NBER WORKING PAPER SERIES

THE EFFECTS OF MENTAL HEALTH INTERVENTIONS ON LABOR MARKET OUTCOMES IN LOW- AND MIDDLE-INCOME COUNTRIES

Crick Lund*
Kate Orkin*
Marc Witte*
John H. Walker
Thandi Davies
Johannes Haushofer
Sarah Murray
Judy Bass
Laura Murray
Wietse Tol
Vikram H. Patel

Working Paper 32423
http://www.nber.org/papers/w32423

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2024

The three authors listed first (Crick Lund, Kate Orkin, and Marc Witte) are jointly the first author. This study was funded by the Wellspring Philanthropic Fund. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

At least one co-author has disclosed additional relationships of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w32423

The Effects of Mental Health Interventions on Labor Market Outcomes in Low- and Middle-Income Countries
Crick Lund, Kate Orkin, Marc Witte, John H. Walker, Thandi Davies, Johannes Haushofer, Sarah Murray, Judy Bass, Laura Murray, Wietse Tol, and Vikram H. Patel
NBER Working Paper No. 32423
May 2024
JEL No. D9,I14,J24,O1

ABSTRACT

Mental health conditions are prevalent but rarely treated in low- and middle-income countries (LMICs). Little is known about how these conditions affect economic participation. This paper shows that treating mental health conditions substantially improves recipients' capacity to work in these contexts. First, we perform a systematic review and meta-analysis of all randomized controlled trials (RCTs) ever conducted that evaluate treatments for mental ill-health and measure economic outcomes in LMICs. On average, treating common mental disorders like depression with psychotherapy improves an aggregate of labor market outcomes made up of employment, time spent working, capacity to work and job search by 0.16 standard deviations. Treating severe mental disorders, like schizophrenia, improves the aggregate by 0.30 standard deviations, but effects are noisily estimated. Second, we build a new dataset, pooling all available microdata from RCTs using the most common trial design: studies of psychotherapy in LMICs that treated depression and measured days participants were unable to work in the past month. We observe comparable treatment effects on mental health and work outcomes in this sub-sample of highly similar studies. We also show evidence consistent with mental health being the mechanism through which psychotherapy improves work outcomes.

Crick Lund
Centre for Global Mental Health
Health Service and Population
Research Department
Institute of Psychiatry,
Psychology and Neuroscience
King's College London
De Crespigny Park
London SE5 8AF
United Kingdom
and University of Cape Town
crick.lund@kcl.ac.uk

Kate Orkin
University of Oxford
Blavatnik School of Government
Walton Street
Oxford, OX26GG
United Kingdom
and CEPR, IZA
kate.orkin@merton.ox.ac.uk

Marc Witte
VU Amsterdam
and IZA
m.j.witte@vu.nl

John H. Walker
University of Oxford
Nuffield College
New Road
Oxford OX1 1NF
United Kingdom
john.walker@nuffield.ox.ac.uk

Thandi Davies
Alan J. Flisher Centre for Public Mental Health
Department of Psychiatry & Mental Health
University of Cape Town
Rondebosch, Cape Town
South Africa
thandi.davies@uct.ac.za

Johannes Haushofer
Department of Economics
National University of Singapore
1 Arts Link AS2 #06-02
Singapore 117570
Singapore
and Stockholm University
and also NBER
johannes.haushofer@nus.edu.sg

Sarah Murray
Johns Hopkins University
Bloomberg School of Public Health
624 N. Broadway
Hampton House 863
Baltimore, MD 21205
sarah.murray@jhu.edu

Judy Bass
Johns Hopkins University
Bloomberg School of Public Health
624 N. Broadway, Room 861
Baltimore, MD 21205
jbass1@jhu.edu

Laura Murray
Johns Hopkins University
Bloomberg School of Public Health
624 N. Broadway
Baltimore, MD 21205
lmurra15@jhu.edu

Wietse Tol
Department of Public Health
University of Copenhagen
Øster Farimagsgade 5 opg. B
1353 København K
and John Hopkins University
and also VU University Amsterdam
wietse.tol@sund.ku.dk

Vikram H. Patel
The Pershing Square Professor of Global Health
Department of Global Health and Social Medicine
Harvard Medical School
vikram_patel@hms.harvard.edu

A data appendix is available at: http://www.nber.org/data-appendix/w32423

1 Introduction

Mental health conditions are highly prevalent: they are one of the ten major causes of disability globally, affecting 12% of the global population at any time (GBD 2019 Collaborators, 2022). In high-income countries (HICs), treating mental health conditions improves symptoms of mental illness and improves employment rates, reduces sick days, and enhances functioning at work, reducing output losses from mental ill-health (Chan et al., 2015, Nieuwenhuijsen et al., 2020, Salomonsson et al., 2018, van Duin et al., 2019).[1]

However, the effect of mental ill-health on economic outcomes in low- and middle-income countries (LMICs) remains poorly understood. Treatments developed in HICs and adapted to LMICs improve symptoms of mental illness in LMICs (Cuijpers et al., 2018, De Silva et al., 2013), but recent economic studies of depression treatments find conflicting effects on economic outcomes.[2] There is also little evidence on the economic effects of treatments for severe mental disorders, like schizophrenia. Evidence from HICs might have limited relevance for economic outcomes in LMICs: mental health treatments in LMICs are often modified to limit costs and labor market characteristics differ substantially.

Hence, this paper studies whether mental health treatments improve work and other economic outcomes in LMICs. We conduct the first systematic review of this literature, compiling a dataset of findings and study characteristics from all studies available online before April 2022 that 1) reported on a randomized controlled trial (RCT) testing a psychosocial or pharmacological (medication) intervention in an LMIC; 2) treated people diagnosed with a mental health disorder; and 3) measured any of a list of pre-specified economic outcomes.[3] We screened 15,031 papers and read 1,128 fully, yielding a sample of 39 eligible interventions. We record all effect sizes on economic outcomes and potential psychological and behavioral mechanisms for economic effects in these studies using standard meta-analysis methods.

The first part of the paper estimates the effect of psychosocial treatments for mental health conditions across all available studies. We conduct separate meta-analyses on our database of effect sizes for two theoretically distinct types of studies. First, studies of sixteen interventions test the effect of psychosocial interventions treating populations

[1]It is estimated global output losses from mental ill-health will total USD 7.3 trillion (in 2010 USD) over the period 2010-30, more than those associated with cardiovascular diseases (Bloom et al., 2011). Roughly two-thirds of these losses are attributed to lost income from the effects of mental ill-health on work outcomes, like employment, productivity and absenteeism.

[2]For example, Angelucci and Bennett (2024), Baranov et al. (2020), Barker et al. (2022), Bhat et al. (2022) and Haushofer et al. (2020).

[3]Our review protocol CRD42017058930 was registered with the Prospective Register of Systematic Reviews (PROSPERO) (`https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=58930`).

experiencing diagnosed common mental disorders (CMDs), depression and anxiety, with 15,444 trial participants.[4] These studies usually compare treatment to no treatment. Second, nine studies test "combination" treatments for populations experiencing severe mental disorders (SMDs), like schizophrenia, with 2,096 participants. These include both a psychosocial treatment and medication and are usually compared to receiving medication alone.[5] Under the assumption that the combination of treatments for SMDs has additive treatment effects, we recover the effect of receiving a psychosocial treatment in both populations. For each population, we separately estimate an average effect, across studies, on each of a set of similarly measured labor market outcomes and a labor market outcome aggregate, as well as other economic outcomes.We conservatively account for heterogeneity between study populations within these two groupings using Bayesian hierarchical models, alongside their traditional frequentist random-effects meta-analysis counterparts. Reported effects are from Bayesian models unless specified. Our search also captured studies testing treatments for post-traumatic stress disorder and substance use disorders and one study that tested pharmacological interventions for depression. However, we had too few studies of each of these types for substantive inference.[6]

Our first finding is that psychosocial interventions improve labor market outcomes in populations experiencing CMDs. On average, these interventions improve a pre-specified set of labor market outcomes comprised of measures of whether recipients are employed, their time spent working, their capacity to work, and their engagement in job search by 0.16 SD (95% CI: [0.03, 0.31]). Point estimates under our frequentist model are similar, but more precisely estimated, consistent with observed heterogeneity in effect sizes. These interventions improve a measure of work capacity: the number of days recipients are unable to work, by 0.08 SD (95% CI: [0.01, 0.17]). Effects on other sub-groupings of work outcomes are positive but not statistically significant. Six interventions reported an effect on the specific measure "days unable to work in the last 30 days", allowing us to report an aggregate effect size without standardization. Among these interventions, the average reduction was 1.42 days, or 13% relative to a control group mean of 11 days.[7]

Our second finding is that the effects on the labor market aggregate are even larger for mental health treatments in populations experiencing SMDs, averaging 0.30 SD. These effects are significant in the frequentist specification (95% CI: [0.06,0.54]) but only sig-

[4]At any one time, 3.4% of adults internationally have depression, 3.8% have anxiety and 1% have severe mental health conditions like schizophrenia (GBD 2019 Collaborators, 2022).

[5]There is now strong evidence that pharmacological treatments (medication) are effective in the treatment of SMDs, so few trials evaluate this intervention alone (Leucht et al., 2013).

[6]For completeness, findings are reported in Appendix Table A12.

[7]Here, we report the coefficient from the frequentist model. The Bayesian point estimate is less representative of the magnitude of the effect due to the small sample size, as outlined on page 19.

nificant at the 90% level (95% CI: [-0.05, 0.67]) under our more conservative Bayesian specification. Standard errors are larger under the Bayesian method as it better accounts for the impact of unexplained heterogeneity in study effects.

We run robustness checks on our main findings of effects on the aggregate of work-related outcomes. First, we control for study characteristics within a meta-regression framework. Next, we compare effects across subsamples disaggregated by potentially important dimensions of heterogeneity. Effects are robust across the income level of the country and the region in which the country falls. Effects are also robust to the use of different study designs. There is some evidence that treatment effects decay over time. For the CMD sample, effects are smaller in survey rounds conducted more than a year after treatment relative to those conducted sooner, but there is limited evidence of any effect in the SMD sample. Treatments for CMDs administered by clinicians and laypeople are both effective at improving work outcomes, although clinician-administered interventions have larger effects. Effects on work outcomes are, unsurprisingly, smaller for CMD treatments delivered to women in the post-natal period and living in countries with low female labor force participation than in other populations. Using a range of methods, we find no evidence of publication bias in our sample of studies .

Studies of psychosocial interventions targeting CMDs also measure non-work economic outcomes: education expenditure, assets, income, consumption, input expenditure, and subjective poverty. Our third finding is that treatment non-significantly improves an aggregate of these outcomes by 0.08 SD (95% CI: [-0.05, 0.21]), with larger effects on education outcomes and subjective poverty measures, and smaller effects on other outcomes. While results are significant at conventional levels under the frequentist specification, they are not under the Bayesian due to effect heterogeneity.

Fourth, we present evidence that improvements in symptoms of mental ill-health and functional impairment due to mental ill-health are mechanisms through which treatments improve work outcomes. Functional impairment occurs when an individual's health condition reduces their capacity to fulfil their normal social and work roles (Edlund et al., 2018). Psychosocial treatments for both CMDs and SMDs lead to large, generally statistically significant improvements in both symptoms of mental health disorders and measures of functional impairment. Larger positive effects on work outcomes reported for a given intervention are strongly correlated with larger improvements in symptoms of mental ill-health ($\beta = 0.70$) and in functioning ($\beta = 0.63$).

The second part of the paper replicates these findings using individual-level data and methods more standard in economics. We also examine if mental ill-health is the mechanism causing poorer work outcomes. Our second econometric strategy leverages the high frequency of studies of psychosocial treatments for depression identified by our review. We generate a unique new dataset, pooling and harmonizing microdata on 10,731 study participants from six studies that treat depression using psychosocial therapies, measure days participants are unable to work, and provide data publicly.[8]

First, we show our main findings are robust in individual-level data using standard economic methods within this subset of highly similar studies. We use OLS to test if psychosocial treatment improves mental health and ability to work in this sample and present our findings in traditional (un-standardized) units. Consistent with our meta-analysis, psychosocial treatment improves depression substantially and reduces days unable to work by 1.57 days per month (SE 0.86), or 24% relative to a control mean of 6.34 days.[9]

Second, we show substantial treatment effect heterogeneity by individual-level characteristics, analysis uniquely enabled by our individual-level data. Treatment yields larger improvements in mental health in groups with more severe baseline depression: treatment reduces depression by 0.12, 0.30 and 0.35 SD for those with mild, moderate and severe depression respectively and differences between groups are significant. We find suggestive evidence that treated individuals more severe depression also see larger reductions in days unable to work than those with mild depression, suggesting that mental health is a mechanism through which mental health treatment improves work outcomes. However, effects are noisily estimated. We find little evidence that treatment effects differ by age.

Finally, we estimate the individual-level elasticity of economic outcomes with respect to changes in depression symptoms. We instrument depression with assignment to a psychosocial intervention. We produce Two-Stage Least Squares (2SLS) estimates: a 0.22 SD reduction in depression symptoms (equivalent to that induced by the average treatment) is significantly associated with being unable to work 1.68 (26%) fewer days per month. In our view, this is the current best feasible test of the causal impact of mental ill-health associated with depression on work outcomes.

We contribute to the literature in economics studying the effects of mental health treatments on work outcomes. We present the first definitive evidence that treating mental health disorders improves work outcomes in LMICs, from the first meta-analysis of the economic impacts of mental health interventions in LMICs and the first analysis of microdata pooled from such interventions in any context.

The question of how mental health interventions affect work outcomes has been stud-

[8]For a subset we have repeated measurements, yielding 15,517 observations of 10,731 unique study participants. This approach mirrors Angrist and Meager (2023), Meager (2019) and Tan and Kremer (2020).

[9]This offers some evidence against violation of exchangeability (included studies plausibly arise from the same data generating process). We observe a consistent result among a more homogeneous subsample, suggesting that it is not extreme heterogeneous cases driving results.

ied in HICs. Meta-analyses find that therapies for mental ill-health improve employment rates, reduce sick days, and enhance functioning at work (Chan et al., 2015, Nieuwenhuijsen et al., 2020, Salomonsson et al., 2018, van Duin et al., 2019), although none use Bayesian methods. Individual economic studies find that improved drug availability for these conditions improves earnings and labor market participation (Biasi et al., 2021, Bütikofer et al., 2020). But treatment effects may differ by context. HIC studies focus on relatively expensive treatments administered by clinicians. In LMICs, therapies are simplified to work cross-culturally and are often administered by non-specialist workers.[10] Labor markets also differ: LMICs tend to have more informal work and weaker labor market regulations, like provisions for sick leave, which people may use to manage their conditions.

An emerging literature in development economics finds mixed effects of mental health treatments on economic outcomes in LMICs. Papers mainly study the effects of psychosocial interventions for depression.[11] Barker et al. (2022) find that a psychosocial intervention improves mental and physical health in rural Ghana after one to three months.[12] In contrast, Haushofer et al. (2020) find no impact of a psychosocial intervention on either mental health or economic outcomes 12 months post-treatment in rural Kenya. Baranov et al. (2020) and Bhat et al. (2022) find persistent improvements in mental health but no labor supply effects from psychosocial interventions among all or mostly female populations in South Asia multiple years after treatment.[13]

We contribute to this literature in three ways. First, we are able to aggregate evidence from all available clinical and economic studies in LMICs, allowing us to reconcile these conflicting findings. We show that on average, psychosocial interventions for depression and anxiety (CMD) have positive and statistically significant impacts on work outcomes, although these are much smaller in contexts with low female labor force participation and decay over time. Effects are remarkably similar to effects from high income countries.[14]

Second, we present the first analysis of microdata pooled from such interventions in any context. We show that meta-analytic findings and standard economic methods produce similar findings. We also provide unique evidence using individual-level data that improvements in mental health are likely to be one of the mechanisms linking mental

[10] 62% of the studies in our sample report on interventions employing lay-counsellors (Table H).

[11] Angelucci and Bennett (2024) find slightly negative effects of antidepressants for CMDs on hours worked and earnings but no other study examines pharmacological treatments for CMDs and measures economic outcomes so we do not provide meta-analytic evidence on these interventions.

[12] They study a general population and measure their mental health at baseline. We include estimates only on the subsample who meet clinical thresholds for having a mental disorder at baseline.

[13] They find effects on other economic outcomes. Baranov et al. (2020) find large effects on women's financial empowerment and parental investments. Bhat et al. (2022) find effects on people's beliefs about themselves, with implications for economic decision-making.

[14] See Section 5.2 for a detailed comparison of effect sizes.

health treatments to improvements in ability to work.

Finally, there is little evidence in either HICs or LMICs on the work effects of interventions to treat SMDs. We provide some of the first, albeit suggestive, evidence that such interventions may have large, economically important effects on work outcomes. However, using Bayesian methods highlights effect heterogeneity in the existing evidence base, indicating the need for more high-quality studies.

Our work is also related to the literature on the causal relationship between mental health and poverty. Whether mental ill-health causes poverty and the mechanisms underlying this relationship remain poorly understood (Ridley et al., 2020). A small quasi-experimental literature finds that mental ill-health causally increases poverty (Alloush, 2024, Stoop et al., 2019). However, these studies rely on strong assumptions about the dynamics of the long-run relationship between mental health and economic outcomes. Relative to these studies, we leverage RCT data to provide strong evidence that mental health worsens the ability to work, which may be the first link in a causal chain leading to lower earnings and poverty. Our findings paint a coherent picture. Studies with larger effects on mental health in our meta-analysis also have larger effects on work outcomes. In individual-level data, treatments have larger effects on work for those with worse mental health and the elasticity of days able to work with respect to mental health is of an economically meaningful magnitude. However, few studies in our sample collect data on earnings, limiting conclusions on later stages of the causal chain.

Our findings have immediate policy implications. Government expenditure on mental health treatment is meager, especially in LMICs.[15] Over 80% of people who need treatment for common mental disorders cannot access it, a substantially higher proportion than those who cannot access treatment for major physical health conditions (Chisholm et al., 2016). Analysing data on costs in our sample where available, we find interventions are of moderate cost, although there is substantial heterogeneity in costs by region and whether interventions use professional or lay-counsellors. Our work shows that treating mental ill health likely has important economic benefits in LMICs, alongside the known positive effects on mental health. Improving access to treatment presents an opportunity to significantly improve the lives of people living with these conditions in poor countries.

Sections 2 and 3 describe the systematic review process and studies captured in the search. Sections 4, 5 and 6 present the empirical strategy, results on economic outcomes and heterogeneity analysis. Section 7 analyzes potential mechanisms. Section 8 presents an analysis of the pooled microdata. Section 9 presents cost data.

[15]Median domestic expenditure on mental health is 2.1% of health expenditure globally, but only 1.05, 1.1 and 1.60% in low, lower-middle and upper-middle income countries (World Health Organization, 2021).

2 Systematic review procedure

A systematic review involves collecting information on and summarising all existing research on a topic. We followed guidelines from the Cochrane Collaboration for such reviews (Higgins et al., 2022). We review all studies published before April 2022 that 1) reported on an RCT testing a psychosocial or pharmacological intervention in an LMIC; 2) treated people diagnosed with a mental disorder; and 3) measured a pre-specified economic outcome. We used the Population, Intervention, Comparison and Outcome (PICO) method to pre-specify study inclusion criteria and minimize subjective inclusion decisions. We provide additional details of the search in Appendix A.

2.1 Inclusion and exclusion criteria

Population: We include only studies in low or middle-income countries, as defined by the World Bank in 2018.[16] We study the effect of treatment for a clinically diagnosed mental health condition. Study participants had to have been screened for a specific mental health condition and meet clinical criteria indicating they were currently living with the disorder. Screening could include an assessment on a self-reported psychological scale measuring symptoms of a mental health condition or a diagnostic assessment based on the Diagnostic and Statistical Manual of Mental Disorders (DSM) or International Classification of Diseases (ICD) criteria. Screening did not have to be done by a clinician. Studies where participants had a history of, but no current, mental illness were excluded. Participants had to be aged 14 years or older to focus on economically active populations.

Intervention: Interventions could include psychotherapy, psychological or psychosocial treatments ("psychosocial interventions"); pharmacological treatment; or interventions that combined psychosocial and pharmacological treatments. We compiled a list of widely used treatments, which we searched for using specific terms. In addition, we searched broadly for terms such as "mental health services" or "psychotherapy". Interventions could vary in dose, duration, mode of delivery, and setting.

Comparison: We initially screened both RCTs and non-randomized evaluations for inclusion in the meta-analysis. However, we found a sufficient number of studies which used an RCT for well-powered inference, so we restricted the sample to only include RCTs.

Outcomes: We searched for any study that measured employment, ability to work, labor force participation, productivity, job search, income, earnings, wages, assets, wealth, consumption, expenditure, calorie count, food security, savings, investments or input expenditure, technology adoption, expenditure on temptation goods, financial outcomes,

[16] The World Bank's classification criteria are outlined at https://datahelpdesk.worldbank.org/knowledgebase/articles/378834-how-does-the-world-bank-classify-countries.

health investment, education spending (children and own), income diversification, agricultural yields, revenue or profit from own employment, social networks or subjective poverty measures.[17] We present results by subgroups of outcomes as well as overall.

2.2 Search strategy

In the primary search, one author searched 21 databases (listed in Appendix A), including all major economics, social science, and clinical databases and repositories of working papers. We included studies published in any language if its abstract was in English.[18] We conducted forward and backward reference tracking of the citation lists of all included papers to identify other eligible studies. To capture trials in progress that might have reported results, we searched trial registries, contacted trial authors on NIH reporter, and contacted trial funders Grand Challenges Canada and the Abdul Latif Jameel Poverty Action Lab. More details are in Figure A1 and Appendix A. We placed no restrictions on the study date, with the earliest study published in 1994. The search ended in April 2022.[19]

3 Intervention and study characteristics

This section outlines the characteristics of the studies captured in our search, and the interventions they report on. Our search identifies 39 interventions evaluated in 35 different RCTs and reported on in 40 papers. Some trials test multiple interventions (e.g. Ran et al., 2003) and in some cases, more than one paper reports on the same intervention (e.g. Nadkarni et al., 2017b,a). Table 1 and Table A2 summarize the characteristics of the 39 interventions. Details of each intervention and study are presented in Table H.

3.1 Interventions, target conditions and control conditions

The search found studies that differed in which conditions they targeted, their choice of treatment and their selected control group. We face a decision about whether to aggregate across all treatments for all mental health conditions or analyze treatments for different conditions separately. Aggregating across more types of studies increases power, as we have more estimates of effect sizes, but also increases heterogeneity between studies being compared. To explain our choice of how to aggregate across interventions and studies into analysis groups, we briefly define some medical terms.

"Target conditions" are the type of mental health condition experienced by the population in which a study was conducted. We find interventions targeting four broad categories

[17] We collected studies measuring contraception use but decided not to include them in the analysis.

[18] We found three studies in Mandarin Chinese which were translated into English and included.

[19] Two articles were identified after the end of the search as part of forward and backward reference tracking of citation lists.

Table 1: Interventions in included randomized controlled trials

	All		Target CMD	Target SMD
	(1) Number of interventions	(2) Share of interventions	(3) N	(4) N
Panel A: All interventions	39	1.00	16	9
Intervention type and targeted condition combination *(mutually exclusive)*				
Psychosocial + common mental disorders (CMD)	14	0.36	14	0
Combination + common mental disorders (CMD)	2	0.05	2	0
Combination + severe mental disorders (SMD)	9	0.23	0	9
Psychosocial + post-traumatic stress disorders (PTSD)	5	0.13	0	0
Psychosocial + substance use disorders (SUD)	5	0.13	0	0
Other combinations	4	0.10	0	0
Panel B: Control conditions *(mutually exclusive)*				
Enhanced usual care	12	0.31	8	1
No treatment	14	0.36	5	0
Treatment as usual (pharmacological)	13	0.33	3	8
Panel C: Outcome measures *(not mutually exclusive)*				
Economic outcomes				
Work-related outcomes	34	0.87	14	9
In employment (dummy)	7	0.18	4	1
Time in work	7	0.18	3	2
Unable to work (dummy)	5	0.13	0	3
Days unable to work	13	0.33	9	1
Functioning at work	13	0.33	5	4
Job search	3	0.08	2	0
Non-work-related outcomes	14	0.36	6	2
Education	3	0.08	1	0
Assets	4	0.10	1	0
Income, consumption and input expenditure	7	0.18	4	0
Subjective poverty measures	4	0.10	2	1
Social networks	2	0.05	0	0
Other	3	0.08	2	1
Mental health outcomes				
Mental health condition symptoms	36	0.92	14	8
Suicide attempts or at risk of suicide	10	0.26	6	1
Relapse (dummy)	8	0.21	3	4
Recovery (dummy)	4	0.10	1	2
Rehospitalisation (dummy)	4	0.10	1	3
Diagnosed with mental health condition (dummy)	7	0.18	4	0
Qualitative assessment of mental health condition	7	0.18	3	2
Substance use	6	0.15	0	0
CMD symptoms	23	0.59	13	0
PTSD symptoms	6	0.15	0	0
SMD symptoms	6	0.15	0	4
Functioning and disability aggregate	31	0.79	12	8
Overall measures of functioning	25	0.64	10	7
Functioning in social interactions	5	0.13	1	3
Self-regulation	4	0.10	3	1
Self-esteem/self-efficacy	5	0.13	2	1
Cognition	5	0.13	2	1
Physical health	4	0.10	1	1

Notes: There are 39 interventions, 119 economic effect sizes and 196 mental health effect sizes. Outcome variable categories are not mutually exclusive (for example, interventions can measure employment and education outcomes), which is why percentages within categories can exceed 100%. Functioning at work measures are qualitative measures of functioning on the job. For example, the IDEAS scale is a rating from one of the interventions which evaluates a patient's disability in work on a 5 point scale. Self-regulation captures ability to set goals, control impulses and structure one's time. Self-esteem or self-efficacy is underlying beliefs about one's ability to carry out actions or achieve desired outcomes. Table A3 lists each unique measurement tool used for each group of outcomes.

of mental health conditions according to the World Health Organization International Classification of Disease (ICD-10) (World Health Organization, 2016): common mental disorders (CMDs, including anxiety and depression), severe mental disorders (SMDs, including schizophrenia and bipolar disorder), substance use disorders (SUDs) and post-traumatic stress disorder (PTSD). The categories are based on shared clinical presentation, functional disability, and treatment approaches.[20]

Treatments can be psychosocial (involving psychotherapy or training to provide education, guidance and support) or pharmacological (involving administration of medication), or combine both psychosocial and pharmacological treatments. We group different types of psychosocial treatments together because these have been found to have similar effects on mental health (Cuijpers et al., 2008, Cleary et al., 2008). The most frequently observed elements of these psychosocial treatments are cognitive behavioral therapy, psychoeducation, problem-solving therapy and interpersonal therapy (Tables H and A2). Pharmacological treatments are usually somewhat specific to a condition. For example, anti-depressants are regularly prescribed for depression while anti-psychotics are prescribed for severe mental disorders or substance use disorders.

Studies use different types of control condition. If a treatment is widely available and known to improve recipient outcomes, it is considered unethical to deprive a control group of the standard of care they could receive outside the trial. In these cases trials test if an experimental treatment performs better than this standard treatment. However, as often there is no widely available treatment provided for mental health conditions in LMICs, no treatment controls are common. Enhanced usual care (EUC) controls involve limited treatment, such as receiving information pamphlets, general health home visits, or referrals to a doctor. Treatment as usual (TAU): pharmacological controls are used when the public health system offers drug treatments.

Two distinct types of interventions and study designs appear in our search with sufficient frequency to conduct a meta-analysis. The first type is studies of 16 interventions targeting common mental disorders (depression and anxiety), described in Column 3, Table 1. The second type is studies of 9 interventions targeting severe mental disorders, which include schizophrenia and bipolar disorder and have more severe impacts on functioning and a longer duration than CMDs, described in Column 3, Table 1. We follow the medical literature which studies the effects of mental health treatments and estimate, for each outcome category, an average effect for interventions targeting CMDs and for interventions targeting SMDs. This also has the advantage that studies of these two intervention types

[20]We retrieve only studies including an economic outcome so the disorders retrieved by the search may not capture all disorders examined in mental health trials in LMICs.

use similar control strategies. We do not report one average across interventions targeting CMDs and interventions targeting SMDs because the interventions and target populations are too different to make aggregation meaningful.

Of the 16 CMD interventions, 14 provide psychosocial treatment only. They are usually compared to limited treatment: 5 interventions are compared to no treatment controls, 8 to enhanced usual care, usually a patient leaflet or one consultation with a doctor, and 1 to anti-depressants (see Table H). Two interventions treat CMDs with a combination of a psychosocial intervention and anti-depressants and are compared to control groups where participants only receive anti-depressants. These two trials thus also identify the effect of psychosocial interventions, under the assumption that treatment effects of psychosocial interventions and anti-depressants are additive. Throughout the paper, we aggregate these two types of studies together as capturing the effects of psychosocial interventions for CMDs. We show findings are robust to accounting for our pooling of studies with different types of control groups in Section 6.1.

All of the 9 SMD interventions are combination treatments including both a psychosocial intervention and a pharmacological treatment (an anti-psychotic). Of these, 8 are compared to a TAU pharmacological control, an anti-psychotic, available as standard treatment in the setting. Again, studies of these interventions identify the effect of psychosocial interventions for SMD under the assumption that treatment effects of psychosocial interventions and anti-psychotics are additive. One is compared to enhanced usual care. We aggregate these studies as capturing the effects of psychosocial interventions for SMDs. Again, we show robustness to pooling studies with different types of control groups in Section 6.1.

We observe studies of interventions for other conditions: post-traumatic stress (5 interventions) and substance abuse (5 interventions).[21] We also observe 4 interventions where the study is the only one of its design in the sample.[22] Results for these three groupings are reported in Appendix Table A12. There are too few studies per category and too much heterogeneity within categories of studies (using measures discussed in Section 4.2) to conduct meaningful analysis.

3.2 Economic outcomes

We pre-specified groupings of outcomes: employment, education, assets, income, consumption, financial behavior, health costs, subjective indicators of poverty, and social net-

[21]This includes Blattman et al. (2017), where all studied individuals are diagnosed with substance abuse problems.

[22]This includes Angelucci and Bennett (2024), who study effects of antidepressants.

works.[23] Table A3 lists each measurement tool used in each group of outcomes.

Panel C of Table 1 reports the frequency with which outcomes are reported. Most relevant are columns 3 and 4, which show the frequency with which effect sizes are reported for the two main intervention-target condition pairs we focus on (psychosocial or combination interventions targeting CMDs, column 3, and combination interventions targeting SMDs, column 4).

Employment or work-related outcomes are the most commonly reported category, reported on in 14 of 16 interventions targeting CMDs and all 9 interventions targeting SMDs. We had sufficient employment outcomes to disaggregate these further. The first two measures capture the intensive and extensive margins of employment. "In employment" captures if someone is employed. "Time in work" is the amount of time worked in hours or months over different recall periods. Being "unable to work" is known in HIC studies as "work-related disability", and indicates when a person is prevented from working by a health-related challenge. "Days unable to work" is similar to measures of disability days or sick leave in HIC studies, and measures the duration for which a person cannot work due to a health-related challenge.[24] "Functioning at work" measures are validated qualitative scales used in medical studies, where a clinician or participant rates the extent to which a participant is able to perform their normal role at work or whether their attendance or performance is impaired.[25] Table A6 provides wording for commonly used measures of functioning at work in our sample. Measures tend to relate to an individual's participation in paid and unpaid work both inside and outside the home. "Job search" captures measures of a person's engagement in, or intensity of, job search.

Relatively few studies of interventions capture non-work-related economic outcomes for our two main study types, so we present findings on these outcomes with caution.

3.3 Psychological and behavioral mechanisms

We also extracted all effects on psychological and behavioral pathways which might act as mechanisms for effects on economic outcomes and report on the frequency with which they appear in Panel C of Table 1. We coded any mental health outcome pre-specified as a primary outcome by the authors, as well as all outcomes which fell into one of 22

[23]We did not include broad measures of financial behavior, such as financial empowerment from Baranov et al. (2020), as these measures may not necessarily represent material economic outcomes. We instead included the sub-aggregate measures of impacts on our pre-specified outcomes where these were available.

[24]See Nieuwenhuijsen et al. (2020) and Salomonsson et al. (2018) for examples of HIC studies of these outcomes.

[25]For example, on the IDEAS scale, a clinician evaluates a patient's disability in work on a 5-point scale from no (0) to profound disability (4). A ranking of moderate disability indicates "Declining work performance, frequent absences, lack of concern about all this. Financial difficulties foreseen."

categories: suicide risk, re-hospitalisation, relapse, diagnosis with a mental health condition, psychiatric morbidity, depression, anxiety, CMD symptoms, alcohol misuse, drug misuse, schizophrenia, SMD symptoms, PTSD symptoms, disability, global functioning, executive functioning, cognitive functioning, social functioning, general health, general mental health, self-efficacy and self-esteem. The measures used to assess behavioral and psychological pathways are listed in Tables A4 and A5.

The bulk of outcome measures capture symptoms of mental ill health through psychological scales reported by the participant which measure the severity, frequency or duration of symptoms of personal distress. For example, for depression, these would include low mood, loss of interest or pleasure, sleep disturbance and difficulty concentrating. Some scales are used only for a particular disorder. Others, such as measures of depression and anxiety, are used across both conditions. Thirteen of 16 interventions targeting CMDs measure a scale of CMD symptoms; 4 of 9 interventions targeting SMDs measure a scale of SMD symptoms. Other studies measure whether a mental health condition is diagnosed or make a qualitative assessment of it. Some outcome measures are from hospital or clinician records, such as whether individuals made any suicide attempts or were at risk of suicide, or whether participants have relapsed, recovered, or been rehospitalized. We list the specific wording for commonly used scales in our sample in Tables A6-A9.

Studies of most interventions – 12 of 16 interventions targeting CMD; 8 of 9 targeting SMDs – report effects on a measure of functioning or disability. Functional impairment occurs when an individual's health condition reduces their capacity to fulfil their normal social and work roles (Edlund et al., 2018). Where only an overall functioning score is reported, we treat this effect as a psychological and behavioral mechanism. Some interventions capture functioning in specific domains of life, including performing daily tasks, personal care, family relationships, broader social interactions and work. Where these different domains are reported separately, we include effects for work-related functioning as economic outcomes, in the "functioning at work" category, and social interactions as psychological and behavioral mechanisms.

3.4 Other intervention characteristics

Table A2 shows other intervention-level characteristics. Most interventions for both CMD (11 of 16) and SMD (5 of 9) are restricted to adults above 17, with others target participants aged 14 and above. Within interventions targeting CMD, 11 target both genders and 5 target women. Within interventions targeting SMD, all target both genders.

There is a degree of spread over regions. For CMD interventions, 10 interventions are in South Asia, 3 in sub-Saharan Africa and the remainder in other regions. For SMD interventions, 6 interventions are in East Asia and the Pacific, with the rest spread over regions. CMD interventions have been mostly in lower-middle income countries, while SMD interventions are mostly in upper-middle income countries.

We include effects measured at any point after the beginning of treatment and often include multiple measurement points per intervention. The average intervention in our sample has 1.5 follow-up rounds. Measurement occurs 15.2 months after treatment, on average. Interventions have various combinations of follow-up periods. For CMD interventions, 10 interventions have one follow-up: 5 of these follow-up before 7 months and 4 between 7 and 12 months. Three interventions have two follow-ups, all before 12 months. Three interventions have three follow-ups, with different durations between follow-ups. SMD interventions in our sample usually have much shorter follow-ups: 7 of 9 interventions have only one round of follow-up before 7 months, while only 2 interventions have more than one follow-up. We average over these outcomes but disaggregate results by length of follow-up in Section 6.1.

4 Empirical strategy

4.1 Aggregating from raw effect sizes to inference datasets

We begin with a dataset of effect sizes of a single intervention on one outcome in one survey round and its associated confidence interval. A study often reports multiple estimates of the effect size of an intervention on an outcome, such as in robustness checks or repeated survey rounds. As is standard in meta-analyses, we average across the multiple effect size estimates (Higgins et al., 2022), with details outlined in Appendix D. Averaging processes allow for dependence between multiple effect sizes reported within a given study (Gleser and Olkin, 2009) and do not give studies with more effect sizes reported more weight. We calculate the standard error of the average effect size following Borenstein et al. (2009).

As discussed in Section 3.1, we follow the medical literature, which studies the effects of mental health treatments, and estimate, for each outcome category, an average effect for interventions targeting CMDs and for interventions targeting SMDs. This also has the advantage that studies of these two intervention types use similar control strategies. As discussed, we observe a few studies of interventions for post-traumatic stress and substance abuse, as well as single studies of pharmacological interventions against no-treatment controls, so present these with caution. In Table A16, we show that results are similar if we do not average across the multiple effect size estimates and instead perform inference on the individual effects reported by studies, while explicitly accounting for correlation between effects estimated for the same intervention.

We also face a decision about how much to pool effect sizes for different economic outcomes. Here, there is a trade-off between statistical power and interpretability. In our preferred specifications, we generate two aggregate outcomes upon which we perform inference. The first aggregates across all work-related outcomes, and the second across all non-work-related economic outcomes. We focus on work-related outcomes, where we have a moderately large sample and substantial power to detect effects, although there is heterogeneity in our estimates. Second, we report average effect sizes for groups of similarly measured outcomes. For example, we include variables capturing "Self-reported employment status" and "Engaged in work in the last week" in an "In employment" aggregate. While tests conducted on treatment effects at this level of aggregation are poorly powered, coefficients are more easily interpreted. Finally, we present findings averaging over work- and non-work-related economic outcomes, but view these estimates with caution given high levels of heterogeneity in the outcomes being measured.

4.2 Model

We expect study-level treatment effect heterogeneity in our sample of effect sizes. Even within groupings of the same intervention type, treating the same condition, we aggregate across effects from interventions with subtly different features in diverse contexts. We therefore follow the random-effects meta-analysis literature (DerSimonian and Laird, 1986). For each study k, we model the observed average treatment effect, $\{\hat{\tau}_k{}_{k=1}^{K}\}$ as the study-specific intervention effect τ_k and a sampling error term ϵ_k.

$$\hat{\tau}_k = \tau_k + \epsilon_k \tag{1}$$

This allows us to estimate the quantity of interest: the average latent treatment effect across studies and contexts, $\tau = E[\tau_k]$. We estimate Equation 1 using two approaches. First, we follow the frequentist meta-analysis literature, computing a weighted average $\hat{\tau}_{RE} = \sum_{k=1}^{K} \hat{\tau}_k \hat{\phi}_k / \sum_{k=1}^{K} \hat{\phi}_k$ to aggregate point estimates of intervention effects across studies. The weight $\hat{\phi}_k$ allocated to a study's estimate is set as the inverse of its variance, which minimizes the variance of the pooled estimate. This approach gives higher weight to more precise estimates, which tend to come from larger studies.

Second, we take a hierarchical Bayesian approach to model treatment effect heterogeneity explicitly and to allow the model to discount information from the marginal study where there is significant heterogeneity in studied effects. We implement the simple Rubin

(1981) model:

$$\hat{\tau}_k | \hat{se}_k, \sigma \sim N(\tau_k, \hat{se}_k^2) \quad \forall k$$
$$\tau_k | \tau, \sigma \sim N(\tau, \sigma^2) \quad \forall k$$

Where $\{\hat{\tau}_{k}{}_{k=1}^{K}\}, \{\hat{se}_{k=1}^{K}\}$ are the observed estimated effects and sampling errors, and setting $\sigma^2 = 0$ recovers the random-effects specification in Equation 1 (Gelman et al., 2009). We assume that the effect τ_k is drawn from a normal distribution of effects across sites governed by (τ, σ_τ^2). In our preferred specification, our priors on τ and σ are only weakly informative:

$$\tau \sim N(0, 1)$$
$$\sigma \sim \mathit{HC}(10)$$

Where N indicates the Normal distribution and HC the Half-Cauchy distribution. These choices allow us to concentrate our estimates of τ in the reasonable space of standard deviation effect sizes without assuming a sign on effects, and to enforce positive $\hat{\sigma}$, while allowing it to vary widely. As our priors are only weakly informative, we have a moderate number of studies, and we observe substantial effect size heterogeneity, we expect less power under the Bayesian approach relative to the frequentist approach. In estimating both models, we winsorize the top 1% of effect sizes to limit the impact of large outliers.

4.3 Quantifying heterogeneity

The approach outlined above offers two means of understanding the extent of and impact of study-level effect heterogeneity on our findings. First, we analyse the difference in point estimates and standard errors under the frequentist and Bayesian approaches. The Bayesian approach directly accounts for study-level heterogeneity in standard errors, while the frequentist approach does not (Higgins et al., 2009). In a zero-heterogeneity environment, the Bayesian hierarchical model pools effect sizes from all studies equally, using information from each study to shrink the estimates $\hat{\tau}_{k=1}^{K}$ towards the average τ, and precisely estimate τ. In contrast, in a high-heterogeneity environment, the hierarchical model will not pool information across sites, generating large credible intervals on τ. If the Bayesian models present larger credible intervals relative to their frequentist analogues, and produce evidence of high heterogeneity, then we should be suspicious that the frequentist confidence intervals are too tight because they fail to account for this heterogeneity.

Second, we estimate three measures of heterogeneity to show whether our effect size estimates are stable across studies, as well as across the pooling decisions we make about outcomes. First, our estimate of σ^2 is an absolute measure of heterogeneity in the distribution of effect sizes. If the Rubin (1981) model returns $\sigma^2 \approx 0$, then there is no unexplained heterogeneity between studies. However, this metric is difficult to interpret for non-zero values: any positive value indicates some degree heterogeneity but it is difficult to identify what constitutes a large amount of heterogneity (Vivalt, 2020). To account for this, secondly we report the average pooling metric, per Meager (2019):

$$\omega(\tau) = \frac{1}{K} \sum_{i=k}^{K} \frac{\hat{se}_k^2}{\hat{\sigma}^2 + \hat{se}_k^2}$$

This has a more obvious interpretation: $\omega(\tau_k) > 0.5$ implies that σ^2 is smaller than the sampling variation and that heterogeneity is "small" (Gelman and Hill, 2006). Third, we report the I^2, where $I^2 \approx \frac{\hat{\sigma}^2}{\hat{\sigma}^2 + \hat{se}_k^2}$, under both frequentist and Bayesian specifications. This measure of heterogeneity is closely related to the pooling factor, but has the opposite interpretation: a higher I^2 indicates a greater degree of heterogeneity. We extend consideration of heterogeneity with additional analysis in Section 6.1.

5 Results

In this section, we present our core meta-analysis findings on the impact of mental health interventions on economic outcomes. We report separate estimates of treatment effects within populations experiencing CMDs and SMDs under both frequentist and Bayesian estimation strategies, and at different levels of outcome pooling.

5.1 Work-related outcomes

Our core findings are summarized in Figure 1, which reports estimates of the latent average treatment effect of psychosocial interventions, $\hat{\tau}$, on a range of work-related outcomes for interventions targeting CMDs (Panel A) and targeting SMDs (Panel B). These are recovered from repeated meta-analyses estimating Equation 1 using both frequentist and Bayesian approaches. Within each panel, each boxplot summarizes the distribution of the estimate recovered from a meta-analysis on a different outcome grouping, arranged by row of the figure. For example, the "In employment" meta-analyses included effect sizes on all dummies that measure employment status. The "Work aggregate" represents the results of a meta-analysis pooling all of the effect sizes used to estimate the rows below. Box edges represent the bounds of a 50% confidence interval or credible intervals of the frequentist and Bayesian estimates, respectively, while whiskers represent their 95% analogues. In reporting findings, CI refers to 95% confidence or credible intervals unless otherwise specified. Details of each meta-analysis, including exact effect sizes and confidence intervals, sample sizes and heterogeneity measures, are reported in Table 2. Findings are reported in standard deviation (SD) units to allow for aggregation of different outcome measures across studies.[26]

5.1.1 Psychosocial interventions in populations experiencing CMDs

We find that psychosocial interventions significantly improve the "Work aggregate" among populations experiencing CMDs (Row 1 of Panel A of Figure 1 and of Table 2). Under the frequentist specification, we observe a moderately large effect of these interventions of 0.16 standard deviations (Column 1 of Table 2) with a 95% confidence interval of [0.05, 0.27] (Columns 2,3). This estimate is constructed from 36 observations of effect sizes (Column 4). These 36 observations are first aggregated to intervention-level average effect estimates of 14 interventions (Column 5) as described in Appendix G.2, before the final pooled estimate is constructed as a weighted average. There is evidence of some heterogeneity in reported effect sizes included in the weighted average, with a reported I^2 of 0.68 (Column 8).

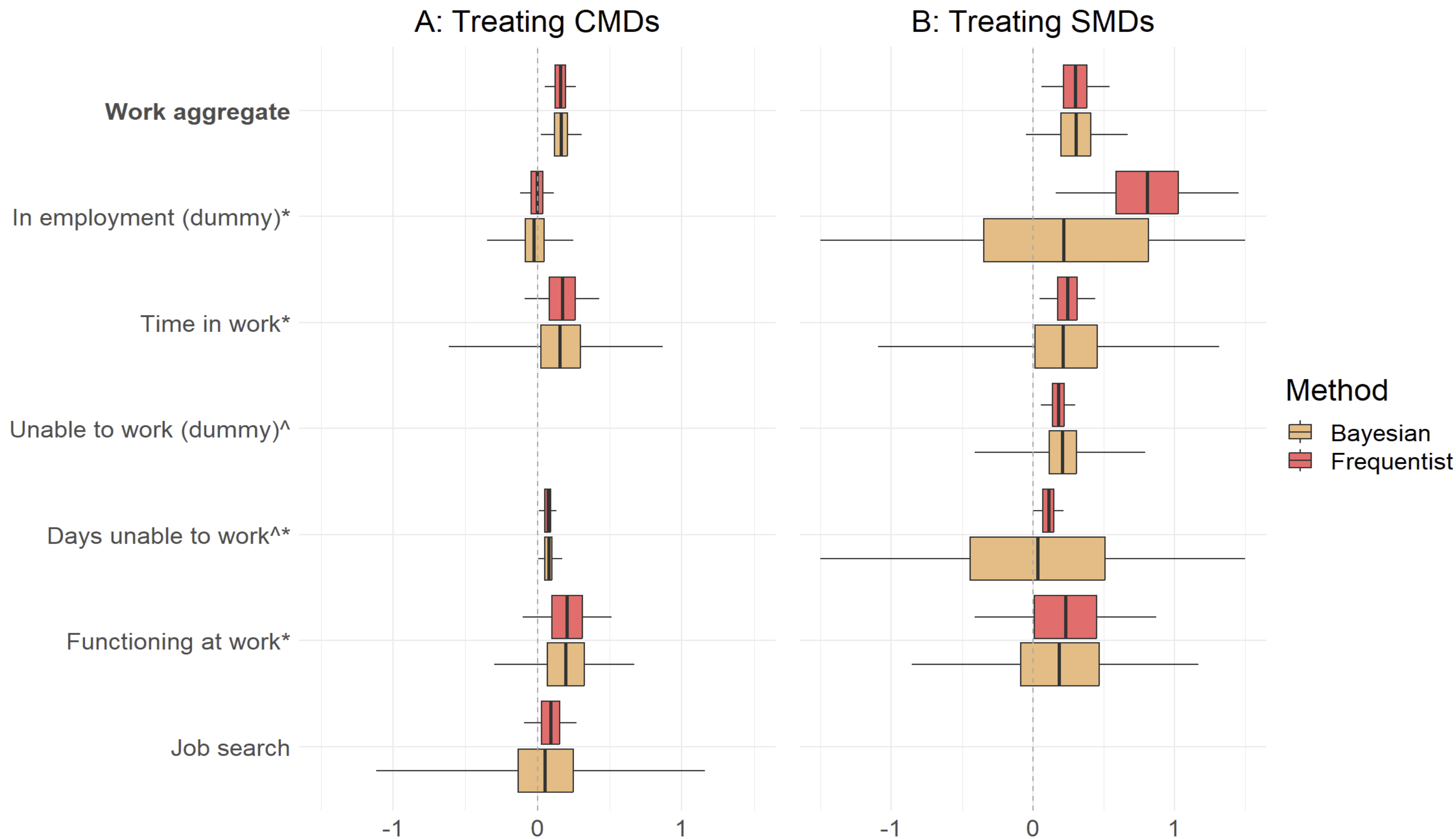
In the rows which follow, we show effects at a lower level of pooling of outcomes measured in different ways, on sub-aggregate measures. We find evidence of an effect on the most commonly measured sub-aggregate of similar outcomes, "Days unable to work", for which there are 17 effect sizes recorded for 9 interventions. Psychosocial interventions targeting CMDs improve "Days unable to work" by 0.07 SD, CI: [0.01,0.13]. For three other sub-aggregates – time in work, functioning at work and job search – effects are positive but not significant at conventional levels, likely due to small effective sample sizes.[27]

Our main findings are robust to adjusting for effect heterogeneity using the Bayesian specification. Under the Bayesian specification, the effect on the work aggregate is 0.16 SD, CI: [0.03,0.31], an identical point estimate and significant at conventional levels. The effect on "Days unable to work" is similarly robust. The Bayesian estimate is 0.08 SD, CI: [0.01,0.17]. Standard errors are not markedly different for this outcome under frequentist and Bayesian methods due to little observed heterogeneity in effect sizes (I^2=0.31 under the Bayesian framework).

[26] We do not adjust for multiple hypothesis testing. However, our main findings are on aggregate outcomes, limiting the number of tests performed.

[27] The reported frequentist I^2 estimates for these treatment effect estimates are likely substantially downward-biased by small sample bias, complicating direct inference on heterogeneity (von Hippel, 2015).

Figure 1: Positively-coded work-related economic impacts of mental health interventions: summary



Notes: Each boxplot represents the distribution of estimates from meta-analyses across study effect sizes captured under the category of the row title. The Work aggregate meta-analysis includes effect sizes from each of the other outcome groupings. There are two meta-analyses per outcome row, performed under both frequentist (red) and Bayesian (yellow) specifications outlined in Section 4.2. The (marginal posterior) maximum likelihood estimator estimate, $\hat{\tau}$, is represented by the line within the box. $\hat{\tau}$ is measured in standard deviations. Box edges represent 50% confidence intervals or shortest credible intervals of the frequentist and Bayesian estimates respectively, while whiskers represent their 95% analogues. Panel A presents estimates of the standard deviation effects of psychosocial interventions targeting common mental disorders, Panel B the effects of combined interventions targeting severe mental disorders. ^ indicates variables that have been reverse-coded such that higher values are positive. Whiskers are trimmed to [-1.5,1.5] if $\tau \notin [-1.5, 1.5]$. Untrimmed results are presented in Table 2. * indicates variables that have at least one trimmed estimate. The average measurement in our sample happens 15.2 months after intervention start. The procedure for aggregating multiple effect sizes from a given intervention is described in subsection 4.1. All individual effect sizes are winsorized at the 99^{th} percentile.

Table 2: Positively-coded work-related economic impacts of mental health interventions: details

	$\hat{\tau}$ (SD)	$\hat{\tau}$ CI		Sample size		Heterogeneity		
	(1) Average	(2) $2.5^{th}pctl.$	(3) $97.5^{th}pctl.$	(4) # obs.	(5) # intrv.	(6) σ	(7) $\omega(\tau)$	(8) I^2
Panel A: Treatment effects of psychosocial interventions on *common* mental disorders								
Frequentist specification								
Work aggregate	0.16***	0.05	0.27	36	14			0.68
In employment (dummy)	-0.00	-0.12	0.11	4	4			0.00
Time in work	0.17	-0.09	0.43	3	3			0.64
Days unable to work^	0.07**	0.01	0.13	17	9			0.13
Functioning at work	0.21	-0.10	0.52	8	5			0.86
Job search	0.09	-0.09	0.27	4	2			0.00
Work-related effects in original units								
Self-reported as employed (dummy)	0.00	-0.04	0.04	4	4			0.00
Days in last 30 days unable to work[1]	-1.42	-3.15	0.30	12	6			0.66
Bayesian specification								
Work aggregate	0.16**	0.03	0.31	36	14	0.21	0.32	0.8
In employment (dummy)	-0.02	-0.35	0.25	4	4	0.18	0.4	0.67
Time in work	0.16	-0.61	0.87	3	3	0.58	0.06	0.95
Days unable to work^	0.08**	0.01	0.17	17	9	0.06	0.72	0.31
Functioning at work	0.19	-0.3	0.67	8	5	0.48	0.08	0.93
Job search	0.05	-1.12	1.16	4	2	1.13	0.01	0.98
Work-related effects in original units								
Self-reported as employed (dummy)	-0.01	-0.15	0.1	4	4	0.08	0.38	0.71
Days in last 30 unable to work[1]	-0.67	-2.05	0.73	12	6	1.9	0.42	0.75
Panel B: Treatment effects of psychosocial interventions on *severe* mental disorders								
Frequentist specification								
Work aggregate	0.30**	0.06	0.54	20	9			0.76
In employment (dummy)	0.81**	0.16	1.45	1	1			1.00
Time in work	0.24**	0.05	0.44	4	2			0.07
Unable to work (dummy)^	0.18***	0.06	0.30	8	3			0.00
Days unable to work^	0.11**	0.00	0.22	1	1			1.00
Functioning at work	0.23	-0.41	0.87	6	4			0.89
Bayesian specification								
Work aggregate	0.30*	-0.05	0.67	20	9	0.45	0.21	0.88
In employment (dummy)	0.22	-1.66	1.85	1	1	5.37	0	NaN
Time in work	0.21	-1.09	1.32	4	2	1.28	0.02	0.98
Unable to work^	0.21	-0.41	0.8	8	3	0.4	0.13	0.9
Days unable to work (dummy)^	0.04	-1.69	1.71	1	1	4.92	0	NaN
Functioning at work	0.19	-0.86	1.17	6	4	1.12	0.07	0.97

Notes: The frequentist and Bayesian specifications are outlined in Section 4.2. In column (1), $\hat{\tau}$ is the estimate of the latent treatment effect in standard deviations. *, ** and *** represent significant at the 10, 5 and 1% levels respectively. In columns (2) and (3), $\hat{\tau}$ CI presents estimates of 95% confidence or credible interval under the frequentist or Bayesian specification, respectively. The Work aggregate meta-analysis includes effect sizes from each of the other outcome groupings below. Panel A presents estimates of the standard deviation effects of psychosocial interventions targeting common mental disorders and Panel B the effects of psychosocial interventions targeting severe mental disorders. ^ indicates variables that have been reverse-coded such that higher values indicate improvements in outcomes. The average measurement in our sample happens 15.2 months after intervention start. The procedure for aggregating multiple effect sizes from a given intervention is described in subsection 4.1. All individual effect sizes are winsorized at the 99^{th} percentile. [1] = based on the WHO Disability Assessment Schedule (WHODAS 2.0). The control mean for "self-reported as employed" in original units is 0.18 and for "days in last 30 days unable to work" in original units is 10.93.

We observe wider confidence intervals under the Bayesian specification relative to the frequentist specification across all other results. This is consistent with the frequentist approach failing to adequately account for study-level heterogeneity in effect sizes.[28]

How large are these effect sizes? Where possible, we report effects in "original units". A subset of interventions for CMD measured work outcomes on identical scales, allowing us to generate aggregate estimates across studies without standardising effects. We report these aggregate effects where at least two studies reported in the same units, with the caveat that this is a selected sample of interventions. Six interventions targeting CMDs report the identical WHODAS measure "Days that participants are unable to work in the last 30 days". These interventions reduced days unable to work by 1.42, CI: [-3.14,0.30] days, or an economically significant 13% reduction, under the frequentist specification, although this result has marginal statistical significance due to a small number of estimates.[29] We observe a null effect on an employment dummy.

5.1.2 Psychosocial interventions within populations experiencing SMDs

In Panel B of Figure 1, we show a large, positive average effect of psychosocial interventions targeting SMDs on the work aggregate under the frequentist specification of 0.30 SD, CI: [0.06,0.54], but are cautious in our interpretation of this effect. Relative to our estimate of the average treatment effects on CMDs, we observe substantially greater variance in our estimated treatment effect within populations experiencing SMDs, attributable to the substantially smaller effective sample size (9 interventions for SMDs compared to 14 for CMDs for which work outcomes are measured). Once we account for heterogeneity under the Bayesian specification, we find that the effect on the work aggregate is significant only at the 90% level (0.30 SD, CI: [-0.05,0.67]). This is consistent with evidence of substantial heterogeneity. The point estimate of the I^2 ranges from 0.76 to 0.88 in the frequentist and Bayesian models, respectively, and aligns with other measures. The Bayesian model factors in the uncertainty associated with this effect heterogeneity into standard errors, generating substantially wider confidence intervals.

[28]Across outcomes, the relative width of each of the credible intervals under the Bayesian specification is consistent with the heterogeneity measures and sample sizes we observe in Table 2. We observe that outcome groupings with the highest estimates of heterogeneity (Columns 6, 7 and 8; note a higher value in Column 7 indicates lower heterogeneity), and smallest sample sizes (Column 5), have the widest credible intervals. We present the distributions of these estimates in Appendix F.2. Our choice of a mean zero prior shrinks the coefficients on estimates generated from small samples towards zero. For example, the coefficient on "Job search" is 0.09 under the frequentist specification, but only 0.05 under the Bayesian specification, and the confidence interval is substantially wider. This is consistent with estimates of high heterogeneity (e.g. $\widehat{\omega(\tau)} = 0.01$).

[29]We report the frequentist, rather than Bayesian, point estimate because it provides a more representative indication of the magnitude of effects. Here, because there are so few intervention observations, the choice of standard normally distributed prior on effect sizes anchors the estimated effect size to be closer to zero.

Turning to the sub-aggregate measures, we find only limited evidence of specific treatment effects. Under the frequentist specification, we do observe positive and significant effects: treatment significantly improves “In employment” (0.81 SD, CI: [0.16,1.45]),“Time in work” (0.24 SD, CI: [0.05, 0.44]), “Unable to work (0.18 SD, CI: [0.06, 0.30]) and “Days unable to work” (0.11 SD, CI: [0.00,0.22]). However, these effects are non-significant under the Bayesian specification. First, the effective sample sizes are very small, ranging from 1-4 interventions. Our choice of mean zero normal priors shrinks the estimate of the latent mean generated from small samples toward zero. Second, there is evidence of very high heterogeneity in these subgroupings, so accounting for it substantially inflates standard errors. We view these effects as suggestive evidence that interventions targeting SMDs may show promise in improving work outcomes in LMICs, but more high-quality studies are needed before meaningful aggregated evidence can be produced.

5.1.3 Comparing effect sizes in populations experiencing CMDs and SMDs

Treatment effects in populations experiencing SMDs are almost twice as large as those in populations experiencing CMDs. However, one should be careful to not over-interpret this finding. First, the trials treating SMDs with psychosocial interventions mainly compare a psychosocial intervention combined with a pharmacological treatment to a control group that received a pharmacological treatment only. This identifies the effect of psychosocial interventions only under the assumption that treatment effects of psychosocial and pharmacological interventions are additive, ruling out plausible synergies between psychosocial and pharmacological interventions. Larger treatment effects in populations experiencing SMDs might be partially explicable by such synergies. In contrast, the trials treating CMDs with psychosocial interventions mainly compare psychosocial interventions to no treatment or enhanced usual care. Second, treatment selection may be endogenously determined. For example, physicians may opt for higher treatment intensity in populations experiencing a higher burden of mental ill-health. We might expect larger treatment effects in populations experiencing SMDs, if the illness is successfully treated. Finally, the trials of interventions targeting SMDs mainly have fairly short-term follow-ups, with 7 of 9 trials following up participants after less than 7 months. In contrast, the trials of interventions targeting CMDs often include at least one round of longer-term follow-up, in which effect sizes may be smaller. Future work must unpack these differences.

5.2 Comparisons to high income country effects

How do these results compare to those from meta-analyses in high-income countries (HIC)? In Table A1, we summarize effects from recent meta-analyses in HIC alongside findings from our study, referred to as “Lund et al., 2024” in the table. We report only HIC meta-analyses that overlap with at least one target condition or intervention and note some caution as studies have different inclusion criteria. Generally, effects are similar in magnitude to our study, consistent with findings that effects of treatment on mental health are similar across contexts (Patel et al., 2018, Singla et al., 2017).

Our estimates of the effects of psychosocial interventions in populations experiencing CMDs are very close to those reported in the HIC literature. At the aggregate level, our estimate of the effect of psychosocial interventions targeting CMDs is the same as that of Timbie et al. (2006) (0.16 SD) on a similar aggregate. At the sub-aggregate level, we observe slightly smaller effects on measures of absence from work (analogous in our study to "Days unable to work") relative to Nieuwenhuijsen et al. (2020), Salomonsson et al. (2018) and Finnes et al. (2019) (0.08 SD, compared to 0.15 SD, SD={0.12,0.17} and 0.17 SD respectively). We observe a non-significant effect on "Functioning at work" of 0.19 SD, which falls squarely between that of Nieuwenhuijsen et al. (2020) (0.05 SD), and Kamenov et al. (2017) (0.43 SD).

Comparisons between HIC results and our own are more difficult for effects in populations experiencing SMDs: no study reports a "Work aggregate" result within these settings, and our estimates on sub-aggregate measures have large standard errors. However, our broad finding that effect sizes on work outcomes tend to be larger among populations experiencing severe mental disorders compared to those experiencing common mental disorders also holds in HICs.

5.3 Non-work economic outcomes

In Tables A10 and A11, we present findings on non-work-related economic outcomes, including education expenditure; assets; income, consumption and input expenditure; and subjective poverty measures. While individual studies may find effects on these outcomes, the small number of studies and the lack of homogeneity across studies prevent us from constructing meaningful aggregate effects. Under the frequentist specification (Table A10), we find that psychosocial interventions improve the aggregate of non-work-related economic outcomes, driven by a large effect on education outcomes (from only a single intervention) and a marginally significant result on subjective poverty measures (from two interventions). However, under the Bayesian specification (Table A11), we see only weak evidence for an effect on the non-work aggregate and no evidence of an effect on education or subjective poverty once we account for heterogeneity.

6 Robustness to heterogeneity and publication bias

This section examines robustness of our core findings to accounting for observed measures of study heterogeneity. We then summarize results of tests for publication bias, concluding there is minimal evidence of publication bias in our sample of studies.

6.1 Heterogeneity

We extend our frequentist meta-analysis framework to explore robustness of our core findings on the "Work aggregate" to heterogeneity. This offers some insight into potential drivers of heterogeneity captured by the Bayesian model. As part of our search, we captured study- and intervention-level data that proxies for heterogeneity in measurement, interventions and context. We perform repeated multi-variate meta-analyses that include these (de-meaned) proxies as controls. The model is detailed in Appendix F.1.

In Table A16, we report findings from these repeated meta-regressions. The first column represents a "base model" in which we regress (non-aggregated) study effects on only a constant and intervention-level fixed effects. Each column that follows includes a collection of specified study-level covariates. In Panel A, we consider psychological interventions targeting CMDs. We find that the intercept, which represents the mean treatment effect on the work-related outcomes aggregate at the mean of the included (de-meaned) covariates, is roughly stable across the models, falling in the range [0.15,0.25] and remaining significant at the 95% level. This provides evidence against the observed measures of heterogeneity "driving" our core findings.

A χ^2 test of modifier relevance indicates whether the covariates in a model predict differential treatment effects. We observe that both a measure of the time between the intervention and outcome measurement, and an indicator for whether interventions are provided by a specialist are statistically significant modifiers. Error term variance, a proxy for "small sample effects" associated with publication bias, is non-significant. Variation in the I^2 statistic across these models is suggestive of how well the model, given its covariates, captures heterogeneity. While interventions costs are not a significant modifier, their inclusion reduces the I^2 from 0.77 to 0.56. The Cochran Q test of residual (unexplained) heterogeneity indicates that none of the observed characteristics capture all of the remaining heterogeneity in effects.

In the severe mental disorders model (Panel B), none of the measures of heterogeneity are explanatory, though there is weak evidence that measurement timing may be relevant (p=0.13). Standard errors on the estimate are large, consistent with the smaller sample size and with findings under the Bayesian model. As for the common mental disorder model, the test for residual heterogeneity is significant across models.

6.2 Sub-group analyses

In Table A17, we present effects on the "Work aggregate" for subgroups of studies where the analysis just presented suggests results may differ from the average effect. We re-estimate the frequentist model of Equation 1, partitioning our dataset into subsamples using indicators of intervention and study characteristics.[30] We urge caution in interpreting differences between groups of studies causally. For example, if studies with different control conditions have different average effects on work outcomes, this could be because of characteristics of the groups of studies other than the control condition. We report on heterogeneity by intervention cost in Section 9.

6.2.1 Measurement timing

For interventions targeting CMD, treatment effects are largest in the six months following exit (0.31 SD, CI: [0.06, 0.55]), before falling in the six months that follow (0.12 SD, CI: [0.01, 0.23]) and becoming undetectable more than a year after exit. In the SMD group, it is more difficult to draw conclusions about effect decay. Treatment effects are large in the six months following exit (0.29 SD, CI: [0.04, 0.54]). Only a single intervention reports on work outcomes after more than a year after exit, although this finds effects may be large and persistent. We caveat these findings by acknowledging that this evidence might be subject to selective reporting: follow-up surveys may only be conducted for interventions initially found to be effective.

6.2.2 Delivery type

For interventions targeting CMD, treatment effects on the work aggregate are larger for interventions delivered by specialists (0.27 SD, CI: [0.03, 0.52]) than laypeople (0.06 SD, CI: [0.01, 0.12]), although both are positive and significant.[31] However, we are cautious in over-interpretation of these findings. A number of studies in South Asia (where delivery by laypeople is common) are with female populations in settings where women have low labor force participation. It may be the labor market context rather than the delivery agent causing differences in effect sizes on the work aggregate.

In the SMD group, there is no difference in coefficient magnitude (0.36 SD), although only effects for specialist delivery are significant (CI: [0.09, 0.63]) due to a small sample of 3 interventions for non-specialist delivery. Trials with treatment by laypeople may differ

[30] As for the results for outcomes in "original units", discussed in footnote 29, we report the frequentist instead of the Bayesian specification results for this subsample analysis.

[31] Consistent with these findings, we also find that treatment effects on mental health and functioning outcomes are larger for interventions delivered by specialists than laypeople.

in multiple ways: interventions may be shorter, cheaper, or delivered at scale through government rather than research institutions. As the evidence base grows, research should unpack which dimensions are relevant for whether mental health treatments improve work outcomes.

6.2.3 Target population

Although we did not find that indicators of target population in general explained heterogeneity in the work aggregate, we do see one important difference by study context. In the right panel of Figure A2, we separate out studies targeting "standard" economically active populations from 1) populations of perinatal mothers and 2) female populations in contexts of low female labor force participation. Unsurprisingly, treatment of CMD among perinatal women induces smaller labor market effects than in the general population, consistent with women being engaged in child-rearing and potentially limiting labor market engagement. However, for this group, we observe effects on education investment (for the participants' children). Interventions in contexts of low female labor force participation have smaller, sometimes negative and very noisily estimated effects than in the general population. This might reflect the lack of labor market attachment of the participants in these samples.

6.2.4 Measurement of work outcomes

In Table A18, we show that the choice to aggregate across the different elements of our work-related outcomes aggregate to a work-related outcome aggregate is broadly reasonable i.e. we can aggregate over some heterogeneity in outcome definitions within this aggregate.[32] Regressing the work-related outcome groupings on indicators of particular sub-aggregate outcomes indicates that no sub-aggregate outcome is systematically different to the reference class of the most commonly studied outcome, "Days unable to work".

6.3 Publication bias

We summarize findings on publication bias in our entire study sample here and provide a detailed discussion in Appendix F.3. Conventional methods suggest little evidence of publication bias in reporting on economic outcomes. In Figure A3, we find little visual evidence of bunching of published results around the usual threshold significance level 5%, or asymmetry in the funnel plot. We formally assess funnel plot asymmetry via the Egger et al. (1997) regression test for small-study effects with standard errors clustered by study, with the Pustejovsky and Rodgers (2019) correction for false positives. We find no

[32] Conceptually, the findings here are identical to a column in Table A16, but here we additionally report coefficients on each of the included covariates.

evidence against the null hypothesis of no small-study effects, $\hat{\beta} = 0.01(0.19)$ (Table A20). We also formally model the impact of publication bias in our study sample, following Andrews and Kasy (2019).[33] We find no evidence that statistically significant findings are reported more often than null findings (Table A21).

7 Mental health effects: a likely mechanism

In this section, we present evidence that improvements in mental health and functioning associated with treatments are an important mechanism through which mental health interventions affect economic outcomes. We first present meta-analyses to estimate the effects of mental health interventions on psychological and behavioral mechanisms, in the same sample of interventions for which we estimated effects on economic outcomes. We then show that the economic effect sizes from included studies are highly correlated with mental health effect sizes. We extend these findings with evidence from microdata in Section 8.

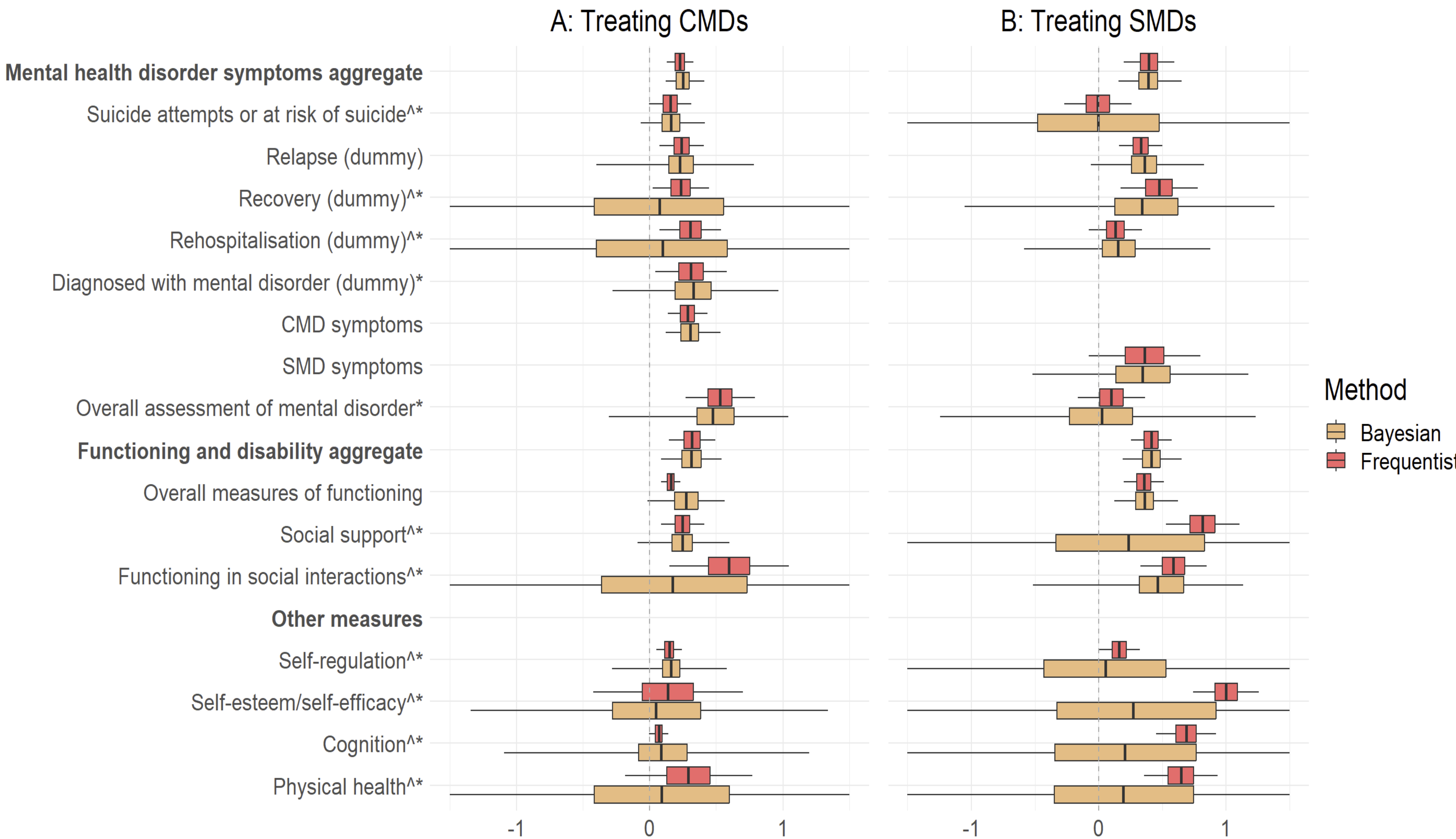
7.1 Treatment effects on psychological and behavioral mechanisms

We summarise our findings on the effects of treatments for mental ill-health on psychological and behavioral mechanisms in Figure 2. This figure reports the distributions of estimates of average treatment effects, $\hat{\tau}$, recovered from repeated meta-analyses estimating Equation 1 under both frequentist and Bayesian specifications. Detailed results from the frequentist and Bayesian specifications are reported in Table A13 and Table A14, respectively.

We organize our findings into three groupings: mental health disorder symptoms measures; functioning and disability measures; and other outcome measures. We report findings on two aggregates, one for mental health disorder symptoms, and a second for functioning and disability. Example wording of a representative sample of commonly reported mental health and functioning measures is provided in Tables A6-A9.

[33] This is valuable as the tests above are known to be underpowered with respect to some types of publication bias and in the presence of multiple reported effects within study (Rodgers and Pustejovsky, 2020).

Figure 2: Positively-coded psychological and behavioral impacts of mental health interventions: summary



Notes: Each boxplot represents the distribution of estimates from meta-analyses across study effect sizes captured under the category of the row title. The Mental health disorder symptoms aggregate is made up of effect sizes captured under each of the sub-aggregate headings between it and the Functioning and disability aggregate. The Functioning and disability aggregate is made up of "Overall measures of functioning", "Social support" and "Functioning in social interactions". There are two meta-analyses per outcome row, performed under both frequentist (red) and Bayesian (yellow) specifications outlined in Section 4.2. The (marginal posterior) maximum likelihood estimator estimate, $\hat{\tau}$, is represented by the line within the box. $\hat{\tau}$ is measured in standard deviations. Box edges represent 50% confidence intervals or shortest credible intervals of the frequentist and Bayesian estimates respectively, while whiskers represent their 95% analogues. Panel A presents estimates of the standard deviation effects of psychosocial interventions targeting common mental disorders, Panel B the effects of combined interventions targeting severe mental disorders. ^ indicates variables that have been reverse-coded such that higher values are positive. To allow visual inspection of small effect sizes, whiskers are trimmed to [-1.5,1.5] if $\tau \notin [-1.5, 1.5]$, while full untrimmed results are presented in Table A13 and Table A14. * indicates variables that have at least one trimmed estimate. The average measurement in our sample happens 15.2 months after intervention start. The procedure for aggregating multiple effect sizes from a given intervention is described in subsection 4.1. All individual effect sizes are winsorized at the 99^{th} percentile.

Across both populations experiencing CMDs and SMDs, we consistently observe improvements in mental health disorder symptoms and functioning and disability. Treatments improve the aggregate of mental health disorder symptoms (made up of each of the measures in the rows below this symptoms aggregate). Under the Bayesian approach, there are large effects both for psychosocial interventions targeting CMDs (0.25 SD, CI: [0.12,0.41) and SMDs (0.39 SD, CI: [0.16,0.65]) (Table A14). These interventions also substantially improve an aggregate of functioning and disability measures, made up of measures of overall functioning, social support and functioning in social interactions. Effects are broadly similar for psychosocial interventions targeting CMDs (0.32 SD, CI: [0.09,0.54]) and SMDs (0.41 SD, CI:[0.19,0.65]).

Turning to sub-aggregate indicators, we find treatment for CMDs reduces CMD symptoms (measured in 13 interventions, 0.31 SD, CI: [0.12,0.53]) and improves both overall measures of functioning and social support (0.28 and 0.25 SD, significant at the 10% level), even under the more conservative Bayesian specification. Treatment for SMDs marginally reduces relapse (measured in 4 of 9 interventions, 0.31 SD, CI: [0.12,0.53]) and improves overall measures of functioning (measured in 7 of 9 interventions, 0.36, CI: [0.12,0.62]).

For both CMD and SMD treatment, there are positive, significant effects on most other sub-aggregates in the frequentist specification (Table A13). But we observe substantial heterogeneity among the sub-aggregate indicators not mentioned above (Columns 6-8 of Tables A13-A14), often because only a subsample of studies collect each type of measure and/or there is heterogeneity in outcome measurement. This makes it difficult to draw conclusions on other sub-aggregate measures, as indicated by the large standard errors under the Bayesian specification compared to the frequentist approach in Figure 2. This highlights the importance of the Bayesian method.

7.1.1 Effect heterogeneity by measuring party

Figure A4 offers an insight into the impact of self-reporting biases on mental health effects. We disaggregate our mental health and functioning outcome findings by the party responsible for measurement, showing outcomes measured by clinicians vs. self-rated by patients. The results are quite similar for interventions targeting both CMD and SMD, with no evidence of a pattern of larger effects for self-reported outcomes. This suggests that findings are not driven by treatment-induced social desirability bias, such as by participants deducing from the content of therapy that an improvement in mental health is desirable to the experimenter and hence reporting improvements.

7.2 Correlation between psychological and behavioral mechanisms and economic outcomes

The sizes of effects of mental health interventions on economic outcomes and on potential psychological and behavioral mechanisms are strongly positively correlated. This indicates that psychological and behavioral factors may play an important role in mediating the effect of mental health interventions on labor market outcomes. Figure 3 reports the unconditional correlations between the aggregate treatment effects on behavioral and psychological pathways and the aggregate treatment effects on economic outcomes, measured at the intervention level.[34] The slope of the blue line corresponds to a β coefficient retrieved from a simple OLS regression, representing the "effect" of a 1 SD increase in behavioral and psychological pathways on economic outcomes. We include effects from all types of treatments we identify in our review, including treatments for substance use disorders and post-traumatic stress disorders, but results are even stronger when only treatments for CMD and SMD are considered.

In each case, there is a strong positive correlation between the effect size on a given economic outcome and the effect size for the potential mechanism. Mental health disorder treatment effects are highly correlated with work-related ($\beta = 0.70$) and non-work-related outcome treatment effects (β=0.38). This provides strong suggestive evidence that mental health is an important mechanism through which mental health interventions affect economic outcomes.

We observe similar correlations between functioning treatment effects and work-related ($\beta = 0.63$) and non-work-related ($\beta = 0.80$) outcomes. Mental ill-health and functioning tend to be highly correlated. Symptoms of mental ill-health, like depressed mood, often worsen functioning by reducing motivation or energy to conduct day-to-day activities or the desire to engage in social interactions. The consistency of correlations between economic outcomes and different measures of an individual's psychological health suggests this pattern is robust.

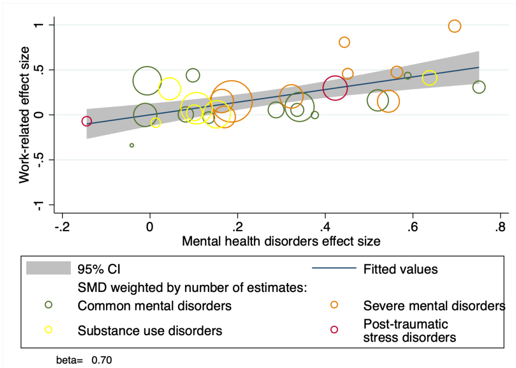
8 New evidence from aggregating microdata

In this section, we present a new dataset and analysis, where the units of observation are individuals within studies, rather than treatment effect estimates aggregated on a study-level. We pool microdata from the largest subset of included studies that measure the im-

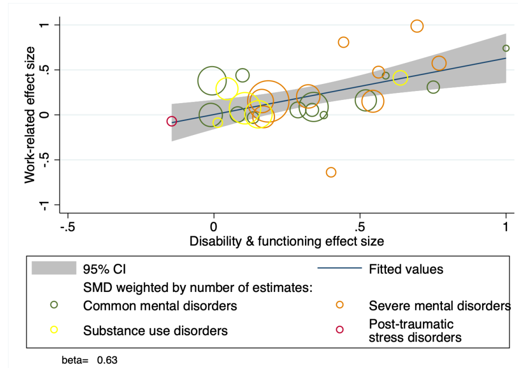
[34] Where a study reports functioning in different domains separately, we include effects for work-related functioning as economic outcomes and social interactions with psychological and behavioral mechanisms. We never include the overall functioning score including work-related functioning among mechanisms if we also include the work-related functioning score as an economic outcome.

Figure 3: Intervention-level correlations between economic outcomes and behavioral and psychological pathways at the individual level

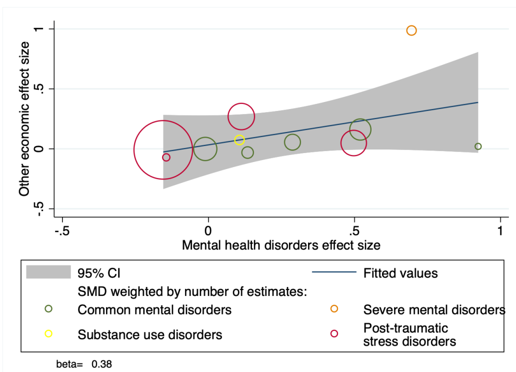
(a) Work-related effects and mental health disorders



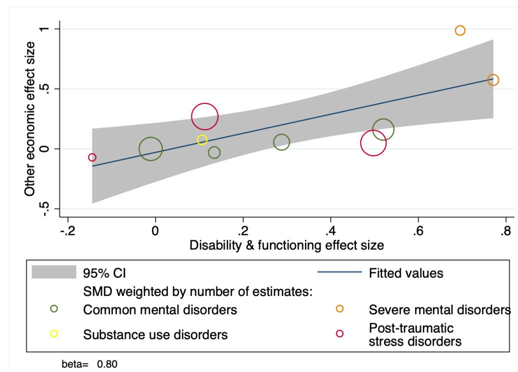
(b) Work-related effects and disability+functioning



(c) Other economic effects and mental health disorders



(d) Other economic effects and disability+functioning



These four scatterplots present average work outcome effect sizes (Hedges' g) by intervention, for a total of 39 interventions. The horizontal axis displays the average mental health disorder effect size (panels a and c) or the average disabiliy/functioning effect size (panels and d). The vertical axis shows the average economic effect for work-related outcomes (panels a and b) or other economic outcomes (panels c and d). The size of the circles indicates the sample size of the respective intervention. The aggregation of individual effect sizes works as described in Section 4.1. Individual effect sizes are winsorized within outcome type (work-related, other economic, mental health disorders, disability/functioning) at the 99^{th} percentile. The blue lines indicate the prediction of the economic effect size from a linear regression of the economic effect size on the behavioral/psychological effect size, along with the 95% confidence interval.

pact of psychosocial interventions in the same target group (populations suffering CMD), on the same economic outcome (days unable to work).[35] By only evaluating an outcome reported by each of the included studies that we can harmonize across studies, we avoid standardising the outcome measure (cf. Vivalt, 2020, Meager, 2019). Our approach balances the competing goals of comparing very similar interventions and capturing sufficient studies to have power to explore the following questions. First, as a test for the impacts of heterogeneity in our meta-analysis sample, do we observe the same findings in this subset of interventions that are "more similar" in terms of intervention characteristics? Second,

[35] Similarly, Tan and Kremer (2020), Meager (2019) and Banerjee et al. (2018) show results from pooling microdata.

is there heterogeneity in treatment effects by participant characteristics? Third, what is the relationship between mental ill-health and work outcomes? To the authors' knowledge, this is the first time that multi-study microdata has been pooled from mental health interventions in low and middle-income countries.

Our sample is made up of all 6 RCTs in our meta-analysis which meet three criteria: i) they evaluate a psychosocial intervention for depression relative to no treatment or enhanced usual care ii) they measure a depression screening questionnaire iii) they measure an outcome that can be harmonized to represent how many days a study participant is able to work per month. These RCTs study 9 interventions, all cognitive behavioral therapy or behavioral activation.[36] We expect relatively little heterogeneity in study population or treatment in this sample relative to the total meta-analysis sample. Variable construction is outlined in Section G.2.

8.1 Average treatment effects

We find that treatment reduces depression by 0.22 standard deviations (standard error 0.06) (Column 1, Table A24) in a simple regression of the combined depression measure on psychosocial treatment and study fixed effects. This is consistent with our finding in the meta-analysis sample that psychosocial treatments for CMDs reduce symptoms of common mental disorders by 0.31 standard deviations. The included depression questionnaires are the PHQ-9, BDI, DSM-IV and Kessler Scale, which are aggregated into an index of depression severity. [37] Findings in Columns (4)-(7) indicate that there is some heterogeneity in the standard deviation effects of interventions across measures of depression, which could reflect differences in interventions or measures.

We find that treatment reduces days unable to work by a significant (at the 10% level) 1.57 days per month (SE 0.85, Column 1, Table A25). Broadly, our results are highly consistent with those under the meta-analysis specification. We find that psychosocial treatments for CMDs improve days unable to work by a statistically significant 0.08 standard deviations ([CI: 0.01,0.17]) in the 9 interventions which measure variants of this variable e.g. with different recall periods (Bayesian specification, Table A10). It is also consistent with our meta-analysis finding that psychosocial treatments for CMDs improve days unable to work by 1.42 days (frequentist specification, Table A10). By contrast, we find no

[36]The six studies are Fuhr et al. (2019), Sikander et al. (2019), Baranov et al. (2020), Barker et al. (2022), Weobong et al. (2017), Patel et al. (2011). Recall we have 16 interventions targeting CMDs in the meta-analysis. We exclude 5 studies of interventions targeting depression because they do not have the days unable to work outcome (see Table G. We also exclude 2 studies of interventions which target depression with a psychosocial intervention and a pharmacological intervention compared to a pharmacological intervention alone, to reduce heterogeneity.

[37]Details on the variable construction can be found in Appendix G.2.

evidence of an effect on "Healthy days", reported by Baranov et al. (2020), though the effect is imprecisely estimated due to the small sample size (n=429, from a single study).

8.2 Sample characteristic heterogeneity

Two theoretically important recipient characteristics that could act as mediators of the impacts of treatments on mental ill-health and economic outcomes are reported at baseline by these studies. They are a categorical measure of depression severity, and age. We explore treatment effect heterogeneity by these characteristics by estimating an interacted regression model with median splits. Model details are provided in Appendix G.3. Unfortunately, we cannot test for heterogeneity by participant gender because the majority of the included studies have a sample of single-gender participants.[38]

We find strong evidence of heterogeneous treatment effects of psychosocial interventions on depression severity at endline, by baseline depression severity. More severe baseline depression predicts larger treatment effects. Treatment reduces depression more among people experiencing moderate (-0.17 SD, $p < 0.01$) and severe (-0.23 SD, $p < 0.01$) depression at baseline compared to the group experiencing mild depression (which has a treatment effect of -0.12 SD, $p < 0.10$). The differences between treatment effects for people with mild vs moderate, and mild vs severe, depression are highly statistically significant (footer of Table A24).[39]

We also find suggestive evidence that treatment for depression leads to stronger reductions in days unable to work (i.e. improvements in work outcomes) for those with higher baseline depression. Coefficients on the interaction terms between having moderate or severe (vs mild) depression and treatment are negative, consistent with effects on work outcomes being larger among populations experiencing more severe mental ill-health at baseline (Column 3, Table A25). However, coefficients are noisily estimated, preventing strong conclusions. Having higher baseline depression is associated with working between 1.05 and 1.42 days per month less (for moderate and severe groups, respectively) relative to participants not diagnosed with depression at baseline (Column 3, Table A25, $p < 0.01$ in both cases).

We see few heterogeneous effects by age, although being above median age is associated with marginally significantly higher depression (Column 2, Table A24) and significantly fewer days worked (Column 2, Table A25).

[38] Baranov et al. (2020), Fuhr et al. (2019) and Sikander et al. (2019) only include female participants.

[39] As might be expected, depression is also highly persistent: relative to people diagnosed with mild depression, those diagnosed with more moderate depression or severe depression experience worse depression at endline (0.24 SD and 0.43 SD, respectively, Column 3, Table A24).

8.3 Instrumenting depression with treatment assignment

Does mental ill-health worsen work outcomes in LMICs? We test this hypothesis by instrumenting changes in depression with random assignment of treatment to estimate the effect of a change in depression on days able to work via 2SLS. Psychosocial interventions used in this sample significantly improve mental health (Table A24). If these interventions have no further effects on economic outcomes that are not mediated by improvements in mental health, then we can estimate the impact that mental ill-health has on ability to work. Under this assumption, 2SLS identifies the "total effect" of mental health on work outcomes, across other potential mediators, such as functioning.

In our view, this is the *best feasible test* of the causal impact of mental ill-health on work outcomes. The (theoretical) first-best test would be to randomize mental health across individuals. This is impossible, but best approximated by random assignment to a treatment that is calibrated to improve mental health, and which has few effects on other potential mechanisms.

Our identifying assumption may be violated if either 1) treatments improve ability to work directly or 2) affect a non-mental health-based mechanism that is correlated with ability to work. To minimize the possibility of reverse causality – changes in economic outcomes causing changes in mental health, and not vice versa – we use data on mental health effects from midline surveys and data on economic effects from endline surveys where available, in four of the five trials.

We estimate the relationship between depression and ability to work via 2SLS, instrumenting depression with treatment.

$$MH_i = \gamma_0 + \gamma_1 T_{is} + S_s + \varepsilon_{is}, \tag{2}$$

where MH_i is participant i's depression outcome (measured at the earliest follow-up) and T_{is} is the indicator for whether i received the randomly allocated therapy treatment (as opposed to being in the control group) in study s. We then estimate the 2SLS equation:

$$y_i = \beta_0 + \beta_1 \widehat{MH}_i + S_s + \varepsilon_{is}, \tag{3}$$

where y_i is participant i's days able to work measure (measured at endline), $\widehat{MH}_i$ is her depression outcome instrumented by therapy treatment, and ε_{is} is a participant-study specific error term. Further model details are provided in Appendix G.

Treatment is a relevant instrument, as indicated by the F-test labeled "Weak identification": $F > 10$ for each instrument (Table 3). This is consistent with our finding that treatments significantly improve mental health in both the meta-analysis and microdata samples.

Table 3: Instrumenting the decrease in depression with random treatment allocation in the pooled sample

	Days unable to work		Healthy days	
	(1)	(2)	(3)	(4)
	OLS	IV	OLS	IV
Depression reduction	-3.282***	-7.615**	0.0762	0.434
	(0.239)	(3.058)	(0.363)	(1.041)
Constant	6.721***		26.27***	
	(0.324)		(0.362)	
Study FE	Yes	Yes	Yes	Yes
Control mean	6.43	6.43	26.16	26.16
Standard deviation	9.86	9.86	7.66	7.66
# of participants	10302	10302	429	429
Obs.	15088	15088	429	429
Studies	5	5	1	1
Underidentification		0.00		0.00
Weak identification		10.70		26.41

Notes: This table shows four different regression of the outcome variable on the depression scale, as well as study fixed effects, the endline round, and the number of months after treatment when the outcome was measured. The odd columns show the (endogenous) OLS regression of the outcome on the depression measure, while in the even columns the depression measure is instrumented by the treatment indicator. Columns 1-2 show the impact on days unable to work in the last month, columns 3-4 on healthy days per month. Outcome definitions and further details of the methodology are provided in Appendix G.2. Standard errors are in parentheses and clustered by the original clustering unit of each study. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Our preferred model aggregates across the five studies that measure “Days unable to work”. We estimate that a one standard deviation improvement in depression symptoms is associated with a 7.62-day decrease in “Days unable to work” (Column 2, Table 3), relative to a control mean of 6.43. Our preferred interpretation is to instead consider the effect of the mean intervention, rather than a hypothetical intervention that generates a one standard deviation improvement. The average intervention in our sample improves depression symptoms by 0.22 standard deviations. A back-of-the-envelope calculation indicates that at the mean, the improvement in days unable to work would be 1.68 (26%) fewer days unable to work.[40] While we cannot rule out violations of the identifying assumption, these

[40]We do not observe a statistically significant effect of depression on healthy days alone, potentially due to the relatively small sample size (n=429) from only a single study.

results indicate that treating depression in these contexts will cause an increase in work days. Moreover, they provide the strongest evidence to date that depression worsens work outcomes in LMICs by an economically meaningful magnitude.

9 Costs

Are the interventions considered by this paper sufficiently cheap to be a scalable or cost-effective method to improve work outcomes? We gained access to cost data on 20 of the 39 interventions covered by this paper either because it was publicly available, or by contacting authors. The studies for which cost data was, or was not, available are listed in section I. The sub-sample of studies for which we have cost data has roughly the same distribution of observable characteristics as the full study sample (Table A27).

Costs of treatments for mental ill-health are moderate on average, but highly heterogeneous. The median per-participant average cost of psychosocial interventions in populations experiencing CMDs and SMDs are reported in Table A26. The median cost of psychosocial interventions for CMDs was USD 105, while that for psychosocial interventions for SMDs was USD 180. However, the least and most costly interventions differ in cost by several orders of magnitude, even within intervention-condition combinations.[41]

Costs differ markedly across intervention regions. On average, interventions that took place in South Asia and East Asia & Pacific had substantially lower costs than those delivered in Europe & Central Asia and in Sub-Saharan Africa. This is consistent with interventions in Asia more commonly being administered by non-specialists.

In Table A17, we report effects on the work aggregate for the subsample of interventions costing less than 100 USD per participant compared to all interventions in our sample, and those costing more than this. For interventions targeting CMD, interventions costing less than 100 USD per participant display a non-significant average treatment effect of only 0.05 SD, while the whole sample average is 0.16 SD. This is consistent with earlier findings that delivery by specialists has larger effects on the work aggregate than delivery by non-specialists, although both are positive and statistically significant (see Section 6.2.2). However, we are cautious in over-interpretation of these findings, because many of the South Asian studies where delivery is by laypeople are with women in settings where women have low labor force participation. It may be the labor market context rather than the intervention cost causing differences in effect sizes.

In contrast, we do not find that more expensive interventions have larger treatment effects in populations experiencing severe mental disorders. If anything, the point estimate

[41]For example, the cheapest psychosocial treatment used among a population experiencing common mental disorders cost 1.43 USD per participant, while the most expensive cost 1226 USD per participant.

for the subsample of relatively "cheap" interventions is larger – 0.46 SD, relative to 0.30 SD – but based on only a single intervention.

These findings, suggest that on average, interventions may be cost-effective in improving economic outcomes. It may also be possible to reduce costs of interventions with only limited impacts on intervention effectiveness. Psychosocial interventions in clinical trials may require administration by doctors or psychiatrists, while this may not be necessary when treatments are scaled. Therapies were successfully administered by lay-counsellors in many of the considered interventions, with strong effects on work outcomes.

10 Conclusion

Our study presents findings from the first meta-analysis of the economic impacts of common treatments for mental ill-health in low- and middle-income countries. Psychosocial interventions generate substantial and economically meaningful improvements in work outcomes in populations experiencing CMDs, even after conservatively accounting for study-level heterogeneity using Bayesian methods. Our findings indicate that they likely have even larger effects in populations experiencing SMDs, though this finding is only significant at the 90% level under our most conservative approach. Impacts on mental health are highly correlated with impacts on economic outcomes.

We conduct an analysis of the economic impacts of mental health interventions in LMICs using microdata pooled from across trials that treat common mental disorders with psychosocial treatments, and measure days able to work. In this sample, we find that populations experiencing more severe mental ill-health at baseline benefit more from treatment in terms of improvements in mental health, and present suggestive evidence that effects on work outcomes are also larger. We instrument depression with random assignment of psychosocial treatment, and find that depression reduces "Days able to work". This is the cleanest feasible test of whether mental ill-health causes poorer work outcomes.

Taken together, our results suggest that work outcomes might be an important channel through which mental ill-health causes or exacerbates poverty. However, further work is needed on the mechanisms through which mental ill-health affects work-related outcomes. Our paper motivates future trials of mental health treatments powered to detect economic effects (see e.g. Angelucci and Bennett (2024), Barker et al. (2022), Blattman et al. (2017)). These could productively measure more potential psychological and behavioral mechanisms through which mental health treatment improves ability to work, such as increasing future orientation in economic decision-making or enabling more realistic appraisals of financial options rather than attention to threat (Haushofer and Fehr, 2014). Future work should also capture measures of labor supply, earnings and wealth, to enable further study of relationships between poor mental health and poverty. Multiple follow-up rounds would allow researchers to leverage the timing of changes in outcomes to explore causal pathways.

Further research could also develop and test multidimensional, integrated interventions targeting both poverty alleviation and mental health. This builds on findings that administering interventions targeting poverty and mental health alongside one another can be more effective than interventions on their own (Angelucci and Bennett, 2024, Blattman et al., 2017) and that the mentorship and handholding components of intensive livelihood programs are important elements of their success (Banerjee et al., 2018).

Our findings provide strong support to other calls to invest in mental health care as an important component of poverty alleviation (Patel et al., 2018). Policy-makers and international agencies focused on economic development have tended to overlook the importance of mental health. Existing, cost-effective interventions targeting mental health conditions both alleviate symptoms and improve recipients' ability to generate a livelihood. Further investment in mental health interventions is an urgent global priority.

References

ALLOUSH, M. (2024): "Income, Psychological Well-being, and the Dynamics of Poverty," *Economic Development and Cultural Change (forthcoming)*.

ANDREWS, I. AND M. KASY (2019): "Identification of and Correction for Publication Bias," *American Economic Review*, 109, 2766–94.

ANGELUCCI, M. AND D. BENNETT (2024): "The Economic Impact of Depression Treatment in India: Evidence from Community-based Provision of Pharmacotherapy," *American Economic Review*, 114, 169–198.

ANGRIST, N. AND R. MEAGER (2023): "Implementation Matters: Generalizing Treatment Effects in Education," *University of Oxford: Blavatnik School of Government Working Paper*, 23-802.

BANERJEE, A., D. KARLAN, D. OSEI, H. TRACHTMAN, AND C. UDRY (2018): "Unpacking a Multifaceted Program to Build Sustainable Income for the Very Poor," *National Bureau of Economic Research Working Paper*, 24271.

BARANOV, V., S. BHALOTRA, P. BIROLI, AND J. MASELKO (2020): "Maternal Depression, Women's Empowerment, and Parental Investment: Evidence from a Randomized Controlled Trial," *American Economic Review*, 110, 824–59.

BARKER, N., G. BRYAN, D. KARLAN, A. OFORI-ATTA, AND C. UDRY (2022): "Cognitive Behavioral Therapy among Ghana's Rural Poor Is Effective Regardless of Baseline Mental Distress," *American Economic Review: Insights*, 4, 527–45.

BHAT, B., J. DE QUIDT, J. HAUSHOFER, V. H. PATEL, G. RAO, F. SCHILBACH, AND P. L. P. VAUTREY (2022): "The Long-run Effects of Psychotherapy on Depression, Beliefs, and Economic Outcomes," *National Bureau of Economic Research Working Paper*, 30011.

BIASI, B., M. DAHL, AND P. MOSER (2021): "Career Effects of Mental Health," *National Bureau of Economic Research Working Paper*, 29031.

BLATTMAN, C., J. C. JAMISON, AND M. SHERIDAN (2017): "Reducing Crime and Violence: Experimental Evidence from Cognitive Behavioral Therapy in Liberia," *American Economic Review*, 107, 1165–1206.

BLOOM, D. E., E. CAFIERO, E. JANÉ-LLOPIS, S. ABRAHAMS-GESSEL, L. R. BLOOM, S. FATHIMA, A. B. FEIGL, T. GAZIANO, A. HAMANDI, M. MOWAFI, D. O'FARRELL, AND EMRE (2011): "The Global Economic Burden of Noncommunicable Diseases," Tech. rep., World Economic Forum, Geneva.

BORENSTEIN, M., H. COOPER, L. HEDGES, AND J. VALENTINE (2009): "Effect Sizes for Continuous Data," *The Handbook of Research Synthesis and Meta-analysis*, 2, 221–235.

BORENSTEIN, M., L. V. HEDGES, J. P. HIGGINS, AND H. R. ROTHSTEIN (2011): *Introduction to Meta-analysis*, John Wiley & Sons.

BÜTIKOFER, A., C. J. CRONIN, AND M. M. SKIRA (2020): "Employment Effects of Healthcare Policy: Evidence from the 2007 FDA Black Box Warning on Antidepressants," *Journal of Health Economics*, 73, 102348.

CHAN, J. Y., H. W. HIRAI, AND K. K. TSOI (2015): "Computer-assisted Cognitive Remediation and Employment/Productivity Outcomes in Severe Mental Illness: A Meta-analysis of Prospective Controlled Trial," *Journal of Psychiatric Research*, 68, 293–300.

CHISHOLM, D., K. SWEENY, P. SHEEHAN, B. RASMUSSEN, F. SMIT, P. CUIJPERS, AND S. SAXENA (2016): "Scaling-up Treatment of Depression and Anxiety: A Global Return on Investment Analysis," *The Lancet Psychiatry*, 3, 415 – 424.

CLEARY, M., G. E. HUNT, S. L. MATHESON, N. SIEGFRIED, AND G. WALTER (2008): "Psychosocial Interventions for People with Both Severe Mental Illness and Substance Misuse," *Cochrane Database of Systematic Reviews*, 12.

COHEN, J. (1988): *Statistical Power Analysis for the Behavioral Sciences, 2nd edn*, Erlbaum Associates, Hillsdale.

CUIJPERS, P., E. KARYOTAKI, M. REIJNDERS, M. PURGATO, AND C. BARBUI (2018): "Psychotherapies for Depression in Low- and Middle-income Countries: A Meta-analysis," *World Psychiatry*, 17, 90–101.

CUIJPERS, P., A. VAN STRATEN, G. ANDERSSON, AND P. VAN OPPEN (2008): "Psychotherapy for Depression in Adults: A Meta-analysis of Comparative Outcome Studies," *Journal of Consulting and Clinical Psychology*, 76, 909.

DE SILVA, M. J., S. COOPER, H. L. LI, C. LUND, AND V. PATEL (2013): "Effect of Psychosocial Interventions on Social Functioning in Depression and Schizophrenia: Meta-analysis," *British Journal of Psychiatry*, 202, 253–260.

DERSIMONIAN, R. AND N. LAIRD (1986): "Meta-analysis in Clinical Trials," *Controlled Clinical Trials*, 7, 177–188.

EDLUND, M. J., J. WANG, K. G. BROWN, V. L. FORMAN-HOFFMAN, S. L. CALVIN, S. L. HEDDEN, AND J. BOSE (2018): "Which Mental Disorders Are Associated With the Greatest Impairment in Functioning?" *Social Psychiatry and Psychiatric Epidemiology*, 53, 1265–1276.

EGGER, M., G. D. SMITH, M. SCHNEIDER, AND C. MINDER (1997): "Bias in Meta-analysis Detected by a Simple, Graphical Test," *British Medical Journal*, 315, 629–634.

FINNES, A., P. ENEBRINK, A. GHADERI, J. DAHL, A. NAGER, AND L.-G. ÖST (2019): "Psychological Treatments for Return to Work in Individuals on Sickness Absence Due to Common Mental Disorders or Musculoskeletal Disorders: A Systematic Review and Meta-analysis of Randomized Controlled Trials," *International Archives of Occupational and Environmental Health*, 92, 273–293.

FUHR, D. C., B. WEOBONG, A. LAZARUS, F. VANOBBERGHEN, H. A. WEISS, D. R. SINGLA, H. TABANA, E. AFONSO, A. DE SA, E. D'SOUZA, ET AL. (2019): "Delivering the Thinking Healthy Programme for Perinatal Depression Through Peers: An Individually Randomised Controlled Trial in India," *The Lancet Psychiatry*, 6, 115–127.

GBD 2019 COLLABORATORS (2022): "Global, Regional, and National Burden of 12 Mental Disorders in 204 Countries and Territories, 1990–2019: A Systematic Analysis for the Global Burden of Disease Study 2019," *The Lancet Psychiatry*, 9, 137–150.

GELMAN, A., J. B. CARLIN, H. S. STERN, AND D. B. RUBIN (2009): *Bayesian Data Analysis (3nd ed.)*, Abingdon: Taylor and Francis.

GELMAN, A. AND J. HILL (2006): *Data Analysis Using Regression and Multilevel/Hierarchical Models*, Analytical Methods for Social Research, Cambridge University Press.

GLESER, L. J. AND I. OLKIN (2009): "Stochastically Dependent Effect Sizes," in *The Handbook of Research Synthesis and Meta-analysis*, ed. by H. Cooper, L. B. Hedges, and J. Valentine, Russell Sage Foundation, 357–376.

HAUSHOFER, J. AND E. FEHR (2014): "On the Psychology of Poverty," *Science*, 344, 862–867.

HAUSHOFER, J., R. MUDIDA, AND J. P. SHAPIRO (2020): "The Comparative Impact of Cash Transfers and a Psychotherapy Program on Psychological and Economic Well-being," *National Bureau of Economic Research Working Paper*, 28106.

HEDGES, L. V. (1981): "Distribution Theory for Glass's Estimator of Effect Size and Related Estimators," *Journal of Educational Statistics*, 6, 107–128.

HIGGINS, J., S. THOMPSON, AND D. SPIEGELHALTER (2009): "A Re-evaluation of Random Effects Meta-analysis," *Journal of the Royal Statistical Society. Series A, (Statistics in Society)*, 172, 137–159.

HIGGINS, J. P. T., J. THOMAS, J. CHANDLER, M. CUMPSTON, T. LI, M. PAGE, AND V. A. WELCH (2022): *Cochrane Handbook for Systematic Reviews of Interventions*, Wiley-Blackwell.

Jackson, D., R. Riley, and I. R. White (2011): "Multivariate Meta-analysis: Potential and Promise," *Statistics in Medicine*, 30, 2481–2498.

Kamenov, K., C. Twomey, M. Cabello, A. M. Prina, and J. L. Ayuso-Mateos (2017): "The Efficacy of Psychotherapy, Pharmacotherapy, and their Combination on Functioning and Quality of Life in Depression: A Meta-analysis," *Psychological Medicine*, 47, 414–425.

Leucht, S., A. Cipriani, L. Spineli, D. Mavridis, D. Örey, F. Richter, M. Samara, C. Barbui, R. R. Engel, J. R. Geddes, et al. (2013): "Comparative Efficacy and Tolerability of 15 Antipsychotic Drugs in Schizophrenia: A Multiple-treatments Meta-analysis," *The Lancet*, 382, 951–962.

Meager, R. (2019): "Understanding the Average Impact of Microcredit Expansions: A Bayesian Hierarchical Analysis of Seven Randomized Experiments," *American Economic Journal: Applied Economics*, 11, 57–91.

Nadkarni, A., H. A. Weiss, B. Weobong, D. McDaid, D. R. Singla, A.-L. Park, B. Bhat, B. Katti, J. McCambridge, P. Murthy, M. King, G. T. Wilson, B. Kirkwood, C. G. Fairburn, R. Velleman, and V. Patel (2017a): "Sustained Effectiveness and Cost-effectiveness of Counselling for Alcohol Problems: A Brief Psychological Treatment for Harmful Drinking in Men, Delivered by Lay Counsellors in Primary Care - 12-month Follow-up of a Randomised Controlled Trial," *PLOS Medicine*, 14, 1–20.

Nadkarni, A., B. Weobong, H. A. Weiss, J. McCambridge, B. Bhat, B. Katti, P. Murthy, M. King, D. McDaid, A.-L. Park, G. T. Wilson, B. Kirkwood, C. G. Fairburn, R. Velleman, and V. Patel (2017b): "Counselling for Alcohol Problems (CAP): A Lay Counsellor-delivered Brief Psychological Treatment for Harmful Drinking in Men, in Primary Care in India: A Randomised Controlled Trial," *The Lancet*, 389, 186–195.

Nieuwenhuijsen, K., J. H. Verbeek, A. Neumeyer-Gromen, A. C. Verhoeven, U. Bültmann, and B. Faber (2020): "Interventions to Improve Return to Work in Depressed People," *Cochrane Database of Systematic Reviews*, 1–160.

Nigatu, Y. T., Y. Liu, M. Uppal, S. McKinney, S. Rao, K. Gillis, and J. Wang (2016): "Interventions for Enhancing Return to Work in Individuals with a Common Mental Illness: Systematic Review and Meta-analysis of Randomized Controlled Trials," *Psychological Medicine*, 46, 3263–3274.

Patel, V., S. Saxena, C. Lund, G. Thornicroft, F. Baingana, P. Bolton, D. Chisholm, P. Y. Collins, J. L. Cooper, J. Eaton, et al. (2018): "The Lancet Commission on Global Mental Health and Sustainable Development," *The Lancet*, 392, 1553–1598.

Patel, V., H. A. Weiss, N. Chowdhary, S. Naik, S. Pednekar, S. Chatterjee, B. Bhat, R. Araya, M. King, G. Simon, H. Verdeli, and B. R. Kirkwood (2011): "Lay Health Worker-Led Intervention for Depressive and Anxiety Disorders in India: Impact on Clinical and Disability Outcomes Over 12 Months," *British Journal of Psychiatry*, 199, 459–466.

Pustejovsky, J. E. and M. A. Rodgers (2019): "Testing for Funnel Plot Asymmetry of Standardized Mean Differences," *Research Synthesis Methods*, 10, 57–71.

Ran, M.-S., M.-Z. Xiang, C. L.-W. Chan, J. Leff, P. Simpson, M.-S. Huang, Y.-H. Shan, and S.-G. Li (2003): "Effectiveness of Psychoeducational Intervention for Rural Chinese Families Experiencing Schizophrenia," *Social Psychiatry and Psychiatric Epidemiology*, 38, 69–75.

Ridley, M., G. Rao, F. Schilbach, and V. Patel (2020): "Poverty, Depression, and Anxiety: Causal Evidence and Mechanisms," *Science*, 370, eaay0214.

Rodgers, M. and J. Pustejovsky (2020): "Evaluating Meta-analytic Methods to Detect Selective Reporting in the Presence of Dependent Effect Sizes," *Psychological Methods*, 26(2), 141–160.

RUBIN, D. B. (1981): “Estimation in Parallel Randomized Experiments,” *Journal of Educational Statistics*, 6, 377–401.

SALOMONSSON, S., E. HEDMAN-LAGERLÖF, AND L.-G. ÖST (2018): “Sickness Absence: A Systematic Review and Meta-Analysis of Psychological Treatments for Individuals on Sick Leave Due to Common Mental Disorders,” *Psychological Medicine*, 48, 1954–1965.

SIKANDER, S., I. AHMAD, N. ATIF, A. ZAIDI, F. VANOBBERGHEN, H. A. WEISS, A. NISAR, H. TABANA, Q. U. AIN, A. BIBI, ET AL. (2019): “Delivering the Thinking Healthy Programme for Perinatal Depression Through Volunteer Peers: A Cluster Randomised Controlled Trial in Pakistan,” *The Lancet Psychiatry*, 6, 128–139.

SINGLA, D. R., B. A. KOHRT, L. K. MURRAY, A. ANAND, B. F. CHORPITA, AND V. PATEL (2017): “Psychological Treatments for the World: Lessons from Low- and Middle-income Countries,” *Annual Review of Clinical Psychology*, 13, 149–181.

STOOP, N., M. LEIBBRANDT, AND R. ZIZZAMIA (2019): “Exploring Psychological Well-being and Poverty Dynamics in South-Africa: Evidence from NIDS waves 1-5,” *Southern African Labour and Development Research Unit Working Paper*, 240.

TAN, B. AND M. KREMER (2020): “Water Treatment and Child Survival: A Meta-analysis,” *National Bureau of Economic Research Working Paper*, 30835.

TIMBIE, J., M. HORVITZ-LENNON, R. G. FRANK, AND S.-L. T. NORMAND (2006): “A Meta-analysis of Labor Supply Effects of Interventions for Major Depressive Disorder,” *Psychiatric Services*, 57, 212–218.

VAN DUIN, D., L. DE WINTER, M. OUD, H. KROON, W. VELING, AND J. VAN WEEGHEL (2019): “The Effect of Rehabilitation Combined with Cognitive Remediation on Functioning in Persons with Severe Mental Illness: Systematic Review and Meta-analysis,” *Psychological Medicine*, 49, 1414–1425.

VIVALT, E. (2020): “How Much Can We Generalize from Impact Evaluations?” *Journal of the European Economic Association*, 18, 3045–3089.

VON HIPPEL, P. (2015): “The Heterogeneity Statistic I Can Be Biased in Small Meta-analyses,” *BioMed Central Medical Research Methodology*, 15, 35.

WEOBONG, B., H. A. WEISS, D. MCDAID, D. R. SINGLA, S. D. HOLLON, A. NADKARNI, A.-L. PARK, B. BHAT, B. KATTI, A. ANAND, S. DIMIDJIAN, R. ARAYA, M. KING, L. VIJAYAKUMAR, G. T. WILSON, R. VELLEMAN, B. R. KIRKWOOD, C. G. FAIRBURN, AND V. PATEL (2017): “Sustained Effectiveness and Cost-effectiveness of the Healthy Activity Programme, a Brief Psychological Treatment for Depression Delivered by Lay Counsellors in Primary Care: 12-month Follow-up of a Randomised Controlled Trial,” *PLOS Medicine*, 14, e1002385.

WORLD HEALTH ORGANIZATION (2016): “International Classification of Diseases (ICD),” *http://www.who.int/classifications/icd/en/*.

——— (2021): “Mental Health Atlas 2020,” Tech. rep.