Received: 30 August 2023 | Revised: 6 October 2023 | Accepted: 5 December 2023

DOI: 10.1002/mpr.2003

ORIGINAL ARTICLE

WILEY

Proof-of-concept of a data-driven approach to estimate the associations of comorbid mental and physical disorders with global health-related disability

Ymkje Anna de Vries[1] | Jordi Alonso[2,3,4] | Somnath Chatterji[5] | Peter de Jonge[6] | Joran Lokkerbol[7] | John J. McGrath[8,9,10] | Maria V. Petukhova[11] | Nancy A. Sampson[11] | Erik Sverdrup[12] | Daniel V. Vigo[13] | Stefan Wager[12] | Ali Al-Hamzawi[14] | Guilherme Borges[15] | Ronny Bruffaerts[16] | Brendan Bunting[17] | Stephanie Chardoul[18] | Elie G. Karam[19,20,21] | Andrzej Kiejna[22] | Viviane Kovess-Masfety[23] | Fernando Navarro-Mateu[24,25,26] | Akin Ojagbemi[27] | Marina Piazza[28] | José Posada-Villa[29] | Carmen Sasu[30] | Kate M. Scott[31] | Hisateru Tachimori[32] | Margreet Ten Have[33] | Yolanda Torres[34] | Maria Carmen Viana[35] | Manuel Zamparini[36] | Zahari Zarkov[37] | Ronald C. Kessler[11] | on behalf of the World Mental Health Survey Collaborators

Correspondence
Ronald C. Kessler, Department of Health Care Policy, Harvard Medical School, 180 Longwood Ave, Ste 215, Boston, MA 02115-5899, USA.
Email: kessler@hcp.med.harvard.edu

Abstract

Objective: The standard method of generating disorder-specific disability scores has lay raters make rankings between pairs of disorders based on brief disorder vignettes. This method introduces bias due to differential rater knowledge of disorders and inability to disentangle the disability due to disorders from the disability due to comorbidities.

Methods: We propose an alternative, data-driven, method of generating disorder-specific disability scores that assesses disorders in a sample of individuals either from population medical registry data or population survey self-reports and uses Generalized Random Forests (GRF) to predict global (rather than disorder-specific) disability assessed by clinician ratings or by survey respondent self-reports. This method also provides a principled basis for studying patterns and predictors of heterogeneity in disorder-specific disability. We illustrate this method by analyzing data for 16 disorders assessed in the World Mental Health Surveys (n = 53,645).

Results: Adjustments for comorbidity decreased estimates of disorder-specific disability substantially. Estimates were generally somewhat higher with GRF than conventional multivariable regression models. Heterogeneity was nonsignificant.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2023 The Authors. International Journal of Methods in Psychiatric Research published by John Wiley & Sons Ltd.

Conclusions: The results show clearly that the proposed approach is practical, and that adjustment is needed for comorbidities to obtain accurate estimates of disorder-specific disability. Expansion to a wider range of disorders would likely find more evidence for heterogeneity.

KEYWORDS
causal forest, comorbidity, disability, global burden of disease, mental disorders

1 | INTRODUCTION

Most common mental disorders are highly comorbid (McGrath et al., 2020; Plana-Ripoll et al., 2019). Indeed, half of all people with a mental disorder in the past year meet criteria for two or more such disorders (Kessler et al., 2005). Chronic physical disorders also tend to be comorbid, as one-third of people with any such disorder have two or more (Nguyen et al., 2019). In addition, mental disorders predict increased risk of subsequent physical disorders (Momen et al., 2020; Scott et al., 2016) and physical disorders predict increased risk of subsequent mental disorders (Cohen et al., 1998; Hotopf et al., 1998).

Both mental and physical disorders are associated with considerable disability (Alonso, Petukhova, et al., 2011; Bruffaerts et al., 2012; Moussavi et al., 2007; Ormel et al., 2008). However, the high comorbidity among disorders makes it difficult to disentangle how much of an individual's disability is due to specific disorders. Failure to account for comorbidity may lead to overestimating disorder-specific disability. In fact, population-based epidemiologic surveys that assess diverse mental and physical disorders show that associations of specific disorders with role impairment (Alonso, Petukhova, et al., 2011; Bruffaerts et al., 2012) perceived health (Alonso, Vilagut, et al., 2011), and overall self-rated health (Moussavi et al., 2007), decrease by up to 70% when adjusting for comorbidity.

It would be useful to expand analyses of disorder-specific disability to account for comorbidity. In the global burden of disease (GBD) framework (Murray, 2022), the most commonly-used system for estimating disorder-specific disability, disability scores are determined by presenting general population samples with pairs of brief vignette describing two different disorders and asking which one the respondent considers more disabling (Burstein et al., 2015). This approach is thought to be superior to asking patients to report the extent to which they believe their own disorders cause disability because patients might have biased perceptions due either to an exaggerated belief that one of their disorders accounts for disabilities when, in fact, the disability is due to other causes (e.g., other disorders or personal characteristics) or due to adaptations leading to under-reports of disability (Stiggelbout & de Vogel-Voogt, 2008). Third-party ratings are thought to be more objective because they avoid these personal biases.

However, a major drawback with the vignette approach is that most people lack enough familiarity with the disorders in question to make well-informed evaluations. Vignette descriptions are designed to address this problem by providing a common core of information about the conditions, but the vignettes are necessarily sparse and raters read in information based on their own differential familiarity with the disorders (Stiggelbout & de Vogel-Voogt, 2008). Consistent with this concern, disability scores assigned to disorders are highly dependent on variations in health state descriptions (Salomon et al., 2015). In addition, some people have biased perceptions about certain disorders that influence their ratings even when they are instructed to consider only the information in the vignettes. For example, the disability score assigned by general population raters to acute schizophrenia (0.78) is implausibly higher than the disability score assigned to severe multiple sclerosis (0.72) (GBD 2019 Diseases and Injuries Collaborators, 2020). In other cases, popular perceptions of the disability associated with a specific disorder are accurate but more proximally due to comorbidities. For example, much of the disability known to be associated with Type 2 diabetes mellitus, a disorder with which many people are familiar, is more proximally due to such comorbidities as cardiovascular disease and chronic kidney disease (Marassi & Fadini, 2023).

GBD assumes that such biases do not exist and that the vignette-based disorder-specific disability estimates produced by lay raters in general population samples are accurate. Estimates of the joint effects of comorbidity based on this assumption are then obtained by multiplying disease-specific disability estimates and discounting the product to prevent the disability score of any one individual from exceeding 1.0 (GBD 2017 Disease and Injury Incidence and Prevalence Collaborators, 2018). The implicit assumption in this approach is that comorbidity influences disability sub-additively and equivalently for all disorders (GBD 2019 Diseases and Injuries Collaborators, 2020). This assumption is untested.

Given the high prevalence of comorbidity, it would be useful from a public health perspective to have a more data-driven approach to estimate the effects of pure disorders in a way that adjusts for the effects of comorbidity empirically. One possible way of doing this would be to expand the GBD vignettes to include common comorbidity profiles rather than use a priori multiplication to approximate the effects of comorbidity (Mansourian et al., 2022). However, this would not resolve the perceptual bias problems noted in the prior paragraph.

Another possibility, and the one that we explore in the current report, would be to collect data from large representative general population samples based on either electronic health records (EHR) and/or self-reports of individual-level disorder and investigate the associations of these disorders with *global* disability (as opposed to

disorder-specific disability). The latter could be assessed either by self-report or, in EHR samples, clinician ratings. Recently developed statistical methods could then be used to estimate average disorder-specific disability adjusting in a principled way for comorbidities as well as to explore the possibility that disorder-specific disability varies systematically as a joint function of comorbid disorders and other potential specifiers.

We illustrate this approach in a secondary analysis of data collected in the World Mental Health (WMH) surveys, a large coordinated series of community epidemiological surveys carried out in countries throughout the world that assessed self-reported prevalence of common mental disorders with a fully structured diagnostic interview (https://www.hcp.med.harvard.edu/wmh/) and chronic physical disorders with a standard conditions checklist. The self-reported outcome was based on a modified version of the WHO Disability Assessment Schedule (WHODAS-II) (Von Korff et al., 2008), a widely used self-report scale designed to assess global disability due to overall "health-related problems" in six domains: cognition, mobility, self-care, getting along, life activities, and participation. The analytic approach applied to these data was the Generalized Random Forests (GRF) (Athey & Wager, 2019; Wager & Athey, 2018) machine learning method. GRF focuses on estimating the effects of a discrete variable (in our case, a focal disorder) on an outcome (in our case, self-reported global disability) adjusting for measured confounders (in our case, age, sex, and comorbid disorders) in a way that controls optimally for the joint (i.e., potentially nonlinear and nonadditive) associations of all measured confounders and simultaneously provides a principled basis for studying heterogeneity in disorder-specific disability associated with these confounders.

2 | METHODS

2.1 | The sample

The WMH surveys are a coordinated set of community epidemiologic surveys of mental disorder prevalence and correlates carried out in countries throughout the world (Scott et al., 2018). Adults are selected using multi-stage clustered area probability sampling methods designed to generate samples representative of the household population. Calibration weights are used to match sample distributions to census population distributions on key socio-demographic/geographic variables. These methods are described elsewhere (Heeringa et al., 2008). We included all WMH surveys that assessed the relevant variables (mental and physical disorders and disability). A total of n = 53,645 respondents from 21 countries participated in these 24 WMH surveys.

2.2 | Field procedures

Informed consent was obtained before interviews using protocols established by local Institutional Review Boards. Interviews were conducted face-to-face by trained lay interviewers in respondents' homes. Consistent interviewer training and quality control monitoring procedures were used across surveys (Pennell et al., 2008). To reduce respondent burden, interviews were administered in two parts. Part I, which assessed core mental disorders, was administered to all respondents. Part II, which assessed other disorders and correlates, was then administered to all respondents with any lifetime Part I disorders and a probability subsample of other Part I respondents. Part II respondents were weighted by the inverse of their probability of selection into Part II to adjust for differential sampling. We used the weighted Part II sample in the present study, as physical disorders and disability were assessed in the Part II samples.

2.3 | Measurement

2.3.1 | Mental disorders

Twelve-month prevalence of mental disorders was assessed with the WHO Composite International Diagnostic Interview (CIDI) (Kessler & Üstün, 2008), a fully structured research diagnostic interview. DSM-IV diagnostic criteria were used in the analyses reported here, which focus on 8 disorders: alcohol abuse (with or without dependence), drug abuse (with or without dependence), Generalized anxiety disorder (GAD), Major depressive episode (MDE) (with or without bipolar disorder), panic disorder and/or agoraphobia, Post-traumatic stress disorder (PTSD), social anxiety disorder, and specific phobia. Clinical reappraisal studies show these CIDI diagnoses have good concordance with diagnoses based on blinded semi-structured clinical research diagnostic reinterviews (Ghimire et al., 2013; Gonzalez et al., 2016; Haro et al., 2006; Kessler et al., 2020; Lu et al., 2015).

2.3.2 | Physical disorders

Prevalence of eight common physical disorders in the 12 months before interview was assessed with a standard chronic disorder checklist adapted from the list used in the US National Health Interview Survey (Schoenborn et al., 2003): arthritis or chronic back/neck pain, cancer, Chronic obstructive pulmonary disease (COPD), diabetes, epilepsy, severe chronic headaches, heart disease, and other pain disorders. Checklists of this sort yield more complete and accurate reports about chronic conditions than open-ended questions (Knight et al., 2001) and have moderate to high concordance with medical records in developed countries (Baker et al., 2004; Galenkamp et al., 2014; van den Akker et al., 2015; Wada et al., 2009).

2.3.3 | Disability

As noted in the introduction, disability was defined using a recently published mapping algorithm (Lokkerbol et al., 2021) that generates

disability scores my calibrating scores on the WHODAS-II to a best-practices health state evaluation measure (Wijnen et al., 2018) obtained in the 2000–2001 WHO Multi-Country Survey Study (MCSS) on Health and Responsiveness (Üstün et al., 2001). The latter measure used multi-method ratings with visual analog, time trade-off, and standard gamble trade-off to define disability (Lokkerbol et al., 2021). The multivariable machine learning model using WHODAS dimension scores as predictors was found to predict the disability scores with high accuracy, justifying the use of this calibration method. The recall period for disability based on WHODAS reports was the past 30 days.

2.4 | Statistical analysis

2.4.1 | Estimating average disorder-specific disability

As noted in the introduction, the GRF method (Athey et al., 2018; Wager & Athey, 2018) was used to carry out the analysis. A separate GRF model was estimated for each of the 16 disorders, with one focal disorder treated as the predictor of primary interest in each model. The association between the focal disorder and the global disability score was the outcome. The other 15 comorbid disorders were treated as covariates that could have associations with the focal disorder and could modify the association between the focal disorder and the outcome. Respondent age and gender were also included as covariates in pooled within-country analyses.

GRF is like Random Forest (RF) (Breiman, 2001) in estimating an ensemble of regression trees and averaging across this ensemble to stabilize results. However, unlike RF, where the goal is to define splits in the regression trees that maximize between-node variation in mean outcome scores, the goal in GRF is to define splits that maximize between-node variation in the association between the focal disorder and the outcome. This is done in the GRF (*grf*) *R* package (Athey et al., 2018) by first estimating propensity scores to adjust for significant associations of the covariates with the focal disorder, then estimating expected marginal outcome scores given baseline covariates, and, finally, estimating associations of the focal disorder with the outcome for each respondent via a localized partial linear modeling estimator that orthogonalizes out the propensity score and marginalizes outcome estimates (Athey & Wager, 2019; Nie & Wager, 2021). A grid search is used to tune hyper-parameters (Feurer & Hutter, 2019).

This approach allows counter-factual logic to be used to estimate a predicted outcome score separately for each respondent in the presence and absence of the focal disorder given the covariates. The individual-level disorder-specific disability is the difference between the two counter-factual estimates (i.e., assuming the presence vs. absence of the focal disorder). The average disorder-specific disability is then estimated as the average of the individual-level disorder-specific disabilities among individuals with the disorder, which implicitly adjusts for the confounding effects of comorbidity because each individual-level estimate is obtained in a subsample defined by the multivariate covariate profile that accounts for all meaningful variation in the association between the focal disorder and the outcome.

We estimated a set of five models for the association of each focal disorder with global disability. The first three and the fifth were linear regression models. The focal disorder was the only predictor in Model 1. Model 2 expanded on Model 1 to include additive controls (i.e., assuming no interactions of the focal disorder with these other predictors) for age and sex. Model 3 added additional additive controls for the 15 other comorbid disorders. Model 4, the GRF model, adjusted for all stable complex interactions of age, sex, and comorbid disorders with the focal disorder. We used the GRF overlap-weighted method, which is recommended when, as in our data, predicted probability of having the focal disorder is close to 0 or 1 for a meaningful proportion of respondents (Athey & Wager, 2019). Model 5, finally, estimated the association of the focal disorder with the outcome controlling age and sex in the subsample of respondents who had none of the other 15 disorders, allowing us to characterize the disability associated with pure disorders in comparison to the average disorder-specific disability estimated in Model 4 across the full range of comorbidity profiles.

2.4.2 | Estimating heterogeneity of disorder-specific disability

Given that GRF generates individual-level estimates of the association between each focal disorder and global disability, these individual-level estimates can be treated as outcomes in analyses of the extent to which covariates are associated with significant inter-individual variation in disorder-specific disability. The *grf* package includes two ways to assess this kind of stability in consistency of estimates of individual-level disorder-specific disability. The first is the rank-weighted average treatment effect (RATE) method (Yadlowsky et al., 2021), which uses individual-level disorder-specific disability estimates based on models estimates in 50% training sample models and then applied to the remaining 50% test sample and parallel estimates based on an independent implementation of *grf* in the test sample. The average disorder-specific disability estimated in the full test sample is subtracted from the average disorder-specific disability estimated in the X% of the test sample with the highest individual-level disorder-specific disability based on the training sample model. This comparison is repeated across a range of X values (e.g., 1%, 2%, …50%, … 99%). If heterogeneity in disorder-specific disability is stable, we would expect the X% of respondents with the highest predicted disorder-specific disability based on the training sample model to have a higher estimated disorder-specific disability estimated in the test sample than the estimate in the full test sample, resulting in a curve of difference scores with increasing values of X, known as the Targeting Operator Characteristic (TOC) curve, that would become smaller as X increases and would decreases to 0 as X approaches 100%. Area under the TOC curve was calculated for

each disorder to quantify this association. If heterogeneity is absent, area under the TOC would have an expected value of 0. For a more detailed explanation, see https://grf-labs.github.io/grf/articles/rate.html.

The second approach to assess stable heterogeneity is to estimate the best linear projection of the estimated conditional association of the focal disorder on the outcome. This function estimates whether a predictor, in this case the disorder-specific disability score based on the training sample model, is associated with the disorder-specific disability score for people with a particular set of values for the covariates. If there is stable heterogeneity based on the covariates in the causal forest models, this association should be significant. An association of 0 implies no evidence for heterogeneity. For a more detailed explanation, see https://grf-labs.github.io/grf/reference/best_linear_projection.html.

All linear regression models were estimated in SAS 9.4 controlling for participant country and using the design-based Taylor series method to account for the clustering and weighting of the WMH data. The GRF models used country-mean-centered global disability scores and were estimated in R using the *grf* package. All tests were evaluated at $\alpha = 0.001$ to reduce false positives due to multiple testing.

3 | RESULTS

3.1 | Sample characteristics

Table 1 provides a description of the included surveys. Mean respondent age was 42.6 (SD = 17.1, range = 18–100). 52.3% of respondents were women. Disorder prevalence ranged from 0.5% for cancer to 25.9% for arthritis and other back/neck pain, with 45.4% of respondents reporting at least one disorder (Table 2). The mean global disability score was 0.12 (SD = 0.03, range = 0.09–0.64) among respondents with no disorders and 0.15 (SD = 0.07, range = 0.09–0.68) among respondents with at least one disorder. The great majority of respondents were clustered at the low end of the global disability scale, with 92.4% having a score below 0.20.

3.2 | Average disorder-specific disability

Table 3 shows estimates of average disorder-specific disability based on the five models described above in the section on analysis methods. In Model 1, the unadjusted model, three mental disorders (GAD, panic disorder, PTSD) and one physical disorder (cancer) had the highest average disorder-specific disabilities (0.043–0.055), whereas another three mental disorders (alcohol abuse, drug abuse, specific phobia) and one physical disorder (diabetes) had the lowest such averages (0.015–0.026). The inter-quartile range of disorder-specific disability estimates was 0.043–0.026, a ratio of about 1.6:1 that far exceeded the standard errors of the estimates.

The results in Model 2, which adjusted for respondent age and gender, were close to Model 1 for all mental disorders other than alcohol/drug abuse and for three physical disorders (epilepsy, headaches, and other pain disorders), but were substantially higher than in Model 1 for the other mental disorders and substantially lower than in Model 1 for the other physical disorders. Because of these changes, the four disorders with highest disorder-specific disabilities were mental disorders (the same three as in Model 1 plus MDE; 0.044–0.054) and the lowest were all physical disorders (the same two as in Model 1 in addition to arthritis and other back/neck pain and heart disease; 0.015–0.025). The inter-quartile range of disorder-specific disability estimates was 0.044–0.025, a ratio of about 1.8:1, again with disability estimates far exceeding their standard errors.

The results in Model 3, which additionally adjusted additively for comorbidity, were consistently and substantially lower than in Model 2, with greater proportional reductions for mental (40%–70%) than physical (17% for cancer and 30%–40% for the other physical disorders) disorders. Two mental disorders (MDE and PTSD) and two physical disorders (cancer and other pain disorders) had the highest (0.023–0.029) and another two mental (drug abuse and specific phobia) and two physical (diabetes and heart disease) disorders the lowest disorder-specific disabilities. The inter-quartile range of disorder-specific disability estimates was 0.0.023–0.012, a ratio of about 1.9:1, again far exceeding the standard errors of the estimates.

The results in Model 4, which adjusted non-additively for age, sex, and comorbidity using the *grf* approach, were either very similar to or somewhat higher than those in Model 4. The disorders with the highest disorder-specific disabilities were three mental (MDE, panic disorder, PTSD) and one physical (cancer) disorder, whereas those with lowest disorder-specific disabilities included one mental (specific phobia) and three physical (arthritis and other back/neck pain, diabetes, and heart disease) disorders. The inter-quartile range of disorder-specific disability estimates was 0.024–0.014, a ratio of about 1.7:1 and again far exceeding the standard errors of the estimates.

The results in Model 5, finally, were for pure disorders, which occur among 11%–40% of individuals with a given disorder. These estimates were either equivalent to or somewhat lower than those in Model 4. The disorders with the highest disorder-specific disabilities in these pure disorder comparisons were all mental disorder (alcohol abuse, GAD, MDE, PTSD), whereas those with lowest disorder-specific disabilities included one mental (specific phobia) and three physical (COPD, diabetes, and heart disease) disorders. The inter-quartile range of disorder-specific disability estimates was 0.016–0.009, a ratio of about 1.8:1, again with this range far exceeding the standard error of these estimates.

3.3 | Heterogeneity in disorder-specific disability

The RATE and best linear projection approaches both failed to find evidence for stable heterogeneity in disorder-specific disabilities with respect to the predictors in the causal forest models (Tables 4 and 5).

TABLE 1 WMH sample characteristics by World Bank income categories.[a]

Country by income category	Survey[b]	Sample characteristics[c]	Field dates	Age range	Sample size			
					Part I	Part II	Part II with WHODAS items[d]	Response rate[e]
I. Low and lower middle income countries								
Colombia	NSMH	All urban areas of the country (approximately 73% of the total national population)	2003	18–65	4426	2381	--	87.7
Iraq	IMHS	Nationally representative	2006-7	18–96	4332	4332	--	95.2
Nigeria	NSMHW	21 of the 36 states in the country, representing 57% of the national population. The surveys were conducted in Yoruba, Igbo, Hausa and Efik languages	2002-4	18–100	6752	2143	--	79.3
Peru	EMSMP	Five urban areas of the country (approximately 38% of the total national population)	2004-5	18–65	3930	1801	--	90.2
Total					(19,440)	(10,657)	(10,657)	86.5
II. Upper-middle income countries								
Brazil–São Paulo	São Paulo Megacity	São Paulo metropolitan area	2005-8	18–93	5037	2942	--	81.3
Bulgaria	NSHS	Nationally representative	2002-6	18–98	5318	2233	--	72.0
Bulgaria 2	NSHS–2	Nationally representative	2016-17	18–91	1508	578	--	61.0
Colombia–Medellin[f]	MMHHS	Medellin metropolitan area	2011-12	19–65	3261	1673	--	97.2
Lebanon	LEBANON	Nationally representative	2002-3	18–94	2857	1031	--	70.0
Mexico	M-NCS	All urban areas of the country (approximately 75% of the total national population)	2001-2	18–65	5782	2362	--	76.6
Romania	RMHS	Nationally representative	2005-6	18–96	2357	2357	--	70.9
Total					(26,120)	(13,176)	(13,176)	76.0
III. High-income countries								
Argentina	AMHES	Eight largest urban areas of the country (approximately 50% of the total national population)	2015	18–98	3927	2116	--	77.3
Belgium	ESEMeD	Nationally representative. The sample was selected from a national register of Belgium residents	2001-2	18–95	2419	1043	702	50.6
France	ESEMeD	Nationally representative. The sample was selected from a national list of households with listed telephone numbers	2001-2	18–97	2894	1436	995	45.9
Germany	ESEMeD	Nationally representative	2002-3	19–95	3555	1323	855	57.8
Italy	ESEMeD	Nationally representative. The sample was selected from municipality resident registries	2001-2	18–100	4712	1779	992	71.3
Japan	WMHJ 2002–2006	Eleven metropolitan areas	2002-6	20–98	4129	1682	--	55.1
Netherlands	ESEMeD	Nationally representative. The sample was selected from municipal postal registries	2002-3	18–95	2372	1094	769	56.4
New Zealand[g]	NZMHS	Nationally representative	2004-5	18–98	12,790	7312	--	73.3
N. Ireland	NISHS	Nationally representative	2005-8	18–97	4340	1986	--	68.4

TABLE 1 (Continued)

Country by income category	Survey[b]	Sample characteristics[c]	Field dates	Age range	Sample size			
					Part I	Part II	Part II with WHODAS items[d]	Response rate[e]
Poland	EZOP	Nationally representative	2010-11	18–65	10,081	4000	--	50.4
Spain	ESEMeD	Nationally representative	2001-2	18–98	5473	2121	1252	78.6
Spain–Murcia	PEGASUS-Murcia	Murcia region. Regionally representative	2010-12	18–96	2621	1459	--	67.4
United States	NCS-R	Nationally representative	2001-3	18–99	9282	5692	--	70.9
Total					(68,595)	(33,043)	(29,812)	63.3
IV. Total					(114,155)	(56,876)	(53,645)	69.1

[a]The World Bank (2012) Data. Accessed May 12, 2012 at: http://data.worldbank.org/country. Some of the WMH countries have moved into new income categories since the surveys were conducted. The income groupings above reflect the status of each country at the time of data collection. The current income category of each country is available at the preceding URL.

[b]NSMH (The Colombian National Study of Mental Health); IMHS (Iraq Mental Health Survey); NSMHW (The Nigerian Survey of Mental Health and Wellbeing); EMSMP (La Encuesta Mundial de Salud Mental en el Peru); CMDPSD (Comorbid Mental Disorders during Periods of Social Disruption); NSHS (Bulgaria National Survey of Health and Stress); MMHHS (Medellín Mental Health Household Study); LEBANON (Lebanese Evaluation of the Burden of Ailments and Needs of the Nation); M-NCS (The Mexico National Comorbidity Survey); RMHS (Romania Mental Health Survey); (Argentina Mental Health Epidemiologic Survey); ESEMeD (The European Study Of The Epidemiology Of Mental Disorders); NHS (Israel National Health Survey); WMHJ2002-2006 (World Mental Health Japan Survey); NZMHS (New Zealand Mental Health Survey); NISHS (Northern Ireland Study of Health and Stress); EZOP (Epidemiology of Mental Disorders and Access to Care Survey); NMHS (Portugal National Mental Health Survey); SNMHS (Saudi National Mental Health Survey); PEGASUS-Murcia (Psychiatric Enquiry to General Population in Southeast Spain-Murcia); NCS-R (The US National Comorbidity Survey Replication).

[c]Most WMH surveys are based on stratified multistage clustered area probability household samples in which samples of areas equivalent to counties or municipalities in the US were selected in the first stage followed by one or more subsequent stages of geographic sampling (e.g., towns within counties, blocks within towns, households within blocks) to arrive at a sample of households, in each of which a listing of household members was created and one or two people were selected from this listing to be interviewed. No substitution was allowed when the originally sampled household resident could not be interviewed. These household samples were selected from Census area data in all countries other than France (where telephone directories were used to select households) and the Netherlands (where postal registries were used to select households). Several WMH surveys (Belgium, Germany, Italy, Poland, Spain-Murcia) used municipal, country resident or universal health-care registries to select respondents without listing households. The Japanese sample is the only totally un-clustered sample, with households randomly selected in each of the 11 metropolitan areas and one random respondent selected in each sample household. Nineteen of the 29 surveys are based on nationally representative household samples.

[d]The WHODAS-II items were assessed in all participants in most countries, with the exception of the ESEMeD surveys in Belgium, France, Germany, Italy, the Netherlands, and Spain, where the WHODAS-II was assessed in all participants reporting some impairment during the screener and a 10% subsample of participants reporting no or very little impairment during the screener.

[e]The response rate is calculated as the ratio of the number of households in which an interview was completed to the number of households originally sampled, excluding from the denominator households known not to be eligible either because of being vacant at the time of initial contact or because the residents were unable to speak the designated languages of the survey. The weighted average response rate is 69.3%.

[f]Colombia moved from the "lower and lower-middle income" to the "upper-middle income" category between 2003 (when the Colombian National Study of Mental Health was conducted) and 2010 (when the Medellin Mental Health Household Study was conducted), hence Colombia's appearance in both income categories.

[g]For the purposes of cross-national comparisons we limit the sample to those 18+.

4 | DISCUSSION

We found significant disorder-specific disability for each of the disorders considered here. These estimates varied significantly across disorders and decreased substantially in models that controlled comorbidity. Three of the four disorders with the highest disorder-specific average treatment effects in the model that controlled most comprehensively for comorbidity (Model 4)—major depressive disorder, PTSD, and cancer—were also among the top four in the model with more conventional additive controls (Model 3). But only two of those four (MDE and cancer) were also among the top four in the univariable models (Model 1) most like those used in GBD. This variation in rank ordering across models illustrates the importance of controlling comorbidity.

We found no evidence for stable heterogeneity in the GRF disorder-specific disability estimates for any disorder. This contrasts with the assumption in GBD (GBD 2019 Diseases and Injuries Collaborators, 2020) that comorbid disorders combine sub-additively in promoting disability. As noted in the introduction, this assumption is mainly pragmatic in that it ensures that a person's total disability

TABLE 2 Prevalence of each disorder and mean global disability score associated with each disorder in the total sample (n = 53,645).

	Prevalence		Global disability				Sample size	
	Est	(SE)	Mean[a]	(SD)	Min	Max	Pure[b]	Total[c]
I. Mental disorders								
Alcohol abuse	2.0	(0.1)	0.15	(0.07)	0.10	0.67	521	1493
Drug abuse	0.6	(0.0)	0.16	(0.08)	0.10	0.53	84	511
Generalized anxiety disorder (GAD)	2.1	(0.1)	0.19	(0.09)	0.10	0.68	215	2004
Major depressive episode (MDE)	5.4	(0.1)	0.17	(0.09)	0.10	0.62	1162	5294
Panic disorder	1.8	(0.1)	0.18	(0.09)	0.09	0.62	186	1698
Post-traumatic stress disorder (PTSD)	1.7	(0.1)	0.19	(0.10)	0.10	0.68	198	1532
Social anxiety disorder (SAD)	2.8	(0.1)	0.17	(0.08)	0.10	0.61	488	2554
Specific phobia	6.3	(0.1)	0.16	(0.07)	0.09	0.68	1464	5410
II. Physical disorders								
Arthritis and other back/neck pain	25.9	(0.3)	0.15	(0.07)	0.09	0.68	6333	15,772
Cancer	0.5	(0.0)	0.18	(0.10)	0.10	0.57	101	334
Chronic obstructive pulmonary disease (COPD)	1.3	(0.1)	0.17	(0.08)	0.10	0.68	170	853
Diabetes	3.9	(0.1)	0.16	(0.08)	0.10	0.68	711	2395
Epilepsy	1.0	(0.1)	0.16	(0.08)	0.10	0.55	151	631
Headaches	11.9	(0.2)	0.16	(0.08)	0.09	0.68	2093	8133
Heart disease	4.6	(0.1)	0.16	(0.08)	0.10	0.68	713	2951
Other pain disorders	5.3	(0.1)	0.17	(0.09)	0.09	0.68	638	3623
III. Number of disorders								
No disorder	54.6	(0.3)	0.12	(0.03)	0.09	0.64	-	24,095
1 or more mental and 0 physical disorders	7.2	(0.1)	0.14	(0.06)	0.09	0.56	4318	6133
1 or more physical and 0 mental disorders	30.1	(0.3)	0.14	(0.06)	0.09	0.66	10,910	16,159
1 or more mental and 1 or more physical disorders	8.1	(0.1)	0.18	(0.09)	0.09	0.68	0	7258
IV. Total	100	-	0.13	(0.05)	0.09	0.68	15,228	53,645

[a]Mean global disability weight in the subsample of respondents who experienced the disorder.

[b]The number of respondents who experienced this disorder but none of the other disorders.

[c]The number of respondents who experienced this disorder whether or not they also experienced any of the other disorders.

score never exceeds 1 no matter how many disorders they have. But our data show empirically that this assumption may not be correct for individuals with low levels of comorbidity. It is almost certainly the case, of course, that individual-level disorder-specific comorbidity is lower among people with many disorders. However, we were unable to see this empirically in our example given that we considered only 16 disorders and few people in the sample experienced more than 3–4 of these. This means that evidence for heterogeneity in disorder-specific disability will be more likely to emerge in studies that consider a larger set of disorders, although such studies would also need to be based on larger samples to obtain stable estimates of heterogeneity. This requirement could easily be achieved by working with population registry data in which information was available on all ICD diagnoses for a very large sample of individuals. Although clinician expert ratings of global disability would be needed for this type of dataset to be practical, it would not be difficult to instruct clinicians to make such ratings as part of ongoing quality improvement initiatives.

Perhaps the most striking of our results is that disorder-specific disability estimates are much smaller than in GBD. There are several possible explanations for this. First, as noted in the introduction, people without a disorder have markedly different ideas about how disabling a disorder is. In the MCSS (Üstün et al., 2001), for example, where the mapping function used in our study was created, lay participants thought arthritis would be associated with moderate to severe impairment in most dimensions, whereas most participants who had arthritis rated themselves as only mildly impaired. Differences such as this may be due to biases on the part of either the lay public or patients (Stiggelbout & de Vogel-Voogt, 2008). Arguably, though, if one is interested in assessing disability burden,

TABLE 3 Association of disability weight with 12-month disorders in unadjusted and adjusted models.[a]

	Model 1[b]		Model 2[c]		Model 3[d]		Model 4[e]		Model 5[f]	
Disorder	**Est**	**(SE)**	**Est**	**(SE)**	**Est**	**(SE)**	**Est**	**(SE)**	**Est**	**(SE)**
I. Mental disorders										
Alcohol abuse	*0.015*	*(0.002)*	0.026	(0.002)	0.014	(0.002)	0.019	(0.004)	**0.016**	**(0.003)**
Drug abuse	*0.024*	*(0.004)*	0.036	(0.004)	*0.011*	*(0.004)*	0.014	(0.006)	0.010	(0.005)
Generalized anxiety disorder (GAD)	**0.052**	**(0.003)**	**0.052**	**(0.003)**	0.021	(0.003)	0.023	(0.004)	**0.024**	**(0.005)**
Major depressive episode (MDE)	0.043	(0.002)	**0.044**	**(0.002)**	**0.025**	**(0.002)**	**0.029**	**(0.003)**	**0.024**	**(0.002)**
Panic disorder	**0.049**	**(0.003)**	**0.050**	**(0.003)**	0.018	(0.003)	**0.024**	**(0.004)**	0.013	(0.004)
Post-traumatic stress disorder (PTSD)	**0.055**	**(0.003)**	**0.054**	**(0.003)**	**0.028**	**(0.003)**	**0.027**	**(0.004)**	**0.019**	**(0.005)**
Social anxiety disorder (SAD)	0.034	(0.002)	0.036	(0.002)	0.013	(0.002)	0.015	(0.003)	0.015	(0.004)
Specific phobia	*0.026*	*(0.001)*	*0.025*	*(0.001)*	*0.010*	*(0.001)*	*0.014*	*(0.002)*	*0.009*	*(0.002)*
II. Physical disorders										
Arthritis and other back/neck pain	0.027	(0.001)	*0.020*	*(0.001)*	0.012	(0.001)	*0.011*	*(0.001)*	0.010	(0.001)
Cancer	**0.047**	**(0.008)**	0.035	(0.007)	**0.029**	**(0.007)**	**0.030**	**(0.011)**	0.013	(0.006)
Chronic obstructive pulmonary disease (COPD)	0.034	(0.004)	0.025	(0.004)	0.014	(0.003)	0.016	(0.005)	*0.006*	*(0.003)*
Diabetes	*0.025*	*(0.002)*	*0.015*	*(0.002)*	*0.010*	*(0.002)*	*0.010*	*(0.003)*	*0.006*	*(0.002)*
Epilepsy	0.029	(0.004)	0.029	(0.004)	0.017	(0.004)	0.021	(0.006)	0.010	(0.004)
Headaches	0.030	(0.001)	0.029	(0.001)	0.017	(0.001)	0.018	(0.002)	0.013	(0.002)
Heart disease	0.032	(0.002)	*0.020*	*(0.002)*	*0.012*	*(0.002)*	*0.013*	*(0.003)*	*0.008*	*(0.002)*
Other pain disorders	0.041	(0.002)	0.037	(0.002)	**0.023**	**(0.002)**	0.023	(0.002)	0.015	(0.002)

[a]All estimates were significant at $p < 0.001$. Bold entries represent the 25% of disorders with the highest average disorder-specific disabilities. Italics entries represent the 25% of disorders with the lowest average disorder-specific disabilities.

[b]Model 1, a linear regression model in which the focal disorder was the only predictor.

[c]Model 2, a linear regression model that expanded on Model 1 to include additive controls (i.e., assuming no interactions of the focal disorder with these other predictors) for age and sex.

[d]Model 3, a linear regression model that added additional additive controls for the 15 other comorbid disorders.

[e]Model 4, the Generalized Random Forest (GRF) model, which adjusted for all stable complex interactions of age, sex, and comorbid disorders with the focal disorder. See the text for more details about GRF models. The GRF average disorder-specific disabilities used the overlap-weighted method, which is recommended when, as in our data, propensities (i.e., likelihood of having the focal disorder) are close to 0 or 1 for a meaningful proportion of respondents.

[f]Model 5, a model comparable to Model 2 that was estimated in the subsample of respondents who had none of the other 15 disorders.

assessments of the burden actually *experienced* should be more valid than lay people's assessments of *imagined* burden.

A second possible explanation for the difference in the disorder-specific disability estimates in our study versus GBD is that we used 12-month diagnoses for mental and physical disorders, while disability was assessed for the past 30 days. This probably led to an underestimate of the disability associated with currently active disorders, which is one good reason for considering our WMH results as providing merely proof-of-concept rather than accurate disorder-specific disability estimates. A third possible explanation for the difference is that GBD disability weights fail to adjust for the substantial effects of comorbidity documented in our study.

One other empirical study also yielded somewhat larger comorbidity-adjusted disorder-specific estimates of disability than in our analysis (Lokkerbol et al., 2013), although the estimates in that other study were also markedly smaller than the GBD estimates. That study included quality-of-life-related factors such as pain and affect in the calculation of disability, whereas we examined eight WHODAS items that assessed more strictly defined disability (i.e. impairments, functional limitations, and participation restrictions), consistent with the WHO International Classification of Functioning, Disability, and Health (ICF) (World Health Organization, 2007). Interestingly, we still found that mental disorders were at least as disabling as physical disorders, even though we did not include affect or other factors that might be confounded with mental disorder symptoms in defining disability.

It would be inappropriate to accept our findings as definitive because of the limitations discussed above related to the limited number of disorders considered and the mismatch between the recall periods for disability and disorder prevalence. It is nonetheless noteworthy that substantial variation was found in disorder-specific disabilities and that little evidence was found for heterogeneity in

TABLE 4 Rank-weighted estimates of heterogeneity in disorder-specific disabilities.[a]

Disorder	AU-TOC[b]	(SE)
Alcohol abuse	0.003	(0.004)
Drug abuse	−0.004	(0.003)
Generalized anxiety disorder (GAD)	0.016	(0.015)
Major depressive episode (MDE)	0.002	(0.003)
Panic disorder	0.003	(0.008)
Post-traumatic stress disorder (PTSD)	0.006	(0.009)
Social anxiety disorder (SAD)	−0.003	(0.011)
Specific phobia	0.002	(0.002)
Arthritis and other back/neck pain	0.001	(0.001)
Cancer	0.011	(0.008)
Chronic obstructive pulmonary disease (COPD)	−0.001	(0.003)
Diabetes	0.001	(0.003)
Epilepsy	0.005	(0.005)
Headaches	0.006	(0.003)
Heart disease	0.003	(0.003)
Other pain disorders	0.002	(0.003)

[a]See the text or https://grf-labs.github.io/grf/articles/rate.html for more details.

[b]AU-TOC, area under the treatment operating characteristic curve.

these disabilities as a function of comorbidity. If these results are confirmed in larger samples that consider a much more extensive set of disorders, it could have important implications for the allocation of disorder-specific intervention resources.

Advantages of the GRF approach are: (1) that it allows disorder-specific disability to be estimated in a principled fashion in the presence of high comorbidity and; (2) that it allows heterogeneity of individual-level disorder-specific disabilities to be examined rigorously. Previous studies on these topics either used simple linear-additive regression methods (Alonso et al., 2013; Bruffaerts et al., 2012) or relied on the GBD methodology (Kruijshaar et al., 2003; Verboom et al., 2011), GRF provides a much more principled and rigorous method of doing this than in previous studies. As noted above, the fact that we found little evidence for heterogeneity could be due to the relatively small number of disorders and only two other covariates (i.e., age and sex) included in the analysis. It might be that a more exhaustive analysis of potentially informative covariates would find evidence for greater heterogeneity.

An important additional consideration is that the disability score we considered here is highly skewed. It is easy to imagine, based on this fact, that estimates of the comparative burden of different disorders might differ depending on the transformations used for that variable. It is not implausible to think, for example, that the disorders associated with the highest probability of having *any* disability are quite different from those associated with the highest probability of *severe* disability. Hence, a different ranking of disorders might emerge

TABLE 5 Best linear projection estimates of heterogeneity in disorder-specific disabilities.[a]

Disorder	Est[b]	(SE)	*p*-value
Alcohol abuse	0.23	(0.20)	0.27
Drug abuse	0.21	(0.29)	0.47
Generalized anxiety disorder (GAD)	−0.52	(0.28)	0.07
Major depressive episode (MDE)	0.02	(0.14)	0.86
Panic disorder	−0.17	(0.24)	0.48
Post-traumatic stress disorder (PTSD)	0.23	(0.42)	0.59
Social anxiety disorder (SAD)	−0.05	(0.33)	0.86
Specific phobia	0.24	(0.18)	0.17
Arthritis and other back/neck pain	0.10	(0.10)	0.33
Cancer	0.04	(0.79)	0.96
Chronic obstructive pulmonary disease (COPD)	−0.24	(0.49)	0.63
Diabetes	0.02	(0.23)	0.93
Epilepsy	0.58	(0.38)	0.12
Headaches	0.35	(0.14)	0.01
Heart disease	0.23	(0.25)	0.35
Other pain disorders	−0.28	(0.22)	0.19

[a]See the text or https://grf-labs.github.io/grf/reference/best_linear_projection.html for more details.

[b]Est, estimate of linear association between individual-level disorder-specific disability estimates based on the *grf* models trained in the training sample and test sample.

depending on whether the analysis focused on any disability, on extreme levels of disability, or, as in the analysis presented here, on *average* disability. The same could be true for nonadditive effects of comorbidity and heterogeneity in estimates of disorder-specific disability. An investigation of these different outcomes is beyond the scope of the current report but should be considered in future research.

4.1 | Strengths and limitations of the illustrative analysis

We noted above that the WMH analysis reported here was presented only to illustrate the potential value of GRF rather than as a serious investigation of the disorder-specific disabilities of the few disorders considered here. The illustration was limited in two noteworthy respects.

First, we included only 8 mental and 8 chronic physical disorders. A much larger number of disorders would be required for a thorough analysis of disorder-specific disability. These two sets of disorders were selected because they represented the largest number of commonly reported disorders that were assessed in the largest number of WMH surveys. Additional WMH surveys exist that were excluded from the analysis because these disorders were not assessed in those surveys. And additional disorders were assessed in some, but not all the WMH surveys considered here.

Second, disability was assessed over a 30-day recall period and disorder prevalence over a 12-month recall period. This mismatch presumably influenced results, but we have no way of evaluating this possibility given that the WHODAS is based on 30-day recall and disorders on 12-month recall.

4.2 | More general strengths and limitations of the GRF approach

The GRF approach is very appealing in that it provides a principled basis for investigating the associations of specific disorders with global health-related disability separately within segments of the population that are defined by all informative multivariate covariate profiles. Subgrouping distinctions defined by splits in regression trees are made only if those distinctions are consequential in predicting variation in disorder-specific associations. Averaging disorder-specific estimates across all such subgroup distinctions consequently controls for all relevant measured baseline characteristics. At the same time, the subgrouping distinctions capture evidence of heterogeneity in the magnitude of disorder-specific disability, allowing individual-level variation in disorder-specific disability to be treated as an outcome in the investigation of both the magnitude and predictors of such heterogeneity.

However, an important limitation of using GRF to study disorder-specific disability is that conclusions about the magnitude disorder-specific disability and about predictors of heterogeneity depend fundamentally on the set of disorders included in the analysis. This means that the results reported here about the effects of comorbidity, based as they were on a consideration of only 8 mental and 8 chronic physical disorders, cannot be assumed to hold more generally. If this method is to be used as a widespread basis for estimating disorder-specific disability, broad agreement among stakeholders would be needed about the range of disorders to consider in the analysis and the types of datasets that would be needed to carry out the analysis.

A related practical challenge in using GRF is that it requires large samples with comprehensive assessments of many disorders to investigate patterns and correlates of disorder-specific disability. This challenge involves both the expense of large-scale data collection and the feasibility of assessing a wide range of disorders comprehensively. One way of addressing these challenges might be to work with EHR in large healthcare systems or, when available, population health registries. These databases are limited in that they capture data only on diagnosed conditions and do not routinely collect self-reported disability data. It might be possible to address these limitations, though, by carrying out comprehensive diagnostic assessments and obtaining patient reports of global health-related disability for a probability sample of individuals in registry databases and then using data-driven methods to impute missing values to the remaining cases.

A different sort of GRF limitation involves the fact that earlier disorders associated with the onset of later disorders and more proximally relate to disability might be thought of as overlooked root causes (Cohen et al., 1998; Hotopf et al., 1998; Momen et al., 2020; Scott et al., 2016). For example, depression may lead to an unhealthy lifestyle that increases risk for later physical conditions. Disability due to these later conditions could therefore also be attributed, at least in part, to earlier depression. It is important to recognize, though, that the search for such distal causes is quite a different undertaking than attempting to determine which *current* disorders are associated with *current* disability. It is only for the latter type of question that GRF is appropriate.

5 | CONCLUSIONS

In this study, we presented proof-of-concept for a novel method of estimating disorder-specific disability. We demonstrated in an illustrative cross-national general population dataset that comorbid disorders account for a large but varying proportion of the disability experienced by people with specific disorders. This result demonstrates that it is critical to adjust for comorbidity when examining disorder-specific disability. Although we expected to find significant heterogeneity in disorder-specific disability due to comorbidity, age, and gender, we found little such evidence in the dataset we considered. It is important to note, though, that heterogeneity might exist due to other characteristics and that different results would likely emerge in studies working with more extensive sets of disorders.

AFFILIATIONS

[1]Department of Child and Adolescent Psychiatry, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

[2]Health Services Research Group, Institut Hospital del Mar d'Investigacions Mediques, Barcelona, Spain

[3]Department of Medicine and Life Sciences, Universitat Pompeu Fabra, Barcelona, Spain

[4]Biomedical Research Networking Center in Epidemiology & Public Health (CIBERESP), Madrid, Spain

[5]Department of Information, Evidence, and Research, World Health Organization, Geneva, Switzerland

[6]Department of Developmental Psychology, University of Groningen, Groningen, Netherlands

[7]Centre of Economic Evaluation, Trimbos Institute (Netherlands Institute of Mental Health), Utrecht, Netherlands

[8]Queensland Centre for Mental Health Research, The Park Centre for Mental Health, Wacol, Queensland, Australia

[9]Queensland Brain Institute, University of Queensland, St Lucia, Queensland, Australia

[10]National Centre for Register-based Research, Aarhus Universitet, Aarhus, Midtjylland, Denmark

[11]Department of Health Care Policy, Harvard Medical School, Boston, Massachusetts, USA

[12]Stanford Graduate School of Business, Stanford University, Stanford, California, USA

[13]Department of Psychiatry, University of British Columbia, Vancouver, British Columbia, Canada

[14]College of Medicine, University of Al-Qadisiya, Diwaniya Governorate, Al Diwaniyah, Iraq

[15]National Institute of Psychiatry Ramón de la Fuente Muñiz, Mexico City, Mexico

[16]Universitair Psychiatrisch Centrum - Katholieke Universiteit Leuven (UPC-KUL), Campus Gasthuisberg, Leuven, Belgium

[17]School of Psychology, Ulster University, Londonderry, UK

[18]Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, Michigan, USA

[19]Institute for Development, Research, Advocacy and Applied Care (IDRAAC), Beirut, Lebanon

[20]Department of Psychiatry and Clinical Psychology, St George Hospital University Medical Center, Beirut, Lebanon

[21]Faculty of Medicine, University of Balamand, Beirut, Lebanon

[22]Faculty of Applied Studies, University of Lower Silesia, Wroclaw, Poland

[23]Institut de Psychologie, EA 4057, Université Paris Cité, Paris, France

[24]Unidad de Docencia, Investigación y Formación en Salud Mental (UDIF-SM), Gerencia Salud Mental, Servicio Murciano de Salud, Murcia, Spain

[25]Murcia Biomedical Research Institute (IMIB-Arrixaca), Murcia, Spain

[26]CIBER Epidemiology and Public Health-Murcia (CIBERESP-Murcia), Murcia, Spain

[27]Department of Psychiatry, University of Ibadan, Ibadan, Nigeria

[28]School of Public Health and Administration, Universidad Cayetano Heredia, Lima, Peru

[29]Faculty of Social Sciences, Colegio Mayor de Cundinamarca University, Bogota, Colombia

[30]National Institute of Health Services Management, Bucharest, Romania

[31]Department of Psychological Medicine, University of Otago, Dunedin, Otago, New Zealand

[32]Keio University School of Medicine, Tokyo, Japan

[33]Department of Epidemiology, Netherlands Institute of Mental Health and Addiction, Trimbos Institute, Utrecht, Netherlands

[34]Center for Excellence on Research in Mental Health, CES University, Medellin, Colombia

[35]Department of Social Medicine, Federal University of Espírito Santo, Vitoria, Brazil

[36]IRCCS Istituto Centro San Giovanni di Dio Fatebenefratelli, Brescia, Italy

[37]National Center of Public Health and Analyses, Sofia, Bulgaria

WORLD MENTAL HEALTH SURVEY COLLABORATORS CONSISTS OF

Sergio Aguilar-Gaxiola, MD, PhD; **Yasmin A. Altwaijri**, PhD; **Laura Helena Andrade**, MD, PhD; **Lukoye Atwoli**, MD, PhD; **Corina Benjet**, PhD; **Evelyn J. Bromet**, PhD; **Jose Miguel Caldas-de-Almeida**, MD, PhD; **Graça Cardoso**, MD, PhD; **Alfredo H. Cía**, MD; **Louisa Degenhardt**, PhD; **Giovanni de Girolamo**, MD; **Oye Gureje**, MD, DSc, FRCPsych; **Josep Maria Haro**, MD, PhD; **Meredith G. Harris**, PhD; **Hristo Hinkov**, MD, PhD; **Chi-yi Hu**, MD, PhD; **Aimee Nasser Karam**, PhD; **Georges Karam**, MD; **Alan E. Kazdin**, PhD; **Norito Kawakami**, MD, DMSc; **Salma Khaled**, PhD; **Maria Elena Medina-Mora**, PhD; **Jacek Moskalewicz**, PhD; **Daisuke Nishi**, MD, PhD; **Juan Carlos Stagnaro**, MD, PhD; **Dan J. Stein**, FRCPC, PhD; **Cristian Vladescu**, MD, PhD; **David R. Williams**, MPH, PhD; **Bogdan Wojtyniak**, ScD; **Peter Woodruff**, MBBS, PhD, FRCPsych; **Miguel Xavier**, MD, PhD; **Alan M. Zaslavsky**, PhD.

AUTHOR CONTRIBUTIONS

Ymkje Anna de Vries: Conceptualization; formal analysis; investigation; resources; visualization; writing – original draft; writing – review & editing; validation. **Jordi Alonso**: Conceptualization; funding acquisition; investigation; writing – review & editing. **Somnath Chatterji**: Conceptualization; investigation; writing – review & editing. **Peter de Jonge**: Conceptualization; funding acquisition; investigation; writing – review & editing. **Joran Lokkerbol**: Conceptualization; investigation; resources; validation; visualization; writing – review & editing. **John J. McGrath**: Conceptualization; investigation; writing – review & editing; funding acquisition. **Maria V. Petukhova**: Data curation; writing – review & editing; investigation; visualization; validation. **Nancy A. Sampson**: Investigation; project administration; supervision; writing – review & editing. **Erik Sverdrup**: Investigation; methodology; supervision; resources; software; writing – review & editing. **Daniel V. Vigo**: Conceptualization; investigation; writing – review & editing. **Stefan Wager**: Investigation; writing – review & editing; software; methodology; resources; supervision. **Ali Al-Hamzawi**: Investigation; writing – review & editing; funding acquisition. **Guilherme Borges**: Investigation; funding acquisition; writing – review & editing. **Ronny Bruffaerts**: Investigation; funding acquisition; writing – review & editing. **Brendan Bunting**: Investigation; writing – review & editing; funding acquisition. **Stephanie Chardoul**: Investigation; funding acquisition; writing – review & editing. **Elie G. Karam**: Investigation; funding acquisition; writing – review & editing. **Andrzej Kiejna**: Investigation; funding acquisition; writing – review &

editing. **Viviane Kovess-Masfety**: Investigation; funding acquisition; writing – review & editing. **Fernando Navarro-Mateu**: Investigation; funding acquisition; writing – review & editing. **Akin Ojagbemi**: Investigation; funding acquisition; writing – review & editing. **Marina Piazza**: Investigation; funding acquisition; writing – review & editing. **José Posada-Villa**: Investigation; writing – review & editing; funding acquisition. **Carmen Sasu**: Investigation; funding acquisition; writing – review & editing. **Kate M. Scott**: Investigation; funding acquisition; writing – review & editing. **Hisateru Tachimori**: Investigation; funding acquisition; writing – review & editing. **Margreet Ten Have**: Investigation; writing – review & editing. **Yolanda Torres**: Investigation; funding acquisition; writing – review & editing. **Maria Carmen Viana**: Investigation; funding acquisition; writing – review & editing. **Manuel Zamparini**: Investigation; funding acquisition; writing – review & editing. **Zahari Zarkov**: Investigation; funding acquisition; writing – review & editing. **Ronald C. Kessler**: Conceptualization; investigation; funding acquisition; writing – original draft; writing – review & editing; supervision.

ACKNOWLEDGMENTS

The World Mental Health (WMH) Survey Initiative is supported by the United States National Institute of Mental Health (NIMH; R01 MH070884), the John D. and Catherine T. MacArthur Foundation, the Pfizer Foundation, the United States Public Health Service (R13-MH066849, R01-MH069864, and R01 DA016558), the Fogarty International Center (FIRCA R03-TW006481), the Pan American Health Organization, Eli Lilly and Company, Ortho-McNeil Pharmaceutical Inc., GlaxoSmithKline, and Bristol-Myers Squibb. We thank the staff of the WMH Data Collection and Data Analysis Coordination Centres for assistance with instrumentation, fieldwork, and consultation on data analysis. None of the funders had any role in the design, analysis, interpretation of results, or preparation of this paper. The views and opinions expressed in this report are those of the authors and should not be construed to represent the views of the sponsoring organizations, agencies, or governments. The Argentina survey–Estudio Argentino de Epidemiología en Salud Mental (EASM)–was supported by a grant from the Argentinian Ministry of Health (Ministerio de Salud de la Nación)–(Grant Number 2002-17270/13-5). The São Paulo Megacity Mental Health Survey is supported by the State of São Paulo Research Foundation (FAPESP) Thematic Project Grant 03/00204-3. The Bulgarian Epidemiological Study of common mental disorders EPIBUL is supported by the Ministry of Health and the National Center for Public Health Protection. EPIBUL 2, conducted in 2016-17, is supported by the Ministry of Health and European Economic Area Grants. The Colombian National Study of Mental Health (NSMH) is supported by the Ministry of Social Protection. The Mental Health Study Medellín—Colombia was carried out and supported jointly by the Center for Excellence on Research in Mental Health (CES University) and the Secretary of Health of Medellín. The ESEMeD project is funded by the European Commission (Contracts QLG5-1999-01042; SANCO 2004123, and EAHC 20081308), the Piedmont Region [Italy]), Fondo de Investigación Sanitaria, Instituto de Salud Carlos III, Spain (FIS 00/0028), Ministerio de Ciencia y Tecnología, Spain (SAF 2000-158-CE), Generalitat de Catalunya (2017 SGR 452; 2014 SGR 748), Instituto de Salud Carlos III (CIBER CB06/02/0046, RETICS RD06/0011 REM-TAP), and other local agencies and by an unrestricted educational grant from GlaxoSmithKline. Implementation of the Iraq Mental Health Survey (IMHS) and data entry were carried out by the staff of the Iraqi MOH and MOP with direct support from the Iraqi IMHS team with funding from both the Japanese and European Funds through United Nations Development Group Iraq Trust Fund (UNDG ITF). The World Mental Health Japan (WMHJ) Survey is supported by the Grant for Research on Psychiatric and Neurological Diseases and Mental Health (H13-SHOGAI-023, H14-TOKUBETSU-026, H16-KOKORO-013, H25-SEISHIN-IPPAN-006) from the Japan Ministry of Health, Labour and Welfare. The Lebanese Evaluation of the Burden of Ailments and Needs of the Nation (L.E.B.A.N.O.N.) is supported by the Lebanese Ministry of Public Health, the WHO (Lebanon), National Institute of Health/Fogarty International Center (R03 TW006481-01), anonymous private donations to IDRAAC, Lebanon, and unrestricted grants from, Algorithm, AstraZeneca, Benta, Bella Pharma, Eli Lilly, Glaxo Smith Kline, Lundbeck, Novartis, OmniPharma, Pfizer, Phenicia, Servier, UPO. The Mexican National Comorbidity Survey (MNCS) is supported by The National Institute of Psychiatry Ramon de la Fuente (INPRFMDIES 4280) and by the National Council on Science and Technology (CONACyT-G30544-H), with supplemental support from the Pan American Health Organization (PAHO). Te Rau Hinengaro: The New Zealand Mental Health Survey (NZMHS) is supported by the New Zealand Ministry of Health, Alcohol Advisory Council, and the Health Research Council. The Nigerian Survey of Mental Health and Wellbeing (NSMHW) is supported by the WHO (Geneva), the WHO (Nigeria), and the Federal Ministry of Health, Abuja, Nigeria. The Northern Ireland Study of Mental Health was funded by the Health & Social Care Research & Development Division of the Public Health Agency. The Peruvian World Mental Health Study was funded by the National Institute of Health of the Ministry of Health of Peru. The Polish project Epidemiology of Mental Health and Access to Care–EZOP Project (PL 0256) was carried out by the Institute of Psychiatry and Neurology in Warsaw in consortium with Department of Psychiatry–Medical University in Wroclaw and National Institute of Public Health-National Institute of Hygiene in Warsaw and in partnership with Psykiatrist Institut Vinderen—Universitet, Oslo. The project was funded by the European Economic Area Financial Mechanism and the Norwegian Financial Mechanism. EZOP project was co-financed by the Polish Ministry of Health. The Romania WMH study projects "Policies in Mental Health Area" and "National Study regarding Mental Health and Services Use" were carried out by National School of Public Health & Health Services Management (former National Institute for Research & Development in Health), with technical support of Metro Media Transilvania, the National Institute of Statistics, Cheyenne Services SRL, Statistics Netherlands and were funded by Ministry of Public Health (former Ministry of Health) with supplemental support of Eli Lilly Romania SRL. The Psychiatric Enquiry to General Population in Southeast Spain—Murcia (PEGASUS-Murcia) Project has

been financed by the Regional Health Authorities of Murcia (Servicio Murciano de Salud and Consejería de Sanidad y Política Social) and Fundación para la Formación e Investigación Sanitarias (FFIS) of Murcia. The US National Comorbidity Survey Replication (NCS-R) is supported by the National Institute of Mental Health (NIMH; U01-MH60220) with supplemental support from the National Institute of Drug Abuse (NIDA), the Substance Abuse and Mental Health Services Administration (SAMHSA), the Robert Wood Johnson Foundation (RWJF; Grant 044708), and the John W. Alden Trust.

CONFLICT OF INTEREST STATEMENT

In the past 3 years, RCK was a consultant for Cambridge Health Alliance, Canandaigua VA Medical Center, Holmusk, Partners Healthcare, Inc., RallyPoint Networks, Inc., and Sage Therapeutics. He has stock options in Cerebral Inc., Mirah, PYM, Roga Sciences and Verisense Health. All other authors declare no conflicts of interest with respect to this manuscript.

DATA AVAILABILITY STATEMENT

Access to the cross-national World Mental Health (WMH) data is governed by the organizations funding and responsible for survey data collection in each country. These organizations made data available to the WMH consortium through restricted data sharing agreements that do not allow us to release the data to third parties. The exception is that the U.S. data are available for secondary analysis via the Inter-University Consortium for Political and Social Research (ICPSR), http://www.icpsr.umich.edu/icpsrweb/ICPSR/series/00527.

ETHICS STATEMENT

Informed consent was obtained before interviews using protocols established by local Institutional Review Boards. At all survey sites, the local ethics or institutional review committee reviewed and approved the protocol to ensure protection of human subjects, in line with appropriate international and local guidelines. Details of the ethics committees and informed consent procedures for the WMH surveys can be viewed at this link: http://www.hcp.med.harvard.edu/wmh/ftpdir/WMH_Ethics_approval.pdf.

ORCID

Jordi Alonso https://orcid.org/0000-0001-8627-9636
Somnath Chatterji https://orcid.org/0000-0002-3085-6550
Margreet Ten Have https://orcid.org/0000-0003-0086-2199
Ronald C. Kessler https://orcid.org/0000-0003-4831-2305

REFERENCES

Alonso, J., Petukhova, M., Vilagut, G., Chatterji, S., Heeringa, S., Ustun, T. B., Alhamzawi, A. O., Viana, M. C., Angermeyer, M., Bromet, E., Bruffaerts, R., de Girolamo, G., Florescu, S., Gureje, O., Haro, J. M., Hinkov, H., Hu, C. Y., Karam, E. G., Kovess, V., … Kessler, R. C. (2011a). Days out of role due to common physical and mental conditions: Results from the WHO world mental health surveys. *Molecular Psychiatry*, *16*(12), 1234–1246. https://doi.org/10.1038/mp.2010.101

Alonso, J., Vilagut, G., Adroher, N. D., Chatterji, S., He, Y., Andrade, L. H., Bromet, E., Bruffaerts, R., Fayyad, J., Florescu, S., de Girolamo, G., Gureje, O., Haro, J. M., Hinkov, H., Hu, C., Iwata, N., Lee, S., Levinson, D., Lépine, J. P., …, & Kessler, R. C. (2013). Disability mediates the impact of common conditions on perceived health. *PLoS One*, *8*(6), e65858. https://doi.org/10.1371/journal.pone.0065858

Alonso, J., Vilagut, G., Chatterji, S., Heeringa, S., Schoenbaum, M., Bedirhan Ustun, T., Rojas-Farreras, S., Angermeyer, M., Bromet, E., Bruffaerts, R., de Girolamo, G., Gureje, O., Haro, J. M., Karam, A. N., Kovess, V., Levinson, D., Liu, Z., Medina-Mora, M. E., Ormel, J., …, & Kessler, R. C. (2011b). Including information about co-morbidity in estimates of disease burden: Results from the world health organization world mental health surveys. *Psychological Medicine*, *41*(4), 873–886. https://doi.org/10.1017/S0033291710001212

Athey, S., Tibshirani, J., & Wager, S. (2018). Generalized random forests. *The Annals of Statistics*, *47*(2), 1148–1178. https://doi.org/10.1214/18-AOS1709

Athey, S., & Wager, S. (2019). Estimating treatment effects with causal forests: An application. *Observational studies*, *5*(2), 37–51. https://doi.org/10.1353/obs.2019.0001

Baker, M., Stabile, M., & Deri, C. (2004). What do self-reported, objective, measures of health measure? *Journal of Human Resources*, *39*(4), 1067–1093. https://doi.org/10.3386/w8419

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Bruffaerts, R., Vilagut, G., Demyttenaere, K., Alonso, J., Alhamzawi, A., Andrade, L. H., Benjet, C., Bromet, E., Bunting, B., de Girolamo, G., Florescu, S., Gureje, O., Haro, J. M., He, Y., Hinkov, H., Hu, C., Karam, E. G., Lepine, J. P., Levinson, D., …, & Kessler, R. C. (2012). Role of common mental and physical disorders in partial disability around the world. *The British Journal of Psychiatry*, *200*(6), 454–461. https://doi.org/10.1192/bjp.bp.111.097519

Burstein, R., Fleming, T., Haagsma, J., Salomon, J. A., Vos, T., & Murray, C. J. (2015). Estimating distributions of health state severity for the global burden of disease study. *Population Health Metrics*, *13*(1), 31. https://doi.org/10.1186/s12963-015-0064-y

Cohen, P., Pine, D. S., Must, A., Kasen, S., & Brook, J. (1998). Prospective associations between somatic illness and mental illness from childhood to adulthood. *American Journal of Epidemiology*, *147*(3), 232–239. https://doi.org/10.1093/oxfordjournals.aje.a009442

Feurer, M., & Hutter, F. (2019). Hyperparameter optimization. In F. Hutter, L. Kotthoff, & J. Vanschoren (Eds.), *Automated machine learning: Methods, systems, challenges* (pp. 3–33). Springer Cham.

Galenkamp, H., Huisman, M., Braam, A. W., Schellevis, F. G., & Deeg, D. J. (2014). Disease prevalence based on older people's self-reports increased, but patient-general practitioner agreement remained stable, 1992-2009. *Journal of Clinical Epidemiology*, *67*(7), 773–780. https://doi.org/10.1016/j.jclinepi.2014.02.002

Ghimire, D. J., Chardoul, S., Kessler, R. C., Axinn, W. G., & Adhikari, B. P. (2013). Modifying and validating the composite international diagnostic interview (CIDI) for use in Nepal. *International Journal of Methods in Psychiatric Research*, *22*(1), 71–81. https://doi.org/10.1002/mpr.1375

Gonzalez, L. E. M., Bernal, D. P. R., Mejía-Montoya, R., Bareño-Silva, J., Sierra-Hincapié, G., de Galvis, Y. T., Marulanda-Restrepo, D., Gómez-Sierra, N., & Gaviria-Arbeláez, S. (2016). Sensitivity and specificity between the composite international diagnostic interview version 3.0 (world mental health, CIDI) and the standardised clinical evaluation version I (SCID-I) in the mental health survey of the city of Medellin, 2012. *Revista Colombiana de Psiquiatria*, *45*(1), 22–27. https://doi.org/10.1016/j.rcp.2015.07.001

Haro, J. M., Arbabzadeh-Bouchez, S., Brugha, T. S., de Girolamo, G., Guyer, M. E., Jin, R., Lepine, J. P., Mazzi, F., Reneses, B., Vilagut, G., Sampson, N. A., & Kessler, R. C. (2006). Concordance of the composite international diagnostic interview version 3.0 (CIDI 3.0) with standardized clinical assessments in the WHO world mental health surveys. *International Journal of Methods in Psychiatric Research*, *15*(4), 167–180. https://doi.org/10.1002/mpr.196

Heeringa, S., Wells, J., Hubbard, F., Mneimneh, Z., Chiu, W., Sampson, N., & Berglund, P. (2008). Sample designs and sampling procedures. In R. C. Kessler & T. B. Ustun (Eds.), *The WHO world mental health surveys: Global perspectives on the epidemiology of mental disorders* (pp. 14–32). Cambridge University Press.

Hotopf, M., Mayou, R., Wadsworth, M., & Wessely, S. (1998). Temporal relationships between physical symptoms and psychiatric disorder: Results from a national birth cohort. *The British Journal of Psychiatry*, *173*(3), 255–261. https://doi.org/10.1192/bjp.173.3.255

GBD 2017 Disease and Injury Incidence and Prevalence Collaborators. (2018). Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: A systematic analysis for the global burden of disease study 2017. *The Lancet*, *392*(10159), 1789–1858. https://doi.org/10.1016/S0140-6736(18)32279-7

Kessler, R. C., Al-Desouki, M., King, A. J., Sampson, N. A., Al-Subaie, A. S., Al-Habeeb, A., Bilal, L., Shahab, M. K., Aradati, M., & Altwaijri, Y. A. (2020). Clinical reappraisal of the composite international diagnostic interview version 3.0 in the Saudi national mental health survey. *International Journal of Methods in Psychiatric Research*, *29*(3), e1828. https://doi.org/10.1002/mpr.1828

Kessler, R. C., Chiu, W. T., Demler, O., Merikangas, K. R., & Walters, E. E. (2005). Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the national comorbidity survey replication. *Archives of General Psychiatry*, *62*(6), 617–627. https://doi.org/10.1001/archpsyc.62.6.617

Kessler, R. C., & Üstün, T. B. (2008). The world health organization composite international diagnostic interview. In R. C. Kessler & T. B. Üstün (Eds.), *The WHO world mental health surveys: Global perspectives on the epidemiology of mental disorders* (pp. 58–90). Cambridge University Press.

Knight, M., Stewart-Brown, S., & Fletcher, L. (2001). Estimating health needs: The impact of a checklist of conditions and quality of life measurement on health information derived from community surveys. *Journal of Public Health*, *23*(3), 179–186. https://doi.org/10.1093/pubmed/23.3.179

Kruijshaar, M. E., Hoeymans, N., Bijl, R. V., Spijker, J., & Essink-Bot, M. L. (2003). Levels of disability in major depression: Findings from The Netherlands mental health survey and incidence study (NEMESIS). *Journal of Affective Disorders*, *77*(1), 53–64. https://doi.org/10.1016/s0165-0327(02)00099-x

Lokkerbol, J., Adema, D., de Graaf, R., ten Have, M., Cuijpers, P., Beekman, A., & Smit, F. (2013). Non-fatal burden of disease due to mental disorders in The Netherlands. *Social Psychiatry and Psychiatric Epidemiology*, *48*(10), 1591–1599. https://doi.org/10.1007/s00127-013-0660-8

Lokkerbol, J., Wijnen, B. F. M., Chatterji, S., Kessler, R. C., & Chisholm, D. (2021). Mapping of the world health organization's disability assessment Schedule 2.0 to disability weights using the multi-country survey study on health and responsiveness. *International Journal of Methods in Psychiatric Research*, *30*(3), e1886. https://doi.org/10.1002/mpr.1886

Lu, J., Huang, Y.-Q., Liu, Z.-R., & Cao, X.-L. (2015). Validity of Chinese version of the composite international diagnostic interview-3.0 in psychiatric settings. *Chinese Medical Journal*, *128*(18), 2462–2466. https://doi.org/10.4103/0366-6999.164930

Mansourian, M., Ghasemi, K., Haghdoost, A., Kopec, J. A., Sarrafzadegan, N., & Shariful Islam, S. M. (2022). Measuring the burden of comorbidity for ischaemic heart disease and four common non-communicable diseases in Iran, 1990–2017: A modelling study based on global burden of diseases data. *BMJ Open*, *12*(11), e054441. https://doi.org/10.1136/bmjopen-2021-054441

Marassi, M., & Fadini, G. P. (2023). The cardio-renal-metabolic connection: A review of the evidence. *Cardiovascular Diabetology*, *22*(1), 195. https://doi.org/10.1186/s12933-023-01937-x

McGrath, J. J., Lim, C. C. W., Plana-Ripoll, O., Holtz, Y., Agerbo, E., Momen, N. C., Mortensen, P. B., Pedersen, C. B., Abdulmalik, J., Aguilar-Gaxiola, S., Al-Hamzawi, A., Alonso, J., Bromet, E. J., Bruffaerts, R., Bunting, B., de Almeida, J. M. C., de Girolamo, G., De Vries, Y. A., Florescu, S., …, & de Jonge, P. (2020). Comorbidity within mental disorders: A comprehensive analysis based on 145,990 survey respondents from 27 countries. *Epidemiology and Psychiatric Sciences*, *29*, e153. https://doi.org/10.1017/S2045796020000633

Momen, N. C., Plana-Ripoll, O., Agerbo, E., Benros, M. E., Borglum, A. D., Christensen, M. K., Dalsgaard, S., Degenhardt, L., de Jonge, P., Debost, J. P. G., Fenger-Gron, M., Gunn, J. M., Iburg, K. M., Kessing, L. V., Kessler, R. C., Laursen, T. M., Lim, C. C. W., Mors, O., Mortensen, P. B., …, & McGrath, J. J. (2020). Association between mental disorders and subsequent medical conditions. *The New England Journal of Medicine*, *382*(18), 1721–1731. https://doi.org/10.1056/NEJMoa1915784

Moussavi, S., Chatterji, S., Verdes, E., Tandon, A., Patel, V., & Ustun, B. (2007). Depression, chronic diseases, and decrements in health: Results from the world health surveys. *The Lancet*, *370*(9590), 851–858. https://doi.org/10.1016/S0140-6736(07)61415-9

Murray, C. J. L. (2022). The global burden of disease study at 30 years. *Nature Medicine*, *28*(10), 2019–2026. https://doi.org/10.1038/s41591-022-01990-1

Nguyen, H., Manolova, G., Daskalopoulou, C., Vitoratou, S., Prince, M., & Prina, A. M. (2019). Prevalence of multimorbidity in community settings: A systematic review and meta-analysis of observational studies. *Journal of Multimorbidity and Comorbidity*, *9*, 223504 2X19870934. https://doi.org/10.1177/2235042X19870934

Nie, X., & Wager, S. (2021). Quasi-oracle estimation of heterogeneous treatment effects. *Biometrika*, *108*(2), 299–319. https://doi.org/10.1093/biomet/asaa076

Ormel, J., Petukhova, M., Chatterji, S., Aguilar-Gaxiola, S., Alonso, J., Angermeyer, M. C., Bromet, E. J., Burger, H., Demyttenaere, K., de Girolamo, G., Haro, J. M., Hwang, I., Karam, E., Kawakami, N., Lepine, J. P., Medina-Mora, M. E., Posada-Villa, J., Sampson, N., Scott, K., …, & Kessler, R. C. (2008). Disability and treatment of specific mental and physical disorders across the world. *The British Journal of Psychiatry*, *192*(5), 368–375. https://doi.org/10.1192/bjp.bp.107.039107

Pennell, B., Mneimneh, Z., Bowers, A., Chardoul, S., Wells, J., Viana, M., Dinkelmann, K., Gebler, N., Florescu, S., & He, Y. (2008). Implementation of the world mental health surveys. In R. Kessler & T. B. Üstün (Eds.), *The WHO world mental health surveys: Global perspectives on the epidemiology of mental disorders* (pp. 33–57). Cambridge University Press.

Plana-Ripoll, O., Pedersen, C. B., Holtz, Y., Benros, M. E., Dalsgaard, S., de Jonge, P., Fan, C. C., Degenhardt, L., Ganna, A., Greve, A. N., Gunn, J., Iburg, K. M., Kessing, L. V., Lee, B. K., Lim, C. C. W., Mors, O., Nordentoft, M., Prior, A., Roest, A. M., …, & McGrath, J. J. (2019). Exploring comorbidity within mental disorders among a Danish national population. *JAMA Psychiatry*, *76*(3), 259–270. https://doi.org/10.1001/jamapsychiatry.2018.3658

Salomon, J. A., Haagsma, J. A., Davis, A., de Noordhout, C. M., Polinder, S., Havelaar, A. H., Cassini, A., Devleesschauwer, B., Kretzschmar, M., Speybroeck, N., Murray, C. J., & Vos, T. (2015). Disability weights for the global burden of disease 2013 study. *The Lancet Global Health*, *3*(11), e712–e723. https://doi.org/10.1016/S2214-109X(15)00069-8

Schoenborn, C. A., Adams, P. F., & Schiller, J. S. (2003). Summary health statistics for the U.S. population: National health interview survey, 2000. *Vital and Health Statistics Series*, *10*(214), 1–83. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/15786609

Scott, K. M., de Jonge, P., Stein, D. J., & Kessler, R. C. (2018). *Mental disorders around the world: Facts and figures from the WHO World Mental Health surveys*. Cambridge University Press.

Scott, K. M., Lim, C., Al-Hamzawi, A., Alonso, J., Bruffaerts, R., Caldas-de-Almeida, J. M., Florescu, S., de Girolamo, G., Hu, C., de Jonge, P., Kawakami, N., Medina-Mora, M. E., Moskalewicz, J., Navarro-Mateu, F., O'Neill, S., Piazza, M., Posada-Villa, J., Torres, Y., & Kessler, R. C. (2016). Association of mental disorders with subsequent chronic physical conditions: World Mental Health Surveys from 17 countries. *JAMA Psychiatry*, *73*(2), 150–158. https://doi.org/10.1001/jamapsychiatry.2015.2688

Stiggelbout, A. M., & de Vogel-Voogt, E. (2008). Health state utilities: A framework for studying the gap between the imagined and the real. *Value in Health*, *11*(1), 76–87. https://doi.org/10.1111/j.1524-4733.2007.00216.x

Üstün, T. B., Chatterji, S., Villanueva, M., Bendib, L., Celik, C., Sadana, R., Valentine, N. B., Ortiz, J. P., Tandon, A., Salomon, J. A., Cao, Y., Xie, W. J., Özaltin, E., Mathers, C. D., & Murray, C. J. L. (2001). WHO multi-country survey study on health and responsiveness 2000-2001. In C. J. L. Murray & D. B. Evans (Eds.), *Health systems performance assessment: Debates, methods and empiricism* (pp. 761–796). World Health Organization.

van den Akker, M., van Steenkiste, B., Krutwagen, E., & Metsemakers, J. F. (2015). Disease or no disease? Disagreement on diagnoses between self-reports and medical records of adult patients. *European Journal of General Practice*, *21*(1), 45–51. https://doi.org/10.3109/13814788.2014.907266

Verboom, C. E., Sentse, M., Sijtsema, J. J., Nolen, W. A., Ormel, J., & Penninx, B. W. (2011). Explaining heterogeneity in disability with major depressive disorder: Effects of personal and environmental characteristics. *Journal of Affective Disorders*, *132*(1–2), 71–81. https://doi.org/10.1016/j.jad.2011.01.016

Von Korff, M., Crane, P. K., Alonso, J., Vilagut, G., Angermeyer, M. C., Bruffaerts, R., de Girolamo, G., Gureje, O., de Graaf, R., Huang, Y., Iwata, N., Karam, E. G., Kovess, V., Lara, C., Levinson, D., Posada-Villa, J., Scott, K. M., & Ormel, J. (2008). Modified WHODAS-II provides valid measure of global disability but filter items increased skewness. *Journal of Clinical Epidemiology*, *61*(11), 1132–1143. https://doi.org/10.1016/j.jclinepi.2007.12.009

GBD 2019 Diseases and Injuries Collaborators. (2020). Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: A systematic analysis for the global burden of disease study 2019. *The Lancet*, *396*(10258), 1204–1222. https://doi.org/10.1016/S0140-6736(20)30925-9

Wada, K., Yatsuya, H., Ouyang, P., Otsuka, R., Mitsuhashi, H., Takefuji, S., Matsushita, K., Sugiura, K., Hotta, Y., Toyoshima, H., & Tamakoshi, K. (2009). Self-reported medical history was generally accurate among Japanese workplace population. *Journal of Clinical Epidemiology*, *62*(3), 306–313. https://doi.org/10.1016/j.jclinepi.2008.04.006

Wager, S., & Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, *113*(523), 1228–1242. https://doi.org/10.1080/01621459.2017.1319839

Wijnen, B. F. M., Mosweu, I., Majoie, M., Ridsdale, L., de Kinderen, R. J. A., Evers, S., & McCrone, P. (2018). A comparison of the responsiveness of EQ-5D-5L and the QOLIE-31P and mapping of QOLIE-31P to EQ-5D-5L in epilepsy. *The European Journal of Health Economics*, *19*(6), 861–870. https://doi.org/10.1007/s10198-017-0928-0

World Health Organization (2007). *International classification of functioning, disability, and health: Children & youth version (ICF-CY)*. WHO Press.

Yadlowsky, S., Fleming, S., Shah, N., Brunskill, E., & Wager, S. (2021). Evaluating treatment prioritization rules via rank-weighted average treatment effects. *arXiv*. https://arxiv.org/abs/2111.07966

How to cite this article: de Vries, Y. A., Alonso, J., Chatterji, S., de Jonge, P., Lokkerbol, J., McGrath, J. J., Petukhova, M. V., Sampson, N. A., Sverdrup, E., Vigo, D. V., Wager, S., Al-Hamzawi, A., Borges, G., Bruffaerts, R., Bunting, B., Chardoul, S., Karam, E. G., Kiejna, A., Kovess-Masfety, V., … Zaslavsky, A. M. (2024). Proof-of-concept of a data-driven approach to estimate the associations of comorbid mental and physical disorders with global health-related disability. *International Journal of Methods in Psychiatric Research*, e2003. https://doi.org/10.1002/mpr.2003