Treating Mental Health Conditions Improves Labor Market and Other Economic Outcomes in Low and Middle-Income Countries*

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

23 December 2022

Abstract

Does treating mental health conditions improve labor market and other economic outcomes in low and middle-income countries? We run a systematic search for all randomized controlled trials (RCTs) which evaluate mental health treatments and measure economic outcomes in these countries. We conduct frequentist and Bayesian meta-analyses on estimates of treatment effects from 39 interventions. Treatments reduce the number of days participants cannot work by 16%, reduce the probability of being unable to work by 9 percentage points (26%), and improve qualitative measures of performance at work. They increase asset wealth and education investment. Results suggest multiple psychological and behavioral mechanisms.

*Affiliations: Lund: King's College London. Orkin, Walker: University of Oxford. Witte: IZA – Institute of Labor Economics, Bonn. Davies: University of Cape Town. Haushofer: Stockholm University. Bass, Murray, Murray: John Hopkins University. Patel: Harvard University. * denotes joint first authors. This study was funded by the Wellspring Philanthropic Fund. We are grateful to Ondine Berland, Carrie Brooke-Sumner, Tim Deisemann, Vimbayi Mafunda, Drummond Orr and Hannah Zillesen for excellent research assistance. Our thanks to Manuela Angelucci, Victoria Baranov, Sonia Bhalotra, Daniel Bennett, Pietro Biroli, Paul Bolton, Jon de Quidt, Sandy Douglas, Michael Gechter, Supreet Kaur, Rachael Meager, Gautam Rao, Matt Ridley, Frank Schilbach, Wietse Tol, Graham Thornicroft, and Eva Vivalt for thoughtful comments.

1 Introduction

Mental health conditions are prevalent and costly: estimates suggest 1 in 6 people (11-18 percent) worldwide have one or more mental health or substance use conditions at any one time (IHME, 2017). Globally, mental health conditions led to income losses and medical costs estimated at US$2.5 trillion in 2010, a similar cost to cancer (Bloom et al., 2011).[1] In high-income countries (HIC), meta-analyses find treating these conditions improves labor market outcomes, reducing sick leave and improving functioning at work (Nieuwenhuijsen et al., 2020, Salomonsson et al., 2018).[2] However, we know less about the effects on patients' economic outcomes of treating such conditions in low and middle-income countries (LMIC). The number of studies of mental health interventions in LMIC measuring economic outcomes has grown in the last ten years, including a growing literature in development economics studying treatments for depression, but results are mixed (Angelucci and Bennett, 2022, Baranov et al., 2020, Barker et al., 2022, Bhat et al., 2022, Patel et al., 2018). On average, treatment might lead to similar improvements in economic outcomes to treatment in richer economies, given treatments lead to similar improvements in mental health in low and high-income settings (Cuijpers et al., 2018, Singla et al., 2017). But treatment may also have different effects on economic outcomes in different economic environments: for example, those treated for mental ill health may still struggle to work in economies with informal work arrangements and few employment services.

In this paper, we estimate the average effect, across all available RCTs, of providing treatment for mental health conditions on patients' employment status, ability to work, time in work, functioning at work, and other economic outcomes in LMIC. We collect a new, comprehensive sample of estimates of treatment effects from a systematic search of studies which 1) test a psychotherapy or pharmacological intervention to treat mental health conditions in an RCT; 2) treat people diagnosed with a mental health condition; 3) measure any economic outcome; and 4) are in an LMIC. We consider published and unpublished work in any language as long as the abstract is in English. To ensure coverage of unpublished work, we search 21 databases and 8 trial registries in economics and health and hand-search references of included articles.

[1] This estimate is based on international prevalence estimates and national cost estimates for the United States, China, Kenya, Australia, Canada, the United Kingdom and France. Income losses account for roughly two thirds of this cost.

[2] Studies in Denmark and the United States find treatment of mental health conditions improves labor market participation and earnings (Biasi et al., 2021, Bütikofer et al., 2020).

The search ends in April 2022. We screen 15,031 papers and read 1,128 fully, yielding a sample of 39 interventions which meet our inclusion criteria. These are studied in 35 trials, involving 24,064 participants, and reported on in 40 papers.[3] The interventions include 22 psychotherapy or psychosocial interventions, 2 pharmacological interventions, and 15 combined interventions (combining both pharmacological and psychosocial treatments). These studies yield 180 estimates of the effect of an intervention on an economic outcome and 335 estimates of effects on mental health, functioning and other psychological outcomes. There are 59 rounds of data collection in these studies: 29 rounds less than six months after treatment, 18 rounds after 6-12 months, and 12 rounds after more than a year.

We find large and positive effects of mental health treatments on a number of economic outcomes. We conduct separate meta-analyses on groups of outcomes measuring similar underlying economic concepts. The most widely measured type of economic outcomes are "work-related", such as whether someone is in employment or unable to work. Such outcomes are measured in 35 of 39 interventions and make up 61 percent of our sample of effect sizes. On average, mental health interventions have a positive and significant effect of 0.14 standard deviations (CI: [0.08,0.22]) across all work-related outcomes, coded so that an increase indicates an improvement in the outcome. We also analyze smaller subgroups of work-related outcomes measured in the same units. We find a 9 percentage point (CI: [0,0.17]) decrease in whether patients are unable to work in studies which measure this outcome. This is a 26% decrease relative to a control group mean of 0.35. Treated participants are able to work on 1.7 more days in the last 30 days (CI: [-3.37,-0.03]), relative to a control mean of 10.69, a 16% increase. Treatment improves qualitative measures of whether participants are functioning normally at work by 0.2 SD (CI:[0.01,0.39]).[4] It has positive but noisily estimated effects on whether patients are employed and the time they spend in work. Treatment has positive, statistically significant effects on education outcomes (0.16 SD, CI: [0.04,0.29]) and asset wealth (0.07 SD, CI:[0.01,0.12]), but only small effects on income, consumption, and spending on inputs.

Second, we examine whether effects are robust to different ways of accounting

[3]Some trials include multiple interventions (e.g. Ran et al., 2003), while some papers report on the same intervention (e.g. Nadkarni et al., 2017a,b).

[4]"Functioning" refers to people's cognitive and social abilities to perform their normal social and economic roles (Edlund et al., 2018). Improvements in functioning are equivalent to reductions in disability. Table A7 shows representative examples of commonly used measures of functioning at work.

for heterogeneity in study-level treatment effects arising from different study populations, treatments or regions. Effects are robust to using a hierarchical Bayesian approach to more explicitly model treatment effect heterogeneity and to using multivariate meta-regression analysis to control for study characteristics. We find little heterogeneity in effects by region, but some heterogeneity if we compare populations with different disorders receiving different types of treatment. Economic effects are somewhat larger for populations with severe than mild disorders and for combination treatments than for drugs or therapy alone, consistent with clinical evidence that combined interventions have larger effects on mental health (Cuijpers et al., 2013, De Silva et al., 2013). We find little evidence of publication bias using a range of techniques and hence little impact of correcting for potential bias.

Third, we examine potential psychological and behavioral mechanisms behind these effects. We find that these economic effects occur alongside reductions in symptoms of mental ill health and improvements in functioning. In this sample of studies, treating a mental health condition improves recovery rates, reduces relapse and hospitalisation, reduces severity of symptoms, and improves functioning in domains of life outside work.[5] These effects hold in the subsample of estimates from clinician assessments of participants: clinicians are usually blind to treatment status. We find interventions with larger effects on mental health or on functioning have larger effects on economic outcomes. Moreover, the elasticity of employment outcomes with respect to variation in mental health is of an economically meaningful magnitude. We conduct analysis on individual-level data from 17,668 participants in seven studies which treat depression, measure days participants are unable to work, and have publicly available data. Employing an instrumental variable design, a 1 SD decrease in depression is associated with a decrease in days unable to work of 3.86. However, treatment may affect days unable to work through additional channels beyond depression, complicating a straightforward causal interpretation of this relationship.

Finally, we collect cost data from 20 of 39 interventions. The mean cost of these 20 interventions is USD 363 per person treated, with variation by intervention type and region.[6] We do not conduct a head-to-head benchmarking as the target populations differ, but impacts on economic outcomes per dollar spent are substantially

[5]Effects on mental health and functioning in our subsample of studies, which measure health and economic outcomes, are similar to effects for all studies of mental health interventions in LMIC (Cuijpers et al., 2018, Singla et al., 2017).

[6]To benchmark, active labor market programmes in LMIC and graduation programs cost an average of USD 566 and 1,468 per person treated, respectively (McKenzie, 2017, Banerjee et al., 2015).

higher than for widely used economic interventions like livelihood interventions and active labor market programs, highlighting that mental health treatments cost-effectively improve economic outcomes for people with mental health conditions.

Our main contribution is to the literature on the effects of mental health interventions on economic outcomes. To the best of our knowledge, this is the first meta-analysis on the effects of mental health interventions on economic outcomes in low- and middle-income countries.[7] A growing strand of work in development economics finds mixed results on economic outcomes of treatments for depression in LMIC settings (Angelucci and Bennett, 2022, Baranov et al., 2020, Barker et al., 2022, Bhat et al., 2022). These papers find strong effects on investment in children's education and perceptions of economic wellbeing, but few effects on labor market outcomes or wealth. We complement these studies by aggregating across all available studies in LMIC. We can examine robustness of effects across different regions, treatments, and conditions: for example, the existing literature focuses on depression, while we study a range of disorders.[8] The main limitation of our method is that we can only draw conclusions for outcome measures captured by the studies in our sample. For example, we capture few observations of earnings and cannot speak to some potential mechanisms, such as preferences or beliefs.

Our work complements studies in HIC, which find links between treatment of mental health conditions and labor market outcomes. Meta-analyses in health find that treatments for depression improve employment rates, reduce sick days and improve functioning at work.[9] Individual studies find exogenous changes in drug availability for mental health conditions are positively related to earnings and labor market participation (Biasi et al., 2021, Bütikofer et al., 2020).[10] Relative to this work, we exploit causal identification from RCTs and examine effects in LMIC economic contexts. Poverty, food insecurity, unemployment and increased exposure to trauma cause or exacerbate mental health conditions in these contexts (Haushofer and Fehr, 2014, Lund et al., 2011, Ridley et al., 2020). We show that available treatments both alleviate symptoms and enable participants' income-generating activities.

[7] A previous systematic review on this question identified 9 studies (Lund et al., 2011) included in this review, but did not include a meta-analysis.

[8] Depression prevalence is 3.4% internationally at any one time, but other mental health conditions are also common. Prevalence is 3.8% for anxiety, 2.3% for substance abuse disorders, and 1% for severe mental health conditions like schizophrenia (IHME, 2017).

[9] Table A20 summarizes outcomes and effect sizes in this literature.

[10] Studies using panel data or instrumental variables have similar findings (Banerjee et al., 2017, Chatterji et al., 2011, Jones and Mitra, 2017, Kessler and Frank, 1997, Peng et al., 2016).

Finally, our work is one of a growing number of meta-analyses in economics (Card et al., 2018, Meager, 2019, Ridley et al., 2020, Vivalt, 2020). We use meta-analysis to test a hypothesis inexpensively by pooling secondary outcomes from many small studies to improve power. Tan and Kremer (2020) similarly pool studies of point-of-use water treatment to examine effects on child mortality, a rare outcome for which few individual studies are powered. Many studies in our sample were powered to test effects on their primary mental health outcomes, rather than the secondary economic outcomes we study, and mental health interventions tend to have larger effects on mental health than economic outcomes.[11]

The rest of the paper proceeds as follows. Section 2 describes the criteria for study inclusion and the search process. Section 3 describes the included studies and the effect size estimates we extracted from them. Sections 4 and 5 present the empirical strategy and main results. Section 6 tests for robustness of results to heterogeneity and describes tests for publication bias. Section 7 examines behavioral and psychological mechanisms. Section 8 presents cost data and exploratory analysis on interventions combining economic and mental health interventions.

2 Search, selection, and coding of primary studies

We conduct the systematic review using standard guidelines from the Cochrane Collaboration (see https://training.cochrane.org/handbook). We used the Population, Intervention, Comparison and Outcome (PICO) method to pre-specify study inclusion criteria and include all papers satisfying them to minimize subjective judgements about which studies are included.[12] A sample set of search terms for PubMed is in Appendix Section A. Our systematic review protocol was registered with the Prospective Register of Systematic Reviews (PROSPERO), Number: CRD42017058930.

2.1 Inclusion and exclusion criteria

Population: We include only studies in low or middle-income countries (defined by the World Bank in 2018).[13] We study the effect of treatment for a clinically diag-

[11] In our sample, effects on mental health outcomes are on average 0.22 standard deviations and effects on economic outcomes are 0.15 standard deviations. The average sample size in our pool of studies, $N = 617$, leads to a minimum detectable effect size of 0.2 standard deviations for a representative outcome, assuming measurement at baseline plus one endline, an intertemporal correlation coefficient of 0.2 and variation absorbed by stratification of 0.25.

[12] https://epoc.cochrane.org/resources/epoc-resources-review-authors.

[13] https://datahelpdesk.worldbank.org/knowledgebase/articles/378834-how-does-the-world-bank-classify-countries.

nosed 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 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 mental illness but no current mental illness were excluded. Participants had to be aged 14 years or older, as we sought to evaluate effects on economically active populations.

Intervention: Interventions could include psychotherapy, psychological or psychosocial treatments (hereafter referred to as psychosocial interventions), pharmacological treatment, or interventions combining these treatments. The clinicians on the author team compiled a list of widely used psychosocial or pharmacological interventions, which we searched for using specific terms (e.g., behavioral activation is a type of therapy; antipsychotics are types of drugs). 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 sufficient studies which used an RCT for well-powered inference, so we restricted the sample to only include RCTs. This makes us more confident about the validity of individual study identification strategies and potentially reduces the likelihood of publication bias in our sample.

Outcomes: We searched for any study which measured employment, labor force participation, productivity, job search, income, earnings, wages, assets, wealth, consumption, expenditure, calorie count, food security, savings, investments, technology adoption, expenditure on temptation goods, financial outcomes, health investment, education spending (children and own), income diversification, agricultural yields, revenue or profit from own employment, and economic empowerment. We also include studies on social networks or contacts related to economic behavior, as these may be important correlates of economic outcomes. We present results by subgroups of outcome as well as overall.[14] We do not exclude studies on the basis of the mental health outcomes they measured.

[14]We collected studies measuring contraception (methods and expenditure) but decided not to include these in the analysis.

2.2 Search strategy and data extraction

In the primary search, one author (TD) searched 21 databases, including all major economics, social science, and clinical databases and repositories of working papers. Databases are listed in Appendix Section B. To capture trials in progress that might have results, we searched trial registries, contacted trial authors for funded RCTs on NIH reporter, and contacted trial funders (Grand Challenges Canada and the Abdul Latif Jameel Poverty Action Lab). We also conducted hand searches of the citation lists of all included articles. We placed no restrictions on study date, with the earliest study published in 1994. The search ended in April 2022.[15] We included studies published in any language if an abstract was available in English. We found 3 studies in Mandarin Chinese which were translated into English and included.

Our sample of effect sizes comes from 39 interventions studied 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., 2017a,b). Table A21 gives details on each intervention and study. We followed a search and screening process shown in Figure A1. Searches of the databases and hand searches of citations of included papers yielded 15,031 potential studies. Three reviewers (TD and either CBS or VM) independently screened abstracts against inclusion criteria using Covidence software. Disagreements were resolved by a third reviewer (CL). After removing 90 duplicates, abstracts of 14,941 selected articles were screened by two reviewers (TD and either CBS or VM), with disagreements again resolved by a third reviewer (CL). 13,813 studies were removed as they did not meet inclusion criteria. We reviewed the full text of 1,128 articles. 1,089 were removed because they did not meet inclusion criteria, with reasons detailed in Figure A1, leaving 40 studies. Five studies combined psychosocial with economic interventions, which we exclude but discuss in Section 8. We assessed inter-rater agreement with the Kappa statistic, which measures the probability of agreement between two raters who each classify N items into C mutually exclusive categories. Our Kappa agreement probability was 0.90, reflecting high agreement.

One author then extracted information into a piloted, pre-populated Excel spreadsheet. A second author checked this information against papers. We coded all economic outcomes matching those in our search criteria, primary mental health outcomes, as defined by the authors, as well as all outcomes which fell into one of 22

[15] Two articles were identified later after hand-searching references of previously included studies.

categories of mental health or functioning outcome (see Section 3.4 for the list). We coded the definition of outcomes measured and statistical information on the effects, including the raw (reported) effect size, its type (continuous vs dichotomous) and standard error, the means and standard deviations in treatment and control groups, and sample sizes. We requested any missing information needed to compute effect sizes for a study from authors.

In terms of study quality, we include only RCTs. We code measures of risk of bias, following the Cochrane Collaboration's recommendations. We observe relatively high study quality among included RCTs on these indicators and did not exclude any studies. We implicitly control for a study's sample size by weighting the original effect sizes with the inverse of their variance during the meta-analysis (more details in Section 4).

3 Study and intervention characteristics

Table 1 and Table A1 summarize the characteristics of the 39 interventions. Interventions are in a mix of middle and low income countries and regions (Table A1). Most papers are published after 2011: 33% of interventions are studied in papers published between 2011 and 2015, 28% in papers published between 2016 and 2020, and 21% in papers published or released since 2021 (Table A1).

3.1 Intervention type

Panel A of Table 1 describes interventions. We classify the interventions into three pre-specified categories: psychosocial, pharmacological, and combined interventions, according to the therapeutic element employed. Pharmacological interventions are medications and include antidepressants; antipsychotics, used to treat psychosis (including schizophrenia), and "opioid medication", to treat opioid misuse. They account for only 2 (5%) of the sample of interventions. Psychosocial interventions are the most common in our sample and account for 22 interventions (56%). Interventions include different types or elements of therapy, with some including more than one element. Cognitive behavioral therapy, psychoeducation, problem-solving therapy and interpersonal therapy are the most common elements (Table A1). Combined interventions, which administer psychosocial and pharmacological interventions alongside each other, are the second most common intervention type (15 interventions, 38%). They are typically used for patients with more severe symptoms. We prespecified we would group different types of drugs together because the older

Table 1: Interventions in included randomized controlled trials

	(1) Number of interventions	(2) Share of interventions
Panel A: All interventions	39	1.00
Main intervention category *(mutually exclusive)*		
Pharmacological	2	0.05
Psychosocial (only)	22	0.56
Pharm.+psych	15	0.38
Targeted condition *(mutually exclusive)*		
Common mental disorders (CMD)	17	0.44
Severe mental disorders (SMD)	11	0.28
Substance use disorders	6	0.15
Post-traumatic stress disorders (PTSD)	5	0.13
Control condition *(mutually exclusive)*		
Enhanced Usual Care	7	0.18
No Treatment	19	0.49
Treatment As Usual (Pharmacological)	13	0.33
Panel B: Outcome measures		
Economic outcomes		
In employment	7	0.18
Time in work	7	0.18
Unable to work	5	0.13
Days unable to work	13	0.33
Functioning at work	13	0.33
Job search	3	0.08
Education	3	0.08
Assets	4	0.10
Income, consumption and input expenditure	7	0.18
Subjective poverty measures	4	0.10
Social networks	2	0.05
Other	3	0.08
Mental health outcomes (all)		
Suicide attempts or at risk of suicide	10	0.26
Relapse (dummy)	8	0.21
Recovery (dummy)	4	0.10
Rehospitalisation	4	0.10
Diagnosed with mental health condition	7	0.18
Qualitative assessment of mental health condition	7	0.18
Substance use	6	0.15
CMD symptoms	23	0.59
PTSD symptoms	6	0.15
SMD symptoms	6	0.15
Overall measures of functioning	25	0.64
Functioning in social interactions	5	0.13
Self-regulation	4	0.10
Self-esteem/self-efficacy	5	0.13
Cognition	5	0.13
Physical health	4	0.10

Notes: There are 39 interventions, 180 economic effect sizes and 335 mental health effect sizes. Some 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. Table A2 lists each unique measurement tool used for each group of outcomes.

medications typically available in LMIC have similar clinical effects within a disorder, for both depression (Cipriani et al., 2018) and schizophrenia (Leucht et al., 2013). We group different types of therapy together because most have similar effects on mental health (Cuijpers et al., 2008, Cleary et al., 2008).

3.2 Target condition

We group together mental health conditions into four broad categories according to the World Health Organization International Classification of Disease (ICD-10) (WHO, 2016), which bases these categories on shared clinical presentation, functional disability, and treatment approaches.[16] Common mental disorders (CMD) include depression and anxiety disorders and are targeted by 17 interventions (44%). Severe mental disorders (SMD), which include schizophrenia and bipolar disorder, have more severe impacts on functioning and longer duration than CMD. They are targeted by 11 interventions (28%). Substance use disorders (SUD) include dependent, harmful use of substances such as alcohol, marijuana, and opioids.[17] Post-traumatic stress disorder (PTSD) is triggered by traumatic events, with symptoms such as intrusive memories and nightmares impacting on functioning in daily life. The latter two disorders are targeted by six and five interventions respectively.

3.3 Economic outcomes

Our sample of studies yields 180 effect sizes which capture the effect of an intervention on an economic outcome. We grouped economic outcomes into the categories shown in Tables 1 and 2.[18] Table A2 lists each measurement tool used in each group of outcomes. There are two ways to examine the frequency with which outcomes are measured. Column 1 of Table 1 describes the number of interventions where an outcome category is measured. Column 4 of Table 2 captures the number of effect sizes for an outcome that appear in the dataset. For some interventions, the same outcome is measured more than once (for example, at different time points). We discuss methods of analysing multiple observations for an intervention in Section 4.1.

[16] 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.

[17] One study targets antisocial behavior, which includes a broad range of behaviors that are contrary to norms, including substance abuse and criminal or violent behavior. All sampled individuals are diagnosed with substance abuse problems, so we classify this study under SUD.

[18] We pre-specified nine groupings: employment, education, income, financial behavior, wealth, consumption, health costs, other indicators of poverty, and social networks. However, we had sufficient employment outcomes to disaggregate these further. We did not find health costs measures.

Most outcomes fall into the categories of employment, education, and financial outcomes. Employment or work-related outcomes are the most common category, measured in 87% of interventions, with 95 effect sizes (53% of the sample of effect sizes). We group these outcomes into outcome categories with similar measurement capturing similar aspects of an individual's involvement in work. "In employment" captures if someone is employed. Time in work measures the amount of time worked in hours or months in different recall periods. Being unable to work, known in HIC studies as work-related disability, captures if someone is unable to work. Days unable to work is similar to measures of disability days or sick leave in HIC studies. Measures of functioning at work are validated qualitative scales, mainly 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.[19] Table A7 provides sample 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.

We also find non-work-related economic outcomes. Outcomes related to the education of respondents or their children, such as school attendance or expenditure on education, are measured in 8% of interventions, with 17 effect sizes. Remaining outcome groupings are consumption, wealth, income, and respondents' subjective perceptions of poverty. These smaller groups together make up half of all effect sizes and are measured in 36% of interventions. We also examine 7 social networks outcomes. Our results are not sensitive to the inclusion or exclusion of these outcomes.

3.4 Mental health, general health and other psychological outcomes

We also extracted all effects on psychological and behavioral pathways which might act as mechanisms for effects on economic outcomes. We coded any primary mental health outcome, as defined by the authors, as well as all outcomes which fell into one of 22 categories of outcome: suicide risk, rehospitalisation, 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. We find

[19] 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."

335 effect sizes related to mental health, general health and other psychological outcomes (143 of which are primary outcomes). The measures used to assess behavioral and psychological pathways are listed in Tables A3 and A4.

Some outcome measures are from hospital or clinician records, such as whether individuals made any suicide attempts or were at risk of suicide (measured in 26% of interventions), or whether participants have relapsed (21%), recovered (10%), or been rehospitalised (10%) (Table 1). 18% of interventions measure a dummy variable for whether an individual is diagnosed with a condition and 18% capture a continuous qualitative measure assessing the severity of the condition.

Other outcome measures are psychological scales reported by the participant which measure the severity, frequency or duration of symptoms of personal distress or functional impairment. Some scales are used only for a particular disorder. For example, PTSD or schizophrenia result in distinct symptom profiles and forms of disability. Others, such as measures of depression and anxiety or general functional impairment, are used across disorders. 15% of interventions measure substance use (using scales or urine tests). The majority of interventions measure CMD symptoms (59%) (Table 1), 15% measure PTSD symptoms, and 15% measure SMD symptoms (mostly schizophrenia-related outcomes). We list the specific wording for commonly used scales in our sample in Tables A7-A10.

The majority of interventions (64%) study effects on an overall measure of functioning across domains of life. Some also 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 and social interactions with psychological and behavioral mechanisms. Where only the overall functioning score is reported, we include this effect as a mechanism. A small number of interventions measure self-regulation (ability to control impulses or structure one's time), self-esteem or self-efficacy (underlying beliefs about one's ability to carry out actions or achieve desired outcomes), cognitive performance, and general health.

3.5 Control conditions

Our search allowed a range of control group conditions. We sort control conditions into three categories: no treatment, treatment as usual (TAU, Pharmacological), or enhanced usual care (EUC). Most interventions in our sample (49%) are compared to no treatment (Table A1). We include one waitlist control, where participants do

not know they are waitlisted, and two placebo medications in this group. Treatment as usual (TAU) (Pharmacological), to which 33% of interventions are compared, is where patients receive the standard medication they would normally receive in their clinical setting, which was an anti-psychotic for all interventions in this study. For conditions where a known treatment exists and is provided by the public health system, it is considered unethical to deprive a control group of this intervention. Trials thus test if the experimental treatment (e.g. a new drug or a combination of drugs and therapy) performs better than the usual treatment. Receiving a TAU (Pharmacological) control would be likely have some positive effects, so effects compared to these controls would tend to be smaller than effects compared to no treatment. Enhanced usual care (EUC), to which 18% of interventions are compared, involves standard care plus an additional component such as receiving information pamphlets, general health home visits, or referrals to a doctor. We discuss the validity of pooling studies with different types of control groups in Section 6.2.

3.6 Location, time of measurement, and target population

Most interventions were in South Asia (36%), East Asia and the Pacific (28%), or sub-Saharan Africa (23%) (Table A1). Most interventions are in lower-middle income countries (49%), with a minority (15%) in low-income countries. We include effects measured at any point after the beginning of treatment and often include multiple measurement points per intervention. Interventions have various combinations of follow-up periods (Table A1). 65% of interventions report effects at one follow-up time point only, 16% at two, and 19% at three time points. The average intervention in our sample has 1.5 follow-up rounds and occurs 15.2 months after treatment. There are a total of 59 rounds of data collection covered in the 39 interventions. 29 rounds (in 25 interventions) are less than 6 months after treatment, 18 rounds measure effects between 6 and 12 months after treatment, 6 rounds measure effects between 12 and 24 months after treatment, and 6 rounds measure effects after two years. Most interventions examine adults, except two, which study youth aged 15-24 in one case and 18-35 in another case.

4 Empirical strategy

4.1 Aggregating from raw effect sizes to inference datasets

To explain how we construct the datasets on which we perform meta-analytic inference and to fix intuitions, we present a subset of our "raw dataset" in Figures A2 and

A3. This shows effect sizes on work-related outcomes from individual studies. Each panel shows a plot of effect sizes from our sample, organised in the figure by type of outcome. Each row of the plots represents the effect size of a single intervention on one outcome in one survey round and its associated confidence interval.

Studies often report multiple estimates for the effect size of an intervention on an outcome, such as in robustness checks or repeated survey rounds. To perform inference, we average across the multiple effect size estimates to generate one average effect at the intervention-outcome level. This is necessary for two reasons. First, if we did not, dependence between multiple effect sizes reported within a given study would bias significance tests and confidence intervals (Gleser and Olkin, 2009). Second, effect sizes would be improperly weighted in the aggregation procedure as studies with more effect sizes would be given more weight. We generate an average effect size in standard deviations, known as Hedges' g (for both continuous and binary outcomes). We calculate the standard error of the average effect size following Borenstein et al. (2009). Details are outlined in Appendix D. In meta-analyses, we then average over these effect sizes at the intervention-outcome level.

We face a decision about whether to pool effect sizes that plausibly measure the same underlying economic concept, but are often measured using different survey questions or recall periods. We face a trade-off between statistical power and interpretability. For example, we could calculate an estimate using only measures of job search hours per week, as reported in panel (b) of Figure A3, but we would only have two observations of the effect of an intervention on this outcome, from Fuhr et al. (2019) and Patel et al. (2017). If we instead pooled all job search outcomes in panel (f), including search hours measured over different recall periods and availability to take on another job, we could expand the sample size to six original effect sizes from three interventions, giving us three average effects at the intervention-outcome level, one for Fuhr et al. (2019), one for Angelucci and Bennett (2022) and one for Patel et al. (2017). However, this comes at the cost of direct interpretability, as it is less clear exactly what the average effect size measures, and we can only compare outcomes in standard deviation terms.

We therefore report results from repeated meta-analyses, each of which employs a different degree of pooling across outcomes, allowing the reader to select the level of aggregation they feel is appropriate. At the lowest level of pooling, we only average across outcomes measured using the same survey tool for the same recall period.

In the second level of aggregation, we report average effect sizes for groups of similar outcomes, such as at the panel-level in Figures A2 and A3. Finally, we aggregate across all work-related or non-work-related outcomes, or all economic outcomes. Of course, besides outcomes, we are also pooling across subtly different mental health interventions, targeted disorders, regions, and other plausibly important determinants of intervention efficacy, as well as across observations from different time points after an intervention. We disaggregate our findings by these dimensions of potential heterogeneity or show robustness to controlling for them in Section 6.

4.2 Model

A priori, we expect significant heterogeneity in study-level treatment effects arising from subtly different study populations, treatments, and outcome measurement. We therefore follow the Random Effects meta-analysis literature. 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 plus a sampling error term ϵ_k.

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

We estimate Equation 1 using two approaches.[20] 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. The Central Limit Theorem ensures that the weighted average is approximately normally distributed above about $k = 30$, enabling inference.

Second, we take a hierarchical Bayesian approach to model treatment effect heterogeneity more explicitly. This is particularly useful at the lowest level of pooling, where we have small sample sizes. We implement the Rubin (1981) model

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

[20] In all cases, we winsorize the top 1% of effect sizes to limit the impact of large outliers on the effect size sample.

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). We use measures of heterogeneity to show whether it is sensible to pool across outcomes. We report three measures of heterogeneity. First, our estimate of σ^2 is a measure of heterogeneity in the distribution of effect sizes. Second, we report the average pooling metric, which 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). We take the simple average across studies as an indicator of heterogeneity, per Meager (2019):

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

Third, we report the I^2, where $I^2 = \frac{\hat{\sigma}^2}{\hat{\sigma}^2 + \hat{se}_k^2}$. 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. In our preferred specification, our priors on τ and σ are weakly informative:

$$\tau \sim N(0, 1)$$
$$\sigma \sim HC(1)$$

Where N indicates the Normal distribution and HC the Half-Cauchy distribution. As our priors are only weakly informative and we have a large number of studies, we expect little power improvement from the Bayesian relative to the frequentist approach. Our results are robust to the choice of a range of reasonable priors, including uninformative priors (e.g. $\tau \sim N(0, 10)$).

5 Effects on economic outcomes

Our core findings are reported in Table 2, which shows results from repeated meta-analyses estimating Equation 1. Each row represents a meta-analysis on a different outcome, where bolded titles represent groups aggregated across the outcome subgroups below. Panel A shows standardized effect sizes aggregating across outcomes measured in different ways, while Panel B shows effect sizes for a subset of work-related outcomes measured in the same raw units. No adjustment is made for multiple hypothesis testing.

The average treatment effect of mental health interventions across all economic outcomes measured in these studies is meaningful ($SD = 0.15$) and precisely estimated (95% CI: [0.09,0.21]). Disaggregating this effect into outcome subgroups, we see positive effects for most outcome subgroups, although there is heterogeneity in magnitude and significance of effects. Across work-related outcomes, the effect size is the same as the aggregate and precisely estimated ($SD = 0.15$, CI: [0.08,0.22]). There is some heterogeneity in the standardized effect size across different measures. There is a small, positive but insignificant ($SD = 0.03$, CI: [-0.1,0.16]) effect on measures which capture whether an individual is in employment. There is a large effect on measures of time in work, but this is only significant at the 10% level ($SD = 0.12$, CI: [-0.02, 0.26]). There is a large, precisely estimated reduction in measures of being unable to work: treatment increases the likelihood an individual is able to work ($SD = 0.17$, CI: [0.06, 0.27]) (the studies report measures of being "unable to work", which we reverse-code so that an increase reflects an improvement). There is a small, positive, but insignificant effect ($SD = 0.03$, CI: [-0.01,0.07]) on measures of the number of days an individual is unable to work. There are large, positive, and significant effects on qualitative scales measuring "functioning at work" ($SD = 0.20$, CI: [0.01,0.39]).[21] There is little evidence of an effect on job search, although only three studies measure this outcome. The lack of effect on employment, despite effects on participants' reported ability to work, may capture that individuals work in environments where employment is difficult to find, conditional on being able to work.

Taken together, these estimates suggest treatment for a mental health condition has economically important effects on whether participants are able to participate in economic activity and their performance at work. We are, however, limited by what studies have measured. Many studies do not use standard labor market survey measures. Measures aggregate over work inside and outside the home and paid and unpaid work. We also cannot compare the magnitude of effects on different outcome groupings because studies do not measure all outcome groupings, so differences between outcome groupings may reflect the pool of studies measuring a particular outcome grouping rather than differences in effects by outcome grouping.

In Panel B, we show that effects on labor market outcomes are also present for the subset of studies which measure work-related outcomes using the same survey mea-

[21] Wording of a representative subset of work-related functioning outcomes is reported in Table A7.

sures and recall periods. Here, we do not standardize effect sizes. There is no effect on dummy variable measures capturing whether individuals are employed, compared to the average of control group means in these studies of 28%. Treated individuals are 9 percentage points (CI: [0.00,0.17]) more likely to self-report currently being able to work on the Social Disability Screening Schedule (SDSS) assessment, relative to the average control mean of 35% of non-treated individuals who are able to work, an increase of 26%.[22] They also report being able to work on 1.7 more days in the last 30 days (CI: [-3.37,-0.03]), relative to a control mean of 10.69, on the WHO-DAS 2.0 measure, an increase of 16%. The effect on this measure of days worked is larger than the effect in Panel A. This estimate is our preferred estimate, as the effect in Panel A includes two studies which measure days of sick leave in the last two years, which is likely to suffer from poor recall and is difficult to compare to measures with a 30-day recall period. The raw effect sizes reported by individual studies are presented in Figures A2 and A3.

In Table A20, we summarize effects from meta-analyses in HIC, although these have different inclusion criteria and often examine only one type of intervention, so are not exactly comparable to our average effects. 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). There are some exceptions which may reflect differences in economic environment. Effects on the employment rate are larger in HIC than LMIC studies (0.15 SD vs 0.03 SD), although effects on the employment rate in HIC meta-analyses are similar to the effect on being unable to work in our study (0.17 SD). This is consistent with employment being harder to find in LMIC settings, conditional on being able to work. Effects on sick leave, equivalent to days unable to work in LMIC, are broadly similar (decreases of 15 and 20 days per year in HIC studies; decreases of 20 days per year in our study). Effects on employment frequency are also similar (increases of 0.15 and 0.31 SD in different HIC meta-analyses; increases of 0.12 SD in our study). Effects on functioning at work are somewhat larger in HIC studies (0.43 and 0.31 SD) than our study (0.2 SD), though the groups of studies in HIC samples and our samples use different measures.

[22] See Table A7 for detailed wording of this scale.

Table 2: Effects of mental health interventions on economic outcomes: Frequentist approach

	(1) Aggregate Hedges' g	(2) 95% lower CI	(3) 95% upper CI	(4) # of obs.	(5) # of intrv.	(6) Control mean
Panel A: Effects by outcome type						
Total	0.15***	0.09	0.21	180	39	
Work-related outcomes						
All work-related outcomes	0.15***	0.08	0.22	95	34	
In employment	0.03	-0.10	0.16	11	7	
Time in work	0.12*	-0.02	0.26	17	7	
Unable to work[1]	0.17***	0.06	0.27	18	5	
Days unable to work[1]	0.03	-0.01	0.07	23	13	
Functioning at work	0.20**	0.01	0.39	20	13	
Job search	0.02	-0.08	0.12	6	3	
Non-work-related outcomes						
All non-work-related outcomes	0.06**	0.00	0.12	85	14	
Education	0.16***	0.04	0.29	17	3	
Assets	0.07**	0.01	0.12	18	4	
Income, consumption and input expenditure	0.01	-0.06	0.09	27	7	
Subjective poverty measures	0.06	-0.11	0.22	11	4	
Social networks	0.02	-0.05	0.08	7	2	
Other	0.17	-0.16	0.51	5	3	
Panel B: Work-related effects in original units						
Self-reported as employed	0.01	-0.04	0.06	11	7	0.28
WHODAS 2.0: self-reported days unable to work in last 30 days	-1.70**	-3.37	-0.03	13	7	10.69
SDSS: Self-reported assessment – able to work part or full time	0.09**	0.00	0.17	4	2	0.35

Notes: Hedges' g is the small-sample-bias-corrected standardized mean difference in the economic outcome between treatment and control. *, ** and *** denote statistical significance at 10 percent, 5 percent and 1 percent level of significance respectively. The average measurement in our sample happens 15.2 months after intervention start. [1] = reverse coded, so higher values mean better employment outcomes. Aggregate Hedges' g represents an estimate from random effects inverse variance weighted meta-analysis. The aggregation of individual effect sizes works as described in subsection 4.1. All individual effect sizes are winsorized at the 99^{th} percentile.

Turning to non-work-related outcomes, there are large, significant impacts on education outcomes ($SD = 0.16$, CI: [0.04,0.29]) and asset wealth ($SD = 0.07$, CI: [0.01,0.12]). However, the number of interventions that reported these specific outcomes was small ($n \leq 4$) and we aggregate over different measures, so we interpret these findings with care. There is no significant effect on measures of income, consumption or input expenditure, subjective poverty measures, or social networks. We report the average across non-work-related economic outcomes and the "other" category for completeness, but these averages capture significant heterogeneity between groups of non-work-related outcomes.

To explicitly model heterogeneity between interventions, we replicate our findings under the Bayesian hierarchical model, reporting posterior means and estimated using the maximum likelihood estimator. The recovered mean estimates are essentially identical to the frequentist specification. The credible intervals are similar to their frequentist counterparts, though slightly wider. We have moderate heterogeneity of roughly the same size as sampling error when pooling across all economic outcomes ($\omega(\tau) = 0.49, I^2 = 0.75$). This is driven by much greater heterogeneity in work-related outcomes ($\omega(\tau) = 0.36$, $I^2 = 0.84$) than in non-work-related outcomes ($\omega(\tau) = 0.79$, $I^2 = 0.19$). We therefore have to be particularly careful in interpreting the overall effect size on work-related outcomes.

Looking at the more granular outcome levels, we see that the credible interval for σ includes zero in most cases, indicating that σ is imprecisely estimated.[23] We therefore interpret findings on the heterogeneity measures with care. Looking only at the point estimates of the pooling factors, among work-related outcomes, we see the lowest heterogeneity in the measures of being in employment, being unable to work, and days unable to work ($\omega(\tau) = \{0.63, 0.61, 0.58\}$, respectively). Among non-work-related outcomes, we see the lowest heterogeneity in income ($\omega(\tau) = 0.72$). We are most confident that the effect sizes drawn from pooling within these categories is capturing a stable underlying effect. We see particularly high heterogeneity on the time in work, functioning at work and job search ($\omega(\tau) = \{0.34, 0.20, 0.22\}$), as well as education and social networks ($\omega(\tau) = \{0.29, 0.15\}$). Overall, we see substantial heterogeneity, motivating the careful investigation of the impact of site-level covariates and heterogeneous treatment effects in the following sections.

[23] The Half-Cauchy prior mechanically bounds the heterogeneity measure to positive values.

Table 3: Effects of mental health interventions on economic outcomes: Bayesian approach

	Effect size			Heterogeneity measures				
	(1) τ posterior mean	(2) τ 95% Lower CI	(3) τ 95% Upper CI	(4) σ posterior mean	(5) σ 95% Lower CI	(6) σ 95% Upper CI	(7) $\omega(\tau)$	(8) I^2
Total	0.15	0.08	0.23	0.16	0.09	0.24	0.47	0.75
Work-related outcomes								
All work-related outcomes	0.16	0.07	0.25	0.21	0.13	0.30	0.36	0.84
In employment	0.03	-0.14	0.21	0.12	0	0.36	0.63	0.26
Time in work	0.12	-0.05	0.33	0.18	0	0.38	0.34	0.67
Unable to work[1]	0.18	0	0.36	0.11	0.00	0.33	0.61	0.25
Days unable to work[1]	0.05	-0.02	0.14	0.07	0.00	0.16	0.58	0.75
Functioning at work	0.20	-0.03	0.43	0.36	0.18	0.57	0.20	0.85
Job search	0.03	-0.31	0.38	0.20	0.00	0.65	0.22	0.60
Non-work-related outcomes								
All non-work-related outcomes	0.07	0	0.14	0.06	0	0.13	0.79	0.19
Education	0.17	-0.22	0.56	0.22	0	0.72	0.29	0.57
Assets	0.06	-0.06	0.17	0.07	0	0.22	0.45	0.35
Income, consumption and inputs	0.01	-0.09	0.12	0.07	0	0.17	0.72	0.20
Subjective poverty measures	0.06	-0.23	0.37	0.22	0	0.55	0.26	0.74
Social networks	0.02	-0.73	0.78	0.45	0	1.41	0.15	0.68
Other	0.19	-0.50	0.97	0.54	0	1.42	0.14	0.88

Notes: Table 3 reports estimates of the Rubin (1981) hierarchical Bayesian model. The marginal posterior mean and credible interval estimated via MLE is reported for the treatment effect τ and heterogeneity measure σ. The average pooling factor $\omega(\tau)$ and I^2 are reported as interpretable summary measures of heterogeneity at each level of disaggregation. The quoted intervals are shortest credible intervals. For simplicity, we do not report Bayesian equivalents to frequentist hypothesis tests (e.g. Bayes factors). The number of of observations and interventions is the same as reported in Table 2. [1] = reverse coded, so higher values mean better employment outcomes. We take $\tau \sim N(0,1)$ and $\sigma \sim HC(1)$ as our priors. The estimate of σ is bounded by zero below by the choice of Half-Cauchy prior.

6 Robustness

In this section, we show robustness of our core findings to common complications in meta-analysis: heterogeneity between studies and publication bias. We first perform subgroup analyses to examine whether theoretically important dimensions of heterogeneity are driving our findings. We then explore robustness of findings to inclusion of study-level covariates. Finally, we show results of tests for publication bias.

6.1 Heterogeneity: Intervention, targeted condition and region

Causal inference in our setting is complicated by study teams endogenously selecting samples, targeted conditions and treatments. We might also suspect significant heterogeneity with respect to intervention region. In Figure 1, we show coefficient plots of the results of disaggregating our overall effect size with respect to each of these dimensions of heterogeneity. This allows us to better understand whether study characteristics drive our core findings. In the next section, we examine robustness of the average latent treatment effect to inclusion of a vector of study-level covariates.

In the second and third panels of Figure 1, we examine effects by intervention type and targeted disorder.[24] It is plausible that the effectiveness of mental health treatments in improving economic outcomes will vary between psychosocial, pharmacological, and combined treatments, given the different mechanisms through which these interventions influence mental health symptoms and daily functioning (Patel et al., 2016). It is also plausible that treating different conditions might have different effects, given the different symptom profiles and forms of disability associated with each condition. In the second panel, we find that there are somewhat larger effect sizes for combined interventions (SD=0.30, [0.18,0.43]), relative to psychosocial interventions.[25] In the third panel, we find that interventions that target severe mental disorders (SMD) (SD=0.24, [0.07,0.40]) have slightly larger effects than the overall effect (SD=0.22, [0.14, 0.30]) and than effects of treatments targeting other disorders. These dual findings are congruous in that more severe mental health conditions tend to be treated by a combination of pharmacological and psychosocial treatments. The larger effect in this subpopulation is likely driven by the relatively large impact of interventions on symptoms of these disorders, discussed in Section 7.

[24] We show the number of effect sizes for each outcome category by intervention type in Table A5.

[25] We do not interpret effects on pharmacological treatments as our sample has only two interventions, making inference poorly powered.

Figure 1: Robustness to disaggregation by intervention, targeted condition and region

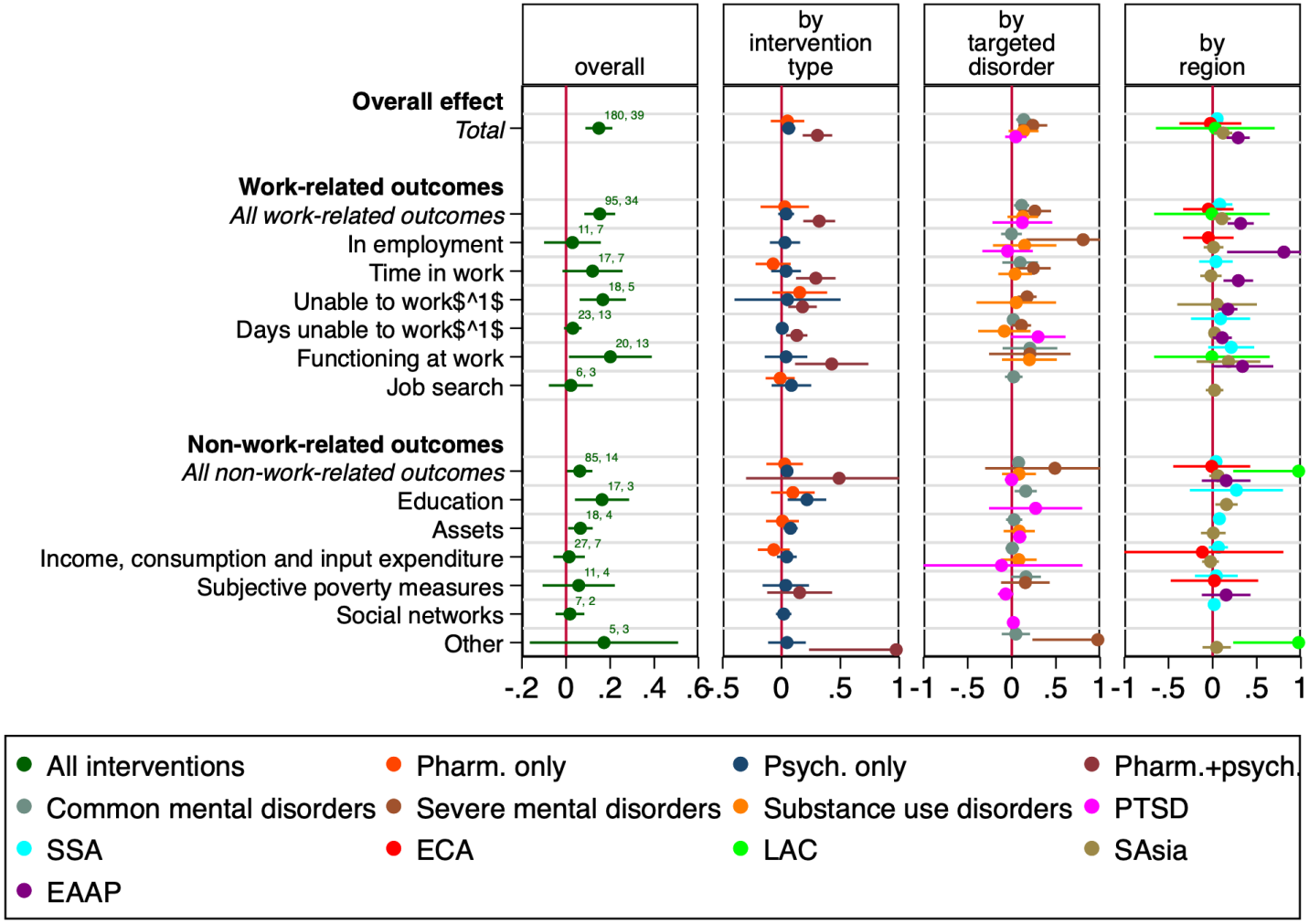


Figure 1 shows aggregate economic meta-effect sizes (Hedges' g) for various economic outcomes, both work-related and non-work-related. The effects are shown for the overall sample of 39 interventions (panel 1, corresponding to table 2), by intervention (panel 2), by targeted disorder (panel 3) and by region (panel 4). The horizontal axis displays the average economic effect size in standard deviations. The aggregation of individual effect sizes works as described in subsection 4.1. Individual effect sizes are winsorized within outcome type at the 99^{th} percentile. In the first panel, the first number next to the effect size marker represents the number of individual effects going into the aggregate meta-effect, the second number represents the number of different interventions from where these individual effect sizes come. SSA = Sub-Saharan Africa, ECA = Europe and Central Asia, LAC = Latin America and the Caribbean, SAsia = South Asia, EAAP = East Asia and Pacific.

This echoes findings in the literature that combination treatments are more effective than either psychosocial or pharmacological treatments alone in treating symptoms for depression, anxiety and schizophrenia (Cuijpers et al., 2013, De Silva et al., 2013). The effect size on common mental disorders is smaller than the effect on more severe disorders, but positive and significant overall, and across multiple outcome subcategories. Effects of substance use disorders and PTSD are noisily measured as they are measured in five and six interventions respectively. The coefficient on substance abuse (SD=0.13, [-0.03, 0.29]) is similar to the overall finding and marginally significant, while that on PTSD is small (SD=0.03, [-0.07, 0.14]), and non-significant. We caution against using these estimates to compare the effectiveness of types of treatment: intervention type and study population are chosen endogenously, so differences may simply capture the effect of treating different subpopulations.

In the fourth panel of Figure 1, we find some evidence of regional variation in effect sizes, but inference is complicated by the large standard errors of the individual estimates. We conclude that there are some differences in effect magnitudes, though these are mostly relatively small. In the next section, we explicitly account for different dimensions of effect size heterogeneity and show robustness of core results.

6.2 Study characteristics (multivariate meta-regression)

We extend Equation 1 to a Mixed Model (meta-regression) by allowing for a vector of de-meaned study-level covariates $\tilde{X}_{es}$:

$$\begin{aligned} \hat{\tau}_{es} &= \tilde{X}_{es}\beta + \epsilon_{es} \\ \tilde{X}_{es} &= X_{es} - \bar{X}_{es} \quad \forall\ X_{es} \end{aligned} \tag{2}$$

Where $\hat{\tau}_{es}$ is the effect size e taken from study s. The vector β includes the intercept on which we perform inference and captures study-level heterogeneity that is explained via the de-meaned covariates, $\tilde{X}_{es}$. This allows us to show robustness of our result to site and study characteristics.

Relative to Equation 1, we aggregate at the effect size level for work-related outcomes ($n = 95$) to retain higher variation with respect to study-level covariates, instead of at the intervention level ($n = 39$, or $n = 34$ with work-related outcomes). The increase in sample size comes at a cost. We expect dependence between multiple effects and need to account for overweighting of studies that report many effect

sizes. We therefore implement a multivariate random effects meta-regression procedure to allow for joint inference on dependent effect sizes. We estimate parameters via restricted maximum likelihood following Jackson et al. (2011).

Our findings are reported in Table A11. Our headline finding is robust to accounting for study characteristics. In each of the specifications, we see an average effect size that is equal to or slightly smaller than that in our unconditional meta-analysis, and each estimated treatment effect is significant at the 5% level. Including the variance of the error term and accounting for measurement timing (namely month between intervention and measurement) decreases the coefficient by small but non-trivial amounts. Due to limited variation, we cannot include all study fixed effects concurrently. In Appendix Section E.2, we further explore robustness of the finding to exclusion of the largest studies, types of control conditions and other dimensions of heterogeneity, finding little impact of accounting for these study features.

6.3 Publication bias

We summarize findings on publication bias here and provided a detailed investigation in Appendix E.3. Conventional methods suggest little evidence of publication bias. In Figure A6, we display a histogram and funnel plot of our effect sizes and their standard errors, showing that there is 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 evidence against the null hypothesis of no small-study effects, $\hat{\beta} = 0.01(0.19)$, with details reported in Table A13.

However, these tests 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). We therefore formally model the impact of publication bias in our study sample, following Andrews and Kasy (2019). We find no evidence that statistically significant findings are reported more often than null findings (Table A14). If anything, the point estimates indicate that the relative probability of publication of studies in the subsets of $Z \in (-\infty, 1.96)$ is higher than that for $Z > 1.96$, though probabilities are noisily estimated. Findings are robust to removal of reported effects that have very large standard errors ($SE > 10$) (Table A15).

7 Mechanism analysis

We have shown that mental health interventions have large, positive and statistically significant impacts on work-related, education and asset outcomes. In this section, we examine potential psychological and behavioral pathways through which mental health interventions might affect economic outcomes. We first show that mental health treatments improve mental health outcomes, measures of overall functioning across domains life, and other psychological outcomes, suggesting a number of plausible mechanisms through which mental health interventions improve economic outcomes. Next, we show that in our study sample, changes in economic outcomes are well-correlated with changes in mental health and changes in functioning. Finally, leveraging individual-level data collected by a subset (n=7) of studies, we perform an instrumental variable analysis to estimate the elasticity of economic outcomes with respect to changes in mental health.

7.1 Treatment effects on psychological and behavioral mechanisms

In Table 4, we report findings of meta-analyses estimating Equation 1 on three broad categories of psychological and behavioral measures: symptoms of mental health conditions; measures of functioning and disability; and physical health and other psychological measures. Example wording of a representative sample of commonly reported mental health and functioning measures is provided in Tables A7-A10. At the aggregate level, we find large and highly statistically significant improvements in mental health disorder symptoms (0.22 SD), measures of functioning (0.27 SD), and measures of physical health and other outcomes (0.15 SD).

We then consider more disaggregated groupings of outcomes measured in similar ways. We label outcome groupings as they are measured in most studies, but outcomes are coded so higher values mean better mental health outcomes. Panel A examines measures of mental health. Treatments improve recovery and reduce relapse and rehospitalisation. They lead to positive, significant improvements in symptom severity scales for all disorders except PTSD, where effects are positive but noisily estimated, given the small number of interventions. Panel B examines measures of functioning. We find improvements in measures of functioning or disability overall. We also find improvements in measures capturing participants' ability to participate as normal in social interactions, and in measures of whether participants feel social support is available to them or if they face stigma and social isolation. Our

findings are consistent with the broader clinical evidence base, which finds mental health interventions are effective in reducing clinical symptoms of mental health conditions and associated functional impairment in LMIC (Singla et al., 2017, Patel et al., 2018). Panel C captures measures of physical health and other psychological outcomes. Treatment improves self-regulation and self-esteem/self-efficacy. There are positive, insignificant effects on cognition measures, which capture concentration, memory, and abstract reasoning. We do not have sufficient observations to conduct analysis in the next two sections for the outcomes in Panel C.

In Figure A4, we disaggregate these findings by the party responsible for measurement, showing outcomes measured by clinicians vs. self-rated by patients. The results are similar. If anything, effect sizes on clinician-rated outcomes are larger. Clinicians are usually blind to treatment status when performing assessments. 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 Elasticity of employment outcomes with respect to psychological and behavioral mechanisms

This section shows that, at the intervention level, changes in economic outcomes and in psychological and behavioral mechanisms in response to mental health interventions are correlated. This provides suggestive evidence of an economically meaningful elasticity of economic treatment effects with respect to mental health or functioning treatment effects. These parameters are policy-relevant, because they provides suggestive evidence on whether interventions to improve mental health and/or functioning are relevant ways to improve economic outcomes.

Figure 2 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.[26] 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. This relationship must be interpreted with caution as it does not adjust for relevant covariates identified in the meta-regression.

[26]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.

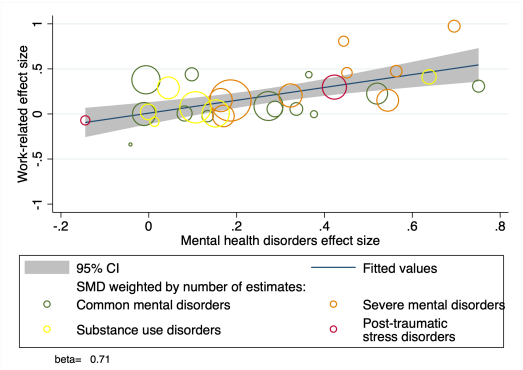
Table 4: Effects of mental health interventions on behavioral and psychological pathways

	(1) Aggregate Hedges' g	(2) 95% lower CI	(3) 95% upper CI	(4) # of observations	(5) # of interventions
Panel A: Measures of mental health disorder symptoms[1]					
All mental health disorder symptoms	0.22***	0.14	0.30	187	36
Suicide attempts or at risk of suicide	0.08	-0.04	0.20	18	10
Relapse (dummy)	0.28***	0.17	0.38	14	8
Recovery (dummy)	0.31***	0.16	0.46	4	4
Rehospitalisation	0.19*	-0.00	0.38	6	4
Diagnosed with mental disorder	0.31***	0.13	0.49	14	7
Substance use	0.17***	0.07	0.26	35	6
CMD symptoms	0.17***	0.07	0.26	60	23
PTSD symptoms	0.09	-0.14	0.32	13	6
SMD symptoms	0.36**	0.04	0.68	8	6
Overall assessment of mental disorder	-0.05	-0.32	0.23	10	7
Panel B: Measures of disability and functioning[1]					
All disability and functioning	0.27***	0.17	0.37	92	31
Overall measures of functioning	0.20***	0.14	0.26	56	25
Social support	0.25***	0.09	0.41	27	10
Functioning in social interactions	0.45***	0.19	0.71	9	5
Panel C: Measures of physical health and other outcomes[1]					
Self-regulation	0.15***	0.07	0.23	7	4
Self-esteem/self-efficacy	0.46**	0.01	0.91	7	5
Cognition	0.16	-0.06	0.37	11	5
Physical health	0.24	-0.08	0.57	10	4

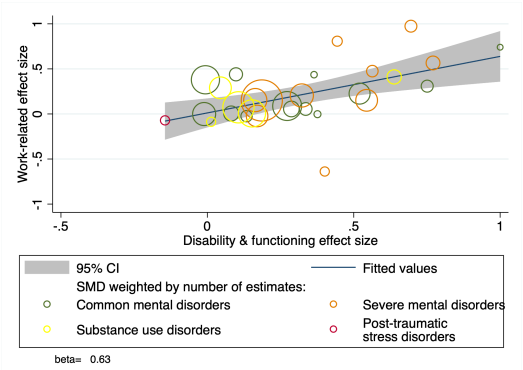
Notes: Hedges' g is the small-sample-bias-corrected standardized mean difference in the economic outcome between treatment and control. *, ** and *** denote statistical significance at 10 percent, 5 percent and 1 percent level of significance respectively. The average measurement in our sample happens 15.2 months after intervention start. [1] = all measures coded so that higher values mean better mental health outcomes. Aggregate Hedges' g represents an estimate from random effects inverse variance weighted meta-analysis. The aggregation of individual effect sizes works as described in subsection 4.1. All individual effect sizes are winsorised at the 95^{th} percentile.

Figure 2: Intervention-level correlations between economic outcomes and behavioral and psychological pathways at the individual level

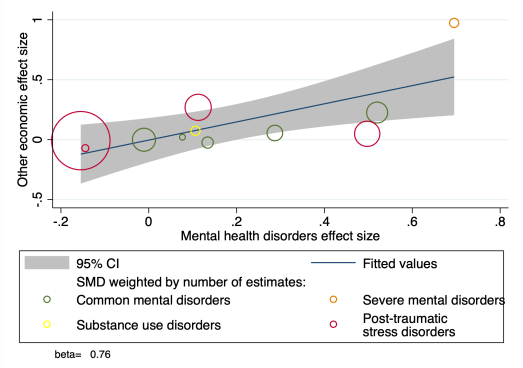
(a) Work-related effects and mental health disorders



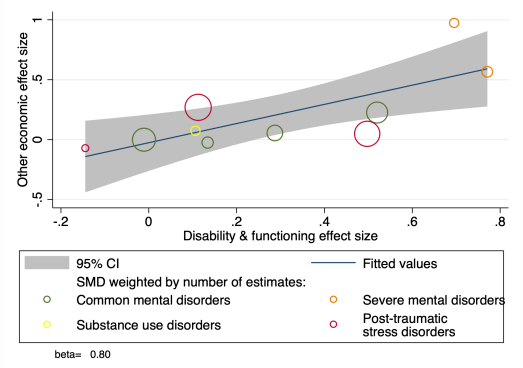
(b) Work-related effects and disability+functioning



(c) Other economic effects and mental health disorders



(d) Other economic effects and disability+functioning



These are four scatterplots of average 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 subsection 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.

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. Work-related treatment effects are well-correlated with both mental health disorder treatment effects ($\beta = 0.71$) and disability and functioning treatment effects ($\beta = 0.63$). Other economic outcomes are even more strongly correlated with each ($\beta = 0.76$, $\beta = 0.80$). This relationship is suggestive evidence that mental health and functioning are among the mechanisms through which mental health interventions affect economic outcomes. However, we cannot rule out that other mechanisms not mea-

sured in these studies contribute to economic effects. We also cannot rule out that economic outcomes are mechanisms for changes in mental health or functioning: for example, performing better at work may promote mental health.

7.3 Elasticity of days able to work measures to changes in depression

To explore this relationship further, we exploit individual-level variation in mental health induced by mental health interventions collected by a subset (n=7) of the included studies. These seven RCTs treated depression using cognitive behavioral therapy and measured similar employment outcomes, so face limited heterogeneity in study population or treatment. We pool individual-level data from these seven trials to conduct an econometric "mega-analysis" of the impact of depression treatments on both mental health symptoms and days participants are able to work, controlling for study-level characteristics.[27]

To conduct a mega-analysis of the pooled individual-level data from the included trials, we use economic outcomes that are present in each of the included studies to avoid standardizing the outcome measure (cf. Vivalt, 2020, Meager, 2019). These are: days unable to work and healthy days, enabling us to construct a measure of days able to work per month.[28] We employ instrumental variable regressions to approximate the impact of variation in mental health on economic outcomes. While the variation in mental health induced by treatment is plausibly exogeneous, we do not interpret the results causally, as there are other potential mediators on the path from intervention to economic impacts, such as functioning.

We instrument individual depression, MH_i, with random treatment assignment T_{is}. We estimate the first stage, Equation 3, with results shown in Table A16:

$$MH_i = \gamma_0 + \gamma_1 T_{is} + S_s + \varepsilon_{is}, \tag{3}$$

where MH_i is participant i's depression outcome and T_{is} is the randomly allocated indicator for whether i received CBT (as opposed to being in the control group) in study s. S_s is a study fixed effect. We estimate the elasticity via two-stage least squares:

[27]The seven studies are Patel et al. (2010), Fuhr et al. (2019), Sikander et al. (2019), Baranov et al. (2020), Meffert et al. (2021), Nadkarni et al. (2019), Barker et al. (2022). Data was either publicly available or shared by the authors upon request.

[28]Details on the variable construction can be found in Appendix F.

$$y_i = \beta_0 + \beta_1 \widehat{MH}_i + S_s + \varepsilon_{is}, \tag{4}$$

where y_i is participant i's days able to work measure, $\widehat{MH}_i$ is her depression outcome instrumented by CBT treatment, and ε_{is} is a participant-study specific error term. The 2SLS results are reported in Table 5. Treatment is a highly relevant instrument ($F > 20$ in the combined days able to work measure and days unable to work 2SLS procedures), which is commensurate with the distribution of effect sizes shown on the X-axis of Figure 2. However, we have some evidence of weak identification in estimation of healthy days, due to the substantially lower sample size ($F = 3.76$).

Table 5: Instrumenting the decrease in depression with random treatment allocation in the pooled sample

	Combined days able to work measure		Healthy days		Days unable to work	
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	IV	OLS	IV	OLS	IV
Depression	-2.437***	-3.784**	-0.638**	-2.989	2.689***	3.855**
	(0.188)	(1.658)	(0.269)	(4.442)	(0.198)	(1.761)
Constant	23.51***		26.55***		5.002***	
	(0.179)		(0.332)		(0.196)	
Wave FE	Yes	Yes	Yes	Yes	Yes	Yes
Study FE	Yes	Yes	No	No	Yes	Yes
Months after treatment	Yes	Yes	Yes	Yes	Yes	Yes
Control mean	22.40	22.40	26.27	26.27	6.18	6.18
Obs.	16884	16884	621	621	16263	16263
Studies	7	7	1	1	6	6
Underidentification		0.00		0.06		0.00
Weak identification		24.32		3.76		21.34

Notes: This table shows six different regression of the outcome variable (days worked per month) 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 a combined days worked per month outcome, columns 3-4 show the impact on healthy days per month, columns 5-6 show the impact on days unable to work in the last month. Standard errors are in parentheses and clustered by the original clustering unit of each study. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

We find that for a 1 SD decrease in depression, the number of days able to work per month increases by 3.78, while days unable to work falls by 3.86. Both results are statistically significant. These effects are sizable, reflecting 17% and 62% improvements compared to the control mean, respectively. We do not observe an effect of depression on healthy days, potentially due to the relatively small sample size (n=621) or weakness of the first stage. In sum, this mechanism analysis offers us tentative evidence that the impact of depression on economic outcomes is plausibly large, so it is a viable mechanism to target to improve economic outcomes.

8 Policy questions

8.1 How cost effective are mental health interventions?

In this section, we analyze the relative costs and benefits of mental health interventions conducted in LMICs compared to economic interventions conducted in the same contexts. We have cost data on 20 of the 39 interventions in this paper, either because it is publicly available or after contacting authors.

We report intervention and region-level per-participant average costs in Table A18. The average cost across all interventions is USD 363 per participant (2011 USD). There is heterogeneity in costs by intervention type and region. In our sample, psychosocial interventions are more costly than pharmacological or combined interventions. However, trials have a limited follow-up period, so cost estimates may not account for the fact that pharmacological interventions tend to be administered for longer periods than psychosocial interventions. Interventions in South Asia or East Asia & Pacific were cheaper than European & Central Asia or Sub-Saharan African programs. As a rough calculation, if we take the intervention average cost and the average impact of 0.15 standard deviations on economic outcomes, we find an improvement in economic outcomes of 0.04 SD for every USD 100 spent per participant.

In a next step, we analyze how these impacts per dollar spent compare to "traditional" economic interventions. We collect cost and economic impact data on active labor market policies (ALMP), unconditional cash transfers (UCT), and livelihood interventions targeting the ultra-poor (LLH), all in LMICs. For ALMPs, we use the sample in McKenzie (2017); for LLH, interventions from Banerjee et al. (2015). The three UCT interventions are De Mel et al. (2008), Haushofer and Shapiro (2016), Blattman et al. (2017). The average per-participant costs of all three types of economic interventions are equal or higher than the costs of the mental health interventions in our sample: the mean costs of the three UCTs are USD 358 per person, the 15 ALMPs on average cost USD 566 per participant, and the six LLH interventions cost on average USD 1,468 per participant. In contrast, the economic impacts are only 60% of those of mental health interventions: UCTs improve economic outcomes by 0.09 SD, ALMPS by 0.06 SD, and LLHs by 0.09 SD. For every USD 100 spent per participant, UCTs have an average impact of 0.025 SD, ALMPs of 0.011 SD, and LLHs of 0.006 SD. Per participant dollar spent, these impacts are smaller than the average improvement in economic outcomes of mental health interventions.

This comparison of mental health and economic interventions is imperfect. The populations targeted for intervention and the outcomes measured are different. The samples of economic programs for which cost data are available are selected and may be particularly expensive. But these crude comparisons underline the cost-effectiveness of mental health interventions.

8.2 Are mental health and economic interventions complementary?

In our final piece of analysis, we explore positive interactions between mental health and economic interventions. Five RCTs identified during our search contain additional treatment arms which combine mental health and economic interventions, which we excluded in all previous analysis. Each combines a psychosocial treatment with an economic intervention, either a cash transfer (in one case including financial literacy training), job search assistance, or a skill-building intervention.

We re-analyze our sample with the five additional interventions. In Table A19, we reproduce Table 2 for the sample of 44 interventions, which now include the five psychosocial plus economic interventions. We note two main results: First, mental health plus economic interventions are more effective than mental health interventions alone and are as effective as combined psychosocial and pharmacological interventions, with an average effect size of 0.29 SDs on all economic outcomes. Second, the effectiveness of mental health plus economic interventions is driven by impacts on employment participation and, to a lesser extent, non-employment outcomes, whereas the effects are zero for time in work.

This analysis, while exploratory, suggests that interventions combining mental health and economic components have larger effects on economic outcomes than psychosocial or pharmacological interventions on their own. This is in some ways unsurprising, as direct economic interventions are likely to improve economic outcomes more than simply treating a mental health condition, and is consistent with causality in both directions in the relationship between mental health and economic outcomes. It highlights potential gains from multi-faceted programs for people with mental health conditions. Furthermore, combined mental health interventions have the same effect on economic outcomes as interventions including both an economic and mental health component, and combined mental health treatments are likely to be considerably cheaper.[29] As discussed in Section 6, combined interventions are

[29]Compare 550 USD per participant for a CBT + cash grant intervention versus 96 USD for a psychoeducation + antidepressant treatment.

often targeted at people with more severe mental health conditions. For these populations, this analysis highlights the importance of mental health service provision.

9 Conclusion

Our study presents evidence from a large body of research on the economic impacts of 39 mental health interventions in 15 LMIC. Most interventions were psychosocial or combined psychosocial and pharmacological treatments. We find that mental health interventions significantly improve work-related outcomes, such as whether participants are unable to work, days they are unable to work, and qualitative measures of whether participants are functioning normally at work. We also see significant improvements in measures of asset wealth and education. There is some heterogeneity in these findings: economic effects are somewhat larger for populations with severe than mild disorders and for combination treatments than for drugs or therapy alone, consistent with clinical evidence. Findings are robust across a range of robustness checks. We find little evidence of publication bias.

We find that these economic effects occur alongside reductions in mental health symptoms and improvements in functioning, which may act as mechanisms. In the studies in our sample, treating a mental health condition promotes recovery, reduces hospitalisation, alleviates symptoms and improves overall functioning in domains outside work. We provide suggestive evidence that the economic improvements are associated with changes in mental health and functioning, by showing that clinically more effective interventions have stronger economic effects.

Our findings suggest a number of directions for future research. Like any meta-analysis, we can only draw conclusions for outcomes and interventions measured by other studies. Many papers in our sample do not use standard, policy-relevant, economic outcome measures, for example of labor supply, earnings or consumption expenditure. Furthermore, many potential mechanisms highlighted by the literature are not measured. For example, it has been hypothesized that clinical improvements yield changes in cognitive and affective styles, such an increased future orientation in economic decision-making, or a more realistic appraisal of financial options rather than attention to threat (Haushofer and Fehr, 2014). Our findings suggest further research studying the effects of existing mental health treatments, either additional survey rounds to follow up mental health intervention trials (as in Baranov et al. (2020) and Bhat et al. (2022)), or large trials of mental health treatments powered to detect economic effects (as in Angelucci and Bennett (2022), Barker et al. (2022)

and Blattman et al. (2017)). 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 other can be more effective than interventions on their own (Angelucci and Bennett, 2022, Blattman et al., 2017) and that the mentorship and handholding components of intensive livelihood programmes 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 (Lund, 2018, Patel et al., 2018). Policy-makers and international agencies focused on economic development have tended to overlook the importance of mental health. People with a mental health condition make up a substantial portion – 11 to 18% – of the general population (IHME, 2017). Existing, cost-effective interventions targeting their mental health conditions both alleviate symptoms and improve their ability to generate a livelihood, accumulate assets, and invest in their children's education. Further investment in these interventions is an urgent global priority.

References

Andrews, Isaiah and Maximilian Kasy, "Identification of and correction for publication bias," *American Economic Review*, August 2019, *109* (8), 2766–94.

Angelucci, Manuela and Daniel Bennett, "The economic impact of depression treatment in India: Evidence from community-based provision of pharmacotherapy," Working Paper 2022.

Banerjee, Abhijit, Dean Karlan, Darko Osei, Hannah Trachtman, and Christopher Udry, "Unpacking a multi-faceted program to build sustainable income for the very poor," *National Bureau of Economic Research Working Paper*, 2018, *24271*.

_ , Esther Duflo, Nathanael Goldberg, Dean Karlan, Robert Osei, William Parienté, Jeremy Shapiro, Bram Thuysbaert, and Christopher Udry, "A multifaceted program causes lasting progress for the very poor: Evidence from six countries," *Science*, 2015, *348* (6236).

Banerjee, Souvik, Pinka Chatterji, and Kajal Lahiri, "Effects of psychiatric disorders on labor market outcomes: a latent variable approach using multiple clinical indicators," *Health Economics*, 2017, *26* (2), 184–205.

Baranov, Victoria, Sonia Bhalotra, Pietro Biroli, and Joanna Maselko, "Maternal depression, women's empowerment, and parental investment: evidence from a randomized controlled trial," *American Economic Review*, 2020, *110* (3), 824–59.

Barker, Nathan, Gharad Bryan, Dean Karlan, Angela Ofori-Atta, and Christopher Udry, "Cognitive Behavioral Therapy among Ghana's Rural Poor Is Effective Regardless of Baseline Mental Distress," *American Economic Review: Insights*, December 2022, *4* (4), 527–45.

Bhat, Bhargav, Jon de Quidt, Johannes Haushofer, Vikram Patel, Gautam Rao, Frank Schilbach, and Pierre Vautrey, "The long-run effects of psychotherapy on depression, beliefs, and economic outcomes," *National Bureau of Economic Research Working Paper*, 2022, *30011*.

Biasi, Barbara, Michael Dahl, and Petra Moser, "Career effects of mental health," Technical Report, National Bureau of Economic Research 2021.

Blattman, Christopher, Julian C. Jamison, and Margaret Sheridan, "Reducing crime and violence: Experimental evidence from Cognitive Behavioral Therapy in Liberia," *American Economic Review*, April 2017, *107* (4), 1165–1206.

Bloom, David E., Elizabeth Cafiero, Eva Jané-Llopis, Shafika Abrahams-Gessel, Lakshmi Reddy Bloom, Sana Fathima, Andrea B. Feigl, Tom Gaziano, Ali Hamandi, Mona Mowafi, Danny O'Farrell, and Emre, "The Global Economic Burden of Noncommunicable Diseases," Technical Report, World Economic Forum, Geneva 2011.

Borenstein, Michael, H Cooper, L Hedges, and J Valentine, "Effect sizes for continuous data," *The handbook of research synthesis and meta-analysis*, 2009, *2*, 221–235.

_ , Larry V Hedges, Julian PT Higgins, and Hannah R Rothstein, *Introduction to meta-analysis*, John Wiley & Sons, 2011.

Bütikofer, Aline, Christopher J. Cronin, and Meghan M. Skira, "Employment effects of healthcare policy: Evidence from the 2007 FDA black box warning on antidepressants," *Journal of Health Economics*, 2020, *73*, 102348.

Card, David, Jochen Kluve, and Andrea Weber, "What works? A meta analysis of recent active labor market program evaluations," *Journal of the European Economic Association*, 2018, *16* (3), 894–931.

Chan, Joyce YC, Hoyee W Hirai, and Kelvin KF Tsoi, "Can computer-assisted cognitive remediation improve employment and productivity outcomes of patients with severe mental illness? A meta-analysis of prospective controlled trials," *Journal of Psychiatric Research*, 2015, *68*, 293–300.

Chatterji, Pinka, Margarita Alegria, and David Takeuchi, "Psychiatric disorders and labor market outcomes: Evidence from the National Comorbidity Survey-Replication," *Journal of Health Economics*, 2011, *30* (5), 858–868.

Cipriani, Andrea, Toshi A Furukawa, Georgia Salanti, Anna Chaimani, Lauren Z Atkinson, Yusuke Ogawa, Stefan Leucht, Henricus G Ruhe, Erick H Turner, Julian Higgins et al., "Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: A systematic review and network meta-analysis.," *The Lancet*, 2018.

Cleary, Michelle, Glenn E Hunt, Sandra L Matheson, Nandi Siegfried, and Garry Walter, "Psychosocial interventions for people with both severe mental illness and substance misuse," *Cochrane database of systematic reviews*, 2008, (1).

Cohen, Jacob, *Statistical power analysis for the behavioral sciences 2nd edn*, Erlbaum Associates, Hillsdale, 1988.

Cuijpers, Pim, Annemieke Van Straten, Gerhard Andersson, and Patricia Van Oppen, "Psychotherapy for depression in adults: a meta-analysis of comparative outcome studies.," *Journal of consulting and clinical psychology*, 2008, *76* (6), 909.

_ , Eirini Karyotaki, Mirjam Reijnders, Marianna Purgato, and Corrado Barbui, "Psychotherapies for depression in low-and middle-income countries: a meta-analysis," *World Psychiatry*, 2018, *17* (1), 90–101.

_ , Marit Sijbrandij, Sander L Koole, Gerhard Andersson, Aartjan T Beekman, and Charles F Reynolds III, "The efficacy of psychotherapy and pharmacotherapy in treating depressive and anxiety disorders: A meta-analysis of direct comparisons," *World Psychiatry*, 2013, *12* (2), 137–148.

Edlund, Mark J., Jiantong Wang, Kristen Gulledge Brown, Valerie L. Forman-Hoffman, Sara L. Calvin, Sarra L. Hedden, and Jonaki Bose, "Which mental disorders are associated with the greatest impairment in functioning?," *Social Psychiatry and Psychiatric Epidemiology*, Nov 2018, *53* (11), 1265–1276.

Egger, Matthias, George Davey Smith, Martin Schneider, and Christoph Minder, "Bias in meta-analysis detected by a simple, graphical test," *BMJ*, 1997, *315* (7109), 629–634.

Finnes, Anna, Pia Enebrink, Ata Ghaderi, JoAnne Dahl, Anna Nager, and Lars-Göran Öst, "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*, 2019, *92* (3), 273–293.

Gelman, Andrew and Jennifer Hill, *Data analysis Using regression and multilevel/hierarchical models* Analytical methods for social research, Cambridge University Press, 2006.

_ , John B. Carlin, Hal S. Stern, and Donald B. Rubin, *Bayesian data analysis 2nd ed.*, Abingdon: Taylor and Francis, 2009.

Gleser, Leon J. and Ingram Olkin, "Stochastically dependent effect sizes," in J. C. Valentine H. Cooper, L. B. Hedges, ed., *The handbook of research synthesis and meta-analysis*, Russell Sage Foundation, 2009, pp. 357–376.

Haushofer, Johannes and Ernst Fehr, "On the psychology of poverty," *Science*, 2014, *344* (6186), 862–867.

__ **and Jeremy Shapiro**, "The short-term impact of unconditional cash transfers to the poor: experimental evidence from Kenya," *The Quarterly Journal of Economics*, 2016, *131* (4), 1973–2042.

Hedges, Larry V., "Distribution theory for Glass's estimator of effect size and related estimators," *Journal of Educational Statistics*, 1981, *6* (2), 107–128.

IHME, "Findings from the Global Burden of Disease Study 2017," Technical Report, Institute for Health Metrics and Evaluation, Seattle 2017.

Jackson, Dan, Richard Riley, and Ian R. White, "Multivariate meta-analysis: Potential and promise," *Statistics in Medicine*, 2011, *30* (20), 2481–2498.

Jones, Kristine and Sophie Mitra, "The impact of recent mental health changes on employment: new evidence from longitudinal data," *Applied Economics*, 2017, *49* (1), 96–109.

Kamenov, K, C Twomey, M Cabello, A M Prina, and J L Ayuso-Mateos, "The efficacy of psychotherapy, pharmacotherapy and their combination on functioning and quality of life in depression: a meta-analysis," *Psychological Medicine*, 2017, *47* (3), 414–425.

Kessler, Ronald C. and Richard G. Frank, "The impact of psychiatric disorders on work loss days," *Psychological Medicine*, 1997, *27* (4), 861–873.

Leucht, Stefan, Andrea Cipriani, Loukia Spineli, Dimitris Mavridis, Deniz Örey, Franziska Richter, Myrto Samara, Corrado Barbui, Rolf R Engel, John R Geddes et al., "Comparative efficacy and tolerability of 15 antipsychotic drugs in schizophrenia: a multiple-treatments meta-analysis," *The Lancet*, 2013, *382* (9896), 951–962.

Lund, Crick, "Mental health in Africa: A neglected link in the development chain," *Africa Policy Review Journal*, 2018.

__ **, Mary De Silva, Sophie Plagerson, Sara Cooper, Dan Chisholm, Jishnu Das, Martin Knapp, and Vikram Patel**, "Poverty and mental disorders: breaking the cycle in low-income and middle-income countries," *The Lancet*, 2011, *378* (9801), 1502 – 1514.

McKenzie, David, "How effective are Active Labor Market Policies (ALMP) in developing countries? A critical review of recent evidence," *The World Bank Research Observer*, 2017, *32* (2), 127–154.

Meager, Rachael, "Understanding the average impact of microcredit expansions: A Bayesian hierarchical analysis of seven randomized experiments," *American Economic Journal: Applied Economics*, January 2019, *11* (1), 57–91.

Mel, Suresh De, David McKenzie, and Christopher Woodruff, "Returns to capital in microenterprises: evidence from a field experiment," *The quarterly journal of Economics*, 2008, *123* (4), 1329–1372.

Nadkarni, Abhijit, Benedict Weobong, Helen A Weiss, Jim McCambridge, Bhargav Bhat, Basavaraj Katti, Pratima Murthy, Michael King, David McDaid, A-La Park, G Terence Wilson, Betty Kirkwood, Christopher G Fairburn, Richard Velleman, and Vikram Patel, "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*, 2017, *389* (10065), 186–195.

_ , Helen A Weiss, Benedict Weobong, David McDaid, Daisy R Singla, A-La Park, Bhargav Bhat, Basavaraj Katti, Jim McCambridge, Pratima Murthy, Michael King, G Terence Wilson, Betty Kirkwood, Christopher G Fairburn, Richard Velleman, and Vikram Patel, "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*, 2017, *14* (9), 1–20.

_ , _ , Richard Velleman, Jim McCambridge, David McDaid, A-La Park, Pratima Murthy, Benedict Weobong, Bhargav Bhat, and Vikram Patel, "Feasibility, acceptability and cost-effectiveness of a brief, lay counsellor-delivered psychological treatment for men with alcohol dependence in primary care: an exploratory randomized controlled trial," *Addiction*, 2019, *114* (7), 1192–1203.

Nieuwenhuijsen, Karen, Jos H Verbeek, Angela Neumeyer-Gromen, Arco C Verhoeven, Ute Bültmann, and Babs Faber, "Interventions to improve return to work in depressed people," *Cochrane Database of Systematic Reviews*, 2020, (10), 1–160.

Patel, Vikram, Dan Chisholm, Tarun Dua, Ramanan Laxminarayan, and María Elena Medina-Mora, *Mental, neurological, and substance use disorders: Disease control priorities, (Volume 4)*, The International Bank for Reconstruction and Development/The World Bank, 2016.

_ , Shekhar Saxena, Crick Lund, Graham Thornicroft, Florence Baingana, Paul Bolton, Dan Chisholm, Pamela Y Collins, Janice L Cooper, Julian Eaton et al., "The Lancet Commission on global mental health and sustainable development," *The Lancet*, 2018, *392* (10157), 1553–1598.

Peng, Lizhong, Chad D Meyerhoefer, and Samuel H Zuvekas, "The short-term effect of depressive symptoms on labor market outcomes," *Health Economics*, 2016, *25* (10), 1223–1238.

Pustejovsky, James E. and Melissa A. Rodgers, "Testing for funnel plot asymmetry of standardized mean differences," *Research Synthesis Methods*, 2019, *10* (1), 57–71.

Ran, Mao-Sheng, Meng-Ze Xiang, Cecilia Lai-Wan Chan, Julian Leff, Peggy Simpson, Ming-Sheng Huang, You-He Shan, and Si-Gan Li, "Effectiveness of psychoeducational intervention for rural Chinese families experiencing schizophrenia," *Social Psychiatry and Psychiatric Epidemiology*, 2003, *38* (2), 69–75.

Ridley, Matthew, Gautam Rao, Frank Schilbach, and Vikram Patel, "Poverty, depression, and anxiety: Causal evidence and mechanisms," *Science*, 2020, *370* (6522).

Rodgers, Melissa and James Pustejovsky, "Evaluating meta-analytic methods to detect selective reporting in the presence of dependent effect sizes," *Psychological Methods*, 07 2020, *26*.

Rubin, Donald B., "Estimation in parallel randomized experiments," *Journal of Educational Statistics*, 1981, *6* (4), 377–401.

Salomonsson, Sigrid, Erik Hedman-Lagerlöf, and Lars-Göran Öst, "Sickness absence: a systematic review and meta-analysis of psychological treatments for individuals on sick leave due to common mental disorders," *Psychological Medicine*, 2018, *48* (12), 1954–1965.

Silva, Mary J De, Sara Cooper, Henry Lishi Li, Crick Lund, and Vikram Patel, "Effect of psychosocial interventions on social functioning in depression and schizophrenia: meta-analysis," *The British journal of psychiatry*, 2013, *202* (4), 253–260.

Singla, Daisy R, Brandon A Kohrt, Laura K Murray, Arpita Anand, Bruce F Chorpita, and Vikram Patel, "Psychological treatments for the world: lessons from low-and middle-income countries," *Annual Review of Clinical Psychology*, 2017, *13*, 149–181.

Tan, Brandon and Michael Kremer, "Water treatment and child survival: A meta-analysis," *Unpublished manuscript: Harvard University*, 2020.

Timbie, Justin, Marcela Horvitz-Lennon, Richard G Frank, and Sharon-Lise T Normand, "A meta-analysis of labor supply effects of interventions for major depressive disorder," *Psychiatric Services*, 2006, *57* (2), 212–218.

van Duin, Daniëlle, Lars de Winter, Matthijs Oud, Hans Kroon, Wim Veling, and Jaap van Weeghel, "The effect of rehabilitation combined with cognitive remediation on functioning in persons with severe mental illness: systematic review and meta-analysis," *Psychological Medicine*, 2019, *49* (9), 1414–1425.

Vivalt, Eva, "How much can we generalize from impact evaluations?," *Journal of the European Economic Association*, 2020.

WHO, "International classification of diseases (ICD)," *http://www.who.int/classifications/icd/en/*, 2016.

Supplementary Appendix
For Online Publication

A Search Strategy for PubMed

Searched: 19 April 2022.
Number of articles retrieved: 6585.

Population

((((((((((((((((((((((((((((((((((”Affective Symptoms”[Mesh]) OR ”Aggression”[Majr:noexp]) OR ”Alcohol Drinking”[Mesh]) OR ”Anxiety”[Mesh]) OR ”Depression”[Mesh]) OR ”Diagnostic and Statistical Manual of Mental Disorders”[Mesh]) OR ”Epilepsy”[Mesh]) OR ”Impulsive Behavior”[Mesh]) OR ”Irritable Mood”[Mesh]) OR ”Mental Disorders”[Mesh]) OR ”Mental Fatigue”[Mesh]) OR ”Mentally Ill Persons”[Mesh]) OR ”Paranoid Behavior”[Mesh]) OR ”Problem Behavior”[Mesh]) OR ”Psychophysiologic Disorders”[Mesh]) OR ”Self-Injurious Behavior”[Mesh]) OR ”Stress, Psychological”[Mesh]) OR ”Mental Health”[Mesh])) OR (((Addiction[Text Word]) OR Alcohol abuse[Text Word]) OR Alcohol use Disorder[Text Word]) OR Anxiety[Text Word]) OR Attention Deficit disorder[Text Word]) OR Bipolar[Text Word]) OR Child Behaviour Disorder[Text Word]) OR Child Development Disorder[Text Word]) OR Conduct Disorder[Text Word]) OR Depression[Text Word]) OR Depressive Disorder[Text Word]) OR Developmental Disability[Text Word]) OR Emotional stress[Text Word]) OR Epilepsies[Text Word]) OR Epilepsy[Text Word]) OR Mania[Text Word]) OR Mental disorder[Text Word]) OR Mental illness[Text Word]) OR Mental stress[Text Word]) OR Mentally ill[Text Word]) OR Mood Disorder[Text Word]) OR Neurodevelopmental Disorder[Text Word]) OR Personality Disorder[Text Word]) OR Posttraumatic stress disorder[Text Word]) OR Post-traumatic stress disorder[Text Word]) OR Psychiatric[Text Word]) OR Psychological distress[Text Word]) OR Psychological stress[Text Word]) OR Psychological trauma[Text Word]) OR Psychosis[Text Word]) OR Psychoses[Text Word]) OR Psychotic[Text Word]) OR psychosomatic[Text Word]) OR Schizophrenia[Text Word]) OR Schizophrenic[Text Word]) OR Somatisation[Text Word]) OR Somatoform[Text Word]) OR Stress Disorder[Text Word]) OR Substance abuse[Text Word]) OR Substance Disorder[Text Word]) OR Substance Withdrawal Syndrome[Text Word]) OR Suicidal[Text Word]) OR Traumatic stress disorder[Text Word])) OR street drugs[Text Word]) OR common mental disorders[Text Word])) OR Mental disorders[Text Word]) OR Mental Health[Text Word]))) OR cocaine[Text Word]) OR heroin[Text Word])) OR Aggression[Text Word]) OR Paranoid Behavior[Text Word]) OR Problem Behavior[Text Word])) AND

Intervention

(((((((((((((((((((((((((”Caregivers”[Mesh]) OR ”Combined Modality Therapy”[Mesh]) OR ”Community Health Workers”[Mesh]) OR ”Comprehensive Health Care”[Mesh]) OR ”Health Education”[Majr:noexp]) OR ”Hospitals, Psychiatric”[Mesh]) OR ”Hospitals, University”[Mesh]) OR ”Interview, Psychological”[Mesh]) OR ”Mental Health Services”[Mesh]) OR ”Psychiatric Rehabilitation”[Mesh]) OR ”Psychiatric Somatic Therapies”[Mesh]) OR ”Psychoanalytic Interpretation”[Mesh]) OR ”Psychological Techniques”[Mesh]) OR ”Psychopharmacology”[Mesh]) OR ”Psychotherapy”[Mesh]) OR ”Psychotropic Drugs”[Mesh]) OR ”Psychotropic Drugs”[Pharmacological Action]) OR ”Rehabilitation Centers”[Mesh]) OR ”Rehabilitation, Vocational”[Mesh]) OR ”Self-Help Groups”[Mesh]) OR ”Ambulatory Care Facilities”[Mesh])) OR ((((((((((((((((((((((((((((((((((((((((((((((((Therapy[Text Word] OR psychotherapy[Text Word]) OR mental health care[Text Word]) OR mental health services[Text Word]) OR mental health service[Text Word]) OR psychotherapeutic[Text Word]) OR addiction recovery[Text Word]) OR counselling[Text

Word]) OR counseling[Text Word]) OR behavioral activation[Text Word]) OR behavioural activation[Text Word]) OR behavioural management[Text Word]) OR behavioral management[Text Word]) OR contingency management[Text Word]) OR epilepsy surgery[Text Word]) OR eye movement desensitization[Text Word]) OR hypnosis[Text Word]) OR prolonged exposure[Text Word]) OR locus of control[Text Word]) OR mindfulness[Text Word]) OR motivational interview[Text Word]) OR neurolinguistic programming[Text Word]) OR pastoral care[Text Word]) OR psychoanalytic[Text Word]) OR psychodynamic[Text Word]) OR psychoeducation[Text Word]) OR psychological feedback[Text Word]) OR psychological interview[Text Word]) OR relaxation training[Text Word]) OR self affirmation[Text Word]) OR supported employment[Text Word]) OR psychotropic drugs[Text Word]) OR psychotropic drug[Text Word]) OR psychopharmacology[Text Word]) OR psychotropic[Text Word]) OR antipsychotic[Text Word]) OR antidepressant[Text Word]) OR anxiolytic[Text Word]) OR antianxiety[Text Word]) OR mood stabilizer[Text Word]) OR mood stabiliser[Text Word]) OR treatment[Text Word]) OR intervention[Text Word]) OR CBT[Text Word]) OR MBCT[Text Word])) OR ((((((((((cognitive therapy[Text Word]) OR interpersonal[Text Word]) OR Rehabilitation[Text Word]) OR psychiatric hospital[Text Word]) OR drug treatment[Text Word]) OR psychological treatment[Text Word]) OR psychoeducational[Text Word]) OR Family-based[Text Word]) OR solution-focused[Text Word]) OR brief[Title])))) OR ((((((((((((((carers[Text Word]) OR caregivers[Text Word]) OR community-based[Text Word]) OR behaviour therapy[Text Word]) OR behavior therapy[Text Word]) OR encounter groups[Text Word]) OR preventive health care[Text Word]) OR preventive programs[Text Word]) OR preventive programmes[Text Word]) OR preventive health service[Text Word]) OR mental hospital[Text Word]) OR mental institution[Text Word]) OR mental health education[Text Word]) OR sensitivity training[Text Word]) OR teaching hospital[Text Word])))))) AND

Location

((((Africa[Title/Abstract] OR Asia[Title/Abstract] OR Caribbean[Title/Abstract] OR West Indies[Title/Abstract] OR South America[Title/Abstract] OR Latin America[Title/Abstract] OR Central America[Title/Abstract] OR "Atlantic Islands"[Title/Abstract] OR "Commonwealth of Independent States"[Title/Abstract] OR "Pacific Islands"[Title/Abstract] OR "Indian Ocean Islands"[Title/Abstract] OR "Eastern Europe"[Title/Abstract] OR Afghanistan[Title/Abstract] OR Albania[Title/Abstract] OR Algeria[Title/Abstract] OR Angola[Title/Abstract] OR Antigua[Title/Abstract] OR Barbuda[Title/Abstract] OR Argentina[Title/Abstract] OR Armenia[Title/Abstract] OR Armenian[Title/Abstract] OR Aruba[Title/Abstract] OR Azerbaijan[Title/Abstract] OR Bahrain[Title/Abstract] OR Bangladesh[Title/Abstract] OR Barbados[Title/Abstract] OR Benin[Title/Abstract] OR Byelarus[Title/Abstract] OR Byelorussian[Title/Abstract] OR Belarus[Title/Abstract] OR Belorussian[Title/Abstract] OR Belorussia[Title/Abstract] OR Belize[Title/Abstract] OR Bhutan[Title/Abstract] OR Bolivia[Title/Abstract] OR Bosnia[Title/Abstract] OR Herzegovina[Title/Abstract] OR Hercegovina[Title/Abstract] OR Botswana[Title/Abstract] OR Brasil[Title/Abstract] OR Brazil[Title/Abstract] OR Bulgaria[Title/Abstract] OR Burkina Faso[Title/Abstract] OR Burkina Fasso[Title/Abstract] OR Upper Volta[Title/Abstract] OR Burundi[Title/Abstract] OR Urundi[Title/Abstract] OR Cambodia[Title/Abstract] OR Khmer Republic[Title/Abstract] OR Kampuchea[Title/Abstract] OR Cameroon[Title/Abstract] OR Cameroons[Title/Abstract] OR Cameron[Title/Abstract] OR Cameroun[Title/Abstract] OR Cape Verde[Title/Abstract] OR Central African Republic[Title/Abstract] OR Chad[Title/Abstract] OR Chile[Title/Abstract] OR China[Title/Abstract] OR Colombia[Title/Abstract] OR Comoros[Title/Abstract] OR Comoro Islands[Title/Abstract] OR Comores[Title/Abstract] OR Mayotte[Title/Abstract] OR Congo[Title/Abstract] OR Zaire[Title/Abstract] OR Costa Rica[Title/Abstract] OR Cote d'Ivoire[Title/Abstract] OR Ivory Coast[Title/Abstract] OR Croatia[Title/Abstract] OR Cuba[Title/Abstract] OR Cyprus[Title/Abstract] OR Czechoslovakia[Title/Abstract] OR Czech Republic[Title/Abstract] OR Slovakia[Title/Abstract] OR Slovak Republic[Title/Abstract] OR Djibouti[Title/Abstract] OR French Somaliland[Title/Abstract] OR Dominica[Title/Abstract] OR Dominican Republic[Title/Abstract] OR East Timor[Title/Abstract] OR East Timur[Title/Abstract] OR Timor Leste[Title/Abstract] OR Ecuador[Title/Abstract] OR Egypt[Title/Abstract] OR United Arab Republic[Title/Abstract] OR El Salvador[Title/Abstract] OR Eritrea[Title/Abstract] OR Estonia[Title/Abstract] OR Ethiopia[Title/Abstract] OR Fiji[Title/Abstract] OR Gabon[Title/Abstract] OR Gabonese Republic[Title/Abstract] OR Gambia[Title/Abstract] OR Gaza[Title/Abstract] OR Georgia Republic[Title/Abstract] OR Georgian Re-

public[Title/Abstract] OR Ghana[Title/Abstract] OR Gold Coast[Title/Abstract] OR Greece[Title/Abstract] OR Grenada[Title/Abstract] OR Guatemala[Title/Abstract] OR Guinea[Title/Abstract] OR Guam[Title/Abstract] OR Guiana[Title/Abstract] OR Guyana[Title/Abstract] OR Haiti[Title/Abstract] OR Honduras[Title/Abstract] OR Hungary[Title/Abstract] OR India[Title/Abstract] OR Maldives[Title/Abstract] OR Indonesia[Title/Abstract] OR Iran[Title/Abstract] OR Iraq[Title/Abstract] OR Isle of Man[Title/Abstract] OR Jamaica[Title/Abstract] OR Jordan[Title/Abstract] OR Kazakhstan[Title/Abstract] OR Kazakh[Title/Abstract] OR Kenya[Title/Abstract] OR Kiribati[Title/Abstract] OR Korea[Title/Abstract] OR Kosovo[Title/Abstract] OR Kyrgyzstan[Title/Abstract] OR Kirghizia[Title/Abstract] OR Kyrgyz Republic[Title/Abstract] OR Kirghiz[Title/Abstract] OR Kirgizstan[Title/Abstract] OR "Lao PDR"[Title/Abstract] OR Laos[Title/Abstract] OR Latvia[Title/Abstract] OR Lebanon[Title/Abstract] OR Lesotho[Title/Abstract] OR Basutoland[Title/Abstract] OR Liberia[Title/Abstract] OR Libya[Title/Abstract] OR Lithuania[Title/Abstract] OR Macedonia[Title/Abstract] OR Madagascar[Title/Abstract] OR Malagasy Republic[Title/Abstract] OR Malaysia[Title/Abstract] OR Malaya[Title/Abstract] OR Malay[Title/Abstract] OR Sabah[Title/Abstract] OR Sarawak[Title/Abstract] OR Malawi[Title/Abstract] OR Nyasaland[Title/Abstract] OR Mali[Title/Abstract] OR Malta[Title/Abstract] OR Marshall Islands[Title/Abstract] OR Mauritania[Title/Abstract] OR Mauritius[Title/Abstract] OR Agalega Islands[Title/Abstract] OR "Melanesia"[Title/Abstract] OR Mexico[Title/Abstract] OR Micronesia[Title/Abstract] OR Middle East[Title/Abstract] OR Moldova[Title/Abstract] OR Moldovia[Title/Abstract] OR Moldovian[Title/Abstract] OR Mongolia[Title/Abstract] OR Montenegro[Title/Abstract] OR Morocco[Title/Abstract] OR Ifni[Title/Abstract] OR Mozambique[Title/Abstract] OR Myanmar[Title/Abstract] OR Myanma[Title/Abstract] OR Burma[Title/Abstract] OR Namibia[Title/Abstract] OR Nepal[Title/Abstract] OR Netherlands Antilles[Title/Abstract] OR New Caledonia[Title/Abstract] OR Nicaragua[Title/Abstract] OR Niger[Title/Abstract] OR Nigeria[Title/Abstract] OR Northern Mariana Islands[Title/Abstract] OR Oman[Title/Abstract] OR Muscat[Title/Abstract] OR Pakistan[Title/Abstract] OR Palau[Title/Abstract] OR Palestine[Title/Abstract] OR Panama[Title/Abstract] OR Paraguay[Title/Abstract] OR Peru[Title/Abstract] OR Philippines[Title/Abstract] OR Philipines[Title/Abstract] OR Phillipines[Title/Abstract] OR Phillippines[Title/Abstract] OR Poland[Title/Abstract] OR Portugal[Title/Abstract] OR Puerto Rico[Title/Abstract] OR Romania[Title/Abstract] OR Rumania[Title/Abstract] OR Roumania[Title/Abstract] OR Russia[Title/Abstract] OR Russian[Title/Abstract] OR Rwanda[Title/Abstract] OR Ruanda[Title/Abstract] OR Saint Kitts[Title/Abstract] OR St Kitts[Title/Abstract] OR Nevis[Title/Abstract] OR Saint Lucia[Title/Abstract] OR St Lucia[Title/Abstract] OR Saint Vincent[Title/Abstract] OR St Vincent[Title/Abstract] OR Grenadines[Title/Abstract] OR Samoa[Title/Abstract] OR Samoan Islands[Title/Abstract] OR Navigator Island[Title/Abstract] OR Navigator Islands[Title/Abstract] OR Sao Tome[Title/Abstract] OR Saudi Arabia[Title/Abstract] OR Senegal[Title/Abstract] OR Serbia[Title/Abstract] OR Montenegro[Title/Abstract] OR Seychelles[Title/Abstract] OR Sierra Leone[Title/Abstract] OR Slovenia[Title/Abstract] OR Sri Lanka[Title/Abstract] OR Ceylon[Title/Abstract] OR Solomon Islands[Title/Abstract] OR Somalia[Title/Abstract] OR Sudan[Title/Abstract] OR South Sudan[Title/Abstract] OR South Africa[Title/Abstract] OR Suriname[Title/Abstract] OR Surinam[Title/Abstract] OR Swaziland[Title/Abstract] OR Syria[Title/Abstract] OR Syrian[Title/Abstract] OR Tajikistan[Title/Abstract] OR Tadzhikistan[Title/Abstract] OR Tadjikistan[Title/Abstract] OR Tadzhik[Title/Abstract] OR Tanzania[Title/Abstract] OR Thailand[Title/Abstract] OR Togo[Title/Abstract] OR Togolese Republic[Title/Abstract] OR Tonga[Title/Abstract] OR Trinidad[Title/Abstract] OR Tobago[Title/Abstract] OR Tunisia[Title/Abstract] OR Turkey[Title/Abstract] OR Turkmenistan[Title/Abstract] OR Turkmen[Title/Abstract] OR Tuvalu[Title/Abstract] OR Uganda[Title/Abstract] OR Ukraine[Title/Abstract] OR Uruguay[Title/Abstract] OR USSR[Title/Abstract] OR Soviet Union[Title/Abstract] OR Union of Soviet Socialist Republics[Title/Abstract] OR Uzbekistan[Title/Abstract] OR Uzbek OR Vanuatu[Title/Abstract] OR New Hebrides[Title/Abstract] OR Venezuela[Title/Abstract] OR Vietnam[Title/Abstract] OR Viet Nam[Title/Abstract] OR West Bank[Title/Abstract] OR Yemen[Title/Abstract] OR Yugoslavia[Title/Abstract] OR Zambia[Title/Abstract] OR Zimbabwe[Title/Abstract] OR Rhodesia[Title/Abstract] OR "developing country"[Title/Abstract] OR "developing countries"[Title/Abstract] OR "developing nation"[Title/Abstract] OR "developing nations"[Title/Abstract] OR "developing population"[Title/Abstract] OR "developing populations"[Title/Abstract] OR "developing world"[Title/Abstract] OR "less developed country"[Title/Abstract] OR "less developed countries"[Title/Abstract] OR "less developed nation"[Title/Abstract] OR "less developed nations"[Title/Abstract] OR "less developed population"[Title/Abstract] OR "less developed populations"[Title/Abstract] OR "less developed world"[Title/Abstract] OR "lesser developed country"[Title/Abstract] OR "lesser developed countries"[Title/Abstract]

OR ”lesser developed nation”[Title/Abstract] OR ”lesser developed nations”[Title/Abstract] OR ”lesser developed population”[Title/Abstract] OR ”lesser developed populations”[Title/Abstract] OR ”lesser developed world”[Title/Abstract] OR ”under developed country”[Title/Abstract] OR ”under developed countries”[Title/Abstract] OR ”under developed nation”[Title/Abstract] OR ”under developed nations”[Title/Abstract] OR ”under developed population”[Title/Abstract] OR ”under developed populations”[Title/Abstract] OR ”under developed world”[Title/Abstract] OR ”underdeveloped country”[Title/Abstract] OR ”underdeveloped countries”[Title/Abstract] OR ”underdeveloped nation”[Title/Abstract] OR ”underdeveloped nations”[Title/Abstract] OR ”underdeveloped population”[Title/Abstract] OR ”underdeveloped populations”[Title/Abstract] OR ”underdeveloped world”[Title/Abstract] OR ”middle income country”[Title/Abstract] OR ”middle income countries”[Title/Abstract] OR ”middle income nation”[Title/Abstract] OR ”middle income nations”[Title/Abstract] OR ”middle income population”[Title/Abstract] OR ”middle income populations”[Title/Abstract] OR ”low income country”[Title/Abstract] OR ”low income countries”[Title/Abstract] OR ”low income nation”[Title/Abstract] OR ”low income nations”[Title/Abstract] OR ”low income population”[Title/Abstract] OR ”low income populations”[Title/Abstract] OR ”lower income country”[Title/Abstract] OR ”lower income countries”[Title/Abstract] OR ”lower income nation”[Title/Abstract] OR ”lower income nations”[Title/Abstract] OR ”lower income population”[Title/Abstract] OR ”lower income populations”[Title/Abstract] OR ”underserved country”[Title/Abstract] OR ”underserved countries”[Title/Abstract] OR ”underserved nation”[Title/Abstract] OR ”underserved nations”[Title/Abstract] OR ”underserved population”[Title/Abstract] OR ”underserved populations”[Title/Abstract] OR ”underserved world”[Title/Abstract] OR ”under served country”[Title/Abstract] OR ”under served countries”[Title/Abstract] OR ”under served nation”[Title/Abstract] OR ”under served nations”[Title/Abstract] OR ”under served population”[Title/Abstract] OR ”under served populations”[Title/Abstract] OR ”under served world”[Title/Abstract] OR ”deprived country”[Title/Abstract] OR ”deprived countries”[Title/Abstract] OR ”deprived nation”[Title/Abstract] OR ”deprived nations”[Title/Abstract] OR ”deprived population”[Title/Abstract] OR ”deprived populations”[Title/Abstract] OR ”deprived world”[Title/Abstract] OR ”poor country”[Title/Abstract] OR ”poor countries”[Title/Abstract] OR ”poor nation”[Title/Abstract] OR ”poor nations”[Title/Abstract] OR ”poor population”[Title/Abstract] OR ”poor populations”[Title/Abstract] OR ”poor world”[Title/Abstract] OR ”poorer country”[Title/Abstract] OR ”poorer countries”[Title/Abstract] OR ”poorer nation”[Title/Abstract] OR ”poorer nations”[Title/Abstract] OR ”poorer population”[Title/Abstract] OR ”poorer populations”[Title/Abstract] OR ”poorer world”[Title/Abstract] OR ”developing economy”[Title/Abstract] OR ”developing economies”[Title/Abstract] OR ”less developed economy”[Title/Abstract] OR ”less developed economies”[Title/Abstract] OR ”lesser developed economy”[Title/Abstract] OR ”lesser developed economies”[Title/Abstract] OR ”under developed economy”[Title/Abstract] OR ”under developed economies”[Title/Abstract] OR ”underdeveloped economy”[Title/Abstract] OR ”underdeveloped economies”[Title/Abstract] OR ”middle income economy”[Title/Abstract] OR ”middle income economies”[Title/Abstract] OR ”low income economy”[Title/Abstract] OR ”low income economies”[Title/Abstract] OR ”lower income economy”[Title/Abstract] OR ”lower income economies”[Title/Abstract] OR ”low gdp”[Title/Abstract] OR ”low gnp”[Title/Abstract] OR ”low gross domestic”[Title/Abstract] OR ”low gross national”[Title/Abstract] OR ”lower gdp”[Title/Abstract] OR ”lower gnp”[Title/Abstract] OR ”lower gross domestic”[Title/Abstract] OR ”lower gross national”[Title/Abstract] OR lmic[Title/Abstract] OR lmics[Title/Abstract] OR ”third world”[Title/Abstract] OR ”lami country”[Title/Abstract] OR ”lami countries”[Title/Abstract] OR ”transitional country”[Title/Abstract] OR ”transitional countries”[Title/Abstract])) OR ((”Developing Countries”[Mesh]) OR ”Rural Population”[Mesh])))) AND

Outcome

((((((((((((((((((((((((((”Absenteeism”[Mesh]) OR ”Achievement”[Mesh]) OR ”Activities of Daily Living”[Mesh]) OR ”Community Networks”[Mesh]) OR ”Costs and Cost Analysis”[Mesh]) OR ”Disability Evaluation”[Mesh]) OR ”Financing, Personal”[Mesh]) OR ”Outcome and Process Assessment (Health Care)”[Mesh]) OR ”Patient Admission”[Mesh]) OR ”Patient Compliance”[Mesh]) OR ”Postoperative Period”[Mesh]) OR ”Problem Solving”[Mesh]) OR ”Quality of Life”[Mesh]) OR ”Social Adjustment”[Mesh]) OR ”Social Environment”[Mesh]) OR ”Socioeconomic Factors”[Mesh]) OR (”Work”[Mesh] OR ”Workload”[Mesh]))) OR ”Financing, Personal”[Mesh]) OR ”Adaptation, Psychological”[Mesh])) OR ((((((((((((((((((

agricultural yield[Text Word] OR access to healthcare[Text Word]) OR access to medical care[Text Word]) OR assets[Text Word]) OR business practices[Text Word]) OR career mobility[Text Word]) OR community networks[Text Word]) OR consumption[Text Word]) OR credit[Text Word]) OR disability[Text Word]) OR earning[Text Word]) OR economic burden[Text Word]) OR economic status[Text Word]) OR economic behavior[Text Word]) OR economic welfare[Text Word]) OR economic behaviour[Text Word]) OR economics[Text Word]) OR employment[Text Word]) OR empowerment[Text Word]) OR enrolment[Text Word]) OR expenditure[Text Word]) OR expense[Text Word]) OR family burden[Text Word]) OR farming[Text Word]) OR finance[Text Word]) OR financial difficulties[Text Word]) OR financial management[Text Word]) OR food security[Text Word]) OR health care utilization[Text Word]) OR hours worked[Text Word]) OR housing[Text Word]) OR household [Text Word]) OR income[Text Word]) OR investment[Text Word]) OR job[Text Word]) OR labour[Text Word]) OR labor[Text Word]) OR labor force[Text Word]) OR labour force[Text Word]) OR land[Text Word]) OR livelihood[Text Word]) OR loan[Text Word]) OR medical care[Text Word]) OR occupation[Text Word]) OR occupational [Text Word]) OR opportunity cost[Text Word]) OR ownership[Text Word]) OR participation[Text Word]) OR pensions[Text Word]) OR poverty[Text Word]) OR productivity[Text Word]) OR re-employment[Text Word]) OR re-entry[Text Word]) OR revenue[Text Word]) OR salary[Text Word]) OR salaries[Text Word]) OR saving[Text Word]) OR social adjustment[Text Word]) OR social class[Text Word]) OR social conditions[Text Word]) OR social contacts[Text Word]) OR social environment[Text Word]) OR social mobility[Text Word] OR social security[Text Word]) OR social support[Text Word]) OR social network[Text Word]) OR social ties[Text Word]) OR socioeconomic[Text Word]) OR socio-economic[Text Word]) OR standard of living[Text Word]) OR subsistence[Text Word]) OR technology adoption[Text Word]) OR unemployment[Text Word]) OR wages[Text Word]) OR wealth[Text Word]) OR work disability[Text Word] OR working[Text Word])) OR Wellbeing[Text Word]) OR Quality of life[Text Word]) OR vocation[Text Word])) OR health status[Text Word]) OR social ecology[Text Word])))) OR coping skills[Text Word]) OR relapse [Text Word]) OR vocational[Text Word]) OR role impairment[Text Word]) OR treatment outcome[Text Word])) AND

Study design

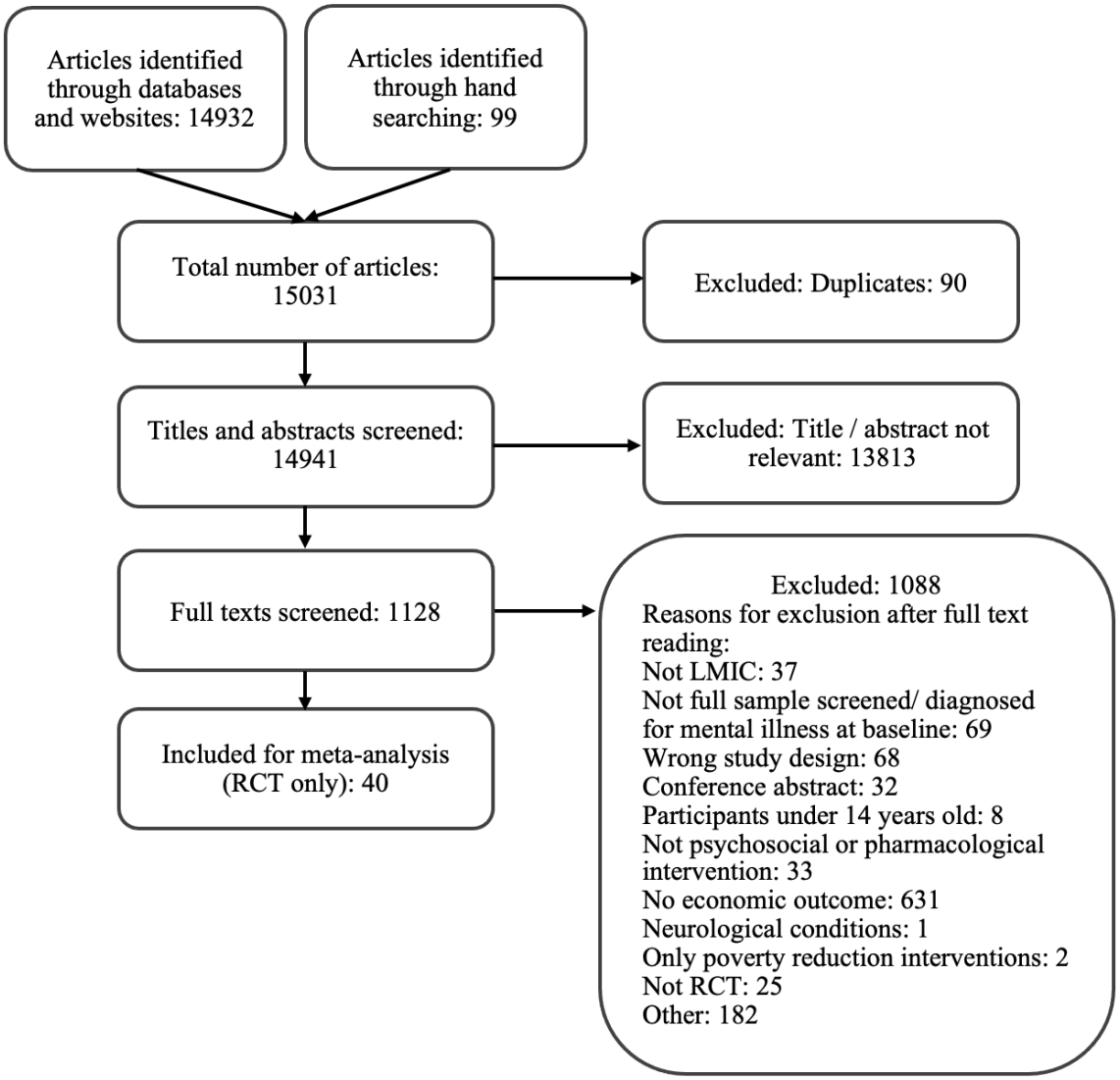
((((((((((”Double-Blind Method”[Mesh] OR ”Single-Blind Method”[Mesh]) OR ”Cohort Studies”[Mesh] OR ”Health Services Research”[Mesh]) OR ”Time Factors”[Mesh]) OR (”Clinical Trials as Topic”[Mesh] OR ”Controlled Clinical Trials as Topic”[Mesh] OR ”Pragmatic Clinical Trials as Topic”[Mesh] OR ”Clinical Trial”[Publication Type] OR ”Non-Randomized Controlled Trials as Topic”[Mesh] OR ”Randomized Controlled Trials as Topic”[Mesh] OR ”Pragmatic Clinical Trial”[Publication Type])) OR ”Comparative Study”[Publication Type]) OR (”Evaluation Studies”[Publication Type] OR ”Evaluation Studies as Topic”[Mesh])) OR ”Program Evaluation”[Mesh])) OR ((((((Clinical Trial[Text Word] OR Comparative study[Text Word]) OR comparative studies[Text Word]) OR Controlled[Text Word]) OR evaluation study[Text Word]) OR evaluation studies[Text Word]) OR follow-up study[Text Word] OR follow-up studies[Text Word] OR longitudinal study[Text Word] OR longitudinal studies[Text Word] OR non-randomised[Text Word] OR non-randomized[Text Word] OR program evaluation[Text Word] OR programme evaluation[Text Word] OR prospective study[Text Word] OR prospective studies[Text Word] OR randomised[Text Word] OR randomized[Text Word] OR quantitative study[Text Word] OR quantitative studies[Text Word] OR quasi experimental[Text Word] OR trial[Text Word] OR trials[Text Word] OR cohort[Text Word])))))) NOT (((animals[mh] NOT humans[mh])))))) NOT ((((((((((((((((plants[Text Word]) OR Pain[Text Word]) OR rats[Text Word]) OR mice[Text Word]) OR stroke[Text Word]) OR qualitative[Text Word]) OR United Kingdom[Text Word]) OR United States[Text Word]) OR America[Text Word]) OR American[Text Word]) OR Britain[Text Word]) OR USA[Text Word]) OR British[Title/Abstract]) OR New Zealand[Text Word]) OR Australia[Text Word]))

B Details on search and screening

List of databases searched: Social Science Research Network (SSRN), Research Papers in Economics (RePEc), the Abdul Latif Jameel Poverty Action Lab (JPAL) Evaluation and Publication Database, the World Bank Poverty Impact Evaluations Database, Research for Development (R4D), ECONLIT, WHO regional databases that cover LMICs, Sociological Abstracts, Applied Social Sciences Index and Abstracts (ASSIA), Public Affairs Information Service (PAIS International), Pubmed (including Medline), Scopus (including Embase), Web of Science, Social Science Citation Index (SSCI), EbscoHost, Africa Wide, Cumulative Index to Nursing and Allied Health Literature (CINAHL), PsycInfo, PROQUEST, and Published International Literature on Traumatic Stress (PILOTS).

List of trial registries searched: Cochrane, ClinicalTrials.gov, the EU clinical trial registry, the Pan African Trials Registry, the ISRCTN Registry, the 3ie Registry for International Development Impact Evaluations (RIDIE), and the American Economic Association trial registry.

Figure A1: Flow of citations reviewed during systematic review



C Details on outcomes and further descriptives on study characteristics

Table A1: Additional characteristics of interventions in included randomized controlled trials, by outcome type and on intervention level

	(1) Number of interventions	(2) Share of interventions
All interventions	39	1.00
Follow-up timing:		
Follow-up over one year later	17	0.44
Measurement during intervention	9	0.23
Additional follow-up (unplanned)	5	0.13
Follow-up:<=6 months after start	25	0.64
Follow-up: >6-12 months after start	18	0.46
Follow-up:>1-2 years after start	6	0.15
Follow-up:>2 years after start	6	0.15
Follow-up combinations (in months after start):		
>24	1	0.03
>12-24	3	0.08
>6-12	7	0.18
>6-12 & >24	3	0.08
<6	13	0.33
<6 & >24	1	0.03
<6 & >12-24	2	0.05
<6 & >6-12 & >12-24	6	0.15
<6 & >6-12 & >24	1	0.03
<6 & >6-12 & >12-24	1	0.03
Type of psychosocial intervention:		
Cognitive-behavioral therapy (CBT)	19	0.49
Problem solving therapy	11	0.28
Interpersonal therapy	10	0.26
Solution-focused therapy	9	0.23
Motivational interviewing	3	0.08
Psychoeducation (only)	15	0.38
Reconciliation	8	0.21
Type of pharmacological intervention:		
Against psychotic disorders	9	0.23
Against mood disorders (depression)	7	0.18
Against substance abuse	1	0.03
Control condition *(mutually exclusive)*:		
Enhanced Usual Care	7	0.18
No Treatment	19	0.49
Treatment As Usual (Pharmacological)	13	0.33
Age target groups:		
Adults (18+)	26	0.67
Youth[1]	2	0.05
All ages[2]	11	0.28
Specific target groups:		
Minority or vulnerable population	7	0.18
Males	7	0.18
Females	8	0.21
Both genders	27	0.69
Publication period:		
Publication: 1990-2000	3	0.08
Publication: 2001-2010	8	0.21
Publication: 2011-2015	13	0.33
Publication: 2016-2020	11	0.28
Publication: 2021-present	8	0.21
Country income *(mutually exclusive)*		
Upper middle income country	14	0.36
Lower middle income country	19	0.49
Low income country	6	0.15
Regions *(mutually exclusive)*		
Sub-Saharan Africa	9	0.23
Europe and Central Asia	2	0.05
Latin America and the Caribbean	3	0.08
South Asia	14	0.36
East Asia and Pacific	11	0.28

Notes: There are 39 interventions, 180 economic effect sizes and 335 mental health effect sizes. Some variable categories are not mutually exclusive (for example, interventions can measure employment and education outcomes), which is why percentages within categories can exceed 100%. The types of psychosocial and pharmacological interventions also cover combined interventions. Psychosocial interventions often include more than one type of therapy. [1] 1 intervention with ages 15-24, 1 intervention with 18-35. [2] 1 intervention with 14+, 1 with 16-45, 7 with 18+, 1 with 16-60, 1 with 25-35, 1 with 16-50.

Table A2: Economic outcomes

Outcome category	# of estimates	Detailed measures
Assets	18	Durable goods index; Index of durable assets; Index of household assets: enumerator assessment.; Level of credit; Level of debt that the household owes to others,; Level of savings; Net worth = savings+credit-debt; Savings stock; Total value of assets owned; Value of business assets.
Education	17	Academic Performance Scale; Binary enrolment; Child investment index; Does index child attend private school: own assessment; Home Observation for Measurement of the Environment (HOME) Inventory: Own assessment of Learning Materials sub-scale; HOME Inventory: Own assessment of Physical Environment sub-scale; Log of the family's educational expenditures in the past month: own assessment; Mother's expected grade attainment for the index child: own assessment; School quality (class size, number of teachers, number of rooms & classroom amenities): Clincian assessment; Homework time; School attendance; School enrolment
Income, consumption and input expenditure	27	Earnings from primary and secondary jobs; Earnings in the past 4 weeks; Food and non-food consumption in past 2 weeks; Investment in the past 2 weeks; Monthly household revenue; Monthly per-capita non-durable consumption; Mother's monthly earnings: own assessment ; Per-capita consumption; Total monthly expenditure; Earnings in the past month.
Other	5	PSFS Occupational Functioning: Own satisfaction with functioning in money-management; Applied for ability-based contract; Reservation wage
Social networks	7	Frequency of borrowing and lending; Integrated Questionnaire for the Measurement of Social Capital: Own assessment of Financial Social Network Size ; Integrated Questionnaire for the Measurement of Social Capital: Own assessment of Instrumental Support Network Size; Monetary value of borrowing and lending; Respondent belongs to an osusu (savings group).
Subjective poverty measures	11	Measure of poverty*; Self-reported economic status; Projected economic status in 5 years time using Cantril's ladder; Satisfaction that household needs are met; Satisfaction with household's economic situation relative to 1 year ago; Self-reported economic status today using Cantril's ladder; WHO QoL work item: own satisfaction with financial resources.
Days unable to work	23	Days of sick leave in last 2 years: own assessment; Days respondent was 'healthy' in the past 30 days: own assessment ; WHODAS 2.0 Participation in Society: self-assessment of days unable to work or reduction in work; WHODAS 2.0: days unable to work + (0.5 x) days with reduced work; Self report of work days missed last month due to poor health
In employment	11	Likelihood of re-employment over the 12 month follow up period; Mother is employed; Self reported 'employed'; Self-reported Employment Status ; Self: Engaged in work in the last week.
Functioning at work	20	Addiction Severity Index (ASI): Employment status*; Addiction Severity Index (ASI): Employment status for respondents receiving relapse prevention treatment*; Addiction Severity Index (ASI): Indication of ideal employment status*; Bracelets made in ten minutes; Indian Disability Evaluation and Assessment Scale (IDEAS) scale: Employment, housework and educational performance*; Independent Living Skills Survey: job maintenance*; KDQOL-SF: work status; Life Chart Schedule: Performance at Work; Morningside Rehabilitation Status Scale (MRSS): activity/inactivity: Rating of functioning in employment and leisure*; Overall occupational disabilities Groningen Social Disability Scale (GSDS-II); Own assessment of capacity to do farming; Own assessment of capacity to do manual labour; Own assessment of capacity to grow food; Psychosocial Functioning Scale (PSFS) Occupational Functioning: Own satisfaction with functioning in occupation; WHO Quality of Life work item: own satisfaction with capacity for work; World Health Organization Disability Assessment Schedule (WHODAS 2.0): Life activities domain 5
Job search	6	Available to take a job opportunity; Job search hours per week; Job search hours,
Time in work	17	Hours per week of work in the last 2 months; Months engaged in normal occupation: family assessment; Percent of time at work: own assessment; SDSS: Ability to Work (full time); Work hours in the last week; Time per 24h in productive activities (converted to weekly value); Child care work hours; Domestic work; Primary and secondary jobs and agricultural work hours.
Unable to work	18	Percent of time on sick leave; SDSS: own assessment of ability to work (full time); SDSS: own assessment of ability to work (part time); SDSS: own assessment of ability to work (unable to work); SDSS: own assessment of ability to work in farm or house work (full time and part-time); WHODAS 2.0: days unable to work; WHODAS 2.0: days unable to work - at least one.

$*$ = outcome measured by a clinician.

Table A3: Mental health outcomes

Outcome category	# of estimates	Detailed measures
Antisocial Behaviour	2	Antisocial behaviour index
Functioning	56	Activites of daily living; Addiction Severity Index (ASI): clinician; Brief Disability Questionnaire (BDQ); Direct Assessment of Functional Status; Direct Assessment of Functional Status: dealing with finances.*; Global Assessment of Functioning (GAF)scale: clinician; Groningen Social Disability Schedule (GSDS-II)*; Indian Disability Evaluation and Assessment Scale (IDEAS) scale: Self-care*; Independent Living Skills Survey; Independent Living Skills Survey: Money management*; Indian Disability Evaluation and Assessment Scale (IDEAS)*; Locally adapted gender specific functional impairment*; Morningside Rehabilitation Status Scale (MRSS) subscale: Current symptoms and deviant behavior*; Personal and Social Performance Scale; Short Inventory of Problems (SIP) mean score; Social Disability Screening Schedule (SDSS): Social dysfunction subscale; World Health Organization Disability Assessment Schedule (WHODAS): Functional impairment; World Health Organization Quality of Life subscale - Independence*; World Health Organization Disability Assessment Schedule 2.0 12 item; World Health Organization Disability Assessment Schedule 2.0 36 item; Local functioning tool; study-specific Psychosocial Functioning Scale (PSFS).
Functioning in social interactions	9	Indian Disability Evaluation and Assessment Scale (IDEAS): Interpersonal activities*; Indian Disability Evaluation and Assessment Scale (IDEAS): Communication and understanding*; Oxford Measure of Psychosocial Adjustment adapted subscale- prosocial behaviour; Revised Social Disability Screening Schedule (SDSS-R); Social Disability Screening Schedule (SDSS); Social Disability Screening Schedule (SDSS): Social dysfunction subscale.
Mental health disorders	4	Proportion currently treated with antipsychotics; Proportion never treated with antipsychotic medication
PTSD	13	Adapted PTSD Symptom Scale; Harvard Trauma Questionnaire (HTQ); Index of locally-relevant post-traumatic stress features; Mini International Neuropsychiatric Interview - PTSD*; Post-Traumatic Stress Disorder Reaction Index (PTSD-RI); Posttraumatic Stress Disorder Checklist Civilian,
Recovery (dummy)	4	Depression recovery on Patient Health Questionnaire (PHQ-9); Full recovery: Social Disability Screening Schedule (SDSS); Full recovery from recorded mental health disorder*.
Relapse (dummy)	14	Depression remission on Patient Health Questionnaire (PHQ-9); Proportion relapse*; Relapse rate*; Schizophrenia relapse rate*.
SMD symptoms	8	Positive and Negative Syndrome Scale (PANSS): General psychopathological health; Positive and Negative Syndrome Scale (PANSS)*; Serious Mental Disability: Social Disability Screening Schedule (SDSS)
Self-regulation	7	Behavioural Assessment of the Dysexecutive Syndrome (BADS) scale; Executive function; Patience index; Self-control scale; Short grit scale
Social support	27	Addiction Severity Index (ASI): Family support*; Addiction Severity Index (ASI): Family/social*; Contact with non-kin social network; Emotional support seeking; Inventory of Socially Supportive Behaviors (ISSB); Kidney Disease and Quality of Life-Short Form (KDQOL-SF): Social interactions; Kidney Disease and Quality of Life-Short Form (KDQOL-SF): Social support; Perceived Social Support; Social network quality index; World Health Organization Quality of Life (WHOQOL) subscale: Social support*.
Anxiety	6	Adapted Zung Anxiety Index; Generalised Anxiety Disorder Assessment (GAD-7):; Hopkins Symptom Checklist: Anxiety; Self Rating Anxiety Scale (SAS)

$*$ = outcome measured by a clinician.

Table A4: Mental health outcomes continued

Outcome category	# of estimates	Detailed measures
Cognition	11	Cognition index; Cognition index*; Digit Span: backwards; Digit span: forwards; Executive function index; Kidney Disease and Quality of Life-Short Form (KDQOL-SF): Cognitive function; Raven's Progressive Matrices; World Health Organization Quality of Life (WHOQOL) subscale: Cognitive function*
Depression	33	Adapted Zung Depression Index; Beck Depression Inventory (BDI); Beck Depression Inventory II; Culturally Grounded Screening for Depression; Patient Health Questionnaire (PHQ-9); Depression status*; Hamilton Depression Rating Scale (HDRS)*; Hopkins Symptom Checklist: Depression; Hopkins Symptom Checklist 25 (HSCL-25) : Depression; Index of locally-relevant depression features; Locally adapted Hopkins Symptom Checklist (HSCL): Depression subscale*; Mini International Neuropsychiatric Interview: Depression*.
Diagnosed with mental disorder	14	Depression Remission on Patient Health Questionnaire (PHQ-9); Diagnostic and Statistical Manual of Mental Disorders IV (DSM-IV): Adapted Depression Index*; Hopkins Symptom Checklist; Mini International Neuropsychiatric Interview: No depression*; No moderate/severe depression on Patient Health Questionnaire (PHQ-9).
General mental health	18	30 minus days in last month with poor mental health; Addiction Severity Index (ASI): Psychiatric status*; Addiction Severity Index (ASI): Legal problems*; Behavioural Risk Factor Surveillance Survey-rating of mental health; Clinical Global Impression Subscale for Severity of Illness (CGIS)*; Kessler Psychological Distress Scale; Mean effect index of psychological health; Mental health index; Mood score; Subjective Well-being Index; Subjective Wellbeing; World Health Organization Quality of Life (WHOQOL) subscale: Mental health*.
Non-specific CMD	21	Clinical Interview Schedule - Revised (CIS-R) for positive screen*; Clinical Interview Schedule - Revised (CIS-R) morbidity score for ICD-10 diagnosis*; Clinical Interview Schedule - Revised (CIS-R) morbidity score for depression diagnosis*; Clinical Interview Schedule - Revised (CIS-R) score for subthreshold*; Hopkins Symptom Checklist 25 (HSCL-25): Anxiety; Hopkins Symptom Checklist; Oxford Measure of Psychosocial Adjustment adapted subscale: Psychological distress.
Overall assessment of mental disorder	10	Addiction Severity Index (ASI): Mental health*; Clinical Global Impression Subscale for Severity of Illness (CGIS)*; Clinical Interview Schedule - Revised (CIS-R) - total score*; General mental health; Kidney Disease and Quality of Life-Short Form (KDQOL-SF): Overall mental health; Positive and Negative Syndrome Scale (PANSS) Scale: General psychopathological health; State of illness* .
Physical health	10	Addiction Severity Index (ASI): Medical*; Addiction Severity Index (ASI): Physical health*; Kidney Disease and Quality of Life-Short Form (KDQOL-SF): Overall health; Kidney Disease and Quality of Life-Short Form (KDQOL-SF): Sexual function; Kidney Disease and Quality of Life-Short Form (KDQOL-SF): Sleep; World Health Organization Quality of Life (WHOQOL) subscale: General physical health*.
Rehospitalisation	6	Days of re-admission*; Days of rehospitalisation*; Rehospitalization rate*.
Self-esteem/self-efficacy	7	General Self-Efficacy Scale; Modified Rosenberg Self-Esteem Scale (SES); Self-Efficacy Scale; Self-esteem index; World Health Organization Quality of Life (WHOQOL) subscale: Self-esteem*
Stress	2	Perceived Stress Scale (PSS)
Substance use	35	Addiction Severity Index (ASI): Alcohol use*; Addiction Severity Index (ASI): Drug*; Any ethanol consumed; Average proportion of negative urine test results*; Daily drinking; Drug positive; Ethanol consumed; Heroin abstinence*; Longest period of abstinence; Non-drinker; Percentage of days abstinent; Percentage of days of heavy drinking; Proportion of days abstinent; Proportion of days heavy drinking; Remission on Alcohol Use Disorders Identification Test (AUDIT)*; Short Inventory of Problems (SIP) mean score; Substance abuse index.
Suicide attempts or at risk of suicide	18	Any suicide attempt or suicidal ideation; Any suicide attempt; Mini International Neuropsychiatric Interview: Suicide risk*; Proportion suicidal behavior for common mental disorder diagnosis (ICD-10) group on CIS-R*; Proportion suicidal behavior for depression diagnosis (ICD-10) group on CIS-R*; Proportion suicidal behavior for positive screen group on CIS-R*; Proportion suicidal behavior for subthreshold group on CIS-R*; Proportion that take their own lives*; Suicidal behaviour; Suicidal thoughts or attempts.

$*$ = outcome measured by a clinician.

Table A5: Economic effect sizes estimates by economic outcome type and by intervention type

Outcome type	Pharmacological	Psychosocial	Pharm.+psych.	Total
Assets	10	8	0	18
Education	8	9	0	17
Income, consumption and input expenditure	4	23	0	27
Other	0	4	1	5
Social networks	0	7	0	7
Subjective poverty measures	0	10	1	11
Days unable to work	0	13	10	23
In employment	0	11	0	11
Functioning at work	0	10	10	20
Job search	2	4	0	6
Time in work	8	4	5	17
Unable to work	6	4	8	18
Total	38	107	35	180

Notes: Table A5 displays the number of included economic effect size estimates by economic outcome type and by intervention type.

Table A6: Mental health effect sizes estimates by mental health outcome type and by intervention type

Outcome type	Pharmacological	Psychosocial	Pharm.+psych.	Total
Antisocial Behaviour	0	2	0	2
Functioning in social interactions	0	2	7	9
PTSD	0	13	0	13
Recovery (dummy)	1	1	2	4
Relapse (dummy)	1	7	6	14
SMD symptoms	2	1	5	8
Self-regulation	0	7	0	7
Anxiety	2	4	0	6
Cognition	2	8	1	11
Depression	2	30	1	33
Diagnosed with mental disorder	4	10	0	14
General mental health	2	12	4	18
Non-specific CMD	0	5	16	21
Overall assessment of mental disorder	1	2	7	10
Physical health	0	7	3	10
Rehospitalisation	0	0	6	6
Self-esteem/self-efficacy	0	6	1	7
Stress	0	0	2	2
Substance use	0	31	4	35
Suicide attempts or at risk of suicide	1	8	9	18
Social support	0	24	3	27
Functioning	3	33	20	56
Mental health disorders	2	0	2	4
Total	23	213	99	335

Notes: Table A6 displays the number of included mental health effect size estimates by mental health outcome type and by intervention type.

Table A7: Scales used to measure mental health and functioning outcomes

Scale name	Survey question example wording (abridged)	Recall Period
Functioning at work		
Revised Social Disability Screening Schedule (SDSS)	I want to know about the functioning of the subject (you) at home and at work, whether or not he/she (you) is able to do what he/she (you) should be able to do. I will ask questions about different types of functioning; please let me know if he/she (you) has had any problems or difficulties in each type of functioning over the last three months: (1) Work (study)(a) None or very mild problems/difficulties; (b) Significant reduction in ability resulting in decreased functioning or complaints; (c) Unable to work/study or risks formal punishment or reprimand at work/school for poor functioning due to psychological problems.	3 months
World Health Organization Disability Assessment Schedule 2.0 (WHODAS)	In the past 30 days, for how many days were you totally unable to carry out your usual activities or work because of any health condition? Not counting the days that you were totally unable, for how many days did you reduce your usual activities or work because of any health	Past 30 days
World Health Organisation Quality of Life	How satisfied are you with your capacity for work?	Not provided
Indian Disability Evaluation and Assesment Scale (IDEAS)	Employment, housework and educational performance. Guiding questions: a. Is he employed/unemployed? b. If employed, does he go to work regularly? c. Does he like his job and coping well with it? d. Can you rely on him financially? e. If unemployed does he make any	Not provided
Overall occupational disabilities (GSDS-II - Groningen Social Disability Scale)	Dimensions: a. daily routine; b. work performance; c. contacts with others; d. (other) daily activities	4 weeks
World Health Organization Disability Assessment Schedule 2.0	In the past 30 days, for how many days were you totally unable to carry out your usual activities or work because of any health condition?	Past 30 days
Kidney Disease Quality of Life Short Form: work status	2 items: q20: during the past 4 weeks, did you work at a paying job? q21: does your health keep you from working at a paying job?	Past 30 days
Global Assessment of Functioning (GAF): occupational	The GAF rating doesn't have explicit wording, but draws on an interview or questionnaire; medical records; information from the person's doctor, care givers, or close relatives; police or court records about violent or illegal behavior.	Contemporaneous
Morningside Rehabilitation Status Scale	Inactivity (occupational and leisure) scale: How well the patient has initiated and sustained activity and performed effectively in: 1. Work; 2. Training programme; 3. Method of looking for work; 4. Daily routine; 5. Leisure time routine, indoor/outdoor (weekdays and weekends); 6. Reading habits and interests, TV, radio, etc.	Not provided
Social functioning		
Kidney Disease Qualiy of Life Short Form (KDQOL-SF): Social Functioning	How much during the past 4 weeks did you: Act irratable toward those around you; Isolate yourself from people around you; Get along well with other people.	4 weeks
Short-Form 36 Health Survey Questionnaire (SF-36)	Have emotional problems interfered with your normal social activities with family, friends, neighbors, or groups?	Not provided
Multidimensional Scale of Perceived Social Support	1. There is a special person who is around when I am in need. 2. There is a special person with whom I can share my joys and sorrows. 3. My family really tries to help me. 4. I get the emotional help and support I need from my family. 5. I have a special person who is a real source of comfort to me. 6. My friends really try to help me. 7. I can count on my friends when things go wrong. 8. I can talk about my problems with my family. 9. I have friends with whom I can share my joys and sorrows. 10. There is a special person in my life who cares about my feelings. 11. My family is willing to help me make decisions. 12. I can talk about my problems with my friends.	Not provided
Indian Disability Evaluation and Assessment Scale (IDEAS): Social Functioning Subscale	Behaviours with others; responsiveness to questions; regulating verbal and physical aggersion; acting independently in social interactions; behaviour with strangers; maintaining friendships. Understanding spoken and written/non-verbal messages and ability to reduce messages in order to commmunicate with others. Avoiding talking to people; when people come home what do they do; do they ever viist others; can they start, maintain and end a conversation? Do they understand body language?	Not provided
Groningen Social Disability Schedule (GSDS-II)	The GSDS-11 assesses 1. Role of self-care; 2. Family role; 3. Kinship role; relationships with parents and siblings; 4. Partner role; relationship with partner in marriage or cohabitation; 5. Parental role; relationship with children; 6. Citizen role; interest and participation in social life.; 7. Social role; relationships with friends and acquaintances.; 8. Occupational role; regular daily activities.	Four weeks

Table A8: Scales used to measure mental health and functioning outcomes

Scale name	Survey question example wording (abridged)	Recall Period
Global Functioning		
Brief Disability Questionnaire (BDQ)	1. Have your health problems limited you in daily activities? 2. Have you had to cut down or stop any activity you used to do, such as a hobby, because of your illness? 3. Have you been unable to do something that your family (or household) expected from you as part of your daily routine? 4. Have your personal problems decreased your motivation to work? 5. Have your personal problems decreased your personal efficiency at home, school or work? 6. Has there been a deterioration in your social relations with friends, workmates, or other people? 7. During the past week how many days in total were you unable to carry out your usual daily activities fully? 8. During the past	1 week
World Health Organization Disability Assessment Schedule 2.0 (WHODAS 2.0)	The 36 item questionnaire asks about difficulty experienced in the domains: Understanding and communicating; Getting around; Self-care; Life activities; Work; and Participation in Society	3-4 days ("several")
Patient Health Questionnaire-9 (PHQ-9)	How often have they been bothered by the following over the past 2 weeks? 1. Little interest or pleasure in doing things? 2. Feeling down, depressed, or hopeless? 3. Trouble falling or staying asleep, or sleeping too much? 4. Feeling tired or having little energy? 5. Poor appetite or overeating? 6. Feeling bad about yourself or that you are a failure or have let yourself or your family down? 7. Trouble concentrating on things, such as reading the newspaper or watching television? 8. Moving or speaking so slowly that other people could have noticed? Or so fidgety or restless that you have been moving a lot more than usual? 9. Thoughts that you would be better off dead, or	2 weeks
Global Assessment of Functioning (GAF) scale	Measures global functioning in two domains: (1) symptom severity, and (2) any serious impairment in psychological, social and occupational functioning on a mental health-illness continuum level of functioning).	Not provided
Personal and Social Performance Scale	Please rate the patient on his/her level of functioning during the reference period (e.g., past month or last 7 days). Consider what the person is doing, taking into account if she needs help or prompting by others. The four main domains of functioning considered in this scale are (a) personal and social relationships; (b) socially useful activities, including work and study; (c) self-care; and (d) aggressive	7 days or one month
Revised Social Disability Screening Schedule	The SDSS assesses social functioning at home and at work, dealing with functioning the domains of work, marital functioning, parental function, loneliness, group activities, physical activities, family function, self care, concern for the outside world, and responsibility and plans for the future. Against each domain, respondents report the number of months among the last 3 in which they experienced a level of difficulty performing a function relative to what they should be able to do.	3 months
Substance abuse		
Alcohol Use Disorders Identification Test (AUDIT)	1. How often do you have a drink containing alcohol? 2. How many drinks containing alcohol do you have on a typical day when you are drinking? 3. How often do you have six or more drinks on one occasion? 4. How often during the last year have you found that you were not able to stop drinking once you had started? 5. How often during the last year have you failed to do what was normally expected from you because of drinking? 6. How often during the last year have you needed a first drink in the morning to get yourself going after a heavy drinking session? 7. How often during the last year have you had a feeling of guilt or remorse after drinking? 8. How often during the last year have you been unable to remember what happened the night before because you had been drinking? 9. Have you or someone else been injured as a result of your drinking? 10. Has a relative, friend, doctor or another health worker been concerned about your drinking or	Not provided
Addiction Severity Index (ASI)	The ASI is a semi-structured interview designed to address seven potential problem areas in substance-abusing patients: medical status, employment and support, drug use, alcohol use, legal status, family/social status, and psychiatric status. The ASI provides an overview of problems related to substance, rather than focusing on any single area.	30 days

Table A9: Scales used to measure mental health and functioning outcomes

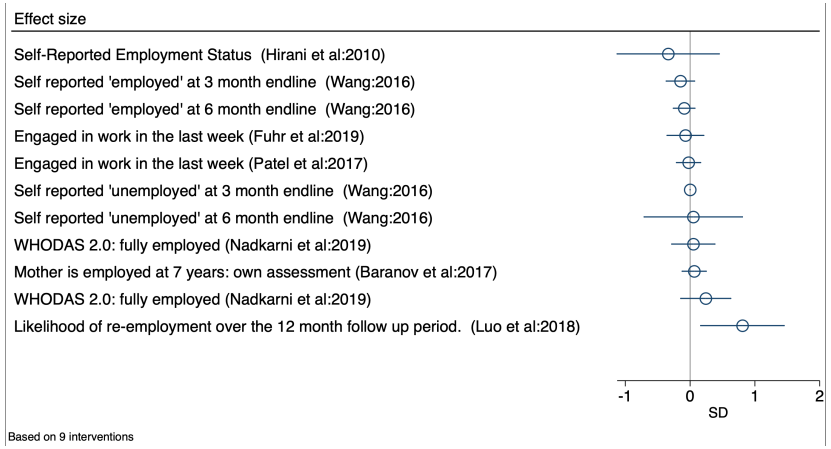
Scale name	Survey question example wording (abridged)	Recall Period
Mental Health - Common mental disorders		
Kessler Psychological Distress Scale	During the last 30 days: 1. how often did you feel tired out for no good reason? 2.how often did you feel nervous? 3 How often did you feel so nervous that nothing could calm you down? How often did you feel hopeless? How often did you feel restless or fidgety? How often did you feel so restless you could not sit still? How often did you feel depressed? How often did you feel that everything was an effort? How often did you feel so sad that nothing could cheer you up? How often did you feel worthless?	30 days
Perceived Stress Scale (PSS)	In the last month, how often have you: 1. Been upset because of something that happened unexpectedly? 2. Felt that you were unable to control the important things in your life? 3. Nervous and stressed? 4. Felt confident about your ability to handle your personal problems? 5. Felt that things were going your way? 6. Found that you could not cope with all the things that you had to do? 7. Been able to control irritations in your life? 8. Felt that you were on top of things? 9. Been angered because of things that happened that were outside of your control? 10.	1 month
Adapted Zung Anxiety Index	In the last month: how much did you feel nervous or anxious or worried? Did you feel fear without cause? Dd you often feel upset or feel sudden panic? How often have you felt that everything is alright and nothing bad will happen in the future? How much did your legs and arms shake and tremble? How much did you have headaches or pain in your neck and back? How often have you felt tired even if you not doing nothing?	4 weeks
Beck Depression Inventory (BDI)	1. Experiencing sadness; 2. Discouraged about the future; 3. Feel like a failure; 4. Feelings of satisfacation from doing things I used to do; 5. Feeling guilty; 6. Feel I am being punished; 7. Feel disappointed in myself; 8. Blame myself for my faults; 9. Suicidal ideation; 10. Cry more than usual; 11. Feeling irritated; 12. Losing interest in other people; 13. Make good decisions; 14. I look worse than I used to; 15. Less effective at work; 16. Poor sleep quality; 17. Easily tired; 18. Worse appetite; 19. Weight gain; 20. Worried about health; 21. Changed in	Not provided
Clinical Interview Schedule - Revised (CIS-R)	The CIS-R is a standardised, structured interview for the measurement and diagnosis of common mental disorders in community and general healthcare settings.	Not provided
Hopkins Symptom Checklist (HSCL-25)	A symptom inventory measuirng symptoms of anxiety and depression, with 10 items for anxiety and 15 for depression.	4 weeks
Hamilton Depression Rating Scale	1. Depressed Mood; 2. Feelings of guilty; 3. Suicidal ideation; 4. Insomnia; 5. Work and activities; 6. Slowness of thought; 7. Agitation; 8; Anxiety; 9 Somatic symptoms; 10 Labido; 11. Hypochondriasis; 12. Insight/self-awareness.	Not provided
Patient Health Questionnaire-9 (PHQ-9)	How often have they been bothered by the following over the past 2 weeks? 1. Little interest or pleasure in doing things? 2. Feeling down, depressed, or hopeless? 3. Trouble falling or staying asleep, or sleeping too much? 4. Feeling tired or having little energy? 5. Poor appetite or overeating? 6. Feeling bad about yourself or that you are a failure or have let yourself or your family down? 7. Trouble concentrating on things, such as reading the newspaper or watching television? 8. Moving or speaking so slowly that other people could have noticed? Or so fidgety or restless that you have been moving a lot more than usual? 9. Thoughts that you would be better off dead, or	2 weeks
Self Rating Anxiety Scale (SAS)	In the last month: 1. How much did you feel nervous or anxious or worried? 2. Did you feel fear without cause? 3. Did you often feel upset or feel sudden panic? 4. How often have you felt that everything is alright and nothing bad will happen in the future? 5. How much did your legs and arms shake and tremble? 6. How much did you have headaches or pain in your neck and back? 7. How often have you felt tired even if you not doing nothing? 8. How much did you feel restless? 9. Has your heart been pounding fast? 10. How much were you	1 month
South Asian Tension Scale	"Over the past 2 weeks, how often have you been bothered by any of these problems:" 1. Feeling sad, feeling like crying, lonely. 2. Feeling a loss of appetite, nausea, stomach pain. 3. Trouble concentrating, loss of memory. 4. Feeling angry or frustrated. 5. Insomnia. 6. Feeling cold in the body (not due to weather). 7. Feeling you would like to run away or escape. 8. Headaches or pain in your eyes. 9. Feeling anxious or afraid. 10. Feeling tired or a lack of energy. 11. Feeling helpless and unsupported. 12. Feelings of shakiness. 13. Sexual problems. 14. Feeling you want to be alone. 15. Problems with your periods. 16. Pains in your arms, legs, or other parts of your body. 17. Feeling homesick or missing family. 18. Feeling hot in parts of your body (not due to weather). 19. Vaginal discharge. 20. Feeling dizzy. 21. Pain or heaviness in your chest, heart palpitations. 22. Feeling a loss of control of your hands or feet. 23. Breathlessness.24. Your hair	2 weeks

Table A10: Scales used to measure mental health and functioning outcomes

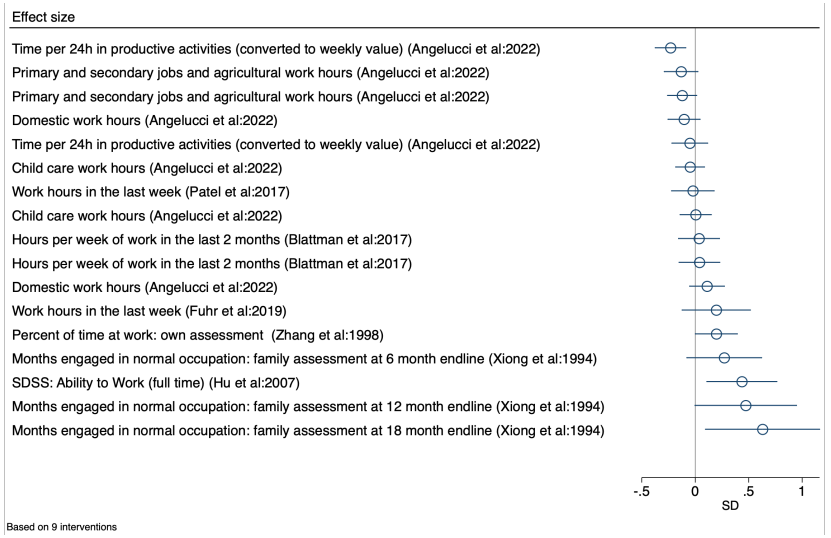
Scale name	Survey question example wording (abridged)	Recall Period
Mental Health - Severe mental disorders, PTSD, substance abuse disorders and other mental health scales		
Posttraumatic Stress Disorder Checklist-Civilian	1. Repeated, disturbing memories, thoughts, or images of a stressful experience from the past? 2. Repeated, disturbing dreams of a stressful experience from the past? 3. Suddenly acting or feeling as if a stressful experience were happening again (as if you were reliving it)? 4. Feeling very upset when something reminded you of a stressful experience from the past? 5. Having physical reactions (e.g., heart pounding, trouble breathing, or sweating) when something reminded you of a stressful experience from the past? 6. Avoid thinking about or talking about a stressful experience from the past or avoid having feelings related to it? 7. Avoid activities or situations because they remind you of a stressful experience from the past? 8. Trouble remembering important parts of a stressful experience from the past? 9. Loss of interest in things that you used to enjoy? 10. Feeling distant or cut off from other people? 11. Feeling emotionally numb or being unable to have loving feelings for those close to you? 12. Feeling as if your future will somehow be cut short? 13. Trouble falling or staying asleep? 14. Feeling irritable or having angry outbursts? 15. Having difficulty concentrating? 16. Being "super alert" or watchful	Not provided
Positive and Negative Syndrome Scale (PANSS)	The positive and negative syndrome scale is made up of three subscales: one positive symptom, one negative symptom and one general psychopathology scale. The positive symptoms measured are: delusions, conceptual disorganization, hallucinations, excitement, grandiosity, suspiciousness and hostility. The negative symptoms measured are: blunted affect, emotional withdrawal, poor rapport, passive apathetic social withdrawal, difficulty in abstract thinking, lack of spontaneity & flow of conversation, stereotyped thinking. The general psychopathology scale measures somatic conern, anxiety, guilt feelings, tension, mannerisms and posturing, depression, motor retardation, uncooperativeness, unusual thought content, disorientation, poor attention, lack of judgement & insight, disturbance of volition, poor impulse control, preoccupation, active social avoidance.	Not provided
Mini International Neuropsychiatric Interview (MINI)	Therapeutic areas: Behaviour and behaviour mechanisms; Mental Disorders; Chemically-induced Disorders. Therapeutic indications: suicidal ideation; psychotic disorders; anxiety disorders; depressive disorder; panic disorder; obsessive-compulsive disorder; alcoholism; bulimia nervosa; anoerxia nervosa; antisocial personality disorder; bipolar disorder; substance-related disorders; binge-eating disorder; generic for mental disorders; stress disorders, post-traumatic stress disorders.	2 weeks
Present State Examination (PSE-9)	The PSE is semi-structured interview, intended to provide an objective evaluation of symptoms associated with mental disorders. It contains 140 items, each scored on a 3-point or 4-point scale, and it is designed for use by experienced clinicians.	Contemporaneous
Harvard Trauma Questionnaire (HTQ)	The HTQ enquires about a variety of trauma events, as well as the emotional symptoms considered to be uniquely associated with trauma.	4 weeks

Figure A2: Original work-related effect size estimates

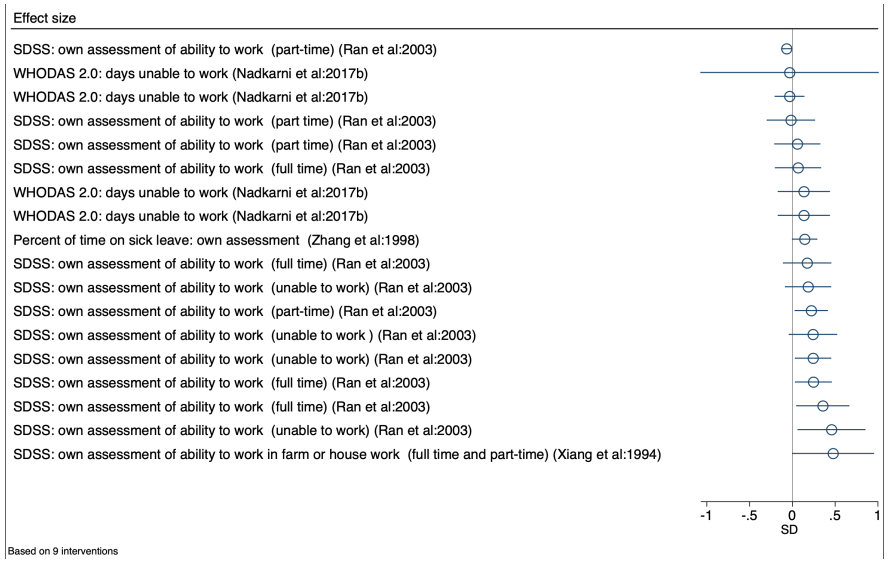
(a) In employment



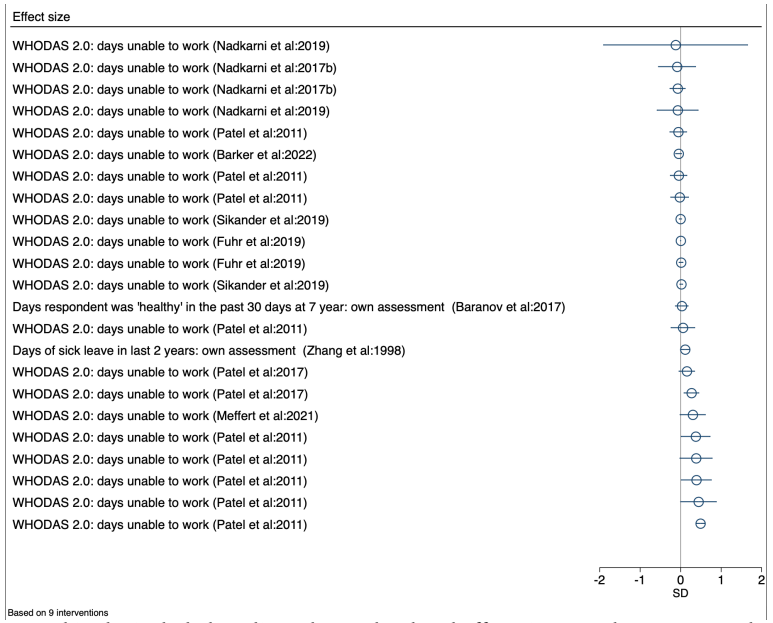
(b) Time in work



(c) Unable to work



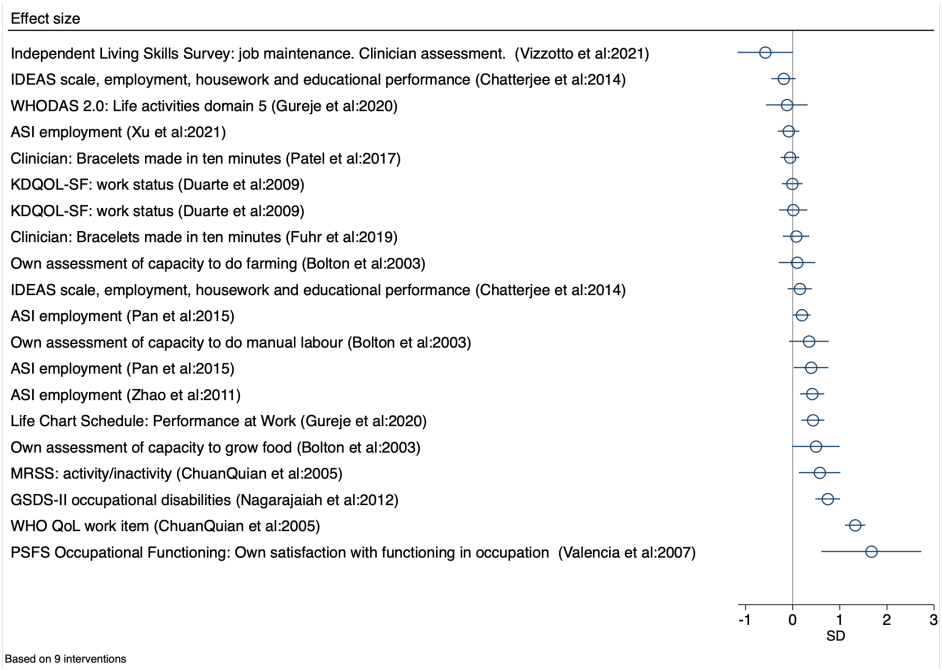
(d) Days unable to work



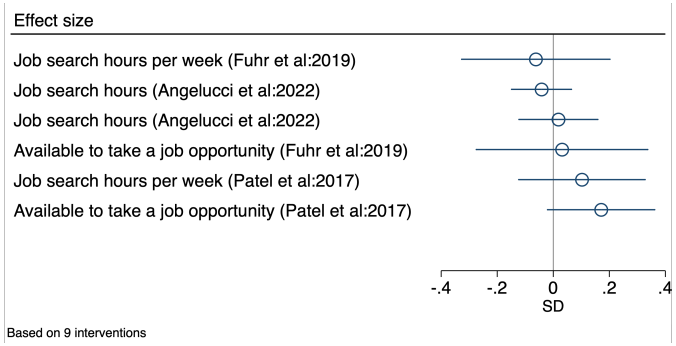
This figure displays a standardized version of the original work-related effect size estimates as reported in the included studies. The work-related effect sizes are shown separately by work-related outcome category and sorted by effect size.

Figure A3: Original work-related effect size estimate (cont.)

(a) Functioning at work



(b) Job search



This figure displays a standardized version of the original work-related effect size estimates as reported in the included studies. The work-related effect sizes are shown separately by work-related outcome category and sorted by effect size.

D Methods for effect size estimates

The true effect size (θ) is the mean difference between the treatment (μ_t) and control groups (μ_c) as a proportion of the standard deviation of the outcome variables (σ):

$$\theta = \frac{\mu_t - \mu_c}{\sigma} \tag{5}$$

An intuitive estimator for θ is Cohen's d (Cohen, 1988) defined by

$$d = \frac{\bar{Y_t} - \bar{Y_c}}{S_p} = \frac{D}{S_p} \tag{6}$$

where $\bar{Y_t}$ is the mean outcome of the treatment group and $\bar{Y_c}$ that of the control group. The numerator of d captures the unstandardized treatment effect and is often reported as a treatment effect parameter estimate, such as an ATT, ITT, or LATE, rather than as differences in means; thus we use D to denote an unstandardized treatment effect estimate. The denominator of d is the pooled standard deviation from the standard deviations of the treatment and control groups and is equivalent to

$$S_p = \sqrt{\frac{(n_t - 1) * S_t^2 + (n_c - 1) * S_c^2}{n_t + n_c - 2}} \tag{7}$$

where n_c and n_t are the sample sizes of the control and treatment groups, respectively, and S_c and S_t are the sample standard deviations of the control and treatment groups, respectively. It has been shown that d has a bias and overestimates the absolute value of the effect in small samples (Hedges, 1981). For this reason, we use a small sample size adjusted estimator referred to as Hedges' g, which is given by

$$g = d\left(1 - \frac{3}{4(n_t + n_c) - 9}\right) \tag{8}$$

The standard error of Hedges' g is given by

$$SE_g = \sqrt{\frac{n_t + n_c}{n_t * n_c} + \frac{g^2}{2 * (n_t + n_c)}} \tag{9}$$

A challenge encountered in the data extraction was the limited information available to compute the standardized mean difference (SD). Standard deviations for the treatment, control, and total sample groups were missing in 3 studies, even after attempting to correspond with authors to acquire this information. In such cases,

the standard deviation of the outcome variable was approximated using the formula from Borenstein et al. (2011):

$$S_p = SE * \sqrt{\frac{n_t * n_c}{n_t + n_c}} \tag{10}$$

where SE is the Standard Error of a comparison of means (e.g. standard error of the regression coefficient estimate). In the case of two studies, we were not even able to compute the standard deviation with the help of the above formula due to a lack of reported standard errors, so we used the standard deviations of the control group instead.

Creating one effect size estimate per intervention

Some studies provided more than one impact estimate for a given outcome type. To arrive at summary effect sizes per intervention and aggregated effect sizes, we combine them to arrive at a single effect size estimate per outcome for each intervention. Estimating summary effect sizes (for example on intervention level, outcome level, target group level, and other types of aggregates) requires a careful procedure to avoid permitting a single group of evaluation survey respondents to influence the aggregate disproportionately. The median number of treatment effect estimates per study was three, with some studies providing more than 20 estimates. In such instances, there can be a multitude of treatment effects reported for the same group where there is no *a priori* reason to give preference to one measure over another.

Where studies reported both pooled effect sizes and effect sizes for subgroups, we dropped those effect sizes that were redundant for the desired level of aggregation. The desired level was always the pooled estimate, except when looking at subgroup effects by gender.

Once redundant effect sizes were removed in some cases we still had multiple effect sizes for one independent group, without clear justification for dropping some over others – for example if an intervention measured one outcome in multiple ways. In order to arrive at one single effect size per intervention, we applied the method for combining effect sizes from the same independent population suggested by Borenstein et al. (2009). The approach is as follows: let g_{ij} and SE_g be the i^{th} effect size, where $i = 1, ..., m$ and its standard error, respectively, for the sample population (e.g. intervention) identified by j. To arrive at a single combined effect size for

intervention we take a simple average:

$$g_j = \frac{1}{m}\sum_{i=1}^{m} g_{ij} \tag{11}$$

and calculate the standard error of g_j by

$$SE_{g,j} = \sqrt{\left(\frac{1}{m}\right)^2 \Big(\sum_{i=1}^{m} SE_{g,i}^2 + \sum_{i\neq k} \rho_{i,k} SE_{g,ij} SE_{g,kj}\Big)}, \tag{12}$$

where $\rho_{i,k}$ is the correlation coefficient between g_{ij} and g_{kj}. Ideally we would estimate $\rho_{i,k}$ from the data. However, due to the lack sufficient number of observations an assumption on $\rho_{i,k}$ was required. The assumption of $\rho_{i,k} = 0$ would likely overestimate precision, while the assumption of $\rho_{i,k} = 1$ would likely underestimate precision. We take the more conservative assumption that $\rho_{i,k} = 1 \forall (i,j)$ where $i \neq k$. In other words, we assume perfect correlation across effect sizes for the same sample population.

Creating aggregate effect sizes for groups of interventions

With one effect size per intervention, we can create aggregate effect sizes for different categories of interventions (such as interventions conducted in high-income countries) as well as an aggregate effect size for the whole sample. Given the range of different interventions included in our sample, it is likely that each intervention's true effect size (θ_i) deviates from the true aggregate effect size for the overall group it belongs to. Furthermore, each observed effect size, estimated by Hedges' g, contains a sampling error. Therefore, g will either be less than or greater than θ_i. This can be expressed as

$$g_i = \mu + \zeta_i + \varepsilon_i = \theta_i + \varepsilon_i, \tag{13}$$

where μ is the true aggregate effect size for the group as a whole, ζ_i is the deviation of the true effect size of intervention i from the group's aggregate effect, and ε_i the sampling error. We estimate the true aggregate effect size for the group as a whole (μ) using a random-effects regression, following equation 13. Moreover, to obtain the most accurate estimate of μ, we estimate a weighted random effects model in which the weights are each study's inverse variance. Note that the study's variance corresponds to the term in equation 7 squared.

E Robustness checks

E.1 Heterogeneity in mental health effects

Figure A4: Robustness to disaggregation by measurement

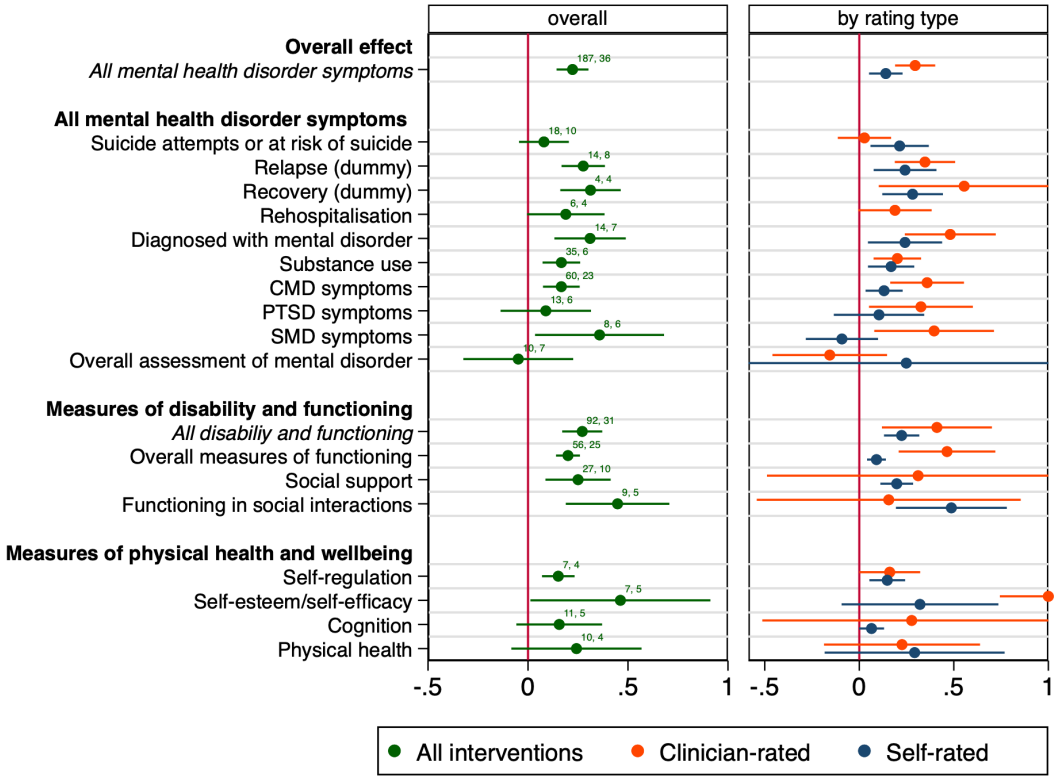


Figure A4 shows aggregate mental health meta-effect sizes (Hedges' g) for various behavioral and psychological pathways. The effects are shown for the overall sample of 39 interventions (panel 1, corresponding to table 4) and by rating type (panel 2). The horizontal axis displays the average mental health effect size in standard deviations. The aggregation of individual effect sizes works as described in subsection 4.1. Individual effect sizes are winsorized within outcome type at the 99^{th} percentile. In the first panel, the first number next to the effect size marker represents the number of individual effects going into the aggregate meta-effect, the second number represents the number of different interventions from where these individual effect sizes come. SSA = Sub-Saharan Africa, ECA = Europe and Central Asia, LAC = Latin America and the Caribbean, SAsia = South Asia, EAAP = East Asia and Pacific.

Figure A5: Mental health meta-analysis: robustness to disaggregation by intervention, targeted condition and region

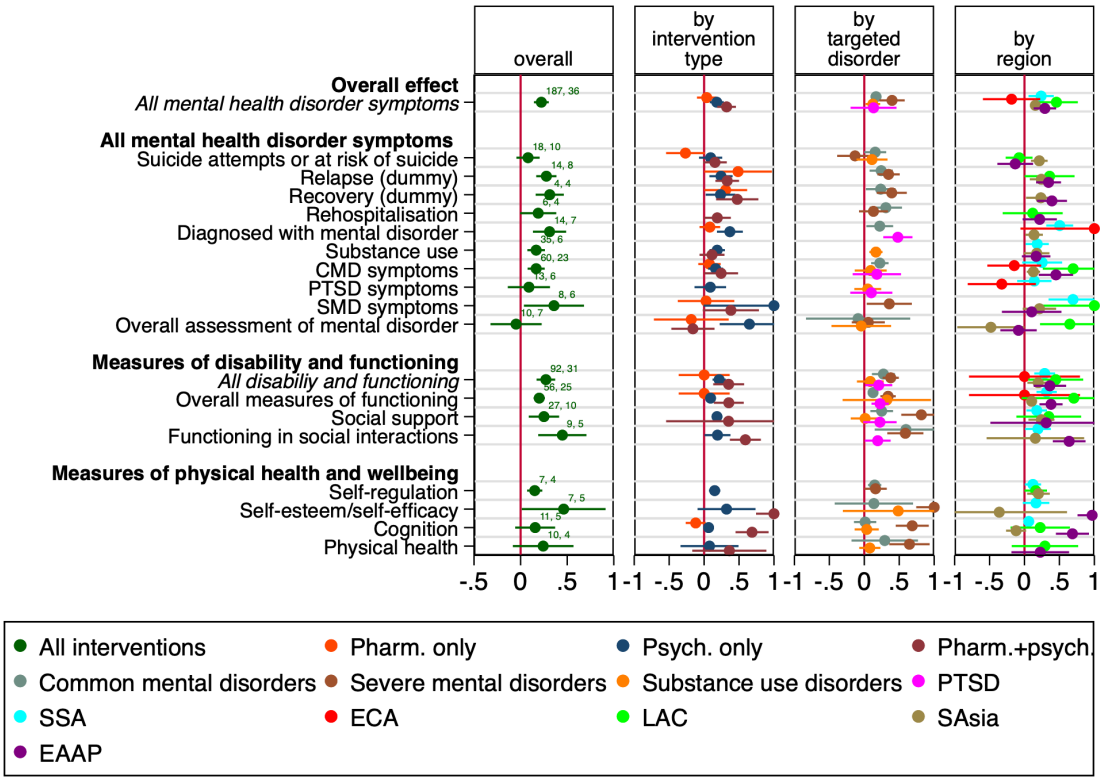


Figure A5 shows aggregate mental health meta-effect sizes (Hedges' g) for various behavioral and psychological pathways. The effects are shown for the overall sample of 39 interventions (panel 1, corresponding to table 4), by intervention (panel 2), by targeted disorder (panel 3) and by region (panel 4). The horizontal axis displays the average mental health effect size in standard deviations. The aggregation of individual effect sizes works as described in subsection 4.1. Individual effect sizes are winsorized within outcome type at the 99^{th} percentile. In the first panel, the first number next to the effect size marker represents the number of individual effects going into the aggregate meta-effect, the second number represents the number of different interventions from where these individual effect sizes come. SSA = Sub-Saharan Africa, ECA = Europe and Central Asia, LAC = Latin America and the Caribbean, SAsia = South Asia, EAAP = East Asia and Pacific.

E.2 Study characteristics

In Table A11, we show robustness of our finding to key study characteristics, as outlined in Section 6.

Table A11: Robustness

	Dep. var.: work-related outcomes (Hedges' g)					
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	0.10***	0.10***	0.09***	0.11***	0.08***	0.07*
	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.04)
Variance of error term	Yes	Yes	Yes	Yes	Yes	Yes
Intervention category	Yes	Yes	Yes	Yes	Yes	Yes
Control conditions	Yes	Yes	Yes	Yes	Yes	Yes
Outcome type	No	Yes	Yes	Yes	No	Yes
Target condition	No	Yes	Yes	Yes	No	Yes
Target sample	No	No	Yes	Yes	No	Yes
Implementer	No	No	No	Yes	No	Yes
Measurement timing	No	No	No	No	Yes	Yes
Observations	95	95	95	95	95	95
Number of interventions	34	34	34	34	34	34

Notes: Table A11 displays the average meta-analysis effect based on a multivariate meta-regression of Hedges' g on various groups of independent variables. Hedges' g is the small-sample-bias-corrected standardized mean difference in the economic outcome between treatment and control. *, ** and *** denote statistical significance at 10 percent, 5 percent and 1 percent level of significance respectively. The average measurement in our sample happens 15.2 months after intervention start. ^{1} = reverse coded, so higher values mean better employment outcomes. In this regression, we use the sample of 95 work-related effect sizes to retain as much variation with respect to covariates as possible. All individual effect sizes are winsorized at the 99^{th} percentile.

In addition to these core robustness checks, we test the robustness of our main meta-regression findings in three different ways: (i) by separately excluding the three studies providing the largest number of effect sizes to our sample, (ii) by grouping interventions in subsamples according to their control conditions, (iii) by looking at various other subgroups.

First, we exclude the studies providing the largest number of effect sizes to our sample, which are Baranov et al. (2020), Blattman et al. (2017) and Angelucci and Bennett (2022). Together, these three studies cover 29% of our effect sizes. Table A12 replicates the first specification of our full sample in Table 2, and then excludes one large study at a time. Our main results (repeated for ease of comparison in Table A12, row 1) remain remarkably stable in rows 2-4: none of the coefficients change in any meaningful way.

Second, in rows 5-6, we run our basic regression on the subsamples of studies only with certain control conditions.

In rows 7-9, we demonstrate that our main finding is robust to excluding various subgroups of studies: those that cost over USD 100 per participant, those reporting social network outcomes, and those without young participants below 18 years of age.

Table A12: Robustness of effects to i) exclusion of largest studies, ii) types of control conditions, and iii) various other subgroups

	(1) Aggregate Hedges' g	(2) 95% lower CI	(3) 95% upper CI	(4) # of observations	(5) # of interventions
All interventions	0.15***	0.09	0.21	180	39
No Baranov	0.15***	0.09	0.21	171	38
No Blattman	0.15***	0.09	0.22	168	38
No Angelucci	0.16***	0.09	0.22	148	38
Only no treatment	0.06**	0.01	0.10	115	19
Only treatment as usual	0.33***	0.15	0.51	22	13
Costs <100 USD	0.10*	-0.01	0.20	52	6
No social network outcomes	0.15***	0.09	0.21	175	38
No part. <18	0.17***	0.08	0.25	149	29

Notes: Hedges' g is the small-sample-bias-corrected standardized mean difference in the economic outcome between treatment and control. *, ** and *** denote statistical significance at 10 percent, 5 percent and 1 percent level of significance respectively. The average measurement in our sample happens 15.2 months after intervention start. ^{1} = reverse coded, so higher values mean better employment outcomes. Aggregate Hedges' g represents an estimate from random effects inverse variance weighted regression. The aggregation of individual effect sizes works as described in subsection 4.1. All individual effect sizes are winsorised at the 99^{th} percentile.

E.3 Publication bias

E.3.1 Funnel plot asymmetry

Table A13: Egger's test

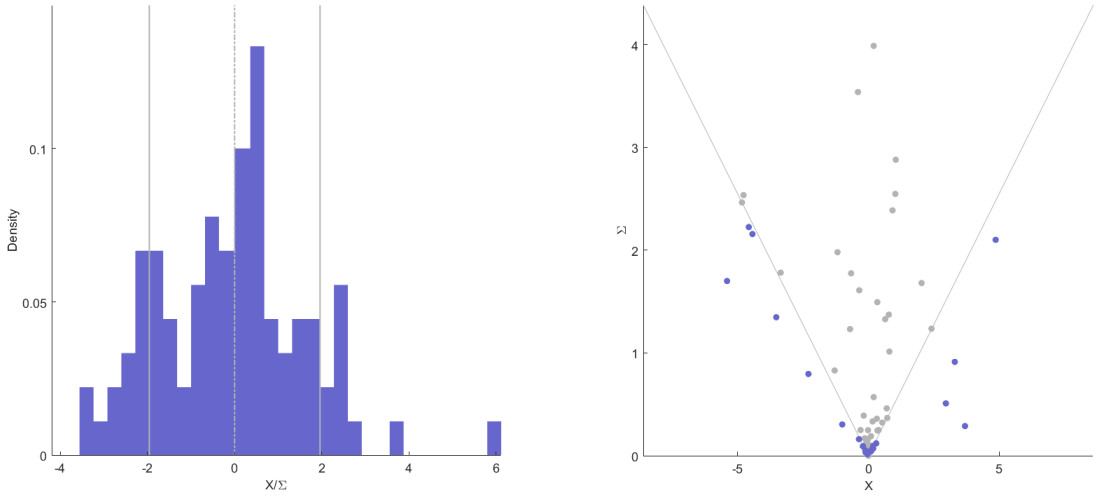
	(1) Egger's test H0: no small-study effects
Beta	0.01
S.E.	0.19
p-value	0.97

Notes: Table A13 displays the results of the Egger's test. We cannot reject the null hypothesis of no small-study effects. The sample size are $N = 95$ effect sizes of mental health treatment impacts on work-related outcomes.

E.3.2 Conditional publication probability model

We follow the maximum likelihood approach of Andrews and Kasy (2019) to formally model the effect of publication bias in our setting. Under the standard independence assumption and under no selectivity, we can write the distribution of estimates for high variance studies as the distribution for low variance studies plus a noise term.

Figure A6: Histogram and funnel plot of reported effect sizes



Notes: Figure A6 displays a binned density plot (histogram) for the Z-statistics recovered from our study sample, X/Σ, while the right panel shows a funnel plot, which plots effect sizes, X, against their standard errors, Σ. To enable us to visually distinguish reported data, we trim "extremely small sample" observations for which $\Sigma > 8$. We show robustness to their inclusion of exclusion in the formal analysis that follows. The grey lines indicate $X/\Sigma = 1.96$, which is the threshold for 95% significance. Substantial bunching around those thresholds would provide tentative evidence for publication bias. Moreover, asymmetry in the funnel plot for higher values of Σ might indicate small sample effects and publication bias.

Deviations from this prediction identify differential publication probabilities conditional on Z-scores. In particular, if we assume $P(pub|Z > 1.96) = 1$, and that the error term follows a t-distribution, which allows for differential publication probabilities whether the result is positive or negative, we can fit the following model using maximum likelihood estimation.

$$\Theta^* \sim \bar{\theta} + t(\tilde{\nu}) \cdot \tilde{\eta}, \quad p(Z) \propto \begin{cases} \beta_{p,1} & \text{if } Z < -1.96 \\ \beta_{p,2} & \text{if } Z \in [-1.96, 0) \\ \beta_{p,3} & \text{if } Z \in [0, 1.96) \\ 1 & \text{if } Z \geq 1.96 \end{cases} \tag{14}$$

Where θ^* is the distribution of latent study effects (whether published or unpublished), $\bar{\theta}$ is the average effect size for large studies, t represents a t-distribution and $\tilde{\nu}$ and $\tilde{\eta}$ its degrees of freedom and scale parameter, respectively. We cluster standard errors by study to account for non-independence of within study-reported outcomes. We report findings for the whole sample in Table A14.

Table A14: Differential publication probability estimates

$\bar{\theta}$	$\tilde{\tau}$	$\tilde{\nu}$	$\beta_{p,1}$	$\beta_{p,2}$	$\beta_{p,3}$
0.195	-0.800	5.017	1.074	1.833	1.912
(0.052)	(0.181)	(4.440)	(0.473)	(0.577)	(0.449)

Notes: Table A14 displays the results of the MLE model for publication bias implemented on the whole sample. $\bar{\theta}$ represents the estimated average effect size for large studies. Publication probability β_p is measured relative to the omitted category of studies which are positive and significant at the 5 percent level. $\beta_{p,1}$ represents the probability of publication given $Z < -1.96$, $\beta_{p,2}$, $Z \in [-1.96, 0]$ and $\beta_{p,3}$, $Z \in [0, 1.96]$. Standard errors clustered by study are reported in parentheses.

Taken literally, our point estimates indicate that relative to the reference category for which $Z > 1.96$, the probability of publication of other effect sizes being published is higher. However, these probabilities are imprecisely estimated, and we interpret them as indicating that we have little evidence of differential publication probabilities conditional on Z-scores in our study sample. That is, we have no direct evidence of publication bias.

We then replicate the model in the sub-sample for which standard errors are less than 10 (we have one observation for which SE = 8). Our findings are broadly similar, but significantly more precisely estimated.

Table A15: Differential publication probability estimates for $SE \leq 8$ subsample

$\bar{\theta}$	$\tilde{\tau}$	$\tilde{\nu}$	$\beta_{p,1}$	$\beta_{p,2}$	$\beta_{p,3}$
0.185	0.320	5.035	0.954	1.421	1.379
(0.040)	(0.070)	(0.390)	(0.247)	(0.385)	(0.294)

Notes: Table A14 displays the results of the MLE model for publication bias implemented on the sub-sample for which $SE \leq 8$. $\bar{\theta}$ represents the estimated average effect size for large studies. Publication probability β_p is measured relative to the omitted category of studies which are positive and significant at the 5 percent level. $\beta_{p,1}$ represents the probability of publication given $Z < -1.96$, $\beta_{p,2}$, $Z \in [-1.96, 0]$ and $\beta_{p,3}$, $Z \in [0, 1.96]$. Standard errors clustered by study are reported in parentheses.

F Variable construction mega-analysis

For the mega-analysis, we construct our main economic outcome of interest, number of days worked per month, in the following way:

- Patel et al. (2010): 'Number of days unable to work in the previous month' is directly measured. We subtract this number from 28, to arrive at the number of days worked per month.
- Fuhr et al. (2019): 'Number of days unable to work in the previous month' is directly measured. We subtract this number from 28, to arrive at the number of days worked per month.
- Sikander et al. (2019): 'Number of days unable to work in the previous month' is directly measured. We subtract this number from 28, to arrive at the number of days worked per month.
- Baranov et al. (2020): 'Number of healthy days in past 30 days' is directly measured – we use it as an approximation for the number of days able to work.
- Meffert et al. (2021): 'Number of days unable to work in the previous month' is directly measured. We subtract this number from 28, to arrive at the number of days worked per month.
- Nadkarni et al. (2019): 'Number of days unable to work in the previous month' is directly measured. We subtract this number from 28, to arrive at the number of days worked per month.
- Barker et al. (2022): 'Number of days unable to work in the previous month' is directly measured. We subtract this number from 28, to arrive at the number of days worked per month.

G Additional results

G.1 Instrumental variable estimation details: first stage and reduced form

Table A16: First stage: effect of CBT treatment on depression

	Depression				
	(1)	(2)	(3)	(4)	(5)
	Combined measure	DSM-IV	PHQ-9	BDI	Kessler
Treatment	-0.195***	-0.167**	-0.190**	-0.497***	-0.198***
	(0.055)	(0.075)	(0.085)	(0.146)	(0.036)
Constant	0.000	-0.000	-0.000	-0.000	0.000
	(0.039)	(0.048)	(0.075)	(0.093)	(0.029)
Wave FE	Yes	Yes	Yes	Yes	Yes
Study FE	Yes	No	Yes	No	No
Months after treatment	Yes	Yes	Yes	Yes	Yes
Control mean	0.00	-0.00	0.00	-0.00	0.00
Obs.	17210	885	8909	195	7221
Studies	7	1	4	1	1

Notes: This table shows five different OLS regression of the outcome variable on the treatment indicator as well as study fixed effects, the endline round, and the number of months after treatment when the outcome was measured. Column 1 shows the impact on a combined depression outcome, columns 2-5 show the impact on depression measured by DSM-IV, PHQ-9, BDI, or Kessler, respectively (all standardized). Standard errors are in parentheses and clustered by original study cluster variable. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A17: Reduced form: effect of CBT treatment on days worked

	(1)	(2)	(3)
	Combined days able to work measure	Healthy days	Days unable to work
Treatment	1.449**	0.491	-1.488**
	(0.689)	(0.657)	(0.717)
Constant	22.151***	26.275***	6.328***
	(0.476)	(0.475)	(0.492)
Wave FE	Yes	Yes	Yes
Study FE	Yes	No	Yes
Months after treatment	Yes	Yes	Yes
Control mean	22.65	26.27	5.89
Obs.	18289	621	17668
Studies	7	1	6

Notes: This table shows three different OLS regression of the outcome variable on the treatment indicator as well as study fixed effects, the endline round, and the number of months after treatment when the outcome was measured. Columns 1 shows the impact on a combined days worked per month outcome, column 2 shows the impact on healhy days per month, column 3 show the impact on days unable to work in the last month. Standard errors are in parentheses and clustered by original study cluster variable. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

G.2 Costs

Table A18: Cost overview

	In 2011 USD	(10^{th} pct)	(90^{th} pct)
All interventions	362.52	23.97	1341.35
By intervention type			
Pharmacological	233.69	227.95	239.43
Psychosocial	509.84	7.11	1599.49
Pharm.+psych.	210.58	55.69	570.96
By region			
East Asia & Pacific	288.69	55.69	570.96
Europe & Central Asia	1456.28	1456.28	1456.28
Latin America & Caribbean			
South Asia	94.21	12.79	180.09
Sub-Saharan Africa	949.36	370.37	1742.69

Notes: This table shows intervention costs per participant in 2011 US-Dollars. Column 1 shows the mean, columns 2 and 3 show the 10th and 90th percentile.

G.3 Combined interventions

Table A19: The impact of psychosocial and economic interventions

	(1) Aggregate Hedges' g	(2) 95% lower CI	(3) 95% upper CI	(4) # of observations	(5) # of interventions
Total	0.17***	0.11	0.23	201	44
Intervention type					
Pharmacological	0.05	-0.09	0.19	38	2
Psychosocial	0.06**	0.01	0.11	109	22
Pharm.+psych	0.30***	0.18	0.43	35	15
Psych.+econ	0.29***	0.11	0.48	19	5
Work-related outcomes, with econ. add-on					
All work-related outcomes	0.40	-0.12	0.92	4	3
In employment	0.66***	0.31	1.00	2	2
Time in work	0.00	-0.20	0.21	2	1

Notes: Hedges' g is the small-sample-bias-corrected standardized mean difference in the economic outcome between treatment and control. *, ** and *** denote statistical significance at 10 percent, 5 percent and 1 percent level of significance respectively. The average measurement in our sample happens 15.2 months after intervention start. The I^2 statistic is the percentage of between-study heterogeneity that is attributable to variability in the true treatment effect, rather than sampling variation. Aggregate Hedges' g represents an estimate from random effects inverse variance weighted regression. The aggregation of individual effect sizes works as described in subsection 4.1. All individual effect sizes are winsorised at the 95^{th} percentile.

H Comparison to high income country effect sizes

Table A20: Comparison to Developed Country Effect Sizes: Economic Outcomes

Outcome	Study	Study population	Effect Size	95% CI
Aggregate				
Labour supply effects[1]	Timbie et al. (2006)	USA	$d = 0.12$	[0.00, 0.24]
Employment Dummy				
Employment rate	Chan et al. (2015)	USA, Italy, Germany, Singapore, Japan	$SD = 0.15$	[0.04, 0.25]
Time in work				
Hours worked [2]	van Duin et al. (2019)	USA, Israel, Germany, UK and Japan	$SD = 0.31$	[0.04, 0.58]
Employment frequency (SD) [3]	Chan et al. (2015)	USA, Italy, Germany, Singapore, Japan	$SD = 0.15$	[0.02, 0.28]
Employment frequency [3]	Chan et al. (2015)	USA, Italy, Germany, Singapore, Japan	19.5 more days / year	[2.5, 36.6]
Days unable to work				
Sick leave [4]	Nieuwenhuijsen et al. (2020)	Europe, USA, Australia, Canada	-14.7 days / year	[-27.6, -3.0]
Sick leave				
Sick leave [4a b]	Nieuwenhuijsen et al. (2020)	Europe, USA, Australia, Canada	SD = 0.15	[0.03, 0.28]
Sick leave [5a]	Salomonsson et al. (2018)	Europe, USA, India	$g = 0.15$	[0.08, 0.22]
Sickness absence [6]	Finnes et al. (2019)	Netherlands, Denmark, USA, Sweden	$g = 0.17$	[-0.03, 0.36]
Functioning at work				
Functioning (Psych) [7]	Kamenov et al. (2017a)	USA, UK,Netherlands	$g = 0.43$	[0.33, 0.54]
Functioning (Pharm) [8]	Kamenov et al. (2017a)	USA, UK, Netherlands	$g = 0.31$	[0.26, 0.36]
Coping with work [9]	Nieuwenhuijsen et al. (2020)	Europe, USA, Australia and Canada	SD = 0.05	[-0.46, 0.57]
Income, etc.[10]				
Wages [11]	van Duin et al. (2019)	USA, Israel, Germany, UK and Japan	SD = 0.25	[-0.07, 0.58]

Notes: g = Hedges' g; d = Cohen's d; and SD = Standard Deviation. Outcomes marked [(a)] have been reverse-coded, so that for all measures higher values can be interpreted as indicating better employment outcomes. [(b)] Nigatu et al. (2016) and Nieuwenhuijsen et al. (2020) did not reverse code Sick Days, so we have given the absolute value of their effect size for comparison. [1] Labour supply effects: includes hours worked per week, odds of being unable to work, days of employment, percent employed; [2] From paper: "number of hours worked in all paid employment settings during the study interval"; [3] total days work in a year; [4] Either measured as sickness absence days during the follow up period or employment status ('off work' or 'at work') after a given period of time; [5] both self-reported and administrative sick-leave data including both individuals on sick leave and those at risk of sick leave; [6] Defined as sickness absence, return to work, or increased working hours; [7] Outcomes of psychological interventions on validated measure of functioning; [8] Outcomes of pharmaceutical interventions on validated measure of functioning; [9] how well people with depression could cope with their work; [10] hourly wages in USD from competitive and all paid employment;

I Included studies

References

Angelucci, Manuela and Daniel Bennett, "The Economic Impact of Depression Treatment in India: Evidence from Community-Based Provision of Pharmacotherapy," 2022.

Ayoughi, Sarah, Inge Missmahl, Roland Weierstall, and Thomas Elbert, "Provision of mental health services in resource-poor settings: a randomised trial comparing counselling with routine medical treatment in North Afghanistan (Mazar-e-Sharif)," *BMC Psychiatry*, 2012, *12* (1), 14.

Baranov, Victoria, Sonia Bhalotra, Pietro Biroli, and Joanna Maselko, "Maternal Depression, Women's Empowerment, and Parental Investment: Evidence from a Randomized Controlled Trial," *American Economic Review*, 2020, *110* (3), 824–59.

Barker, Nathan, Gharad Bryan, Dean Karlan, Angela Ofori-Atta, and Christopher Udry, "Cognitive Behavioral Therapy among Ghana's Rural Poor Is Effective Regardless of Baseline Mental Distress," *American Economic Review: Insights*, December 2022, *4* (4), 527–45.

Betancourt, Theresa S, Ryan McBain, Elizabeth A Newnham, Adeyinka M Akinsulure-Smith, Robert T Brennan, John R Weisz, and Nathan B Hansen, "A Behavioral Intervention for War-Affected Youth in Sierra Leone: A Randomized Controlled Trial," *Journal of the American Academy of Child Adolescent Psychiatry*, 2014, *53* (12), 1288–1297.

Bhat, Bhargav, Jon de Quidt, Johannes Haushofer, Vikram Patel, Gautam Rao, Frank Schilbach, and Pierre Vautrey, "The Long-Run Effects of Psychotherapy on Depression, Beliefs, and Economic Outcomes," *National Bureau of Economic Research Working Paper*, 2022, *30011*.

Blattman, Christopher, Julian C. Jamison, and Margaret Sheridan, "Reducing Crime and Violence: Experimental Evidence from Cognitive Behavioral Therapy in Liberia," *American Economic Review*, April 2017, *107* (4), 1165–1206.

Bolton, Paul, Judith Bass, Richard Neugebauer, Helen Verdeli, Kathleen F Clougherty, Priya Wickramaratne, Liesbeth Speelman, Lincoln Ndogoni, and Myrna Weissman, "Group Interpersonal Psychotherapy for Depression in Rural UgandaA Randomized Controlled Trial," *JAMA*, jun 2003, *289* (23), 3117–3124.

Buttorff, Christine, Rebecca S Hock, Helen A Weiss, Smita Naik, Ricardo Araya, Betty R Kirkwood, Daniel Chisholm, and Vikram Patel, "Economic evaluation of a task-shifting intervention for common mental disorders in India," *Bulletin of the World Health Organization*, 2012, *90*, 813–821.

Chang, Wei-Lun, Carit Jacques Andersen, Besa Shatri Berisha, Olena Estrup, and Shr-Jie Wang, "Economic evaluation of a bio-psycho-social intervention for comorbid disorders in a traumatized population in post-war Kosovo," *Journal of rehabilitation medicine*, 2018, *50* (5), 444–450.

Chatterjee, Sudipto, Smita Naik, Sujit John, Hamid Dabholkar, Madhumitha Balaji, Mirja Koschorke, Mathew Varghese, Rangaswamy Thara, Helen A Weiss, Paul Williams, Paul McCrone, Vikram Patel, and Graham Thornicroft, "Effectiveness of a community-based intervention for people with schizophrenia and their caregivers in India (COPSI): a randomised controlled trial," *The Lancet*, 2014, *383* (9926), 1385–1394.

Chuan-qian, Qiu, Zeng Zhao-xiang, and Li Zhi-cheng, "Influence of family intervention by physician on the social function and quality of life in community schizophrenic patients: a follow-up for half a year," *Chinese Journal of Clinical Rehabilitation*, 2005, *12*, 75–77.

Cilliers, Jacobus, Oeindrila Dube, and Bilal Siddiqi, "Reconciling after civil conflict increases social capital but decreases individual well-being," *Science*, may 2016, *352* (6287), 787–794.

Duarte, Priscila Silveira, Maria Cristina Miyazaki, Sergio Luís Blay, and Ricardo Sesso, "Cognitive–behavioral group therapy is an effective treatment for major depression in hemodialysis patients," *Kidney International*, 2009, *76* (4), 414–421.

Fuhr, Daniela C, Benedict Weobong, Anisha Lazarus, Fiona Vanobberghen, Helen A Weiss, Daisy Radha Singla, Hanani Tabana, Ejma Afonso, Aveena De Sa, Ethel D'Souza et al., "Delivering the Thinking Healthy Programme for perinatal depression through peers: an individually randomised controlled trial in India," *The Lancet Psychiatry*, 2019, *6* (2), 115–127.

Gureje, Oye, John Appiah-Poku, Toyin Bello, Lola Kola, Ricardo Araya, Dan Chisholm, Oluyomi Esan, Benjamin Harris, Victor Makanjuola, Caleb Othieno, LeShawndra Price, and Soraya Seedat, "Effect of collaborative care between traditional and faith healers and primary health-care workers on psychosis outcomes in Nigeria and Ghana (COSIMPO): a cluster randomised controlled trial," *The Lancet*, 2020, *396* (10251), 612–622.

Hall, Brian J, Paul A Bolton, Jeannie Annan, Debra Kaysen, Katie Robinette, Talita Cetinoglu, Karin Wachter, and Judith K Bass, "The Effect of Cognitive Therapy on Structural Social Capital: Results From a Randomized Controlled Trial Among Sexual Violence Survivors in the Democratic Republic of the Congo," *American Journal of Public Health*, jul 2014, *104* (9), 1680–1686.

Haushofer, Johannes, Robert Mudida, and Jeremy P Shapiro, "The Comparative Impact of Cash Transfers and a Psychotherapy Program on Psychological and Economic Well-being," Working Paper 28106, National Bureau of Economic Research November 2020.

Hirani, Saima Shams, Rozina Karmaliani, Judith McFarlane, Nargis Asad, Farhana Madhani et al., "Testing a community derived intervention to promote women's health: preliminary results of a 3-arm randomized controlled trial in Karachi, Pakistan."

Jothimani, Gurusamy, Ramakrishnan Parthasarathi, Konduru Reddemma, Amruthur Thimmaiah Shesh Giri et al., "Efficacy of nursing interventions in reducing social and occupational disabilities among patients with neurosis," *Nursing Journal of India*, 2012, *103* (6), 256.

Luo, Xingwei, Samuel F Law, Xiang Wang, Jingzheng Shi, Wu Zeng, Xiaoqian Ma, Wendy Chow, Shiyan Liu, Wei Zhao, Xiaoli Liu, Shuqiao Yao, and Michael R Phillips, "Effectiveness of an Assertive Community Treatment program for people with severe schizophrenia in mainland China – a 12-month randomized controlled trial," *Psychological Medicine*, 2019, *49* (6), 969–979.

Meffert, Susan M, Thomas C Neylan, Charles E McCulloch, Kelly Blum, Craig R Cohen, Elizabeth A Bukusi, Helen Verdeli, John C Markowitz, James G Kahn, David Bukusi, Harsha Thirumurthy, Grace Rota, Ray Rota, Grace Oketch, Elizabeth Opiyo, and Linnet Ongeri, "Interpersonal psychotherapy delivered by nonspecialists for depres-

sion and posttraumatic stress disorder among Kenyan HIV–positive women affected by gender-based violence: Randomized controlled trial," *PLOS Medicine*, 2021, *18* (1), 1–22.

Nadkarni, Abhijit, Benedict Weobong, Helen A Weiss, Jim McCambridge, Bhargav Bhat, Basavaraj Katti, Pratima Murthy, Michael King, David McDaid, A-La Park, G Terence Wilson, Betty Kirkwood, Christopher G Fairburn, Richard Velleman, and Vikram Patel, "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*, 2017, *389* (10065), 186–195.

_ **, Helen A Weiss, Benedict Weobong, David McDaid, Daisy R Singla, A-La Park, Bhargav Bhat, Basavaraj Katti, Jim McCambridge, Pratima Murthy, Michael King, G Terence Wilson, Betty Kirkwood, Christopher G Fairburn, Richard Velleman, and Vikram Patel**, "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*, 2017, *14* (9), 1–20.

_ **,** _ **, Richard Velleman, Jim McCambridge, David McDaid, A-La Park, Pratima Murthy, Benedict Weobong, Bhargav Bhat, and Vikram Patel**, "Feasibility, acceptability and cost-effectiveness of a brief, lay counsellor-delivered psychological treatment for men with alcohol dependence in primary care: an exploratory randomized controlled trial," *Addiction*, 2019, *114* (7), 1192–1203.

Pan, Shujun, Haifeng Jiang, Jiang Du, Hanhui Chen, Zhibin Li, Walter Ling, and Min Zhao, "Efficacy of Cognitive Behavioral Therapy on Opiate Use and Retention in Methadone Maintenance Treatment in China: A Randomised Trial," *PLOS ONE*, jun 2015, *10* (6), e0127598.

Patel, Vikram, Benedict Weobong, Helen A Weiss, Arpita Anand, Bhargav Bhat, Basavraj Katti, Sona Dimidjian, Ricardo Araya, Steve D Hollon, Michael King, Lakshmi Vijayakumar, A-La Park, David McDaid, Terry Wilson, Richard Velleman, Betty R Kirkwood, and Christopher G Fairburn, "The Healthy Activity Program (HAP), a lay counsellor-delivered brief psychological treatment for severe depression, in primary care in India: a randomised controlled trial," *The Lancet*, 2017, *389* (10065), 176–185.

_ **, Helen A Weiss, Neerja Chowdhary, Smita Naik, Sulochana Pednekar, Sudipto Chatterjee, Bhargav Bhat, Ricardo Araya, Michael King, Gregory Simon, Helena Verdeli, and Betty R Kirkwood**, "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*, 2011, *199* (6), 459–466.

_ **,** _ **,** _ **,** _ **,** _ **,** _ **, Mary J De Silva, Bhargav Bhat, Ricardo Araya, Michael King et al.**, "Effectiveness of an intervention led by lay health counsellors for depressive and anxiety disorders in primary care in Goa, India (MANAS): a cluster randomised controlled trial," *The Lancet*, 2010, *376* (9758), 2086–2095.

Ran, Mao-Sheng, Cecilia Lai-Wan Chan, Siu Man Ng, Ling-Tao Guo, and Meng-Ze Xiang, "The effectiveness of psychoeducational family intervention for patients with schizophrenia in a 14-year follow-up study in a Chinese rural area," *Psychological Medicine*, 2015, *45* (10), 2197–2204.

_ **, Meng-Ze Xiang, Cecilia Lai-Wan Chan, Julian Leff, Peggy Simpson, Ming-Sheng Huang, You-He Shan, and Si-Gan Li**, "Effectiveness of psychoeducational intervention

for rural Chinese families experiencing schizophrenia," *Social Psychiatry and Psychiatric Epidemiology*, 2003, *38* (2), 69–75.

Sikander, Siham, Ikhlaq Ahmad, Najia Atif, Ahmed Zaidi, Fiona Vanobberghen, Helen A Weiss, Anum Nisar, Hanani Tabana, Qurat Ul Ain, Amina Bibi et al., "Delivering the Thinking Healthy Programme for perinatal depression through volunteer peers: a cluster randomised controlled trial in Pakistan," *The Lancet Psychiatry*, 2019, *6* (2), 128–139.

Valencia, Marcelo, Maria Luisa Rascon, Francisco Juarez, and Esther Murow, "A psychosocial skills training approach in Mexican out-patients with schizophrenia," *Psychological Medicine*, 2007, *37* (10), 1393–1402.

Vizzotto, Adriana, Diego Celestino, Patricia Buchain, Alexandra Oliveira, Graça Oliveira, Elaine Di Sarno, Isabel Napolitano, and Helio Elkis, "Occupational Goal Intervention Method for the Management of Executive Dysfunction in People With Treatment-Resistant Schizophrenia: A Randomized Controlled Trial," *The American Journal of Occupational Therapy*, apr 2021, *75* (3), 7503180050.

Weobong, Benedict, Helen A Weiss, David McDaid, Daisy R Singla, Steven D Hollon, Abhijit Nadkarni, A-La Park, Bhargav Bhat, Basavraj Katti, Arpita Anand, Sona Dimidjian, Ricardo Araya, Michael King, Lakshmi Vijayakumar, G Terence Wilson, Richard Velleman, Betty R Kirkwood, Christopher G Fairburn, and Vikram Patel, "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*, sep 2017, *14* (9), e1002385.

Xiang, Mengze, Maosheng Ran, and Sigan Li, "A Controlled Evaluation of Psychoeducational Family Intervention in a Rural Chinese Community," *British Journal of Psychiatry*, 1994, *165* (4), 544–548.

Xiong, Hu, Wang Yu-ling, and Fu Hui-peng, "Synthetical family treatment for depression: a randomized-controlled single-blind study among 76 cases," *Zhongguo Zuzhi Gongcheng Yanjiu Yu Linchuang Linchuang Kangfu*, 2007, *11* (39), 7787.

Xiong, Wei, M R Phillips, Xiong Hu, Ruiwen Wang, Qinqing Dai, J Kleinman, and A Kleinman, "Family-Based Intervention for Schizophrenic Patients in China: A Randomised Controlled Trial," *British Journal of Psychiatry*, 1994, *165* (2), 239–247.

Xu, Xiaomin, Shujuan Chen, Junning Chen, Zhikang Chen, Liming Fu, Dingchen Song, Min Zhao, and Haifeng Jiang, "Feasibility and Preliminary Efficacy of a Community-Based Addiction Rehabilitation Electronic System in Substance Use Disorder: Pilot Randomized Controlled Trial," *JMIR Mhealth Uhealth*, 2021, *9* (4), e21087.

Zhang, Mingyuan, Yangling He, Martin Gittelman, Zhen Wong, and Heqin Yan, "Group psychoeducation of relatives of schizophrenic patients: Two-year experiences," *Psychiatry and Clinical Neurosciences*, dec 1998, *52* (S6), S344–S347.

Zhao, Wei, Samuel Law, Xingwei Luo, Wendy Chow, Jinqiang Zhang, Ye Zhu, Shiyan Liu, Xiaoqian Ma, Shuqiao Yao, and Xiang Wang, "First Adaptation of a Family-Based ACT Model in Mainland China: A Pilot Project," *Psychiatric Services*, jan 2015, *66* (4), 438–441.

Table A21: Included studies

Author, year	Country	Sample & age	Intervention category	Therapeutic type	Control group category	Description of intervention	Follow-up time points*	Target mental disorder	Economic outcomes**
Common Mental Disorders (CMD)									
Angelucci et al. 2022	India	602 adults 18+	Pharmacological	Medication: Antidepressants	No treatment	8 sessions monthly	-2, 13 months post intervention	Depression	Assets, Education, Income, Consumption & input expenditure, Job search, Time in work
Ayoughi et al. 2012	Afghanistan	61 women, 14+	Psychosocial	Problem solving therapy	TAU Pharm: Antidepressants	5-8 sessions over 2 months	0.5 months	Depression	Subjective poverty measures
Baranov et al. 2017 (Rahman et al. 2008)	Pakistan	903 women, 16-45	Psychosocial	Cognitive Behavioural Therapy	EUC: Monthly home visits by Lady Health Workers	16 sessions over 11 months	8, 14 and 86 months	Depression	Financial, Education, Employment, Health (fertility), Wellbeing
Barker et al. 2022	Ghana	7227 adults 18+	Psychosocial	Cognitive Behavioural Therapy	No treatment	1 session weekly for 12 weeks (3 months)	1-3 months after intervention	Common mental disorders	Subjective poverty measures, Days unable to work
Bolton et al. 2003	Uganda	216 adults, 18+	Psychosocial	Interpersonal Therapy	No treatment	1 session weekly for 16 weeks (4 months)	0 and 6 months	Depression	Agriculture, Education, Social networks
Duarte et al. 2009	Brazil	90 adults, 18+	Psychosocial	Cognitive Behavioural Therapy	No treatment	1 session weekly for 12 weeks (3 months)	0 and 6 months	Depression	Social Networks, Employment
Fuhr et al. 2019; Bhat et al. 2022	India	250 women, 18+	Psychosocial	Behavioural Activation	EUC: Patient information leaflet	6-14 sessions over 7-12 months	-3.5; -6.5; 38.5 months	Depression	Income, Consumption & input expenditure, Days unable to work, Employment, Functioning at work, Job search, Time in work
Haushofer et al. 2022	Kenya	2122 adults 18+	Psychosocial	Problem Solving Therapy, Behavioural Activation	No treatment	1 session weekly for 5 weeks	13	Common Mental Disorders	Assets, Income, consumption & input expenditure
Hirani et al. 2010	Pakistan	24 women, 25-35	Psychosocial	Problem Solving Therapy, Stress & Anger management, Communication skills	No treatment	1 session weekly for 8 weeks (2 months)	0.5 months	Depression	Employment
Hu et al. 2007	China	76 adults, 18+	Psychosocial & Pharmacological	Motivational Interviewing, Family therapy, Social support, & Medication: Antidepressants	TAU Pharm: Antidepressants	Unstated number of sessions for 24 months	0 months	Depression	Time in work
Nagarajaiah et al. 2012	India	60 adults, 18-65	Psychosocial & Pharmacological	Interpersonal therapy, Problem Solving Therapy, Family therapy, Social support & Medication	TAU Pharm: Antidepressants	10 sessions over 3 months	0 months	Anxiety	Functioning at work
Patel et al. 2011	India	213 adults 17+	Psychosocial & Pharmacological (Public facility)	Interpersonal Therapy, Psychoeducation, & Medication: Antidepressants	No treatment	90 days medication & 6 sessions over 12 months	9 months	Common mental disorders	Days unable to work
Patel et al. 2011	India	341 adults 17+	Psychosocial & Pharmacological (Private facility)	Interpersonal Therapy, Psychoeducation, & Medication: Antidepressants	No treatment	90 days medication & 6 sessions over 12 months	9 months	Common mental disorders	Days unable to work
Patel et al. 2011; Buttorff et al. 2012	India	1648 adults, 17+	Psychosocial & Pharmacological (Public facility)	Interpersonal Therapy, Psychoeducation, & Medication: Antidepressants	No treatment	90 days medication & 6 sessions over 12 months	9 months	Common mental disorders	Days unable to work

Patel et al. 2011	India	1148 adults, 17+	Psychosocial & Pharmacological (Private facility)	Interpersonal Therapy, Psychoeducation, & Medication: Antidepressants	No treatment	90 days medication & 6 sessions over 12 months	9 months	Common mental disorders	Days unable to work
Patel et al. 2017; Weobong et al. 2017; Bhat et al. 2022	India	495 adults, 18-65	Psychosocial	Behavioural Activation	EUC: Consultation with PHC physician	6-8 sessions over 3-4 months	10 months	Depression	Income, Consumption & input expenditure, Days unable to work, Employment, Functioning at work, Job search, Time in work
Sikander et al. 2019	Pakistan	570 women, 18+	Psychosocial	Behavioural Activation	EUC: Patient information leaflet	14 sessions over 9 months	-3 and -6 months	Depression	Days unable to work

Author, year	Country	Sample & age	Intervention category	Therapeutic type	Control group category	Description of intervention	Follow-up time points*	Target mental disorder	Economic outcomes**
Substance Use Disorders (SUD)									
Blattman et al. 2017	Liberia	999 men, 18-35	Psychosocial	Cognitive Behavioural Therapy	No treatment	3 sessions every week for 2 months (24 sessions)	1 month after start, 10.5 after end	Antisocial Behaviour	Assets, Income, Consumption & input expenditure, Time in work
Nadkarni et al. 2017a; 2017b	India	377 men, 18+	Psychosocial	Counselling for alcohol problems (CAP)	EUC: Consultation with PHC physician	Up to 4 sessions weekly or fortnightly	2 and 11 months	Substance dependence (Alcohol)	Days unable to work
Nadkarni et al. 2019	India	135 men, 18+	Psychosocial	Counselling for alcohol problems (CAP)	EUC: Consultation with PHC physician	Up to 4 sessions weekly or fortnightly	-1 and 8 months	Substance dependence (Alcohol)	Days unable to work, Employment
Pan et al. 2015	China	195 adults, 18-65	Psychosocial & Pharmacological	Cognitive Behavioural Therapy & Methadone Maintenance Therapy (MMT)	TAU Pharm: MMT only	1 session weekly for 26 weeks (5 months)	3 & 6.5 months after start, 0.5 after end	Substance dependence (Heroin)	Employment
Xu et al. 2021	China	40 adults, 20+	Psychosocial	Community-based addiction rehabilitation electronic system (CAREs) using a smartphone app	TAU: Community based care	1 session weekly for 6 months	0 months	Substance dependence (Methamphetamine & Heroin)	Functioning at work
Zhao et al. 2011	China	100 adults, 18+	Psychosocial	Cognitive Behavioural Therapy	EUC: Inpatient drug rehabilitation centre	20 sessions over 2 months	1 month	Substance dependence (Heroin)	Employment
Post Traumatic Stress Disorder (PTSD)									
Betancourt et al. 2014	Sierra Leone	436 youth, 15-24	Psychosocial	Cognitive Behavioural Therapy & Interpersonal therapy	No treatment	1 session weekly for 10 weeks (2.5 months)	0, 6 and 8 months	PTSD	Education
Cilliers et al. 2016	Sierra Leone	2383 adults, 18+	Psychosocial	Community reconciliation & trauma healing	No treatment	2 day-long workshops	9 and 31 months	PTSD	Wealth, Employment, Financial, Wellbeing, Consumption
Hall et al. 2014	DRC	405 women, 18+	Psychosocial	Cognitive Behavioural Therapy based	EUC: Invitation to access existing psychosocial services	1 sessions weekly for 11 weeks (3 months)	2 and 7 months	PTSD	Assets, Social networks, Subjective poverty measures

Meffert et al. 2021	Kenya	206 women, 18+	Psychosocial	Interpersonal therapy	TAU plus Waitlist	12 sessions weekly for 12 weeks (3 months)	0 months	PTSD	Social networks
Wang 2016	Kosovo	34 adults, 18+	Psychosocial	Cognitive Behavioural Therapy & Prolonged exposure therapy	No treatment	1 session weekly for 10 weeks (2.5 months)	0 and 3 months	PTSD	Income, Consumption & input expenditure, Employment

Author, year	Country	Sample & age	Intervention category	Therapeutic type	Control group category	Description of intervention	Follow-up time points*	Target mental disorder	Economic outcomes**
Severe Mental Disorders (SMD)									
Chatterjee et al. 2014	India	282 adults, 16-60	Psychosocial & Pharmacological	Psychoeducation & Medication: Antipsychotics	TAU Pharm: Antipsychotics	22 sessions over 12 months	6 months after start, 0 after end	Schizophrenia	Functioning at work
ChuanQuian et al. 2005	China	112 adults, 18+	Psychosocial & Pharmacological	Psychoeducation & Medication: Antipsychotics	TAU Pharm: Antipsychotics	1 session monthly for 6 months	0 months	Schizophrenia	Subjective poverty measures, Func-tioning at work
Gureje et al. 2020	Ghana and Nigeria	286 adults, 18+	Psychosocial & Pharmacological	Collaborative shared care: Traditional & Faith Healers & PHC; & Medication: Antipsychotics	TAU: No collaboration	At least 1 visit weekly over 3-6 months	0 months	Schizophrenia	Functioning at work
Luo et al. 2018	China	58 adults, 16+	Psychosocial & Pharmacological	Assertive Community Treatment (ACT)	TAU: Standard community treatment	2 sessions weekly & 1 family session monthly for 12 months	0 months	Schizophrenia	Employment
Ran et al. 2003; 2015	China	326 adults, 18+	Psychosocial & Pharmacological	Psychoeducation & Medication: Antipsychotics	No treatment	1 session monthly & 3 family workshops for 9 months	0 and 159 months	Schizophrenia	Unable to work
Ran et al. 2003; 2015	China	326 adults, 18+	Pharmacological	Medication: Antipsychotics	No treatment	Medication for 9 months	0 and 159 months	Schizophrenia	Unable to work
Valencia et al. 2007	Mexico	82 adults, 16-50	Psychosocial & Pharmacological	Problem Solving Therapy, Psychoeducation, Family Therapy, & Medication: Antipsychotics	TAU Pharm: Antipsychotics	1 session weekly for 12 months (48 sessions)	0 months	Schizophrenia	Subjective poverty measures, Func-tioning at work
Vizzotto et al. 2021	Brazil	48 adults, 18-55	Psychosocial	Occupational Goal Intervention (OGI)	TAU Pharm: Antipsychotics	30 sessions for 15 weeks	6 months	Schizophrenia	Functioning at work
Xiang et al. 1994	China	77 adults, 18+	Psychosocial & Pharmacological	Psychoeducation & Medication: Antipsychotics	TAU Pharm: Antipsychotics	1 session monthly for 4 months	0 months	Schizophrenia	Employment
Xiong et al. 1994	China	63 adults, 18+	Psychosocial & Pharmacological	Problem Solving Therapy, Psychoeducation, & Medication: Antipsychotics	TAU Pharm: Antipsychotics	1 session monthly for 12-24 months	6, 12 and 18 months after start	Schizophrenia	Employment
Zhang et al. 1998	China	1048 adults, 18+	Psychosocial & Pharmacological	Psychoeducation & Medication: Antipsychotics	TAU Pharm: Antipsychotics	14 lectures & 5 discussions over 24 months	0 months	Schizophrenia	Employment

* Assessment time point in months after intervention ends (not after intervention start/baseline)
** See Table A4 for full explanation of economic outcomes