

46 **Keywords:** psychiatry, mental health, biobank, real-world data, broad consent,
47 representativeness, electronic health records, medical informatics initiative

48

49 **Funding**

50 **C.F.W., A.N., and N.S.** were supported by the DFG (German Research Council) (grant
51 numbers: FA 876/3-1; 876/5-1).

52 **U.H.** was supported by European Union's Horizon 2020 Research and Innovation
53 Programme (PSY-PGx, grant agreement No 945151).

54 **D.L.** was supported by the Deutsche Forschungsgemeinschaft, Emmy Noether fellowship
55 (LA4126/1-1)

56 **E.C.S.** was supported by the Munich Clinician Scientist Program.

57

58

59 **Conflicts of Interest**

60 **R.M.** has received speakers honoraria from Otsuka Pharmaceuticals. He serves or has served
61 as PI in clinical trials sponsored by Abide Therapeutics (Lundbeck), Böhringer-Ingelheim,
62 Emalex Biosciences, Oryzon, Otsuka Pharmaceuticals.

63 **F.P.** is a member of the European Scientific Advisory Board of Brainsway Inc., Jerusalem,
64 Israel, and the International Scientific Advisory Board of Sooma, Helsinki, Finland. He has
65 received speaker's honoraria from Mag&More GmbH, the neuroCare Group, Munich,
66 Germany, and Brainsway Inc. His lab has received support with equipment from neuroConn
67 GmbH, Ilmenau, Germany, Mag&More GmbH and Brainsway Inc.

68

69

70 **Author contributions**

71 Concept and design: Kalman, Schulte, Wichert, Zill, Schulze, Falkai

72 Analysis and interpretation of data: Kalman, Burkhardt, Schulte

73 Drafting of the manuscript: Kalman, Schulte, Burkhardt, Zill, Wichert

74 All other authors provided data, contributed ideas and suggestions for analyses, interpreted
75 results and revised the final manuscript.

76

77 **Abstract**

78 Translational research on complex, multifactorial mental health disorders, such as bipolar
79 disorder, major depressive disorder, schizophrenia, and substance use disorders requires
80 databases with large-scale, harmonized, and integrated real-world and research data.

81 The Munich Mental Health Biobank (MMHB) is a mental health-specific biobank that was
82 established in 2019 to collect, store, connect, and supply such high-quality phenotypic data
83 and biospecimens from patients and study participants, including healthy controls, recruited
84 at the Department of Psychiatry and Psychotherapy and the Institute of Psychiatric
85 Phenomics and Genomics, University Hospital of the Ludwig-Maximilians-University (LMU),
86 Munich, Germany. Participants are asked to complete a questionnaire that assesses
87 sociodemographic and cross-diagnostic clinical information, provide blood samples, and grant
88 access to their existing medical records. The generated data and biosamples are available to
89 both academic and industry researchers. In the current manuscript, we outline the workflow
90 and infrastructure of the MMHB, describe the clinical characteristics and representativeness
91 of the sample collected so far, and reveal future plans for expansion and application.

92 As of October 31, 2021, the MMHB contains a continuously growing set of data from 578
93 patients and 104 healthy controls (46.37% female; median age, 38.31 years). The five most
94 common mental health diagnoses in the MMHB are recurrent depressive disorder (38.78%;
95 ICD-10: F33), alcohol-related disorders (19.88%; ICD-10: F10), schizophrenia (19.69%; ICD-10:
96 F20), depressive episode (15.94%; ICD-10: F32), and personality disorders (13.78%; ICD-10:
97 F60). Compared with the average patient treated at the recruiting hospitals, MMHB
98 participants have significantly more mental health-related contacts, less severe symptoms,
99 and a higher level of functioning. The distribution of diagnoses is also markedly different in
100 MMHB participants compared with individuals who did not participate in the biobank.

101 After establishing the necessary infrastructure and initiating recruitment, the major tasks for
102 the next phase of the MMHB project are to improve the pace of participant enrollment,
103 diversify the sociodemographic and diagnostic characteristics of the sample, and improve the
104 utilization of real-world data generated in routine clinical practice.

105

106 **1. Background**

107 In recent years, technological and methodological advances have enabled diseases to be
108 studied at an unprecedented depth and expanse. Biobanks, which collect, store, manage, and
109 share large sets of biosamples and deep phenotyping information, are critical infrastructures
110 for the clinical translation of these advances into patient care.

111 Establishing such rich datasets is especially important for mental health disorders because
112 current diagnostic categories have low prognostic value, disease biomarkers are largely
113 lacking, and novel therapeutic approaches have often failed to reduce the global burden of
114 these disorders (1,2). Nevertheless, mental health disorders are underrepresented in
115 population-based biobanking efforts, and even the largest and most successful biobanks, such
116 as the UK Biobank and the Estonian Biobank, began to collect extensive mental health-related
117 phenotypes only years after their initiation (3,4). The reasons for this delay may include the
118 significant resources and commitment required for mental health phenotyping and the
119 difficulties in utilizing real-world data for mental health research because of the lack of data
120 structure, standardization, and interoperability (5). Social stigma attached to many mental
121 disorders and the special legal implications, such as legal guardianship, may further reduce
122 the willingness and ability of patients to participate in biobanking efforts.

123 Similar trends are seen in Germany: Currently, only a few, large mental health facilities
124 routinely collect phenotype and biological data, and most of these efforts focus on individuals
125 participating in clinical studies and do not include patients in routine clinical practice (6–8).
126 Therefore, to facilitate biobanking in mental health, we established the Munich Mental Health
127 Biobank (MMHB), which intimately entwines the collection of both routine clinical and
128 research phenotyping data and biological specimens at a large tertiary care center in Munich,
129 Germany, comprising the Department of Psychiatry and Psychotherapy (DPP) and the
130 Institute of Psychiatric Phenomics and Genomics (IPPG). In the current manuscript, we
131 describe the workflow and information technology infrastructure of the MMHB, provide a
132 first glimpse into the clinical characteristics of the sample collected so far, analyze the
133 representativeness of the sample with regard to a variety of key clinical and
134 sociodemographic characteristics, and reveal future plans for expansion and application.

135

136 **2. The MMHB**

137 The DPP and IPPG are part of the hospital system of the University Hospital of the Ludwig-
138 Maximilians-University (LMU) Munich, Germany. With 3222 unique cases (inpatient: 1058;
139 day hospital: 130; and outpatient: 2034) annually (2020 data), these two institutes are among
140 the largest university-based mental health facilities in Germany. Both institutes are
141 internationally renowned strongholds of mental health research and are currently hosting
142 over 43 ongoing clinical studies. However, until recently, neither the DPP nor the IPPG had
143 access to an organized, standardized, high-throughput hospital-wide collection of phenotypes
144 and biosamples, markedly limiting large-scale translational research. This changed in 2019
145 with the establishment of the MMHB. Since then, all individuals treated at the DPP and IPPG,
146 as well as study participants (both patients with a mental health diagnosis and healthy
147 controls with no self-reported lifetime mental health diagnosis participating in clinical or non-
148 clinical trials), are eligible for inclusion in the MMHB and can be invited to provide informed
149 consent to the collection of phenotypic information and biosamples as part of the biobanking
150 effort.

151 2.1. Informed Consent

152 For the following activities, the MMHB aims to obtain modular informed consent from
153 participants that conforms with the General Data Protection Regulation of the European
154 Union: 1) the collection, storage, analysis, scientific utilization, and distribution of deep
155 phenotypic data and biosamples in a double-pseudonymized form to academic and non-
156 academic research institutions, national and international data archives, and companies, and
157 the linkage of this data to routine clinical data; 2) re-contacting participants to request
158 additional data and biosamples and/or to invite them to participate in new studies; 3) re-
159 contacting participants to seek consent for linking the collected data to other databases; and
160 4) re-contacting participants to inform them about actionable incidental findings. By providing
161 their consent, MMHB participants allow researchers to use their phenotypic data and
162 biosamples for any type of research, including omics studies and the creation of induced
163 pluripotent stem cell (iPSC) lines, for 30 years. MMHB participants can withdraw their consent
164 at any time without providing a specific reason. In case of withdrawal of consent, they can
165 decide whether their data and biosamples should be deleted and destroyed or anonymized.

166 2.2. The MMHB phenotyping battery

167 The MMHB phenotyping battery consists of three components: 1) a **basic phenotyping**
168 **module**, which builds on previous national multi-centric phenotyping efforts, such as the
169 PsyCourse Study and the PD-CAN study; 2) a module consisting of seven **standardized self-**
170 **rating measurements** with cross-disorder utility; and 3) a **variable module** that highlights a
171 specific topic of research in more detail and changes every two years (*Table 1*) (9). In 2020
172 and 2021, the variable module focused on metabolic risk, eating behavior, and stress related
173 to the COVID-19 pandemic. To reduce the biobanking-associated clinician workload, self-
174 rather than observer-rating tools are used to obtain patient characteristics and symptom
175 scores. The MMHB phenotyping battery is complemented by data collected in association
176 with specific scientific projects or clinical trials.

177

178 **Table 1:** Overview of the phenotypes assessed with the Munich Mental Health Biobank
 179 (MMHB) phenotyping battery

Module	Questionnaire	Reference
Basic phenotyping	Sociodemographic information	
	Psychiatric medical history	
	Family history	
	Pharmacologic treatment	
	Self-destructive behavior & suicidality	
	Substance abuse, including the Alcohol Use Disorders Identification Test (AUDIT-C)	Bush et al. 1998 (10)
Standardized self-rating instruments	Childhood Trauma Screener	Grabe et al. 2021 (11)
	Brief Resilience Scale	Smith et al. 2008 (12)
	UCLA Loneliness Scale (3-item version)	Hughes et al. 2004 (13)
	Lubben Social Network Scale	Lubben et al. 2006 (14)
	WHO-5 Well-Being Index	Topp et al. 2015 (15)
	Patient Health Questionnaire (PHQ-9)	Kroenke et al. 2001 (16)
	Munich Chronotype Questionnaire	Roenneberg et al. 2003 (17)
Variable module (2020/2021)	Questions on metabolic risk	Adapted from Barton et al. 2020 (18)
	Questions on eating behavior	Adapted from Cappelleri et al. 2009 (19)
	COVID-19 Pandemic Mental Health Questionnaire (CoPaQ)	Rek et al. 2021 (20)

180

181 2.3. Biosamples

182 A limited amount of blood (maximum 70 mL) and, if available, cerebrospinal fluid (maximum
 183 10 mL) is collected from consented individuals during routine clinical examinations or study
 184 visits (*Supplementary Table S2*). Participants may also be asked to provide saliva, hair, and
 185 stool samples.

186 2.4. Linked routine clinical data

187 Routine clinical data stored in electronic health records or other databases can be linked with
188 the phenotypic information and biosamples collected from MMHB participants and then used
189 for research purposes.

190 2.5. Data and sample processing

191 The collected phenotypic data and biosamples are processed by specially trained staff
192 according to standard operating procedures for data collection, handling, and storage. For
193 each MMHB participant, an electronic record, which includes general and disease-specific
194 clinical information, dates of phenotype and biosample acquisitions, and information on
195 sample processing (i.e., identification code, quantity, location of storage, and type of
196 preparation) is captured in CentraXX (Kairos GmbH, Bochum, Germany), a standardized
197 laboratory information management system. CentraXX is fully integrated into the laboratory
198 automation infrastructure and, besides the administration of biosamples, enables the
199 acquisition and management of all clinical and phenotypic data of participants. Biosamples
200 (DNA, RNA, serum, plasma, etc.) are processed on site with automated systems
201 (ChemagicStar, Decapper/Barcode Reader, and a StarPlus System; Hamilton Robotics GmbH,
202 Martinsried, Germany) and are stored at -80°C in an automated sample storage and
203 management system (BIOS M System, Hamilton Storage GmbH, Bonaduz, Switzerland) (*Figure*
204 *1*).

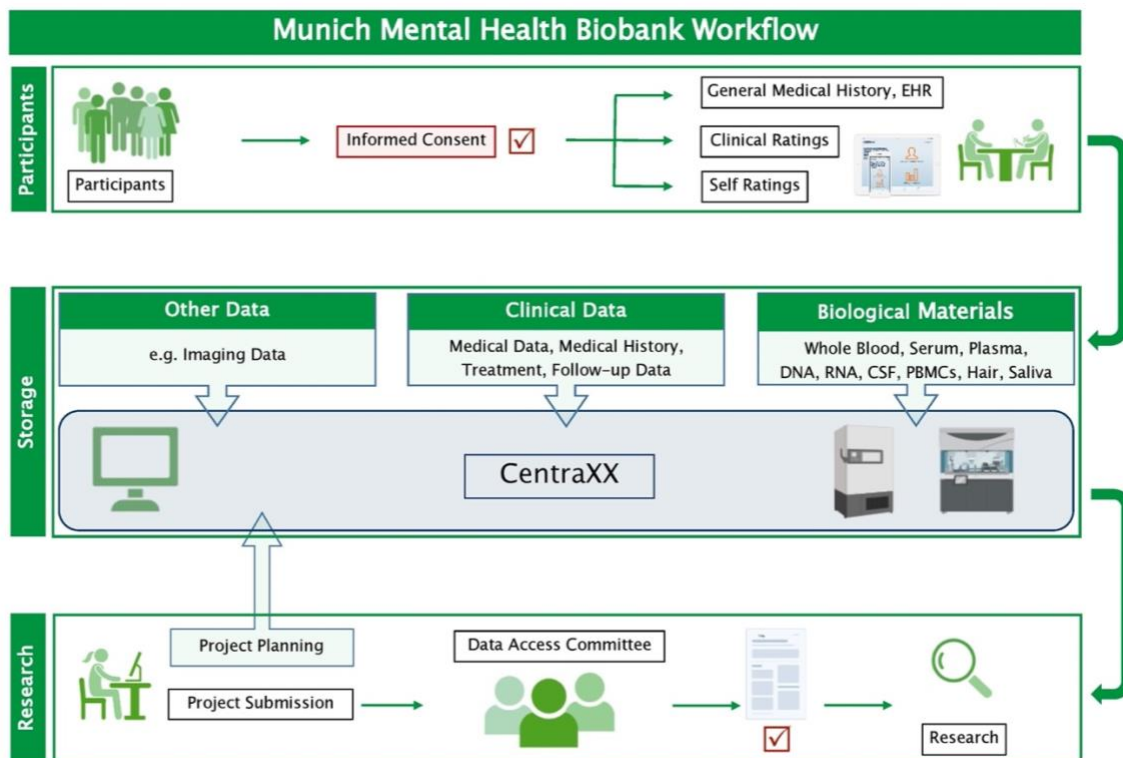
205

206 **Figure 1: Workflow of the Munich Mental Health Biobank**

207 Clinical data and biological materials are stored in pseudonymized form in the information
208 management system CentraXX (Kairos GmbH, Bochum, Germany). Biological materials are
209 stored at -80°C, except for peripheral blood mononuclear cells, which are stored in liquid
210 nitrogen (*Supplementary Table S2*). Researchers can request data access by completing a
211 request form. If the Data Access Committee approves the request, a material and/or data
212 transfer agreement is signed, and the data and biological materials can be provided.

213 **Abbreviations:** CSF, cerebrospinal fluid; EHR, electronic health record; PBMCs, peripheral
214 blood mononuclear cells

215



216

217 2.6. Data availability

218 The collected phenotypes and biomaterials are available to both academic and industry
219 researchers upon request. The process for requesting data and biomaterials is as follows: 1)
220 Requestors send an email with a description of the planned scientific project, the type and
221 quantity of the requested data, and their data protection policies to the MMHB
222 (psy_mmhb@med.uni-muenchen.de) to inquire about data availability, and in the same
223 email, they may also propose collaborations with local researchers; 2) the Data Access
224 Committee evaluates the feasibility of the project and provides a cost estimate for sample
225 extraction, processing, and shipment; 3) at this point, requestors provide documentation
226 confirming ethical approval of the proposed project; 4) the MMHB provides a quote, data-
227 and material transfer agreements (DTA and MTA, respectively); and 5) once the DTA and/or
228 MTA is signed by both parties, the biological materials and double-pseudonymized
229 phenotypic and omics data are sent to the requestor. Depending on the legal process,
230 approximately four to six months should be allowed from submission of the project proposal
231 to delivery of the data and/or biomaterials.

232 3. Sample characteristics

233 Detailed phenotype definitions of the MMHB Sample and Clinical Sample (i.e., the target
234 population of all patients treated at the DPP and IPPG during the same period who were
235 eligible for inclusion in the MMHB but were not included) are provided in the description of
236 the individual analyses and in *Supplementary Table S1*. Qualitative traits were analyzed with
237 non-parametric pairwise Mann-Whitney U tests, and quantitative traits, with χ^2 tests. The
238 significance threshold was set at 0.05. The study protocol was approved by the local ethics
239 committee, and the study was performed in accordance with the Declaration of Helsinki. All
240 participants in the MMHB Sample provided written informed consent. The Clinical Sample
241 served as naturalistic comparator: the data was generated in the clinical routine and
242 irreversibly anonymized for further analysis.

243 3.1. The MMHB Sample

244 Recruitment for the MMHB started on April 11, 2019. As of October 31, 2021, 578 patients
245 and 104 healthy controls had been included in the MMHB. *Table 2* and *Figure 2* provide an
246 overview of the demographic and clinical characteristics of the sample. Participants were
247 more likely to be recruited in an inpatient setting (67.47%) than in the day hospital (3.46%) or
248 outpatient clinic (29.07%). The analysis included all (independent) clinical (ICD-10) diagnoses
249 the patient received during the observational period (April 11, 2019, to October 31, 2021).
250 The most frequent ICD-10 diagnoses were recurrent depressive disorder (38.78%; ICD-10:
251 F33), alcohol-related disorders (19.88%; ICD-10: F10), schizophrenia (19.69%; ICD-10: F20),
252 depressive episode (15.94%; ICD-10: F32), and personality disorders (13.78%; ICD-10: F60). A
253 full list of the diagnoses and the respective frequencies is provided in *Supplementary Table*
254 *S3*.

255 3.2. Comparison of the MMHB Sample and the Clinical Sample

256 Next, we compared the basic sociodemographic and clinical characteristics of the MMHB
257 Sample with the Clinical Sample (i.e., the target population of all patients treated at the DPP
258 and IPPG during the same period who were eligible for inclusion in the MMHB but were not
259 included). The two samples had similar age and sex distributions (*Table 2*). Clinicians at the
260 DPP and IPPG routinely assess the seven-day functioning and disease severity of inpatients

261 and day hospital patients at admission and discharge by using the Global Assessment of
262 Functioning (GAF) and Clinical Global Impression (CGI) scales, respectively (21,22). According
263 to these measures, patients in both the MMHB and Clinical samples had severe symptoms
264 and major impairments in functioning at admission and moderate symptoms and improved
265 functioning at discharge. However, compared with the Clinical Sample, MMHB participants
266 who had been recruited as inpatients had significantly higher functioning at both admission
267 ($GAF-A_{MMHB} = 40$ vs. $GAF-A_{Clinical\ Sample} = 35$, Mann-Whitney-U-Test, $p = 8.10 \times 10^{-15}$) and
268 discharge ($GAF-D_{MMHB} = 55$ vs. $GAF-D_{Full\ Clinical\ Sample} = 55$, $p = 5.18 \times 10^{-3}$) and less severe
269 symptoms at admission ($CGI-A_{MMHB} = 5$ vs. $CGI-A_{Clinical\ Sample} = 5$, $p = 3.77 \times 10^{-5}$). Symptom
270 severity at discharge (CGI-D) was similar in both groups. No significant differences were
271 observed between the MMHB and Clinical samples of day hospital patients. GAF and CGI
272 values were not available for the outpatients. During the observation period, patients in the
273 MMHB Sample were hospitalized more frequently than those in the Clinical Sample (number
274 of hospitalizations: 2 vs. 1; Mann-Whitney U test, $p = 8.04 \times 10^{-41}$), for longer periods of time
275 (total length of stay, 71 vs. 37 days; $p = 2.94 \times 10^{-37}$), and had more outpatient contacts (6 vs.
276 1; $p = 4.35 \times 10^{-30}$) (Table 2).

277 Figure 2 provides an overview of the prevalence in the two samples of the 15 most frequent
278 diagnoses derived from the Clinical Sample. In the MMHB Sample vs the Clinical Sample, we
279 observed the most pronounced differences in the prevalence of recurrent depressive disorder
280 (38.78% vs 32.12%; ICD-10: F33), adjustment disorder (9.65% vs 15.17%; ICD-10: F43),
281 alcohol-related disorders (19.88% vs 11.44%; ICD-10: F10), schizophrenia (19.69% vs 9.64%;
282 ICD-10: F20), personality disorders (13.78% vs 7.72%; ICD-10: F60), and schizoaffective
283 disorder (8.27% vs 3.32%; ICD-10: F25). A full list of diagnoses and the respective frequencies
284 is provided in *Supplementary Table S3*.

285

286

287 **Table 2.** Sample characteristics

Variable	Group	MMHB Sample	Clinical Sample	p
N, (%) ^a	Healthy controls	104		
	Inpatients	390 (67.47%)	2274 (22.93%)	
	Day hospital patients	20 (3.46%)	383 (3.86%)	
	Outpatients	168 (29.07%)	7252 (73.20%)	
Age, median (MAD; range) ^b	Healthy controls	32.42 (11.76; 19.62-64.97)		
	Patients	38.31 (18.30; 18.08-93.07)	39.00 (19.27; 4-98)	0.39
Sex, % male	Healthy controls	53.85		
	Patients	53.63	52.07	0.50
GAF-A, median (MAD; range) ^{b,c}	Inpatients	40.0 (7.41; 7-65)	35 (7.41; 0-75)	8.10×10⁻¹⁵
	Day hospital patients	46 (6.67; 30-61)	45 (7.41; 25-65)	0.89
GAF-D, median (MAD; range) ^{b,c}	Inpatients	55 (7.41; 21-90)	55 (10.38; 0-95)	5.18×10⁻³
	Day hospital patients	60 (7.41; 31-70)	55 (7.41; 25-80)	0.06
CGI-A, median (MAD; range) ^{b,c}	Inpatients	5 (0; 3-7)	5 (0; 2-7)	3.77×10⁻⁵
	Day hospital patients	5 (1.48; 3-6)	4 (0; 3-6)	0.23
CGI-D, median (MAD; range) ^{b,c}	Inpatients	4 (0; 2-6)	4 (0; 1-7)	0.73
	Day hospital patients	4 (0; 3-6)	3 (0; 2-6)	0.36
Days hospitalized, median (MAD; range) ^{b,d}		71 (62.27; 1-502)	34 (34.10; 1-585)	2.94×10⁻³⁷
Number of hospitalizations, median (MAD; range) ^{b,d}		2 (1.48; 1-49)	1 (0; 1-62)	8.04×10⁻⁴¹

Number of outpatient visits, median (MAD; range) ^{b,d}		6 (7.41; 1-165)	1 (0; 1-279)	4.35×10⁻³⁰
---	--	-----------------	--------------	------------------------------

288 **Abbreviations:** CGI-A & CGI-D, Clinical Global Impression at admission & discharge, respectively; GAF-A & GAF-D, Global Assessment of
 289 Functioning at admission & discharge, respectively; MAD, median absolute deviation; MMHB, Munich Mental Health Biobank; p, unadjusted p
 290 value (significant p values are indicated in bold).

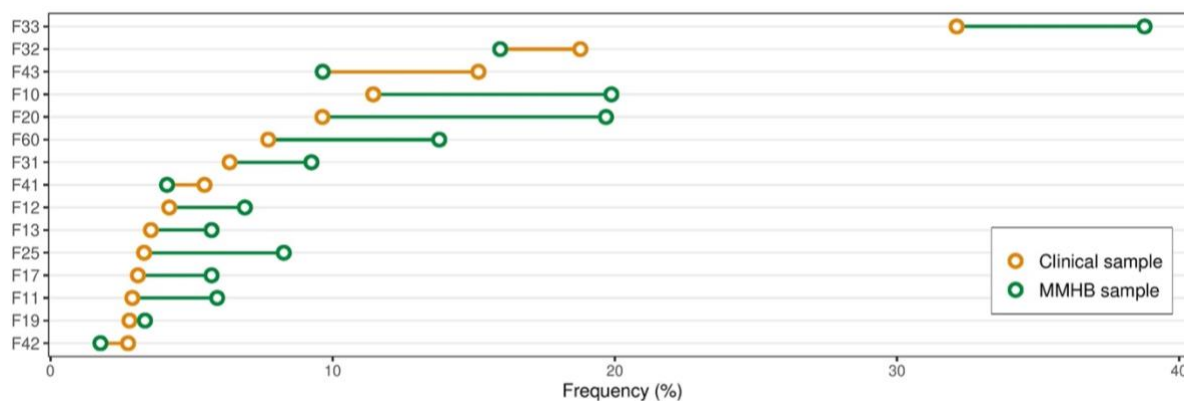
291 **Note:** See *Supplementary Material* for detailed phenotype descriptions. ^a Proportion of inpatients, day hospital patients, and outpatients among
 292 all patients (N = 578). ^b The MAD was calculated by using 1.4826 as a constant. ^c GAF and CGI values were available only for a subset of inpatients
 293 (N_{MMHB}=376, N_{Clinical Sample}= 2117) and day hospital (N_{MMHB}=11, N_{Clinical Sample}= 254) patients. ^d Days hospitalized (as either an inpatient or a day
 294 hospital patient) and the number of outpatient visits between April 11, 2019, and October 31, 2021, to the Department of Psychiatry and
 295 Psychotherapy or Institute of Psychiatric Phenomics and Genomics. Patients with no hospitalizations and/or outpatient contacts were excluded
 296 from the respective analyses.

297

298

299 **Figure 2:** Differences in diagnosis frequency between the Munich Mental Health Biobank
300 (MMHB) and Clinical samples

301 Comparison of the prevalence of the 15 most frequent diagnoses derived from the Clinical
302 Sample in the Munich Mental Health Biobank and the Clinical Sample. The diagnoses were
303 defined as all (independent) clinical (ICD-10) diagnoses the patient received during the
304 observational period (April 11, 2019, to October 31, 2021). A full list of diagnoses and their
305 respective frequencies is provided in *Supplementary Table S3*.



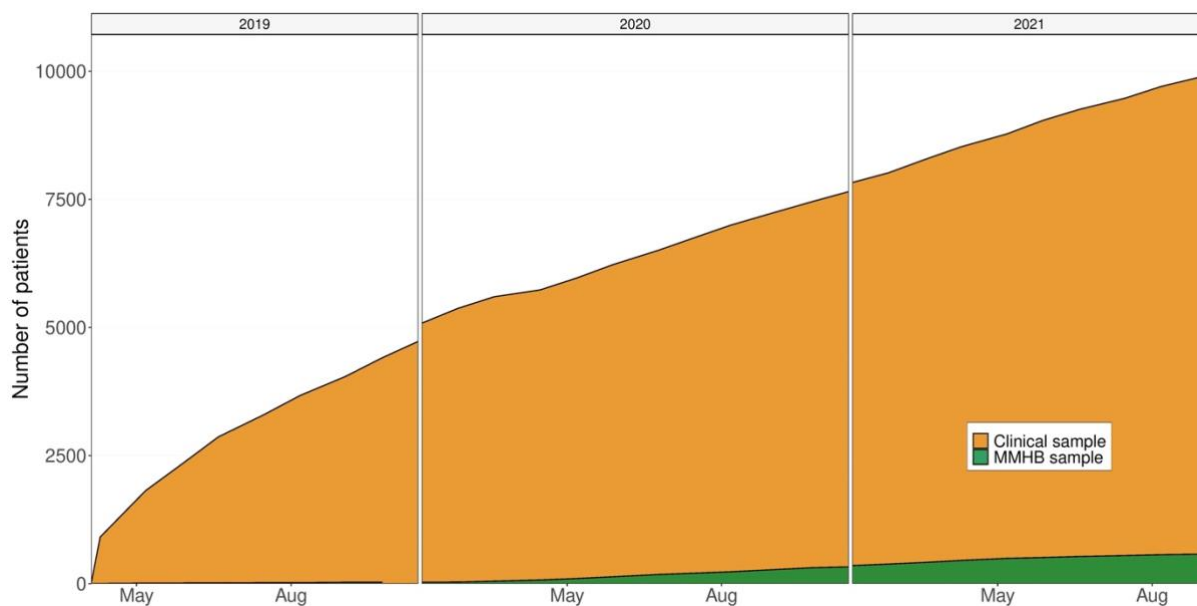
306

307 **Abbreviations:** F33, recurrent depressive disorder; F32, depressive episode; F43, adjustment
308 disorder; F10, alcohol-related disorders; F20, schizophrenia; F60, personality disorders; F31,
309 bipolar disorder; F41, other anxiety disorders; F12, mental and behavioural disorders due to
310 use of cannabinoids; F13, mental and behavioural disorders due to use of sedatives or
311 hypnotics; F25, schizoaffective disorder; F17, mental and behavioural disorders due to use of
312 tobacco; F11, mental and behavioural disorders due to use of opioids; F19, mental and
313 behavioural disorders due to multiple drug use and use of other psychoactive substances;
314 F42, obsessive-compulsive disorder; MMHB, Munich Mental Health Biobank

315

316

317 **Figure 3:** Growth of the Munich Mental Health Biobank (MMHB) and Clinical samples over
318 time



319

320 **Abbreviation:** MMHB, Munich Mental Health Biobank

321 4. Discussion

322 The MMHB is a mental health-specific biobank that aims to integrate multimodal research
323 and real-world phenotypic information with biosamples of patients with mental health
324 disorders and study participants, including healthy controls, to facilitate translatable mental
325 health research and, ultimately, improve patient care. Here, we outlined the information
326 technology and laboratory infrastructure and the administrative processes of the MMHB,
327 described the clinical and sociodemographic characteristics of the 578 patients and 104
328 healthy controls included in the MMHB so far and compared this sample with the overall
329 target population treated at the recruitment facilities. This initial evaluation allows us to
330 discuss the representativeness and scope of the MMHB sample, identify limitations of our
331 current recruitment strategy, and derive action points that could further improve the quality
332 of the MMHB data.

333 4.1. Representativeness

334 There is increasing evidence that individuals with certain non-random characteristics, such as
335 lower socioeconomic and education status, poorer health, non-European ancestry, and
336 increased cumulative genetic burden for schizophrenia, neuroticism, and attention-deficit
337 hyperactivity disorder, are less likely to participate in clinical studies and more likely to drop
338 out during the follow-up period (9,23–25). These data are concerning because non-
339 participation and attrition not only are associated with a loss of statistical power, but, if non-
340 random, can also influence sample representativeness and thereby bias the generalizability
341 and real-life utility of research findings, health policy decision-making, and ultimately, the
342 equity of health care provision (26). Compared with patients treated at the DPP and IPPG
343 since April 11, 2019 who were not included in the MMHB (Clinical Sample), we found that
344 certain mental health diagnoses, such as schizophrenia, schizoaffective disorder, alcohol-
345 related disorders, and personality disorders, are overrepresented in the MMHB sample,
346 whereas adjustment disorders, for example, are relatively underrepresented, most likely
347 because recruitment so far has been propagated by study teams interested in answering
348 specific research questions. Interestingly, we found evidence that the frequency and length
349 of contact with the health care system differed between the target population (i.e., those not
350 included in the biobank) and the MMHB sample: MMHB participants were not only more
351 frequently hospitalized (median number of hospitalizations in MMHB Sample vs Clinical
352 Sample, 2 vs 1), but they also spent more days in hospital (median, 71 vs 37 days) and had
353 more outpatient appointments (6 vs 1) during the 30-month observational period. The
354 proportion of participants recruited in an inpatient setting was also higher in the MMHB
355 Sample than in the Clinical Sample (67.47% vs 22.93%). These differences could be explained
356 by potential systematic sampling biases in that study personnel and clinicians are more likely
357 to ask “known” patients to consent to being included in the MMHB and/or that patients with
358 repeated contacts with the same institution are more likely to trust that institution and
359 participate in its research efforts. Also, longer hospitalizations and more frequent visits in
360 general represent more opportunities to recruit participants. These observations emphasize
361 the importance of interpersonal factors (such as trust, rapport, and compliance) and the type
362 and frequency of contact with the health care system in study participation. These factors are
363 potential sources of recruitment bias that should be taken into consideration when analyzing
364 the MMHB Sample. Interestingly, despite the more frequent health care utilization, the
365 MMHB Sample had less severe symptoms and higher levels of functioning (measured with
366 CGI and GAF, respectively) at inclusion in the biobank than the Clinical Sample in general.

367 Although this difference is likely not clinically meaningful (the difference in functioning was
368 within a 10-point GAF range), it still suggests that patients who were successfully recruited
369 had better overall social, occupational, and psychological functioning.

370 4.2. Availability of real-world clinical data

371 As outlined, the current MMHB phenotyping battery includes detailed sociodemographic and
372 self-reported transdiagnostic information that enables research across multiple disorders and
373 patient subgroups. However, like most other mental health-specific biobanks, currently the
374 MMHB includes only a limited set of features from routine clinical care (patient age, sex, zip
375 code, GAF, CGI, diagnosis, and laboratory results). This limited information is a direct result of
376 the limitations of current electronic health record systems, which are markedly restricted in
377 their ability to present relevant mental health information in a standardized, structured,
378 interoperable, and thus machine-readable format. Therefore, our current data mainly include
379 only cross-sectional information and lack real-world clinical data and longitudinal information
380 on disease trajectories. These data modalities, however, represent a main target of current
381 data-driven precision medicine initiatives, which aim at developing reliable, generalizable
382 prediction models of important mental health-related outcomes, such as antidepressant
383 response, suicide attempts, and the transition from clinical high-risk states to psychosis (27–
384 29).

385 4.3. Outlook

386 A key strength of the MMHB is its close integration into the clinical and scientific
387 infrastructure of a large mental health care center, which not only ensures the necessary
388 domain knowledge for collecting high-quality mental health-specific phenotypes, but also
389 provides direct access to the more than 3000 individual patients treated annually at the DPP
390 and IPPG and thus a strong growth potential (*Figure 3*). Moreover, this target population has
391 good sociodemographic representativeness because Germany has universal health insurance,
392 ensuring that treatment is offered to all individuals in need. A further asset of the MMHB is
393 the wide age range of eligible participants and the diversity of diagnoses, which enables the
394 study of all stages of the disease trajectory across the whole lifespan and diagnostic spectrum.

395 Limitations of the MMHB include the observed differences between the MMHB and Clinical
396 samples, the current sparsity of available real-world data, and the moderate recruitment pace
397 (only 5.56% of the 10,494 patients who were treated at the DPP and IPPG during the 30-
398 month observational period and were thus eligible for inclusion were recruited into the
399 MMHB, *Figure 3*).

400 To address these limitations, in the next phase of our biobanking effort, we will move from
401 active, interest-driven biobank recruitment towards a more passive, automatized
402 participatory system. To achieve this, we will first introduce the Broad Consent of the German
403 Medical Informatics Initiative, for which we have developed an additional Psychiatry Module
404 tailored to the needs of mental health research (30). This modified Broad Consent, which will
405 be valid for 5 years, will enable the scientific use of prospectively collected routine clinical and
406 insurance data, and, via the Psychiatry Module, additionally collected mental health-specific
407 phenotypes, magnetic resonance imaging, electroencephalogram data, and biosamples.
408 Next, we will introduce the Clinical Phenotyping Platform (CliPP), a software solution that
409 enables the collection of structured and standardized (Fast Healthcare Interoperability
410 Resource (FHIR) format) routine mental health data directly from patients and health care
411 providers. Owing to the Broad Consent and the CliPP, the consent process and the phenotypic

412 assessments will be incorporated into routine admission procedures. Their introduction will
413 broaden the available data modalities, reduce the time spent on patient recruitment, and
414 decrease the risk of sampling bias due to physician and/or patient preferences. Furthermore,
415 we will continuously monitor the representativeness of the MMHB Sample, analyze the
416 factors that influence biobank participation, and, if necessary, revise study design and
417 recruitment strategies. Through these measures, we will facilitate recruitment of participants
418 into the MMHB, improve clinician and patient commitment and the representativeness of our
419 sample, and ultimately, increase the scientific and clinical value of the MMHB Sample. Large
420 future transdiagnostic cohorts from the MMHB will leverage forward and reverse
421 translational research in an informative framework spanning preclinical models via disorder
422 focused clinical trials to real-world applications in a transdiagnostic spectrum of mental health
423 disorders including their developmental dynamics and co-morbidities.

424

425

426 5. References

- 427 1. Liu Q, He H, Yang J, Feng X, Zhao F, Lyu J. Changes in the global burden of depression
428 from 1990 to 2017: Findings from the Global Burden of Disease study. *J Psychiatr Res.*
429 2020 Jul 1;126:134–40.
- 430 2. John A, McGregor J, Fone D, Dunstan F, Cornish R, Lyons RA, et al. Case-finding for
431 common mental disorders of anxiety and depression in primary care: an external
432 validation of routinely collected data. *BMC Med Inform Decis Mak.* 2016 Mar 15;16:35.
- 433 3. Biobanking.com. Large Study on Mental Health Genetics Launched by Estonian Biobank
434 [Internet]. Biobanking.com. 2021 [cited 2022 Apr 12]. Available from:
435 [https://www.biobanking.com/large-study-on-mental-health-genetics-launched-by-](https://www.biobanking.com/large-study-on-mental-health-genetics-launched-by-estonian-biobank/)
436 [estonian-biobank/](https://www.biobanking.com/large-study-on-mental-health-genetics-launched-by-estonian-biobank/)
- 437 4. Davis K, Hotopf M. Mental health phenotyping in UK Biobank. *Prog Neurol Psychiatry.*
438 2019;23(1):4–7.
- 439 5. Gehring S, Eulenfeld R. German Medical Informatics Initiative: Unlocking Data for
440 Research and Health Care. *Methods Inf Med.* 2018 Jul;57(Suppl 1):e46–9.
- 441 6. Witt S, Dukal H, Hohmeyer C, Radosavljevic-Bjelic S, Schendel D, Frank J, et al. Biobank
442 of Psychiatric Diseases Mannheim – BioPsy. *Open J Bioresour.* 2016 Jul 8;3(1):e2.
- 443 7. Klingler C, von Jagwitz-Biegnitz M, Hartung ML, Hummel M, Specht C. Evaluating the
444 German Biobank Node as Coordinating Institution of the German Biobank Alliance:
445 Engaging with Stakeholders via Survey Research. *Biopreservation Biobanking.* 2020
446 Apr;18(2):64–72.
- 447 8. Hummel M, Specht C. Biobanks for future medicine. *J Lab Med.* 2019 Dec 1;43(6):383–
448 8.
- 449 9. Budde M, Anderson-Schmidt H, Gade K, Reich-Erkelenz D, Adorjan K, Kalman JL, et al. A
450 longitudinal approach to biological psychiatric research: The PsyCourse study. *Am J Med*
451 *Genet Part B Neuropsychiatr Genet Off Publ Int Soc Psychiatr Genet.* 2019
452 Mar;180(2):89–102.
- 453 10. Bush K, Kivlahan DR, McDonell MB, Fihn SD, Bradley KA. The AUDIT alcohol consumption
454 questions (AUDIT-C): an effective brief screening test for problem drinking. *Ambulatory*
455 *Care Quality Improvement Project (ACQUIP). Alcohol Use Disorders Identification Test.*
456 *Arch Intern Med.* 1998 Sep 14;158(16):1789–95.
- 457 11. Grabe HJ, Schulz A, Schmidt CO, Appel K, Driessen M, Wingefeld K, et al. [A brief
458 instrument for the assessment of childhood abuse and neglect: the childhood trauma
459 screener (CTS)]. *Psychiatr Prax.* 2012 Apr;39(3):109–15.
- 460 12. Smith BW, Dalen J, Wiggins K, Tooley E, Christopher P, Bernard J. The brief resilience
461 scale: assessing the ability to bounce back. *Int J Behav Med.* 2008;15(3):194–200.

- 462 13. Hughes ME, Waite LJ, Hawkey LC, Cacioppo JT. A Short Scale for Measuring Loneliness
463 in Large Surveys: Results From Two Population-Based Studies. *Res Aging*.
464 2004;26(6):655–72.
- 465 14. Lubben J, Blozik E, Gillmann G, Iliffe S, von Renteln Kruse W, Beck JC, et al. Performance
466 of an abbreviated version of the Lubben Social Network Scale among three European
467 community-dwelling older adult populations. *The Gerontologist*. 2006 Aug;46(4):503–
468 13.
- 469 15. Topp CW, Østergaard SD, Søndergaard S, Bech P. The WHO-5 Well-Being Index: a
470 systematic review of the literature. *Psychother Psychosom*. 2015;84(3):167–76.
- 471 16. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity
472 measure. *J Gen Intern Med*. 2001 Sep;16(9):606–13.
- 473 17. Roenneberg T, Wirz-Justice A, Meroo M. Life between clocks: daily temporal patterns
474 of human chronotypes. *J Biol Rhythms*. 2003 Feb;18(1):80–90.
- 475 18. Barton BB, Zagler A, Engl K, Rihs L, Musil R. Prevalence of obesity, metabolic syndrome,
476 diabetes and risk of cardiovascular disease in a psychiatric inpatient sample: results of
477 the Metabolism in Psychiatry (MiP) Study. *Eur Arch Psychiatry Clin Neurosci*. 2020
478 Aug;270(5):597–609.
- 479 19. Cappelleri JC, Bushmakin AG, Gerber RA, Leidy NK, Sexton CC, Lowe MR, et al.
480 Psychometric analysis of the Three-Factor Eating Questionnaire-R21: results from a large
481 diverse sample of obese and non-obese participants. *Int J Obes* 2005. 2009
482 Jun;33(6):611–20.
- 483 20. Rek SV, Bühner M, Reinhard MA, Freeman D, Keeser D, Adorjan K, et al. The COVID-19
484 Pandemic Mental Health Questionnaire (CoPaQ): psychometric evaluation and
485 compliance with countermeasures in psychiatric inpatients and non-clinical individuals.
486 *BMC Psychiatry*. 2021 Aug 31;21(1):426.
- 487 21. Guy W. ECDEU assessment manual for psychopharmacology. US Department of Health,
488 Education, and Welfare, Public Health Service ...; 1976.
- 489 22. Jones SH, Thornicroft G, Coffey M, Dunn G. A brief mental health outcome scale-
490 reliability and validity of the Global Assessment of Functioning (GAF). *Br J Psychiatry J*
491 *Ment Sci*. 1995 May;166(5):654–9.
- 492 23. Pirastu N, Cordioli M, Nandakumar P, Mignogna G, Abdellaoui A, Hollis B, et al. Genetic
493 analyses identify widespread sex-differential participation bias. *Nat Genet*. 2021 May
494 1;53(5):663–71.
- 495 24. Taylor AE, Jones HJ, Sallis H, Euesden J, Stergiakouli E, Davies NM, et al. Exploring the
496 association of genetic factors with participation in the Avon Longitudinal Study of
497 Parents and Children. *Int J Epidemiol*. 2018 Aug 1;47(4):1207–16.

- 498 25. Tyrrell J, Zheng J, Beaumont R, Hinton K, Richardson TG, Wood AR, et al. Genetic
499 predictors of participation in optional components of UK Biobank. *Nat Commun.* 2021
500 Feb 9;12(1):886.
- 501 26. Larsson H. The importance of selection bias in prospective birth cohort studies. *JCPP Adv.*
502 2021;1(3):e12043.
- 503 27. Koutsouleris N, Dwyer DB, Degenhardt F, Maj C, Urquijo-Castro MF, Sanfelici R, et al.
504 Multimodal Machine Learning Workflows for Prediction of Psychosis in Patients With
505 Clinical High-Risk Syndromes and Recent-Onset Depression. *JAMA Psychiatry.* 2021 Feb
506 1;78(2):195–209.
- 507 28. Barak-Corren Y, Castro VM, Nock MK, Mandl KD, Madsen EM, Seiger A, et al. Validation
508 of an Electronic Health Record-Based Suicide Risk Prediction Modeling Approach Across
509 Multiple Health Care Systems. *JAMA Netw Open.* 2020 Mar 2;3(3):e201262.
- 510 29. Chekroud AM, Zotti RJ, Shehzad Z, Gueorguieva R, Johnson MK, Trivedi MH, et al. Cross-
511 trial prediction of treatment outcome in depression: a machine learning approach.
512 *Lancet Psychiatry.* 2016 Mar;3(3):243–50.
- 513 30. Semler SC, Wissing F, Heyder R. German Medical Informatics Initiative. *Methods Inf Med.*
514 2018 Jul;57(Suppl 1):e50–6.
- 515