Supplementary Appendix For Online Publication

A Search and included studies

A.1 Search strategy

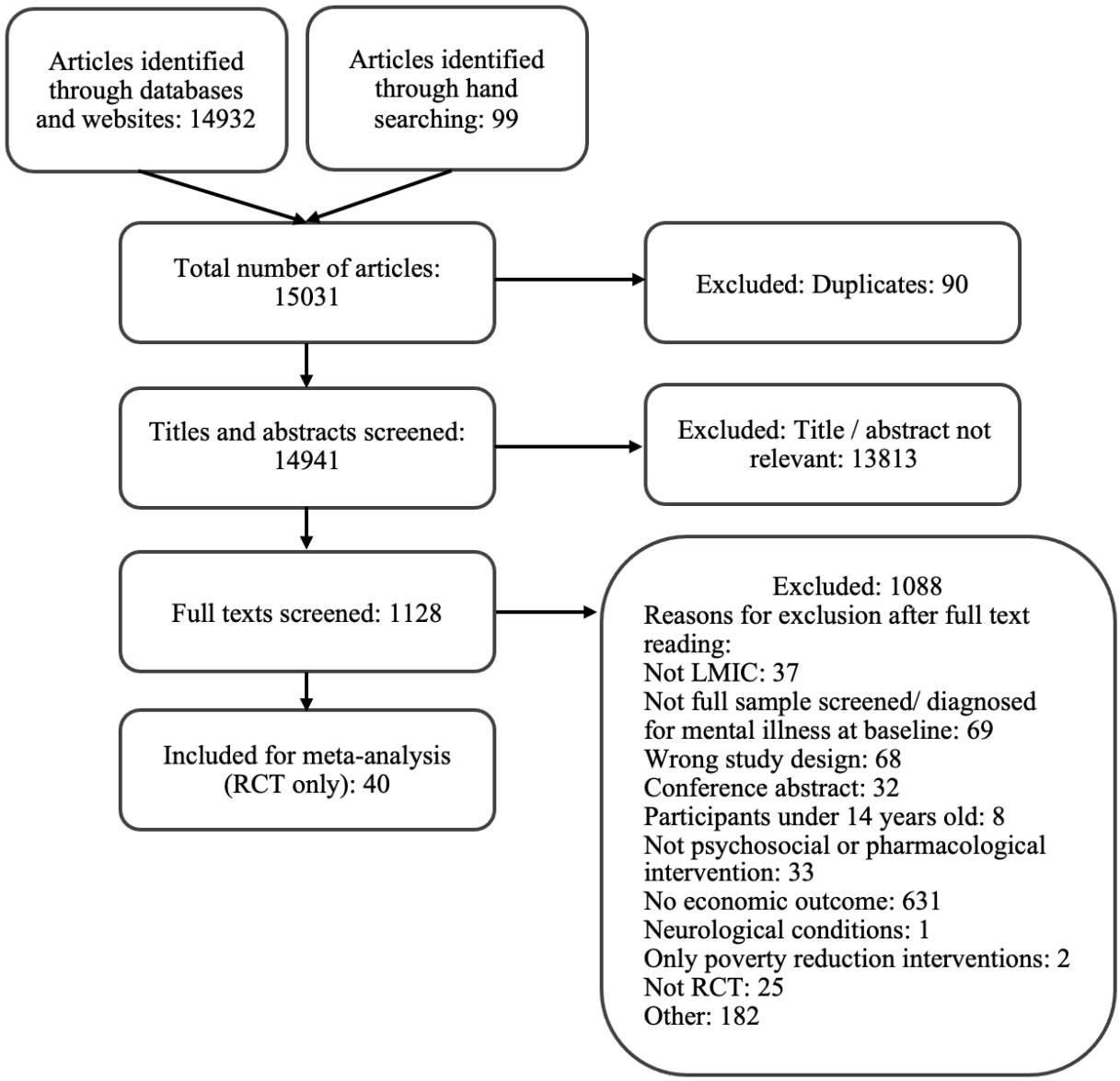
Searches of the databases and forward and backward reference tracking of citations of included papers yielded 15,031 potential studies. A sample set of search terms for PubMed can be found online under the following link: http://tiny.cc/0duyvz. Three reviewers independently screened abstracts against inclusion criteria using Covidence software. Disagreements were resolved by a third reviewer. After removing 90 duplicates, abstracts of 14,941 selected articles were screened by two reviewers, with disagreements again resolved by a third reviewer. 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. 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.

Information was then input into a piloted, pre-populated Excel spreadsheet by multiple members of the study team. A different 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 categories of mental health or functioning outcome (see Section 3.3 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.

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



A.2 Included studies

References

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

AYOUGHI, S., I. MISSMAHL, R. WEIERSTALL, AND T. ELBERT (2012): "Provision of Mental Health Services in Resource-poor Settings: A Randomised Trial Comparing Counselling with Routine Medical Treatment in North Afghanistan (Mazar-e-Sharif)," *BioMed Psychiatry*, 12, 14.

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

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

BETANCOURT, T. S., R. MCBAIN, E. A. NEWNHAM, A. M. AKINSULURE-SMITH, R. T. BRENNAN, J. R. WEISZ, AND N. B. HANSEN (2014): "A Behavioral Intervention for War-affected Youth in Sierra Leone: A Randomized Controlled Trial," *Journal of the American Academy of Child & Adolescent Psychiatry*, 53, 1288–1297.

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

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

BOLTON, P., J. BASS, R. NEUGEBAUER, H. VERDELI, K. F. CLOUGHERTY, P. WICKRAMARATNE, L. SPEELMAN, L. NDOGONI, AND M. WEISSMAN (2003): "Group Interpersonal Psychotherapy for Depression in Rural Uganda: A Randomized Controlled Trial," *Journal of the American Medical Association*, 289, 3117–3124.

BUTTORFF, C., R. S. HOCK, H. A. WEISS, S. NAIK, R. ARAYA, B. R. KIRKWOOD, D. CHISHOLM, AND V. PATEL (2012): "Economic Evaluation of a Task-shifting Intervention for Common Mental Disorders in India," *Bulletin of the World Health Organization*, 90, 813–821.

CHATTERJEE, S., S. NAIK, S. JOHN, H. DABHOLKAR, M. BALAJI, M. KOSCHORKE, M. VARGHESE, R. THARA, H. A. WEISS, P. WILLIAMS, P. MCCRONE, V. PATEL, AND G. THORNICROFT (2014): "Effectiveness of a Community-based Intervention for People with Schizophrenia and Their Caregivers in India (COPSI): A Randomised Controlled Trial," *The Lancet*, 383, 1385–1394.

CHUAN-QIAN, Q., Z. ZHAO-XIANG, AND L. ZHI-CHENG (2005): "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*, 12, 75–77.

CILLIERS, J., O. DUBE, AND B. SIDDIQI (2016): "Reconciling After Civil Conflict Increases Social Capital but Decreases Individual Well-being," *Science*, 352, 787–794.

DUARTE, P. S., M. C. MIYAZAKI, S. L. BLAY, AND R. SESSO (2009): "Cognitive-behavioral Group Therapy: An Effective Treatment for Major Depression in Hemodialysis Patients," *Kidney International*, 76, 414–421.

FUHR, D. C., B. WEOBONG, A. LAZARUS, F. VANOBBERGHEN, H. A. WEISS, D. R. SINGLA, H. TABANA, E. AFONSO, A. DE SA, E. D'SOUZA, ET AL. (2019): "Delivering the Thinking Healthy Pro-

gramme for Perinatal Depression Through Peers: An Individually Randomised Controlled Trial in India," *The Lancet Psychiatry*, 6, 115–127.

GUREJE, O., J. APPIAH-POKU, T. BELLO, L. KOLA, R. ARAYA, D. CHISHOLM, O. ESAN, B. HARRIS, V. MAKANJUOLA, C. OTHIENO, L. PRICE, AND S. SEEDAT (2020): "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*, 396, 612–622.

HALL, B. J., P. A. BOLTON, J. ANNAN, D. KAYSEN, K. ROBINETTE, T. CETINOGLU, K. WACHTER, AND J. K. BASS (2014): "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*, 104, 1680–1686.

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

HIRANI, S. S., R. KARMALIANI, J. MCFARLANE, N. ASAD, F. MADHANI, ET AL. (2010): "Testing a Community-derived Intervention to Promote Women's Health: Preliminary Results of a 3-arm Randomized Controlled Trial in Karachi, Pakistan," *Southern Online Journal of Nursing Research*, 10, 1.

HU, T.-W., Y. HE, M. ZHANG, AND N. CHEN (2007): "Economic costs of depression in China," *Social psychiatry and psychiatric epidemiology*, 42, 110–116.

LUO, X., S. F. LAW, X. WANG, J. SHI, W. ZENG, X. MA, W. CHOW, S. LIU, W. ZHAO, X. LIU, S. YAO, AND M. R. PHILLIPS (2019): "Effectiveness of an Assertive Community Treatment Program for People with Severe Schizophrenia in Mainland China – A 12-month Randomized Controlled Trial," *Psychological Medicine*, 49, 969–979.

MEFFERT, S. M., T. C. NEYLAN, C. E. MCCULLOCH, K. BLUM, C. R. COHEN, E. A. BUKUSI, H. VERDELI, J. C. MARKOWITZ, J. G. KAHN, D. BUKUSI, H. THIRUMURTHY, G. ROTA, R. ROTA, G. OKETCH, E. OPIYO, AND L. ONGERI (2021): "Interpersonal Psychotherapy Delivered by Non-specialists for Depression and Posttraumatic Stress Disorder Among Kenyan HIV-positive Women Affected by Gender-based Violence: A Randomized Controlled Trial," *PLOS Medicine*, 18, 1–22.

MIN, Z., L. XU, H. CHEN, X. DING, Z. YI, AND Z. MINGYUANG (2011): "A Pilot Assessment of Relapse Prevention for Heroin Addicts in a Chinese Rehabilitation Center," *American Journal of Drug and Alcohol Abuse*, 37, 141–147.

NADKARNI, A., H. A. WEISS, R. VELLEMAN, J. MCCAMBRIDGE, D. MCDAID, A.-L. PARK, P. MURTHY, B. WEOBONG, B. BHAT, AND V. PATEL (2019): "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*, 114, 1192–1203.

NADKARNI, A., H. A. WEISS, B. WEOBONG, D. MCDAID, D. R. SINGLA, A.-L. PARK, B. BHAT, B. KATTI, J. MCCAMBRIDGE, P. MURTHY, M. KING, G. T. WILSON, B. KIRKWOOD, C. G. FAIRBURN, R. VELLEMAN, AND V. PATEL (2017a): "Sustained Effectiveness and Cost-effectiveness of Counselling for Alcohol Problems: A Brief Psychological Treatment for Harmful Drinking in Men, Delivered by Lay Counsellors in Primary Care: 12-month Follow-up of a Randomised Controlled Trial," *PLOS Medicine*, 14, 1–20.

NADKARNI, A., B. WEOBONG, H. A. WEISS, J. MCCAMBRIDGE, B. BHAT, B. KATTI, P. MURTHY, M. KING, D. MCDAID, A.-L. PARK, G. T. WILSON, B. KIRKWOOD, C. G. FAIRBURN, R. VELLEMAN, AND V. PATEL (2017b): "Counselling for Alcohol Problems (CAP), A Lay Counsellor-delivered Brief Psychological Treatment for Harmful Drinking in Men, in Primary Care in India: A Randomised Controlled Trial," *The Lancet*, 389, 186–195.

NAGARAJAIAH, G. JOTHIMANI, R. PARTHASARATHI, K. REDDEMMA, AND A. GIRI (2013): "Efficacy of Nursing Interventions in Reducing Social and Occupational Disabilities among Patients with Neurosis," *The Nursing journal of India*, 103, 256–60.

PAN, S., H. JIANG, J. DU, H. CHEN, Z. LI, W. LING, AND M. ZHAO (2015): "Efficacy of Cognitive Behavioral Therapy on Opiate Use and Retention in Methadone Maintenance Treatment in China: A Randomised Trial," *PLOS ONE*, 10, e0127598.

PATEL, V., H. A. WEISS, N. CHOWDHARY, S. NAIK, S. PEDNEKAR, S. CHATTERJEE, B. BHAT, R. ARAYA, M. KING, G. SIMON, H. VERDELI, AND B. R. KIRKWOOD (2011): "Lay Health Worker-led Intervention for Depressive and Anxiety Disorders in India: Impact on Clinical and Disability Outcomes Over 12 Months," *British Journal of Psychiatry*, 199, 459–466.

PATEL, V., B. WEOBONG, H. A. WEISS, A. ANAND, B. BHAT, B. KATTI, S. DIMIDJIAN, R. ARAYA, S. D. HOLLON, M. KING, L. VIJAYAKUMAR, A.-L. PARK, D. MCDAID, T. WILSON, R. VELLEMAN, B. R. KIRKWOOD, AND C. G. FAIRBURN (2017): "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*, 389, 176–185.

RAN, M.-S., C. L.-W. CHAN, S. M. NG, L.-T. GUO, AND M.-Z. XIANG (2015): "The Effectiveness of Psychoeducational Family Intervention for Patients with Schizophrenia in a 14-year Follow-up Study in a Chinese Rural Area," *Psychological Medicine*, 45, 2197–2204.

RAN, M.-S., M.-Z. XIANG, C. L.-W. CHAN, J. LEFF, P. SIMPSON, M.-S. HUANG, Y.-H. SHAN, AND S.-G. LI (2003): "Effectiveness of Psychoeducational Intervention for Rural Chinese Families Experiencing Schizophrenia," *Social Psychiatry and Psychiatric Epidemiology*, 38, 69–75.

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

VALENCIA, M., M. L. RASCON, F. JUAREZ, AND E. MUROW (2007): "A Psychosocial Skills Training Approach in Mexican Out-patients with Schizophrenia," *Psychological Medicine*, 37, 1393–1402.

VIZZOTTO, A., D. CELESTINO, P. BUCHAIN, A. OLIVEIRA, G. OLIVEIRA, E. DI SARNO, I. NAPOLITANO, AND H. ELKIS (2021): "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*, 75, 7503180050.

WANG, S.-J., A. BYTYÇI, S. IZETI, M. KALLABA, F. RUSHITI, E. MONTGOMERY, AND J. MODVIG (2017): "A Novel Bio-psycho-social Approach for Rehabilitation of Traumatized Victims of Torture and War in the Post-conflict Context: A Pilot Randomized Controlled Trial in Kosovo," *Conflict and Health*, 10, 34.

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

XIANG, M., M. RAN, AND S. LI (1994): "A Controlled Evaluation of Psychoeducational Family Intervention in a Rural Chinese Community," *British Journal of Psychiatry*, 165, 544–548.

XIONG, H., W. YU-LING, AND F. HUI-PENG (2007): "Synthetical Family Treatment for Depression: A Randomized-controlled Single-blind Study Among 76 Cases," *Zhongguo Zuzhi Gongcheng Yanjiu Yu Linchuang Linchuang Kangfu*, 11, 7787.

XIONG, W., M. R. PHILLIPS, X. HU, R. WANG, Q. DAI, J. KLEINMAN, AND A. KLEINMAN (1994):

"Family-Based Intervention for Schizophrenic Patients in China: A Randomised Controlled Trial," *British Journal of Psychiatry*, 165, 239–247.

XU, X., S. CHEN, J. CHEN, Z. CHEN, L. FU, D. SONG, M. ZHAO, AND H. JIANG (2021): "Feasibility and Preliminary Efficacy of a Community-Based Addiction Rehabilitation Electronic System in Substance Use Disorder: Pilot Randomized Controlled Trial," *JMIR Mhealth Uhealth*, 9, e21087.

ZHANG, M., Y. HE, M. GITTELMAN, Z. WONG, AND H. YAN (1998): "Group Psychoeducation of Relatives of Schizophrenic Patients: Two-year Experiences," *Psychiatry and Clinical Neurosciences*, 52, S344–S347.

B Comparison to high income country effect sizes

Table A1: Comparison to developed country effect sizes: work outcomes

Outcome	Study	Target Condition	Intervention	Effect Size	95% CI
Aggregate work effects					
Work aggregate	Lund et al. (2024)	Common Mental Disorders	Psychosocial	g = 0.16	[0.03, 0.31]
Work aggregate	Lund et al. (2024)	Severe Mental Disorders	Psychosocial	g = 0.30	[-0.05, 0.67]
Labour supply effects[1]	Timbie et al. (2006)	Depression	Pharmacological	d = -0.39	[-1.01, 0.24]
			Psychosocial	d = 0.16	[0.02, 0.29]
Absence from Work					
Days unable to work	Lund et al. (2024)	Common Mental Disorders	Psychosocial	g = 0.08	[0.01, 0.17]
Sick leave[2]	Nieuwenhuijsen et al. (2020)	Depression	Psychosocial	SMD = 0.15	[0.03,0.28]
			Psych.+pharm.	SMD = 0.38	[-0.24, 0.99]
Sick leave[a]	Salomonsson et al. (2018)	Common Mental Disorders	Problem Solving Therapy	g = 0.12	[0.02, 0.22]
			Cognitive behavioural therapy	g = 0.17	[0.02, 0.32]
Sickness absence	Finnes et al. (2019)	Common Mental Disorders + Musculoskeletal Disorders	Psychosocial (CBT, PST, PDT, MMCBT, MI)	g = 0.17	[-0.03, 0.36]
Functioning at work					
Functioning at work	Lund et al. (2024)	Common Mental Disorders	Psychosocial	g = 0.01	[-0.11, 0.13]
Functioning	Kamenov et al. (2017)	Depression	Psychosocial	g = 0.43	[0.33, 0.54]
Coping with work[3]	Nieuwenhuijsen et al. (2020)	Depression	Psychological	SMD = 0.05	[-0.46, 0.57]
Employment Dummy					
In employment	Lund et al. (2024)	Common Mental Disorders	Psychosocial	g = -0.02	[-0.35, 0.25]
In employment	Lund et al. (2024)	Severe Mental Disorders	Psychosocial	g = 0.22	[-1.66, 1.85]
Employment rate	van Duin et al. (2019)	Severe Mental Disorders	Psychiatric Rehab + Cognitive Rehab	SD = 0.41	[0.10, 0.72]
Employment rate	Chan et al. (2015)	Severe Mental Disorders	Computer Assisted Cognitive Remediation	SD = 0.15	[0.04, 0.25]
Time in work					
Time in work	Lund et al. (2024)	Common Mental Disorders	Psychosocial	g = 0.16	[-0.61, 0.87]
Time in work	Lund et al. (2024)	Severe Mental Disorders	Psychosocial	g = 0.21	[-1.09,1.32]
Hours worked	van Duin et al. (2019)	Severe Mental Disorders	Psychiatric Rehab + Cognitive Rehab	SD = 0.31	[0.04, 0.58]
Employment frequency	Chan et al. (2015)	Severe Mental Disorders	Computer Assisted Cognitive Remediation	SD = 0.15	[0.02, 0.28]
Employment frequency	Chan et al. (2015)	Severe Mental Disorders	Computer Assisted Cognitive Remediation	19.5 more days / year	[2.50, 36.6]

Notes: Effect sizes are denoted: g = Hedges' g; d = Cohen's d; SD = Standard Deviation; SMD = Standardized Mean Difference and comparable, differing only in small sample corrections. [1] The aggregate of labor supply effects includes hours worked per week, odds of being unable to work, days of employment, percent employed, and aligns closely with our work aggregate. [2] Most studies reverse-coded effects such that a positive effect indicates an improvement in work outcomes as in our study. However, Nieuwenhuijsen et al. (2020) did not reverse code Sick Days, so we have given the absolute value of their effect size for comparison. [3] Each of the functioning at work scales are validated measures of functioning except "Coping with work", which measured people's capacity to copy with their work given their depression symptoms.

C Outcome details and study characteristics

Table A2: Additional characteristics of interventions in included RCTs

	All		Target CMD	Target SMD
	(1) Number of interventions	(2) Share of interventions	(3) N	(4) N
Age target groups:	39	1	16	9
Adults (17+)	26	0.67	11	5
Youth[1]	3	0.08	0	0
Other ranges[2]	10	0.26	5	4
Specific target groups:	39	1	16	9
Males	3	0.08	0	0
Females	7	0.18	5	0
Both genders	29	0.74	11	9
Publication period: *(First publication)*	39	1	16	
Publication: 1990-2000	3	0.08	0	3
Publication: 2001-2010	8	0.21	4	3
Publication: 2011-2015	12	0.31	6	1
Publication: 2016-2020	11	0.28	4	1
Publication: 2021-present	6	0.15	2	1
Country income *(mutually exclusive)*	39	1	16	9
Upper middle income country	14	0.36	2	8
Lower middle income country	19	0.49	12	1
Low income country	6	0.15	2	0
Regions *(mutually exclusive)*	39	1	16	9
Sub-Saharan Africa	9	0.23	3	0
Europe and Central Asia	2	0.05	1	0
Latin America and the Caribbean	3	0.08	1	2
South Asia	14	0.36	10	1
East Asia and Pacific	11	0.28	1	6
Follow-up timing:	39	1	16	0
Follow-up:1-6 months after start	25	0.64	9	3
Follow-up: >6-12 months after start	18	0.46	8	1
Follow-up:>1-2 years after start	6	0.15	1	1
Follow-up:>2 years after start	6	0.15	2	1
Follow-up combinations (in months after start):	39	1	16	9
1 round, < 7 months	16	0.41	5	7
1 round, 7 to 12 months	5	0.31	4	0
1 round, 13 to 24 months	1	0.20	1	0
2 rounds, < 7 months	2	2.00	1	0
2 rounds, <7 & 7-12 months	7	3.50	2	0
2 rounds, <7 & >12 months	3	0.43	0	1
2 rounds, 7-12 & >24 months	1	0.33	0	0
3 rounds, <7 & 7-12 & >12 months	3	3.00	2	1
3 rounds, 7-12 & 13-24 & >24 months	1	0.33	1	0
Type of psychosocial intervention: *(not mutually exclusive)*				
Cognitive-behavioral therapy (CBT)	9	0.23	3	0
Problem solving therapy	9	0.23	4	3
Interpersonal therapy	6	0.15	4	0
Psychoeducation (only)	7	0.18	2	5
Behavioural activation	4	0.1	4	0
Other	9	0.23	0	0
Type of pharmacological intervention:	18	0.46	2	9
Against psychotic disorders	10	0.26	0	9
Against mood disorders (depression)	7	0.18	2	0
Against substance abuse	1	0.02	0	0

Notes: There are 39 interventions. Psychosocial interventions often include more than one type of therapy. [1] 1 intervention with ages 15-24, 1 with 18-35, 1 with 25-35. [2] 1 intervention with 14+, 1 with 16+, 1 with 20+, 1 with 16-45, 3 with 18-65, 1 with 16-60, 1 with 16-50, 1 with 18-55.

Table A3: Economic outcomes

Outcome category	# of estimates	Detailed measures (throughout, "self" indicates a self-reported measure)
Assets	18	Durable goods index; Index of durable assets; Index of household assets: enumerator assessment. Uses data from 9 month endline in wave one. ; Index of household assets: enumerator assessment. Uses pooled endline data from both waves and rounds. ; Index of household assets: enumerator assessment. Uses data from 31 month endline in wave one.; Level of credit measured in rupees; Level of debt that the household owes to others, measured in rupees; Level of savings measured in rupees; Net worth = savings+credit-debt; Savings stock; Total value of assets owned (USD PPP)
Days unable to work	23	Days of sick leave in last 2 years: own assessment ; Days respondent was 'healthy' in the past 30 days at 7 year: own assessment ; WHODAS 2.0 Participation in Society: self-assessment of days unable to work or reduction in work; WHODAS 2.0: Number of days unable to work (12 months); WHODAS 2.0: Number of days unable to work (3 months); WHODAS 2.0: days unable to work + (0.5 x) days with reduced work; WHODAS 2.0: days unable to work and days with reduced work: self assessment ; WHODAS 2.0: days unable to work: self assessment ; WHODAS 2.0: self-reported days unable to work at 3 month endline; WHODAS 2.0: self-reported days unable to work at 6 month endline ; WHODAS 2.0: self-reported days unable to work at 6 month endline ; a self report of work days missed last month due to poor health
Education	17	Academic Performance Scale: Clinician (teacher) assessed; Binary enrollment: Clinician (teacher) assessed; 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 Observation for Measurement of the Environment (HOME) Inventory: Own assessment of Physical Environment sub-scale; Likert Scale: Clinician (teacher) assessed; Log of the family'seducational 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): cliincian assessment; homework time; school attendance; school enrollment
Employment (dummy)	11	Likelihood of re-employment over the 12 month follow up period. ; Mother is employed at 7 years: own assessment; Self reported 'employed' at 3 month endline ; Self reported 'employed' at 6 month endline ; Self reported 'unemployed' at 3 month endline ; Self reported 'unemployed' at 6 month endline ; Self reported employment status ; Self: Engaged in work in the last week; WHODAS 2.0: individual was able to work every day: self assessment
Functioning at work	20	ASI index: clinician employment status; ASI index: clinician employment status for respondents receiving relapse prevention treatment ; ASI: Indication of ideal employment status; Clinician: Bracelets made in ten minutes; IDEAS scale: Clinician employment, housework and educational performance; 12 months; IDEAS scale: Clinician employment, housework and educational performance; 6 months; Independent Living Skills Survey: job maintenance. Clinician assessment. ; KDQOL-SF: work status; Life Chart Schedule: Performance at Work; MRSS: activity/inactivity: clinician rating of functioning in employment and leisure ; Overall occupational disabilities (GSDS-II - Groningen Social Disability Scale); Own assessment of capacity to do farming; Own assessment of capacity to do manual labour; Own assessment of capacity to grow food; PSFS Occupational Functioning: Own satisfaction with functioning in occupation ; WHO QoL work item: own satisfaction with capacity for work; WHODAS 2.0: Life activities domain 5
Income, consumption and input expenditure	27	Earnings are from primary and secondary jobs measured in Rupees; Earnings in the past 4 weeks; Food and non-food consumption in past 2 weeks; Investment in the past 2 weeks; Monthly household revenue (USD PPP); Monthly per-capita non-durable consumption (USD PPP); Mother's monthly earnings: own assessment ; Per capita consumption; Self reported monthly income at 3 month endline ; Self reported monthly income at 6 month endline ; Self: Durable goods expenditure; Self: Food expenditure; Self: Medical expenditure; Self: Other expenditures; Self: Total monthly expenditure; Self: earnings in the past month; Value of business assets
Job search	6	Self: available to take a job opportunity; Self: job search hours per week; job search hours
Other	5	PSFS Occupational Functioning: Own satisfaction with functioning in money-management ; Self: Applied for ability-based contract; Self: Reservation wage
Social networks	7	Frequency of borrowing and lending: Self-reported. Uses pooled endline data from both waves and rounds. ; IntegratedQuestionnaire for the Measurement of SocialCapital: Own assessment of Financial Social Network Size ; IntegratedQuestionnaire for the Measurement of SocialCapital: Own assessment of Instrumental Support Network Size; Monetary value of borrowing and lending: Own assessment. Uses pooled endline data from both waves and rounds. ; Respondent belongs to an osusu (savings group): Self-reported. Uses pooled endline data from both waves and rounds.
Subjective poverty measures	11	Clinician-rated measure of poverty; Composed of self-reported economic status and projected economic status in five years time; Proejcted economic status in 5 years time using Cantril's ladder; Satisfaction that household needs are met: Self-reported. Uses pooled endline data from both waves and rounds. ; Satisfaction that household needs are met: Self-reported. Uses data from 31 month endline in wave one.; Satisfaction that household needs are met: Self-reported. Uses data from 9 month endline in wave one.; Satisfaction with household's economic situation relative to 1 year ago: Self-reported. Uses pooled endline data from both waves and rounds. ; Satisfaction with household's economic situation relative to 1 year ago: Self-reported. Uses data from 31 month endline in wave one.; Satisfaction with household's economic situation relative to 1 year ago: Self-reported. Uses data from 9 month endline in wave one.; Self-reported economic status today using Cantril's ladder. ; WHO QoL work item: own satisfaction with financial resources and condition
Time in work	17	Hours per week of work in the last 2 months; Months engaged in normal occupation: family assessment at 12 month endline; Months engaged in normal occupation: family assessment at 18 month endline; Months engaged in normal occupation: family assessment at 6 month endline; Percent of time at work: own assessment ; SDSS: Ability to Work (full time); Self: Work hours in the last week; Time per 24h in productive activities (converted to weekly value); child care work hours; domestic work hours; primary and secondary jobs and agricultural work hours
Unable to work	18	Percent of time on sick leave: own assessment ; 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 (part-time); SDSS: own assessment of ability to work (unable to work); 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 - None (12 months); WHODAS 2.0: days unable to work - None (3 months); WHODAS 2.0: days unable to work - at least one (12 months); WHODAS 2.0: days unable to work - at least one (3 months)

Table A4: Mental health outcomes

Outcome category	# of estimates	Detailed measures (throughout, "self" indicates a self-reported measure)
Antisocial Behaviour	2	Antisocial behaviour index: self
Anxiety	6	Adapted Zung Anxiety Index: self; Generalised Anxiety Disorder Assessment (GAD-7): self; Hopkins Symptom Checklist - anxiety: self; Self Rating Anxiety Scale (SAS): self
Cognition	11	Cognition index; Cognition index: clinician; Digit Span: backwards; Digit span: forwards; Executive function index: self; Kidney Disease and Quality of Life-Short Form - cognitive function: self; Raven's Progressive Matrices; World Health Organization Quality of Life (WHOQOL) subscale - Cognitive function: Clinician
Depression	33	Adapted Zung Depression Index: self; Beck Depression Inventory (BDI): self; Beck Depression Inventory II- Depression score: self; Beck Depression Inventory Version II (BDI-II): self; Culturally Grounded Screening for Depression: self; Depression Remission on Patient Health Questionnaire (PHQ-9): self; Depression score on Patient Health Questionnaire (PHQ-9): self; Depression status: Clinician; Hamilton Depression Rating Scale (HDRS): clinician; Hopkins Symptom Checklist - depression: self; Hopkins Symptom Checklist 25 (HSCL-25) - depression: self; Index of locally-relevant depression features: self; Locally adapted Hopkins Symptom Checklist (HSCL) - depression subscale: Clinician; Mini International Neuropsychiatric Interview - depression: clinician; Patient Health Questionnaire (PHQ-9): self; Patient Health Questionnaire 9(PHQ-9): self
Diagnosed with mental disorder	14	Depression Remission on Patient Health Questionnaire (PHQ-9): self; Diagnostic and Statistical Manual of Mental Disorders IV (DSM-IV) - Adapted Depression Index: Clinician; Diagnostic and Statistical Manual of Mental Disorders IV (DSM-IV) - Adapted Depression Index: clinician; Hopkins Symptom Checklist: self; Mini International Neuropsychiatric Interview - depression: clinician; Mini International Neuropsychiatric Interview: self; No depression on Patient Health Questionnaire (PHQ-9): self; No moderate/severe depression on Patient Health Questionnaire (PHQ-9): self
Functioning	56	Activites of daily living: self; Addiction Severity Index (ASI): clinician; Brief Disability Questionnaire (BDQ): self; Direct Assessment of Functional Status; Direct assessment of functional status: dealing with finances. Clinician assessment. ; Global Assessment of Functioning (GAF)scale: clinician; Groningen Social Disability Schedule (GSDS-II): clinician; IDEAS scale - self-care : Clinician; Independent Living Skills Survey; Independent Living Skills Survey: money management. Clinician assessment. ; Indian Disability Evaluation and Assessment Scale (IDEAS) - overall: Clinician; Locally adapted gender specific functional impairment: Clinician; Morningside Rehabilitation Status Scale (MRSS) subscale - Current symptoms and deviant behavior: Clinician; Personal and Social Performance Scale; Short Inventory of Problems (SIP) mean score: self; Social Disability Screening Schedule (SDSS) - social dysfunction subscale: self ; WHO Disabillity Assessment Schedule- Functional impairment: self; WHOQOL subscale - Independence: Clinician; World Health Organization Disability Assessment Schedule 2.0 - 12 item for ICD-10 diagnosis at baseline subgroup: self; World Health Organization Disability Assessment Schedule 2.0 - 12 item for depression diagnosis at baseline subgroup: self; World Health Organization Disability Assessment Schedule 2.0 - 12 item for positive screen at baseline subgroup: self; World Health Organization Disability Assessment Schedule 2.0 - 12 item for subthreshold at baseline subgroup: self; World Health Organization Disability Assessment Schedule 2.0 - 12 item: self; World Health Organization Disability Assessment Schedule 2.0 - 36 item: self; local functioning tool; study-specific Psychosocial Functioning Scale (PSFS): self
Functioning in social interactions	9	IDEAS scale - Interpersonal activities: clinician; IDEAS scale - communication and understanding: Clinician; Oxford Measure of Psychosocial Adjustment adapted subscale- prosocial behaviour: self; Revised Social Disability Screening Schedule (SDSS-R) ; Social Disability Screening Schedule (SDSS) ; Social Disability Screening Schedule (SDSS) - social dysfunction subscale: self
General mental health	18	30 minus days in last month with poor mental health; Addiction Severity Index (ASI) - psychiatric status: clinician; Addiction Severity Index (ASI) - legal problems: clinician; Behavioural Risk Factor Surveillance Survey: Self-rating of mental health; Clinical Global Impression Subscale for Severity of Illness (CGIS): clinician; Kessler Psychological Distress Scale; Mean effect index of psychological health: self; Mental health index; Mood score; Quality Adjusted Life Years; Subjective well-being index; Subjective wellbeing: self; World Health Organization Quality of Life (WHOQOL) subscale - Mental health: Clinician
Mental health disorders	4	Proportion currently treated with antipsychotics: self; Proportion never treated with antipsychotic medication: self
Non-specific CMD	21	Clinical Interview Schedule - Revised (CIS-R) for positive screen at baseline subgroup: clinician; Clinical Interview Schedule - Revised (CIS-R) morbidity score for ICD-10 diagnosis at baseline subgroup: clinician; Clinical Interview Schedule - Revised (CIS-R) morbidity score for depression diagnosis at baseline subgroup: clinician; Clinical Interview Schedule - Revised (CIS-R) score for subthreshold at baseline subgroup: clinician; Hopkins Symptom Checklist 25 (HSCL-25) - anxiety: self; Hopkins Symptom Checklist: self; Oxford Measure of Psychosocial Adjustment adapted subscale- psychological distress: self; Proportion with common mental disorder diagnosis (ICD-10) for positive screen at baseline subgroup on CIS-R: clinician; Proportion with common mental disorder diagnosis (ICD-10) for subthreshold at baseline subgroup on CIS-R: clinician; Proportion with common mental disorder diagnosis (ICD-10) on CIS-R for ICD-10 diagnosis at baseline subgroup: clinician; Proportion with common mental disorder diagnosis (ICD-10) on CIS-R for depression diagnosis at baseline subgroup: clinician
Overall assessment of mental disorder	10	Addiction Severity Index (ASI) - mental health: clinician; Clinical Global Impression Subscale for Severity of Illness (CGIS): clinician; Clinical Interview Schedule - Revised (CIS-R) - total score: clinician; General mental health; Kidney Disease and Quality of Life-Short Form - Overall mental health: self; PANSS Scale: General psychopathological health ; state of illness: clinician

Table A5: Mental health outcomes

Outcome category	# of estimates	Detailed measures (throughout, "self" indicates a self-reported measure)
PTSD	13	Adapted PTSD Symptom Scale: self; Harvard Trauma Questionnaire (HTQ): self; Index of locally-relevant posttraumatic stress features: self; Mini International Neuropsychiatric Interview - PTSD: clinician; Post-Traumatic Stress Disorder Reaction Index (PTSD-RI): self; Posttraumatic Stress Disorder Checklist-Civilian
Physical health	10	Addiction Severity Index (ASI) - medical: clinician; Addiction Severity Index (ASI) - physical health: clinician; Kidney Disease and Quality of Life-Short Form - overall health: self; Kidney Disease and Quality of Life-Short Form - sexual function: self; Kidney Disease and Quality of Life-Short Form - sleep: self; World Health Organization Quality of Life (WHOQOL) subscale - general physical health: Clinician
Recovery (dummy)	4	Depression Recovery on Patient Health Questionnaire (PHQ-9): self; Full Recovery: Social Disability Screening Schedule (SDSS) ; Full recovery from recorded mental health disorder: clinician
Rehospitalisation	6	Days of re-admission: clinician; Days of rehospitalisation: clinician; Rehospitalization rate: clinician
Relapse (dummy)	14	Depression Remission on Patient Health Questionnaire (PHQ-9): self; Proportion relapse: clinician; Relapse rate: clinician; Schizophrenia relapse rate: clinician; Schizophrenia relapse: clinician
SMD symptoms	8	PANSS Scale: General psychopathological health ; PANSS Scale: clinician; Positive and Negative Syndrome Scale (PANSS): clinician; Serious Mental Disability: Social Disability Screening Schedule (SDSS)
Self-esteem/ self-efficacy	7	General Self-Efficacy Scale: self; Modified Rosenberg Self-Esteem Scale (SES): self; Self-Efficacy Scale (SE): self; Self-esteem index: self; World Health Organization Quality of Life (WHOQOL) subscale - self-esteem: Clinician
Self-regulation	7	Behavioural Assessment of the DysexecutiveSyndrome scale; Executive function; Patience Anderson index; Self-control scale; Short Grit Scale
Social support	27	Addiction Severity Index (ASI) - family support: clinician; Addiction Severity Index (ASI) - family/social: clinician; Contact with nonkin social network; Emotional support seeking: self; Inventory of Socially Supportive Behaviors (ISSB); KDQOL-SF: social interactions; KDQOL-SF: social support; Kidney Disease and Quality of Life-Short Form - quality of social interaction: self; Kidney Disease and Quality of Life-Short Form - social support: self; Perceived social support: self; Social network quality index: self; World Health Organization Quality of Life (WHOQOL) subscale - social support: Clinician
Stress	2	Perceived Stress Scale (PSS): self
Substance use	35	Addiction Severity Index (ASI) - alcohol use: clinician; Addiction Severity Index (ASI) - alcohol: clinician; Addiction Severity Index (ASI) - drug: clinician; Any ethanol consumed: self; Average proportion of negative urine test results: clinician; Daily drinking: self; Drug positive; Ethanol consumed: self; Heroin abstinence: clinician; Longest period of abstinence; Non-drinker; Percentage of days abstinent; Percentage of days of heavy drinking; Proportion of days abstinent: self; Proportion of days heavy drinking: self; Remission on Alcohol Use Disorders Identification Test (AUDIT): clinician; Remission on Alcohol Use Disorders Identification Test (AUDIT): self; Short Inventory of Problems (SIP) mean score: self; Substance abuse index: self
Suicide attempts or at risk of suicide	18	Any suicide attempt or suicidal ideation: self; Any suicide attempt: self; Mini International Neuropsychiatric Interview - suicide risk: clinician; Proportion suicidal behavior for common mental disorder diagnosis (ICD-10) group on CIS-R: clinician; Proportion suicidal behavior for depression diagnosis (ICD-10) group on CIS-R: clinician; Proportion suicidal behavior for positive screen group on CIS-R: clinician; Proportion suicidal behavior for subthreshold group on CIS-R: clinician; Proportion that take their own lives: clinician; Suicidal behaviour; Suicidal thoughts or attempts

Table A6: 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 condition?	Past 30 days
World Health Organisation Quality of Life (WHOQoL)	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 efforts to find a job?	Not provided
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.	Simultaneous
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 A7: 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 week how many days in total did you stay in bed all or most of the day because of your illness?	1 week
World Health Organization Disability Assessment Schedule 2.0 (WHODAS	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 thoughts of hurting yourself in some way?	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 behaviors.	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 suggested you cut down?	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 A8: 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. Felt difficulties were piling up so high that you could not overcome them?	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? How much did you feel restless? Has your heart been pounding fast? How much were you bothered by pain in your stomach?	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 interest in sex	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 thoughts of hurting yourself in some way?	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 bothered by pain in your stomach?	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 turning white or falling out.	2 weeks

Table A9: 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 on guard? 17. Feeling jumpy or easily startled?	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.	Simultaneous
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

D Approach to inference on effect sizes

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):

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

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

$$d = \frac{\bar{Y}_t - \bar{Y}_c}{S_p} = \frac{D}{S_p}$$

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

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

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

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

The standard error of Hedges' g is given by

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

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):

$$S_p = SE * \sqrt{\frac{n_t * n_c}{n_t + n_c}} \tag{4}$$

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:

$$g_j = \frac{1}{m} \sum_{i=1}^{m} g_{ij} \tag{5}$$

and calculate the standard error of g_j by

$$SE_{g,j} = \sqrt{\left(\frac{1}{m}\right)^2 \left(\sum_{i=1}^{m} SE_{g,i}^2 + \sum_{i \neq k} \rho_{i,k} SE_{g,ij} SE_{g,kj}\right)}, \tag{6}$$

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

$$g_i = \mu + \zeta_i + \varepsilon_i = \theta_i + \varepsilon_i, \tag{7}$$

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 7. 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 D squared.

E Additional results

E.1 Economic outcomes

Table A10: Effects of mental health interventions on all economic outcomes: Frequentist approach

	(1) Aggregate Hedges' g	(2) 95% lower CI	(3) 95% upper CI	(4) # of obs.	(5) # of intrv.	(6) I^2
Panel A: Psychosocial interventions targeting common mental disorders						
All economic outcomes aggregate	0.16***	0.06	0.26	66	16	0.62
Work aggregate	0.16***	0.05	0.27	36	14	0.68
In employment (dummy)	-0.00	-0.12	0.11	4	4	0.00
Time in work	0.17	-0.09	0.43	3	3	0.64
Days unable to work^	0.07**	0.01	0.13	17	9	0.13
Functioning at work	0.21	-0.10	0.52	8	5	0.86
Job search	0.09	-0.09	0.27	4	2	0.00
Non-work aggregate	0.09**	0.01	0.17	30	6	0.00
Education	0.21**	0.04	0.38	6	1	1.00
Assets	0.04	-0.08	0.16	1	1	1.00
Income, consumption and input expenditure	0.04	-0.05	0.13	15	4	0.00
Subjective poverty measures	0.16*	-0.01	0.33	4	2	0.00
Other	0.05	-0.11	0.21	4	2	0.00
Panel B: Psychosocial interventions targeting severe mental disorders						
All economic outcomes aggregate	0.27**	0.06	0.49	22	9	0.70
Work aggregate	0.30**	0.06	0.54	20	9	0.76
In employment (dummy)	0.81**	0.16	1.45	1	1	1.00
Time in work	0.24**	0.05	0.44	4	2	0.07
Unable to work (dummy) ^	0.18***	0.06	0.30	8	3	0.00
Days unable to work^	0.11**	0.00	0.22	1	1	1.00
Functioning at work	0.23	-0.41	0.87	6	4	0.89
Non-work aggregate	0.49	-0.30	1.28	2	2	0.76
Subjective poverty measures	0.15	-0.12	0.43	1	1	1.00
Other	0.97**	0.23	1.72	1	1	1.00

Notes: Table A10 reports estimates of the effect of mental health interventions on all studied economic outcomes under the frequentist specification outlined in Section 4.2. In Column (1), Hedges' g is the small-sample-bias-corrected standardized mean difference in the economic outcome between treatment and control. The aggregation of individual effect sizes is described in Section 4.1. Panel A presents estimates of the standard deviation effects of interventions targeting common mental disorders, Panel B interventions targeting severe mental disorders. *, ** 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. ^ indicates variables that have been reverse-coded such that higher values indicate improvements in 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.

Table A11: Effects of mental health interventions on economic outcomes: Bayesian approach

	τ estimate	τ posterior quantiles				Heterogeneity		
	(1) mean	(2) 2.5^{th}	(3) 25^{th}	(4) 75^{th}	(5) 97.5^{th}	(6) σ	(7) $\omega(\tau)$	(8) I^2
Panel A: Psychosocial interventions targeting *common* mental disorders								
All economic outcomes aggregate	0.04	-0.02	0.02	0.06	0.13	0.06	0.74	0.33
Work aggregate	0.16**	0.03	0.12	0.21	0.31	0.21	0.32	0.8
In employment (dummy)	-0.02	-0.35	-0.08	0.05	0.25	0.18	0.4	0.67
Time in work	0.16	-0.61	0.03	0.3	0.87	0.58	0.06	0.95
Days unable to work^	0.08**	0.01	0.05	0.1	0.17	0.06	0.72	0.31
Functioning at work	0.19	-0.3	0.07	0.32	0.67	0.48	0.08	0.93
Job search	0.05	-1.12	-0.13	0.25	1.16	1.13	0.01	0.98
Non-work aggregate	0.08	-0.05	0.05	0.12	0.21	0.09	0.59	0.44
Education	0.07	-1.68	-0.42	0.55	1.73	4.95	0	1
Assets	0.01	-1.69	-0.46	0.49	1.7	4.92	0	1
Income, consumption, input expenditure	0.03	-0.18	-0.01	0.08	0.22	0.12	0.42	0.6
Subjective poverty measures^	0.09	-1.13	-0.11	0.3	1.2	1.18	0.02	0.97
Other	0.03	-1.09	-0.14	0.2	1.11	1.05	0.01	0.99
Panel B: Psychosocial interventions targeting *severe* mental disorders								
All economic outcomes aggregate	0.28*	-0.03	0.18	0.37	0.61	0.39	0.25	0.85
Work aggregate	0.30*	-0.05	0.2	0.41	0.67	0.45	0.21	0.88
In employment (dummy)	0.22	-1.66	-0.35	0.82	1.85	5.37	0	NaN
Time in work	0.21	-1.09	0.01	0.46	1.32	1.28	0.02	0.98
Unable to work (dummy)^	0.21	-0.41	0.11	0.31	0.8	0.4	0.13	0.9
Days unable to work^	0.04	-1.69	-0.44	0.51	1.71	4.92	0	NaN
Functioning at work	0.19	-0.86	-0.09	0.47	1.17	1.12	0.07	0.97
Non-work aggregate	0.27	-1.27	-0.07	0.66	1.58	1.95	0.02	0.98
Subjective poverty measures^	0.05	-1.68	-0.44	0.53	1.72	4.94	0	1
Other	0.24	-1.66	-0.34	0.87	1.89	5.55	0	NaN

Notes: Table A11 reports estimates of the effect of mental health interventions on all studied economic outcomes under the Bayesian specification outlined in Section 4.2. In column (1), $\hat{\tau}$ is the estimate of the latent treatment effect in standard deviations. *, ** and *** represent significant at the 10, 5 and 1% levels respectively. Columns (2) through (5) present posterior quantile estimates, summarising the distribution of $\hat{\tau}$. Columns (6) through (8) present summary indicators of heterogeneity. The Work aggregate and Non-work aggregate meta-analyses include effect sizes from each of the other outcome groupings below. Panel A presents estimates of the standard deviation effects of psychosocial interventions targeting common mental disorders and Panel B the effects of psychosocial interventions targeting severe mental disorders. ^ indicates variables that have been reverse-coded such that higher values indicate improvements in outcomes. The average measurement in our sample happens 15.2 months after intervention start. The procedure for aggregating multiple effect sizes from a given intervention is described in subsection 4.1. All individual effect sizes are winsorized at the 99^{th} percentile. 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. The number of observations and interventions for each of the meta-analyses is identical to that in Table A10.

Table A12: Effects of mental health interventions on economic outcomes: Frequentist approach, additional intervention type - target condition groupings

	(1) Aggregate Hedges' g	(2) 95% lower CI	(3) 95% upper CI	(4) # of obs.	(5) # of intrv.	(6) I^2
Panel A: Psychosocial interventions targeting post-traumatic stress disorder						
All economic outcomes aggregate	0.05	-0.08	0.17	26	5	0.12
Work aggregate	0.12	-0.22	0.46	5	2	0.61
In employment (dummy)	-0.05	-0.33	0.24	4	1	1.00
Days unable to work^	0.30*	-0.01	0.61	1	1	1.00
Non-work aggregate	-0.00	-0.08	0.08	21	4	0.00
Education	0.27	-0.26	0.80	3	1	1.00
Assets	0.09**	0.01	0.16	3	1	1.00
Income, consumption and input expenditure	-0.12	-1.03	0.80	2	1	1.00
Subjective poverty measures	-0.07	-0.16	0.02	6	1	1.00
Social networks	0.02	-0.05	0.08	7	2	0.00
Panel B: Psychosocial interventions targeting substance use disorders						
All economic outcomes aggregate	0.10	-0.10	0.29	24	5	0.58
Work aggregate	0.09	-0.11	0.28	14	5	0.59
In employment (dummy)	0.15	-0.21	0.50	2	1	1.00
Time in work	0.04	-0.15	0.23	2	1	1.00
Unable to work (dummy)[1]	0.05	-0.40	0.50	4	1	1.00
Days unable to work^	-0.08	-0.38	0.21	4	2	0.00
Functioning at work	0.16	-0.33	0.65	2	2	0.89
Non-work aggregate	0.08	-0.11	0.27	10	1	1.00
Assets	0.08	-0.09	0.26	4	1	1.00
Income, consumption and input expenditure	0.08	-0.12	0.28	6	1	1.00
Panel C: Other intervention type-target condition groupings						
All economic outcomes aggregate	0.11*	-0.02	0.25	42	4	0.24
Work aggregate	0.11	-0.06	0.28	20	4	0.54
Time in work	-0.07	-0.22	0.08	8	1	1.00
Unable to work (dummy)^	0.15	-0.08	0.39	6	1	1.00
Functioning at work	0.24**	0.03	0.44	4	2	0.00
Job search	-0.01	-0.14	0.11	2	1	1.00
Non-work aggregate	0.03	-0.13	0.18	22	1	1.00
Education	0.10	-0.09	0.28	8	1	1.00
Assets	0.01	-0.13	0.15	10	1	1.00
Income, consumption and input expenditure	-0.07	-0.20	0.07	4	1	1.00

Notes: Table A12 reports estimates of the effect of mental health interventions on economic outcomes for additional intervention type-target condition groupings under the frequentist specification outlined in Section 4.2. In Column (1), Hedges' g is the small-sample-bias-corrected standardized mean difference in the economic outcome between treatment and control. The aggregation of individual effect sizes is described in subsection 4.1. The Work aggregate and Non-work aggregate meta-analyses include effect sizes from each of the other outcome groupings below. Panel A presents estimates of the standard deviation effects of interventions targeting common mental disorders, Panel B interventions targeting severe mental disorders. *, ** 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. ^ indicates variables that have been reverse-coded such that higher values indicate improvements in outcomes. All individual effect sizes are winsorized at the 99^{th} percentile.

E.2 Behavioral and psychological pathway outcomes

Table A13: Frequentist estimates of effects of mental health interventions on behavioral and psychological pathways (positively-coded)

	(1) Aggregate Hedges' g	(2) 95% lower CI	(3) 95% upper CI	(4) # of obs.	(5) # of intrv.	(6) I^2
Panel A: Psychosocial interventions targeting common mental disorders						
All mental health disorder symptoms^	0.23***	0.13	0.33	76	14	0.53
Suicide attempts or at risk of suicide	0.16*	-0.00	0.31	12	6	0.34
Relapse (dummy)	0.24***	0.08	0.41	7	3	0.58
Recovery (dummy)	0.23**	0.02	0.45	1	1	1.00
Rehospitalisation (dummy)	0.31***	0.08	0.54	1	1	1.00
Diagnosed with mental disorder (dummy)	0.31**	0.04	0.58	7	4	0.51
CMD symptoms	0.29***	0.14	0.43	38	13	0.73
Overall assessment of mental disorder	0.53***	0.27	0.79	4	3	0.00
All disability and functioning^	0.32***	0.15	0.49	37	12	0.86
Overall measures of functioning	0.16***	0.09	0.23	21	10	0.91
Social support	0.25***	0.09	0.41	15	4	0.54
Functioning in social interactions	0.60***	0.15	1.04	1	1	1.00
Other outcomes^						
Self-regulation	0.15***	0.05	0.24	5	3	0.00
Self-esteem/self-efficacy	0.14	-0.42	0.70	2	2	0.39
Cognition	0.07*	-0.00	0.14	6	2	0.00
Physical health	0.29	-0.18	0.77	6	1	1.00
Panel B: Psychosocial interventions targeting severe mental disorders						
All mental health disorder symptoms^	0.39***	0.20	0.59	23	8	0.64
Suicide attempts or at risk of suicide	-0.01	-0.27	0.26	1	1	1.00
Relapse (dummy)	0.33***	0.16	0.50	6	4	0.00
Recovery (dummy)	0.47***	0.17	0.78	2	2	0.00
Rehospitalisation (dummy)	0.13	-0.08	0.34	5	3	0.36
SMD symptoms	0.36	-0.08	0.80	5	4	0.68
Overall assessment of mental disorder	0.10	-0.16	0.37	3	2	0.27
All disability and functioning^	0.41***	0.25	0.57	27	8	0.17
Overall measures of functioning	0.36***	0.20	0.51	20	7	0.18
Social support	0.82***	0.53	1.10	1	1	1.00
Functioning in social interactions	0.59***	0.33	0.85	6	3	0.01
Other outcomes^						
Self-regulation	0.16**	0.00	0.32	2	1	1.00
Self-esteem/self-efficacy	1.00***	0.74	1.26	1	1	0.00
Cognition	0.69***	0.45	0.92	1	1	1.00
Physical health	0.65***	0.36	0.94	1	1	1.00

Notes: Table A13 reports estimates of the effect of mental health interventions on behavioral and psychological pathway outcomes under the frequentist specification outlined in Section 4.2. In Column (1), Hedges' g is the small-sample-bias-corrected standardized mean difference in the economic outcome between treatment and control. Panel A presents estimates of the standard deviation effects of interventions targeting common mental disorders, Panel B interventions targeting severe mental disorders. The "Mental health disorder symptoms" aggregate and "Functioning and disability" aggregate include effect sizes from each of the other outcome groupings below. *, ** 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. ^ indicates variables that have been reverse-coded such that higher values indicate improvements in 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.

Table A14: Effects of mental health interventions on psychological and behavioral pathways: Bayesian approach (positively-coded)

	τ estimate	τ posterior quantiles				Heterogeneity		
	(1) mean	(2) 2.5^{th}	(3) 25^{th}	(4) 75^{th}	(5) 97.5^{th}	(6) σ	(7) $\omega(\tau)$	(8) I^2
Panel A: Treatment effects of psychosocial interventions on *common* mental disorders								
Mental health disorder symptoms[1]	0.25**	0.12	0.2	0.3	0.41	0.16	0.55	0.72
Suicide attempts or at risk of suicide	0.16	-0.07	0.09	0.23	0.42	0.19	0.47	0.61
Relapse (dummy)	0.23	-0.4	0.14	0.33	0.78	0.41	0.06	0.95
Recovery (dummy)	0.08	-1.68	-0.41	0.56	1.73	4.96	0	1
Rehospitalisation (dummy)	0.1	-1.67	-0.4	0.59	1.74	4.98	0	NaN
Diagnosed with mental disorder (dummy)	0.33	-0.28	0.19	0.46	0.97	0.48	0.22	0.87
CMD symptoms	0.31**	0.12	0.24	0.37	0.53	0.26	0.43	0.85
Overall assessment of mental disorder	0.48	-0.3	0.36	0.64	1.04	0.45	0.21	0.79
Functioning and disability[1]	0.32**	0.09	0.25	0.39	0.54	0.35	0.19	0.92
Overall measures of functioning	0.28*	-0.01	0.19	0.37	0.57	0.42	0.12	1
Social support	0.25*	-0.09	0.17	0.32	0.6	0.25	0.21	0.84
Functioning in social interactions	0.18	-1.66	-0.36	0.73	1.8	5.17	0	NaN
Other outcomes[1]								
Self-regulation	0.16	-0.28	0.1	0.23	0.58	0.27	0.14	0.88
Self-esteem/self-efficacy	0.05	-1.34	-0.28	0.39	1.34	1.67	0.04	0.95
Cognition	0.09	-1.09	-0.08	0.28	1.2	1.14	0.02	0.98
Physical health	0.09	-1.68	-0.41	0.6	1.74	5	0	1
Panel B: Treatment effects of psychosocial interventions on *severe* mental disorders								
Mental health disorder symptoms[1]	0.39**	0.16	0.31	0.46	0.65	0.25	0.4	0.73
Suicide attempts or at risk of suicide	0	-1.7	-0.48	0.48	1.7	4.93	0	1
Relapse (dummy)	0.36*	-0.06	0.26	0.45	0.83	0.29	0.41	0.66
Recovery (dummy)	0.34	-1.05	0.13	0.62	1.38	1.26	0.03	0.97
Rehospitalisation (dummy)	0.15	-0.58	0.03	0.29	0.88	0.54	0.11	0.91
SMD symptoms	0.34	-0.52	0.14	0.56	1.18	0.79	0.14	0.92
Overall assessment of mental disorder	0.02	-1.24	-0.23	0.27	1.23	1.38	0.02	0.98
Functioning and disability[1]	0.41**	0.19	0.35	0.48	0.65	0.19	0.61	0.45
Overall measures of functioning	0.36**	0.12	0.29	0.43	0.62	0.2	0.56	0.51
Social support	0.23	-1.66	-0.33	0.83	1.84	5.35	0	NaN
Functioning in social interactions	0.46	-0.52	0.32	0.67	1.14	0.62	0.15	0.87
Other outcomes[1]								
Self-regulation	0.05	-1.68	-0.43	0.53	1.72	4.93	0	NaN
Self-esteem/self-efficacy	0.27	-1.66	-0.33	0.92	1.89	5.56	0	NaN
Cognition	0.2	-1.66	-0.34	0.77	1.81	5.23	0	1
Physical health	0.19	-1.66	-0.35	0.75	1.8	5.19	0	NaN

Notes: Table A14 reports estimates of the effects of mental health interventions on psychological and behavioral pathways under the Bayesian specification outlined in Section 4.2. In column (1), $\hat{\tau}$ is the estimate of the latent treatment effect in standard deviations. *, ** and *** represent significant at the 10, 5 and 1% levels respectively. Columns (2) through (5) present posterior quantile estimates, summarising the distribution of $\hat{\tau}$. Columns (6) through (8) present summary indicators of heterogeneity. The Work aggregate and Non-work aggregate meta-analyses includes effect sizes from each of the other outcome groupings below. Panel A presents estimates of the standard deviation effects of psychosocial interventions targeting common mental disorders and Panel B the effects of psychosocial interventions targeting severe mental disorders. The "Mental health disorder symptoms" aggregate and "Functioning and disability" aggregate include effect sizes from each of the other outcome groupings below. Panel A presents estimates of the standard deviation effects of psychosocial interventions targeting common mental disorders, Panel B the effects of combined interventions targeting common mental disorders and Panel C combined interventions targeting severe mental disorders. $\hat{}$ indicates variables that have been reverse-coded such that higher values indicate improvements in outcomes. The procedure for aggregating multiple effect sizes from a given intervention is described in subsection 4.1. All individual effect sizes are winsorized at the 99^{th} percentile. 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. The average measurement in our sample happens 15.2 months after intervention start.

Table A15: Effects of mental health interventions on behavioral and psychological pathways: Frequentist approach, additional categories (positively-coded)

	(1) Aggregate Hedges' g	(2) 95% lower CI	(3) 95% upper CI	(4) # of obs.	(5) # of intrv.	(6) I^2
Panel A: Psychosocial interventions targeting post-traumatic stress disorder						
All mental health disorder symptoms	0.13	-0.20	0.46	22	5	0.86
Diagnosed with mental disorder (dummy)	0.48***	0.27	0.69	3	2	0.00
CMD symptoms	0.18	-0.17	0.53	11	5	0.84
PTSD symptoms	0.10	-0.20	0.40	11	5	0.86
All disability and functioning	0.21**	0.01	0.40	14	3	0.00
Overall measures of functioning	0.23***	0.10	0.35	6	3	0.00
Social support	0.22*	-0.02	0.46	6	2	0.00
Functioning in social interactions	0.19**	0.01	0.38	2	1	1.00
Panel B: Psychosocial interventions targeting substance use disorders						
All mental health disorder symptoms	0.14**	0.03	0.26	45	5	0.00
Suicide attempts or at risk of suicide	0.09	-0.14	0.31	4	2	0.00
Substance use	0.20***	0.08	0.31	31	5	0.00
CMD symptoms	0.13	-0.09	0.35	7	4	0.65
PTSD symptoms	0.05	-0.14	0.24	2	1	1.00
All disability and functioning	0.16	-0.03	0.36	8	4	0.34
Overall measures of functioning	0.32	-0.31	0.96	5	2	0.78
Social support	0.15	-0.04	0.33	3	2	0.00
Other outcomes[1]						
Self-esteem/self-efficacy	0.49	-0.31	1.28	4	2	0.93
Cognition	0.03	-0.14	0.21	2	1	1.00
Physical health	-0.13	-0.59	0.33	1	1	1.00
Panel C: Other intervention-type target condition groupings						
All mental health disorder symptoms	0.19	-0.08	0.47	21	4	0.75
Suicide attempts or at risk of suicide	-0.27*	-0.54	0.01	1	1	1.00
Relapse (dummy)	0.49**	0.00	0.97	1	1	1.00
Recovery (dummy)	0.31**	0.00	0.62	1	1	1.00
Diagnosed with mental disorder (dummy)	0.08	-0.07	0.23	4	1	1.00
Substance use	0.11	-0.07	0.29	4	1	1.00
CMD symptoms	0.08	-0.08	0.24	4	1	1.00
SMD symptoms	0.37	-0.28	1.02	3	2	0.83
Overall assessment of mental disorder	-0.10	-0.43	0.24	3	2	0.00
All disability and functioning	0.06	-0.17	0.30	6	4	0.83
Overall measures of functioning	0.14	-0.24	0.52	4	3	0.87
Social support	-0.10	-0.21	0.02	2	1	1.00
Other outcomes[1]						
Cognition	-0.12	-0.26	0.02	2	1	1.00
Physical health	0.10	-0.06	0.26	2	1	1.00

Notes: Table A15 reports estimates of the effect of mental health interventions on behavioral and psychological pathway outcomes for additional intervention type-target condition groupings under the frequentist specification outlined in Section 4.2. In Column (1), Hedges' g is the small-sample-bias-corrected standardized mean difference in the economic outcome between treatment and control. The aggregation of individual effect sizes is described in Section 4.1. Panel A presents estimates of the standard deviation effects of interventions targeting common mental disorders, Panel B interventions targeting severe mental disorders. The "Mental health disorder symptoms" aggregate and "Functioning and disability" aggregate include effect sizes from each of the other outcome groupings below. *, ** 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. ˆ indicates variables that have been reverse-coded such that higher values indicate improvements in outcomes. All individual effect sizes are winsorized at the 99^{th} percentile.

F Robustness checks

F.1 Meta-regression and sub-group analyses

We leverage a meta-regression framework to test robustness of our results to study-level heterogeneity and explore determinants of heterogeneity. We extend Equation 1 by allowing for a vector of de-meaned covariates $\tilde{X}_{es}$:

$$\begin{aligned} \hat{\tau}_{es} &= \tilde{\tau}_{es} + \tilde{X}_{es}\beta + \epsilon_{es} \\ \tilde{X}_{es} &= X_{es} - \bar{X}_{es} \quad \forall\, X_{es} \end{aligned} \tag{8}$$

Where $\hat{\tau}_{es}$ is the observed average treatment effect for effect size e taken from study s.

Relative to estimation of Equation 1, we do not aggregate effect sizes within intervention to retain higher variation with respect to study-level covariates. The increase in sample size comes at a cost: we expect dependence between multiple effects from a given intervention, and need to account for overweighting of studies that report many effect sizes. We therefore implement a multivariate random-effects meta-regression procedure, controlling for intervention-level fixed effects. We estimate parameters via restricted maximum likelihood following Jackson et al. (2011).

Table A16: Meta-regression: robustness to study-level covariates

	Dep. var.: work-related outcomes (Hedges' g)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Psychosocial interventions targeting common mental disorders								
Constant ($\tilde{\tau}$ estimate)	0.176	0.247	0.170	0.165	0.186	0.162	0.118	0.236
Standard error of $\tilde{\tau}$	(0.058)	(0.082)	(0.064)	(0.057)	(0.068)	(0.072)	(0.052)	(0.110)
I^2	0.771	0.820	0.776	0.754	0.793	0.828	0.562	0.934
Cochran Q	128	128	127	123	117	97	83	128
Residual heterogeneity (p-value)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Modifier relevance (p-value)	0.003	0.161	0.712	0.048	0.732	0.761	0.294	0.000
Degrees of freedom (#obs-k)	35	34	34	34	31	29	25	34
Number of interventions	14	14	14	14	14	14	8	14
Panel B: Psychosocial interventions targeting severe mental disorders								
Constant ($\tilde{\tau}$ estimate)	0.325	0.245	0.302	0.338	0.360	0.337	0.173	0.339
Standard error of $\tilde{\tau}$	(0.153)	(0.187)	(0.185)	(0.156)	(0.144)	(0.169)	(0.177)	(0.157)
I^2	0.913	0.917	0.925	0.915	0.902	0.926	0.793	0.917
Cochran Q	99	93	98	99	86	96	17	90
Residual heterogeneity (p-value)	0.000	0.000	0.000	0.000	0.000	0.000	0.044	0.000
Modifier relevance (p-value)	0.033	0.442	0.746	0.131	0.281	0.275	0.841	0.384
Degrees of freedom	19	18	18	18	17	17	9	18
Number of interventions	9	9	9	9	9	9	3	9
Included covariates								
Measurement heterogeneity								
Error term variance	No	Yes	No	No	No	No	No	No
Control conditions	No	No	Yes	No	No	No	No	No
Measurement timing	No	No	No	Yes	No	No	No	No
Intervention and context heterogeneity								
Region and income level	No	No	No	No	Yes	No	No	No
Sample characteristics	No	No	No	No	No	Yes	No	No
Intervention costs (USD 2011)	No	No	No	No	No	No	Yes	No
Delivery type	No	No	No	No	No	No	No	Yes
Fixed effects								
Intervention-level	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Table A16 presents the estimated latent treatment effect (Hedges' g) at the mean of the included covariates for a given model ($\tilde{\tau}$), and its standard error from the meta-regression model described in Section F.1. Each column represents a separate meta-regression with the included covariates indicated by "Yes". The error term variance captures the unexplained variation in the included study estimate, and is a commonly used proxy for small sample biases. Control conditions are described in Section 3.1. Measurement timing captures the number of months from the intervention to measurement. "Region and income level" indicates World Bank income groupings and regional classifications. Sample characteristics capture participant age range, gender, a proxy for female labor force participation, and rural status, where available. Intervention costs are average costs measured in 2011 USD. Delivery type is a dummy for whether the intervention was administered by a specialist, or not. "I^2" is the percentage of variation across studies arising due to heterogeneity rather than sampling variance. Residual heterogeneity presents the p-value from the Cochran Q χ^2 test of residual heterogeneity. Modifier relevance presents the p-value from a χ^2 test of joint significance of included moderators (excluding the intercept), or the p-value on the intercept in the univariate model. Degrees of freedom is the degrees of freedom from the test of modifier relevance, and is equal to the number of observed effect sizes, minus the number of included moderators, *k*. All individual effect sizes are winsorized at the 99^{th} percentile.

Table A17: Work effects by other theoretically important dimensions of heterogeneity

	(1) Aggregate Hedges' g	(2) 95% lower CI	(3) 95% upper CI	(4) # of observations	(5) # of interventions
Panel A: Psychosocial interventions targeting common mental disorders					
All interventions	0.16***	0.05	0.27	36	14
Measurement timing					
<6 months after exit	0.31**	0.06	0.55	8	6
6-12 months after exit	0.12**	0.01	0.23	11	6
>1 year after exit	0.04	-0.08	0.16	12	3
Delivery type					
Specialist delivery	0.27**	0.03	0.52	12	7
Non-specialist delivery	0.06**	0.01	0.12	32	11
Costs					
Costs >=100 USD	0.12*	-0.02	0.25	11	5
Costs <100 USD	0.05	-0.06	0.17	16	3
Panel B: Psychosocial interventions targeting severe mental disorders					
All interventions	0.30**	0.06	0.54	20	9
Measurement timing					
<6 months after exit	0.29**	0.04	0.54	13	8
>1 year after exit	0.31**	0.04	0.57	3	1
Delivery type					
Specialist delivery	0.36***	0.09	0.63	18	8
Non-specialist delivery	0.36	-0.14	0.85	4	3
Costs					
Costs >=100 USD	0.09	-0.14	0.31	8	2
Costs <100 USD	0.46**	0.01	0.91	3	1

Notes: Table A17 reports estimates of effect of mental health interventions on the "Work aggregate" within various subsamples. Hedges' g is the small-sample-bias-corrected standardized mean difference in the economic outcome between treatment and control. Panel A presents estimates of the standard deviation effects of interventions targeting common mental disorders, Panel B interventions targeting severe mental disorders. *, ** 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 aggregation of individual effect sizes works as described in subsection 4.1. All individual effect sizes are winsorized at the 99^{th} percentile.

Figure A2: Effects of psychosocial interventions by targeted sample

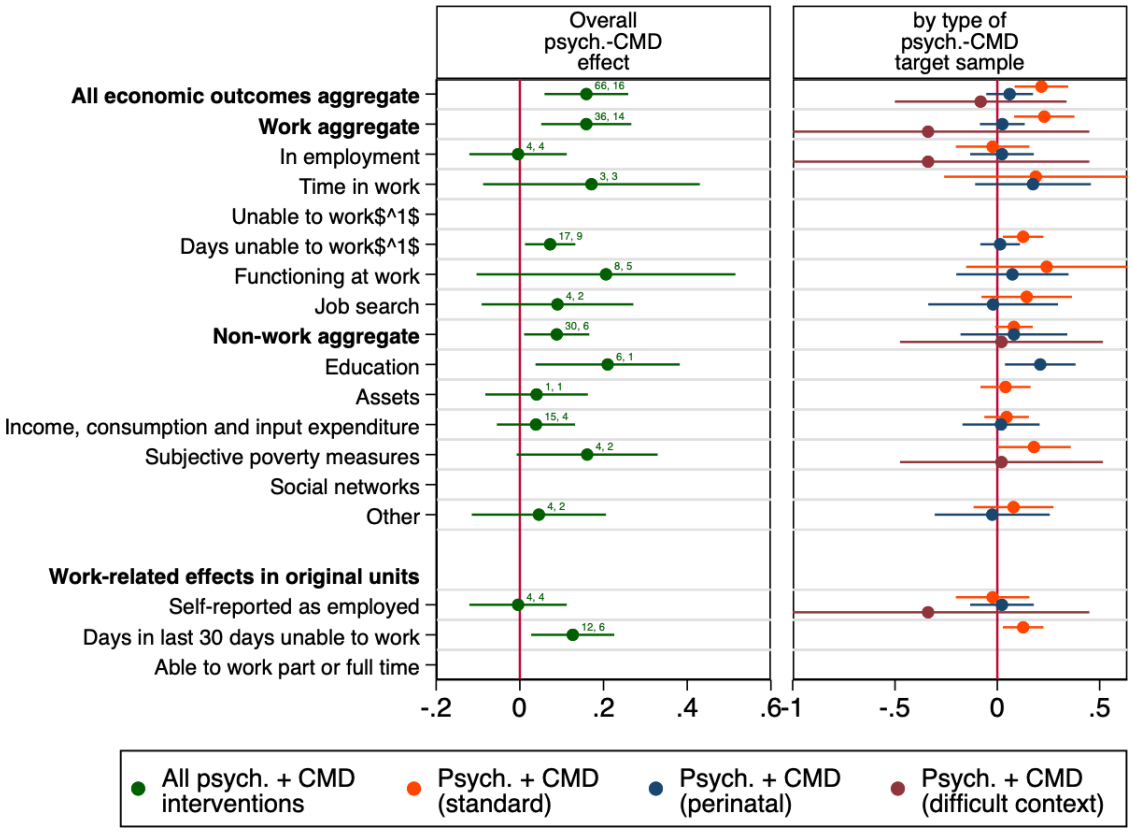


Figure A2 shows aggregate meta-effect sizes (Hedges' g) for various economic outcomes. Colors represent findings from the subsample of interventions implemented in a given target population, namely "standard" populations, those suffering from perinatal depression or "difficult context", where there is very low labor force participation. 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 left 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 which these individual effect sizes come.

Table A18: Meta-regression: differential effects by measured labor market outcome

	Dep. var.: work-related outcomes (Hedges' g)		
	Intercept ($\tilde{\tau}$, SD)	Standard error	# of obs.
Panel A: Psychosocial interventions targeting common mental disorders			
Regression 1, omitted category: "Days unable to work"			
Intercept	0.176	(0.061)	17
In employment	-0.097	(0.061)	4
Time in work	-0.018	(0.090)	3
Functioning at work	-0.054	(0.077)	8
Job search	0.005	(0.078)	4
Residual heterogeneity (p-value)	0.000		
Modifier relevance (p-value)	0.685		
Degrees of freedom (# obs-k)	31		
Number of interventions	14		
Panel B: Psychosocial interventions targeting severe mental disorders			
Regression 2, omitted category: "Days unable to work"			
Intercept	0.312	(0.176)	1
In employment	0.565	(0.616)	1
Time in work	0.091	(0.113)	4
Unable to work	0.036	(0.091)	8
Functioning at work	0.030	(0.361)	2
Residual heterogeneity (p-value)	0.000		
Modifier relevance (p-value)	0.832		
Degrees of freedom (# obs-k)	15		
Number of interventions	9		

Table A18 presents coefficients from two meta-regressions of work-related outcomes on indicators for sub-aggregate work-related outcomes. Each panel represents a separate meta-regression. The intercept term represents the estimated latent treatment effect (Hedges' g) for the omitted category, "Days unable to work". Coefficients on the remaining terms represent differential effects with respect to other sub-aggregate outcomes.

F.2 Heterogeneity in economic effects

We report the distribution of each of the heterogeneity parameters under the Bayesian specification for our core work-related results in Table A19.[42] In Panel A, we present evidence of low-to-moderate heterogeneity in the effect sizes used to estimate the latent effect of psychosocial interventions targeting CMDs on the work aggregate. Heterogeneity is small ($\sigma = 0.05$) and precisely estimated (50% CI: [0.02, 0.06]) and heterogeneity relative to sample variance is low ($\omega(\tau) = 0.81$, $I^2 = 0.20$).

In contrast, we observe moderate-to-high heterogeneity in the interventions targeting SMDs grouping ($\sigma = 0.40$) that is moderately precisely estimated (50% CI: [0.28,0.48]). Again, this is substantial relative relative to sample variance ($\omega(\tau) = 0.24$, $I^2 = 0.85$).

Table A19: Summary statistics of Bayesian posteriors of heterogeneity measures from meta-analyses of the effects of mental health interventions on work-related outcomes

	σ Distribution			$\omega(\tau)$ Distribution			I^2 Distribution		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	mean	25% CI	75% CI	mean	25 pctl. σ	75 pctl. σ	mean	25 pctl. σ	75 pctl. σ
Panel A: Psychosocial interventions targeting common mental disorders									
Work aggregate	0.21	0.16	0.25	0.32	0.26	0.42	0.8	0.7	0.85
In employment (dummy)	0.18	0.05	0.22	0.4	0.35	0.87	0.67	0.12	0.75
Time in work	0.58	0.18	0.65	0.06	0.04	0.36	0.95	0.65	0.96
Days unable to work$^{\wedge}$	0.06	0.02	0.08	0.72	0.59	0.94	0.31	0.05	0.47
Functioning at work	0.48	0.29	0.57	0.08	0.06	0.19	0.93	0.84	0.95
Job search	1.13	0.17	1.21	0.01	0.01	0.38	0.98	0.6	0.99
Panel B: Psychosocial interventions targeting severe mental disorders									
Work aggregate	0.45	0.32	0.55	0.21	0.17	0.33	0.88	0.79	0.92
In employment (dummy)	5.37	1.02	5.87	0	0	0.09	NaN	NaN	NaN
Time in work	1.28	0.22	1.4	0.02	0.02	0.34	0.98	0.61	0.98
Unable to work (dummy)$^{\wedge}$	0.4	0.09	0.44	0.13	0.11	0.67	0.9	0.29	0.92
Days unable to work$^{\wedge}$	4.92	0.83	5.32	0	0	0	NaN	NaN	NaN
Functioning at work	1.12	0.62	1.34	0.07	0.05	0.18	0.97	0.9	0.98

Notes: The Frequentist and Bayesian specifications are outlined in Section 4.2. $\hat{\tau}$ is the MLE estimator under the Frequentist specification, or posterior mean (most likely value) under the Bayesian specification of the treatment effects and is measured in standard deviations. The Work aggregate meta-analysis includes effect sizes from each of the other outcome groupings below. Panel A presents estimates of the standard deviation effects of psychosocial interventions targeting common mental disorders, Panel B the effects of interventions targeting severe mental disorders. Details of scale wording is provided in Table A6. $^{\wedge}$ indicates variables that have been reverse-coded such that higher values indicate improvements in outcomes. NaN represents the estimated I^2 value "Not a number" as reported by $\mathcal{R}$. It can be taken as ≈ 1.

[42] In small samples, the I^2 statistic is biased and tends to have wide confidence intervals, complicating inference on heterogeneity (von Hippel, 2015).

F.3 Publication bias

In this section, we present results from tests for publication bias on the full sample of 40 studies captured by our systematic review. This best reflects the extent of publication bias in the literature defined by our pre-specified search criteria.

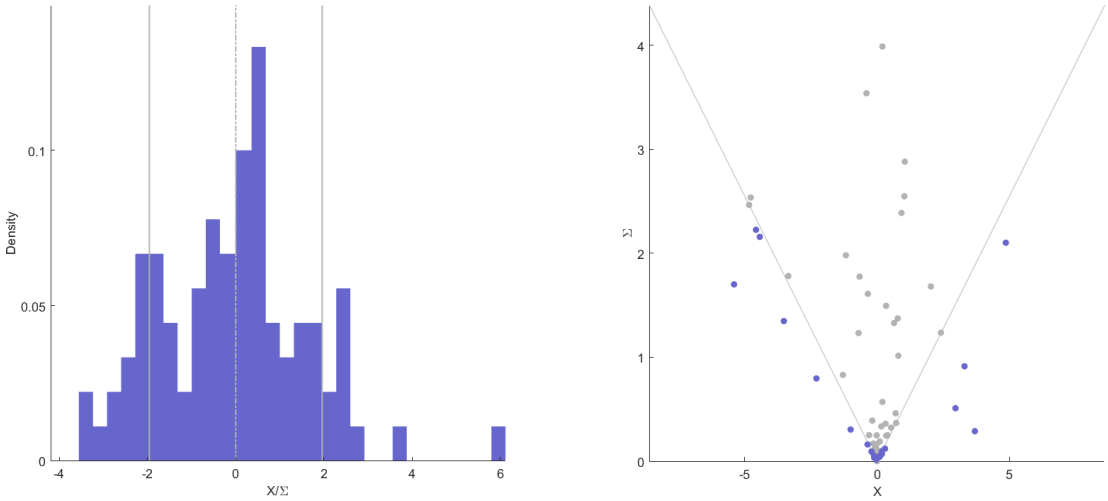
F.3.1 Funnel plot asymmetry

Table A20: 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 A20 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.

Figure A3: Histogram and funnel plot of reported effect sizes



Notes: Figure A3 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 or 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.

F.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. 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.

$$\Theta^* \sim \bar{\theta} + t(\tilde{\nu}) \cdot \tilde{\eta}, \quad p(Z) \propto \begin{cases} \beta_{p,1} & \text{if } Z < -1.96 \\ \beta_{p,2} & \text{if } Z \in [-1.96, 0) \\ \beta_{p,3} & \text{if } Z \in [0, 1.96) \\ 1 & \text{if } Z \geq 1.96 \end{cases} \tag{9}$$

Where θ^* is the distribution of latent study effects, modeled as a t-distribution with degrees of freedom $\tilde{\nu}$ location parameter $\bar{\theta}$ and scale parameter $\tilde{\eta}$. 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 A21.

Table A21: 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 A21 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 find no 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 substantially

more precisely estimated.

Table A22: 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 A21 displays the results of the MLE model for publication bias implemented on the subsample 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.4 Heterogeneity in mental health effects

F.4.1 Heterogeneity by measuring party

Figure A4: Robustness to disaggregation by party responsible for measurement

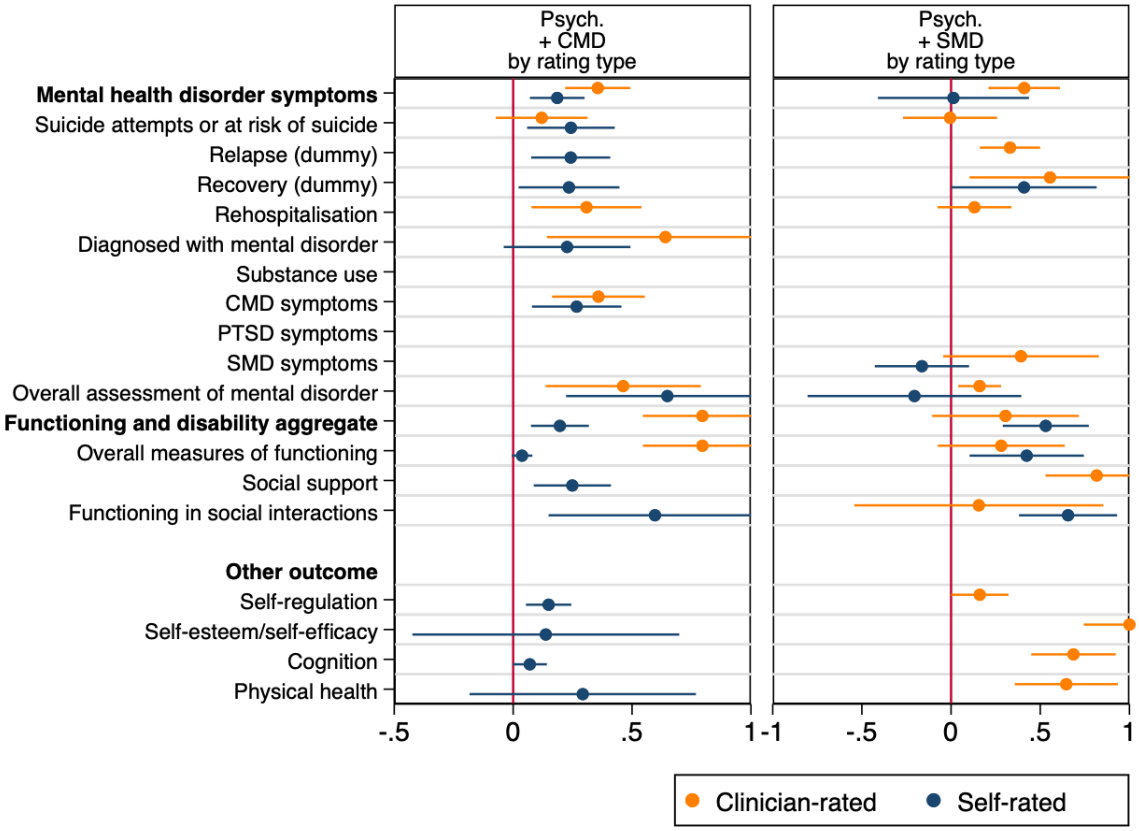


Figure A4 shows aggregate mental health meta-effect sizes (Hedges' g) on various psychological and behavioral pathways outcomes and aggregates. The two panels present 95% confidence intervals for effects of each intervention-target condition combination. The horizontal axis displays the average effect size in standard deviations. Colors represent findings from the subsample of interventions that were measured by a given party, namely the trial participant (self-rated) or a clinician (clinician-rated). 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.

G Pooled microdata analysis details

G.1 Data availability

Table A23: Data inclusion attempts for interventions in microdata sample

Study name	Contact with authors	Data available	Depression?	Work days?	In pooled analysis
Study did not have all relevant variables					
Ayoughi et al. (2012)	No	Unknown	Yes	No	No
Bolton et al. (2003)	No	Unknown	Yes	No	No
Haushofer et al. (2020)	No	Unknown	No	No	No
Hirani et al. (2010)	Unsuccessful attempt	Unknown	Yes	No	No
Duarte et al. (2009)	Yes	Do not want to share	Yes	No	No
Included in sample of microdata studies					
Baranov et al. (2020)	Yes	Yes	Yes	Yes	Yes
Barker et al. (2022)	Yes	Yes	Yes	Yes	Yes
Fuhr et al. (2019) / Bhat et al. (2022)	Yes	Yes	Yes	Yes	Yes
Patel et al. (2017) / Weobong et al. (2017)	Yes	Yes	Yes	Yes	Yes
Patel et al. (2011) / Buttorff et al. (2012)	Yes	Yes	Yes	Yes	Yes
Sikander et al. (2019) / Bhat et al. (2022)	Yes	Yes	Yes	Yes	Yes

G.2 Variable construction details

In our microdata analysis, reported on in Table 3 and Table A25, we make use of the following directly comparable measures of "Days unable to work" from across included studies:

- Fuhr et al. (2019): 'Number of days unable to work in the previous month'
- Sikander et al. (2019): 'Number of days unable to work in the previous month'
- Barker et al. (2022): 'Number of days unable to work in the previous month'
- Weobong et al. (2017): 'Number of days unable to work in the previous month'
- Patel et al. (2011): 'Number of days unable to work in the previous month'
- Baranov et al. (2020): This study reports on 'Number of healthy days in past 30 days'. While this is a close proxy for number of days able to work, to ensure comparability, we report findings against this outcome separately.

In Table 3, we report findings from a regression of "Days unable to work" on a combined measure of depression, instrumented by treatment status. Our combined depression measure is constructed from the main/preferred depression measure captured by a given study as determined by the study's authors. We standardise each of these measures within each study sample, then aggregate them across studies. The main depression measure from each of the included studies is as follows:

- Patient Health Questionnaire-9 (PHQ-9): Fuhr et al. (2019); Sikander et al. (2019); Weobong et al. (2017)
- Beck Depression Index (BDI): Baranov et al. (2020)
- Diagnostic and Statistical Manual of Mental Disorders - IV (DSM-IV): Patel et al. (2011)
- Kessler Psychological Distress K10 Scale (Kessler): Barker et al. (2022)

In Table A24, we report findings from a first stage regression of measures of depression on treatment status. In Columns (1) to (3), the combined measure is encoded as described above. In Column (4) we report effects on the Patient Health Questionnaire-9 (PHQ-9) which are aggregated from across three studies: Fuhr et al. (2019); Sikander et al. (2019); Weobong et al. (2017). In Columns (5) to (7) we report effects of treatment on

the BDI scale from Weobong et al. (2017), the DSM-IV scale from Patel et al. (2011) and the Kessler Psychological Distress K10 Scale from Barker et al. (2022).

In Table A24 and Table A25, we report interaction effects between treatment and categorical measures of depression. The cutoff used in each of the continuous depression measures to construct the categories of mild, moderate and severe depression is taken directly from the assessment criteria for each of the measures.

G.3 Details of the median split model

For continuous dimension of heterogeneity X_i we estimate:

$$Y_{is} = \beta_0 + \beta_1 T_i + \beta_2 I[X_i > \text{M}(X_i)] + \beta_3 T_{is} * I[X_i > \text{M}(X_i)] + S_i + \epsilon_{is}, \quad \text{for M}(X) = \text{Median}(X) \tag{10}$$

where Y_{is} the outcome for participant i from study s', T_i is an indicator for whether i was randomly allocated to CBT (as opposed to being in the control group) and X_i is a continuous measure of a dimension of heterogeneity. S_s is a study fixed effect. The median of each of the dimensions of heterogeneity is calculated *across* (not within) studies.

G.4 Results from analysis of microdata

Table A24: First stage: effect of psychosocial treatments on depression

	Combined measure			PHQ-9	BDI	DSM-IV	Kessler
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treatment	-0.222***	-0.236***	-0.124*	-0.194**	-0.309***	-0.662***	-0.186***
	(0.061)	(0.058)	(0.069)	(0.096)	(0.098)	(0.129)	(0.037)
Above median age		0.059*					
		(0.032)					
Treatment=1 × Above		0.033					
median age		(0.053)					
Moderate depr.			0.235***				
			(0.043)				
Severe depr.			0.428***				
			(0.056)				
Treatment=1 × Moderate			-0.173***				
depr.			(0.052)				
Treatment=1 × Severe			-0.230***				
depr.			(0.076)				
Constant	-0.002	-0.031	-0.138***	-0.000	-0.000	0.000	-0.000
	(0.043)	(0.041)	(0.048)	(0.084)	(0.066)	(0.116)	(0.029)
Study FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control mean	-0.00	-0.00	-0.00	0.00	-0.00	-0.00	-0.00
p(T × Mild=T × Mod)			0.00				
p(T × Mild=T × Sev)			0.00				
p(T × Mod=T × Sev)			0.40				
# of participants	10731	10731	10731	3138	447	429	6717
Obs.	15517	15517	15517	7924	447	429	6717
Studies	6	6	6	3	1	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). Variable construction is further detailed in Appendix G.2. Standard errors are in parentheses and clustered by original study cluster variable. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A25: Heterogeneous treatment effects of psychosocial treatments on days able to work measures

	Days unable to work			Healthy days
	(1)	(2)	(3)	(4)
Treatment	-1.571*	-2.133***	-1.296	0.288
	(0.855)	(0.783)	(0.843)	(0.661)
Above median age		1.555***		
		(0.464)		
Treatment=1 × Above		1.195		
median age		(0.859)		
Moderate depr.			1.046***	
			(0.343)	
Severe depr.			1.421***	
			(0.398)	
Treatment=1 × Moderate			-0.656	
depr.			(0.468)	
Treatment=1 × Severe			-0.182	
depr.			(0.648)	
Constant	7.027***	6.234***	6.507***	26.155***
	(0.588)	(0.617)	(0.569)	(0.476)
Study FE	Yes	Yes	Yes	Yes
Control mean	6.43	6.43	6.43	26.16
Standard deviation	9.86	9.86	9.86	7.66
p(T × Mild=T × Mod)			0.16	
p(T × Mild=T × Sev)			0.78	
p(T × Mod=T × Sev)			0.48	
# of participants	10302	10302	10302	429
Obs.	15088	15088	15088	429
Studies	5	5	5	1

Notes: This table shows four different OLS regressions 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 (1-3) show the impacts on days unable to work in the last month, column (4) shows the impact on healhy days per month. Columns (2) and (3) show heterogenous impacts by age median splits and depression status. Variable construction is further detailed in Appendix G.2. Standard errors are in parentheses and clustered by original study cluster variable. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

H Details of 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**	Costs (per cap.)
Psychosocial – Common Mental Disorders (CMD)										
Ayoughi et al. 2012	Afghanis-tan	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. 2020 (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	10 USD, 2005/2006
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	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	1.36 USD, 2016
Haushofer et al. 2020	Kenya	2122 adults 18+	Psychosocial	Problem Solving Therapy, Behavioural Activation	No treatment	1 session weekly for 5 weeks	13 months	Common mental disorders	Assets, Income, consumption & input expenditure	1189 USD, 2017
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	-
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	1; 10; 58 months	Depression	Income, Consumption & input expenditure, Days unable to work, Functioning at work, Employment, Job search, Time in work	66 USD, 2015
Patel et al. 2011; Buttorff et al. 2012	India	213 adults 17+	Psychosocial (Public facility)	Interpersonal Therapy, Psychoeducation	EUC: Facility in patients' community received their screening results	6 sessions over 12 months	9 months	Common mental disorders	Days unable to work	89 USD, 2009
	India	341 adults 17+	Psychosocial (Private facility)	Interpersonal Therapy, Psychoeducation		6 sessions over 12 months	9 months	Common mental disorders	Days unable to work	89 USD, 2009
	India	1648 adults, 17+	Psychosocial (Public facility)	Psychoeducation		6 sessions over 12 months	9 months	Common mental disorders	Days unable to work	89 USD, 2009
	India	1148 adults, 17+	Psychosocial (Private facility)	Psychoeducation		6 sessions over 12 months	9 months	Common mental disorders	Days unable to work	89 USD, 2009

Author, year	Country	Sample & age	Intervention category	Therapeutic type	Control group category	Description of intervention	Follow-up time points*	Target mental disorder	Economic outcomes**	Costs (per cap.)
Psychosocial – Common Mental Disorders (CMD) (cont.)										
Sikander et al. 2019; Bhat et al. 2022	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	133.55 USD, 2016
Combination – Common Mental Disorders (CMD)										
Hu et al. 2007	China	76 adults, 18+	Psychosocial	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. 2013	India	60 adults, 18-65	Psychosocial	Interpersonal therapy, Problem Solving Therapy, Family Therapy, Social Support & Medication: Antidepressants	TAU Pharm: Antidepressants	10 sessions over 3 months	0 months	Anxiety	Functioning at work	-
Combination – 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	6825 INR, 2009/2010
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, Functioning at work	-
Luo et al. 2019	China	58 adults, 16+	Psychosocial & Pharmacological	Problem Solving Therapy, Psychoeducation, Family Therapy, & Medication: Antipsychotics	TAU Pharm: Antipsychotics	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	EUC	1 session monthly & 3 family workshops for 9 months	0 and 159 months	Schizophrenia	Unable to work	3100 RMB, 1994
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, Functioning at work	-
Vizzotto et al. 2021	Brazil	48 adults, 18-55	Psychosocial & Pharmacological	Occupational Goal Intervention (OGI) & Medication: Antipsychotics	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	-

Author, year	Country	Sample & age	Intervention category	Therapeutic type	Control group category	Description of intervention	Follow-up time points*	Target mental disorder	Economic outcomes**	Costs (per cap.)
Combination – Severe Mental Disorders (SMD) (cont)										
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	31.5 USD,1990-1992
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	-
Psychosocial – 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	314 USD, 2009-2012
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	33 USD, 2015
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	39.93 USD, 2015
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	-
Min et al. 2011	China	100 adults, 18+	Psychosocial	Cognitive Behavioural Therapy	EUC: Inpatient drug rehab. centre	20 sessions over 2 months	1 month	Substance dependence (Heroin)	Employment	-
Psychosocial – 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	No treatment	2 day-long workshops	9 and 31 months	PTSD	Assets, Employment, Financial, Consumption	-
Hall et al. 2014	DRC	405 women, 18+	Psychosocial	Cognitive Behavioural Therapy based	EUC: Invitation to access existing services	1 sessions weekly for 11 weeks (3 months)	2 and 7 months	PTSD	Assets, Social networks, Subjective poverty measures	1530.03 USD, 2011
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 2017	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	1019 EUR, 2012
Author, year	**Country**	**Sample & age**	**Intervention category**	**Therapeutic type**	**Control group category**	**Description of intervention**	**Follow-up time points***	**Target mental disorder**	**Economic outcomes****	**Costs (per cap.)**
Other										
Angelucci et al. 2024	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	221 USD, 2017
Gureje et al. 2020	Ghana and Nigeria	286 adults, 18+	Psychosocial & Pharmacological	Collaborative shared care: Traditional & Faith Healers & PHC; & Medication: Antipsychotics	No treatment	At least 1 visit weekly over 3-6 months	0 months	Schizophrenia	Functioning at work	444 USD, 2017/2018
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	-
Ran et al. 2003; 2015	China	326 adults, 18+	Pharmacological	Medication: Antipsychotics	EUC	Medication for 9 months	0 and 159 months	Schizophrenia	Unable to work	1300 RMB, 1994

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

I Costs

Cost data was available for the following studies: Blattman et al. (2017), Nadkarni et al. (2017b,a), Ran et al. (2003, 2015), Patel et al. (2017), Weobong et al. (2017), Buttorff et al. (2012), Hall et al. (2014), Xiong et al. (1994), Wang et al. (2017), Baranov et al. (2020), Fuhr et al. (2019), Sikander et al. (2019), Angelucci and Bennett (2024), Nadkarni et al. (2019), Bhat et al. (2022), Haushofer et al. (2020), Chatterjee et al. (2014), Luo et al. (2019), Gureje et al. (2020), Patel et al. (2011).

For the remaining papers, cost data was not available: Valencia et al. (2007), Xiang et al. (1994), Betancourt et al. (2014), Bolton et al. (2003), Duarte et al. (2009), Ayoughi et al. (2012), Hirani et al. (2010), Min et al. (2011), Nagarajaiah et al. (2013), Pan et al. (2015), Cilliers et al. (2016), Zhang et al. (1998), Xiong et al. (2007), Chuan-qian et al. (2005), Barker et al. (2022), Meffert et al. (2021), Vizzotto et al. (2021), Xu et al. (2021).

Table A26: Cost overview

	Median	(10^{th} pct)	(90^{th} pct)
By intervention-condition combination			
Psych. + CMD	104.98	1.43	1226.41
Combination + SMD	180.09	55.69	570.96
Psych. + PTSD	1599.49	1456.28	1742.69
Psych. + SUD	42.52	35.14	370.37
Other interventions	239.43	227.95	457.97
By region			
East Asia & Pacific	239.43	55.69	570.96
Europe & Central Asia	1456.28	1456.28	1456.28
Latin America & Caribbean			
South Asia	104.98	12.79	180.09
Sub-Saharan Africa	842.19	370.37	1742.69

Notes: This table shows intervention costs per participant in 2011 US-Dollars. Column 1 shows the median, columns 2 and 3 show the 10th and 90th percentile. No cost data is available for studies in Latin America.

Table A27: Study characteristics by whether cost data is available

	(1) Share (with cost data)	(2) (SD)	(3) Share (without cost data)	(4) (SD)	(5) Cost data difference	(6) (SE)
Panel A: All interventions						
Intervention-condition combination *(mutually exclusive)*						
Psychosocial + common mental disorders (CMD)	0.45	(0.50)	0.37	(0.50)	0.08	(0.16)
Psychosocial + severe mental disorders (SMD)	0.15	(0.37)	0.32	(0.48)	-0.17	(0.14)
Psychosocial + post-traumatic stress disorders (PTSD)	0.10	(0.31)	0.16	(0.37)	-0.06	(0.11)
Psychosocial + substance use disorders (SUD)	0.15	(0.37)	0.11	(0.32)	0.04	(0.11)
Other interventions	0.15	(0.37)	0.05	(0.23)	0.10	(0.10)
Control condition *(mutually exclusive)*						
Enhanced Usual Care	0.55	(0.50)	0.05	(0.23)	0.50	(0.13)***
No Treatment	0.35	(0.49)	0.37	(0.50)	-0.02	(0.16)
Treatment As Usual (Pharmacological)	0.10	(0.31)	0.58	(0.50)	-0.48	(0.14)***
Panel B: Outcome measures						
Economic outcomes						
Employment (dummy)	0.25	(0.44)	0.11	(0.32)	0.14	(0.12)
Time in work	0.25	(0.44)	0.11	(0.32)	0.14	(0.12)
Unable to work	0.15	(0.37)	0.11	(0.32)	0.04	(0.11)
Days unable to work	0.50	(0.50)	0.16	(0.37)	0.34	(0.14)**
Functioning at work	0.20	(0.41)	0.47	(0.50)	-0.27	(0.15)*
Job search	0.15	(0.37)	0.00	(0.00)	0.15	(0.08)*
Education	0.10	(0.31)	0.05	(0.23)	0.05	(0.09)
Assets	0.15	(0.37)	0.05	(0.23)	0.10	(0.10)
Income, consumption and input expenditure	0.35	(0.49)	0.00	(0.00)	0.35	(0.11)***
Subjective poverty measures	0.00	(0.00)	0.21	(0.42)	-0.21	(0.10)**
Social networks	0.05	(0.22)	0.05	(0.23)	-0.00	(0.07)
Other	0.10	(0.31)	0.05	(0.23)	0.05	(0.09)
Mental health outcomes (all)						
Suicide attempts or at risk of suicide	0.45	(0.50)	0.05	(0.23)	0.40	(0.13)***
Relapse (dummy)	0.30	(0.47)	0.11	(0.32)	0.19	(0.13)
Recovery (dummy)	0.15	(0.37)	0.05	(0.23)	0.10	(0.10)
Rehospitalisation	0.05	(0.22)	0.16	(0.37)	-0.11	(0.10)
Diagnosed with mental disorder	0.20	(0.41)	0.16	(0.37)	0.04	(0.13)
Overall assessment of mental disorder	0.15	(0.37)	0.21	(0.42)	-0.06	(0.13)
Substance use	0.15	(0.37)	0.16	(0.37)	-0.01	(0.12)
CMD symptoms	0.70	(0.47)	0.47	(0.50)	0.23	(0.16)
PTSD symptoms	0.15	(0.37)	0.16	(0.37)	-0.01	(0.12)
SMD symptoms	0.20	(0.41)	0.11	(0.32)	0.09	(0.12)
Overall measures of functioning	0.80	(0.41)	0.47	(0.50)	0.33	(0.15)**
Functioning in social interactions	0.05	(0.22)	0.21	(0.42)	-0.16	(0.11)
Self-regulation	0.10	(0.31)	0.11	(0.32)	-0.01	(0.10)
Self-esteem/self-efficacy	0.05	(0.22)	0.21	(0.42)	-0.16	(0.11)
Cognition	0.10	(0.31)	0.16	(0.37)	-0.06	(0.11)
Physical health	0.00	(0.00)	0.21	(0.42)	-0.21	(0.10)**

Notes: This table shows the mean prevalence for each intervention characteristic listed in the rows separately by whether the intervention reports cost data (columns 1 and 2) or not (columns 3 and 4). The mean difference in study characteristic by whether the intervention has cost data or not is calculated by OLS regression in columns 5 and 6. Each row is based on a separate regression over all N=39 interventions.