4)	4.0(1.5)	4.0(1.4)	3.0(1.2)	3.8(1.3)
ICD-10	MDD	89	83	17	25	4	8
	BD	18	27	88	91	13	34
	SCZ	2	0	7	2	97	95
	SZA	0	0	5	2	5	12
	DD	0	0	0	1	2	5
Symptoms	Suicide attempt	41	34	22	26	12	17
	Suicide idea tion	75	60	46	47	31	36
	delusions	22	19	65	57	93	86
	hallucinations	31	26	56	55	89	90
	grandiosity	1	0	34	20	32	19
Medication Class	Antidepressant	74	71	39	51	29	31
	Antipsychotics	22	19	61	55	88	88
	Benzodiazepines	7	11	12	16	7	10
	Hypnotics	67	58	68	73	67	75
	Mood Stabilizer	29	36	81	83	49	64
	Hyperthyroidism	0	0	0	0	0	0
	Hypothyroidism	2	6	9	20	6	15
Hospital Usage	ER visits	76	70	80	76	75	73
	Inpatient visits	79	73	85	83	89	80
Family History	BD	5	10	18	22	6	5
	MDD	21	25	15	19	8	5
	SCZ	3	4	5	4	12	10
	Psych	0	0	1	1	2	2
Substance Use	Tobacco	49	36	55	44	55	42
	Alcohol	40	27	49	34	35	31
	Cannabis	14	7	27	9	42	7
	Cocaine	9	4	16	5	20	5
	Others	48	42	60	49	66	51

Table 4.2 *Demographic and clinical summary of the study population. Summaries are presented by sex and diagnosis, for the three most frequent SMI diagnoses: MDD, BD and SCZ. Age is in decades; other values are in percentages. Individuals with other psychoses are not included (N=98).*

4.3.2 High Agreement between ICD-10 Code and Diagnosis from SCID Interview

We observed very high agreement between the SCID diagnoses and those obtained from either of the two rule-based definitions of EHR diagnosis. The first definition (most recent SMI code) reached an overall kappa value of 0.70 (95% CI: 0.68-0.72). The second definition (most common SMI code) had a similarly high kappa value of 0.68. Below, we detail the results from the first definition.

Diagnoses of BD and MDD have a “very high” [9,10] agreement between SCID and EHR (kappa values are 0.68, 95% CI: 0.65-0.70, and 0.71, 95% CI 0.68-0.73, respectively). At the same time, diagnoses of SCZ have a "near perfect" [9,10] agreement (kappa = 0.84, 95% CI 0.81-0.87). Positive Predictive Values for these diagnoses are similarly high. On this metric, the EHR-based diagnosis of MDD ranks highest with a PPV of 0.88 (95% CI 0.87-0.89), followed by BD (PPV =0.78, 95% CI: 0.76-0.79), SCZ (PPV =0.79, 95% CI 0.77-0.80). SZA has lower agreement scores, likely due to its comparatively smaller sample size (kappa = 0.19, 95% CI: 0.10-0.29, and PPV = 0.54, 95% CI: 0.53-0.56). For DD, kappa is -0.001 and PPV 0.

Table 4.3 shows the confusion matrix comparing these two sets of diagnoses. The largest disagreements between the diagnostic approaches were observed in patients with mood disorder diagnoses, including those with an ICD-10 code of BD and a SCID diagnosis of MDD (n=317), and those with an ICD-10 code of MDD and a SCID diagnosis of BD (n=156).

		Last SMI ICD-10 codes					**Mode SMI ICD-10 codes**				
		MDD	BD	SCZ	SZA	DD	MDD	BD	SCZ	SZA	DD
SCID Diagnosis	MDD	**1207**	*317*	9	1	1	**1232**	*295*	6	1	1
	BD	*156*	**1308**	15	7	1	*201*	**1245**	25	13	3
	SCZ	2	18	**288**	2	0	3	22	**279**	2	4
	SZA	4	34	46	**12**	0	8	33	44	**11**	0
	DD	1	0	1	0	0	1	0	0	0	**1**

Table 4.3 *Confusion matrix of the comparison between SCID and EHR-based diagnoses of SMI.*

4.3.3 Features Associated with Diagnostic Disagreements

We aimed to understand the features that are associated with discrepancies between the two approaches for diagnosing mood disorders by conducting a comparison of two scenarios (Table 4.4). In the first scenario, from the group of patients diagnosed with MDD via EHR, we contrasted those who received a SCID diagnosis of MDD with those who received a SCID diagnosis of BD. In the second scenario, among patients diagnosed with BD through EHR, we compared those who received a SCID diagnosis of MDD with those who received a SCID diagnosis of BD.

Individuals in the first scenario differed in some key characteristics. Those with the BD diagnosis (SCID) were twice as likely to have received at least one BD ICD-10 code in their records than those with the MDD diagnosis (SCID) (9.6% vs. 4.3%, respectively). They were also twice as likely to have used mood stabilizers (44.2% and 22.1%, respectively). Individuals in the second scenario also showed some notable differences. Those with an MDD diagnosis (SCID) were more than twice as likely to have received at least one MDD ICD-10 code in their EHR when compared with those with a BD diagnosis (SCID) (30.91% vs. 13.38%, respectively). On the other hand, they were more than eight times as likely to have received at least one ICD-10 code for a manic or hypomanic episode (52.0% vs. 6.3%, respectively) compared to those with a BD diagnosis (SCID). In this scenario, both groups had a high frequency of use of mood stabilizers (86.93% and 77.29%, respectively). Finally, in both scenarios, the interval between the most recent visit to the hospital and the SCID interview was greater for individuals with mismatches between interview and EHR diagnoses than for individuals in whom these diagnoses matched (17.8 and 11.7 vs. 9.2 and 9.2, respectively).

EHR	MDD			BD			
	N= 1370			N= 1680			
SCID	MDD	BD		MDD	BD		
	1207	156	p-value *	317	1308	p-value *	p-value **
BD + MDD (%)	4.3	9.6	0.0039	30.9	13.4	6.77E-14	5.76E-43
Hypomania or Mania (%)	0.3	1.9	0.0089	6.3	52.0	<2e-16	<2e-16
Mood Stabilizers (%)	22.1	44.2	1.64E-09	77.3	86.9	1.58E-05	<2e-16
Months since most recent visit	9.16 (17.32)	17.82 (22.28)	1.78E-08	11.69 (20.44)	9.28 (17.15)	0.03	4.32E-08

Table 4.4 *Informative features in the disagreements between SCID and EHR-based diagnoses for MDD and BD. BD + MDD refers to individuals who had both diagnostic codes on their record. * z-test for % and t-test for number of months. ** chi2 for proportions and ANOVA for months.*

4.3.4 Machine Learning Models

Next, we tested whether machine learning models improve the alignment between EHR and SCID compared to the rule-based algorithms. Random forest achieved the highest accuracy, with an overall kappa of 0.69; 95% CI 0.67-0.71). However, the other three models (XGBoost, LASSO, and elastic-net logistic regression) performed comparatively well, with kappa values ranging from 0.66 to 0.68. Two additional benefits of random forest are its efficient runtime and its ability to leverage non-linear relationships. Therefore, we used this model for downstream analyses. Random forest reached an AUPRC of 0.86, 0.89, 0.95 and an AUROC of 0.91, 0.90, 0.99 for MDD, BD, and SCZ respectively (Figures 4.1A and 4.1B), and diagnosis-specific kappa values of 0.71, 0.67 and 0.83 for MDD, BD, and SCZ, respectively. Despite these high values, the random forest did not significantly improve the agreement observed in the rule-based model.

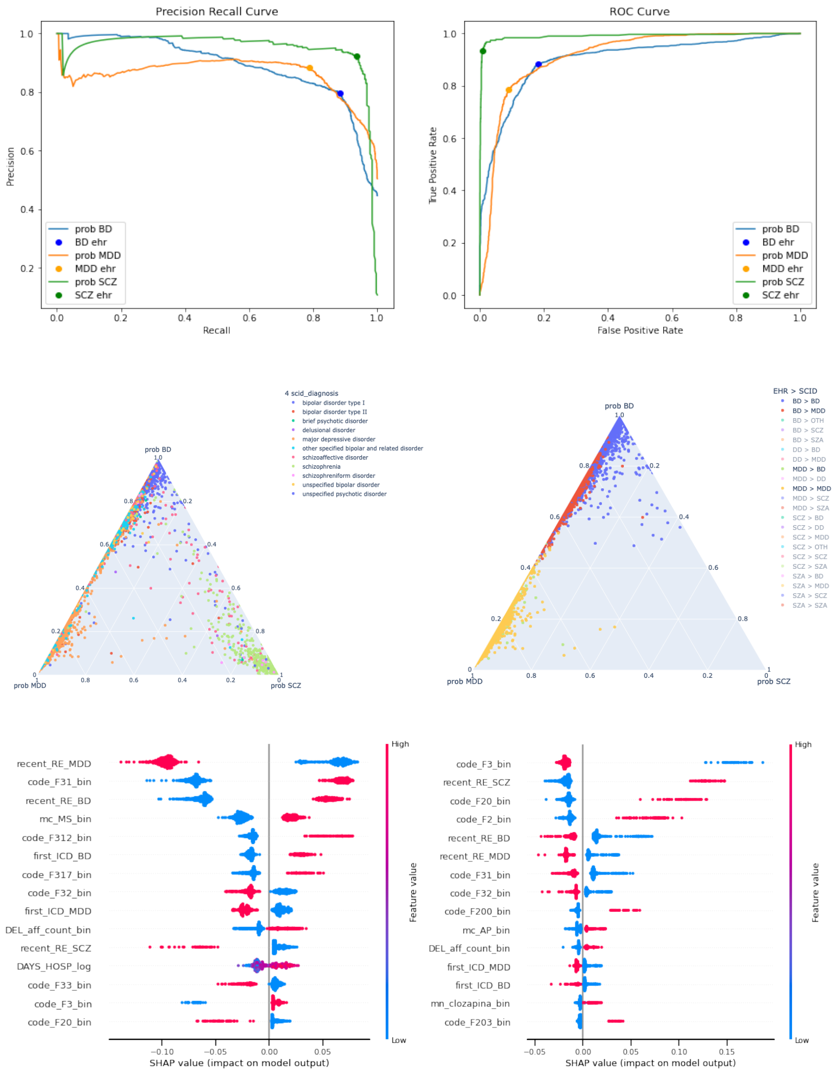


***Figure 4.1** Random Forest model for diagnosis of SMI based on 162 EHR-extracted features. A and B) Precision recall and ROC curves. The performance of the rule-based model (most recent SMI ICD-10 code) is shown by the dots. AUPRC: MDD: 0.86; BD: 0.891; SCZ: 0.95. AUROC: MDD: 0.913; BD: 0.90; SCZ: 0.991. C and D) probabilities for diagnosis of MDD, BD and SCZ. Dots represent 3-way probabilities for each patient. Individuals are colored in Figure 1C according to their SCID diagnosis and in Figure 1D to their combination of EHR-*

SCID diagnoses from Table 4. E and F) SHAP value plots of the 15 most important predictors for diagnoses of BD and SCZ, respectively.

The ternary plots in Figures 4.1C and 4.1D provide insight into the distinctiveness of different diagnostic categories. It is noteworthy that while SCZ patients are clearly differentiated from those with MDD and BD, the separation between the latter two groups is less apparent, consistent with the expectation that mood disorder diagnoses exist on a spectrum. [11–14]

Figures 4.1E and 4.1F show the SHAP values for the 15 most important features for the diagnoses of BD and SCZ in the RF model. As expected, the most recent SMI codes are among the most influential predictors. Also contributing to the final probabilities are other disorder and episode-specific ICD-10 codes, as well as specific medications (e.g., antidepressants and mood stabilizers) and the presence of clinical features, in particular, delusions. Additionally, the SHAP value plots show that the top episode-specific ICD-10 predictive of BD is F312 (i.e., manic episode with psychotic features), suggesting that, as would be expected, BD type I lies at the extreme of the probability distribution for this diagnosis.

4.3.5 Interpretation of Diagnostic Probabilities

We ran two experiments to evaluate the relationship between specific clinical features and the probability space of SMI diagnoses. Our findings align with our preliminary hypotheses. For example, hallucinations are frequent in regions with a high probability of SCZ; delusions are frequent along the spectrum between BD and SCZ but are less frequent near MDD; the use of mood stabilizers is indicative of a high probability for BD. Interestingly, features like suicidal ideation and suicide attempt, the use of antidepressants, and the comorbidity with neurotic and stress-related disorders (F4) are more characteristic of intermediate rather than extreme diagnostic probabilities in the three-way distribution – they don't overlap substantially with any diagnosis. Our second experiment revealed that a "worsening" trajectory is associated with an increased probability of a BD or SCZ diagnosis.

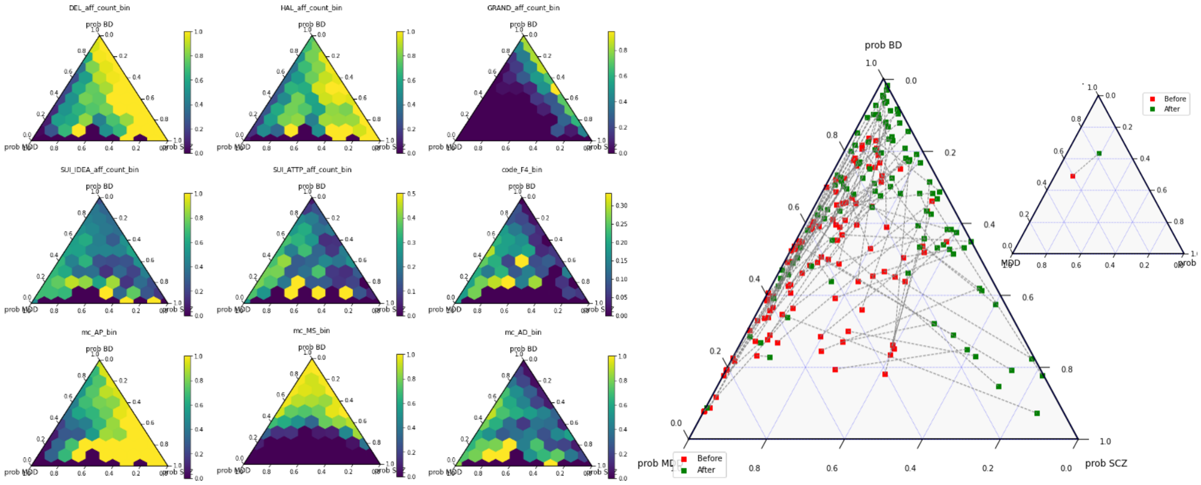


Figure 4.2 *Ternary plots interpreting diagnostic probabilities. A) shows the frequency of nine features across the three-way probability space when ICD-10 codes are not part of the prediction model. In order, the features are delusions, hallucinations, grandiosity, suicidal ideation, suicide attempt, ICD-10 codes of neurotic, stress-related and somatoform disorders (F4), use of antipsychotics, use of mood stabilizers, and use of antidepressants. B) positions of "worsening" individuals based on their EHR features in their first (red) and last (green) five visits. They had no hospitalization in the first five visits and two or more in the last five visits (N=107). Small ternary plot shows centroids for the first and last five visits.*

4.3.6 A Mood-Disorder Spectrum Defined from EHR Features

We fit a binary classifier (separating BD type I from MDD) to obtain a better resolution of the mood disorder spectrum and gain additional insights into transdiagnostic phenotypic overlaps. Under this binary classification, BD type II (N=350) and other specified BD (N=168) displayed bimodal distributions (Figure 4.3) with a point of density in the intermediate region between MDD and BD type I and another close to MDD. This latter group could potentially be composed by patients who are more likely to seek help during depressive episodes than during hypomanic ones.[15,16]

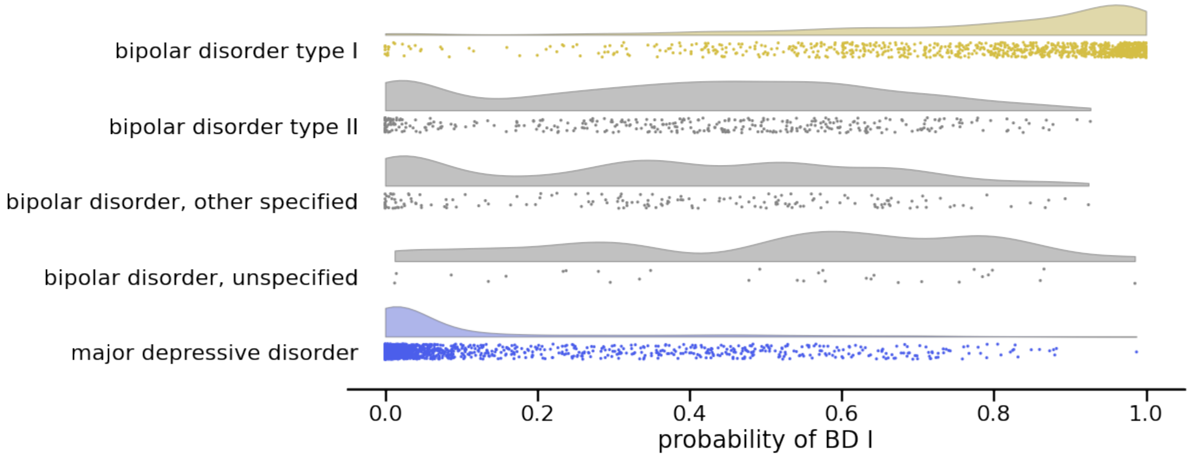


Figure 4.3 *Density plot of the probability distribution of the binary classifier for BD type I and MDD, stratified by SCID diagnosis: BD type II (N=350), other specified BD (N=168), and unspecified BD (N=31).*

4.4 Discussion

In this study, we characterized the concordance between diagnoses of SMI derived from EHRs and those obtained through structured diagnostic interviews. Two distinct phenotyping strategies were employed: rule-based, harnessing only ICD-10 codes, and machine learning-based, integrating extensive longitudinally EHR data. Both exhibited robust agreement with the SCID for the three more common diagnoses of MDD, BD, and SCZ. Agreement for diagnoses with smaller samples, such as SZA and DD, was limited. We, however, don't consider these results a direct validation for EHR diagnoses, as the patient selection was based on their EHR, and interviewers were privy to these records. As anticipated, some level of discrepancies between diagnoses was present. At the same time, a close evaluation of the longitudinal EHR data revealed evidence of phenotypic overlap between traditional categorical diagnoses, potentially accounting for some of these discrepancies.

Machine learning models did not outperform the rule-based baseline regarding the alignment of EHR and SCID. However, these approaches enabled us to define continuous phenotypic scores for SMI, suggesting that further development of these models may be a useful strategy for constructing dimensional SMI phenotypes based on the assignment of probabilistic diagnostic profiles to each patient. Using this continuous phenotypic space, we could describe the distribution of clinical features, the course of illness of patients with worsening outcomes, and the clustering of specific SCID diagnoses in distinct regions of this space. For instance, while BD type I is nestled closer to the top of the BD distribution, BD type II sprawls out more, sharing a heightened MDD probability.

This study is a first step in probing these probabilities as a more holistic portrayal of a patient's condition. We anticipate that this strategy may be particularly valuable in enabling transdiagnostic genetic research. Currently, we are evaluating the robustness of continuous EHR-driven phenotypes under different modeling designs and their behavior when used across hospitals. Ultimately, we aim to integrate these phenotypes into upcoming genetic studies.

4.4.1 Limitations

One significant limitation of our study stems from the fact that the SCID interviews were conducted by raters with access to EHRs, introducing the possibility of confirmation bias. This prior knowledge of a patient's diagnosis could influence the outcomes of the structured diagnostic interviews. Additionally, some steps for feature selection in our machine learning models depend on the frequency of the features themselves, potentially compromising the generalizability of the models.

4.5 Conclusion

Using rule-based algorithms, we extracted SMI diagnoses from thousands of patients' EHRs with very high precision. While integrating additional EHR data and using machine learning models doesn't significantly improve diagnostic accuracy, these models provide valuable quantitative insights into the phenotypic overlap between psychiatric diagnoses.

4.6 References

1. Service, S. K. et al. Distinct and shared contributions of diagnosis and symptom domains to cognitive performance in severe mental illness in the Paisa population: a case-control study. Lancet Psychiatry 7, 411–430 (2020).

2. Brodey, B. B. et al. Validation of the NetSCID: An automated web-based adaptive version of the SCID. Compr Psychiatry 66, 67–70 (2016).

3. World Health Organization. The International Statistical Classification of Diseases and Health Related Problems ICD-10: Tenth Revision. vol. 2 (2004).

4. De la Hoz, J. F. et al. Electronic health records reveal transdiagnostic clinical features and diverse trajectories of serious mental illness. medRxiv (2023) doi:10.1101/2022.08.20.22279007.

5. Sara, G. et al. Comparing algorithms for deriving psychosis diagnoses from longitudinal administrative clinical records. Soc Psychiatry Psychiatr Epidemiol 49, 1729–1737 (2014).

6. Davis, K. A. S. et al. Using data linkage to electronic patient records to assess the validity of selected mental health diagnoses in English Hospital Episode Statistics (HES). PLoS One 13, (2018).

7. Cohen, J. A Coefficient of Agreement for Nominal Scales. Educ Psychol Meas 20, 37–46 (1960).

8. Lundberg, S. M. et al. From local explanations to global understanding with explainable AI for trees. Nat Mach Intell 2, 56–67 (2020).

9. Regier, D. A. et al. DSM-5 field trials in the United States and Canada, part II: Test-retest reliability of selected categorical diagnoses. American Journal of Psychiatry 170, 59–70 (2013).

10. Kraemer, H. C., Kupfer, D. J., Clarke, D. E., Narrow, W. E. & Regier, D. DSM-5: How reliable is reliable enough? American Journal of Psychiatry 169, 13–15 (2012).

11. Angst, J. & Cassano, G. The mood spectrum: Improving the diagnosis of bipolar disoder. Bipolar Disorders, Supplement 7, 4–12 (2005).

12. Benvenuti, A. et al. Mood Spectrum Model: Evidence reconsidered in the light of DSM-5. World J Psychiatry 5, 126 (2015).

13. Coleman, J. R. I. et al. The Genetics of the Mood Disorder Spectrum: Genome-wide Association Analyses of More Than 185,000 Cases and 439,000 Controls. Biol Psychiatry 88, 169–184 (2020).

14. Whitton, A. E. et al. Mapping Disease Course Across the Mood Disorder Spectrum Through a Research Domain Criteria Framework. Biol Psychiatry Cogn Neurosci Neuroimaging 6, 706–715 (2021).

15. Benazzi, F. A prediction rule for diagnosing hypomania. Prog Neuropsychopharmacol Biol Psychiatry 33, 317–322 (2009).

16. Drancourt, N. et al. Duration of untreated bipolar disorder: Missed opportunities on the long road to optimal treatment. Acta Psychiatr Scand 127, 136–144 (2013).

CHAPTER 5

Extraction of Symptom-Level Phenotypes from Clinical Notes

5.1 Introduction

Clinical notes offer a rich repository of patient information not captured by structured data alone [1,2.] Still, most research involving Electronic Health Records (EHRs) leans heavily on structured data, particularly ICD codes [3,4]. Such data aim to standardize the representation of disorders, treatments, and procedures, bolstering interoperability for administrative tasks. In contrast, free-text clinical notes favor detailed descriptions of individual symptoms, their combinations and severity, family history, and environmental influences.[2] Although such information could be invaluable for phenotyping studies of serious mental illness (SMI), the unstructured nature of such data poses challenges for efficient, high-throughput phenotyping [1,3,5].

Several studies have shown the efficacy of clinical Natural Language Processing (cNLP) for detailed phenotyping of psychiatric disorders in English-language clinical notes [6–8]. For example, Jackson et al. (2017) [7] employed the Clinical Record Interactive Search (CRIS-CODE) system to extract SMI symptoms from clinical notes, noting symptom overlap across multiple diagnoses. Similarly, McCoy et al. (2018) [8] developed an NLP method to phenotype five domains defined by a widely used research framework (Research Domain Criteria, RDoC [9]) from clinical notes. These findings underscore the potential of cNLP in enhancing our understanding of psychiatric disorders beyond conventional diagnostic categories. Yet these approaches have had limited application in either Spanish-language contexts or in hospitals in low- and middle-income countries (LMIC), providing a rationale for the work reported in this chapter.

This chapter describes the initial stages of development of an NLP algorithm tailored to extract psychiatric phenotypes from Spanish clinical notes. We based our approach on MedspaCy, a versatile cNLP tool that combines rule-based and traditional machine learning methods [10]. Rule-based methods stand out for their efficacy even with limited labeled data—a common limitation of clinical datasets. These methods are interpretable and adaptable to unique text nuances, especially for tasks like Named Entity Recognition (NER) or contextual analysis. Data for this project come from the Hospital Mental de Antioquia (HOMO), a public psychiatric institution that serves the entire population of Medellín, Colombia. Here, we present our methodological approach, focused on defining an extensive list of EHR-based phenotypes and developing and validating the cNLP algorithm designed for their extraction.

5.2 Methods

5.2.1 Hospital Medical Records

As part of the phenotyping for Misión Origen, we accessed psychiatrists' clinical notes from the EHR at the HOMO mental hospital in Bello, Antioquia. This access followed protocols approved by the IRBs of UCLA, UdeA, and HOMO. Clinical notes at HOMO are categorized into four areas, each further subdivided into smaller sections:

1. Physical Exam and Mental Status Exam (MSE).

    a. Physical Exam sections encompass the abdomen, head, neck, throat, ears, nose, skin, and specific systems such as cardiovascular, genitourinary, respiratory, neurologic, etc.
    b. MSE sections include affect, intelligence, introspection, judgment and reasoning, motor activity, language, memory, alertness, orientation, thought content and processes, demeanor and attitude, sensory perception, and other positive findings.
2. Risk Scales. These scales contain sections relating to multiple topics, including aggression, risk of falling, absconding, suicide, non-adherence to treatment, pressure ulcers, and misunderstandings of the provided information.
3. Patient History. This area has sections related to:
    a. Medical history includes cancer, diabetes, neurologic conditions, respiratory conditions, cardiovascular conditions, surgeries, etc.
    b. Family history includes psychiatric diagnoses, epilepsy, diabetes, cancer, etc.
    c. Social history includes substance and tobacco use, diet, and exercise.
4. Visit-specific sections. These sections include the chief complaint, a description of the current illness, an overall analysis, and a treatment plan.

5.2.2 Selection of Concepts to be Extracted

We started by designing an initial list of 82 concepts to be extracted from the EHR (Table 5.1). This list was a combination of concepts from the literature, e.g., from CRIS-CODE [7], concepts identified in a set of 20 randomly selected notes from the hospital's EHR, and other additional concepts of interest to our research group. Some of these are associated with psychiatric comorbidities, such as binge eating and compulsions, while others are related to observational features like poor personal hygiene and lack of visual contact.

To generate a comprehensive list of concepts, we allowed the inclusion of 48 new concepts during the early stages of the annotation. The complete list of 130 concepts and the phase of the study in which they were added is detailed in Table 5.1. We included the label "Other" to capture relevant but infrequent concepts. Finally, we annotated five labels for cues used by the ConText algorithm.

Lastly, we proposed a concept hierarchy to aggregate concepts in downstream analyses. (Table 5.2). Based on this hierarchy, a less-specific concept is considered present if its constituent (more-specific) concepts are marked as present. For example, the concept of "*hallucinations*" (less-specific) is present if the concept of "*visual hallucinations*" (more-specific) is present.

5.2.3 Clinical Note Sections to Annotate

As previously described, clinical notes are segmented into multiple sections, each serving a distinct purpose. Our objective was to select those sections that include the densest concentration of pertinent concepts for annotation. For this, we executed a targeted search using a regular expression for each of the initial 82 concepts across all sections. Table 5.3 displays the outcome of this search, expressed as the percentage of concept matches per section; we opted to annotate the Chief Complaint, Affect, Analysis, Thought Content, and Sensory Perception sections, all of which display high concentrations of concepts.

5.2.4 Selection of Text Entries for Annotation

From all the clinical notes in the EHR database (N=2,651,447), we focused on those associated with SMI ICD-10 codes as defined in Chapter Three (section 3.2.2), resulting in 899,624 notes. Subsequently, the five sections mentioned above were isolated. Entries with a length below 30 characters were discarded, resulting in the following: 219,668 for Chief Complaint, 18,000 for affect, 803,111 for Analysis, 54,832 for Thought Content, and 13,059 for Sensory Perception. 400 entries were randomly sampled from each section, totaling 2,000, hereafter referred to as "documents."

5.2.5 Annotation Process

Annotations were made by two clinicians—an MD and a clinical psychologist—using WebAnno (version 3.6.5), as recommended by [11]. They independently annotated 2,000 documents with 136 labels described in 5.2.2.

5.2.6 Annotator Training Protocol

The training began with a group session where we reviewed 20 purposefully chosen documents. This session led to identifying 14 more concepts, enriching the initial list. Review sessions were scheduled after 50, 100, and 500 documents. These sessions addressed ambiguities encountered during annotations and clarified concept definitions. An additional 34 concepts were incorporated into the list based on reviewer feedback. Each added concept was substantiated and deliberated upon with three senior psychiatrists in the team – each with over 25 years of clinical and research experience.

5.2.7 Contextual Signifiers with the ConText Algorithm

The ConText algorithm [12] is used to identify negation, experiencer, and temporality modifiers around concepts found in clinical notes. Specific cues help in ascertaining this context, including terms signaling negation ("no evidence of…"), uncertainty ("suspicious for…"), hypothetical situations ("at risk of…"), historical events ("history of…"), or references to third parties ("the father…").

The default implementation of the ConText algorithm uses cues in English. To integrate it with HOMO's notes, annotators tagged these cues within the annotated document set. For an event to be classified as 'historical,' it had to predate the note by at least one year or specify a time before the current episode. The annotated cues informed the development of search patterns for using ConText. This refinement ensured that the identified concepts reflected the patient's condition at the time of documentation.

5.2.8 Inter-Annotator Agreement Metrics

Following the annotation of the 2000 documents, we refined the list of concepts to include only those frequently and consistently annotated. To start, we excluded concepts with fewer than ten instances, as indicated by either of the two annotators. Then, we calculated the inter-annotator agreement for each concept utilizing Cohen's kappa [13]. Concepts with kappa values below 0.6 were removed.

5.2.9 Selection of Documents for Development and Test Sets

We aimed to split the annotated documents into development and test sets in a way that balanced the number of examples of each concept between both sets. A simple random split of the 2000-document dataset risked having some concepts missing from either test set. Therefore, we set the goal that 20-50% of the examples of each concept were found in the test set. For this purpose, we implemented the following three-step method as described below:

1. Merge annotations from both annotators. If either of the annotators identified a concept in a document, the document was counted as having the concept.
2. Flag documents for inclusion in the test set. For each concept, 20% of documents containing it were randomly flagged for inclusion in the test set with a maximum cap of five documents per concept (i.e., documents with 25 documents or more could flag only 5). A single document could be flagged multiple times by different concepts.
3. Remove flags from documents. If more than 50% of a concept's documents were in the test set, we eliminated the flags from enough of them to bring this proportion down to 50%. This step was done one concept at a time, starting from the least frequent one. After 11 iterations, all concepts had 20-50% representation in the test set.

The resulting testing set contained 309 documents, leaving 1,691 for the development set.

5.2.10 Generation of a Gold Standard from the Test Set

For an optimal test of the algorithm, both annotators jointly reviewed the 309 documents and resolved any annotation discrepancies in consensus. This process involved merging their annotations, identifying new concepts that were initially missing, and removing conflicting annotations. From here on, we refer to this set of documents as the 'Gold Standard' (GS).

5.2.11 Patient Selection for Chart Review

Since 2020, our group has recruited SMI patients from HOMO for several genetic studies. Out of 4,608 SMI patients recruited by June 30th, 2023, we picked 120, ensuring an equal representation of the three major SMI diagnoses – 40 each of MDD, BD, and SCZ. The diagnostic label was sourced from each patient's recruitment form.

For each patient, we singled out a clinical note for evaluation, focusing on the most recent one. To have this note reflect heightened symptom severity, we selected it in this order: 1st) most recent hospital intake note if ever hospitalized. 2nd) If not hospitalized, see the latest emergency department note. 3rd) If neither was available, the most recent outpatient note.

5.2.12 Annotation of Patient Charts

Both annotators reviewed each note independently, labeling the 130 concepts with attributes: A) Unmentioned. B) Ambiguous presence. C) Confirmed absence. D) Mild/moderate intensity. E) Severe intensity. F) Chief complaint. Annotations were done at the level of the complete note. Afterward, we grouped the annotations to label concepts in a note as 'absent' (A-C) or 'present' (D-F). Cohen's kappa was used to measure the inter-annotator agreement from these binary categories. Lastly, to test the NLP algorithm, we made a consensus annotation where both annotators solved inconsistencies. Furthermore, we expanded the list of attributes to ten, including G) Recent improvement. H) Basis on diagnosis. I) Historical event. J) Hypothetical event. For later testing, G, H, and I were considered 'present' and J absent.

5.2.13 NLP Pipeline Overview

All analyses were done in Python (v3.8.6), and data were handled using pandas (v1.2.1). The core of our NLP pipeline was built using spaCy (v3.5.4) and its clinical extension, medspaCy (v1.1.2) [10]. spaCy is a framework for constructing high-performance NLP pipelines that combine machine-learning and rule-based methods and has a large ecosystem of users and packages.

medspaCy is a package in the spaCy ecosystem that provides an array of components for cNLP operations, such as NER and context analysis. It processes clinical text using four sets of rules: tokenizer, sentencizer (RuSH [14]), NER, and ConText. Given our work's emphasis on Spanish clinical notes, we had to modify the sentencizer and ConText rules. Our complete pipeline uses the following components: tok2vec, morphologizer, attribute_ruler, lemmatizer, medspacy_pyrush, medspacy_target_matcher, medspacy_context.

5.2.14 Rule Generation Process

1. Sentencizer Rules: using the RuSH algorithm [14], we crafted rules to identify sentence start and end points in clinical text. Spanish rules were formulated based on the English-language ones and refined using our development-set documents.
2. NER Rules: We manually constructed search patterns for each concept based on the development set. These patterns identified different word forms, including typos, spacing, punctuations, and wildcards. Each concept's patterns were stored in an individual JSON file.
3. ConText Rules: The development set annotations informed the creation of 103 unique ConText Rule patterns. These replaced the standard 97 English-language patterns. We evaluated our new set against specific examples from the development set.

5.2.15 Document-Level Evaluation

To assess the performance of the NER patterns in identifying concepts within individual documents, we tested them on the 309 documents of the GS. Precision, recall, and F1 scores were used to assess performance. We established a performance benchmark: concepts achieving an F1 score of 0.80 or above were deemed high-performing.

5.2.16 Patient-Level Evaluation

At this level, we aimed to determine if a particular concept had been documented concerning any patient. This evaluation required applying the complete pipeline, from concept detection using NER to verifying context via ConText to integrating supplemental structured data. The latter comes from a set of checklists inaccessible to the NLP algorithm.

5.2.17 Ethical Approval

This study was executed in alignment with protocols approved by the IRBs of UCLA, UdeA, and HOMO.

5.3 Results

We present results at the document level first (see 5.2.15) and then at the patient level (see 5.2.16).

5.3.1 Document-level Annotations and Inter-Annotator Concordance

Over six weeks, 11,900 labeled text segments were annotated across 2,000 documents (Figure 5.1 A). Out of the 130 concepts targeted for annotation, seven were never identified (echolalia, stupor, waxy flexibility, binge eating, avolition, excessive exercise, and specific purging behaviors), 24 were labeled less than ten times between both annotators (Figure 5.1 B), and out of the remaining 99, 18 didn't reach the specified level of inter-annotator concordance of kappa >= 0.6. (Figure 5.2 A; For more details on these results, see Table 5.4).

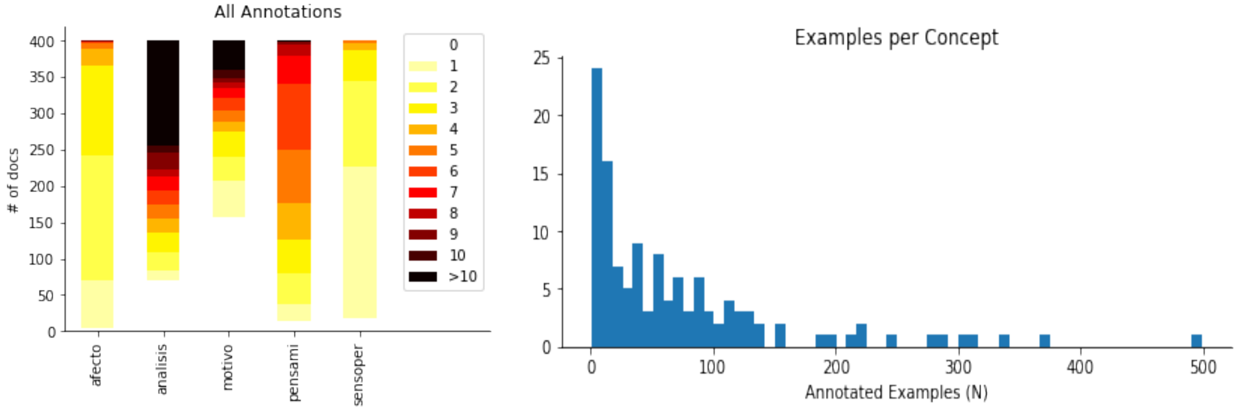


__Figure 5.1__ A) Density of concepts per document in the five sections of origin. B) Distribution of concept occurrences (examples) in the 2000 documents annotated.

5.3.2 Patient-level Annotations and Inter-Annotator Concordance

Over four weeks, 120 patient charts were annotated independently and in consensus. Out of the 130 concepts targeted for annotation, nine were never observed (stupor, waxy flexibility, binge eating, avolition, excessive exercise, emotional withdrawal, apathy, altered prospection, and nightmares), 13 were found in fewer than three patient charts, and of the remaining 108 concepts, 17 did not achieve the predetermined kappa value of 0.60, leading to their exclusion from later evaluations (Figure 5.2B). An in-depth breakdown is provided in Table 5.4.

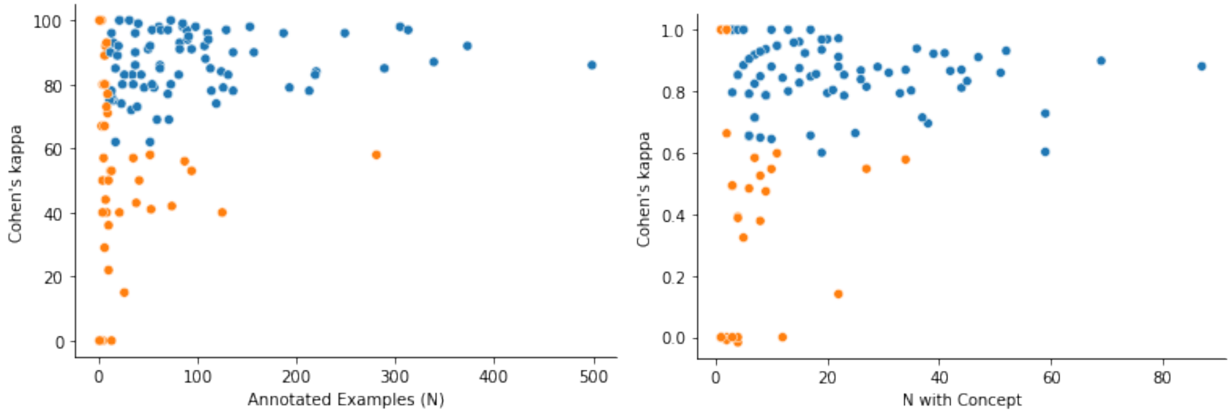


Figure 5.2 *Illustration of the concurrent distribution of concept frequency and kappa values across the sentence and patient datasets. Blue dots represent concepts with high-quality annotations.*

5.3.3 Frequency and Agreement of Concepts Passing QC

69 concepts passed both frequency and inter-annotator agreement thresholds in document- and patient-level evaluations. Their average Cohen's kappa was 0.87 at the document level and 0.83 at the patient level. These concepts were found on average in 33.2 of the 309 documents and 20.5 of the 120 patient charts. The most recurrent concepts at the document level were delusions (n=144), aggressivity (n=135), psychomotor agitation (n=90), thoughts of death (n=86), and general psychotic symptoms (n=80). Conversely, within the patient charts, the most prevalent concepts were denoted as poor insight (n=87), poor judgment (n=66), substance use (n=60), poor treatment adherence (n=55), and general depressive symptoms (n=51). A more detailed breakdown of these statistics can be found in **Table 5.3**.

5.3.4 Overview of the Document Gold Standard

The GS comprised 309 documents, distributed over different sections: 53 from the Chief Complaint, 52 from the affect, 115 from the Analysis, 63 from the Thought Content, and 26 from the Sensory Perception. Even though the individual concepts appeared less frequently in the GS than in the development set, the sampling design intentionally enriched the GS with concepts. As a result, the average number of concepts present per document in the GS surpassed that of the development set, with 7.4 and 2.9, respectively. (Figure 5.3)

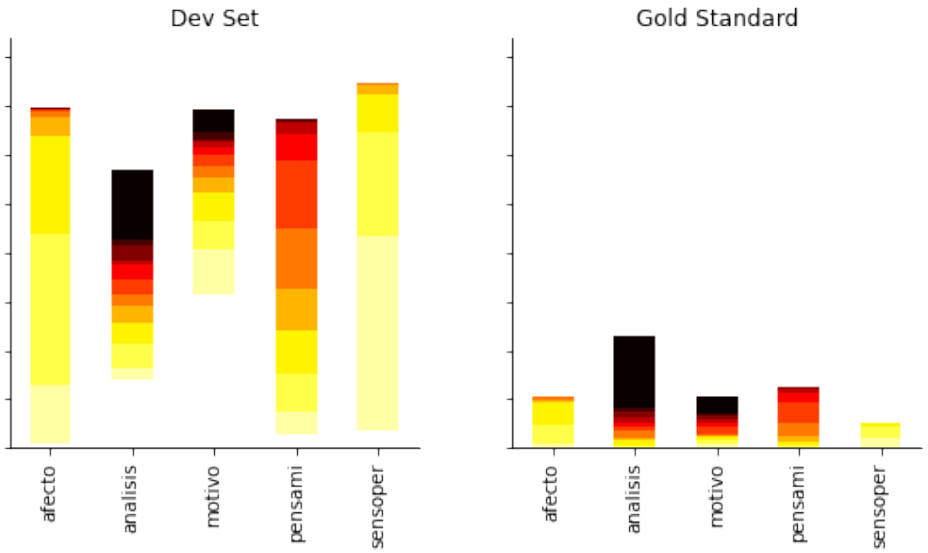


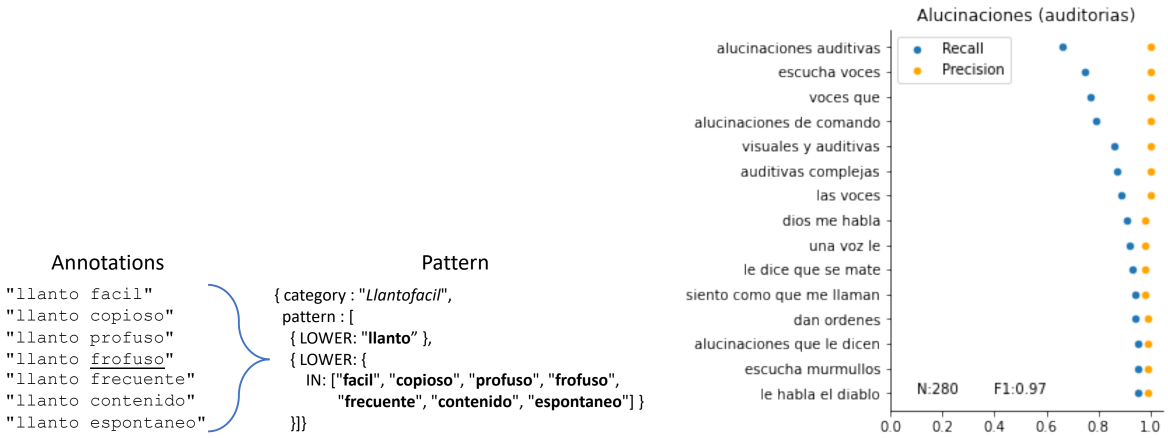
Figure 5.3 Concept density per document across the five origination sections, further differentiated by development set and the gold standard.

5.3.5 Overview of the Patient Chart Test Set

The mean number of observed concepts was 25.8 Per patient. Observations from this dataset include that “sensory perception” was frequently documented but often negated, rendering it non-relevant in 65 of its 71 occurrences. The concept of “cognitive decline” was most commonly reported in an uncertain context (13/19). Additionally, “aggressivity” was frequently noted in a hypothetical context (n=18/62).

5.3.6 Search Patterns Generated for Sentencizer, NER, and ConText

We formulated 56 rules for the Sentencizer (RuSH) component of medspaCy, 324 for the NER component (from 2,348 distinct phrases labeled in the development set. Figure 5.4), and 103 for the ConText component. The number of patterns within each ConText category was 24 for negated existence, 27 for possible existence, 18 for hypothetical context, 24 for historical occurrences, and nine for individuals other than the patient.



```
"llanto facil"
"llanto copioso"
"llanto profuso"
"llanto frofuso"
"llanto frecuente"
"llanto contenido"
"llanto espontaneo"
```

```
{ category : "Llantofacil",
  pattern : [
   { LOWER: "llanto" },
   { LOWER: {
      IN: ["facil", "copioso", "profuso", "frofuso",
           "frecuente", "contenido", "espontaneo"] }
   }]}
```

Figure 5.4 A) Illustration delineates the process of deriving patterns for NER from annotations within the document development set. B) saturation curve of recall in the development set for every new pattern (e.g., "escucha voces") added to a concept (i.e., "alucinaciones auditorias").

5.3.7 Document-Level Evaluation

We ran the "target_matcher" component using the patterns designed for NER on the 309 testing documents. Across the 69 concepts, the mean precision was 0.92, the recall was 0.83, and the mean F1 score was 0.87. Overall, 56 concepts had an F1 score above 0.80 and are considered adequate quality. (Table 5.5)

5.3.8 Patient-Level Evaluation

We ran the complete NLP algorithm (NER and ConText) and the supplementary (checklist) data to identify relevant concepts in the patient charts. The mean precision for the 69 concepts was 0.88, the mean recall was 0.74, and the mean F1 score was 0.78. Overall, 37 concepts had F1 scores that exceeded our stringent threshold of 0.80. These concepts are therefore considered of high quality for use in future studies. (Table 5.5)

5.3.9 Emergent Concepts from the Hierarchy

From the proposed concept hierarchy, we derived an additional nine *less-specific* concepts from the groupings of *more-specific* concepts. Of these, three exhibited robust performance at the patient chart level: “altered thought/speech” (n= 56, F1=0.87), “psychomotor alteration” (n= 33, F1=0.84), and “altered sleep” (n= 46, F1=0.89).

5.4 Discussion

This study represents the first use of cNLP for extracting psychiatric phenotypes from a Spanish-language EHR. Our phenotyping method integrates multi-modal data to boost recall of specific concepts. A distinctive feature of this study is that, unlike other phenotyping studies that use exclusively a top-down[7] approach to generate the list of concepts for annotation, we employed a mixed approach, with a top-down (concepts from the literature and hand-picked features) and a bottom-up components (concepts added during the initial phases of annotation). Other studies using only bottom-up [15] approaches had slightly lower F1 scores as the ones presented here (0.38-0.86). Additionally, a strength of our study lies in its inclusion of observational features frequently documented in clinical notes yet often omitted from standard assessment tools, underscoring the latent potential of routinely recorded clinical narratives.

Our process of annotation revealed interesting nuances of this process. The example of ‘delusions’ is particularly instructive, demonstrating a continuum of annotations from patient narratives like “people are following me” to clinical observations such as “delusional ideas.” This example emphasizes the need to train NLP tools in congruence with local clinical expertise. This study contributes to the broader clinical NLP community by introducing an effective cNLP algorithm for Spanish EHRs and offering an extensive annotation of psychiatric clinical notes and complete charts. This resource paves the way for the future development of machine learning cNLP algorithms.

5.4.1 Future Directions

This study is part of a larger ongoing effort. First, initiatives are underway to evaluate the algorithm’s portability, with plans to replicate our study at another psychiatric institution in Colombia (CSJDM). Second, error analysis of underperforming concepts will inform the algorithm’s iterative optimization. This refinement can potentially “rescue” concepts, amplifying the spectrum of extractable phenotypes. Third, we foresee a transition towards deep learning techniques. This development would open the opportunity to finetune the phenotyping algorithm through real-time user feedback, creating an adaptive system that learns continuously from new data.

Developing innovative phenotyping algorithms is a necessity. However, the deployment of these algorithms in real-world settings is equally valuable. We foresee our phenotype extraction system's role in developing future clinical tools. Similar initiatives are starting to yield promising new applications for improving clinical care [16,17].

5.5 Conclusion

Our investigation delineates the potential for devising clinical NLP algorithms tailored for extracting psychiatric phenotypes from clinical narratives. The capacity to capture such nuanced phenotypes at scale promises to be a cornerstone for upcoming genetic studies of SMI.

5.6 References

1. Forbush, T. B. et al. ‘Sitting on pins and needles’: characterization of symptom descriptions in clinical notes". AMIA Jt Summits Transl Sci Proc 2013, 67–71 (2013).

2. Assale, M., Dui, L. G., Cina, A., Seveso, A. & Cabitza, F. The Revival of the Notes Field: Leveraging the Unstructured Content in Electronic Health Records. Front Med (Lausanne) 6, 1–23 (2019).

3. Sarwar, T. et al. The Secondary Use of Electronic Health Records for Data Mining: Data Characteristics and Challenges. ACM Computing Surveys vol. 55 Preprint at https://doi.org/10.1145/3490234 (2023).

4. Bastarache, L. Using Phecodes for Research with the Electronic Health Record: From PheWAS to PheRS. Annu Rev Biomed Data Sci 4, 1–19 (2021).

5. Carrell, D. S. et al. Challenges in adapting existing clinical natural language processing systems to multiple, diverse health care settings. Journal of the American Medical Informatics Association 24, 986–991 (2017).

6. Liu, Q. et al. Symptom-based patient stratification in mental illness using clinical notes. J Biomed Inform 98, 103274 (2019).

7. Jackson, R. G. et al. Natural language processing to extract symptoms of severe mental illness from clinical text: The Clinical Record Interactive Search Comprehensive Data Extraction (CRIS-CODE) project. BMJ Open 7, 1–10 (2017).

8. McCoy, T. H. et al. High Throughput Phenotyping for Dimensional Psychopathology in Electronic Health Records. Biol Psychiatry 83, 997–1004 (2018).

9. Cuthbert, B. N. & Insel, T. R. Toward the future of psychiatric diagnosis: The seven pillars of RDoC. BMC Med 11, (2013).

10. Eyre, H. et al. Launching into clinical space with medspaCy: a new clinical text processing toolkit in Python. AMIA Annu Symp Proc 438–447 (2022).

11. Neves, M. & Ševa, J. An extensive review of tools for manual annotation of documents. Briefings in Bioinformatics vol. 22 146–163 Preprint at https://doi.org/10.1093/bib/bbz130 (2021).

12. Harkema, H., Dowling, J. N., Thornblade, T. & Chapman, W. W. ConText: An algorithm for determining negation, experiencer, and temporal status from clinical reports. J Biomed Inform 42, 839–851 (2009).

13. Cohen, J. A Coefficient of Agreement for Nominal Scales. Educ Psychol Meas 20, 37–46 (1960).

14. Jianlin Shi, Danielle L. Mowery, Kristina M. Doing-Harris & John F. Hurdle. RuSH: a Rule-based Segmentation Tool Using Hashing for Extremely Accurate Sentence Segmentation of Clinical Text. American Medical Informatics Association Annual Symposium (2016).

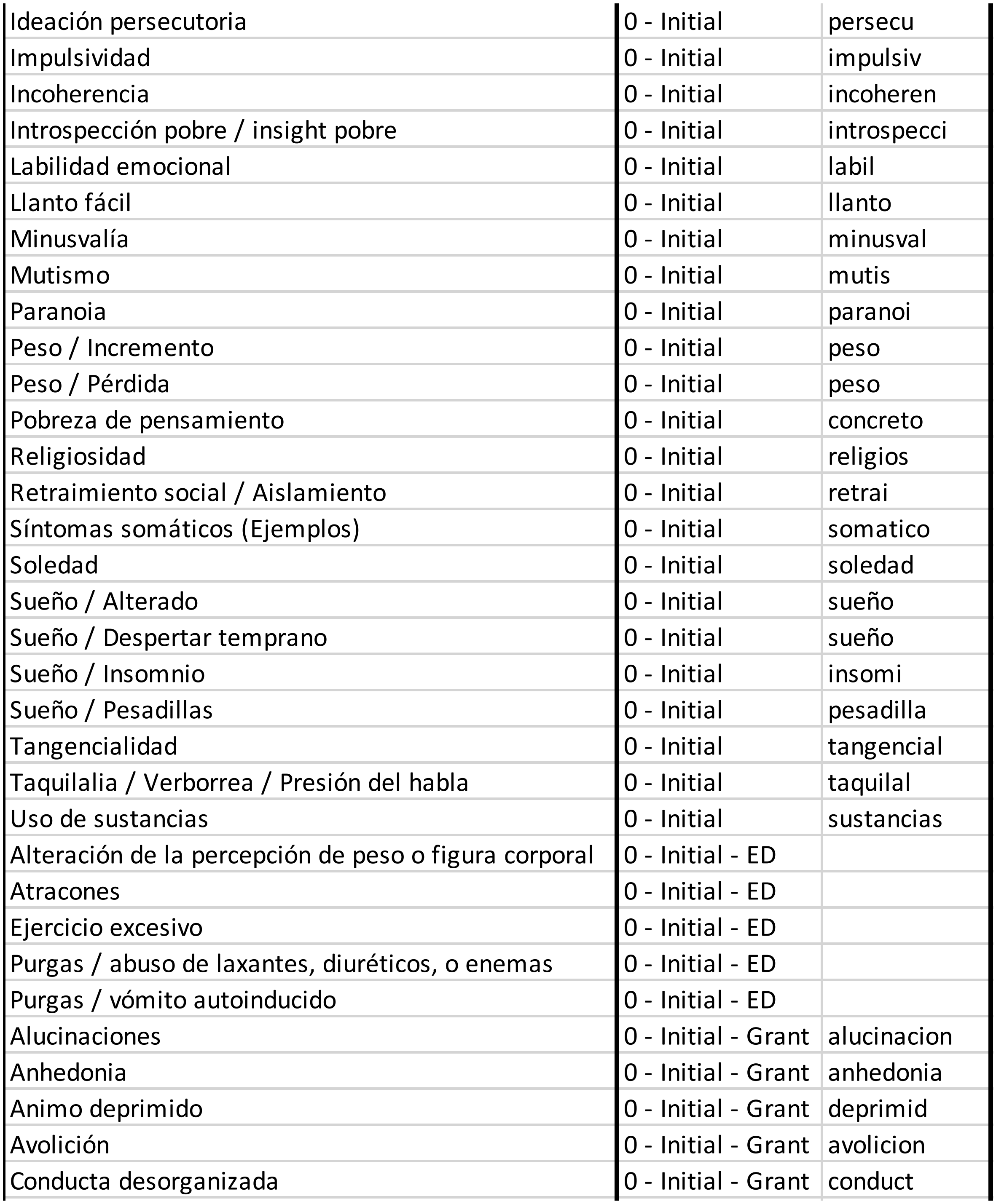
15. Turner, R. J. et al. Information extraction from free text for aiding transdiagnostic psychiatry: constructing NLP pipelines tailored to clinicians' needs. BMC Psychiatry 22, (2022).

16. Oliver, D. et al. Real-world implementation of precision psychiatry: Transdiagnostic risk calculator for the automatic detection of individuals at-risk of psychosis. Schizophr Res 227, 52–60 (2021).

17. Patel, R. et al. Associations of presenting symptoms and subsequent adverse clinical outcomes in people with unipolar depression: A prospective natural language processing (NLP), transdiagnostic, network analysis of electronic health record (EHR) data. BMJ Open 12, (2022).

Table 5.1 List of 130 concepts, the phase in which they were added and the regular expression used to search for them

Concept Name	Added	RegEx
Abulia	0 - Initial	abulia
Afecto embotado	0 - Initial	embota
Agitación psicomotora	0 - Initial	agita
Agresividad	0 - Initial	agresiv
Aislamiento emocional	0 - Initial	aislamient
Alcohol	0 - Initial	alcohol
Alogia	0 - Initial	alogia
Alucinaciones (auditorias)	0 - Initial	auditivas
Alucinaciones (visual)	0 - Initial	visuales
Alucinógenos	0 - Initial	alucinogen
Animo expansivo	0 - Initial	expansiv
Ansiedad	0 - Initial	ansi
Apatía	0 - Initial	apatia
Apetito / aumento de	0 - Initial	apetito
Apetito / disminución de	0 - Initial	apetito
Baja concentración	0 - Initial	concentracion
Baja energía	0 - Initial	energia
Cannabis	0 - Initial	marihuan
Circunstancialidad	0 - Initial	circunstancial
Cocaína	0 - Initial	cocain
Comportamiento catatónico / Catatonia	0 - Initial	catatoni
Culpa	0 - Initial	culpa
Desesperanza	0 - Initial	desesperanz
Desorden formal del pensamiento	0 - Initial	pensamiento
Despersonalización / Desrealización	0 - Initial	despersonal
Discapacidad cognitiva (Ejemplos)	0 - Initial	cognitiv
Ecolalia	0 - Initial	ecolalia
Estupor	0 - Initial	estupor
Flexibilidad cérea	0 - Initial	flexibilidad
Hipersexualidad	0 - Initial	hipersexual
Hostilidad	0 - Initial	hostil

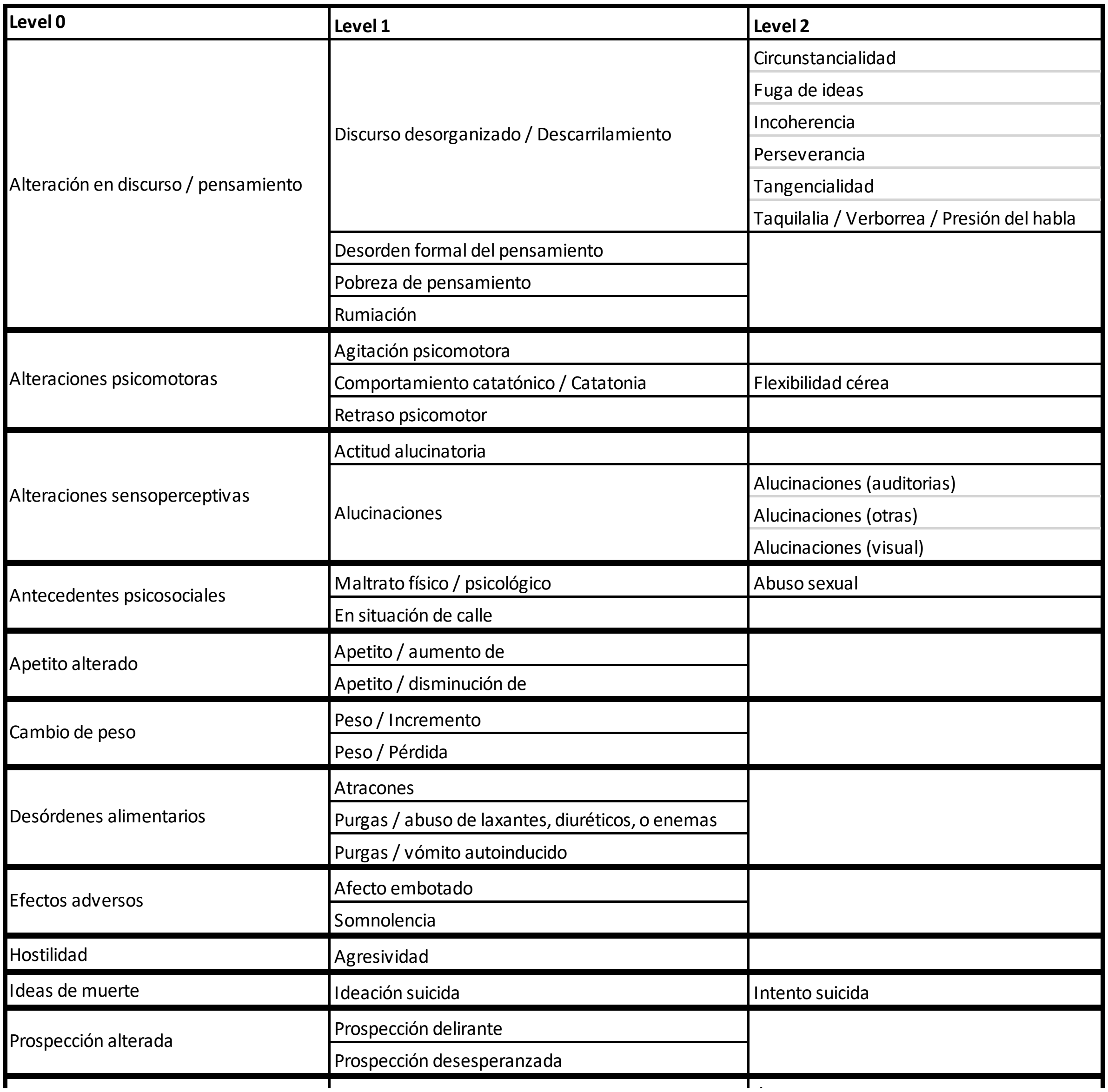


Ideación persecutoria	0 - Initial	persecu
Impulsividad	0 - Initial	impulsiv
Incoherencia	0 - Initial	incoheren
Introspección pobre / insight pobre	0 - Initial	introspecci
Labilidad emocional	0 - Initial	labil
Llanto fácil	0 - Initial	llanto
Minusvalía	0 - Initial	minusval
Mutismo	0 - Initial	mutis
Paranoia	0 - Initial	paranoi
Peso / Incremento	0 - Initial	peso
Peso / Pérdida	0 - Initial	peso
Pobreza de pensamiento	0 - Initial	concreto
Religiosidad	0 - Initial	religios
Retraimiento social / Aislamiento	0 - Initial	retrai
Síntomas somáticos (Ejemplos)	0 - Initial	somatico
Soledad	0 - Initial	soledad
Sueño / Alterado	0 - Initial	sueño
Sueño / Despertar temprano	0 - Initial	sueño
Sueño / Insomnio	0 - Initial	insomi
Sueño / Pesadillas	0 - Initial	pesadilla
Tangencialidad	0 - Initial	tangencial
Taquilalia / Verborrea / Presión del habla	0 - Initial	taquilal
Uso de sustancias	0 - Initial	sustancias
Alteración de la percepción de peso o figura corporal	0 - Initial - ED	
Atracones	0 - Initial - ED	
Ejercicio excesivo	0 - Initial - ED	
Purgas / abuso de laxantes, diuréticos, o enemas	0 - Initial - ED	
Purgas / vómito autoinducido	0 - Initial - ED	
Alucinaciones	0 - Initial - Grant	alucinacion
Anhedonia	0 - Initial - Grant	anhedonia
Animo deprimido	0 - Initial - Grant	deprimid
Avolición	0 - Initial - Grant	avolicion
Conducta desorganizada	0 - Initial - Grant	conduct

Delirios	0 - Initial - Grant	delirio
Discurso desorganizado / Descarrilamiento	0 - Initial - Grant	discurso
Fatiga	0 - Initial - Grant	fatiga
Fuga de ideas	0 - Initial - Grant	fuga
Grandiosidad	0 - Initial - Grant	grandios
Ideación suicida	0 - Initial - Grant	suicida
Intento suicida	0 - Initial - Grant	suicidio
Irritabilidad	0 - Initial - Grant	irritab
Sueño / Disminución de necesidad	0 - Initial - Grant	sueño
Sueño / Hipersomnio	0 - Initial - Grant	hipersomni
Dromomanía	0 - Initial - Juan	
Mala higiene personal	0 - Initial - Juan	
No hace contacto visual	0 - Initial - Juan	
Compulsiones	0 - Initial - Victor	
Negativismo	0 - Initial - Victor	
Obsesiones	0 - Initial - Victor	
Pánico	0 - Initial - Victor	
Quitarse la ropa / Desnudarse	0 - Initial - Victor	
Abuso de sustancias	1 - Training	
Adicción (otras)	1 - Training	
Afecto plano	1 - Training	
Alucinaciones (otras)	1 - Training	
Angustia / Miedo / Temor	1 - Training	
Autolesión	1 - Training	
Deterioro cognitivo	1 - Training	
Disforia	1 - Training	
Eutimia	1 - Training	
Hipertimia	1 - Training	
Hipotimia	1 - Training	
Ideas de muerte	1 - Training	
Inhalantes	1 - Training	
Referencial	1 - Training	
Actitud alucinatoria	2 - review 50	

Actitud pueril	2 - review 50	
Alteraciones sensoperceptivas	2 - review 50	
Cigarrillo / tabaco	2 - review 50	
Efectos adversos	2 - review 50	
En situación de calle	2 - review 50	
Experiencias de pasividad	2 - review 50	
Ideas sobrevaloradas	2 - review 50	
Juicio comprometido	2 - review 50	
Memoria alterada	2 - review 50	
Modulación afectiva alterada	2 - review 50	
Orientación alterada	2 - review 50	
Pobre adherencia	2 - review 50	
Pobre respuesta a psicofármacos	2 - review 50	
Prospección delirante	2 - review 50	
Prospección desesperanzada	2 - review 50	
Resonancia emocional alterada	2 - review 50	
Retraso psicomotor	2 - review 50	
Rumiación	2 - review 50	
Síntomas ansiosos (generales)	2 - review 50	
Síntomas cognitivos (generales)	2 - review 50	
Síntomas depresivos (generales)	2 - review 50	
Síntomas maníacos (generales)	2 - review 50	
Síntomas psicóticos (generales)	2 - review 50	
Somnolencia	2 - review 50	
Abuso sexual	3 - review 100	
Actitud de extrañeza	3 - review 100	
Falsos reconocimientos	3 - review 100	
Maltrato físico / psicológico	3 - review 100	
Perseverancia	3 - review 100	
Síntomas afectivos	3 - review 100	
Síntomas hipomaníacos	3 - review 100	
Síntomas residuales	3 - review 100	
TECAR	3 - review 100	

Table 5.2 hierarchy of concepts



Level 0	Level 1	Level 2
Alteración en discurso / pensamiento	Discurso desorganizado / Descarrilamiento	Circunstancialidad
		Fuga de ideas
		Incoherencia
		Perseverancia
		Tangencialidad
		Taquilalia / Verborrea / Presión del habla
	Desorden formal del pensamiento	
	Pobreza de pensamiento	
	Rumiación	
Alteraciones psicomotoras	Agitación psicomotora	
	Comportamiento catatónico / Catatonia	Flexibilidad cérea
	Retraso psicomotor	
Alteraciones sensoperceptivas	Actitud alucinatoria	
	Alucinaciones	Alucinaciones (auditorias)
		Alucinaciones (otras)
		Alucinaciones (visual)
Antecedentes psicosociales	Maltrato físico / psicológico	Abuso sexual
	En situación de calle	
Apetito alterado	Apetito / aumento de	
	Apetito / disminución de	
Cambio de peso	Peso / Incremento	
	Peso / Pérdida	
Desórdenes alimentarios	Atracones	
	Purgas / abuso de laxantes, diuréticos, o enemas	
	Purgas / vómito autoinducido	
Efectos adversos	Afecto embotado	
	Somnolencia	
Hostilidad	Agresividad	
Ideas de muerte	Ideación suicida	Intento suicida
Prospección alterada	Prospección delirante	
	Prospección desesperanzada	

Síntomas afectivos	Síntomas depresivos	Ánimo deprimido
	Síntomas ansiosos	Obsesiones
		Pánico
		Ansiedad
		Compulsiones
	Síntomas maníacos	Dromomanía
	Síntomas hipomaníacos	
	Animo expansivo	
	Culpa	
	Desesperanza	
	Disforia	
	Hipertimia	
	Hipotimia	
	Irritabilidad	
	Labilidad emocional	
	Llanto fácil	
	Minusvalía	
	Modulación afectiva alterada	
	Resonancia emocional alterada	
Síntomas cognitivos	Discapacidad cognitiva (Ejemplos)	
	Deterioro cognitivo	Memoria alterada
Síntomas negativos	Abulia	
	Afecto plano	
	Aislamiento emocional	
	Alogia	
	Retraimiento social / Aislamiento	
Síntomas psicóticos	Experiencias de pasividad	
	Delirios	Grandiosidad
		Ideación persecutoria
		Paranoia
		Religiosidad
Sueño alterado	Sueño / Despertar temprano	
	Sueño / Disminución de necesidad	
	Sueño / Hipersomnio	
	Sueño / Insomnio	
	Sueño / Pesadillas	
Uso de sustancias	Abuso de sustancias	
	Alcohol	
	Alucinógenos	
	Cannabis	
	Cigarrillo / tabaco	
	Cocaína	
	Inhalantes	

Table 5.3 *Frequency of regular expression matching across all sections of the intake note.*

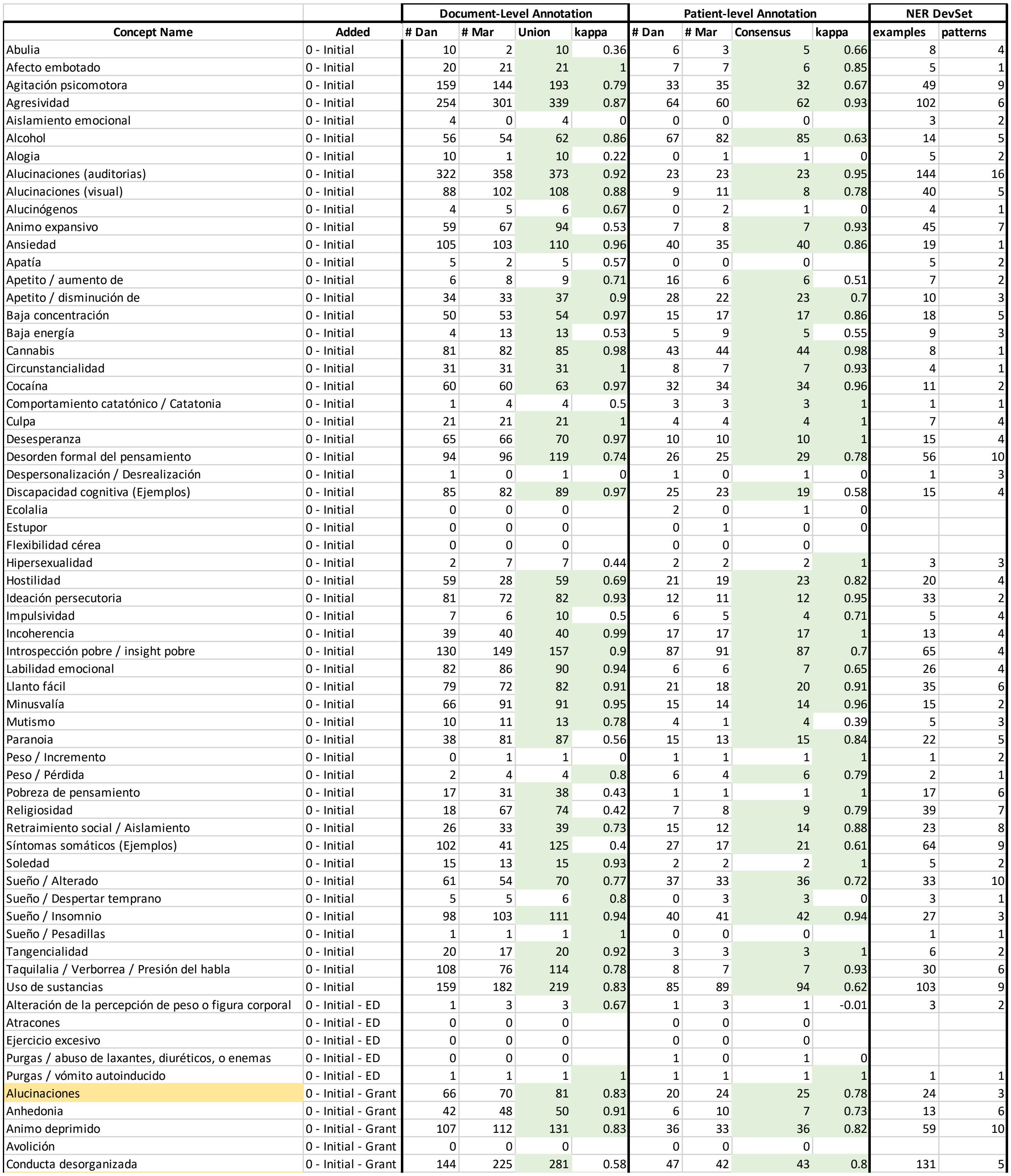
	Intake Note													
High frequency	🟦													🟦
Downloaded for annotation														🟩
	evolucion_examen_mental	nivel_alerta	porte_y_actitud	actividad_motora	afecto	sensopercepciones	lenguaje	pensamiento	memoria	juicio_y_raciocinio	introspecion	otros hallazgos	motivo	analisis_profesional
Anhedonia	0.01	0	0	0	0.1	0	0	0.02	0	0	0	0	0.04	0.28
Apatía	0	0	0	0	0.01	0	0	0	0	0	0	0	0	0.08
Apetito / aumento	0	0	0	0	0	0	0	0	0	0	0	0.02	0.02	0.08
Avolición	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Hipotimia	0.05	0	0	0	0.19	0	0	0	0	0	0	0	0	0.17
Elación (¿expansivo?)	0.11	0	0.05	0	0.17	0	0	0	0	0	0	0	0	0.59
Fatiga	0	0	0	0.01	0	0	0	0	0	0	0	0	0.01	0.06
Grandiosidad	1.1	0	0	0	0.01	0	0.01	0.62	0	0	0	0.02	0.03	2.85
Culpa	0.22	0	0	0	0	0	0	0.13	0	0	0	0	0.02	0.61
Desesperanza	0.32	0	0	0	0	0	0	0.18	0	0	0	0.01	0.02	0.68
Hipersexualidad	0.02	0	0	0	0	0	0	0	0	0	0	0	0.03	0.24
Irritabilidad	1.62	0	0.04	0	1.04	0	0	0.01	0	0	0	0.01	0.85	4.81
Baja motivación (¿Abulia?)	0.01	0	0	0.02	0	0	0	0.01	0	0	0	0.01	0.02	0.19
Pobreza de discurso	0.12	0	0	0	0	0	0.01	0.06	0	0	0	0	0	0.34
Pobreza de pensamiento	0	0	0	0	0	0	0	0	0	0	0	0	0	0.01
Presión del habla	0.16	0	0	0	0	0	0.03	0	0	0	0	0	0	0.52
Sueño / Hipersomnio	0	0	0	0	0	0	0	0	0	0	0	0.03	0.01	0.29
Sueño / Insomnio	0	0	0	0	0	0	0	0	0	0	0	0	0.01	0.05
Retraimiento social	0.07	0	0	0	0.02	0	0	0	0	0	0	0.01	0.06	1.11
Ideación suicida	0.09	0	0	0	0	0	0	0.29	0	0	0	0.01	0.04	0.56
Intento suicida	0.01	0	0	0	0	0	0	0	0	0	0	0	0.06	0.29
Llanto fácil	0.16	0	0.02	0	0.18	0	0.01	0	0	0	0	0	0.22	0.87

Llanto fácil	0.16	0	0.02	0	0.18	0	0.01	0	0	0	0	0	0.22	0.87
Peso / Incremento	0	0	0	0	0	0	0	0	0	0	0	0	0.02	0.2
Peso / Pérdida	0	0	0	0	0	0	0	0	0	0	0	0	0.01	0.15
Minusvalía	0.35	0	0	0	0	0	0	0.52	0	0	0	0.02	0.03	0.8
Agresividad	0.45	0	0.02	0.06	0.04	0	0	0.03	0	0	0	0.03	2.05	8.84
Agitación	0.33	0.01	0.02	0.14	0.01	0	0	0.01	0	0	0	0.01	0.32	6.12
Afecto plano	0.37	0	0	0	0.01	0	0	0	0	0	0	0	0.01	0.77
Pensamiento concreto	0.01	0	0	0	0	0	0	0.09	0	0	0	0.01	0	0.05
Delirios	0.55	0.03	0	0	0.01	0.01	0	2.89	0	0	0	0.02	0.21	2.83
Retraimiento emocional	0.01	0	0	0	0	0	0	0	0	0	0	0	0.01	0.1
Alucinaciones	2.64	0.01	0	0	0.01	2.48	0	0.02	0	0	0	0.02	0.47	6.06
Hostilidad	0.17	0	0.32	0	0.06	0	0	0.01	0	0	0	0.01	0.2	1.2
Paranoia	0.18	0	0.01	0	0	0.01	0	0.47	0	0	0	0.01	0.66	4.33
Ideación persecutoria	0.41	0	0	0	0	0.01	0	0.38	0	0	0	0	0.24	1.13
Comportamiento catatónico	0.06	0	0	0.01	0	0	0	0	0	0	0	0	0.01	0.83
Circunstancialidad	1.18	0	0	0	0	0	0.14	0.29	0	0	0	0.01	0.01	1.92
Ecolalia	0.04	0	0	0	0	0	0	0	0	0	0	0	0	0.05
Fuga de ideas	0.23	0	0	0	0	0	0.02	0.03	0	0	0	0	0.01	0.32
Incoherencia	0.18	0	0	0	0	0	0.11	0.13	0	0	0	0	0.6	1.52
Mutismo	0.08	0	0.01	0	0	0	0.1	0.01	0	0	0	0	0.06	0.33
Estupor	0	0	0	0	0	0	0	0	0	0	0	0	0	0.01
Tangencialidad	0.55	0	0	0	0	0	0.04	0.18	0	0	0	0	0.01	1.07
Ansiedad	0.77	0.01	0.04	0.01	1.14	0	0	0.24	0	0	0	0.03	0.32	2.45
Impulsividad	0.01	0	0	0	0	0	0	0	0	0	0	0	0.02	0.3
Solitud	0	0	0	0	0	0	0	0	0	0	0	0	0	0.03
Instabilidad del humor	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Pobre introspección	0.11	0	0	0	0	0	0	0	0	0	0.69	0	0	0.25
Religiosidad	0.15	0	0	0	0	0	0.01	0.08	0	0	0	0	0.04	0.49
Sueño / Pesadillas	0	0	0	0	0	0	0	0	0	0	0	0.01	0.02	0.06
Uso de sustancias	0.02	0	0	0	0	0	0	0	0	0	0	0	0.16	0.64
Alcohol	0.03	0	0	0	0	0	0	0.01	0	0	0	0	0.42	2.02
Cannabis	0.01	0	0	0	0	0	0	0	0	0	0	0	0.09	0.77
Cocaína	0.04	0	0	0	0	0	0	0.01	0	0	0	0	0.24	2.53
Alucinógenos	0	0	0	0	0	0	0	0	0	0	0	0	0.01	0.01
Sums	13	0.1	0.5	0.3	3	2.5	0.5	6.7	0	0	0.7	0.3	7.7	63

	Motivation			Mental Status Exam											
High frequency	🟦	🟦			🟦								🟦		🟦
Downloaded for annotation	🟩				🟩								🟩		🟩
	motivo_consulta	enfermedad_actual	revision_por_sistema	ACTIVIDAD MOTORA	AFECTO	INTELIGENCIA	INTROSPECCION	JUICIO Y RACIOCINIO	LENGUAJE	MEMORIA	NIVEL ALERTA, ORIENTACION	OTROS HALLAZGOS POSITIVOS	PENSAMIENTO	PORTE Y ACTITUD	SENSOPERCEPCIONES
Anhedonia	0.18	2.11	0.05	0	0.36	0	0	0	0	0	0	0	0	0	0
Apatía	0.01	0.26	0	0	0	0	0	0	0	0	0	0.02	0	0	0
Apetito / aumento	0.06	0.57	0.05	0	0	0	0	0	0	0	0	0.08	0	0	0
Avolición	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Hipotimia	0.02	0.23	0.01	0.01	0.8	0	0	0	0	0	0.01	0	0	0	0
Elación (¿expansivo?)	0.01	0.18	0	0.02	2.1	0	0	0	0	0	0	0.02	0.01	0.46	0
Fatiga	0.07	0.5	0.03	0.04	0	0	0	0	0	0	0	0.02	0	0	0
Grandiosidad	0.08	1.14	0.01	0	0.07	0	0	0	0.07	0	0.05	0.16	6.74	0	0.04
Culpa	0.08	1.24	0.02	0	0	0	0	0	0	0	0	0.01	0.51	0	0
Desesperanza	0.09	1.1	0.02	0.02	0.02	0.02	0.02	0.02	0.03	0.02	0.02	0.05	0.65	0.02	0.02
Hipersexualidad	0.09	0.28	0	0	0	0	0	0	0	0	0	0.04	0.01	0.01	0
Irritabilidad	2.6	17.2	0.1	0	8.56	0.02	0	0	0	0	0.08	0.07	0.04	0.42	0
Baja motivación (¿Abulia?)	0.07	0.97	0	0.03	0	0	0	0	0	0	0.01	0.07	0.01	0.01	0
Pobreza de discurso	0	0.08	0	0.01	0	0	0	0	0.03	0	0.03	0	0.44	0	0
Pobreza de pensamiento	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Presión del habla	0	0.04	0	0	0	0	0	0	0.46	0	0.01	0.02	0.04	0	0
Sueño / Hipersomnio	0.06	0.37	0.02	0.01	0	0	0	0	0	0	0	0.2	0	0	0
Sueño / Insomnio	0.03	0.39	0.01	0	0	0	0	0	0	0	0	0	0	0	0
Retraimiento social	0.21	1.73	0.02	0	0.07	0	0	0	0	0	0	0.06	0	0	0
Ideación suicida	0.11	0.63	0.02	0	0	0	0	0	0	0	0	0.06	0.4	0	0
Intento suicida	0.27	0.67	0.03	0	0	0	0	0	0	0	0	0	0.02	0	0
Llanto fácil	0.79	4.83	0.05	0.01	0.99	0.01	0.01	0.01	0.03	0.01	0.04	0	0.01	0.1	0.01

Llanto fácil		0.79	4.83	0.05		0.01	0.99	0.01	0.01	0.01	0.03	0.01	0.04	0	0.01	0.1	0.01
Peso / Incremento		0.09	0.59	0.09		0	0	0	0	0	0	0	0	0.01	0	0	0
Peso / Pérdida		0.02	0.16	0.07		0	0	0	0	0	0	0	0	0.02	0	0	0
Minusvalía		0.14	1.62	0.02		0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.07	1.81	0.06	0.01
Agresividad		7.02	19.9	0.17		0.54	0.2	0	0.02	0	0.02	0	0.06	0.36	0.21	0.25	0.01
Agitación		1.04	2.36	0.04		0.92	0.04	0.03	0	0	0	0.05	0.15	0.08	0.05	0.17	0
Afecto plano		0.03	0.16	0.02		0	0.1	0	0	0	0	0	0.09	0.02	0	0	0
Pensamiento concreto		0	0	0		0	0	0	0	0	0	0	0	0.02	0.71	0	0
Delirios		0.53	4.05	0.03		0.01	0.01	0	0	0	0.01	0.02	0.11	0.21	17.7	0	0.09
Retraimiento emocional		0.03	0.26	0		0	0	0	0	0	0	0	0	0	0	0.03	0
Alucinaciones		1.28	9.61	0.07		0.02	0	0	0	0	0	0.01	0.07	0.23	0.1	0	13.4
Hostilidad		0.59	1.93	0.09		0.08	0.42	0	0	0	0.02	0	0.06	0.06	0.09	3.17	0
Paranoia		2.02	5.43	0.04		0	0	0	0	0	0	0	0.01	0.01	3.71	0.18	0.04
Ideación persecutoria		0.66	2.51	0.01		0	0	0	0	0	0	0	0.01	0.07	1.13	0	0.04
Comportamiento catatónico		0.05	0.26	0		0.11	0	0	0	0	0.02	0	0	0.02	0	0	0
Circunstancialidad		0.03	0.28	0.01		0	0	0	0	0	1.06	0	0.03	0.12	2.57	0	0
Ecolalia		0	0.02	0		0	0	0	0	0	0.06	0	0	0	0	0	0
Fuga de ideas		0.02	0.07	0		0	0	0	0	0	0.29	0	0.01	0.03	0.56	0	0
Incoherencia		1.69	5.84	0.03		0	0	0	0	0	1.4	0	0.02	0.04	3.12	0	0
Mutismo		0.19	0.68	0		0	0	0	0	0	1.49	0	0.03	0	0.09	0.03	0
Estupor		0	0	0		0	0	0	0	0	0	0	0	0	0	0	0
Tangencialidad		0.02	0.19	0.01		0	0	0	0	0	0.51	0	0.05	0.06	1.89	0	0
Ansiedad		1.38	8.78	0.12		0.03	6.39	0	0	0	0.01	0	0.06	0.07	0.74	0.12	0
Impulsividad		0.09	1.08	0.02		0.06	0	0	0	0.01	0	0	0	0.03	0.02	0	0
Solitud		0	0.12	0.01		0	0	0	0	0	0	0	0	0	0	0	0
Instabilidad del humor		0	0	0		0	0	0	0	0	0	0	0	0	0	0	0
Pobre introspección		0	0.1	0		0	0	0	4.49	0	0	0	0.01	0.02	0.01	0	0
Religiosidad		0.12	1.19	0		0.02	0	0	0	0	0.05	0	0	0	0.74	0	0
Sueño / Pesadillas		0.07	0.57	0.01		0	0	0	0	0	0	0	0	0.03	0	0	0
Uso de sustancias		0.66	2.6	0.08		0	0	0	0	0	0	0	0.02	0	0.03	0	0
Alcohol		2.37	3.31	0.08		0	0	0	0.01	0	0	0	0	0	0.03	0.01	0
Cannabis		0.46	1.86	0.04		0	0	0	0	0	0	0	0	0	0	0	0
Cocaína		1.05	4.06	0.08		0	0	0.01	0	0	0	0	0	0.05	0.04	0	0
Alucinógenos		0.02	0.05	0		0	0	0	0	0	0	0	0	0	0	0	0
Sums		27	114	1.6		2	20	0.1	4.6	0.1	5.6	0.1	1.1	2.5	44	5	14

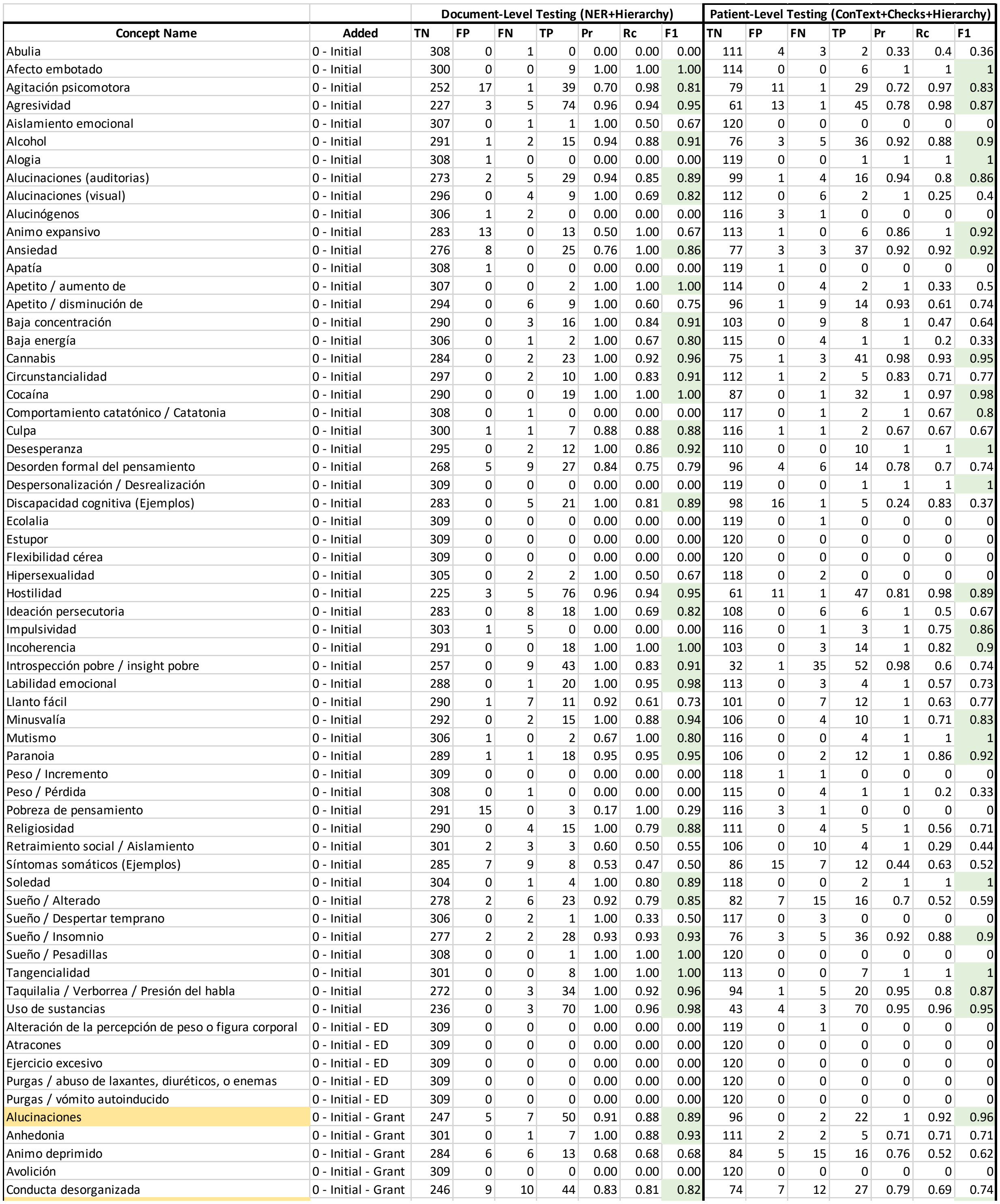
***Table 5.4** Number of annotations per concept and inter-annotator concordance for clinical documents and patient charts*



		Document-Level Annotation				Patient-level Annotation				NER DevSet	
Concept Name	Added	# Dan	# Mar	Union	kappa	# Dan	# Mar	Consensus	kappa	examples	patterns
Abulia	0 - Initial	10	2	10	0.36	6	3	5	0.66	8	4
Afecto embotado	0 - Initial	20	21	21	1	7	7	6	0.85	5	1
Agitación psicomotora	0 - Initial	159	144	193	0.79	33	35	32	0.67	49	9
Agresividad	0 - Initial	254	301	339	0.87	64	60	62	0.93	102	6
Aislamiento emocional	0 - Initial	4	0	4	0	0	0	0		3	2
Alcohol	0 - Initial	56	54	62	0.86	67	82	85	0.63	14	5
Alogia	0 - Initial	10	1	10	0.22	0	1	1	0	5	2
Alucinaciones (auditorias)	0 - Initial	322	358	373	0.92	23	23	23	0.95	144	16
Alucinaciones (visual)	0 - Initial	88	102	108	0.88	9	11	8	0.78	40	5
Alucinógenos	0 - Initial	4	5	6	0.67	0	2	1	0	4	1
Animo expansivo	0 - Initial	59	67	94	0.53	7	8	7	0.93	45	7
Ansiedad	0 - Initial	105	103	110	0.96	40	35	40	0.86	19	1
Apatía	0 - Initial	5	2	5	0.57	0	0	0		5	2
Apetito / aumento de	0 - Initial	6	8	9	0.71	16	6	6	0.51	7	2
Apetito / disminución de	0 - Initial	34	33	37	0.9	28	22	23	0.7	10	3
Baja concentración	0 - Initial	50	53	54	0.97	15	17	17	0.86	18	5
Baja energía	0 - Initial	4	13	13	0.53	5	9	5	0.55	9	3
Cannabis	0 - Initial	81	82	85	0.98	43	44	44	0.98	8	1
Circunstancialidad	0 - Initial	31	31	31	1	8	7	7	0.93	4	1
Cocaína	0 - Initial	60	60	63	0.97	32	34	34	0.96	11	2
Comportamiento catatónico / Catatonia	0 - Initial	1	4	4	0.5	3	3	3	1	1	1
Culpa	0 - Initial	21	21	21	1	4	4	4	1	7	4
Desesperanza	0 - Initial	65	66	70	0.97	10	10	10	1	15	4
Desorden formal del pensamiento	0 - Initial	94	96	119	0.74	26	25	29	0.78	56	10
Despersonalización / Desrealización	0 - Initial	1	0	1	0	1	0	1	0	1	3
Discapacidad cognitiva (Ejemplos)	0 - Initial	85	82	89	0.97	25	23	19	0.58	15	4
Ecolalia	0 - Initial	0	0	0		2	0	1	0		
Estupor	0 - Initial	0	0	0		0	1	0	0		
Flexibilidad cérea	0 - Initial	0	0	0		0	0	0			
Hipersexualidad	0 - Initial	2	7	7	0.44	2	2	2	1	3	3
Hostilidad	0 - Initial	59	28	59	0.69	21	19	23	0.82	20	4
Ideación persecutoria	0 - Initial	81	72	82	0.93	12	11	12	0.95	33	2
Impulsividad	0 - Initial	7	6	10	0.5	6	5	4	0.71	5	4
Incoherencia	0 - Initial	39	40	40	0.99	17	17	17	1	13	4
Introspección pobre / insight pobre	0 - Initial	130	149	157	0.9	87	91	87	0.7	65	4
Labilidad emocional	0 - Initial	82	86	90	0.94	6	6	7	0.65	26	4
Llanto fácil	0 - Initial	79	72	82	0.91	21	18	20	0.91	35	6
Minusvalía	0 - Initial	66	91	91	0.95	15	14	14	0.96	15	2
Mutismo	0 - Initial	10	11	13	0.78	4	1	4	0.39	5	3
Paranoia	0 - Initial	38	81	87	0.56	15	13	15	0.84	22	5
Peso / Incremento	0 - Initial	0	1	1	0	1	1	1	1	1	2
Peso / Pérdida	0 - Initial	2	4	4	0.8	6	4	6	0.79	2	1
Pobreza de pensamiento	0 - Initial	17	31	38	0.43	1	1	1	1	17	6
Religiosidad	0 - Initial	18	67	74	0.42	7	8	9	0.79	39	7
Retraimiento social / Aislamiento	0 - Initial	26	33	39	0.73	15	12	14	0.88	23	8
Síntomas somáticos (Ejemplos)	0 - Initial	102	41	125	0.4	27	17	21	0.61	64	9
Soledad	0 - Initial	15	13	15	0.93	2	2	2	1	5	2
Sueño / Alterado	0 - Initial	61	54	70	0.77	37	33	36	0.72	33	10
Sueño / Despertar temprano	0 - Initial	5	5	6	0.8	0	3	3	0	3	1
Sueño / Insomnio	0 - Initial	98	103	111	0.94	40	41	42	0.94	27	3
Sueño / Pesadillas	0 - Initial	1	1	1	1	0	0	0		1	1
Tangencialidad	0 - Initial	20	17	20	0.92	3	3	3	1	6	2
Taquilalia / Verborrea / Presión del habla	0 - Initial	108	76	114	0.78	8	7	7	0.93	30	6
Uso de sustancias	0 - Initial	159	182	219	0.83	85	89	94	0.62	103	9
Alteración de la percepción de peso o figura corporal	0 - Initial - ED	1	3	3	0.67	1	3	1	-0.01	3	2
Atracones	0 - Initial - ED	0	0	0		0	0	0			
Ejercicio excesivo	0 - Initial - ED	0	0	0		0	0	0			
Purgas / abuso de laxantes, diuréticos, o enemas	0 - Initial - ED	0	0	0		1	0	1	0		
Purgas / vómito autoinducido	0 - Initial - ED	1	1	1	1	1	1	1	1	1	1
Alucinaciones	0 - Initial - Grant	66	70	81	0.83	20	24	25	0.78	24	3
Anhedonia	0 - Initial - Grant	42	48	50	0.91	6	10	7	0.73	13	6
Animo deprimido	0 - Initial - Grant	107	112	131	0.83	36	33	36	0.82	59	10
Avolición	0 - Initial - Grant	0	0	0		0	0	0			
Conducta desorganizada	0 - Initial - Grant	144	225	281	0.58	47	42	43	0.8	131	5

Delirios	0 - Initial - Grant	385	457	499	0.86	83	84	85	0.86	172	4
Discurso desorganizado / Descarrilamiento	0 - Initial - Grant	25	25	35	0.57	8	7	8	0.64	13	4
Fatiga	0 - Initial - Grant	7	3	8	0.4	4	0	5	0	5	3
Fuga de ideas	0 - Initial - Grant	11	10	13	0.76	2	2	2	1	5	4
Grandiosidad	0 - Initial - Grant	100	94	107	0.92	6	5	6	0.71	40	3
Ideación suicida	0 - Initial - Grant	290	299	305	0.98	58	57	56	0.88	95	13
Intento suicida	0 - Initial - Grant	36	51	52	0.92	26	25	25	0.88	25	6
Irritabilidad	0 - Initial - Grant	178	178	187	0.96	43	41	43	0.96	34	4
Sueño / Disminución de necesidad	0 - Initial - Grant	7	8	8	0.93	5	5	5	1	1	1
Sueño / Hipersomnio	0 - Initial - Grant	5	5	6	0.8	2	2	2	1	3	2
Dromomanía	0 - Initial - Juan	23	17	24	0.8	6	6	6	0.82	8	3
Mala higiene personal	0 - Initial - Juan	18	52	52	0.58	21	29	22	0.8	28	9
No hace contacto visual	0 - Initial - Juan	7	11	11	0.75	5	6	5	0.33	6	2
Compulsiones	0 - Initial - Victor	7	5	7	0.8	5	4	6	0.65	4	3
Negativismo	0 - Initial - Victor	21	33	41	0.5	8	4	5	0.48	22	8
Obsesiones	0 - Initial - Victor	36	35	37	0.96	8	8	10	0.73	13	3
Pánico	0 - Initial - Victor	7	6	8	0.73	5	4	4	0.88	5	2
Quitarse la ropa / Desnudarse	0 - Initial - Victor	6	9	9	0.77	4	6	6	0.79	5	3
Abuso de sustancias	1 - Training	231	226	289	0.85	76	92	93	0.57	132	18
Adicción (otras)	1 - Training	4	12	12	0.53	0	22	21	0	7	4
Afecto plano	1 - Training	60	60	61	0.98	12	11	12	0.86	20	1
Alucinaciones (otras)	1 - Training	29	45	46	0.79	2	4	1	-0.02	33	4
Angustia / Miedo / Temor	1 - Training	118	92	124	0.84	17	18	17	0.77	53	5
Autolesión	1 - Training	27	47	52	0.62	36	35	41	0.78	25	7
Deterioro cognitivo	1 - Training	20	25	26	0.83	18	13	15	0.67	6	1
Disforia	1 - Training	61	45	62	0.85	9	9	9	1	21	3
Eutimia	1 - Training	0	13	13	0	2	22	22	0.14	5	1
Hipertimia	1 - Training	84	87	94	0.91	5	5	5	1	21	3
Hipotimia	1 - Training	96	96	98	0.98	16	16	16	1	13	2
Ideas de muerte	1 - Training	305	299	313	0.97	60	62	64	0.9	61	5
Inhalantes	1 - Training	6	7	7	0.92	9	7	10	0.87	3	2
Referencial	1 - Training	127	121	129	0.97	15	16	16	0.96	52	1
Actitud alucinatoria	2 - review 50	148	151	153	0.98	34	34	33	0.96	28	5
Actitud pueril	2 - review 50	13	12	13	0.96	3	3	3	1	3	1
Alteraciones sensoperceptivas	2 - review 50	106	104	126	0.79	66	64	71	0.66	45	9
Cigarrillo / tabaco	2 - review 50	26	32	34	0.83	59	80	85	0.52	9	4
Efectos adversos	2 - review 50	103	119	136	0.78	45	44	48	0.84	64	13
En situación de calle	2 - review 50	24	26	33	0.72	3	4	4	0.85	13	2
Experiencias de pasividad	2 - review 50	28	29	35	0.8	4	4	3	0.74	15	7
Ideas sobrevaloradas	2 - review 50	8	13	13	0.76	6	7	7	0.92	9	1
Juicio comprometido	2 - review 50	84	84	85	0.99	67	79	69	0.76	21	1
Memoria alterada	2 - review 50	17	12	17	0.85	47	43	33	0.57	12	4
Modulación afectiva alterada	2 - review 50	100	95	113	0.85	18	16	12	0.45	27	6
Orientación alterada	2 - review 50	15	22	23	0.74	14	23	14	0.65	8	2
Pobre adherencia	2 - review 50	158	202	220	0.84	56	54	56	0.66	125	10
Pobre respuesta a psicofármacos	2 - review 50	40	62	71	0.69	20	24	26	0.61	42	9
Prospección delirante	2 - review 50	14	53	53	0.41	26	23	27	0.82	14	4
Prospección desesperanzada	2 - review 50	1	4	4	0.4	1	1	0	-0.01	3	2
Resonancia emocional alterada	2 - review 50	48	43	56	0.79	5	8	7	0.76	25	8
Retraso psicomotor	2 - review 50	5	24	26	0.15	3	2	2	0.8	10	3
Rumiación	2 - review 50	10	11	12	0.9	9	7	7	0.87	3	1
Síntomas ansiosos (generales)	2 - review 50	58	61	73	0.8	28	30	32	0.86	28	4
Síntomas cognitivos (generales)	2 - review 50	17	8	21	0.4	4	5	5	0.88	7	4
Síntomas depresivos (generales)	2 - review 50	170	172	213	0.78	64	60	62	0.93	62	8
Síntomas maníacos (generales)	2 - review 50	113	127	136	0.9	18	15	19	0.82	32	1
Síntomas psicóticos (generales)	2 - review 50	237	233	249	0.96	66	64	68	0.9	75	3
Somnolencia	2 - review 50	17	17	19	0.89	9	8	10	0.68	7	3
Abuso sexual	3 - review 100	9	12	15	0.75	3	2	3	0.39	6	4
Actitud de extrañeza	3 - review 100	1	6	6	0.29	3	2	3	0.8	3	2
Falsos reconocimientos	3 - review 100	6	5	6	0.89	0	1	1	0	4	2
Maltrato físico / psicológico	3 - review 100	11	16	17	0.62	4	4	4	0.74	8	4
Perseverancia	3 - review 100	30	33	37	0.86	5	5	5	0.79	13	4
Síntomas afectivos	3 - review 100	41	50	54	0.8	30	26	34	0.67	19	3
Síntomas hipomaníacos	3 - review 100	3	3	3	1	5	8	5	0.59	1	1
Síntomas residuales	3 - review 100	28	43	43	0.83	1	5	5	0.32	13	5
TECAR	3 - review 100	68	65	73	1	25	23	25	0.9	17	5

Table 5.5 *Document and patient-level evaluation.*



		Document-Level Testing (NER+Hierarchy)							Patient-Level Testing (ConText+Checks+Hierarchy)						
Concept Name	**Added**	**TN**	**FP**	**FN**	**TP**	**Pr**	**Rc**	**F1**	**TN**	**FP**	**FN**	**TP**	**Pr**	**Rc**	**F1**
Abulia	0 - Initial	308	0	1	0	0.00	0.00	0.00	111	4	3	2	0.33	0.4	0.36
Afecto embotado	0 - Initial	300	0	0	9	1.00	1.00	1.00	114	0	0	6	1	1	1
Agitación psicomotora	0 - Initial	252	17	1	39	0.70	0.98	0.81	79	11	1	29	0.72	0.97	0.83
Agresividad	0 - Initial	227	3	5	74	0.96	0.94	0.95	61	13	1	45	0.78	0.98	0.87
Aislamiento emocional	0 - Initial	307	0	1	1	1.00	0.50	0.67	120	0	0	0	0	0	0
Alcohol	0 - Initial	291	1	2	15	0.94	0.88	0.91	76	3	5	36	0.92	0.88	0.9
Alogia	0 - Initial	308	1	0	0	0.00	0.00	0.00	119	0	0	1	1	1	1
Alucinaciones (auditorias)	0 - Initial	273	2	5	29	0.94	0.85	0.89	99	1	4	16	0.94	0.8	0.86
Alucinaciones (visual)	0 - Initial	296	0	4	9	1.00	0.69	0.82	112	0	6	2	1	0.25	0.4
Alucinógenos	0 - Initial	306	1	2	0	0.00	0.00	0.00	116	3	1	0	0	0	0
Animo expansivo	0 - Initial	283	13	0	13	0.50	1.00	0.67	113	1	0	6	0.86	1	0.92
Ansiedad	0 - Initial	276	8	0	25	0.76	1.00	0.86	77	3	3	37	0.92	0.92	0.92
Apatía	0 - Initial	308	1	0	0	0.00	0.00	0.00	119	1	0	0	0	0	0
Apetito / aumento de	0 - Initial	307	0	0	2	1.00	1.00	1.00	114	0	4	2	1	0.33	0.5
Apetito / disminución de	0 - Initial	294	0	6	9	1.00	0.60	0.75	96	1	9	14	0.93	0.61	0.74
Baja concentración	0 - Initial	290	0	3	16	1.00	0.84	0.91	103	0	9	8	1	0.47	0.64
Baja energía	0 - Initial	306	0	1	2	1.00	0.67	0.80	115	0	4	1	1	0.2	0.33
Cannabis	0 - Initial	284	0	2	23	1.00	0.92	0.96	75	1	3	41	0.98	0.93	0.95
Circunstancialidad	0 - Initial	297	0	2	10	1.00	0.83	0.91	112	1	2	5	0.83	0.71	0.77
Cocaína	0 - Initial	290	0	0	19	1.00	1.00	1.00	87	0	1	32	1	0.97	0.98
Comportamiento catatónico / Catatonia	0 - Initial	308	0	1	0	0.00	0.00	0.00	117	0	1	2	1	0.67	0.8
Culpa	0 - Initial	300	1	1	7	0.88	0.88	0.88	116	1	1	2	0.67	0.67	0.67
Desesperanza	0 - Initial	295	0	2	12	1.00	0.86	0.92	110	0	0	10	1	1	1
Desorden formal del pensamiento	0 - Initial	268	5	9	27	0.84	0.75	0.79	96	4	6	14	0.78	0.7	0.74
Despersonalización / Desrealización	0 - Initial	309	0	0	0	0.00	0.00	0.00	119	0	0	1	1	1	1
Discapacidad cognitiva (Ejemplos)	0 - Initial	283	0	5	21	1.00	0.81	0.89	98	16	1	5	0.24	0.83	0.37
Ecolalia	0 - Initial	309	0	0	0	0.00	0.00	0.00	119	0	1	0	0	0	0
Estupor	0 - Initial	309	0	0	0	0.00	0.00	0.00	120	0	0	0	0	0	0
Flexibilidad cérea	0 - Initial	309	0	0	0	0.00	0.00	0.00	120	0	0	0	0	0	0
Hipersexualidad	0 - Initial	305	0	2	2	1.00	0.50	0.67	118	0	2	0	0	0	0
Hostilidad	0 - Initial	225	3	5	76	0.96	0.94	0.95	61	11	1	47	0.81	0.98	0.89
Ideación persecutoria	0 - Initial	283	0	8	18	1.00	0.69	0.82	108	0	6	6	1	0.5	0.67
Impulsividad	0 - Initial	303	1	5	0	0.00	0.00	0.00	116	0	1	3	1	0.75	0.86
Incoherencia	0 - Initial	291	0	0	18	1.00	1.00	1.00	103	0	3	14	1	0.82	0.9
Introspección pobre / insight pobre	0 - Initial	257	0	9	43	1.00	0.83	0.91	32	1	35	52	0.98	0.6	0.74
Labilidad emocional	0 - Initial	288	0	1	20	1.00	0.95	0.98	113	0	3	4	1	0.57	0.73
Llanto fácil	0 - Initial	290	1	7	11	0.92	0.61	0.73	101	0	7	12	1	0.63	0.77
Minusvalía	0 - Initial	292	0	2	15	1.00	0.88	0.94	106	0	4	10	1	0.71	0.83
Mutismo	0 - Initial	306	1	0	2	0.67	1.00	0.80	116	0	0	4	1	1	1
Paranoia	0 - Initial	289	1	1	18	0.95	0.95	0.95	106	0	2	12	1	0.86	0.92
Peso / Incremento	0 - Initial	309	0	0	0	0.00	0.00	0.00	118	1	1	0	0	0	0
Peso / Pérdida	0 - Initial	308	0	1	0	0.00	0.00	0.00	115	0	4	1	1	0.2	0.33
Pobreza de pensamiento	0 - Initial	291	15	0	3	0.17	1.00	0.29	116	3	1	0	0	0	0
Religiosidad	0 - Initial	290	0	4	15	1.00	0.79	0.88	111	0	4	5	1	0.56	0.71
Retraimiento social / Aislamiento	0 - Initial	301	2	3	3	0.60	0.50	0.55	106	0	10	4	1	0.29	0.44
Síntomas somáticos (Ejemplos)	0 - Initial	285	7	9	8	0.53	0.47	0.50	86	15	7	12	0.44	0.63	0.52
Soledad	0 - Initial	304	0	1	4	1.00	0.80	0.89	118	0	0	2	1	1	1
Sueño / Alterado	0 - Initial	278	2	6	23	0.92	0.79	0.85	82	7	15	16	0.7	0.52	0.59
Sueño / Despertar temprano	0 - Initial	306	0	2	1	1.00	0.33	0.50	117	0	3	0	0	0	0
Sueño / Insomnio	0 - Initial	277	2	2	28	0.93	0.93	0.93	76	3	5	36	0.92	0.88	0.9
Sueño / Pesadillas	0 - Initial	308	0	0	1	1.00	1.00	1.00	120	0	0	0	0	0	0
Tangencialidad	0 - Initial	301	0	0	8	1.00	1.00	1.00	113	0	0	7	1	1	1
Taquilalia / Verborrea / Presión del habla	0 - Initial	272	0	3	34	1.00	0.92	0.96	94	1	5	20	0.95	0.8	0.87
Uso de sustancias	0 - Initial	236	0	3	70	1.00	0.96	0.98	43	4	3	70	0.95	0.96	0.95
Alteración de la percepción de peso o figura corporal	0 - Initial - ED	309	0	0	0	0.00	0.00	0.00	119	0	1	0	0	0	0
Atracones	0 - Initial - ED	309	0	0	0	0.00	0.00	0.00	120	0	0	0	0	0	0
Ejercicio excesivo	0 - Initial - ED	309	0	0	0	0.00	0.00	0.00	120	0	0	0	0	0	0
Purgas / abuso de laxantes, diuréticos, o enemas	0 - Initial - ED	309	0	0	0	0.00	0.00	0.00	120	0	0	0	0	0	0
Purgas / vómito autoinducido	0 - Initial - ED	309	0	0	0	0.00	0.00	0.00	120	0	0	0	0	0	0
Alucinaciones	0 - Initial - Grant	247	5	7	50	0.91	0.88	0.89	96	0	2	22	1	0.92	0.96
Anhedonia	0 - Initial - Grant	301	0	1	7	1.00	0.88	0.93	111	2	2	5	0.71	0.71	0.71
Animo deprimido	0 - Initial - Grant	284	6	6	13	0.68	0.68	0.68	84	5	15	16	0.76	0.52	0.62
Avolición	0 - Initial - Grant	309	0	0	0	0.00	0.00	0.00	120	0	0	0	0	0	0
Conducta desorganizada	0 - Initial - Grant	246	9	10	44	0.83	0.81	0.82	74	7	12	27	0.79	0.69	0.74

Delirios	0 - Initial - Grant	183	0	6	120	1.00	0.95	0.98	78	3	3	36	0.92	0.92	0.92
Discurso desorganizado / Descarrilamiento	0 - Initial - Grant	244	3	7	55	0.95	0.89	0.92	75	3	9	33	0.92	0.79	0.85
Fatiga	0 - Initial - Grant	307	0	2	0	0.00	0.00	0.00	115	0	5	0	0	0	0
Fuga de ideas	0 - Initial - Grant	303	3	0	3	0.50	1.00	0.67	117	1	1	1	0.5	0.5	0.5
Grandiosidad	0 - Initial - Grant	279	1	2	27	0.96	0.93	0.95	114	0	1	5	1	0.83	0.91
Ideación suicida	0 - Initial - Grant	242	1	12	54	0.98	0.82	0.89	89	7	2	22	0.76	0.92	0.83
Intento suicida	0 - Initial - Grant	297	1	5	6	0.86	0.55	0.67	99	2	7	12	0.86	0.63	0.73
Irritabilidad	0 - Initial - Grant	255	0	7	47	1.00	0.87	0.93	77	4	6	33	0.89	0.85	0.87
Sueño / Disminución de necesidad	0 - Initial - Grant	303	0	5	1	1.00	0.17	0.29	115	0	3	2	1	0.4	0.57
Sueño / Hipersomnio	0 - Initial - Grant	307	0	1	1	1.00	0.50	0.67	118	0	1	1	1	0.5	0.67
Dromomanía	0 - Initial - Juan	299	2	2	6	0.75	0.75	0.75	111	3	2	4	0.57	0.67	0.62
Mala higiene personal	0 - Initial - Juan	292	0	3	14	1.00	0.82	0.90	98	0	10	12	1	0.55	0.71
No hace contacto visual	0 - Initial - Juan	306	0	1	2	1.00	0.67	0.80	116	0	0	4	1	1	1
Compulsiones	0 - Initial - Victor	306	0	1	2	1.00	0.67	0.80	113	1	3	3	0.75	0.5	0.6
Negativismo	0 - Initial - Victor	295	3	7	4	0.57	0.36	0.44	114	1	1	4	0.8	0.8	0.8
Obsesiones	0 - Initial - Victor	294	0	2	13	1.00	0.87	0.93	111	0	6	3	1	0.33	0.5
Pánico	0 - Initial - Victor	307	0	0	2	1.00	1.00	1.00	117	0	1	2	1	0.67	0.8
Quitarse la ropa / Desnudarse	0 - Initial - Victor	307	0	2	0	0.00	0.00	0.00	114	0	3	3	1	0.5	0.67
Abuso de sustancias	1 - Training	246	11	4	48	0.81	0.92	0.86	56	7	8	49	0.88	0.86	0.87
Adicción (otras)	1 - Training	300	3	6	0	0.00	0.00	0.00	108	1	8	3	0.75	0.27	0.4
Afecto plano	1 - Training	302	0	0	7	1.00	1.00	1.00	108	0	0	12	1	1	1
Alucinaciones (otras)	1 - Training	301	0	6	2	1.00	0.25	0.40	120	0	0	0	0	0	0
Angustia / Miedo / Temor	1 - Training	284	0	2	23	1.00	0.92	0.96	97	6	1	16	0.73	0.94	0.82
Autolesión	1 - Training	288	5	8	8	0.62	0.50	0.55	104	1	10	5	0.83	0.33	0.48
Deterioro cognitivo	1 - Training	293	2	3	11	0.85	0.79	0.81	97	3	10	10	0.77	0.5	0.61
Disforia	1 - Training	295	3	0	11	0.79	1.00	0.88	107	4	1	8	0.67	0.89	0.76
Eutimia	1 - Training	303	1	0	5	0.83	1.00	0.91	97	1	2	20	0.95	0.91	0.93
Hipertimia	1 - Training	283	0	5	21	1.00	0.81	0.89	115	0	3	2	1	0.4	0.57
Hipotimia	1 - Training	292	0	2	15	1.00	0.88	0.94	106	1	0	13	0.93	1	0.96
Ideas de muerte	1 - Training	225	0	7	77	1.00	0.92	0.96	70	14	6	30	0.68	0.83	0.75
Inhalantes	1 - Training	309	0	0	0	0.00	0.00	0.00	110	0	3	7	1	0.7	0.82
Referencial	1 - Training	278	0	3	28	1.00	0.90	0.95	103	1	2	14	0.93	0.88	0.9
Actitud alucinatoria	2 - review 50	271	0	7	31	1.00	0.82	0.90	99	2	0	19	0.9	1	0.95
Actitud pueril	2 - review 50	304	0	0	5	1.00	1.00	1.00	117	0	0	3	1	1	1
Alteraciones sensoperceptivas	2 - review 50	201	1	14	93	0.99	0.87	0.93	75	6	1	38	0.86	0.97	0.92
Cigarrillo / tabaco	2 - review 50	291	2	2	14	0.88	0.88	0.88	75	5	10	30	0.86	0.75	0.8
Efectos adversos	2 - review 50	256	3	10	40	0.93	0.80	0.86	64	7	21	28	0.8	0.57	0.67
En situación de calle	2 - review 50	298	2	7	2	0.50	0.22	0.31	117	0	0	3	1	1	1
Experiencias de pasividad	2 - review 50	301	0	4	4	1.00	0.50	0.67	117	0	2	1	1	0.33	0.5
Ideas sobrevaloradas	2 - review 50	306	0	0	3	1.00	1.00	1.00	115	0	1	4	1	0.8	0.89
Juicio comprometido	2 - review 50	271	0	3	35	1.00	0.92	0.96	52	2	29	37	0.95	0.56	0.7
Memoria alterada	2 - review 50	304	1	4	0	0.00	0.00	0.00	101	1	8	10	0.91	0.56	0.69
Modulación afectiva alterada	2 - review 50	286	2	5	16	0.89	0.76	0.82	107	1	3	9	0.9	0.75	0.82
Orientación alterada	2 - review 50	299	0	3	7	1.00	0.70	0.82	100	6	5	9	0.6	0.64	0.62
Pobre adherencia	2 - review 50	259	4	12	34	0.89	0.74	0.81	48	17	20	35	0.67	0.64	0.65
Pobre respuesta a psicofármacos	2 - review 50	281	6	11	11	0.65	0.50	0.56	95	2	16	7	0.78	0.3	0.44
Prospección delirante	2 - review 50	277	2	7	23	0.92	0.77	0.84	93	0	15	12	1	0.44	0.62
Prospección desesperanzada	2 - review 50	307	2	0	0	0.00	0.00	0.00	120	0	0	0	0	0	0
Resonancia emocional alterada	2 - review 50	290	5	1	13	0.72	0.93	0.81	112	2	1	5	0.71	0.83	0.77
Retraso psicomotor	2 - review 50	294	11	2	2	0.15	0.50	0.24	115	3	1	1	0.25	0.5	0.33
Rumiación	2 - review 50	305	0	0	4	1.00	1.00	1.00	114	0	2	4	1	0.67	0.8
Síntomas ansiosos (generales)	2 - review 50	260	0	3	46	1.00	0.94	0.97	65	0	8	47	1	0.85	0.92
Síntomas cognitivos (generales)	2 - review 50	267	3	8	31	0.91	0.79	0.85	80	12	10	18	0.6	0.64	0.62
Síntomas depresivos (generales)	2 - review 50	246	1	8	54	0.98	0.87	0.92	61	2	8	49	0.96	0.86	0.91
Síntomas maníacos (generales)	2 - review 50	262	2	2	43	0.96	0.96	0.96	102	4	1	13	0.76	0.93	0.84
Síntomas psicóticos (generales)	2 - review 50	161	1	6	141	0.99	0.96	0.98	62	5	2	51	0.91	0.96	0.94
Somnolencia	2 - review 50	304	0	0	5	1.00	1.00	1.00	108	2	4	6	0.75	0.6	0.67
Abuso sexual	3 - review 100	304	1	1	3	0.75	0.75	0.75	116	1	1	2	0.67	0.67	0.67
Actitud de extrañeza	3 - review 100	305	3	1	0	0.00	0.00	0.00	112	6	2	0	0	0	0
Falsos reconocimientos	3 - review 100	308	0	0	1	1.00	1.00	1.00	119	1	0	0	0	0	0
Maltrato físico / psicológico	3 - review 100	302	1	3	3	0.75	0.50	0.60	110	3	5	2	0.4	0.29	0.33
Perseverancia	3 - review 100	295	2	7	5	0.71	0.42	0.53	113	2	1	4	0.67	0.8	0.73
Síntomas afectivos	3 - review 100	104	3	12	190	0.98	0.94	0.96	14	2	4	100	0.98	0.96	0.97
Síntomas hipomaníacos	3 - review 100	307	0	1	1	1.00	0.50	0.67	118	1	1	0	0	0	0
Síntomas residuales	3 - review 100	294	2	2	11	0.85	0.85	0.85	115	0	5	0	0	0	0
TECAR	3 - review 100	301	0	0	8	1.00	1.00	1.00	117	0	1	2	1	0.67	0.8

CHAPTER 6

Concluding Remarks

The search for scalable approaches to phenotyping Serious Mental Illness (SMI) has become increasingly relevant thanks to the emergence of large research biobanks and hypothesis-free study designs such as GWAS. It is widely recognized that medical records, as repositories of data passively collected throughout healthcare delivery, offer an untapped wealth of phenotypic information ripe for research repurposing. However, leveraging these resources requires new phenotyping approaches and tools.

This dissertation proposes a novel framework for phenotyping psychiatric disorders from Electronic Health Records (EHR) for large-scale biomedical research. This framework embraces the complexity of SMI by delineating a set of approaches that broaden the spectrum of phenotypes that can, accurately and efficiently, be extracted from EHRs. Combining these phenotypes with the rigorous data cleaning, de-identification, and integration with external databases described in Chapter Two, we lay the groundwork for future studies into the biological underpinnings of SMI in the Paisa population of Colombia.

A critical component of the work developed in Chapter Three was the validation of diagnostic codes for SMI diagnoses used by psychiatrists in clinical practice. This validation highlights the value of EHRs in representing real-world clinical diagnoses and their potential for secondary use in biomedical research. Furthermore, in Chapter Four, our framework aligns with recent initiatives in psychiatric research and practice that move away from traditional diagnostic categories and towards dimensional measures of illness severity. By leveraging longitudinal EHR data, our approach defines a continuous space shared by the most common SMI diagnoses:

Major Depression, Bipolar Disorder, and Schizophrenia. This three-dimensional phenotype illuminates the overlap and distinctiveness of clinical features across diagnoses and time.

In Chapter Five, we achieve a significant leap in precision by moving beyond broad diagnostic categories and incorporating nuanced, symptom-level data that can be extracted from narrative clinical notes using clinical NLP (cNLP). Before extracting, we annotated one of the largest and most comprehensive datasets of psychiatric EHRs in Spanish, with an extensive array of relevant and highly granular phenotypic information. With this resource, we generated the first cNLP algorithm to extract psychiatric phenotypes from Spanish-written clinical notes. Our hope is that this tool will expand the scope of psychiatric research to a broader Spanish-speaking population. At the same time, the training data generated here will allow for future developments to integrate large language models to improve clinical data extraction.

Medical records provide a unique perspective into the illness trajectories of individuals with SMI. Our framework leverages this new perspective to characterize the immense diversity of illness trajectories within a real-world population and learn what clinical features affect the stability of psychiatric diagnoses. Future efforts will aim to reduce the dimensionality of these trajectories, thereby enhancing their research utility and clinical value.

While these advances mark significant progress, the journey toward defining and extracting psychiatric phenotypes at scale is ongoing. Limitations regarding the standardization and completeness of clinical data pose significant challenges to our ability to characterize SMI fully.

As the availability of EHR-linked biobanks continues to expand, we expect that future studies will continue to build upon the foundation of our phenotyping framework. Upcoming genetic studies will affirm the validity of the phenotypes delineated in this research and assess the transferability of our findings, both within the Paisa population and beyond. The work presented here, therefore, does not mark an end but a beginning. It is a stepping stone towards a more nuanced and inclusive field of the genetic study of psychiatric disorders. We hope this work contributes to the evolving field of precision psychiatry, ultimately benefiting diverse populations worldwide.

APPENDIX A

Supplementary Materials for Chapter 3

SUPPLEMENTARY NOTES

N.1 Supplementary Note 1. NLP algorithm for symptom extraction

N.1.1 Overview: Our algorithm aims to extract clinical features from Spanish text, which involves two distinct tasks: Named Entity Recognition (NER) and Negation Detection (ND). NER involves identifying instances of a particular clinical feature in the text, for which we utilized the EntityRuler component of the spaCy NLP library (v2.4). ND assesses whether a feature is affirmed or negated within a sentence. To perform this task, we implemented the NegEx algorithm.

N.1.2 Sampling of sentences: We focused on extracting four specific clinical features: Suicide Attempt, Suicidal Ideation, Delusions, and Hallucinations. In the EHR, we identified nine sections most likely to contain these features. These sections span across three types of notes. **Intake note:** (1) chief complaint, (2) thought content, (3) sensory perception, and (4) assessment. **Progress note**: (5) subjective, (6) objective, and (7) assessment. **Outpatient note**: (8) subjective and (9) assessment.

Notes are represented in the EHR database in the form of tables, where each type of note is stored in its corresponding table. In each of these three tables, a row represents an individual note taken at a specific time for a specific patient, while columns represent distinct sections of a note. For example, in the table for Intake Notes, a row may be the first note of patient X and the columns will be the chief complaint, the thought content, and so on. Each cell in the table, then, contains a brief text describing one section of a patient's note. These texts may be as small as a single sentence. Additional columns in the table contain other relevant information, such as the patient ID and the date of the note. Lastly, one of the columns contains the primary ICD-10 code associated with each note (i.e., with each row). The rows in each table were selected to keep only those associated with an ICD-10 code of an SMI diagnosis (F20X, F22X, F25X, F301, F302, F310, F311, F312, F313, F314, F315, F316, F317, F322, F323, F331, F332, F333, or F334), resulting in 19,713 intake notes, 225,362 progress notes, and 26,673 outpatient notes.

From these filtered tables, we isolated each of the nine columns that represent the sections identified above. Finally, from each column we randomly sampled 400 cells for annotation, i.e., a total of 3600 cells across the nine columns. The text contained in each of the 3600 cells is hereafter referred to as a "sentence".

N.1.3 Annotation of clinical features*:* Two clinicians independently annotated each sentence for the presence of the four clinical features, identifying the specific span of text inside the sentence in which each feature was mentioned. For example, in the sentence "the patient is presenting with auditory hallucinations", the text "auditory hallucinations" would be identified, and the sentence flagged for the presence of the feature – Hallucinations.

N.1.5 Gold Standard and Development Set Creation*:* Of the 3600 sentences, at least one annotator flagged 83 for Suicide Attempt, 523 for Suicidal Ideation, 317 for Delusions, and 495 for Hallucinations. We selected at random 30% of each one of these four sets of sentences to generate a Gold Standard. The sentences in these sets may overlap with each other, i.e., sentences selected based on one feature may contain additional features; for example, "the patient is presenting with persecutory *delusions* and *suicidal ideation*".

In total, 290 sentences were included in the gold standard; in these sentences all inconsistencies between annotators were resolved, and each clinical feature was labeled as either affirmed or negated. The remaining 3310 sentences were used as the development set. For each of the four clinical features in the development set, we counted how many sentences were flagged by either of the annotators, or by both of them, and estimated the inter-annotator agreement using Cohen's Kappa [21].

N.1.6 Algorithm for Clinical Feature Extraction: The first step was to develop the list of search patterns to be used in NER. For this, we used the spans of text identified inside the sentences of the development set. Concretely, these spans of text are sequences of words that represent a clinical feature. We converted each span to lowercase and formatted it using two different components of spaCy's medium-sized Spanish language model (sp_news_md), as follows: First, we used the "tokenizer" component to split the span of text into a sequence of words and punctuation marks (jointly known as tokens). Then, we used the "tagger" component to assign a Part-of-Speech (POS) tag to each token. Within each sequence of tokens, we replaced the tokens with their POS-tags, unless their tags were one of the following: noun, verb, adjective, adverb, pronoun, auxiliary or subordinating conjunction. In such cases, the tokens were replaced by their lemma. As a result, a search pattern is a sequence of POS-tags and lemmas of the same length and order as the sequence of tokens identified during annotation. (Supplementary Table 2). The final list of search patterns was manually curated to increase the coherence of the clinical feature. The procedures for this manual curation included: first, removing patterns that, by themselves, were not sufficiently complete to ensure that they indicated the presence of the feature (e.g., the pattern "visual" is not sufficient to indicate Hallucinations); then, delineating the boundaries of the features by excluding patterns that could not be unambiguously interpreted as indicating the presence of the feature (e.g., separating Suicide Attempt from self-harm). These procedures reduced the initial number of patterns for each feature, from 44 to 23 in Suicide Attempt, from 176 to 78 in Suicidal Ideation, from 127 to 119 in Delusions and from 154 to 122 in Hallucinations. Finally, the curated list of patterns was passed to the "EntityRuler" component of spaCy to complete the NER task.

Subsequently, we detected the negation status of each clinical feature using the NegEx algorithm. Briefly, this algorithm assumes each feature to be "affirmed" by default and its status is only changed to "negated" when it is located within five tokens of a negation cue. Negation cues were manually identified in the development set (Supplementary Table 3) and their location in the text was determined using the "EntityRuler" component of spaCy.

To improve the accuracy of both patterns and negation terms, we ran our full pipeline (NER and ND) with each of the four features, on the entire EHR database and randomly selected 100 instances (50 affirmed and 50 negated) of each feature for manual review. Two clinicians evaluated the 400 instances and recommended adjustments to the search patterns and negation cues. Although this process can be iterated until the expected performance is achieved, we considered a single iteration to be sufficient for all features.

N.1.7 Validation of extracted features: To evaluate the performance of our algorithm for each one of the four clinical features, we report precision, recall, and F1. First, at the level of individual sentences, we used the 290 sentences of the gold standard described above. At the patient level, we used the item checklist from the clinician's chart review of 120 patients described in the main text. For any given patient, a "lifetime" phenotype extracted by our NLP pipeline was defined as follows: by default, the phenotype is absent, and it is changed to present only if the patient has two or more notes with affirmed instances of the clinical feature. One patient out of the 120 had to be excluded from this evaluation because they had only one note on record.

N.1.8 Threshold for a "lifetime" phenotype from NLP extracted features*:* The threshold of "two notes or more" is arbitrary. We hypothesized that requiring more affirmative mentions of a feature to classify a patient would result in increased precision of the phenotype, while at the same time reducing its recall; that is, increasing the two-note threshold could result in a narrower and likely more severe phenotype, albeit with a smaller sample size. We tested this hypothesis by varying the number of notes required to change a patient's "lifetime" phenotype from absent to present from 1 to 10 and determined the optimal balance of precision and recall using F1.

N.1.9 Addressing human error in chart review*:* Some degree of human error can be expected when performing clinical chart review. Specifically, features may be overlooked in reviewing an extensive clinical history. We therefore conducted an additional manual chart review (by two independent clinicians blinded to the patients' original classification) of the 29 instances of apparent discordance between the algorithm output (which reported one or more features) and the initial manual review (which had not reported these features). In 16 of these instances both of the subsequent reviewers reported features that the initial reviewer had apparently overlooked. In 4 additional instances the output was ambiguous, meaning that the two raters disagreed on whether the concept was present. The remaining 9 instances were false positives. Agnostic to the observed concordance/discordance of each instance, reviewers also independently examined 29 randomly selected instances for which the algorithm output and the initial review were concordant (both interview and review reported a feature); their reports confirmed the concordance for 28 of these instances, with the last one being ambiguous. We then incorporated the information obtained from this second review in a *post hoc* analysis comparing the algorithm output to manual review, requiring agreement between both raters; we observed an increase in concordance and F1 for all four clinical features (Suicide Attempt:

94/104, F1 = 0.75; Suicidal Ideation: 95/104, F1 = 0.89; Delusions: 88/104, F1 = 0.86 and Hallucinations: 91/104, F1 = 0.88), Supplementary Table 8C).

N.2 Supplementary Note 2. Patient-level data validation: NLP features and ICD-10 diagnoses

We evaluated the relationship between the presence of *lifetime* clinical features with gender and with the most recent diagnoses of MDD, BD, and SCZ, at the *individual level*.

As established above in our validation of NLP clinical features using chart review, we defined "lifetime" features for patients with at least two notes (n=20,658). We define a feature as present in a patient if at least two notes in their records have an affirmative mention of the feature.

To test for associations between diagnosis, gender, and features, we used logistic regression to model the logit of the probability of a feature being present as a function of gender (female=1), and current diagnosis (BD as reference).

In this model, we accounted for the number of notes a patient has on record using the $\log_{10}$ transformed variable $N\ notes$, since the likelihood of a feature being present is expected to increase with the number of notes. We also account for illness severity by including a binary variable that is 1 if the individual has had a history of hospitalization (*Inpatient*).

Consider the probability that feature *s* is present to be P_{Sx_s}. The resulting model is:

$$ln\left(\frac{P_{Sx_s}}{1-P_{Sx_s}}\right) \sim \beta_0 + \beta_1 MDD + \beta_2 SCZ + \beta_3 Gender + \ \beta_4 log_{10}(N\ notes) + \beta_5 Inpatient$$

We fit four models in total – one for each feature. In each model we tested for associations with gender and with two diagnoses (MDD and SCZ), for a Bonferroni-corrected alpha of 0.05/12=0.0041.

Then, to explore the interactions between gender and diagnosis, we expanded the model to include interaction terms. This model is expressed as:

$$ln\left(\frac{P_{Sx_s}}{1-P_{Sx_s}}\right) \sim \beta_0 + \beta_1 MDD + \beta_2 SCZ + \beta_3 Gender + \ \beta_4 log_{10}(N\ notes) + \beta_5 Inpatient$$
$$+ \beta_6 Gender\text{: MDD} + \beta_7 Gender\text{: SCZ}$$

To evaluate the relationship between different clinical features that may occur in a patient over the entire course of their EHR, we used the above modelling framework, but added as predictors, for each feature Sx_s, the lifetime presence of the three remaining features ($Sx_j\ ;\ j \neq s\ ;\ s, j \in \{1,2,3,4\}$). The resulting model is:

$$ln\left(\frac{P_{Sx_s}}{1-P_{Sx_s}}\right) \sim \beta_0 + \beta_1 MDD + \beta_2 SCZ + \beta_3 Gender + \ \beta_4 log_{10}(N\ notes) + \beta_5 Inpatient$$
$$+ \sum_{j=1}^{4} \gamma_{j\neq s} Sx_{j\neq s}$$

We fit four models in total – one for each feature. In each model, we tested for associations with gender and with the three remaining features (e.g., in the case of delusions, we tested for associations with suicide attempt, suicidal ideation, and hallucinations). This procedure results in a Bonferroni-corrected alpha of 0.05/16=0.0031.

Finally, to explore the interactions between gender and co-occurring features, we expanded the model to include an interaction term, as follows:

$$ln\left(\frac{P_{Sx_s}}{1-P_{Sx_s}}\right) \sim \beta_0 + \beta_1 MDD + \beta_2 SCZ + \beta_3 Gender + \ \beta_4 log_{10}(N\ notes) + \beta_5 Inpatient$$

$$+\sum\nolimits_{j=1}^{4} \gamma_{j\neq s} Sx_{j\neq s} + \sum\nolimits_{j=1}^{4} \alpha_{j\neq s} Gender{:}\, Sx_{j\neq s}$$

N.3 Supplementary Note 3. Definition of diagnostic trajectories and examples

To define a diagnostic trajectory, we map the progression of a patient's diagnoses over time using the sequence of ICD-10 codes extracted from their EHR. Starting with a patient's initial diagnosis, each subsequent visit contributes to this trajectory by either introducing a new diagnosis or indicating a switch from a previous diagnosis. Specifically, we followed these steps:

Record the diagnosis from the patient's first visit.

At every subsequent visit, add the new diagnosis to the patient’s cumulative record unless the diagnosis is already there.

If the diagnosis is found to be incompatible with a pre-existing one (as determined in Supplementary Table 4), the prior diagnosis is replaced, marking a diagnostic switch.

In the resulting trajectory, consecutive visits that do not introduce new diagnoses are condensed to avoid redundancy, ensuring that the final trajectory primarily represents either the acquisition of new comorbidities or diagnostic switches. This approach offers a concise and chronological representation of a patient's diagnostic journey over time. We provide below, two working examples of this procedure:

Consider the following sequence of ICD-10 diagnoses in the EHR of a patient with five visits.

```
F32 -> F32 -> F31 -> F31 -> F31
```

Here, the patient’s initial diagnosis of F32 (MDD) switched to F31 (BD) by the third visit. This trajectory incorporating the switch can be represented by:

```
F32 -> F31
```

Consider now the following sequence of ICD-10 diagnoses in the EHR of a patient with six visits.

```
F32 -> F41 -> F32 -> F32 -> F31 -> F31
```

Here, the patient had an initial diagnosis of F32 (MDD). On their second visit, they acquired the comorbidity of F41 (anxiety disorders) and finally the diagnosis switched from F32 to F31 (BD) on the fifth visit. This trajectory incorporating both comorbidity and a switch can be represented by:

```
F32 -> F32,F41 -> F31,F41
```

N.4 Supplementary Note 4. Prospective and retrospective diagnostic stability

To assess the stability over time of individual SMI diagnoses, we considered a diagnosis unstable if a patient with a given diagnosis switched to a different one. For these analyses we included only patients with 10 or more visits ($k_{Last} \geq 10$). We calculated the following stability metrics for three diagnoses, MDD, BD, SCZ, evaluating differences in stability across diagnoses using z-tests.

Prospective stability is the probability that a patient's first diagnosis is the same as their last one. Formally, it is defined as the proportion of patients with diagnosis *x* on their first visit ($ICD10_0^{Dx}$) that also have diagnosis *x* on their last visit ($ICD10_{k_{Last}}^{Dx}$)

$$P(ICD10_{k_{Last}}^{Dx} = x \mid ICD10_0^{Dx} = x\, ; k_{Last} \geq 10)$$

$$= \frac{\#\ of\ patients\ with\ (ICD10_0^{Dx} = x\ \cap\ ICD10_{k_{Last}}^{Dx} = x)}{\#\ of\ patients\ with\ (ICD10_0^{Dx} = x)}\ ; k_{Last} \geq 10$$

Retrospective stability is the probability that a patient's final diagnosis is the same as their first one. Formally, it is defined as the proportion of patients with diagnosis x on their last visit, who also had diagnosis x on their first one.

$$P(ICD10_0^{Dx} = x \mid ICD10_{k_{Last}}^{Dx} = x\, ; k_{Last} \geq 10)$$

$$= \frac{\#\ of\ patients\ with\ (ICD10_{0}^{Dx} = x\ \cap\ ICD10_{k_{Last}}^{Dx} = x)}{\#\ of\ patients\ with\ (ICD10_{k_{Last}}^{Dx} = x)}\ \ ; k_{Last}\ \geq 10$$

To test whether age is a driver of diagnostic instability, we split our cohort into two (approximately equally sized groups); those who were younger than 30 years of age at their first visit and those who were older than 30 at their first visit. We then conducted the above prospective and retrospective stability analyses separately for these two groups and used z-tests to compare the stability metrics, first across diagnoses and then across age groups for each individual diagnosis.

Prospective stability was lower for MDD compared to BD and SCZ in both age-at-first-visit groups (p-values: $2.9e^{-14}$ and $2.6e^{-15}$ respectively in <30, and p-values: $3.6e^{-68}$ and $2.2e^{-11}$ respectively in >30) Retrospective stability of MDD compared to BD and SCZ was greater in the younger age-at-first-visit group, but was not significantly different in the older age-at-first-visit group (p-values: $1.9e^{-4}$ and $5.9e^{-6}$ respectively in <30, and p-values: 0.5 and 0.7 respectively in >30). Overall, measures of stability are significantly lower in the younger age-at first-visit group (Supplementary Table 13; z-test p-values $<7e^{-5}$), with two exceptions: the prospective stabilities of SCZ (p-value= 0.08) and MDD (p-value= 0.01).

N.5 Supplementary Note 5. Factors contributing to visit-to-visit diagnostic stability

We used visit-level data to characterize the rate at which diagnoses stabilize over time and the factors that increase or decrease diagnostic stability.

To do this, we modeled a switch in diagnosis using the binary variable $Switch_{k+1}$. A value of 1 indicates that a patient's diagnosis at their next visit (*k+1)* is different from their current one:

$$Switch_{k+1}: ICD10_{k}^{Dx} \neq ICD10_{k+1}^{Dx}$$

We fit a mixed-effect logistic regression with the logit of the probability of a diagnostic switch in visit k+1 as the outcome and (log-transformed) visit number *k* as the predictor. We accounted for repeated observations of patients using random intercepts. We define $P_{Switch_{k+1,i}}$ as the probability of a diagnostic switch in visit k+1 for patient i. The resulting model is:

$$ln\left(\frac{P_{Switch_{k+1,i}}}{1 - P_{Switch_{k+1,i}}}\right) \sim \beta_0 + \beta_1 log_{10}(k_i) + u_{0i} + e_i$$

Where $u_{0i} \sim N(0, {\sigma_u}^2)$ is a random intercept with mean 0 and variance ${\sigma_u}^2$ and $e_i \sim N(0, {\sigma_e}^2)$ is residual error. In this framework, if a patient has *K* total visits, they contribute *K – 1* observations to the analysis, since their final visit doesn't have a $Switch_{k+1}$.

We used this flexible framework to understand the short-term stability of diagnoses. For this, we included dummy variables to indicate the ICD-10 diagnosis at visit k, with BD as the reference. Possible diagnoses were *MDD, SCZ, other*. We extended this model with four additional explanatory variables, two at the visit level and two at the patient level. At the visit level these explanatory variables are represented by two binary indicators: inpatient status ($Inpatient_k$) and an indicator representing an ICD-10 diagnosis of “Not Otherwise Specified” (NOS; $ICD10_k^{NOS}$). These were defined as all codes with the form FXX8 ("Other ...") or FXX9 ("Unspecified ..."), as well as those with the form FX8 and FX9 that are explicitly named "Other [...] Disorder" or "Unspecified [...] Disorder", respectively. At the patient level, we included gender and age at first visit. Additionally, we included a binary variable for each of the four clinical features ($Sx_{j,k}$, see Supplementary Note 2), indicating if they were present during the current visit.

The resulting model is:

$$ln\left(\frac{P_{Switch_{k+1,i}}}{1-P_{Switch_{k+1,i}}}\right) \sim \beta_0 + \beta_1 log_{10}(k_i) + \beta_2 MDD_{k,i} + \beta_3 SCZ_{k,i} + \beta_4 other_{k,i} + \beta_5 Inpatient_{k,i} + \beta_6 ICD10_{k,i}^{NOS} + \beta_7 Gender_i + \beta_8 Age_i + \sum_{j=1}^{4} \gamma_{j\neq s} Sx_{j\neq s,k,i} + u_{0i} + e_i$$

Finally, to assess the evidence of sustained diagnostic instability, we fit a second model where we include the switch indicator for the previous visit – $Switch_k$ – to estimate the effect of a previous switch on a future switch. In this analysis, each patient contributes *K– 2* observations to the analysis, since their first visit doesn’t have a $Switch_k$. This model is:

$$ln\left(\frac{P_{Switch_{k+1,i}}}{1-P_{Switch_{k+1,i}}}\right) \sim \beta_0 + \beta_1 log_{10}(k_i) + \ \beta_2 MDD_{k,i} + \beta_3 SCZ_{k,i} + \beta_4 other_{k,i} + \beta_5 Inpatient_{k,i} + \beta_6 ICD10_{k,i}^{NOS}$$

$$+ \beta_7 Gender_i + \beta_8 Age_i + \beta_9 Switch_{k,i} + \sum_{j=1}^{4} \gamma_{j\neq s} Sx_{j\neq s,k,i} + u_{0i} + e_i$$

For this section, we use a Bonferroni-corrected alpha of 0.05/13=0.0038.

N.6 Supplementary Note 6. Visit-level data validation: NLP features and ICD-10 diagnoses

For the major mood disorder diagnoses (MDD and BD), the 3-digit ICD codes recorded at each visit qualify an episode according to severity and according to the presence of absence of psychotic features. We used these qualifiers to evaluate the relationship between the information recorded in the codes with the clinical features extracted from the free-text hospital visit notes. Specifically, we tested whether the likelihood of extracting a clinical feature varied between visits labeled as being for severe episodes versus those labeled as being for mild or moderate episodes. Similarly, we tested whether the likelihood of extracting a psychotic feature (Delusions or Hallucinations) varied between visits labeled as being for severe episodes *with* psychotic symptoms versus those labeled as being for severe episodes *without* psychotic symptoms.

We employed a mixed-effect logistic regression to model the logit of the probability of the presence of a clinical feature. Specifically, $P_{Sx_{s,k,i}}$ represents the probability of feature *s* during visit *k* of patient *i*. We account for repeated observations on a patient with a random intercept for every individual.

The clinical features that we investigated are more frequently found in the notes from inpatient visits than in notes from outpatient or emergency department visits; this observation likely reflects not only the increased severity of symptoms experienced by individuals in association with inpatient hospitalization, but also the larger number of notes recorded during an inpatient stay. To account for the latter factor, we adjust for the binary variable $Inpatient_k$.

First, we tested for differences between two levels of severity using the binary variable $ICD10_k^{severe}$. Severe episodes are represented by codes: F301, F302, F311, F312, F314, F315, F322, F323, F332, or F333. Mild or moderate episodes are denoted by codes: F300, F310, F313, F320, F321, F330, or F331. The corresponding mixed model is:

$$ln\left(\frac{P_{Sx_{s,k,i}}}{1-P_{Sx_{s,k,i}}}\right) \sim \beta_0 + \beta_1 ICD10_{k,i}^{severe} + \beta_2 Inpatient_{k,i} + u_{0i} + e_i$$

Where $u_{0i} \sim N(0, {\sigma_u}^2)$ is a random intercept with mean 0 and variance ${\sigma_u}^2$ and $e_i \sim N(0, {\sigma_e}^2)$ is residual error.

Second, we focused on the participants with codes for severe episodes and tested for differences between episodes with and without psychotic symptoms with the binary variable

$ICD10_k^{psychosis}$. Episodes with psychotic symptoms are represented by codes: F302, F312, F315, F323, or F333. Episodes without psychotic symptoms are denoted by codes: F301, F311, F314, or F322. The corresponding mixed model is:

$$ln\left(\frac{P_{Sx_{s,k,i}}}{1-P_{Sx_{s,k,i}}}\right) \sim \beta_0 + \beta_1 ICD10_{k,i}^{psychosis} + \beta_2 Inpatient_{k,i} + u_{0i} + e_i$$

In total, we fit six models – four in the first group of tests (all clinical features) and two in the second (Delusions and Hallucinations only). Considering the six tests, we applied a Bonferroni-corrected alpha of 0.0083 (0.05/6). After removing patients with only one visit to avoid issues of model convergence, we were left with an N of 47,186 visits in 9,203 people for the first model and 15,120 visits in 4,075 people for the second.

Prediction of diagnostic codes from NLP-extracted clinical features:

After examining the relationship between clinical features and ICD-10 codes within a single visit, we explored if these features could also forecast future diagnostic codes. Our objective was to assess if clinical features extracted from free-text during a patient's current visit (k) can predict ICD-10 codes for the subsequent visit (k+1).

In this model, we defined a new clinical feature, "Psychosis" (Sx_p), to include the presence of either Delusions or Hallucinations. Then, we evaluated whether the presence of this feature during a visit *k* has a correlation with the logit of the probability of an ICD-10 code indicative of an episode *with* psychotic symptoms in the following visit *k+1*, using a logistic regression.

To fit this model, we adjusted for the presence of psychosis in visit k ($ICD10_k^{psychosis}$, defined as above), the clinical feature Sx_p on visit k+1 ($Sx_{p,k+1}$), and the inpatient status in both visits ($Inpatient_k$ and $Inpatient_{k+1}$). Let $P_{ICD10_{k+1,i}^{psychosis}}$ be the probability that patient *i* has a psychosis ICD-10 code at visit *k+1*. Finally, we account for repeated observations on a patient with a random intercept as above. The resulting model is:

$$ln\left(\frac{P_{ICD10_{k+1,i}^{psychosis}}}{1-P_{ICD10_{k+1,i}^{psychosis}}}\right) \sim \beta_0 + \beta_1 Sx_{p,k,i}$$

$$+ \beta_2 ICD10_{k,i}^{psychosis} + \beta_3 Sx_{p,k+1,i} + \beta_4 Inpatient_{k,i} + \beta_5 Inpatient_{k+1,i} + u_{0i} + e_i$$

N.7 Supplementary Note 7. Evaluation of errors in the gold standard.

Considering both false positives and false negatives, the NLP algorithm failed in 28 instances of the gold standard. These errors happened in both the NER step and in the ND step.

Most errors occurred in the NER step (25/28) and were due to missing search patterns. We identified three reasons for missing patterns: 1) the text contains spelling errors that were not observed in the development set; 2) the specific pattern did not appear in the development set; and 3) the pattern was observed in the development set but was removed because it was not specific enough and would have generated numerous false positives.

Three false positives were caused by failure to identify the negation in the sentence. In two of these instances, the feature was part of a list of negated terms and was located beyond the scope of the negation cue (five tokens). In the last instance, the error was caused by a spelling error in the negation cue that was missing from the development set.

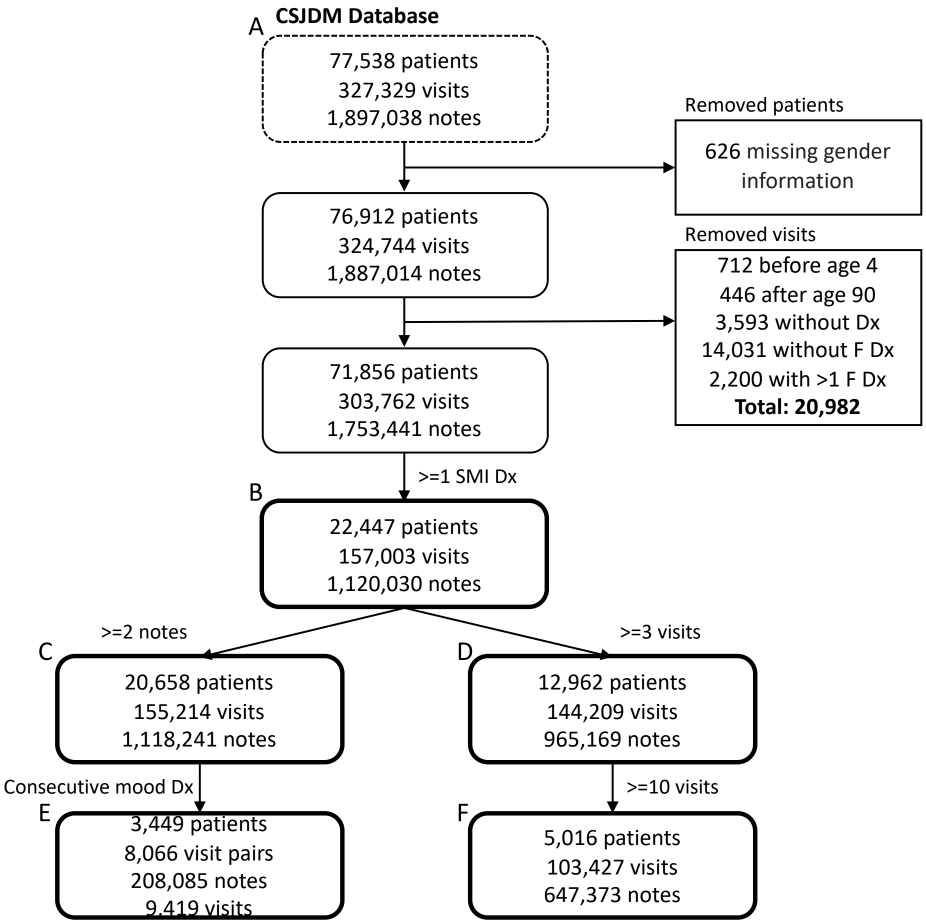
N.8 Supplementary Note 8. Suicidality and psychosis as distinct dimensions of severity in MDD

When comparing the frequency of clinical features across different episodes in MDD, we came across an interesting observation. Visits labeled with the ICD-10 code for severe episodes with psychosis (F323/F333) were less likely to contain the NLP extracted clinical features of Suicide Attempt (OR=0.49, p-value $2e^{-8}$) and Suicidal Ideation (OR 0.43, p-value $4e^{-6}$) compared to visits labeled as severe episodes without psychosis (F322/F332).

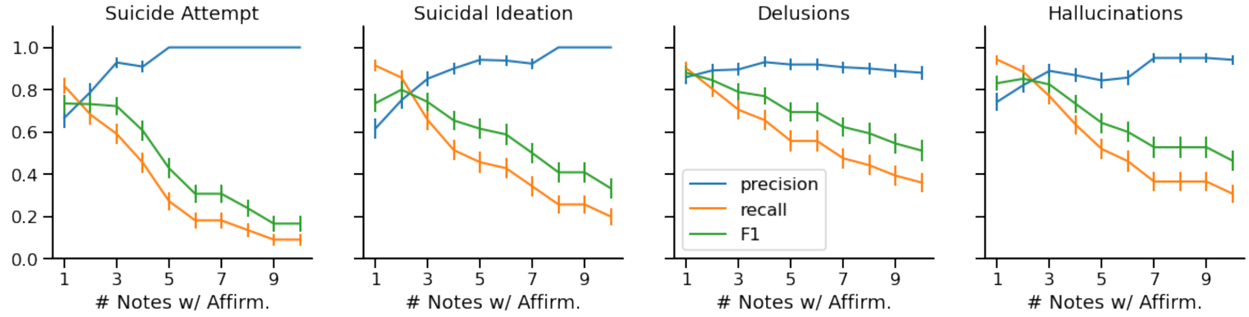
Rather than reflecting disease pathology, this may be a consequence of analyzing visits rather than patients as the primary unit of analysis: psychosis and suicidality being two main dimensions of care-seeking for patients with severe MDD. Consequently, this may constitute an artifact of phenotyping clinical features based on the EHR.

SUPPLEMENTARY FIGURES

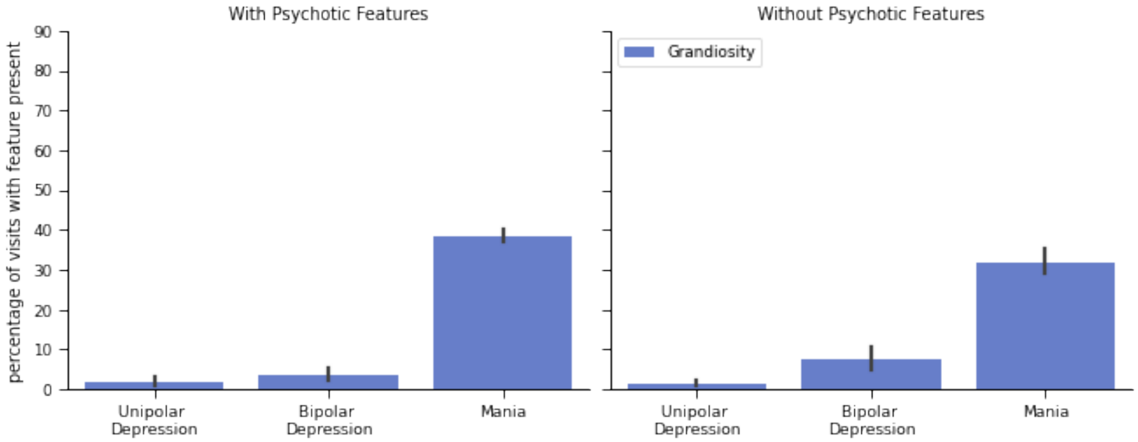
F.1 Supplementary Figure 1. *Flow diagram of sample selection from the CSJDM EHR database indicating: (A) the steps used to remove patients and patient visits not meeting criteria for any of our analyses; (B) the complete SMI cohort from which we selected subsets for different analyses as described in the Methods; (C) the cohort used for evaluating patient-level association between clinical features and ICD-10 diagnoses; (D) the cohort used for the trajectory analyses exploring diagnostic switches and comorbidities (E) the cohort used to test if clinical features identified at one visit anticipate changes in ICD-10 codes at the subsequent visit; (F) the cohort used for estimating long-term diagnostic stability*



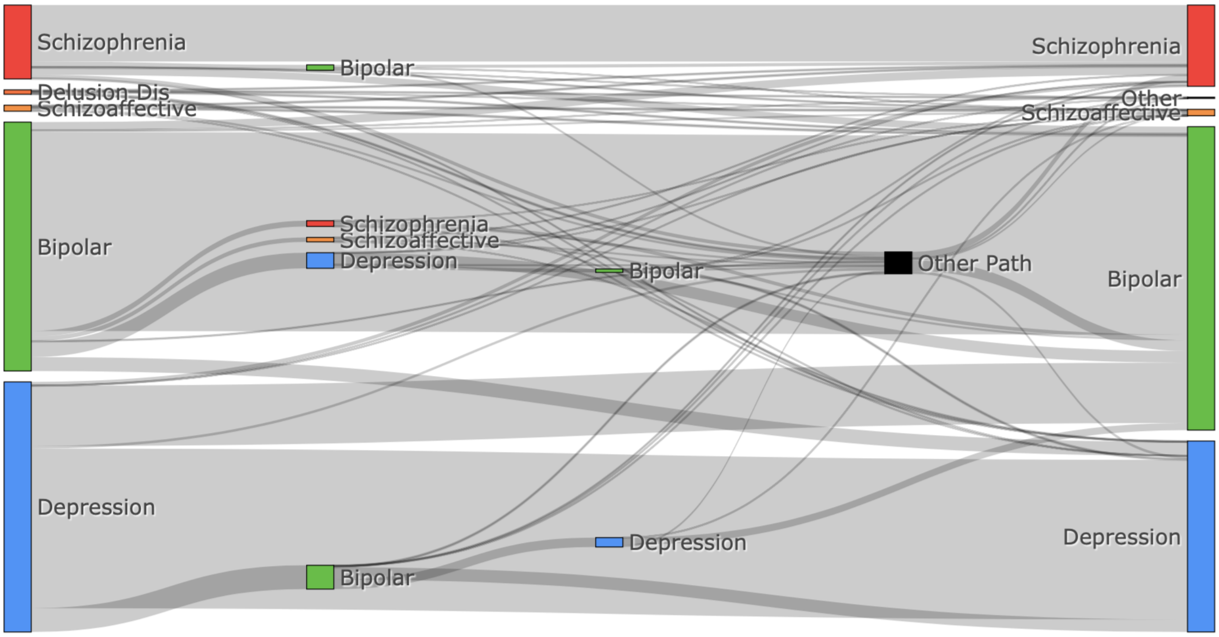
F.2 Supplementary Figure 2. *The performance of the NLP algorithm at different thresholds for the number of affirmative mentions required to classify patients as positives or negatives for each clinical feature. We considered all thresholds between one to ≥10 affirmative mentions per patient; we could only evaluate such mentions if a patient had at least that number of different notes in their EHR, and therefore the sample size of patients who could be evaluated decreases with increasing thresholds (from 105 patients with ≥ one note to 88 with ≥ 10 or more notes). At each threshold we evaluated the performance of the algorithm in terms of precision, recall and F1. We selected a threshold of ≥ two affirmative mentions to designate a patient as positive for a clinical feature, as this threshold (evaluated in the119 patients with at ≥ two notes) yields the highest F1 across the four features.*



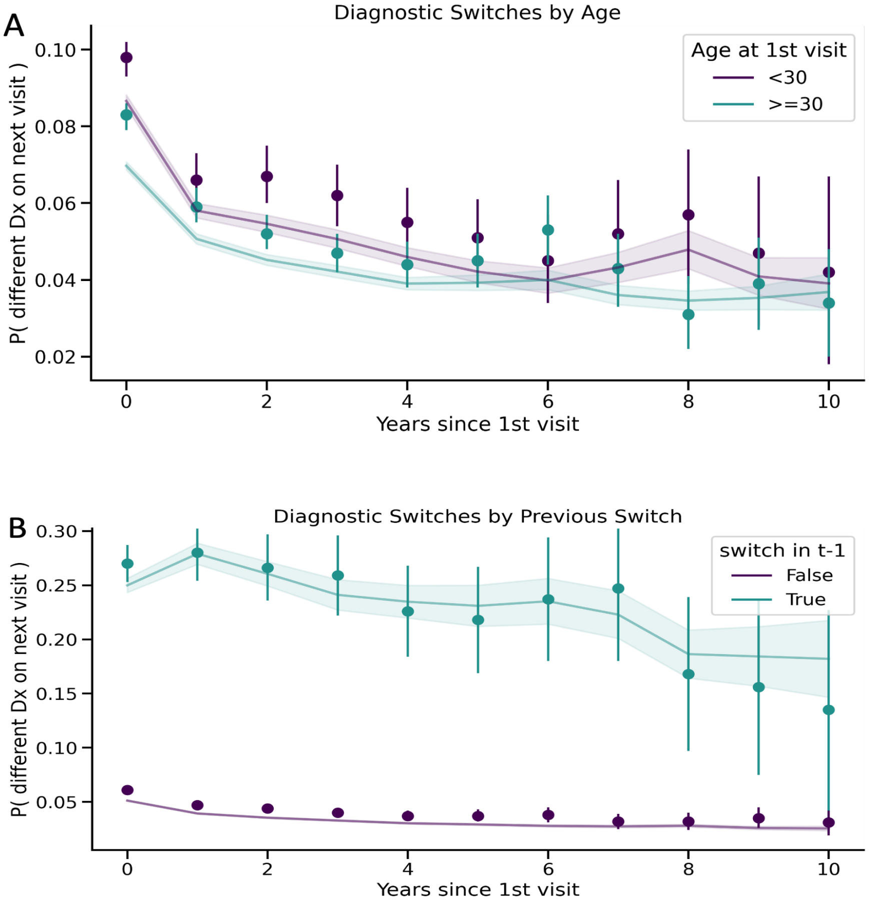
F.3 Supplementary Figure 3. *The proportion of delusions representing grandiosity varies by diagnostic code. The overall proportion is 18%. Specific search patterns were: "Grandiosi", "Grandeza" and "Megaloma".*



F.4 Supplementary Figure 4. *Sankey diagram of ICD-10 code trajectories. The figure shows switches between SMI diagnoses in patients with 3 or more visits (n=12,962).*



F.5 Supplementary Figure 5. *Diagnostic stability over time. Using time since the first encounter instead of visit number. For every year, the observed proportion of visits that will have a diagnostic switch on the next visit is plotted as a dot with 95% confidence intervals. The solid line is the average probability of switching at any given visit during that year, as estimated by the model. The corresponding shaded area is the 95% confidence interval. A) Patients are stratified by the age of their first visit: before and after 30 years. B) Visits are stratified by having switched diagnoses from visit k-1. N=12,962 patients.*

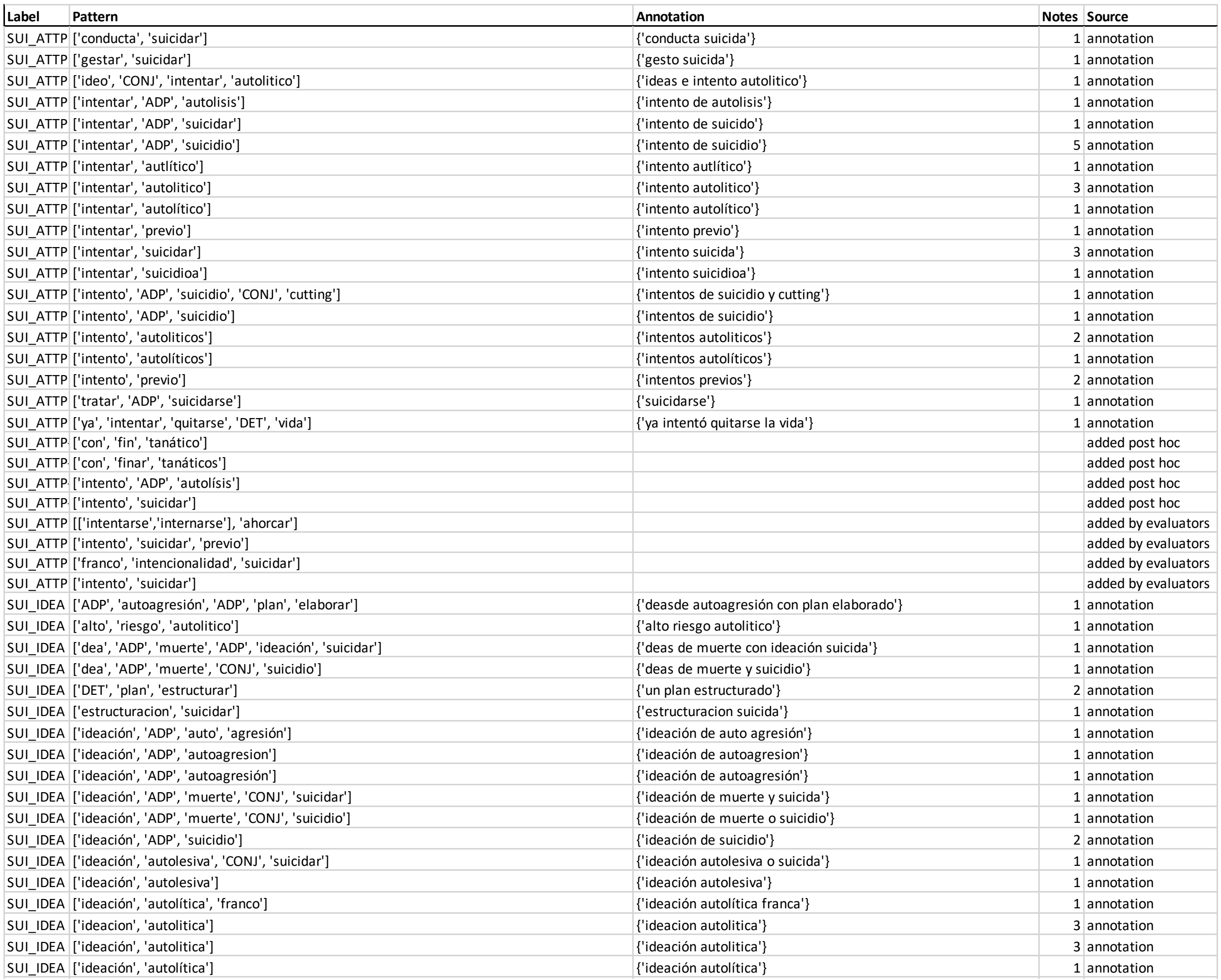


SUPPLEMENTARY TABLES

T.1 Supplementary Table 6. *ICD-10 codes used in this study. Codes meeting our definition of SMI are marked with *.*

Dx	ICD-10	Diagnosis	Specifier	SMI
SCZ	F200	Schizophrenia	paranoid schizophrenia	*
SCZ	F201	Schizophrenia	hebephrenic schizophrenia	*
SCZ	F202	Schizophrenia	catatonic schizophrenia	*
SCZ	F203	Schizophrenia	undifferentiated schizophrenia	*
SCZ	F204	Schizophrenia	post-schizophrenic depression	*
SCZ	F205	Schizophrenia	residual schizophrenia	*
SCZ	F206	Schizophrenia	simple schizophrenia	*
SCZ	F208	Schizophrenia	other schizophrenia	*
SCZ	F209	Schizophrenia	unspecified	*
Psych	F25X	Persistent delusional disorders		*
Psych	F25X	Schizoaffective Disorder		*
BD	F301	Manic episode	mania without psychotic symptoms	*
BD	F302	Manic episode	mania with psychotic symptoms	*
BD	F310	Bipolar disorder	current episode hypomanic	*
BD	F311	Bipolar affective disorder	current episode manic without psychotic symptoms	*
BD	F312	Bipolar affective disorder	current episode manic with psychotic symptoms	*
BD	F313	Bipolar affective disorder	current episode mild or moderate depression	*
BD	F314	Bipolar affective disorder	current episode severe depression without psychotic symptoms	*
BD	F315	Bipolar affective disorder	current episode severe depression with psychotic symptoms	*
BD	F316	Bipolar affective disorder	current episode mixed	*
BD	F317	Bipolar affective disorder	currently in remission	*
BD	F318	Bipolar affective disorder	other bipolar affective disorders	
BD	F319	Bipolar affective disorder	unspecified	
MDD	F320	Depressive episode	mild depressive episode	
MDD	F321	Depressive episode	moderate depressive episode	
MDD	F322	Depressive episode	severe depressive episode without psychotic symptoms	*
MDD	F323	Depressive episode	severe depressive episode with psychotic symptoms	*
MDD	F328	Depressive episode	other depressive episodes	
MDD	F329	Depressive episode	depressive episode, unspecified	
MDD	F330	Recurrent depressive disorder	current episode mild	
MDD	F331	Recurrent depressive disorder	current episode moderate	*
MDD	F332	Recurrent depressive disorder	current episode severe without psychotic symptoms	*
MDD	F333	Recurrent depressive disorder	current episode severe with psychotic symptoms	*
MDD	F334	Recurrent depressive disorder	currently in remission	*
MDD	F338	Recurrent depressive disorder	other recurrent depressive disorders	
MDD	F339	Recurrent depressive disorder	unspecified	

T.2 Supplementary Table 2. *Patterns that the NLP algorithm uses for identifying clinical features in the notes. Label: label used for annotating clinical features. SUI_ATTP: Suicide Attempt, SUI_IDEA: Suicidal Ideation, DEL: Delusions, HAL: Hallucinations; Pattern: sequence of tokens used by the EntityRuler component of Spacy to perform Named Entity Recognition; Annotation: specific span of text highlighted by the annotators and used to generate the pattern (as described in Supplementary Note 1); Notes: number of notes represented by the same annotation; Source: indicates if the pattern derives from the annotation or from the subsequent curation process*



Label	Pattern	Annotation	Notes	Source
SUI_ATTP	['conducta', 'suicidar']	{'conducta suicida'}	1	annotation
SUI_ATTP	['gestar', 'suicidar']	{'gesto suicida'}	1	annotation
SUI_ATTP	['ideo', 'CONJ', 'intentar', 'autolitico']	{'ideas e intento autolitico'}	1	annotation
SUI_ATTP	['intentar', 'ADP', 'autolisis']	{'intento de autolisis'}	1	annotation
SUI_ATTP	['intentar', 'ADP', 'suicidar']	{'intento de suicido'}	1	annotation
SUI_ATTP	['intentar', 'ADP', 'suicidio']	{'intento de suicidio'}	5	annotation
SUI_ATTP	['intentar', 'autlítico']	{'intento autlítico'}	1	annotation
SUI_ATTP	['intentar', 'autolitico']	{'intento autolitico'}	3	annotation
SUI_ATTP	['intentar', 'autolítico']	{'intento autolítico'}	1	annotation
SUI_ATTP	['intentar', 'previo']	{'intento previo'}	1	annotation
SUI_ATTP	['intentar', 'suicidar']	{'intento suicida'}	3	annotation
SUI_ATTP	['intentar', 'suicidioa']	{'intento suicidioa'}	1	annotation
SUI_ATTP	['intento', 'ADP', 'suicidio', 'CONJ', 'cutting']	{'intentos de suicidio y cutting'}	1	annotation
SUI_ATTP	['intento', 'ADP', 'suicidio']	{'intentos de suicidio'}	1	annotation
SUI_ATTP	['intento', 'autoliticos']	{'intentos autoliticos'}	2	annotation
SUI_ATTP	['intento', 'autolíticos']	{'intentos autolíticos'}	1	annotation
SUI_ATTP	['intento', 'previo']	{'intentos previos'}	2	annotation
SUI_ATTP	['tratar', 'ADP', 'suicidarse']	{'suicidarse'}	1	annotation
SUI_ATTP	['ya', 'intentar', 'quitarse', 'DET', 'vida']	{'ya intentó quitarse la vida'}	1	annotation
SUI_ATTP	['con', 'fin', 'tanático']			added post hoc
SUI_ATTP	['con', 'finar', 'tanáticos']			added post hoc
SUI_ATTP	['intento', 'ADP', 'autolísis']			added post hoc
SUI_ATTP	['intento', 'suicidar']			added post hoc
SUI_ATTP	[['intentarse','internarse'], 'ahorcar']			added by evaluators
SUI_ATTP	['intento', 'suicidar', 'previo']			added by evaluators
SUI_ATTP	['franco', 'intencionalidad', 'suicidar']			added by evaluators
SUI_ATTP	['intento', 'suicidar']			added by evaluators
SUI_IDEA	['ADP', 'autoagresión', 'ADP', 'plan', 'elaborar']	{'deasde autoagresión con plan elaborado'}	1	annotation
SUI_IDEA	['alto', 'riesgo', 'autolitico']	{'alto riesgo autolitico'}	1	annotation
SUI_IDEA	['dea', 'ADP', 'muerte', 'ADP', 'ideación', 'suicidar']	{'deas de muerte con ideación suicida'}	1	annotation
SUI_IDEA	['dea', 'ADP', 'muerte', 'CONJ', 'suicidio']	{'deas de muerte y suicidio'}	1	annotation
SUI_IDEA	['DET', 'plan', 'estructurar']	{'un plan estructurado'}	2	annotation
SUI_IDEA	['estructuracion', 'suicidar']	{'estructuracion suicida'}	1	annotation
SUI_IDEA	['ideación', 'ADP', 'auto', 'agresión']	{'ideación de auto agresión'}	1	annotation
SUI_IDEA	['ideación', 'ADP', 'autoagresion']	{'ideación de autoagresion'}	1	annotation
SUI_IDEA	['ideación', 'ADP', 'autoagresión']	{'ideación de autoagresión'}	1	annotation
SUI_IDEA	['ideación', 'ADP', 'muerte', 'CONJ', 'suicidar']	{'ideación de muerte y suicida'}	1	annotation
SUI_IDEA	['ideación', 'ADP', 'muerte', 'CONJ', 'suicidio']	{'ideación de muerte o suicidio'}	1	annotation
SUI_IDEA	['ideación', 'ADP', 'suicidio']	{'ideación de suicidio'}	2	annotation
SUI_IDEA	['ideación', 'autolesiva', 'CONJ', 'suicidar']	{'ideación autolesiva o suicida'}	1	annotation
SUI_IDEA	['ideación', 'autolesiva']	{'ideación autolesiva'}	1	annotation
SUI_IDEA	['ideación', 'autolítica', 'franco']	{'ideación autolítica franca'}	1	annotation
SUI_IDEA	['ideacion', 'autolitica']	{'ideacion autolitica'}	3	annotation
SUI_IDEA	['ideación', 'autolitica']	{'ideación autolitica'}	3	annotation
SUI_IDEA	['ideación', 'autolítica']	{'ideación autolítica'}	1	annotation

SUI_IDEA	['ideacion', 'suicidar', 'ADP', 'plan', 'parcialmente', 'estructurar']	{'ideacion suicida con plan parcialmente estructurado'}	1	annotation
SUI_IDEA	['ideación', 'suicidar', 'ADP', 'plan', 'parcialmente', 'estructurar']	{'ideación suicida con plan parcialmente estructurado'}	2	annotation
SUI_IDEA	['ideacion', 'suicidar', 'estructurar']	{'ideacion suicida estructurada'}	1	annotation
SUI_IDEA	['ideación', 'suicidar', 'estructurar']	{'ideación suicida estructurada'}	1	annotation
SUI_IDEA	['ideacion', 'suicidar', 'no', 'estructurar']	{'ideacion suicida no estructurada'}	1	annotation
SUI_IDEA	['ideación', 'suicidar', 'no', 'estructurar']	{'ideación suicida no estructurada', 'ideación suicida no estructuradas'}	2	annotation
SUI_IDEA	['ideacion', 'suicidar', 'poco', 'estructurar']	{'ideacion suicida poco estructurada'}	1	annotation
SUI_IDEA	['ideacion', 'suicidar']	{'ideacion suicida'}	31	annotation
SUI_IDEA	['ideación', 'suicidar']	{'ideación suicida'}	26	annotation
SUI_IDEA	['ideación', 'suicidia', 'estructurar']	{'ideación suicidia estructurada'}	1	annotation
SUI_IDEA	['ideacion', 'suicidia']	{'ideacion suicidia'}	1	annotation
SUI_IDEA	['ideación', 'suicidia']	{'ideación suicidia'}	2	annotation
SUI_IDEA	['ideo', 'ADP', 'ADP', 'muerte', 'ADP', 'plan', 'estructurar']	{'ideas de de muerte con plan estructurado'}	1	annotation
SUI_IDEA	['ideo', 'ADP', 'auto', 'agresión']	{'ideas de auto agresión'}	1	annotation
SUI_IDEA	['ideo', 'ADP', 'auto', 'PUNCT', 'agresión']	{'ideas de auto-agresión'}	1	annotation
SUI_IDEA	['ideo', 'ADP', 'auto']	{'ideas de auto'}	5	annotation
SUI_IDEA	['ideo', 'ADP', 'autoagresión']	{'ideas de autoagresión'}	9	annotation
SUI_IDEA	['ideo', 'ADP', 'muerte', 'ADP', 'plan', 'estructurar']	{'ideas de muerte con plan estructurado'}	3	annotation
SUI_IDEA	['ideo', 'ADP', 'muerte', 'CONJ', 'ADP', 'suicidio', 'estructurar']	{'ideas de muerte y de suicidio estructuradas'}	1	annotation
SUI_IDEA	['ideo', 'ADP', 'muerte', 'CONJ', 'ADP', 'suicidio']	{'ideas de muerte y de suicidio', 'ideas de muerte o de suicidio'}	8	annotation
SUI_IDEA	['ideo', 'ADP', 'muerte', 'CONJ', 'suicidar']	{'ideas de muerte y suicidas', 'ideas de muerte o suicida', 'ideas de muerte o suicidas'}	4	annotation
SUI_IDEA	['ideo', 'ADP', 'muerte', 'CONJ', 'suicidio', 'ADP', 'plan', 'ADP', 'estructuración']	{'ideas de muerte y suicidio con plan en estructuración'}	1	annotation
SUI_IDEA	['ideo', 'ADP', 'muerte', 'CONJ', 'suicidio', 'ADP', 'plan', 'estructurar']	{'ideas de muerte y suicidio con plan estructurado', 'ideas de muerte o suicidio con plan estructurado'}	2	annotation
SUI_IDEA	['ideo', 'ADP', 'muerte', 'CONJ', 'suicidio', 'estructurar']	{'ideas de muerte y suicidio estructuradas'}	3	annotation
SUI_IDEA	['ideo', 'ADP', 'muerte', 'CONJ', 'suicidio', 'no', 'estructurar']	{'ideas de muerte y suicidio no estructuradas'}	1	annotation
SUI_IDEA	['ideo', 'ADP', 'muerte', 'CONJ', 'suicidio']	{'ideas de muerte o suicidios', 'ideas de muerte o suicidio', 'ideas de muerte y suicidio'}	50	annotation
SUI_IDEA	['ideo', 'ADP', 'muerte', 'CONJ', 'suicidio']	{'no ideas de muerte o suicidio'}	1	annotation
SUI_IDEA	['ideo', 'ADP', 'muerte', 'plan', 'estructurar']	{'ideas de muerte plan estructurado'}	1	annotation
SUI_IDEA	['ideo', 'ADP', 'sicidio']	{'ideas de sicidio'}	1	annotation
SUI_IDEA	['ideo', 'ADP', 'suicidio', 'estructurar']	{'ideas de suicidio estructuradas'}	2	annotation
SUI_IDEA	['ideo', 'ADP', 'suicidio']	{'idea de suicidio', 'ideas de suicidio'}	39	annotation
SUI_IDEA	['ideo', 'autolíticas']	{'ideas autolíticas'}	2	annotation
SUI_IDEA	['ideo', 'claro', 'ADP', 'suicidio']	{'ideas claras de suicidio'}	1	annotation
SUI_IDEA	['ideo', 'plan', 'CONJ', 'intencion', 'autolesivo', 'CONJ', 'suicidar']	{'ideas plan o intencion autolesivo o suicida'}	1	annotation
SUI_IDEA	['ideo', 'suicidar']	{'idea suicidas', 'ideas suicidas'}	12	annotation
SUI_IDEA	['ideo', 'suicidias']	{'ideas suicidias'}	1	annotation
SUI_IDEA	['ideo', 'tanaticas', 'CONJ', 'suicidar']	{'ideas tanaticas y suicidas'}	3	annotation
SUI_IDEA	['ideo', 'tantico', 'CONJ', 'suicidar']	{'ideas tanticas y suicidas'}	1	annotation
SUI_IDEA	['intención', 'ADP', 'matarme']	{'intención para matarme'}	1	annotation
SUI_IDEA	['intencion', 'suicidar']	{'intencion suicida'}	1	annotation
SUI_IDEA	['intención', 'suicidar']	{'intención suicida'}	2	annotation
SUI_IDEA	['matarme']	{'matarme'}	1	annotation
SUI_IDEA	['pensamiento', 'ADP', 'muerte', 'CONJ', 'ideacion', 'suicidar']	{'pensamientos de muerte o ideacion suicida'}	1	annotation
SUI_IDEA	['pensamiento', 'ADP', 'muerte', 'CONJ', 'suicidio']	{'pensamiento de muerte o suicidio'}	1	annotation
SUI_IDEA	['pensar', 'ADP', 'quitarme', 'DET', 'vida']	{'pensar en quitarme la vida'}	1	annotation
SUI_IDEA	['plan', 'ADP', 'suicidio']	{'plan de suicidio'}	1	annotation
SUI_IDEA	['plan', 'autolitico', 'estructurar']	{'plan autolitico estructurado'}	1	annotation
SUI_IDEA	['plan', 'autolítico', 'estructurar']	{'plan autolítico estructurado'}	2	annotation
SUI_IDEA	['plan', 'autolitico']	{'plan autolitico'}	1	annotation
SUI_IDEA	['plan', 'estructurar', 'ADP', 'mismo']	{'plan estructurado del mismo'}	1	annotation
SUI_IDEA	['plan', 'estructurar', 'ADP', 'suicidio']	{'plan estructurado de suicidio'}	2	annotation
SUI_IDEA	['plan', 'estructurar']	{'plan estructurado'}	3	annotation
SUI_IDEA	['plan', 'suicidar', 'estructurar']	{'plan suicida estructurado'}	2	annotation
SUI_IDEA	['plan', 'suicidar', 'poco', 'estructurar']	{'planes suicidas poco estructurados'}	1	annotation
SUI_IDEA	['plan', 'suicidar']	{'planes suicidas'}	1	annotation
SUI_IDEA	['plan', 'suicidia']	{'plan suicidia'}	1	annotation
SUI_IDEA	['riesgo', 'ADP', 'suicidio']	{'riesgo de suicidio'}	2	annotation
SUI_IDEA	['riesgo', 'suicidar']	{'riesgo suicida'}	3	annotation
SUI_IDEA	['suicidar', 'estructurar']	{'suicidas estructuras'}	1	annotation
SUI_IDEA	['suicidio', 'estructurar']	{'suicidio estructuradas'}	1	annotation
SUI_IDEA-	['ideo', 'ADP', 'muerte', 'ADP', 'plan', 'esturado']			added post hoc
SUI_IDEA-	['ideo', 'suicidio']			added post hoc
SUI_IDEA-	['querer', 'suicidarme']			added post hoc

Label	Pattern	Annotation	Notes	Source
DEL	['acrividad', 'delirante']	{'acrividad delirante'}	1	annotation
DEL	['activiad', 'delirante']	{'activiad delirante'}	1	annotation
DEL	['actividad', 'delirante', 'ADP', 'tipo', 'megalomaniaco']	{'actividad delirante de tipo megalomaniaco'}	1	annotation
DEL	['actividad', 'delirante']	{'actividad delirante'}	10	annotation
DEL	['ADP', 'actividad', 'delirante']	{'con actividad delirante del núcleo del daño "vea que me tengo que esconder un vigilante ayer si me estaba echando ojo'}	1	annotation
DEL	['celotipia']	{'celotipia'}	1	annotation
DEL	['compromiso', 'delirante']	{'compromiso delirante'}	1	annotation
DEL	['dea', 'delirante', 'CONJ', 'sobrevalorar', 'ADP', 'grandiosidad', 'CONJ', 'misticas']	{'deas delirantes y sobrevaloradas de grandiosidad y misticas'}	1	annotation
DEL	['dea', 'delirante']	{'deas delirantes'}	1	annotation
DEL	['deirio', 'ADP', 'grandiosidad']	{'deirio de grandiosidad'}	1	annotation
DEL	['delirante']	{'delirantes', 'delirante'}	10	annotation
DEL	['delirio', 'ADP', 'grandiosidad']	{'delirios de grandiosidad'}	1	annotation
DEL	['delirio', 'ADP', 'persecucion']	{'delirios de persecucion'}	1	annotation
DEL	['delirio', 'ADP', 'persecución']	{'delirios de persecución'}	1	annotation
DEL	['delirio', 'ADP', 'tipo', 'persecutorio']	{'delirios de tipo persecutorio'}	1	annotation
DEL	['delirio', 'cronificado']	{'delirios cronificado'}	1	annotation
DEL	['delirio', 'cronificados']	{'delirios cronificados'}	1	annotation
DEL	['delirio', 'estructurar']	{'delirios estructurados', 'delirios estructuradas'}	2	annotation
DEL	['delirio', 'florido']	{'delirios floridos'}	1	annotation
DEL	['delirio', 'grandioso']	{'delirios grandiosos'}	1	annotation
DEL	['delirio', 'persecutorio', 'paranodides']	{'delirios persecutorios paranodides'}	1	annotation
DEL	['delirio', 'persecutorio', 'paranoides', 'poco', 'estructurar']	{'delirios persecutorios paranoides poco estructurados'}	1	annotation
DEL	['delirio', 'persecutorio']	{'delirios persecutorios'}	1	annotation
DEL	['delirio', 'somaticos']	{'delirios somaticos'}	1	annotation
DEL	['delirio']	{'delirio', 'delirios'}	8	annotation
DEL	['delrios', 'persecutorio']	{'delrios persecutorios'}	1	annotation
DEL	['distorsionar', 'cognitivo', 'ADP', 'tipo', 'religioso']	{'distorsiones cognitivas de tipo religioso'}	1	annotation
DEL	['elemento', 'místico', 'delirante']	{'elementos místicos delirantes'}	1	annotation
DEL	['elemento', 'misticos', 'delirante']	{'elementos misticos delirantes'}	1	annotation
DEL	['expresion', 'delirante']	{'expresion delirante'}	1	annotation
DEL	['expresión', 'delirante']	{'expresión delirante'}	1	annotation
DEL	['ideacion', 'deilrante']	{'ideacion deilrante'}	1	annotation
DEL	['ideación', 'delirante', 'ADP', 'característico', 'megalomaníaco']	{'ideación delirante de características megalomaníacas'}	1	annotation
DEL	['ideación', 'delirante', 'ADP', 'característico', 'paranoides']	{'ideación delirante de características paranoides'}	1	annotation
DEL	['ideación', 'delirante', 'ADP', 'dañar']	{'ideación delirante de daño referencial al diablo'}	1	annotation
DEL	['ideación', 'delirante', 'ADP', 'grandeza']	{'ideación delirante de grandeza'}	2	annotation
DEL	['ideacion', 'delirante', 'ADP', 'grandiosidad']	{'ideacion delirante de grandiosidad'}	1	annotation
DEL	['ideación', 'delirante', 'ADP', 'magalomania']	{'ideación delirante de magalomania'}	1	annotation
DEL	['ideacion', 'delirante', 'ADP', 'nucleo', 'ADP', 'dañar']	{'ideacion delirante de nucleo de daño'}	1	annotation
DEL	['ideación', 'delirante', 'ADP', 'nucleo', 'ADP', 'dañar']	{'ideación delirante del nucleo del daño'}	3	annotation
DEL	['ideacion', 'delirante', 'ADP', 'nucleo', 'dela', 'dañar']	{'ideacion delirante del nucleo dela daño'}	1	annotation
DEL	['ideacion', 'delirante', 'ADP', 'tipo', 'somatico']	{'ideacion delirante de tipo somatico'}	1	annotation
DEL	['ideacion', 'delirante', 'grandioso']	{'ideacion delirante grandiosa'}	1	annotation
DEL	['ideación', 'delirante', 'megalomana', 'ADP', 'riqueza']	{'ideación delirante megalomana de riqueza'}	1	annotation
DEL	['ideación', 'delirante', 'paranoide']	{'ideación delirante paranoide'}	1	annotation
DEL	['ideacion', 'delirante', 'referencial', 'CONJ', 'persecutorio']	{'ideacion delirante referencial y persecutoria'}	1	annotation
DEL	['ideacion', 'delirante']	{'ideacion delirante'}	13	annotation
DEL	['ideación', 'delirante']	{'ideación delirante'}	21	annotation
DEL	['ideo', 'ADP', 'contener', 'delirante']	{'ideas de contenido delirante'}	1	annotation
DEL	['ideo', 'ADP', 'dañar', 'CONJ', 'ADP', 'persecución']	{'ideas de daño y de persecución'}	1	annotation
DEL	['ideo', 'ADP', 'grandiosidad']	{'ideas de grandiosidad'}	1	annotation
DEL	['ideo', 'ADP', 'nucleo', 'ADP', 'dañar']	{'ideas de nucleo de daño'}	1	annotation
DEL	['ideo', 'ADP', 'referenciar']	{'ideas de referencia'}	1	annotation
DEL	['ideo', 'deiirantes']	{'ideas deiirantes'}	1	annotation
DEL	['ideo', 'deirantes']	{'ideas deirantes'}	1	annotation
DEL	['ideo', 'delirante', 'ADP', 'celotipia']	{'ideas delirante de celotipia'}	2	annotation
DEL	['ideo', 'delirante', 'ADP', 'comandar', 'CONJ', 'ADP', 'nucleo', 'ADP', 'dañar']	{'ideas delirantes de comando y del nucleo de daño'}	1	annotation
DEL	['ideo', 'delirante', 'ADP', 'comandar']	{'ideas delirantes de comando'}	1	annotation
DEL	['ideo', 'delirante', 'ADP', 'contener', 'grandioso', 'CONJ', 'persecutorio']	{'ideas delirantes de contenido grandioso y persecutorias'}	1	annotation
DEL	['ideo', 'delirante', 'ADP', 'contener', 'grandioso', 'CONJ', 'persucotorias']	{'ideas delirantes de contenido grandioso y persucotorias'}	1	annotation
DEL	['ideo', 'delirante', 'ADP', 'contener', 'magico', 'PUNCT', 'religioso']	{'ideas delirantes de contenido magico - religioso'}	1	annotation

DEL	['ideo', 'delirante', 'ADP', 'contener', 'mágico', 'PUNCT', 'religioso']	{'ideas delirantes de contenido mágico - religioso'}	1	annotation
DEL	['ideo', 'delirante', 'ADP', 'contener', 'paranoide']	{'ideas delirantes de contenido paranoide'}	1	annotation
DEL	['ideo', 'delirante', 'ADP', 'dañar']	{'ideas delirantes de daño'}	1	annotation
DEL	['ideo', 'delirante', 'ADP', 'gandiosidad', 'CONJ', 'mistico', 'religioso']	{'ideas delirante de gandiosidad y mistico religiosas'}	1	annotation
DEL	['ideo', 'delirante', 'ADP', 'gandiosidad', 'mistica']	{'ideas delirantes de gandiosidad mistica'}	1	annotation
DEL	['ideo', 'delirante', 'ADP', 'grandeza', 'CONJ', 'misticas']	{'ideas delirantes de grandeza y misticas'}	1	annotation
DEL	['ideo', 'delirante', 'ADP', 'grandeza', 'CONJ', 'místico']	{'ideas delirantes de grandeza y mística'}	1	annotation
DEL	['ideo', 'delirante', 'ADP', 'grandeza']	{'ideas delirantes de grandeza'}	2	annotation
DEL	['ideo', 'delirante', 'ADP', 'grandiosidad', 'misitca']	{'ideas delirantes de grandiosidad misitca'}	1	annotation
DEL	['ideo', 'delirante', 'ADP', 'grandiosidad', 'mistica']	{'ideas delirantes de grandiosidad mistica'}	1	annotation
DEL	['ideo', 'delirante', 'ADP', 'grandiosidad', 'místico', 'religioso']	{'ideas delirantes de grandiosidad místico religiosas'}	1	annotation
DEL	['ideo', 'delirante', 'ADP', 'grandiosidad']	{'ideas delirantes de grandiosidad'}	2	annotation
DEL	['ideo', 'delirante', 'ADP', 'minusvalia']	{'ideas delirante de minusvalia'}	1	annotation
DEL	['ideo', 'delirante', 'ADP', 'nucleo', 'ADP', 'año']	{'ideas delirantes del nucleo del año'}	1	annotation
DEL	['ideo', 'delirante', 'ADP', 'núcleo', 'ADP', 'dañar', 'ADP', 'tipo', 'persecutorio']	{'ideas delirantes del núcleo de daño de tipo persecutorios'}	1	annotation
DEL	['ideo', 'delirante', 'ADP', 'núcleo', 'ADP', 'dañar', 'pobremente', 'estructurar']	{'ideas delirantes del núcleo del daño pobremente estructuradas'}	1	annotation
DEL	['ideo', 'delirante', 'ADP', 'nucleo', 'ADP', 'dañar', 'tipo', 'persecutorio']	{'ideas delirantes del nucleo del daño tipo persecutorias'}	1	annotation
DEL	['ideo', 'delirante', 'ADP', 'nucleo', 'ADP', 'dañar']	{'ideas delirantes del nucleo del daño', 'ideas delirantes de nucleo de daño'}	8	annotation
DEL	['ideo', 'delirante', 'ADP', 'núcleo', 'ADP', 'dañar']	{'ideas delirantes del núcleo de daño', 'ideas delirantes de núcleo de daño', 'ideas delirantes del núcleo del daño'}	8	annotation
DEL	['ideo', 'delirante', 'ADP', 'perjuicio']	{'ideas delirantes de perjuicio'}	1	annotation
DEL	['ideo', 'delirante', 'ADP', 'persecución']	{'ideas delirantes de persecución'}	1	annotation
DEL	['ideo', 'delirante', 'ADP', 'referenciar', 'CONJ', 'ADP', 'persecución']	{'ideas delirantes de referencia y de persecución'}	1	annotation
DEL	['ideo', 'delirante', 'ADP', 'referenciar', 'CONJ', 'grandioso']	{'ideas delirantes de referencia y grandiosas'}	1	annotation
DEL	['ideo', 'delirante', 'ADP', 'tipo', 'grandioso']	{'ideas delirantes de tipo grandioso'}	1	annotation
DEL	['ideo', 'delirante', 'ADP', 'tipo', 'metacognitivos', 'ADP', 'inserción', 'ADP', 'pensamiento']	{'ideas delirantes de tipo metacognitivos de inserción del pensamiento'}	1	annotation
DEL	['ideo', 'delirante', 'ADP', 'tipo', 'metacognitivos']	{'ideas delirantes de tipo metacognitivos'}	1	annotation
DEL	['ideo', 'delirante', 'ADP', 'tipo', 'místico']	{'ideas delirantes de tipo místico'}	1	annotation
DEL	['ideo', 'delirante', 'ADP', 'tipo', 'nihilista', 'CONJ', 'ADP', 'persecución']	{'ideas delirantes de tipo nihilista y de persecución'}	1	annotation
DEL	['ideo', 'delirante', 'ADP', 'tipo', 'paranoides']	{'ideas delirantes de tipo paranoides'}	1	annotation
DEL	['ideo', 'delirante', 'ADP', 'tipo', 'persecutorio', 'CONJ', 'misticas']	{'ideas delirantes de tipo persecutorio y misticas'}	1	annotation
DEL	['ideo', 'delirante', 'ADP', 'tipo', 'persecutorio']	{'ideas delirantes de tipo persecutoria', 'ideas delirantes de tipo persecutorias', 'ideas delirantes de tipo persecutorio', 'ideas delirante de tipo persecutorio'}	5	annotation
DEL	['ideo', 'delirante', 'ADP', 'tipo', 'PROPN']	{'ideas delirantes de tipo paranoide'}	1	annotation
DEL	['ideo', 'delirante', 'autoreferenciales']	{'ideas delirantes autoreferenciales'}	1	annotation
DEL	['ideo', 'delirante', 'grandioso']	{'ideas delirantes grandiosas'}	1	annotation
DEL	['ideo', 'delirante', 'mistico', 'religioso']	{'ideas delirantes mistico religiosas'}	2	annotation
DEL	['ideo', 'delirante', 'místico', 'religioso']	{'ideas delirantes místico religiosas'}	1	annotation
DEL	['ideo', 'delirante', 'nihilista']	{'ideas delirante nihilista'}	1	annotation
DEL	['ideo', 'delirante', 'paranoides']	{'ideas delirantes paranoides'}	1	annotation
DEL	['ideo', 'delirante', 'tipo', 'persecutorio', 'CONJ', 'ideo', 'mágico', 'religioso']	{'ideas delirantes tipo persecutorio e ideas mágico religiosas'}	1	annotation
DEL	['ideo', 'delirante']	{'ideas delirantes', 'ideas delirante', 'idea delirante'}	102	annotation
DEL	['ideo', 'delirantess']	{'ideas delirantess'}	1	annotation
DEL	['ideo', 'delireantes']	{'ideas delireantes'}	1	annotation
DEL	['ideo', 'delirentes']	{'ideas delirentes'}	1	annotation
DEL	['ideo', 'delirtantes']	{'ideas delirtantes'}	1	annotation
DEL	['ideo', 'dellirantes']	{'ideas dellirantes'}	1	annotation
DEL	['ideo', 'delriantes']	{'ideas delriantes'}	1	annotation
DEL	['ideo', 'grandioso']	{'ideas grandiosas'}	1	annotation
DEL	['ideo', 'misticas']	{'ideas misticas'}	1	annotation
DEL	['ides', 'delirante', 'ADP', 'nucleo', 'ADP', 'dañar']	{'ides delirantes del nucleo del daño'}	1	annotation
DEL	['influenciar', 'CONJ', 'metacognitivas']	{'influencia y metacognitivas'}	1	annotation
DEL	['inserción', 'ADP', 'pensamiento']	{'inserción del pensamiento'}	1	annotation
DEL	['interpretación', 'delirante', 'ADP', 'perseccuón']	{'interpretaciones delirantes de perseccuón'}	1	annotation
DEL	['interpretación', 'delirante', 'ADP', 'persecución', 'CONJ', 'referencialidad']	{'interpretaciones delirantes de persecución y referencialidad'}	1	annotation
DEL	['interpretación', 'delirante', 'ADP', 'referencialidad']	{'interpretaciones delirantes de referencialidad'}	1	annotation
DEL	['interpretación', 'delirante', 'ADP', 'tipo', 'persecutorio']	{'interpretaciones delirantes de tipo persecutorio'}	1	annotation
DEL	['interpretación', 'delirante']	{'interpretaciones delirantes'}	1	annotation
DEL	['ir', 'grandioso']	{'idas grandiosas'}	1	annotation
DEL	['nucleos', 'delirante']	{'nucleos delirantes'}	1	annotation

Label	Pattern	Annotation	Notes	Source
HAL	['actitud', 'alucinartoria']	{'actitud alucinartoria'}	1	annotation
HAL	['actitud', 'alucinatorio', 'auditivo']	{'actitud alucinatoria auditiva'}	1	annotation
HAL	['actitud', 'alucinatorio']	{'actitud alucinatoria', 'actitudes alucinatorias'}	45	annotation
HAL	['actitud', 'aluinatorias']	{'actitudes aluinatorias'}	1	annotation
HAL	['activida', 'alucinatorio']	{'activida alucinatoria'}	2	annotation
HAL	['actividad', 'aclucinatoria']	{'actividad aclucinatoria'}	2	annotation
HAL	['actividad', 'alcunatoria']	{'actividad alcunatoria'}	1	annotation
HAL	['actividad', 'alucinaroria']	{'actividad alucinaroria'}	1	annotation
HAL	['actividad', 'alucinartoria', 'auditivo']	{'actividad alucinartoria auditiva'}	1	annotation
HAL	['actividad', 'alucinatora']	{'actividad alucinatora'}	1	annotation
HAL	['actividad', 'alucinatorio', 'ADP', 'tipo', 'auditivo']	{'actividad alucinatoria de tipo auditivo'}	2	annotation
HAL	['actividad', 'alucinatorio', 'auditivo', 'complejo']	{'actividad alucinatoria auditiva compleja'}	1	annotation
HAL	['actividad', 'alucinatorio', 'visual']	{'actividad alucinatoria visual'}	2	annotation
HAL	['actividad', 'alucinatorio']	{'actividad alucinatoria'}	85	annotation
HAL	['actividad', 'alucnatoria']	{'actividad alucnatoria'}	1	annotation
HAL	['actividad', 'aucinatoria']	{'actividad aucinatoria'}	1	annotation
HAL	['actividad', 'aulcinatoria']	{'actividad aulcinatoria'}	1	annotation
HAL	['actividiad', 'alucinatorio']	{'actividiad alucinatoria'}	1	annotation
HAL	['ADP', 'ctitudes', 'alucinatorio']	{'a ctitudes alucinatorias'}	1	annotation
HAL	['alt', 'sensoperceptivas']	{'alt sensoperceptivas'}	1	annotation
HAL	['alteracion', 'ADP', 'DET', 'sensopercepcion']	{'alteracion en la sensopercepcion'}	2	annotation
HAL	['alteracion', 'ADP', 'DET', 'sensopercepción']	{'alteracion en la sensopercepción'}	3	annotation
HAL	['alteración', 'ADP', 'DET', 'sensopercepcion']	{'alteraciones en la sensopercepcion', 'alteraciones en su sensopercepcion', 'alteración en la sensopercepcion', 'alteraciones de la sensopercepcion'}	12	annotation
HAL	['alteración', 'ADP', 'DET', 'sensopercepción']	{'alteración en la sensopercepción', 'alteraciones en la sensopercepción', 'alteraciones de la sensopercepción', 'alteración de la sensopercepción'}	20	annotation
HAL	['alteración', 'ADP', 'DET', 'sensopersepción']	{'alteración en la sensopersepción'}	1	annotation
HAL	['alteracion', 'ADP', 'sensopercepcion']	{'alteracion de sensopercepcion'}	1	annotation
HAL	['alteración', 'ADP', 'sensopercepcion']	{'alteraciones en sensopercepcion', 'alteración de sensopercepcion'}	3	annotation
HAL	['alteracion', 'alucinatorio']	{'alteracion alucinatoria'}	1	annotation
HAL	['alteración', 'alucinatorio']	{'alteraciones alucinatorias'}	2	annotation
HAL	['alteración', 'auditivo', 'ADP', 'DET', 'sensopercepcion']	{'alteraciones auditivas de la sensopercepcion'}	1	annotation
HAL	['alteración', 'auditivo']	{'alteraciones auditivas'}	1	annotation
HAL	['alteración', 'senoperceptivas']	{'alteraciones senoperceptivas'}	1	annotation
HAL	['alteración', 'sensopeceptivas']	{'alteraciones sensopeceptivas'}	1	annotation
HAL	['alteracion', 'sensoperceptiva']	{'alteracion sensoperceptiva'}	1	annotation
HAL	['alteración', 'sensoperceptiva']	{'alteración sensoperceptiva'}	1	annotation
HAL	['alteración', 'sensoperceptivas']	{'alteraciones sensoperceptivas'}	40	annotation
HAL	['alteración', 'sensoperceptuales']	{'alteraciones sensoperceptuales'}	1	annotation
HAL	['alteracions', 'ADP', 'DET', 'sensopercepción']	{'alteracions de la sensopercepción'}	1	annotation
HAL	['alterciones', 'ADP', 'DET', 'sensopercepción']	{'alterciones en la sensopercepción'}	1	annotation
HAL	['aluciaciones', 'tactiles']	{'aluciaciones tactiles'}	1	annotation
HAL	['alucianaciones']	{'alucianaciones'}	1	annotation
HAL	['alucianciones', 'visual', 'simple']	{'alucianciones visuales simples'}	1	annotation
HAL	['alucianciones', 'visual', 'terrofícas', 'CONJ', 'propioceptivo']	{'alucianciones visuales terrofícas y propioceptivas'}	1	annotation
HAL	['alucinación', 'ADP', 'tipo', 'auditivo']	{'alucinaciones de tipo auditivo'}	1	annotation
HAL	['alucinación', 'ADP', 'tipo', 'táctil']	{'alucinaciones de tipo táctiles'}	1	annotation
HAL	['alucinación', 'ADP', 'tipo', 'visual', 'CONJ', 'auditivo']	{'alucinaciones de tipo visual o auditivas'}	1	annotation
HAL	['alucinación', 'auditivo', 'complejo', 'ADP', 'nucleo', 'ADP', 'comandar']	{'alucinaciones auditivas complejas del nucleo del comando'}	1	annotation
HAL	['alucinación', 'auditivo', 'complejo', 'CONJ', 'visual']	{'alucinaciones auditivas complejas y visuales'}	1	annotation
HAL	['alucinación', 'auditivo', 'complejo']	{'alucinaciones auditivas complejas'}	6	annotation
HAL	['alucinación', 'auditivo', 'CONJ', 'visual']	{'alucinaciones auditivas y visuales', 'alucinaciones auditivas o visuales'}	4	annotation
HAL	['alucinación', 'auditivo', 'formar']	{'alucinaciones auditivas formadas'}	2	annotation
HAL	['alucinación', 'auditivo', 'poco', 'estructurar']	{'alucinaciones auditivas poco estructuradas'}	1	annotation
HAL	['alucinación', 'auditivo', 'simple']	{'alucinaciones auditivas simples'}	1	annotation
HAL	['alucinación', 'auditivo']	{'alucinaciones auditivas'}	38	annotation
HAL	['alucinación', 'auditivo']	{'alucinaciones auditivas complejas "me están diciendo cosas pero no le quiero contar mas'}	1	annotation
HAL	['alucinación', 'auditvas']	{'alucinaciones auditvas'}	1	annotation
HAL	['alucinación', 'CONJ', 'iluciones']	{'alucinaciones e iluciones'}	1	annotation
HAL	['alucinación', 'CONJ', 'ilusionar']	{'alucinaciones e ilusiones'}	1	annotation
HAL	['alucinación', 'olfativo']	{'alucinaciones olfativas'}	2	annotation

HAL	['alucinación', 'senestesicas', 'visual', 'CONJ', 'auditivo']	{'alucinaciones senestesicas visuales y auditivas'}	1	annotation
HAL	['alucinación', 'tactiles']	{'alucinaciones tactiles'}	1	annotation
HAL	['alucinación', 'tipo', 'auditivo']	{'alucinaciones tipo auditivo'}	1	annotation
HAL	['alucinación', 'tipo', 'tactiles']	{'alucinaciones tipo tactiles'}	1	annotation
HAL	['alucinación', 'visales', 'complejo']	{'alucinaciones visales complejas'}	1	annotation
HAL	['alucinación', 'visual', 'cmplejas']	{'alucinaciones visuales cmplejas'}	1	annotation
HAL	['alucinación', 'visual', 'compleas', 'CONJ', 'auditivo', 'simple']	{'alucinaciones visuales compleas y auditivas simples'}	1	annotation
HAL	['alucinación', 'visual', 'complejo', 'CONJ', 'auditivo', 'simple']	{'alucinaciones visuales complejas y auditivas simples'}	1	annotation
HAL	['alucinacion', 'visual', 'complejo']	{'alucinacion visual compleja'}	1	annotation
HAL	['alucinación', 'visual', 'complejo']	{'alucinaciones visuales complejas', 'alucinación visual compleja'}	7	annotation
HAL	['alucinación', 'visual', 'CONJ', 'auditivo', 'complejo', 'tipo', 'comandar']	{'alucinación visual y auditiva compleja tipo comando'}	1	annotation
HAL	['alucinación', 'visual', 'CONJ', 'auditivo', 'complejo']	{'alucinaciones visuales y auditivas complejas'}	3	annotation
HAL	['alucinación', 'visual', 'CONJ', 'auditivo', 'estrucruradas']	{'alucinaciones visuales y auditivas estrucruradas'}	1	annotation
HAL	['alucinación', 'visual', 'CONJ', 'auditivo', 'simple']	{'alucinaciones visuales y auditivas simples'}	1	annotation
HAL	['alucinacion', 'visual', 'CONJ', 'auditivo']	{'alucinacion visual y auditiva'}	1	annotation
HAL	['alucinación', 'visual', 'CONJ', 'auditivo']	{'alucinaciones visuales o auditivas', 'alucinaciones visuales y auditivas'}	10	annotation
HAL	['alucinación', 'visual', 'simple']	{'alucinaciones visuales simples'}	1	annotation
HAL	['alucinación', 'visual', 'terrorificas']	{'alucinaciones visuales terrorificas'}	1	annotation
HAL	['alucinación', 'visual']	{'alucinaciones visuales'}	20	annotation
HAL	['alucinación', 'vsuales']	{'alucinaciones vsuales'}	1	annotation
HAL	['alucinacion']	{'alucinacion'}	1	annotation
HAL	['alucinación']	{'alucinación', 'alucinaciones'}	18	annotation
HAL	['alucinacions', 'auditivo']	{'alucinacions auditivas'}	1	annotation
HAL	['alucinacios', 'auditivo']	{'alucinacios auditivas'}	1	annotation
HAL	['alucinacones']	{'alucinacones'}	1	annotation
HAL	['alucinar']	{'alucinando'}	3	annotation
HAL	['alucinatorio', 'ADP', 'tipo', 'cenestecico']	{'alucinatorio de tipo cenestecico'}	1	annotation
HAL	['alucinosis']	{'alucinosis'}	1	annotation
HAL	['alucionaciones', 'auditivo']	{'alucionaciones auditivas'}	1	annotation
HAL	['aparentar', 'alucinación']	{'aparenta alucinaciones'}	1	annotation
HAL	['ateraciones', 'sesoperceptivas']	{'ateraciones sesoperceptivas'}	1	annotation
HAL	['auditicas', 'CONJ', 'somaticas']	{'auditicas y somaticas'}	1	annotation
HAL	['auditivo', 'complejo']	{'auditivas complejas'}	1	annotation
HAL	['auditivo', 'CONJ', 'somático', 'formar']	{'auditivas y somáticas formadas'}	1	annotation
HAL	['auditivo', 'CONJ', 'táctil']	{'auditiva y táctil'}	1	annotation
HAL	['auditivo', 'estructurar']	{'auditivas estructuradas'}	1	annotation
HAL	['conducta', 'alucinatorio']	{'conducta alucinatoria'}	1	annotation
HAL	['diablo', 'le', 'decir']	{'en el momento el diablo le dice "no le preste atención al médico'}	1	annotation
HAL	['escuchar', 'DET', 'voz']	{'escucho la voz'}	1	annotation
HAL	['escuchar', 'voz']	{'escucho voces'}	1	annotation
HAL	['ilusionar', 'visual']	{'ilusiones visuales'}	2	annotation
HAL	['intomas', 'alucinatorio', 'auditivo']	{'intomas alucinatorios auditivos'}	1	annotation
HAL	['lucinaciones', 'ADP', 'tipo', 'auditivo']	{'lucinaciones de tipo auditivo'}	1	annotation
HAL	['percepción', 'ADP', 'sombrar', 'ADP', 'manera', 'alucinatorio']	{'percepciones de sombras de manera alucinatoria'}	1	annotation
HAL	['percepción', 'ADP', 'tipo', 'aluciantorio']	{'percepciones de tipo aluciantorio'}	1	annotation
HAL	['percepción', 'ADP', 'tipo', 'alucinatorio']	{'percepciones de tipo alucinatorio'}	3	annotation
HAL	['percepción', 'alucinatorio']	{'percepciones alucinatorias'}	1	annotation
HAL	['PUNCT', 'escuchar', 'voz', 'que', 'decir', 'hola', 'CONJ', 'INTJ']	{'"escucha voces que dicen hola y adiós'}	1	annotation
HAL	['sensopercepcion', 'alteración']	{'sensopercepcion alteraciones'}	1	annotation
HAL	['sensopercepción', 'alteración']	{'sensopercepción alteraciones'}	1	annotation
HAL	['sensopercepcion']	{'sensopercepcion'}	1	annotation
HAL	['sensopercepción']	{'sensopercepción'}	4	annotation
HAL	['sensoperceptivas']	{'sensoperceptivas'}	1	annotation
HAL	['sensoperceptuales']	{'sensoperceptuales'}	1	annotation
HAL	['síntoma', 'sensoperceptivos']	{'síntomas sensoperceptivos'}	1	annotation
HAL	['trastorno', 'ADP', 'DET', 'sensopercepción']	{'trastornos en la sensopercepción'}	1	annotation
HAL	['ver', 'ADP', 'diablo']	{'yo veo al diablo en forma de imagénes y el infierno en todo lado y a mi abuela en él diciendo que me mate y mate al resto'}	1	annotation
HAL	['visual', 'complejo']	{'visuales complejas'}	1	annotation
HAL	['voz', 'que', 'decir', 'hola', 'CONJ', 'INTJ']	{'voces que dicen hola y adiós'}	1	annotation
HAL	['voz', 'que', 'le', 'mandar', 'ADP', 'auto', 'agredirse']	{'voces que le mandan a auto agredirse'}	1	annotation
HAL	['voz', 'que', 'le', 'mandar']	{'con voces que le mandan a auto agredirse'}	1	annotation
HAL	['voz', 'que', 'lo', 'querer', 'matar']	{'voces que la quieren matar'}	1	annotation
HAL	['yo', 'oír', 'ADP', 'DET', 'virgen']	{'yo oigo a la virgen'}	1	annotation

T.3 Supplementary Table 3. *Patterns used by the NegEx algorithm. Type: either negation, pseudo-negation or termination of scope. Scope: directionality of negation.*

Pattern	Type	Scope
[["sin","nunca","no","ni","negar","-no","-sin","sn","tampoco","jamás"]]	negation	forward
["ya:?","no",["referir","impresionar","aparentar","manifiesto","verbalizar","evidenciar"]]	negation	forward
["ya","no"]	negation	forward
["nulo"]	negation	backward
["se", "resolver"]	negation	backward
["haber", "ceder"]	negation	backward
["no", "debilitado"]	negation	backward
["se:?","descartar"]	termination	
[["minimizacion","minimización","minimizar"]]	termination	
["ceder"]	termination	
["embargar"]	termination	
["embargos"]	termination	
["pero"]	termination	
["se:?","porque"]	termination	
["con"]	termination	
[',']	termination	
["sin",["alteración","alteracion"]]	negation	bidirectional
["normal"]	negation	backward
["ausencia","de"]	negation	forward
["no", "alterar"]	negation	backward
["sin","cambio"]	pseudonegation	

T.4 Supplementary Table 4. *List of diagnostic pairs from the F chapter of ICD-10 that are, by definition, incompatible with each other and, therefore, represent diagnostic switches. All other combinations of diagnoses are considered comorbidities. There are two exceptions to this rule: the pairs F30-F31 and F32-F33. These are neither switches nor comorbidities.*

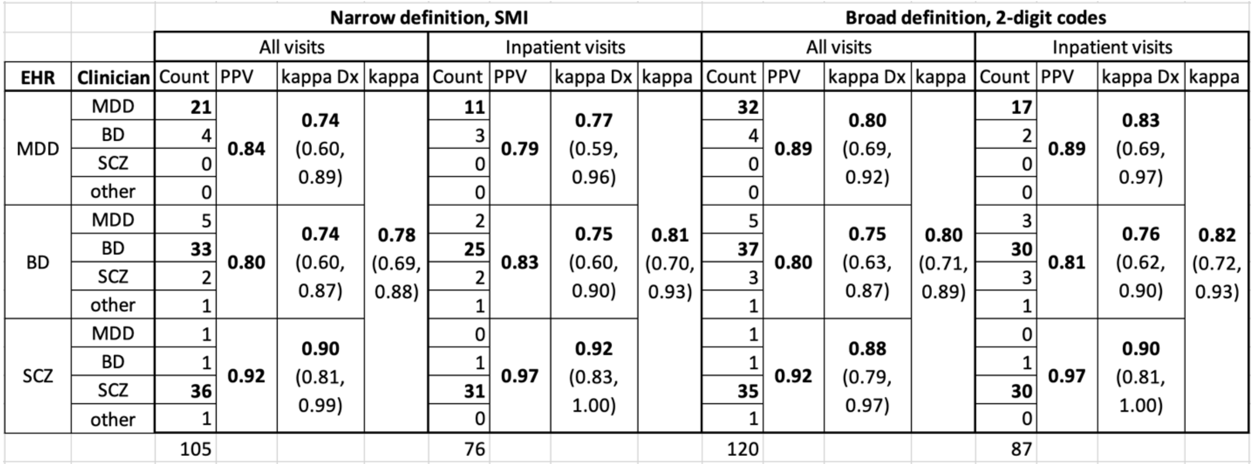
			Schizophrenia, schizotypal and delusional disorders								**Mood [affective] disorders**						
		t+1	Schizophrenia	Schizotypal disorder	Persistent delusional disorders	Acute and transient psychotic disorders	Induced delusional disorder	Schizoaffective disorders	Other nonorganic psychotic disorders	Unspecified nonorganic psychosis	Manic episode	Bipolar affective disorder	Depressive episode	Recurrent depressive disorder	Persistent mood [affective] disorders	Other mood [affective] disorders	Unspecified mood [affective] disorder
	t		F20	F21	F22	F23	F24	F25	F28	F29	F30	F31	F32	F33	F34	F38	F39
Schizophrenia, schizotypal and delusional disorders	Schizophrenia	F20		1	1	1	1	1	1	1	1	1		1	1	1	1
	Schizotypal disorder	F21	1		1	1	1	1	1	1	1	1		1	1	1	1
	Persistent delusional disorders	F22	1	1		1	1	1	1	1	1	1		1	1	1	1
	Acute and transient psychotic disorders	F23	1	1	1		1	1	1	1	1	1		1	1	1	1
	Induced delusional disorder	F24	1	1	1	1		1	1	1	1	1		1	1	1	1
	Schizoaffective disorders	F25	1	1	1	1	1		1	1	1	1		1	1	1	1
	Other nonorganic psychotic disorders	F28	1	1	1	1	1	1		1	1	1		1	1	1	1
	Unspecified nonorganic psychosis	F29	1	1	1	1	1	1	1		1	1		1	1	1	1
Mood [affective] disorders	Manic episode	F30	1	1	1	1	1	1	1	1		-	1	1	1	1	1
	Bipolar affective disorder	F31	1	1	1	1	1	1	1	1	-		1	1	1	1	1
	Depressive episode	F32									1	1		-	1	1	1
	Recurrent depressive disorder	F33	1	1	1	1	1	1	1	1	1	1	-		1	1	1
	Persistent mood [affective] disorders	F34	1	1	1	1	1	1	1	1	1	1	1	1		1	1
	Other mood [affective] disorders	F38	1	1	1	1	1	1	1	1	1	1	1	1	1		1
	Unspecified mood [affective] disorder	F39	1	1	1	1	1	1	1	1	1	1	1	1	1	1	

T.5 Supplementary Table 5. *Demographic and clinical characteristics of the study cohort. Patients are classified by their most recent SMI diagnosis. Medians with interquartile range (IQR) are presented for: visits per patient, age at the most recent visit, length of stay, and length of the medical record. Tests comparing these values across the three main diagnoses (MDD, BD and SCZ) are provided in the bottom part of the table.* ***Test*:*** *across the three diagnoses, differences in percentages are tested with a chi-squared test and differences in distributions with a Kruskal-Wallis test.* ***Test**:*** *between pairs of diagnoses, differences in percentages are tested with z-tests and differences in distributions with Mann-Whitney tests. Asterisks mark significant results at the Bonferroni-corrected alpha threshold of 0.05/8=0.006.*

	MDD	BD	SCZ	Psych	All
Patients, N	10862	8662	2652	271	22447
Visits, N	46209	82328	25964	2502	157003
Female, %	66.5	64.6	23.9	42.8	60.4
Any visits under age 18, %	19.7	12.6	16.6	7	16.4
Any inpatient visits, %	56	71.7	74.4	65.7	64.4
Any ER visits, %	53.7	57.1	61.9	42.4	55.8
Visits per patient	2 (1, 5)	5 (2, 12)	5 (2, 13)	4 (2, 12)	3 (1, 9)
Age at most recent visit, years	37.6 (21.7, 54.7)	45.0 (27.2, 58.7)	33.7 (23.6, 51.8)	47.4 (32.1, 60.3)	40.0 (24.1, 56.2)
Length of stay, days	8 (4, 12)	12 (7, 18)	14 (9, 21)	14 (9, 22)	11 (6, 17)
Length of medical record, years	7.0 (4.1, 10.7)	10.1 (6.3, 14.0)	9.0 (5.8, 14.3)	12.9 (9.1, 15.6)	8.4 (5.1, 12.8)

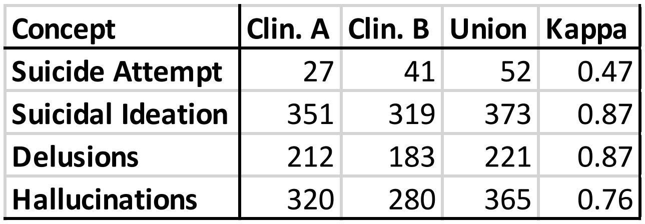
Comparisons	**MDD-BD-SCZ**			**MDD-BD**			**BD-SCZ**			**MDD-SCZ**		
	test*	**p-value**		**test****	**p-value**		**test****	**p-value**		**test****	**p-value**	
Female, %	1762.2	0.0	*	1.9	0.054		37.4	<2e-16	*	40.2	0.0	*
Any visits under age 18, %	192.4	<2e-16	*	13.9	<2e-16	*	-6.5	9.6e-11	*	3.1	1.7e-3	*
Any inpatient visits, %	617.5	<2e-16	*	-21.5	<2e-16	*	-3.4	0.0008	*	-17.6	<2e-16	*
Any ER visits, %	65.9	4.8e-15	*	-3.5	0.0004	*	-5.5	4.6e-8	*	-8.0	1.1e-15	*
Visits per patient	2373.5	<2e-16	*	2.9e+7	0.0	*	1.1e+7	0.42		9.6e+6	<2e-16	*
Age at most recent visit, years	380.7	<2e-16	*	3.9e+7	<2e-16	*	1.4e+7	<2e-16	*	1.6e+7	0.07	
Length of stay, days	1785.6	<2e-16	*	1.2e+7	<2e-16	*	4.8e+6	<2e-16	*	3.1e+6	<2e-16	*
Length of medical record, years	1242.9	<2e-16	*	2.6e+7	<2e-16	*	1.1e+7	5.2e-7	*	9.7e+6	<2e-16	*

T.6 Supplementary Table 6. *Estimates of kappa and PPV from comparisons of ICD-10 diagnoses extracted from the EHR and clinician diagnoses obtained from chart review, considering all visits and considering inpatient visits only. The narrow definition refers to the SMI codes: F20X (SCZ), F301, F302, F310, F311, F312, F313, F314, F315, F316, F317 (BD), F322, F323, F331, F332, F333, F334 (Severe/Recurrent MDD). The broad definition encompasses, additionally, all F31X (including F318 and F319), F32X (including F320, F321, F328 and F329) and F33X (including F330, F338 and F339). Kappa values are estimated both for individual diagnoses and across all diagnoses. 95% confidence intervals for kappa values are shown in parentheses. Kappa values between 0.6-0.8 are considered "very good", while those > 0.8 are considered "excellent".*



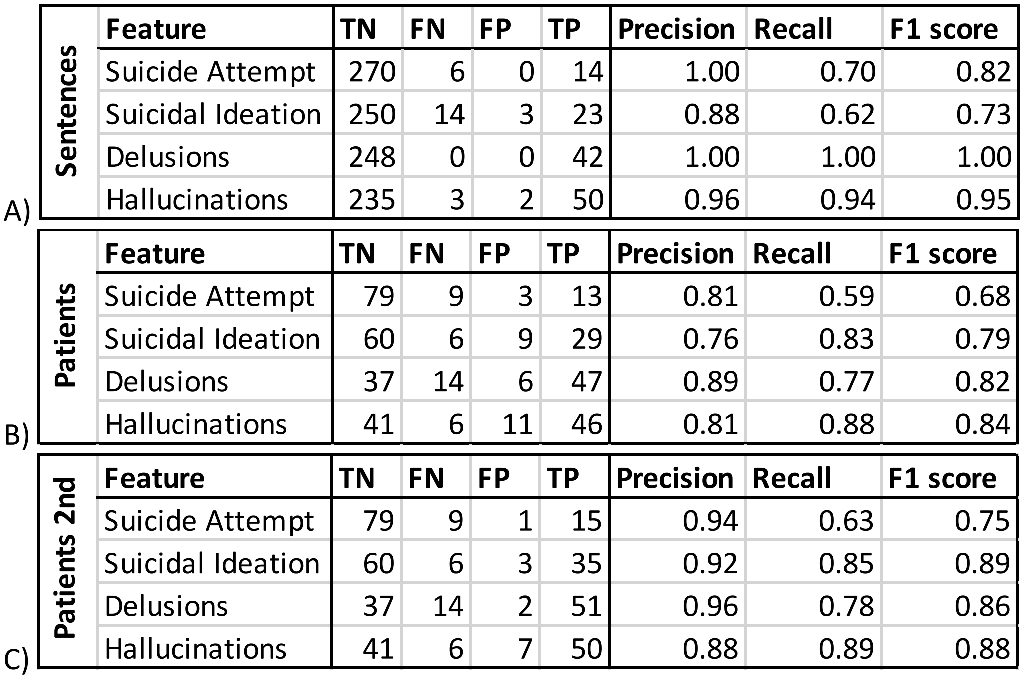
		Narrow definition, SMI								**Broad definition, 2-digit codes**							
		All visits				Inpatient visits				All visits				Inpatient visits			
EHR	**Clinician**	Count	PPV	kappa Dx	kappa	Count	PPV	kappa Dx	kappa	Count	PPV	kappa Dx	kappa	Count	PPV	kappa Dx	kappa
MDD	MDD	**21**	**0.84**	**0.74** (0.60, 0.89)	**0.78** (0.69, 0.88)	**11**	**0.79**	**0.77** (0.59, 0.96)	**0.81** (0.70, 0.93)	**32**	**0.89**	**0.80** (0.69, 0.92)	**0.80** (0.71, 0.89)	**17**	**0.89**	**0.83** (0.69, 0.97)	**0.82** (0.72, 0.93)
	BD	4				3				4				2			
	SCZ	0				0				0				0			
	other	0				0				0				0			
BD	MDD	5	**0.80**	**0.74** (0.60, 0.87)		2	**0.83**	**0.75** (0.60, 0.90)		5	**0.80**	**0.75** (0.63, 0.87)		3	**0.81**	**0.76** (0.62, 0.90)	
	BD	**33**				**25**				**37**				**30**			
	SCZ	2				2				3				3			
	other	1				1				1				1			
SCZ	MDD	1	**0.92**	**0.90** (0.81, 0.99)		0	**0.97**	**0.92** (0.83, 1.00)		1	**0.92**	**0.88** (0.79, 0.97)		0	**0.97**	**0.90** (0.81, 1.00)	
	BD	1				1				1				1			
	SCZ	**36**				**31**				**35**				**30**			
	other	1				0				1				0			
		105				76				120				87			

T.7 ***Supplementary Table 7.*** *Annotation results for four clinical features. Two clinicians independently reviewed and annotated 3,600 sentences. The columns show the number of sentences containing the clinical features identified by Clinician A (Clin. A), Clinician B (Clin. B) or either clinician (Union), and their level of agreement as estimated by Cohen's kappa (Kappa). Kappa values between 0.4-0.6 are considered "good", those between 0.6-0.8 are considered "very good", and those > 0.8 are considered "excellent".*



Concept	Clin. A	Clin. B	Union	Kappa
Suicide Attempt	27	41	52	0.47
Suicidal Ideation	351	319	373	0.87
Delusions	212	183	221	0.87
Hallucinations	320	280	365	0.76

T.8 Supplementary Table 8. *Performance of the NLP algorithm for extraction of clinical features. A) Sentence-level performance on the annotated gold standard. B) Patient-level performance on patient records manually reviewed by a clinician (n=104, as one patient was removed for having only one clinical note). C) Patient-level performance after post-hoc review of true and false positives. The average affirmative and negative instances of each feature per patient are, respectively: 1 and 0 for Suicide Attempt, 4 and 12 for Suicidal Ideation, 17 and 19 for Delusions, 10 and 25 for Hallucinations.*



A)

	Feature	TN	FN	FP	TP	Precision	Recall	F1 score
Sentences	Suicide Attempt	270	6	0	14	1.00	0.70	0.82
	Suicidal Ideation	250	14	3	23	0.88	0.62	0.73
	Delusions	248	0	0	42	1.00	1.00	1.00
	Hallucinations	235	3	2	50	0.96	0.94	0.95

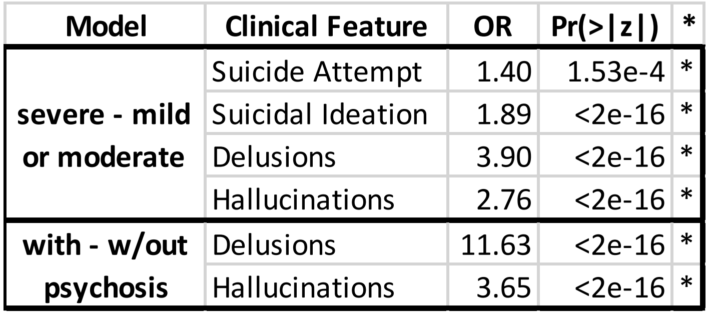
B)

	Feature	TN	FN	FP	TP	Precision	Recall	F1 score
Patients	Suicide Attempt	79	9	3	13	0.81	0.59	0.68
	Suicidal Ideation	60	6	9	29	0.76	0.83	0.79
	Delusions	37	14	6	47	0.89	0.77	0.82
	Hallucinations	41	6	11	46	0.81	0.88	0.84

C)

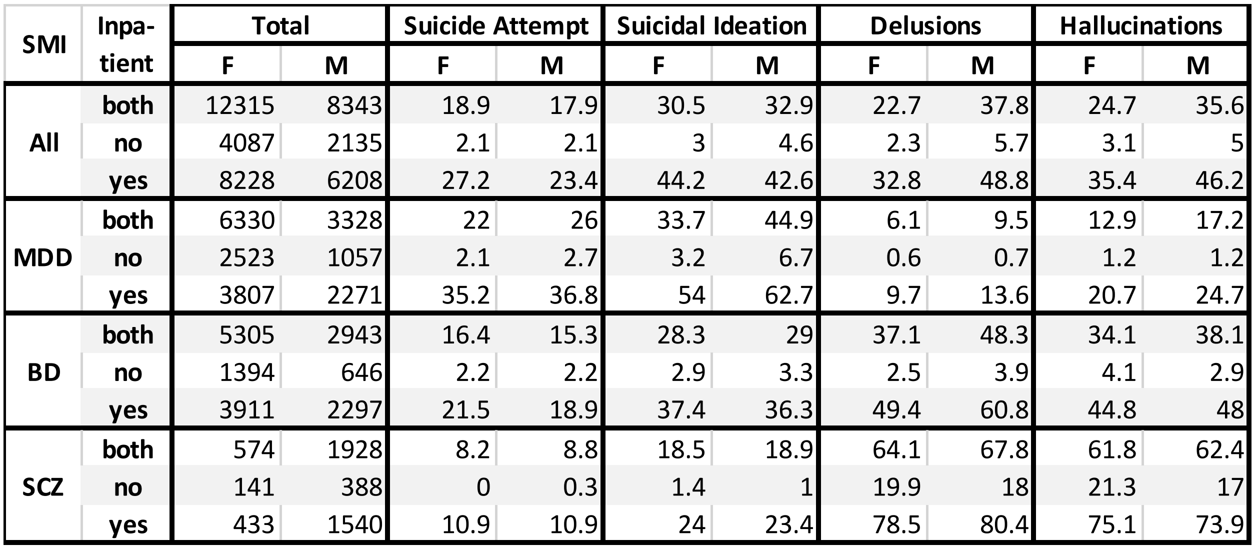
	Feature	TN	FN	FP	TP	Precision	Recall	F1 score
Patients 2nd	Suicide Attempt	79	9	1	15	0.94	0.63	0.75
	Suicidal Ideation	60	6	3	35	0.92	0.85	0.89
	Delusions	37	14	2	51	0.96	0.78	0.86
	Hallucinations	41	6	7	50	0.88	0.89	0.88

T.9 Supplementary Table 7. *ICD-10 code severity and psychosis qualifiers, recorded at each visit for individuals receiving an ICD-10 mood disorder diagnosis demonstrate strong association with clinical feature profiles extracted from the notes during the same visit. The top section shows the association of clinical features with codes representing mood disorder diagnoses designated as severe compared to those designated as either mild or moderate. The binary variable "severe" is defined as 1 when the visit code is one of F301, F302, F311, F312, F314, F315, F322, F323, F332, or F333, and is 0 when the visit code is one of F300, F310, F313, F320, F321, F330, or F331. For this section, N is 47,186 visits in 9,203 people. The bottom section shows the association of psychotic features (Delusions and Hallucinations) with codes designating the presence of psychotic symptoms during visits in which a code designating a mood disorder episode as severe has been recorded. The binary variable "psychosis" is defined as 1 when the visit code is one of F302, F312, F315, F323, or F333 and is 0 when the visit code is one of F301, F311, F314, F322, or F332. For this section, N is 15,120 visits in 4,075 people. Analyses are described in Supplementary Note 6.*



Model	Clinical Feature	OR	Pr(>|z|)	*
severe - mild or moderate	Suicide Attempt	1.40	1.53e-4	*
	Suicidal Ideation	1.89	<2e-16	*
	Delusions	3.90	<2e-16	*
	Hallucinations	2.76	<2e-16	*
with - w/out psychosis	Delusions	11.63	<2e-16	*
	Hallucinations	3.65	<2e-16	*

***T.10 Supplementary Table 8**. Frequency of each clinical feature (in percentages), for patients with SMI diagnoses, stratified by gender and inpatient history (yes: patients with a history of at least one inpatient hospitalization; no: individuals without any history of inpatient hospitalization). The first two columns show the total number of individuals included in this analysis, while the other columns show the frequencies of each clinical feature. The rows display the total numbers and frequencies considering all SMI diagnoses ("All") and then considering each of the three diagnoses separately. All patients included in this table had at least two clinical notes in their EHR.*



SMI	Inpa-tient	Total		Suicide Attempt		Suicidal Ideation		Delusions		Hallucinations	
		F	M	F	M	F	M	F	M	F	M
All	both	12315	8343	18.9	17.9	30.5	32.9	22.7	37.8	24.7	35.6
	no	4087	2135	2.1	2.1	3	4.6	2.3	5.7	3.1	5
	yes	8228	6208	27.2	23.4	44.2	42.6	32.8	48.8	35.4	46.2
MDD	both	6330	3328	22	26	33.7	44.9	6.1	9.5	12.9	17.2
	no	2523	1057	2.1	2.7	3.2	6.7	0.6	0.7	1.2	1.2
	yes	3807	2271	35.2	36.8	54	62.7	9.7	13.6	20.7	24.7
BD	both	5305	2943	16.4	15.3	28.3	29	37.1	48.3	34.1	38.1
	no	1394	646	2.2	2.2	2.9	3.3	2.5	3.9	4.1	2.9
	yes	3911	2297	21.5	18.9	37.4	36.3	49.4	60.8	44.8	48
SCZ	both	574	1928	8.2	8.8	18.5	18.9	64.1	67.8	61.8	62.4
	no	141	388	0	0.3	1.4	1	19.9	18	21.3	17
	yes	433	1540	10.9	10.9	24	23.4	78.5	80.4	75.1	73.9

T.11 Supplementary Table 9. *Odds ratios for patient-level associations of each clinical feature with gender, diagnosis, and the other clinical features. Bonferroni-corrected alpha is 0.05/12=0.0041 for table A and 0.05/16=0.0031 for table B. Analyses are described in Supplementary Note 2.*

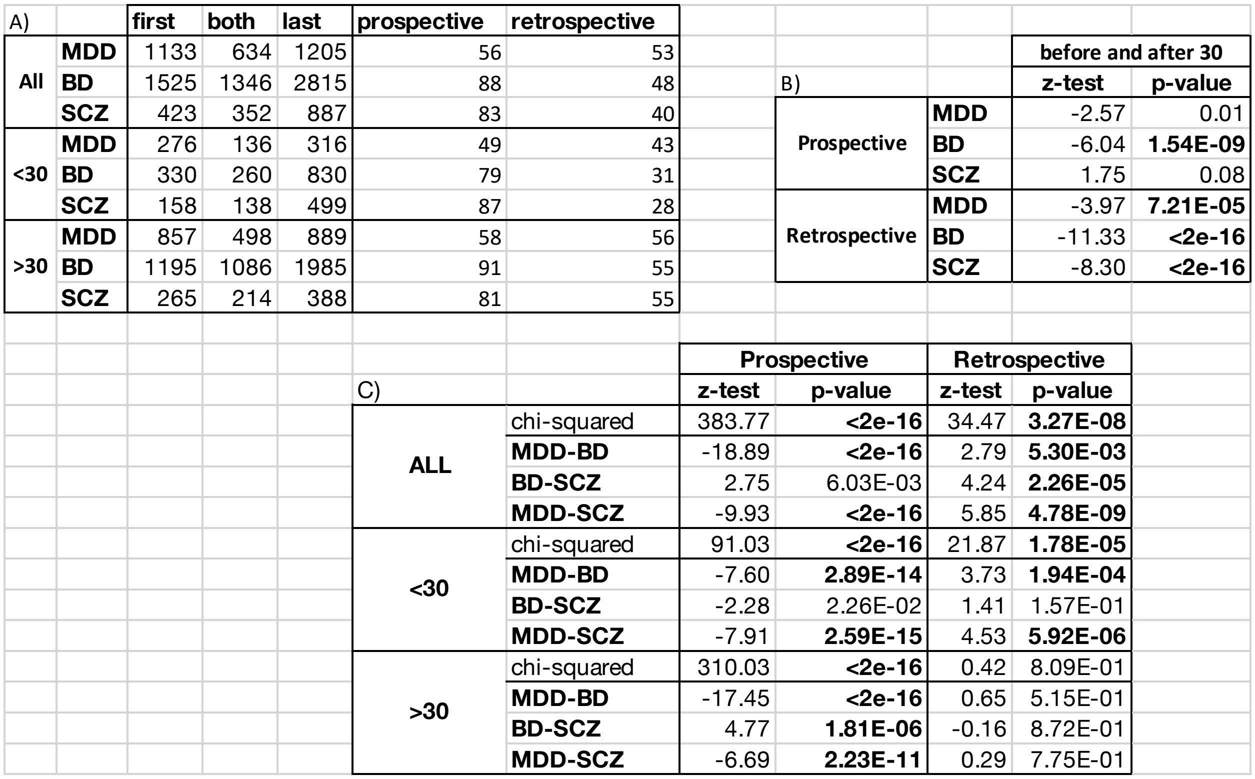
A) Gender-diagnosis interactions		**Suicide Attempt**			**Suicidal Ideation**			**Delusions**			**Hallucinations**		
Interaction	**Coefficient**	**OR**	**Pr(>|z|)**	*****	**OR**	**Pr(>|z|)**	*****	**OR**	**Pr(>|z|)**	*****	**OR**	**Pr(>|z|)**	*****
No	**Female**	1.01	7.16E-01		0.84	8.42E-07	*	0.67	<2e-16	*	0.88	5.90E-04	*
	MDD	2.01	<2e-16	*	2.57	<2e-16	*	0.14	<2e-16	*	0.46	<2e-16	*
	SCZ	0.46	<2e-16	*	0.48	<2e-16	*	3.33	<2e-16	*	3.61	<2e-16	*
Yes	**Female**	1.17	1.53E-02		1.04	4.46E-01		0.62	<2e-16	*	0.88	1.39E-02	
	MDD	2.34	<2e-16	*	3.38	<2e-16	*	0.13	<2e-16	*	0.48	<2e-16	*
	SCZ	0.51	3.97E-12	*	0.52	<2e-16	*	2.89	<2e-16	*	3.38	<2e-16	*
	Female:MDD	0.78	3.99E-03	*	0.65	4.20E-09	*	1.13	2.20E-01		0.91	2.67E-01	
	Female:SCZ	0.84	3.43E-01		1.01	9.61E-01		1.58	7.12E-04	*	1.33	2.61E-02	

B) Gender-clinical feature interactions		**Suicide Attempt**			**Suicidal Ideation**			**Delusions**			**Hallucinations**		
Interaction	**Coefficient**	**OR**	**Pr(>|z|)**	*****	**OR**	**Pr(>|z|)**	*****	**OR**	**Pr(>|z|)**	*****	**OR**	**Pr(>|z|)**	*****
No	**Female**	1.05	2.62E-01		0.81	1.24E-08	*	0.67	<2e-16	*	1.01	8.96E-01	
	MDD	1.4	4.76E-12	*	2.17	<2e-16	*	0.18	<2e-16	*	0.65	<2e-16	*
	SCZ	0.6	2.05E-09	*	0.5	<2e-16	*	2.14	<2e-16	*	3.21	<2e-16	*
	Suicide Attempt				3.85	<2e-16	*	0.62	<2e-16	*	1.29	2.89E-07	*
	Suicidal Ideation	3.85	<2e-16	*				0.62	<2e-16	*	2.05	<2e-16	*
	Delusions	0.61	<2e-16	*	0.59	<2e-16	*				5.14	<2e-16	*
	Hallucinations	1.29	1.27E-07	*	1.98	<2e-16	*	5.14	<2e-16	*			
Yes	**Female**	1.19	1.93E-02		0.81	5.13E-05	*	0.72	1.91E-07	*	0.87	4.52E-02	
	MDD	1.39	8.21E-12	*	2.17	<2e-16	*	0.17	<2e-16	*	0.65	<2e-16	*
	SCZ	0.6	1.74E-09	*	0.5	<2e-16	*	2.21	<2e-16	*	3.21	<2e-16	*
	Suicide Attempt				4.13	<2e-16	*	0.72	8.91E-05	*	1.27	2.06E-03	*
	Suicidal Ideation	4.21	<2e-16	*				0.73	1.63E-05	*	1.92	<2e-16	*
	Delusions	0.69	2.30E-06	*	0.64	7.45E-11	*				4.59	<2e-16	*
	Hallucinations	1.24	4.09E-03	*	1.73	<2e-16	*	4.57	<2e-16	*			
	Female:Suicide Attempt				0.89	1.76E-01		0.77	2.26E-02		1.03	7.73E-01	
	Female:Suicidal Ideation	0.86	8.59E-02					0.73	1.10E-03	*	1.12	2.07E-01	
	Female:Delusions	0.8	2.64E-02		0.84	5.64E-02					1.24	9.13E-03	
	Female:Hallucinations	1.08	4.40E-01		1.27	5.44E-03		1.24	1.28E-02				

__T.12 Supplementary Table 10.__ Percentage of individuals with comorbidities within each SMI diagnosis, as observed in patients with at least 3 encounters (n=12,962). The ICD codes for the 20 most frequent diagnoses are shown.

ICD-10 codes		MDD	BD	SCZ	All
F00	Dementia in Alzheimer disease	1.2	1	1.1	137
F03	Unspecified dementia	1.3	1.2	1.5	165
F06	Other mental disorders due to brain damage and dysfunction and to physical disease	4.1	4.4	7.2	597
F10	Mental and behavioural disorders due to use of alcohol	1	1	0.6	125
F12	Mental and behavioural disorders due to use of cannabinoids	0.9	1.2	5.2	215
F19	Mental and behavioural disorders due to multiple drug use and use of other psychoactive substances	3.1	4	13.3	637
F32	Depressive episode			5.4	100
F41	Other anxiety disorders	28	12	3.3	2126
F42	Obsessive-compulsive disorder	1.5	0.9	0.8	140
F43	Reaction to severe stress, and adjustment disorders	9	3.9	1.9	702
F45	Somatoform disorders	1.6	0.5	0.1	107
F52	Sexual dysfunction, not caused by organic disorder or disease	1.9	0.8	0.1	143
F60	Specific personality disorders	2.3	2	0.9	247
F70	Mild mental retardation	0.8	2.2	7.7	317
F71	Moderate mental retardation	0.4	1.6	9.9	304
F72	Severe mental retardation	0	0.4	4	102
F90	Hyperkinetic disorders	2.1	2.4	2.7	300
F91	Conduct disorders	3.3	4	5.8	512
F92	Mixed disorders of conduct and emotions	3.1	1.6	0.6	259
F99	Mental disorder, not otherwise specified	0.5	1.2	2.1	138
		4704	6226	1855	

T.13 Supplementary Table 11. *Stability of diagnoses. A) Counts and comparative statistics for long-term stability of SMI diagnoses. First: number of individuals with that diagnosis on their first visit. Last: number of individuals with that diagnosis on their last visit. Both: number of individuals with that diagnosis on both visits. Prospective stability is calculated as 100*both/first, and retrospective as 100*both/last. B) comparison of stability values between two groups: those whose first visit was before age 30 and those whose first visit was after age 30. C) comparison between prospective and retrospective stability for all patients and stratified by age group.*



A)		**first**	**both**	**last**	**prospective**	**retrospective**
All	**MDD**	1133	634	1205	56	53
	BD	1525	1346	2815	88	48
	SCZ	423	352	887	83	40
<30	**MDD**	276	136	316	49	43
	BD	330	260	830	79	31
	SCZ	158	138	499	87	28
>30	**MDD**	857	498	889	58	56
	BD	1195	1086	1985	91	55
	SCZ	265	214	388	81	55

B)		**before and after 30**	
		z-test	**p-value**
Prospective	**MDD**	-2.57	0.01
	BD	-6.04	**1.54E-09**
	SCZ	1.75	0.08
Retrospective	**MDD**	-3.97	**7.21E-05**
	BD	-11.33	**<2e-16**
	SCZ	-8.30	**<2e-16**

C)		**Prospective**		**Retrospective**	
		z-test	**p-value**	**z-test**	**p-value**
ALL	chi-squared	383.77	**<2e-16**	34.47	**3.27E-08**
	MDD-BD	-18.89	**<2e-16**	2.79	**5.30E-03**
	BD-SCZ	2.75	6.03E-03	4.24	**2.26E-05**
	MDD-SCZ	-9.93	**<2e-16**	5.85	**4.78E-09**
<30	chi-squared	91.03	**<2e-16**	21.87	**1.78E-05**
	MDD-BD	-7.60	**2.89E-14**	3.73	**1.94E-04**
	BD-SCZ	-2.28	2.26E-02	1.41	1.57E-01
	MDD-SCZ	-7.91	**2.59E-15**	4.53	**5.92E-06**
>30	chi-squared	310.03	**<2e-16**	0.42	8.09E-01
	MDD-BD	-17.45	**<2e-16**	0.65	5.15E-01
	BD-SCZ	4.77	**1.81E-06**	-0.16	8.72E-01
	MDD-SCZ	-6.69	**2.23E-11**	0.29	7.75E-01