scientific reports

OPEN

Data harmonization for the analysis of personalized treatment of psychosis with metacognitive training

Caroline König[1,10✉], Pedro Copado[1,10], Maria Lamarca[3,4,5,10], Wafaa Guendouz[1,10], Rabea Fischer[6], Merle Schlechte[6], Vanessa Acuña[7], Fabrice Berna[8], Łucasz Gawęda[9], Alfredo Vellido[1], Àngela Nebot[1], Cecilio Angulo[2] & Susana Ochoa[3,4]

Personalized medicine is a data-driven approach that aims to adapt patients' diagnostics and therapies to their characteristics and needs. The availability of patients' data is therefore paramount for the personalization of treatments on the basis of predictive models, and even more so in machine learning-based analyses. Data harmonization is an essential part of the process of data curation. This study presents research on data harmonization for the development of a harmonized retrospective database of patients in Metacognitive Training (MCT) treatment for psychotic disorders. This work is part of the European ERAPERMED 2022-292 research project entitled 'Towards a Personalized Medicine Approach to Psychological Treatment of Psychosis' (PERMEPSY), which focuses on the development of a personalized medicine platform for the treatment of psychosis. The study integrates information from 22 studies into a common format to enable a data analytical approach for personalized treatment. The harmonized database comprises information about 698 patients who underwent MCT and includes a wide range of sociodemographic variables and psychological indicators used to assess a patient's mental health state. The characteristics of patients participating in the study are analyzed using descriptive statistics and exploratory data analysis.

Traditional healthcare generalizes treatment from population averages. Personalized approaches are becoming increasingly feasible and hold the combined promises of predicting, preventing, and treating illnesses according to individual needs[1]. This is due to the availability of a data-driven approach, often fuelled by machine learning (ML)-based analytical pipelines that leverage predictive capabilities and enable the simultaneous handling of extensive predictor datasets[2,3]. The concept of personalization of healthcare in the field of psychology and its application to treating mental illnesses is already used in different clinical applications[4,5]. Since multiple treatment alternatives are available for psychological disorders, the clinical efficacy and cost-effectiveness of treatments are essential factors to consider for policymakers, therapists, and patients alike, thereby significantly influencing healthcare decision-making[6]. At the onset of an intervention, identifying the specific individuals

[1]Soft Computing Research Group (SOCO) at Intelligent Data Science and Artificial Intelligence (IDEAI-UPC) Research Centre, Universitat Politècnica de Catalunya (UPC Barcelona Tech), Jordi Girona 1-3, 08034 Barcelona, Spain. [2]Knowledge Engineerig Research Group (GREC) at Intelligent Data Science and Artificial Intelligence (IDEAI-UPC) Research Centre, Universitat Politècnica de Catalunya (UPC Barcelona Tech), Jordi Girona 1-3, 08034 Barcelona, Spain. [3]MERITT Group, Institut de Recerca Sant Joan de Déu, Parc Sanitari Sant Joan de Déu, 08830 Sant Boi de Llobregat, Barcelona, Spain. [4]Consorcio de Investigación Biomédica en Red de Salud Mental (CIBERSAM), Instituto de Salud Carlos III, Madrid, Spain. [5]Clinical and Health Psychology Department, School of Psychology, Universitat Autònoma de Barcelona, Bellaterra, 08193 Barcelona, Spain. [6]Department of Psychiatry and Psychotherapy, University Medical Center Hamburg-Eppendorf, 20246 Hamburg, Germany. [7]Departamento de Psiquiatría, Escuela de Medicina, Facultad de Medicina, Universidad de Valparaíso, Valparaíso, Chile. [8]Inserm, University Hospital of Strasbourg, University of Strasbourg, 67091 Strasbourg, France. [9]Experimental Psychopathology Lab, Institute of Psychology, Polish Academy of Sciences, Warsaw, Poland. [10]Caroline König, Pedro Copado, Maria Lamarca and Wafaa Guendouz contributed equally to this work. ✉email: caroline.leonore.konig@upc.edu

who will benefit most from a particular treatment option, as well as predicting the distribution of costs at the individual patient level, is often challenging[7]. Such predictions enable the provision of personalized treatment recommendations.

Mental illnesses are a significant public health issue worldwide and are an important burden in terms of both health and economic losses[8]. Schizophrenia, in particular, has become a public health issue in Europe, with about 2% of the population affected[9]. The core symptoms of this severe mental illness are delusions and hallucinations, but the cognitive and affective domains are also affected. Schizophrenia leads to significant disability and the corresponding societal costs, particularly due to work absence and early retirement[10]. While antipsychotic medications are commonly prescribed, they often fail to improve functional outcomes, and psychological treatment has shown to be beneficial for symptom improvement even without medication[11]. From a psychological perspective to dealing with schizophrenia was introduced an intervention called Metacognitive Training (MCT)[12,13], aimed at reducing positive symptoms in schizophrenia by targeting cognitive biases[14]. This training consists of 10 modules covering different items such as attribution style, jumping to conclusions bias, confirmation bias, social cognition, false memories, affective symptoms, and extra modules on stigma and self-esteem. Recently, a systematic review and meta-analysis of 43 studies[15] found that MCT was effective in reducing delusions, hallucinations, and cognitive biases, reducing negative symptoms to some extent and improving self-esteem and functioning. Some studies suggest that variables such as gender, anxiety, self-esteem, quality of life and level of severity could act as moderators of response to treatment[16–18]. However, the use of ML in the context of personalized MCT requires sizeable volumes of data to build predictive models, so that the diversity of patient profiles and their possible heterogeneous responses to the treatment are represented[19]. Although several meta-analyses have been conducted, incorporating results from various samples and studies[20–24], including the most recent one by[15] gathering 43 studies with a total of 1816 participants, to the best of our knowledge, no prior studies have developed ML-based models for predicting the effectiveness of MCT and personalizing treatment.

The MCT related data considered in this study should ideally comprise clearly labeled outcomes and relevant information about the patients before and after the treatment. Gathering these data involves pooling retrospective heterogeneous data sets and transforming them into one large unified dataset, which could subsequently be used for analysis. This is the domain of data harmonization, which concerns the unification of data of disparate nature that cannot be processed with simple tools and techniques[25,26]. Data harmonization has successfully been used in many health domains, such as, for instance, oncology[27], epidemiology[28] or mental health[29]. Specific projects have researched the development of data harmonization tools in general for medical applications, such as the BioSHaRE project[30] or for secure data sharing in healthcare, such as, for example in the Datashield project[31].

This study describes the data harmonization process conducing to the generation of a harmonized database in the context of the European project ERAPERMED 2022-292 entitled '*Towards a Personalized Medicine Approach to Psychological Treatment of Psychosis*', henceforth referred to as PERMEPSY (https://www.permepsy.org). The participants of the project are five clinical partners, namely Parc Sanitari San Joan de Déu (PSSJD) from Spain as project coordinator, University Medical Center Hamburg-Eppendorf (UKE) from Germany, Polish Academy of Sciences (PAoS) from Poland, the Universidad de Valparaiso (UV) from Chile, University Hospital of Strasbourg (Inserm) from France, and as a technical partner the Universitat Politècnica de Catalunya (UPC) from Spain. In this project, 22 retrospective studies with information from patients who had received MCT were acquired from clinical partners. These datasets contain information on patients with psychosis, before and after undergoing MCT. Despite the digital dataset format and data structure varying from one partner to another, three distinct parts can be distinguished. The first part consists of sociodemographic information describing personal data from the patient; the second part corresponds to the precondition or the patient's clinical, cognitive and metacognitive state prior to MCT; while the third part contains the post-condition, describing the patient's state after treatment. The study comprises the presentation of the harmonized dataset in the Result section and its discussion subsequently. The methods for data harmonization are described in the Method section in detail.

Results

The *PERMEPSY MCT* database integrates information from 22 international retrospective studies with information about the evolution of patients receiving MCT. These studies are from 4 internal partners of the PERMEPSY project and 14 external collaborators. The harmonized database integrates records of 698 patients and 563 attributes. Records contain sociodemographic information, such as age, gender, diagnosis, and employment status as well as the psychological evaluation of the patients before starting treatment (pre-evaluation) and after finishing treatment (post-evaluation), using a variety of psychological indicators. The reader can find an item catalog of the harmonized dataset in the supplementary material. Tables A2–A32 provide information about the nomenclature of variables, data types, admitted values and a categorization by searchable tags.

The harmonization process aimed at integrating the variables of interest by transforming the information to a common format. Nevertheless, given the diversity of data collected in the different studies in origin due to their different aims, the identified variables of interest were not present in all the databases, and in some cases, they were not always complete after harmonization and included missing values. Figure 1 shows the number of MCT patients by source dataset and according to psychological variable indicator.

Sociodemographic variables

The *Permepsy MCT* database registers information about age, gender, marital status, living status, employment status, years spent in education as well as information about the illness, that is, diagnosis itself as well as years of evolution, i.e., the period between the time when the patient was first diagnosed with that illness until the time of evaluation (*length-of-illness*). Furthermore, there is information regarding the consumption of substances, namely caffeine, tobacco, alcohol, cannabis, and illicit substances.
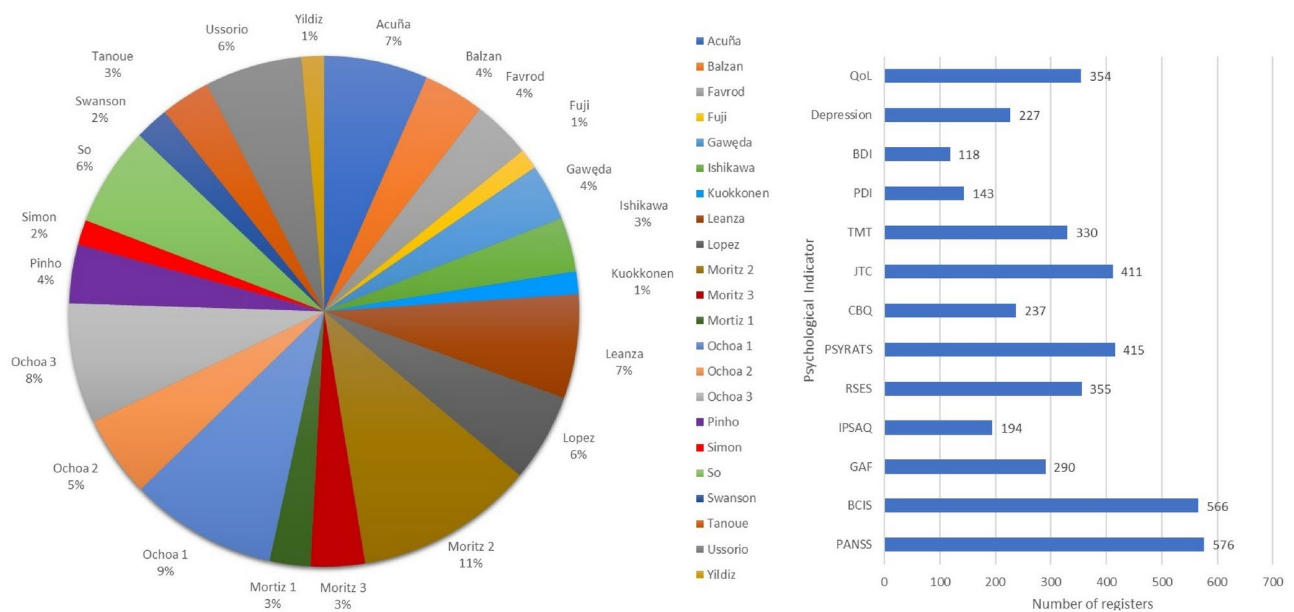
Fig. 1. Data distribution of the harmonized dataset. Number of patients according to source dataset (left) and number of patients with information according to psychological indicator (right).

In detail, the harmonized dataset comprises: (I) *Gender*: male, or female. (II) *Age*: in years. (III) *Education*: years spent in education. (IV) *Marital status*: single, married, separated, widowed, and other. (V) *Living status*: alone, resident with family/parents , living in a residence, and other situations. (VI) *Employment*: inactive, active, temporal illness, retired, disabled, student, and others. (VII) Substance consumption (Yes/No) for *Caffeine*, *Tobacco*, *Alcohol*, *Cannabis*, and *Illicit substances*. (VIII) *Diagnosis*: Schizophrenia, Schizophreniform, Schizoaffective, Delusional Disorder, First Episode of Psychosis (FEP), Brief Psychotic Disorder, Other Psychoses, and Other Disorders. (IX) *Length of illness*: Numerical value describing the length of illness (in years), which is the time between the appearance of the first psychiatric symptoms, for which the patient is now being treated, and the date of assessment.

Psychological indicators

The database comprises information of psychological baseline and post-treatment, before and after the implementation of the MCT therapy, respectively. This study comprises a variety of psychological indicators obtained from the following questionnaires, in detail: (I) *Positive and Negative Syndrome Scale* (PANSS)[32] measuring positive, negative and general symptoms severity. (II) *Beck Cognitive Insight Scale* (BCIS)[33] evaluating the patients' self-reflectiveness and self-certainty in their interpretations of their experiences. (III) *Global Assessment of Functioning* (GAF)[34] assessing the impact of the illness on daily life with a focus on symptomatology and its social impact. IV) The *Internal, Personal, Situational Attributions Questionnaire* (IPSAQ) as a measure of attributional style[35]. (V) The *Rosenberg Self-Esteem Scale* (RSES)[36] evaluating patients' self-esteem. (VI) The *Psychotic Symptom Rating Scale* (PSYRATS) evaluates the severity of positive symptoms, namely delusions and hallucinations[37]. (VII) The *Cognitive Biases Questionnaire* for psychosis (CBQp)[38] captures five cognitive distortions affecting psychotic symptomatology (jumping to conclusions, intentionalising, catastrophising, emotional reasoning, and dichotomous thinking). (VIII) The *Jumping to Conclusion* (JTC) bias tasks testing the individual's tendency to make hasty decisions[39]. (IX) *Trail Making Test* (TMT)[40], a neuropsychological test for attention and task switching. (X) *Peters et al. Delusions Inventory* (PDI)[41] evaluates the presence of several types of delusions and the distress, preoccupation, and conviction of those delusions. (XI) *Beck Depression Inventory* (BDI)[42] for evaluating the severity of depression. (XII) *Harmonized Depression indicator* (DEP) calculated from other depressive symptomatology questionnaires. (XIII) *Harmonized Quality of Life* indicator (QoL). (XIV) *Completion* indicates whether the patient has completed MCT.

Table 1 provides an overview of the available information about the psychological indicators by source. Figure 1 represents graphically the number of registers for each indicator. Table A1 available in the supplementary information details the number of patients collected from each study.

Descriptive statistics and distribution

This section presents a basic overview of the characteristics of sociodemographic information through descriptive statistics and visualizations. A detailed description of the variables Age, Gender, Marital Status, Employment and Living Situation can be found in the supplementary material (Section B).

Data set	PANSS	BCIS	GAF	IPS	RSE	PSY	CBQ	JTC	TMT	PDI	BDI	DEP	QoL
Acuña	X	X					X						X
Balzan	X	X				X		X		X	X		
Favrod	X												
Fuji	X	X	X										
Gawęda			X			X	X	X	X				
Ishikawa	X	X	X		X		X				X		
Kuokkonen	X	X	X			X		X					X
Leanza	X	X		X	X	X			X				
Lopez	X	X	X					X				X	X
Moritz 1	X	X											
Moritz 2	X	X		X	X	X	X	X	X				X
Moritz 3	X	X											
Ochoa 1	X	X	X	X	X	X		X	X		X	X	X
Ochoa 2	X	X	X	X	X	X	X	X	X		X	X	X
Ochoa 3		X	X	X	X			X					
Pinho		X				X							
Simon	X	X											
So	X					X							
Swanson	X		X									X	X
Tanoue	X	X	X		X		X			X			
Ussorio	X	X						X					
Yildiz	X		X										

Table 1. List of psychological indicators present in each source dataset. Please note that the columns IPS, RSE, PSY, CBQ, DEP, and QoL refer to the indicators IPSAQ, RSES, PSYRATS, and CBQp, respectively. Finally, column DEP is the abbreviation for depression, and QoL is the acronym for quality of life.
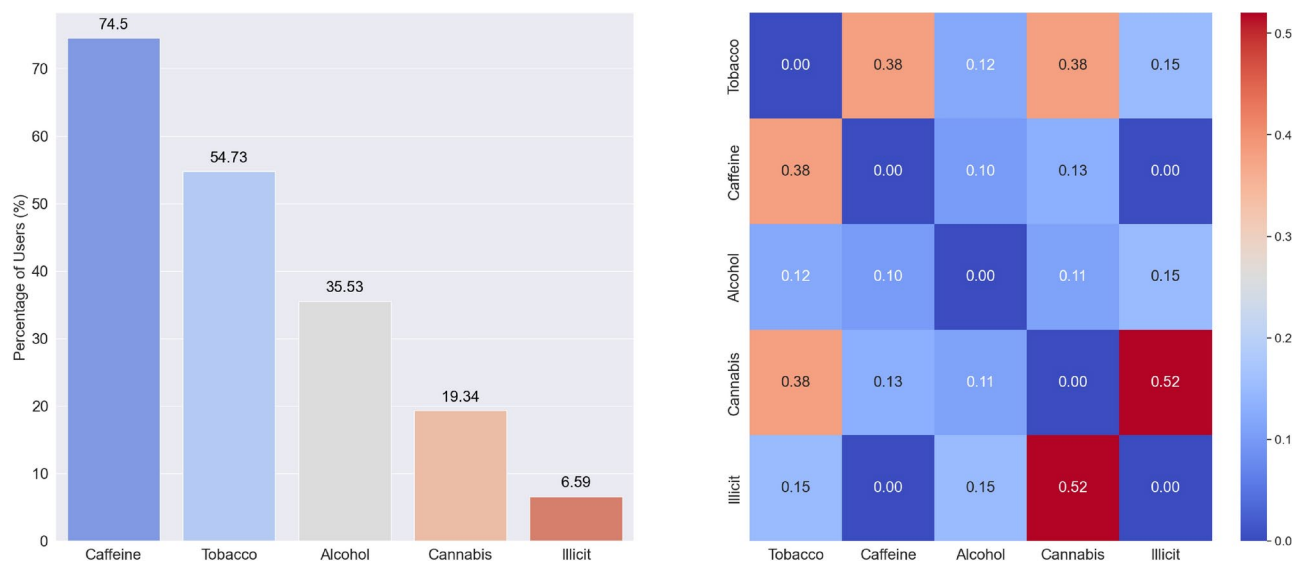
Fig. 2. Overall Prevalence of substance use (left) and Cramér's V heatmap of substance use variables (right). Autocorrelations in heatmap set to zero for sake of visualization.

Substance use profile
This section explores the patterns of substance use among the study participants. For *Caffeine Usage*, the clear majority of the study population, 74.5%, reported using caffeine. Focusing on *Alcohol Consumption*, a significant proportion, 35.53%, reported consuming alcohol. With regard to *Tobacco Usage*, Tobacco use was reported by 54.73% of participants. For *Cannabis Usage*, a smaller section of the studied population, 19.34%, reported using cannabis. For *Illicit Substance Usage*, a very small proportion, namely a 6.59%, reported using illicit substances. For further analysis, the overall prevalence of substance use is shown in Fig. 2. This chart provides an aggregated view of the percentage of users for each substance.

In examining the associations between various substance use behaviors within the study cohort, we decided to calculate Cramér's V statistics[43], revealing several noteworthy relationships, as visualized in Fig. 2. A moderate association exists between tobacco and caffeine usage (V = 0.378), as well as between tobacco and cannabis use (V = 0.381), indicating that individuals who smoke tobacco might also be inclined to consume caffeine and use cannabis. Conversely, tobacco use shows only a weak association with alcohol (V = 0.121) and illicit substances (V = 0.153), suggesting that these substance use behaviors may not be as closely linked in our sample. Caffeine consumption, while prevalent, is not strongly associated with either alcohol (V = 0.100) or illicit substance use (V = 0.000), positioning it as a potentially independent variable in the context of this study. Furthermore, alcohol and cannabis use are weakly associated (V = 0.106), but there is a noticeable moderate to strong relationship between cannabis and the use of other illicit substances (V = 0.516), which could indicate a pattern of poly-substance use.

Diagnosis and illness length analysis

This section examines the types of psychotic disorder diagnoses present in the harmonized dataset and the duration of these mental illnesses among the participants. This analysis provides insights into how common each diagnosis is and the chronicity and variability of psychotic disorders within the study population by plotting the range of time individuals have lived with their conditions. This first chart in Fig. 3 outlines the distribution of various mental disorders, from the more commonly diagnosed Schizophrenia to the less frequent Schizoaffective and Brief Psychotic Disorder. Schizophrenia is the most common diagnosis, significantly surpassing 400 cases. Psychosis Other follows with 94 cases, Schizoaffective with 64 cases and First Episode Psychosis (FEP) with 51 cases. While FEP is not a diagnostic label for any psychotic disorder, in this sample, when specific diagnostic information was unavailable, any patient with a length of illness below 5 years who did not have a different specified diagnosed mental disorder was given the label of FEP. On the other hand, Psychosis—Other refers to labels of psychotic disorders that were either single-case or unspecified. The other diagnoses fall below 50 cases.

Moreover, the length of illness associated with each diagnosis is investigated in the second chart, showing the chronicity and persistence of symptoms over time. Some diagnoses, like 'Other Psychosis, Schizophrenia and Schizoaffective disorders, show a broader spread and potential upper outliers, indicating that there are patients with these diagnoses who have had symptoms for a significantly longer time than others.

The data was grouped by diagnosis and the average length of illness for each diagnosis was calculated. Conditions like Schizophreniform and Brief Psychotic Disorder are on the shorter end of the spectrum (from 0.3 to 1.5 years), indicating acute phases or shorter duration conditions. First Episode Psychosis (FEP) and Delusional Disorder show moderate durations (1.8 to 3.2 years), suggesting these conditions can have a significant, though not necessarily prolonged, impact on individuals. Other and Psychosis-Other represent a broader category of mental health conditions with varied durations, pointing towards the diverse nature of psychosis-related disorders. Schizoaffective Disorder and Schizophrenia are at the longer end of the spectrum (from 9 to 10.4 years), reflecting their chronic nature and the ongoing impact they have on patients' lives.

Statistical analysis of psychological indicators

The correlations between the main psychological indicators both in baseline and post-treatment were analyzed using Pearson's correlation coefficient. The heatmap representation of the correlations is provided in Fig. 4. The focus is on the correlation between independent types of psychological indicators. The analysis shows that the PANSS positive scores are strongly associated with the PSYRATS Delusion and PSYRATS Hallucination indicators (0.61 and 0.74 in pre-evaluation and 0.54 and 0.74 in post-evaluation). The GAF indicator shows a strong negative association with PANSS scores with values in the range of –0.8 to –0.88 in pre-evaluation and –0.79 to –0.82 in post-evaluation. These strong associations are congruent with the indicators measured by the scales. The PANSS Positive subscale measures, among other positive symptoms, the presence of hallucinations and delusions, while the PSYRATS measures the presence, frequency and severity of hallucinations and delusions. As a results, strong positive associations between the Positive PANSS subscale and both PSYRATS subscales is to be expected. Likewise, the GAF scale measures the impact of symptomatology on patients' daily
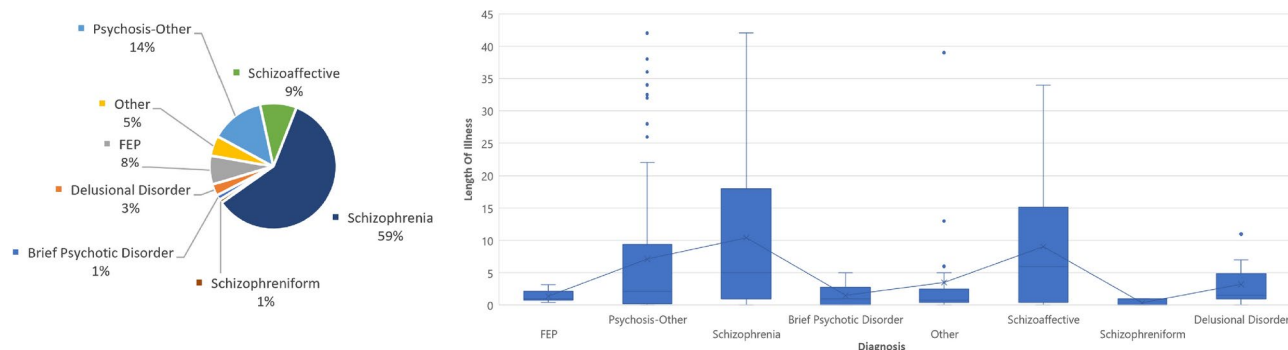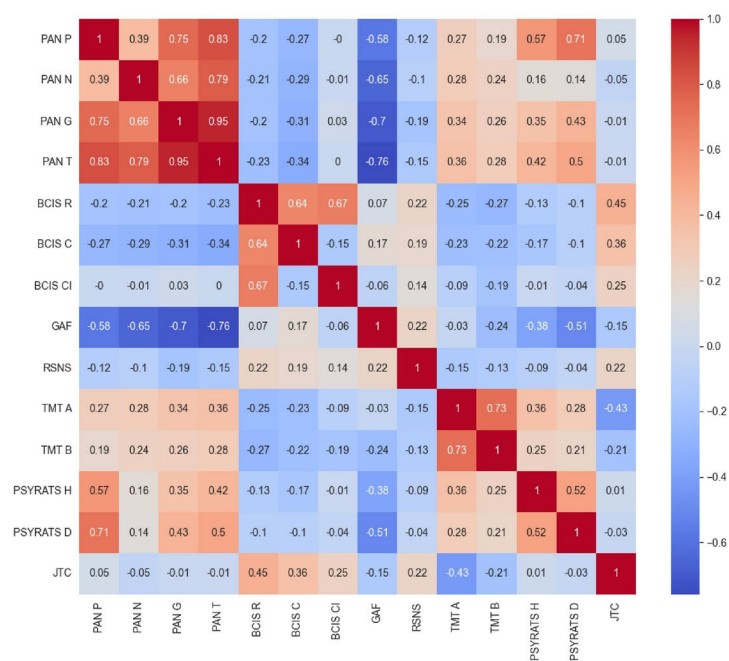

Fig. 3. Distribution of diagnosis (left) and boxplot representation of the distribution of the illness length across type of diagnosis (right).
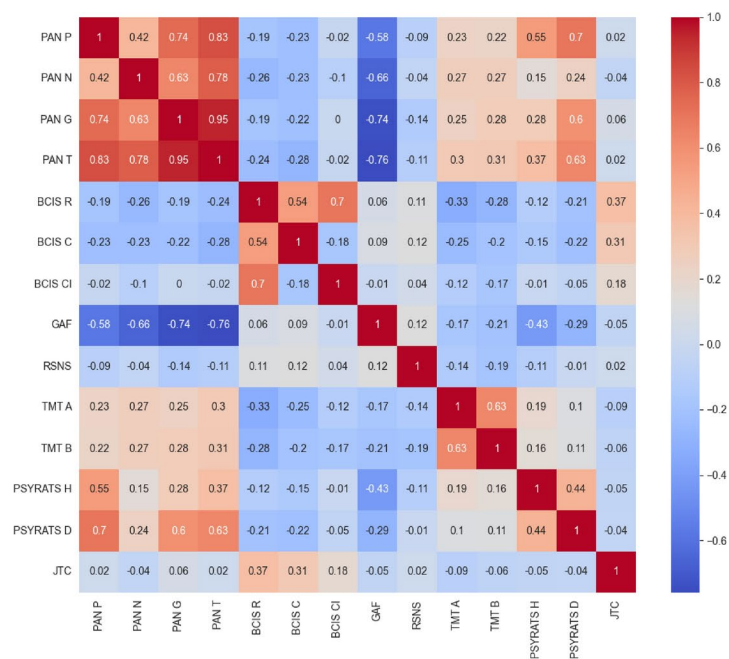
(a) Baseline

(b) Post-treatment

Fig. 4. Heatmap representation of the correlations of the psychological indicators at baseline and post-treatment.

functioning, as a result, increased symptom presence and severity is expect to negatively impact functioning, resulting in lower GAF scores.

Discussion

The resulting harmonized dataset comprises information from 698 patients who have undergone MCT. A common set of sociodemographic attributes was derived from the 22 source datasets and the possibilities to

integrate information into a common format were considered. The information availability criterion was key to decide on the integration of a variable. For example, even though the information about handedness of patients was of interest, only few studies collected this information and it was not possible to integrate it in the harmonized dataset. Regarding the availability of psychological indicators, the harmonization process comprised the analysis of 12 indicators and the calculation of two derived psychological indicators (QoL and depression indexes). Additionally, the variable used to assess the completion of MCT was also included in the harmonization process. Integration of psychological indicators required the greatest effort during the data harmonization process due to the high number of variables and scores included in each indicator (see Table A2 of the supplementary material section for a summary). As presented in Fig. 1, the harmonized dataset included information from individual indicators for each patient in as much detail as possible. Nevertheless, information from psychological indicators differed in the number of available data points, leading to registers with high amounts of unknown values. In general, the PANSS, BCIS, and GAF were the indicators with the largest amount of available information, while information on the PDI indicator was the scarcest one.

The exploratory data analysis of participant sociodemographic information provides interesting insights on their most common characteristics. This analysis will be extended in future research with more advanced data analytical approaches, such as clustering, to discover meaningful patient groups.

The systematic approach to data harmonization presented here has enabled the successful integration of a large amount of diverse information from a very specific field of research. The identification of variables of interest and the definition of the common nomenclature of variables was key to successfully integrate the 563 variables from 22 source datasets. The implementation of the logic of variable mapping and transformation by means of Python programs allowed to efficiently handle such amount of information. The benefits of using a programmatical approach include enabling data harmonization process reproducibility with the possibility of correcting errors in variable mapping, transformation, and data quality assessment. Nevertheless, data understanding and domain knowledge were essential to define the mapping of variables and the design of transformation rules.

The characterization of depression as a binary variable (the patient either suffers or does not suffer depression) may seem an oversimplification of a nuanced pathology. That is, we are not assessing it in terms of its severity, or the subtleties of depressive symptomatology. This is, in fact, a trade-off for analytical purposes: in this multi-center study, the instruments used to measure depression in the original studies were different at different centers. From the point of view of a data harmonization process oriented towards the quantitative analysis of the resulting unified dataset, the binary characterization of depression, using the cutoff points established in the literature[44–46], was deemed to be the most adequate approach to guarantee the harmonization of a large enough dataset for subsequent analysis. It is a particularly useful approach for diagnostic and screening purposes in its practical application, but precludes a more in-detail analysis of an heterogeneous condition with a wide spectrum of severity.

Methods

Data harmonization is a process of conciliating various types, levels, and sources of data into compatible and comparable formats to ensure that the data can be effectively utilized for better decision-making[47]. It is a complex process involving several steps to resolve heterogeneity at the level of syntax, structure, and semantics and bring the conceptually similar information into a common taxonomy or ontology[48]. A common approach of unifying conceptual information is merging[49] where a single global taxonomy of the concepts is developed.

The methodological approach for the data harmonization applied in this study is that of retrospective data harmonization, as already existing conceptually similar data is analyzed. The approach comprises several steps: The selection of data sources, the identification of variables of interest (definition of the taxonomy), the mapping of variables, data transformation by transformation rules, and a postprocessing step for data cleaning (Fig. 5). The steps of variable mapping and transformation rules mainly comprise merging techniques as the conceptually related information is mapped to a unique global taxonomy.

Data harmonization requires not only technical skills to handle data but also knowledge of the significance and taxonomy of the information to select relevant datasets and specify the necessary transformation rules to bring the data to a common format. In this research, the harmonized database was reached through the collaborative effort of all PERMEPSY stakeholders, both technical and clinical, in an 8-month-long collaboration.
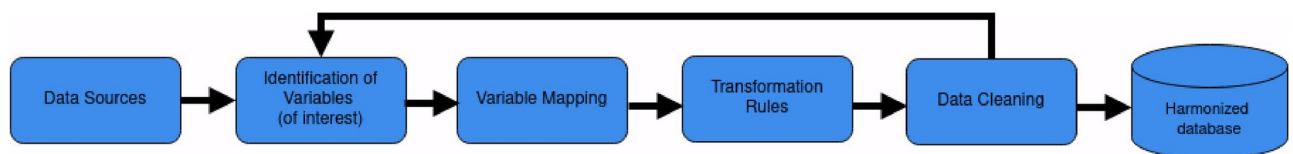
Fig. 5. Harmonization process, which is composed of these steps: Selection of potential data sources or datasets, thorough revision of the datasets identifying the variables of interest, mapping the variables of each dataset with the aggregated harmonized variable, applying transformations to each of the variables to homogenize them or also implementing new variables, and finally cleaning the data. Note that the last three steps have been automated using a Python program on Jupyter Notebooks. Therefore, all phases are reviewable, and errors can be quickly rectified.

DB Name	Origin	Country	Ref	Rows	Cols	MCT
Acuña	UV	Chile	–	47	136	47
Balzan	Outside Consortium	Australia	[50]	55	607	27
Favrod	Outside Consortium	Switzerland	[51]	52	126	52
Fuji	Outside Consortium	Japan	[52]	17	77	8
Gawęda	PAoS	Poland	–	48	343	25
Ishikawa	Outside Consortium	Japan	[53]	50	256	24
Kuokkanen	Outside Consortium	Finland	[54]	10	150	10
Leanza	Outside Consortium	Germany	[55]	93	1548	46
Lopez	Outside Consortium	Spain/France/Chile/UK	[56]	77	147	39
Moritz 1	UKE	Germany	[57]	45	447	18
Moritz 2	UKE	Germany	[58]	202	3054	79
Moritz 3	UKE	Germany	[57]	48	595	24
Ochoa 1	PSSJD	Spain	[59]	122	1897	65
Ochoa 2	PSSJD	Spain	[60]	70	1560	36
Ochoa 3	PSSJD	Spain	-	54	824	54
Pinho	Outside Consortium	Portugal/Spain/Turkey/USA	[61]	52	509	27
Simon	Outside Consortium	Spain	[62]	22	110	11
So	Outside Consortium	China	[63]	44	520	44
Swanson	Outside Consortium	UK/Sweden/Germany	[64]	15	884	15
Tanoue	Outside Consortium	Japan	-	23	445	23
Ussorio	Outside Consortium	Italy/Germany	[65]	44	82	44
Yildiz	Outside Consortium	Turkey	[66]	20	47	10

Table 2. Summary of initially collected datasets: from left to right, the partnership with the project, associated journal publication, dataset size (rows and columns), and an estimate number of MCT patients (MCT). *PAoS* Polish Academy of Sciences, *UKE* University Medical Center Hamburg-Eppendorf, *UV* Universidad de Valparaíso, *PSSJD* Parc Sanitari Sant Joan de Déu.

Technically, computational tools such as Python (http://www.python.org) and Jupyter Notebooks (jupyter.org) have been used for automated data processing. Python programming language is a well-suited data analytical environment for processing and transforming datasets due to their versatility, interactivity, and extensive ecosystem of tools and libraries, such as Pandas. On the one hand, Jupyter Notebooks provide an interactive computing environment where users can seamlessly integrate code execution, visualization, and documentation, facilitating an iterative and exploratory approach to data processing. On the other hand, the combination of Python and Jupyter Notebooks promotes reproducibility, transparency, and efficiency in data processing and harmonization. The Python code used for data harmonization is freely available in a public repository at https://gitlab-rdlab.cs.upc.edu/soco-permepsy/harmonization.

Data sources

The first phase of the data harmonization approach focuses on the selection of MCT-related patient datasets. The entities providing the MCT-related patient datasets as specified in Table 2 confirm that data collection and data sharing methods were carried out in accordance with relevant guidelines and regulations, as well as informed consent was obtained from all subjects and/or their legal guardian(s). The data has been appropriately anonymized before sharing. The study of the MCT-related retrospective data has been approved by the respective ethical committees, namely the Research Ethics Comitee of Fundació San Joan de Déu on 27/04/2023 by approval PIC-68-23, the 'Lokale Psychologische Ethikkommission am Zentrum für Pyschosoziale Medizin' of UKE on 29/03/2023 by approval LPEK-0603, the 'Comitè Ético Científico del Servicio de Salud Valparaíso San Antonio' by approval $N°54/2023$ on 27/09/2023, and the 'Ethics Comitee of the UPC' on 11/12/2023 by approval 2023.13.

A set of 24 datasets from different studies were collected for potential integration in the harmonized database. All selected datasets underwent a thorough review, compiling an initial inventory of all potential variables appearing in each dataset. After this review, datasets with no relevant information were excluded from further consideration. Thus, initially, there were 24 datasets, and eventually, two were discarded, with 22 datasets remaining. Table 2 shows a summary of the MCT studies used for this research in the harmonized database and preliminary counts of patient numbers and the relevant variables associated with each dataset.

In this phase, the data sources have undergone different types of preprocessing, such as a conversion to a standard data file format (csv) and transformation to a uniform encoding. Preprocessing comprised the filtering and removal of non-relevant records. As the aim of this work is the investigation about response to MCT treatment, only records of patients having received MCT were selected for integration into the harmonized dataset.

Target variables

The second step of the data harmonization approach is the identification of the variables of interest or target variables to become part of the harmonized database and its location in the different individual datasets. Research about the most commonly used information in MCT assessment was done using a similar approach as proposed in[67] for the identification of constructs of interest. This analysis revealed two types of target variables, sociodemographic information of patients and clinical questionnaires. Sociodemographic variables refer to a set of characteristics that describe individuals in terms of their identity and situation in society. On the other hand, some clinical questionnaires such as the PANSS have the purpose of assessing and measuring symptoms related to psychotic disorders.

A systematical analysis for each target variable revealed information about the availability and type of representation of the variable under study in each individual dataset. For sociodemographic variables, the goal was to identify the present attributes and the corresponding categories. For the case of the questionnaires, the inquiry was about the number of questions, responses, number of categories in responses, number of scores, how scores are calculated, and the number of evaluations. Regarding the evaluations, the initial evaluation (baseline or pre-evaluation) and the final evaluation (post-treatment or post-evaluation) were found as most commonly used in most studies and therefore selected as relevant for harmonization discarding intermediate and follow-up evaluations. Another important aspect of harmonization was the discrimination between MCT patients and non-MCT patients (controls), as only the former ones were of interest for the harmonized database as variability in control interventions did not allow for valid harmonization of non-MCT patient data.

Although the variable identification approach was automated as most as possible, based on text search assuming a common nomenclature of variables, the final identification of attributes was conducted manually using a suitable tool to visualize and inspect the dataset and with the assistance of clinical partners, determine which data was relevant for harmonization. The sociodemographic variables of interest comprise 13 variables describing age, gender, and other information about patients' education, living status, and use of substances. For the assessment of psychological status, a range of 12 diverse psychological questionnaires of interest were identified as described in the section of Results. These 12 psychological indicators comprise 274 variables. In consequence, the dataset comprises this number of variables each at pre-evaluation and at post-evaluation, thus yielding a total of 563 attributes.

Variable mapping

In this phase, a pre-harmonized or intermediate dataset is generated. This task involves extracting relevant information related to variables of interest from each dataset and aggregating it into a single dataset.

The variable mapping is a crucial phase, as it defines the reference model[68] by establishing, for each dataset and each variable, the 'correspondence' of names between the attributes of the dataset and those of the harmonized dataset. Additionally, it indicates the name of the attribute (or column) from where to extract the columns for adding them to the pre-harmonized dataset. Before making this correspondence, the harmonized variable names were agreed on. In general, the applied rules for variable naming were to use the prefix 'pre' or 'post' defining whether the variables were collected in pre or post-evaluation, followed by the primary name of the questionnaire (PANSS, PSYRATS, etc.) and followed by a suffix describing whether the variable was a total score or a subscore. Once the taxonomy of harmonized variables was defined, the mapping from source variables was then defined for each variable of interest. Please refer to Figure B2 in the supplementary materials for a description of the taxonomy of psychological indicators of interest in this work.

For variable mapping, a Python program was developed to automate this task. This program takes all the datasets and the mapping of each attribute-dataset pair to its corresponding harmonized attribute as input, and outputs a single pre-harmonized or intermediate dataset. Therefore, it is necessary to specify this mapping so that this program considers it. This mapping process has the advantage of being reviewable, so any detected errors can be quickly corrected. The aggregated dataset (pre-harmonized dataset) comprised 828 patients and 1350 aggregated variables. The number of variables in the aggregated dataset is higher than the number of target variables due to the differences in registering information. In many instances, it was necessary to map and aggregate various variables of a source dataset to the harmonized variable in the pre-harmonized dataset, so that the relevant information in the pre-harmonized dataset was aggregated for its further transformation in a subsequent step. For example, in the case of the IPSAQ questionnaires, some studies used only one attribute to register the choice the patient made with the most confidence, while other studies used three attributes to register the patient's confidence in the three answers related to one IPSAQ question.

Variable transformation

From the previous phase, an intermediate or pre-harmonized dataset in raw data format is available. In this phase, the aggregated variables are analyzed by different transformation rules, which aim to bring the variable into a common format. The transformation rules focus on the analysis of the range of values validating whether variables are inside the theoretical range or applying the necessary transformations to bring data to the correct range. For example for the PANSS questionnaire, it was quite frequent to see studies use a 7-point Likert-type scale from 0 to 6 to describe the severity of symptoms from absent to extreme, while other studies used the official 7-point Likert scale ranging from 1 to 7. There were specifically implemented transformation rules to ensure the accuracy of attributes, such as total scores summing individual scores correctly, for example, PANSS positive scores summing the seven positive subscores. Another type of transformation rules are those related to the creation of new indicators from the source data. For example, a harmonized depression or harmonized quality of life indicator was created to integrate the information in a unified format.

Sociodemographic variables

This section summarizes the transformations for the sociodemographic variables. For *Age* the original values were retained representing the age in years. For *Gender*, the categories 1 for Males and 2 for Females were applied. For *Civil Status* categorical variables ranked from 1 to 5 were agreed with categories Single (1), Married (2), Separated (3), Widow (4), and Other (5). The *Living Status* is a categorical variable ranked from 1 to 4 with the following definitions: Alone (1), Resident with family/parents, other relatives (2), residence/assisted living facility (3), and Other situations (4). *Employment* is a categorical variable ranked from 1 to 7 with the following definitions: Inactive (1), Active (2), Temporal illness (3), Pensionist (4), Disabled (5), Others (6), and Students (7). The variable *Years of studies* appears to have different meanings depending on the dataset. Therefore, it was necessary to divide it into two numerical harmonised variables. These variables: *Basic education* (Edu_basic) describes the number of years for education at school, ranked by 0-18. *Further education* (Edu_ampl) describes the number of years for education at school plus years of professional education if this information is available in the original dataset, ranked by 0-28. The information about the illness is a categorical variable for the *Diagnosed Illness* comprising the categories: Schizophrenia (1), Schizophreniform (2), Schizoaffective (3), Delusional disorder (4), First Episode of Psychosis (FEP) (5), Brief Psychotic Disorder (6), Other Psychosis (7) and Others (8). The criteria applied for deciding on the diagnosis during data harmonization was to use Schizophrenia, Schizoaffective Disorder, Delusional Disorder, and Schizophreniform Disorder according to the DSM-5 categories[69]. The category 'Other-Psychosis' (other psychotic disorders) was assigned as a response to the variety of other diagnoses that either did not match the DSM-5 categories or were not found more than once in the full sample during data harmonization, as was the case for example for Psychosis not otherwise specified or schizotypal disorder. The category 'Brief Psychotic Disorder' refers to cases where the episode was less than 1 month long and the category 'First Episode of Psychosis' (FEP) refers to cases where the episode was more than one month but still without a definitive diagnosis and there was not available other information about the diagnosis. *Length-of-illness* is a numerical value describing the number of years that have passed since the patient was first diagnosed with the illness they were being treated for. Furthermore, there was information regarding the consumption of substances, namely, *Caffeine*, *Tobacco*. *Alcohol*, *Cannabis* and *Illicit substances*, coded as a binary variable with values NO (1) and YES (2).

Psychological indicators

This section explains the transformation rules that have been applied to the psychological indicators presented in the previous section in the harmonization process. In general, in the transformation process, the following three operations were applied to each questionnaire whenever possible: checking the ranges of the attributes, recalculating the scores, and the replacement of missing values and erroneous entries with 'NA' (Not Available). In summary, these operations were performed to ensure the accuracy and reliability of the data. By checking the ranges of the attributes, it was possible to identify and correct any invalid data. By recalculating the scores, it was possible to ensure that the scores were based on the correct values of the attributes. Table A2 summarises the ranges of values for the subscores (items) as well as the total scores (Aggregates).

<u>Positive and negative syndrome scale</u>   The PANSS[32] questionnaire stands out in our study for being the one that appears most frequently. It is composed of 30 questions and 4 scores (see Eq. 1). The 30 questions are divided into 3 sections, each one intended to evaluate different spectra of symptoms. The severity of the symptoms measured by each item is rated with a 7-point scale, where 1 means absent and 7 means extreme.

- Positive symptoms, named PANSS Positive ($PANSS_P$), includes items 1 to 7.
- Negative symptoms are also referred to as PANSS Negative ($PANSS_N$) items 8 to 14.
- Finally, the General symptoms or PANSS General ($PANSS_G$), which involves summation items 15 to 30.

Three of the sub-total scores correspond to the sum of each of the sections of the $PANSS_P$, $PANSS_N$ and $PANSS_G$, except the last one, which adds up the total score for the PANSS Total $PANSS_T$. Therefore, the ranges for each of the scores are $7 - 49$ for $PANSS_P$ and $PANSS_N$, 16 to 112 for $PANSS_G$, and 30 to 210 for $PANSS_T$. In the harmonization process of the PANSS variable, some datasets were found where all items were present, while others contained only the aggregated values ($PANSS_P$, $PANSS_N$, $PANSS_G$, $PANSS_T$). In the former scenario, only items $1 - 30$ are extracted, from which aggregated values are recalculated. This recalculation is conducted after verifying the correctness of the ranges (1-7) and missing values. Conversely, in the latter case, where only aggregated PANSS scores are present, only the aggregated values are extracted and the correctness of their ranges is checked (see Table A2). For $PANSS_P$ and $PANSS_N$, the valid range of values is 7 to 49; for $PANSS_G$ the valid range is 16 to 112; and for $PANSS_T$ the valid range are 30 to 210.

$$PANSS_P = \sum_{i=1}^{7} item_i$$
$$PANSS_N = \sum_{i=8}^{15} item_i$$
$$PANSS_G = \sum_{i=15}^{30} item_i$$
$$PANSS_T = PANSS_P + PANSS_N + PANSS_G$$

(1)

The PANSS indicators extracted from the datasets have been presented in two forms: (1) including all items, including the subscales; (2) only the subscales. All datasets have maintained the range and number of items; only

some datasets have retained just the subscales. In both cases, to process the data, the ranges of each item have been checked (see Table A2), and the Eq. (1) has been applied to calculate the subscales.

Beck cognitive insight scale

The BCIS questionnaire[33] consists of 15-items, 2 sub-scales ($BCIS_R$, $BCIS_C$) and a total score ($BCIS_T$). Subscale $BCIS_R$ measures the self-reflectiveness, which is calculated as the sum of 9-items given in this set $S = \{1, 3, 4, 5, 6, 8, 12, 14, 15\}$, whereas subscale $BCIS_C$, called self-certainty, is calculated as the sum of the remaining 6-items. The total score $BCIS_T$ is calculated as the difference between $BCIS_R$ and $BCIS_C$ as shown in Eq. (2).

$$
\begin{aligned}
BCIS_C &= \sum_{i \in S} item_i \qquad\qquad \text{where } S = \{1, 3, 4, 5, 6, 8, 12, 14, 15\} \\
BCIS_R &= \sum_{i \notin S} item_i \\
BCIS_T &= BCIS_R - BCIS_C
\end{aligned} \tag{2}
$$

The BCIS questionnaire employs a 4-point scale, with responses ranging from 0 ("do not agree at all") to 3 ("agree completely"). This scale determines the possible score ranges for each subscale and the total score. The $BCIS_R$, scores range from 0 to 24, $BCIS_C$, scores range from 0 to 36, and $BCIS_T$ scores range from -24 to 36. The indicator was implemented identically in almost all datasets, including both the response ranges and the scale calculation. Therefore, the value ranges and scales have been checked and recalculated to harmonize the indicator. Only the value ranges have been checked in the datasets where only the sub-scales were available. Finally, erroneous values are replaced with 'NA." The BCIS questionnaires extracted from the datasets, similarly to PANSS, have been presented in two formats: encompassing all items, including the subscales, and featuring only the subscales. Therefore, in each dataset, verifying the range of each item (see Table A2) and employing the corresponding Eq. (2) for subscale calculation.

Beck depression inventory

The BDI indicator[42] consists of 21 questions, each with only three responses. Responses are scored from 0 to 3, with only one total score. Therefore, this total score varies between 0 and 63, as shown in Eq. (3).

$$
BDI_{total} = \sum_{i=1}^{21} item_i \tag{3}
$$

In the case of the BDI, all datasets containing it presented all the items. Therefore, as with previous indicators, the range of values for the items, which is 0-3, has been checked, replacing values outside this range with 'NA.' Additionally, the total score is calculated using Eq. 3. In the following section, BDI is used to construct the depression indicator.

Cognitive biased questionnaire for psychosis

The CBQp questionnaire[38] comprises 30 questions, with responses ranging from 1 to 3, and encompasses the calculation of 6 scales. Among these scales, 5 scales are categorized into the 5 measured biases, labeled I (Intentionalising), C (Catastrophising), DT (Dichotomous Thinking), JTC (Jumping to Conclusions), and ER (Emotional Reasoning), while the final scale represents the total sum $CBQp_t$.

$$
\begin{aligned}
CQB_I &= \sum_{i \in I} item_i & \text{where } I &= \{1, 3, 20, 22, 23, 28\} \\
CQB_C &= \sum_{i \in C} item_i & \text{where } C &= \{2, 4, 7, 10, 12, 25\} \\
CQB_{DT} &= \sum_{i \in DT} & \text{where } DT &= \{5, 11, 14, 15, 27, 30\} \\
CQB_{JTC} &= \sum_{i \in JTC} & \text{where } JTC &= \{6, 13, 16, 19, 24, 26\} \\
CQB_{ER} &= \sum_{i \in ER} & \text{where } ER &= \{8, 13, 16, 19, 24, 26\} \\
CQB_{total} &= CQB_I + CQB_C + CQB_{DT} + CQB_{JTC} + CQB_{ER}
\end{aligned} \tag{4}
$$

For the case of CBQp, similar to the previous questionnaires, no major transformation was required. Simply, the range of items was checked in those datasets where all items were stored. However, there were also datasets that only stored the subscales; for this scenario, the ranges were also checked. Likewise, as in previous sections, the subscales were recalculated using Eq. (4).

Global assessment of functioning

The GAF questionnaire, measures how much a person's symptoms affect their day-to-day life on a scale of 0 to 100. It was devised to help psychiatrists and psychologists understand how "well" a patient can carry out everyday activities. The score can determine what intensity of treatments might work. In the case of the GAF, which consists of only a single item, only the range of values between 0 and 100 has been verified.

Psychotic symptom rating scale

The PSYRATS questionnaire[37] comprises 17 questions with responses categorized on a scale from 0 to 4. This metric computes two scores, $PSYRATS\_H$ and $PSYRATS\_D$, as indicated in Equation 5. As with the previous indicators, ranges were checked, and scores were recalculated.

$$PSYRATS\_H = \sum_{i=1}^{11} item_i$$
$$PSYRATS\_D = \sum_{i=12}^{17} item_i$$

(5)

Trail making test

The Trail Making Test (TMT)[40] assesses the subject's neurofunctioning through two tests, the numerical one called TMT_A, and the alphanumeric one called TMT_B. Usually, the scores achieved in the tests are recorded, but the time taken is also available. Therefore, the variable may contain either 2 or 4 items depending on the dataset. However, we only retain the obtained score.

Jumping to conclusion bias

The JTC indicator[39] is binary, with values between 0 and 1. Thus, 0 represents 'No,' and 1 represents 'Yes.' In the datasets, it is not directly available in binary form but rather as a measure how many turns of the JTC task it took for the patient to reach a decision, where JTC bias is present when the patient made a decision in less than three turns. Therefore, it was necessary to establish a threshold L , beyond which JTC was set to 1.

Peters et al. delusion inventory

The Peters et al. Delusion Indicador[41] indicator consists of 21 questions. Each question has four sub-questions: one with a binary response (0 or 1) and three with responses ranging from 1 to 5. Note that if the response of the first sub-question is 0, then responses to the other three sub-questions are invalidated, taking 0 as their value (see Eq. 6). Four scores are calculated corresponding to each sub-question. The procedure followed for this indicator is the same as that for previous indicators.

$$PDI_{tot} = \sum_{i=1}^{21} b_i$$
$$PDI_{ans} = \sum_{i=1}^{21} item_i^a \cdot b_i$$
$$PDI_{preo} = \sum_{i=1}^{21} item_i^b \cdot b_i$$
$$PDI_{conv} = \sum_{i=1}^{21} item_i^c \cdot b_i$$

(6)

Rosenberg self-esteem scale

The RSES is composed of 10 questions with 1–4 values as a response. In order to compute the total score, it is required to apply transformation to each items that belong to set I (see Eq. 7). After this transformation, the total score is a simple sum.

$$RSES_{total} = \sum_{i \in I} item_i + \sum_{i \notin I} (5 - item_i) \quad \text{where } I = \{2, 5, 6, 8, 9\} \subseteq \{1 \dots 10\}$$

(7)

Internal, personal, situational attributions questionnaire

The IPSAQ indicator[35] consists of 32 questions with responses categorized as 1, 2, and 3. Eight scores are calculated, four computed by dividing the 32 questions into two sets, *P* (Positive events) and *N* (Negative events), and counting the number of responses in each category depending on whether the response referred to an internal personal attribution *I*, an external personal attribution *P* or an external situational attribution *S*, and considering whether the question belongs to *P* or *N* ($IPSAQ_{IP}$, $IPSAQ_{IN}$, $IPSAQ_{PP}$, $IPSAQ_{PN}$, $IPSAQ_{SP}$, $IPSAQ_{SN}$). The remaining two scores are calculated based on the previous as indicated in Eq. (8).

$$IPSAQ_{IP} = \sum_{i \in P, \, item_i=1} 1$$
$$IPSAQ_{IN} = \sum_{i \in N, \, item_i=1} 1$$
$$IPSAQ_{PP} = \sum_{i \in P, \, item_i=2} 1$$
$$IPSAQ_{PN} = \sum_{i \in N, \, item_i=2} 1$$
$$IPSAQ_{SP} = \sum_{i \in P, \, item_i=3} 1$$
$$IPSAQ_{SN} = \sum_{i \in N, \, item_i=3} 1$$
$$IPSAQ_{SESG\_EXT} = IPSAQ_{IP} - IPSAQ_{IN}$$
$$IPSAQ_{SESG\_PER} = \frac{IPSAQ_{PP} + IPSAQ_{PN}}{IPSAQ_{PN} + IPSAQ_{SN}}$$
$$P = \{1, 4, 5, 7, 8, 11, 14, 15, 16, 10, 22, 25, 26, 29, 31, 32\}$$
$$N = \{1 \dots 32\} \setminus P$$

(8)

Quality of life

The QoL (Quality of Life) variable is a derived indicator reflecting the patient's quality of life. According to the World Health Organization (WHO), the term quality of life is defined as "an individual's perception of their position in life in the context of the culture and value systems in which they live and about their goals, expectations, standards, and concerns." The assessment of quality of life is found in several datasets that utilize different psychological indicators: (I) Quality of Life Enjoyment and Satisfaction Questionnaire-18-item (Q-LES-Q-18)[70]: Found in the dataset of Swanson. (II) WHO Quality of Life Scale-Brief (WHOQOL-BREF)[71]: Found in UKE 2, UKE 3, and Leanza. (III) Satisfaction with Life Domains Scale (SLDS)[72]: Found in SJD 2, SJD 3, UV, Lopez. (IV) Quality of Life Scale (QoLS)[73]: Found in Yildiz.

To align with a common criterion for quality of life, all these indicators have been consolidated into a single metric named QoL. This derived QoL metric consists of three categories: 0, 1, and 2, which correspond to low, medium, and high levels of quality of life, respectively. To accomplish this transformation, each indicator was reviewed, and thresholds were determined based on the literature[72,74] to convert them into the categories 0, 1, and 2. For example, for the SLDS indicator presented in the datasets Ochoa 1-3, UV, and Lopez, all patients were aggregated to obtain a distribution of the SLDS indicator. Based on this criteria it was inferred that the patients with critical levels below the 33% and 66% percentiles were classified accordingly. Therefore: (I) Below the 33% percentile corresponds to 0. (II) Between the 33% and 66% percentiles corresponds to 1. (III) Above the 66% percentile corresponds to 2.

For SLDS scores, the 33% percentile threshold was 72, and the 66% percentile threshold was 84. Concerning to WHOQOL-BREF indicator, we applied similar reasoning based on the following source[14]. The 33% and 66% percentiles were calculated to classify the values of WHOQOL-BREF. The corresponding thresholds are 81 and 94 for the 33rd and 66th percentiles, respectively. Regarding the Q-LES-Q-18 indicator, thresholds of 54 and 60.74 were determined according to[75]. Finally, for the QoLS indicator, which is ranked from 21 to 126, it is assumed that values greater than 105 equal 2, values between 22 and 104 equal 1, and values less than 21 equal 0 according to[76].

Depression

According to the World Health Organization (WHO), Depression or Depressive disorder, is a common mental health condition that can happen to anyone. It is characterized by a low mood or loss of pleasure or interest in activities for long periods. This variable is considered relevant but has not been implemented directly in any dataset. Therefore, an indicator called 'Depression' was intended to be created from some of the indicators available in the source datasets, namely from the following indicator: (I) the Beck Depression Inventory (BDI) indicator[42] in *Ochoa 2, Ochoa 3, Ishikawa, Tanou*. (II) the Depression Anxiety Stress Scale -21 item (DASS-21) indicator[77] in *Balzan*. (III) the Calgary Depression Scale for Schizophrenia (CDSS) indicator[78] in *Lopez* and *Swanson*.

In order to harmonize information from different scales with varying thresholds and ranges, the Depression binary indicator is represented by the categories 0 and 1, which are interpreted as YES if the patient presents with Depressive symptomatology or NO if the patient does not present with Depressive symptomatology according to the established thresholds of each scale used in the different databases harmonised. To transform these previous indicators into a single binary indicator, each of these indicators has undergone a rigorous review process, and a threshold δ has been inferred to serve as the separation point between the categories 0 and 1. The BDI indicator is present in the datasets Ochoa 1–3, Ishikawa, and Tanoue. The BDI, as mentioned before, has a total score within the range of 0–63. According to the National Institutes of Health[44], a score between 20–63 on the BDI can suggest moderate to severe levels of depressive symptomatology, so if the total score is strictly less than 20, the Depression indicator takes the value of 0. Otherwise, Depression takes the value of 1. Similarly, for the DASS-21 and CDSS indicators, their ranges are 0–42 and 0–27, respectively. A score above 14 on the DASS-21 depression subscale suggests moderate to extremely severe (28+) levels of depressive symptomatology[45], while a score above 6 in the CDSS suggests moderate to severe levels of symptoms of depression[46]. Thus, the proposed classification thresholds are 14 ($14 \geq$ YES, and NO ≤ 14) and 6, respectively.

Completion of MCT

The Completion indicator determines whether a patient has completed MCT treatment. It is a relevant datum that allows for characterizing the patient who will abandon the treatment and determining possible causes of treatment abandonment. Despite the relevance of the datum, it was not included directly in any dataset. Therefore, an indicator called "Completion" has been derived, composed of two categories 0 and 1, interpreted as 0 "not completed" treatment and 1 as "completed" treatment. To develop this indicator, we start from the harmonized dataset, which is divided into three sections: sociodemographic, baseline, and post-treatment. Given a patient with a post-treatment section with all indicators with *unknown* values (NA value), it is assumed that Completion takes the value 0, while if there is at least one non-NA value, Completion takes the value of 1.

Data cleaning

After data transformation, a harmonized dataset is created; however, it may contain instances where all columns have NA values, requiring careful examination and appropriate handling to ensure data integrity and consistency. After applying data cleaning, a dataset with 698 patients is obtained.

Data availability

The source code for the data harmonization steps outlined in this study is available from a public github repository at github.com/pjcopadoCS/permepsy_harmonization.

References

1. Goetz, L. H. & Schork, N. J. Personalized medicine: Motivation, challenges, and progress. *Fertil. Steril.* **109**, 952–963 (2018).
2. Zhang, S., Bamakan, S. M. H., Qu, Q. & Li, S. Learning for personalized medicine: A comprehensive review from a deep learning perspective. *IEEE Rev. Biomed. Eng.* **12**, 194–208 (2018).
3. Egger, J. et al. Medical deep learning-A systematic meta-review. *Comput. Methods Programs Biomed.* **221**, 106874 (2022).
4. Fusar-Poli, P. & Schultze-Lutter, F. Predicting the onset of psychosis in patients at clinical high risk: Practical guide to probabilistic prognostic reasoning. *BMJ Ment. Health* **19**, 10–15 (2016).
5. Dwyer, D. B., Falkai, P. & Koutsouleris, N. Machine learning approaches for clinical psychology and psychiatry. *Annu. Rev. Clin. Psychol.* **14**, 91–118 (2018).
6. Stumpp, N. E. & Sauer-Zavala, S. Evidence-based strategies for treatment personalization: A review. *Cognit. Behav. Pract.* **29**, 902–913 (2022).
7. Fisher, A. J. Toward a dynamic model of psychological assessment: Implications for personalized care. *J. Consult. Clin. Psychol.* **83**, 825 (2015).
8. Arias, D., Saxena, S. & Verguet, S. Quantifying the global burden of mental disorders and their economic value. *EClinicalMedicine* **54** (2022).
9. Mancuso, A. et al. Economic burden of schizophrenia: the European situation. A scientific literature review. *Eur. J. Public Health* **24**, cku166–129. https://doi.org/10.1093/eurpub/cku166.129 (2014).
10. Pares-Badell, O. et al. Cost of disorders of the brain in Spain. *PLoS ONE* **9**, e105471. https://doi.org/10.1371/journal.pone.0105471 (2014).
11. Morrison, A. P. et al. Cognitive therapy for people with schizophrenia spectrum disorders not taking antipsychotic drugs: A single-blind randomised controlled trial. *Lancet* **383**, 1395–1403. https://doi.org/10.1016/S0140-6736(13)62246-1 (2014).
12. Moritz, S., Woodward, T. S. & Balzan, R. Is metacognitive training for psychosis effective?. *Exp. Rev. Neurother.* **16**, 105–107. https://doi.org/10.1586/14737175.2016.1135737 (2016).
13. Moritz, S., Menon, M., Balzan, R. & Woodward, T. S. Metacognitive training for psychosis (MCT): Past, present, and future. *Eur. Arch. Psychiatry Clin. Neurosci.* **273**, 811–817 (2023).
14. Metacognitive Training (MCT) for Psychosis-Clinical Neuropsychology Unit. https://clinical-neuropsychology.de/metacognitive_training-psychosis.
15. Penney, D. et al. Immediate and sustained outcomes and moderators associated with metacognitive training for psychosis a systematic review and meta-analysis supplemental content. *JAMA Psychiatry* **79**, 417–429. https://doi.org/10.1001/jamapsychiatry.2022.0277 (2022).
16. Moritz, S., Menon, M., Andersen, D., Woodward, T. S. & Gallinat, J. Moderators of symptomatic outcome in metacognitive training for psychosis (MCT). Who benefits and who does not? *Cognit. Ther. Res.* **42**, 80–91 (2018).
17. Leanza, L., Studerus, E., Bozikas, V. P., Moritz, S. & Andreou, C. Moderators of treatment efficacy in individualized metacognitive training for psychosis (MCT+). *J. Behav. Ther. Exp. Psychiatry* **68**, 101547 (2020).
18. González-Blanch, C. et al. Moderators of cognitive insight outcome in metacognitive training for first-episode psychosis. *J. Psychiatric Res.* **141**, 104–110 (2021).
19. Ferrer-Quintero, M. et al. Heterogeneity in response to MCT and psychoeducation: A feasibility study using 'latent class mixed models in first-episode psychosis. *Healthcare* **10**. https://doi.org/10.3390/healthcare10112155 (2022).
20. Eichner, C. & Berna, F. Acceptance and efficacy of metacognitive training (MCT) on positive symptoms and delusions in patients with schizophrenia: a meta-analysis taking into account important moderators. *Schizophr. Bull.* **42**, 952–962 (2016).
21. Jiang, J., Zhang, L., Zhipei, Z., Wei, L. & Chunbo, L. Metacognitive training for schizophrenia: A systematic review. *Shanghai Arch. Psychiatry* **27**, 149 (2015).
22. Liu, Y.-C., Tang, C.-C., Hung, T.-T., Tsai, P.-C. & Lin, M.-F. The efficacy of metacognitive training for delusions in patients with schizophrenia: A meta-analysis of randomized controlled trials informs evidence-based practice. *Worldviews Evid. Based Nurs.* **15**, 130–139 (2018).
23. Philipp, R. et al. Effectiveness of metacognitive interventions for mental disorders in adults-a systematic review and meta-analysis (METACOG). *Clin. Psychol. Psychother.* **26**, 227–240 (2019).
24. Lopez-Moriningo, J.-D. et al. Can metacognitive interventions improve insight in schizophrenia spectrum disorders? A systematic review and meta-analysis. *Psychol. Med.* **50**, 2289–2301 (2020).
25. Kraus, J. M. et al. Big data and precision medicine: Challenges and strategies with healthcare data. *Int. J. Data Sci. Anal.* **6**, 241–249. https://doi.org/10.1007/S41060-018-0095-0/METRICS (2018).
26. Kourou, K. D. et al. Cohort harmonization and integrative analysis from a biomedical engineering perspective. *IEEE Rev. Biomed.* **12**, 303–318. https://doi.org/10.1109/RBME.2018.2855055 (2019).
27. Rolland, B. et al. Toward rigorous data harmonization in cancer epidemiology research: One approach. *Am. J. Epidemiol.* **182**, 1033–1038. https://doi.org/10.1093/aje/kwv133 (2015).
28. Fortier, I. et al. Maelstrom research guidelines for rigorous retrospective data harmonization. *Int. J. Epidemiol.* **46**, 103–105. https://doi.org/10.1093/ije/dyw075 (2017).
29. Hutchinson, D. M. et al. How can data harmonisation benefit mental health research? an example of the cannabis cohorts research consortium. *Aust. N. Z. J. Psychiatry* **49**, 317–323. https://doi.org/10.1177/0004867415571 (2015).
30. Doiron, D. et al. Data harmonization and federated analysis of population-based studies: The BioSHaRE project. *Emerg. Themes Epidemiol.* **10**, 1–12. https://doi.org/10.1186/1742-7622-10-12 (2013).
31. Wilson, R. C. et al. Datashield-New directions and dimensions. *Data Sci. J.* **16**, 1–21. https://doi.org/10.5334/dsj-2017-021 (2017).
32. Kay, S. R., Fiszbein, A. & Opler, L. A. The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizoph. Bull.* **13**, 261–276. https://doi.org/10.1093/schbul/13.2.261 (1987).
33. Beck, A. T., Baruch, E., Balter, J. M., Steer, R. A. & Warman, D. M. A new instrument for measuring insight: The Beck cognitive insight scale. *Schizophr. Res.* **68**, 319–329. https://doi.org/10.1016/S0920-9964(03)00189-0 (2004).
34. Luborsky, L. Clinicians' judgments of mental health: A proposed scale. *Arch. Gen. Psychiatry* **7**, 407–417. https://doi.org/10.1001/archpsyc.1962.01720060019002 (1962).
35. Kinderman, P. & Bentall, R. P. A new measure of causal locus: The internal, personal and situational attributions questionnaire. *Pers. Individ. Differ.* **20**, 261–264. https://doi.org/10.1016/0191-8869(95)00186-7 (1996).
36. Rosenberg, M. Rosenberg Self-Esteem Scale (RSE): Acceptance and Commitment Therapy. Measures Package. Vol. 61. (Society and the Adolescent Self-Image, 1965).
37. Haddock, G., McCarron, J., Tarrier, N. & Faragher, E. Scales to measure dimensions of hallucinations and delusions: The psychotic symptom rating scales (PSYRATS). *Psychol. Med.* **29**, 879–889. https://doi.org/10.1017/s0033291799008661 (1999).
38. Peters, E. R. et al. Cognitive biases questionnaire for psychosis. *Schizophr. Bull.* **40**, 300–313. https://doi.org/10.1093/schbul/sbs199 (2014).
39. Speechley, W. J., Whitman, J. C. & Woodward, T. S. The contribution of hypersalience to the "jumping to conclusions" bias associated with delusions in schizophrenia. *J. Psychiatry Neurosci.* **35**, 7–17. https://doi.org/10.1503/jpn.090025 (2010).

40. Reitan, R. M. & Wolfson, D. Category test and trail making test as measures of frontal lobe functions. *Clin. Neuropsychol.* **9**, 50–56. https://doi.org/10.1080/13854049508402057 (1995).
41. Peters, E. R., Joseph, S. A. & Garety, P. A. Measurement of delusional ideation in the normal population: introducing the PDI (Peters et al. delusions inventory). *Schizophr. Bull.* **25**, 553–576. https://doi.org/10.1093/oxfordjournals.schbul.a033401 (1999).
42. Beck, A. T., Ward, C. H., Mendelson, M., Mock, J. & Erbaugh, J. An inventory for measuring depression. *Arch. Gen. Psychiatry* **4**, 561–571. https://doi.org/10.1001/archpsyc.1961.01710120031004 (1961).
43. Cramer, H. *Mathematical Methods of Statistics* (Princeton University Press, 1946).
44. NINDS CDE Notice of Beck Depression Inventory-II (BDI-II). http://commondataelements.ninds.nih.gov.
45. Lovibond, S., Lovibond, P. & Psychology Foundation of Australia. *Manual for the Depression Anxiety Stress Scales* (Psychology Foundation of Australia, 1995).
46. Addington, D., Addington, J. & Patten, S. Depression in people with first-episode schizophrenia. *Br. J. Psychiatry* **172**, 90–92. https://doi.org/10.1192/S0007125000297729 (1998).
47. Zeb, A., Soininen, J.-P. & Sozer, N. Data harmonisation as a key to enable digitalisation of the food sector: A review. *Food Bioprod. Process.* **127**, 360–370. https://doi.org/10.1016/j.fbp.2021.02.005 (2021).
48. Cheng, C. et al. A general primer for data harmonization. *Sci. Data* **11**, 152. https://doi.org/10.1038/s41597-024-02956-3 (2024).
49. Stuckenschmidt, H. *Ontology-Based Information Sharing in Weakly Structured Environments*. Ph.D. Thesis, Vrije Universiteit Amsterdam (2003).
50. Balzan, R. P., Mattiske, J. K., Delfabbro, P., Liu, D. & Galletly, C. Individualized metacognitive training (MCT+) reduces delusional symptoms in psychosis: A randomized clinical trial. *Schizophr. Bull.* **45**, 27–36. https://doi.org/10.1093/schbul/sby152 (2019).
51. Favrod, J., Maire, A., Bardy, S., Pernier, S. & Bonsack, C. Improving insight into delusions: A pilot study of metacognitive training for patients with schizophrenia. *J. Adv. Nurs.* **67**, 401–407. https://doi.org/10.1111/J.1365-2648.2010.05470.X (2011).
52. Fujii, K., Kobayashi, M., Funasaka, K., Kurokawa, S. & Hamagami, K. Effectiveness of metacognitive training for long-term hospitalized patients with schizophrenia: A pilot study with a crossover design. *Asian J. Occup. Ther.* **17**, 45–52. https://doi.org/10.11596/ASIAJOT.17.45 (2021).
53. Ishikawa, R. et al. The efficacy of extended metacognitive training for psychosis: A randomized controlled trial. *Schizophr. Res.* **215**, 399–407. https://doi.org/10.1016/J.SCHRES.2019.08.006 (2020).
54. Kuokkanen, R., Lappalainen, R., Repo-Tiihonen, E. & Tiihonen, J. Metacognitive group training for forensic and dangerous non-forensic patients with schizophrenia: A randomised controlled feasibility trial. *Crim. Behav. Ment. Health* **24**, 345–357. https://doi.org/10.1002/CBM.1905 (2014).
55. Leanza, L., Studerus, E., Bozikas, V. P., Moritz, S. & Andreou, C. Moderators of treatment efficacy in individualized metacognitive training for psychosis (MCT+). *J. Behav. Ther. Exp. Psychiatry* **68**, 101547. https://doi.org/10.1016/J.JBTEP.2020.101547 (2020).
56. Lopez-Morinigo, J. D. et al. A pilot 1-year follow-up randomised controlled trial comparing metacognitive training to psychoeducation in schizophrenia: Effects on insight. *Schizophrenia* **9**, 1–13. https://doi.org/10.1038/s41537-022-00316-x (2023).
57. Moritz, S., Veckenstedt, R., Randjbar, S., Vitzthum, F. & Woodward, T. S. Antipsychotic treatment beyond antipsychotics: Metacognitive intervention for schizophrenia patients improves delusional symptoms. *Psychol. Med.* **41**, 1823–1832. https://doi.org/10.1017/S0033291710002618 (2011).
58. Moritz, S. et al. Complementary group metacognitive training (MCT) reduces delusional ideation in schizophrenia. *Schizophr. Res.* **151**, 61–69. https://doi.org/10.1016/J.SCHRES.2013.10.007 (2013).
59. Ochoa, S. et al. Randomized control trial to assess the efficacy of metacognitive training compared with a psycho-educational group in people with a recent-onset psychosis. *Psychol. Med.* **47**, 1573–1584. https://doi.org/10.1017/S0033291716003421 (2017).
60. Ochoa, S. et al. S34. Effectiveness of individual metacognitive training (MCT+) in first-episode psychosis. *Schizophr. Bull.* **46**, S44 (2020).
61. de Pinho, L. M. G. et al. Assessing the efficacy and feasibility of providing metacognitive training for patients with schizophrenia by mental health nurses: A randomized controlled trial. *J. Adv. Nurs.* **77**, 999–1012. https://doi.org/10.1111/JAN.14627 (2021).
62. Simón-Expósito, M. & Felipe-Castaño, E. Effects of metacognitive training on cognitive insight in a sample of patients with schizophrenia. *Int. J. Environ. Res. Public Health* **16**, 4541. https://doi.org/10.3390/IJERPH16224541 (2019).
63. So, S. H. W. et al. Metacognitive training for delusions (MCTd): Effectiveness on data-gathering and belief flexibility in a Chinese sample. *Front. Psychol.* **6**, 143010. https://doi.org/10.3389/fpsyg.2015.00730 (2015).
64. Swanson, L. et al. Metacognitive training modified for negative symptoms: A feasibility study. *Clin. Psychol. Psychother.* **29**, 1068–1079. https://doi.org/10.1002/CPP.2692 (2022).
65. Ussorio, D. et al. Metacognitive training for young subjects (MCT young version) in the early stages of psychosis: Is the duration of untreated psychosis a limiting factor?. *Psychol. Psychother. Theory Res. Pract.* **89**, 50–65. https://doi.org/10.1111/papt.12059 (2016).
66. Yildiz, M. et al. The effect of psychosocial skills training and metacognitive training on social and cognitive functioning in schizophrenia. *Noropsikiyatri Ars.* **56**, 139–143. https://doi.org/10.29399/NPA.23095 (2018).
67. Verhage, M. L. et al. Conceptual comparison of constructs as first step in data harmonization: Parental sensitivity, child temperament, and social support as illustrations. *MethodsX* **9**, 101889. https://doi.org/10.1016/j.mex.2022.101889 (2022).
68. Chondrogiannis, E., Karanastasis, E., Andronikou, V. & Varvarigou, T. Bridging the gap among cohort data using mapping scenarios. *J. Adv. Inf. Technol.* **12**. https://doi.org/10.12720/jait.12.3.179-188 (2021).
69. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders: DSM-5* Vol. 5 (American Psychiatric Association, 2013).
70. Ritsner, M., Kurs, R., Gibel, A., Ratner, Y. & Endicott, J. Validity of an abbreviated quality of life enjoyment and satisfaction questionnaire (Q-LES-Q-18) for schizophrenia, schizoaffective, and mood disorder patients. *Qual. Life Res.* **14**, 1693–1703. https://doi.org/10.1007/s11136-005-2816-9 (2005).
71. World Health Organization. *The World Health Organization Quality of Life (WHOQOL) - BREF. Technical Documents* (2004).
72. Carlson, J. et al. Adaptation and validation of the quality-of-life scale: Satisfaction with life domains scale by baker and intagliata. *Comprehens. Psychiatry* **50**, 76–80. https://doi.org/10.1016/j.comppsych.2008.05.008 (2009).
73. Heinrichs, D. W., Hanlon, T. E. & Carpenter, J., William T. The quality of life scale: An instrument for rating the schizophrenic deficit syndrome. *Schizophr. Bull.* **10**, 388–398. https://doi.org/10.1093/schbul/10.3.388 (1984). https://academic.oup.com/schizophreniabulletin/article-pdf/10/3/388/5229121/10-3-388.pdf.
74. Ausín, B. & Muñoz, M. *Guía Práctica de Detección de Problemas de Salud Mental. Anexo. Instrumentos de Detección: Escala de Satisfacción con las Dimensiones de la Vida (SLDS)* (Ed. Pirámide, 2018).
75. Ritsner, M., Kurs, R., Gibel, A., Ratner, Y. & Endicott, J. Validity of an abbreviated quality of life enjoyment and satisfaction questionnaire (q-les-q-18) for schizophrenia, schizoaffective, and mood disorder patients. *Qual. Life Res.* **14**, 1693–1703. https://doi.org/10.1007/s11136-005-2816-9 (2005).
76. Heinrichs, D. W., Hanlon, T. E. & Carpenter, W. T. The quality of life scale: An instrument for rating the schizophrenic deficit syndrome. *Schizophr. Bull.* **10**, 388–398. https://doi.org/10.1093/SCHBUL/10.3.388 (1984).
77. Henry, J. D. & Crawford, J. R. The short-form version of the depression anxiety stress scales (DASS-21): Construct validity and normative data in a large non-clinical sample. *Br. J. Clin. Psychol.* **44**, 227–239. https://doi.org/10.1348/014466505X29657 (2005).
78. Addington, D., Addington, J. & Maticka-Tyndale, E. Assessing depression in schizophrenia: The Calgary depression scale. *Br. J. Psychiatry* **163**, 39–44 (1993).

Declaration

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-94815-3.

Correspondence and requests for materials should be addressed to C.K.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.