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

Crick Lund[*], *Kate Orkin*[*,**], *Marc Witte*[*], *Thandi Davies*, **John Walker**[**], *Johannes Haushofer*, *Judy Bass*, *Paul Bolton*, *Sarah Murray*, *Laura Murray*, *Wietse Tol*, *Graham Thornicroft*, *Vikram Patel*

[*] Joint first author, [**]University of Oxford

April 14, 2023

Treating mental health disorders is cheap, highly effective and might plausibly affect economic outcomes

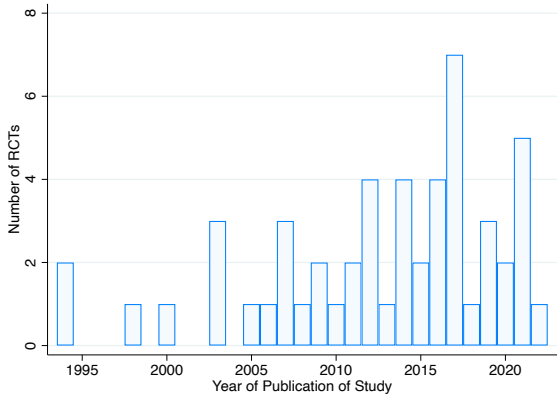
- Clinical improvements in mental health disorders are associated with improved functioning in daily living (Thornicroft and Tansella, 2010)
  - Cognition / Sleep / Stable affect / Improved social functioning
- In high-income countries, treating disorders reduces sick leave and absenteeism and improves likelihood of return to work (Nigatu et al. 2016, Nieuwenhuijsen et al., 2020, Salomonsson et al., 2018)
- Psychosocial and pharmacological treatments, as well as their combination, have been effective in reducing symptoms of mental health disorders in low income settings (Cuijpers et al. 2018, Scott et al., 2016, Singla et al., 2017)

Treating mental health disorders is cheap, highly effective and might plausibly affect economic outcomes

- Clinical improvements in mental health disorders are associated with improved functioning in daily living (Thornicroft and Tansella, 2010)
  - Cognition / Sleep / Stable affect / Improved social functioning
- In high-income countries, treating disorders reduces sick leave and absenteeism and improves likelihood of return to work (Nigatu et al. 2016, Nieuwenhuijsen et al., 2020, Salomonsson et al., 2018)
- Psychosocial and pharmacological treatments, as well as their combination, have been effective in reducing symptoms of mental health disorders in low income settings (Cuijpers et al. 2018, Scott et al., 2016, Singla et al., 2017)
- However, an open question is whether mental health interventions also produce *economic* effects in low-income settings

The recent increase in RCTs in LMIC has enabled meta-studies

Increase in mental health studies in low and middle income countries



This Paper

Does receiving treatment for a mental health disorder affect economic outcomes in LMIC?

Systematic search: studies in LMIC that screen for a mental health condition, test a mental health intervention and measure an economic outcome until Aug 2022

- Screen 15,022 papers, full reading of 1,119
- **147 RCT estimates of effect sizes of interventions from 39 RCT papers**

We run meta-analyses comparing studies that share:

- an intervention type (psychosocial, pharmacological or a combination) &
- a target condition (common mental disorders, severe mental disorders)

Headline results

Our meta-analysis estimates within intervention-type target condition pairs indicate large positive effects

- Combined psychosocial and pharmacological interventions have large effects on an aggregate of work outcomes
- These effects are qualitatively similar whether targeting common mental disorders ($SD = 0.35$, $p < 0.05$) or severe mental disorders ($SD = 0.28$, $p < 0.05$)
- Combined interventions targeting CMD reduce days unable to work by 2.5 days ($p < 0.05$), or 24% relative to a control mean of 10 days
- Psychosocial interventions alone targeting common mental disorders have modest effects ($SD = 0.05$, $p < 0.1$)

We aren't powered to look at pharmacological treatments alone (n=2) or PTSD/SUD subgroupings (n=5)

We see little evidence of publication bias in our sample, and results are robust to controls for study characteristics and accounting for study-level heterogeneity

Discussion preview

Which economic outcomes are affected?

- We aren't powered to look at individual outcomes, but see particularly large positive coefficients on employment status, time in work and functioning at work

Mechanisms - how are mental health interventions driving these improvements?

- Interventions in this sample improve mental health symptoms and functioning
- Interventions with larger +ve effects on mental health have larger effects on work-related outcomes
- Using microdata collected from 5 studies, at the individual level, treatment reduces midline depression and endline days unable to work (not reverse causality)

How costly are the 19 interventions on which we could access cost data?

- Psychosocial interventions targeting CMD costed USD 290.27, combined intervention for CMD (105 USD), for SMD (316 USD)

Related literature

Economic literature on economic effects of mental health treatments (e.g. Biasi et al. 2021, Butikofer et al. 2020, Blattman et al. 2017, Baranov et al. 2020...) and health (e.g. Nigatu et al. 2016, Nieuwenhuijsen et al., 2020, Salomonsson et al., 2018...)

$\longrightarrow$ **First systematic review and meta-analysis across disciplines and languages. Complements work on effect of economic interventions on mental health**

Economic literature on meta-analysis (Ridley et al. 2020, Card et al. 2018, Meager 2019, Vivalt, 2020, Tan and Kremer 2020)

$\longrightarrow$ **Highlight efficacy for aggregating across many underpowered studies**

Economic literature on effect of poor physical health on work days and productivity

$\longrightarrow$ **Examine interventions which are relatively underprioritised by global development funders**

Search strategy

- Systematic review registered with the Prospective Register of Systematic reviews (PROSPERO), Protocol number: CRD42017058930, on 18 April 2017
- 21 databases including SSRN, RePEc, JPAL Evaluation and Publication Database, the World Bank Poverty Impact Evaluations Database, Research for Development (R4D), ECON-LIT; 8 trial registries
- Reference snowballing: hand searches of the citation lists of all included articles for working papers (90 papers)
- The search period ended August 2022
- Included papers in any language if an English abstract was provided

Inclusion criteria

An effect size from a given study sub-sample was included if:

- The study was an RCT testing the impact of a psychosocial, pharmacological or combined intervention on at least one of our pre-specified economic outcomes in a low or middle income country
- The study sub-sample for which an effect was calculated screened positive for a mental disorder at baseline and was over 14
- We excluded interventions with an economic component

Most RCT control groups received no treatment (49% of interventions) or enhanced usual care (18%). Some in psychological treatments received a drug (13%).

Extraction

- 2 reviewers independently screened abstracts against inclusion criteria
- Disagreements were resolved by a third reviewer
- Assessed inter-rater agreement with the Kappa statistic, which measures the probability of agreement between raters $= 0.90$
- 15022 abstracts screened, 1119 full papers screened, 39 papers, 39 interventions
- Data was then extracted. We requested information for computing effect sizes from authors where needed

Study sample characteristics by geography

	(1) Number of interventions	(2) Share of interventions
Country income *(mut. excl.)*		
Upper middle income country	14	0.36
Lower middle income country	19	0.49
Low income country	6	0.15
Regions *(mut. excl.)*		
Sub-Saharan Africa	8	0.21
Europe and Central Asia	2	0.05
Latin America and the Caribbean	3	0.08
South Asia	15	0.38
East Asia and Pacific	11	0.28

Notes: There are 30 interventions, 115 economic effect sizes and 256 mental health effect sizes. Categories are mutually exclusive.

Study sample characteristics by intervention type

Control condition

	(1) Number of interventions	(2) Share of interventions
Main intervention category *(mut. excl.)*		
Pharmacological	2	0.05
Psychosocial (only)	22	0.56
Pharm.+psych	15	0.38
Targeted condition *(mut. excl.)*		
Common mental disorders	17	0.44
Severe mental disorders	11	0.28
Substance use disorders	6	0.15
Post-traumatic stress disorders	5	0.13

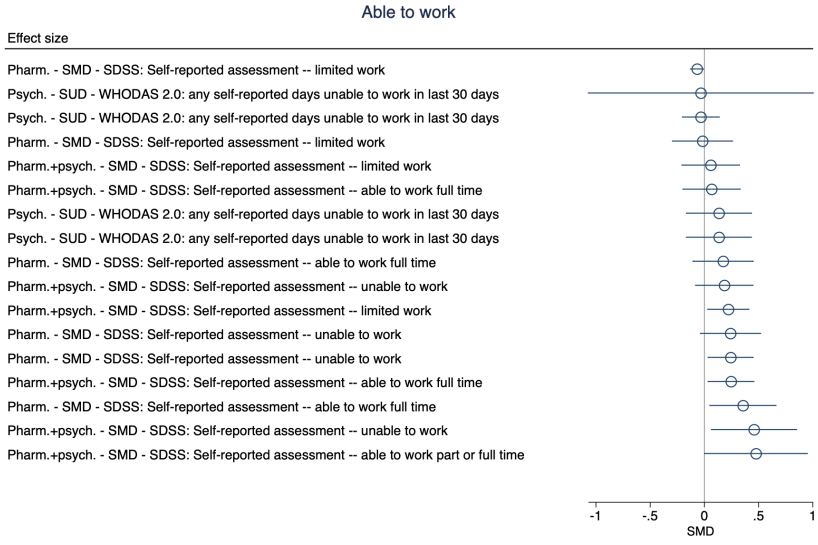
Notes: There are 30 interventions, 115 economic effect sizes and 256 mental health effect sizes. Categories are mutually exclusive.

Study sample

Details

- Main outcome measure: Standardised mean difference (SMD): difference between treatment and control mean divided by standard deviation, with a small sample adjustment
- Within studies and across studies, we consider multiple observations of economics outcomes of a single intervention. We aggregate across these to avoid "double counting"
- Sample: 147 economic effect sizes and 336 mental disorder/functioning effect sizes
- To perform meta-analyses, we aggregate across groups of outcomes that could be argued to measure the same underlying construct and be affected by treatment in similar ways
  - e.g. "Time worked": aggregate amount worked over different recall periods
  - e.g. Don't aggregate "Time worked" and "Days unable to work"

Example: unable to work dummy



Based on 4 interventions

Empirical strategy overview

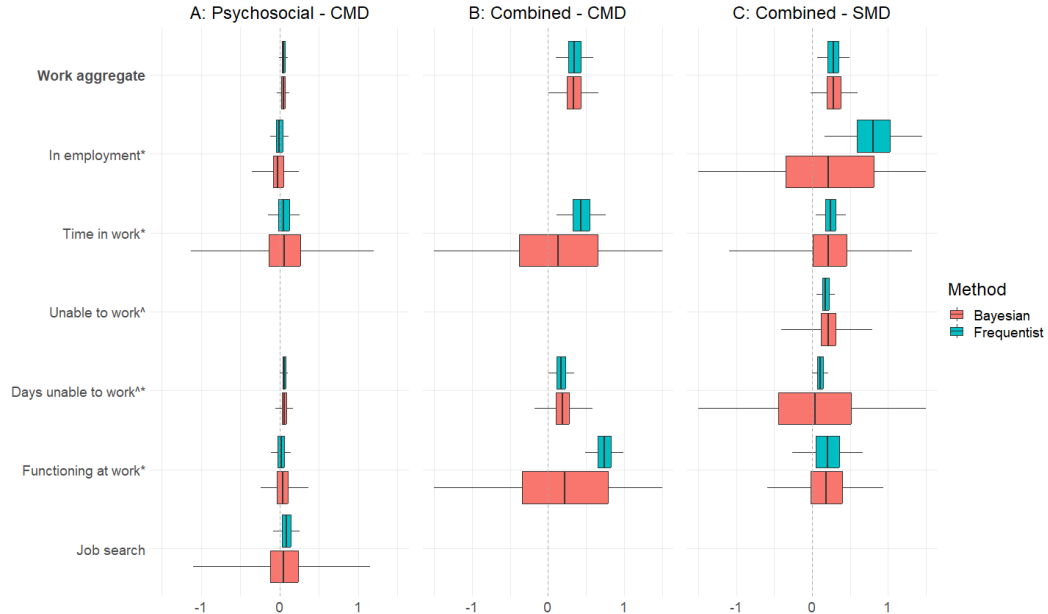
Performing repeated meta-analyses for each outcome of interest (disaggregated by intervention - target condition pair)

1. Standardise outcomes within study
2. Generate a dataset of effect sizes at intervention level: if more than one effect size, aggregate these so studies with more measurements are not weighted more
3. Perform a random effects meta-analysis (weighting by precision)
4. Perform a Bayesian hierarchical meta-analysis (weighting by heterogeneity)

In each figure that follows, each row will represent the results of a meta-analysis of an outcome group of interest (specified by the row) on an intervention of interest, within a sample affected by a given mental disorder

Frequentist estimates of treatment effects

	Estimate	95% CI lower	95% CI upper	# of effect sizes	# of interventions	I^2
Panel A: Psychosocial interventions targeting common mental disorders						
Work aggregate	0.05*	-0.01	0.10	25	8	0.00
In employment	-0.00	-0.12	0.11	4	4	0.00
Time in work	0.05	-0.15	0.25	2	2	0.23
Days unable to work[1]	0.05*	-0.00	0.11	8	5	0.00
Functioning at work	0.02	-0.11	0.14	7	4	0.00
Job search	0.08	-0.08	0.25	4	2	0.00
Panel B: Combined interventions targeting common mental disorders						
Work aggregate	0.35***	0.10	0.59	11	6	0.76
Time in work	0.44***	0.11	0.76	1	1	1.00
Days unable to work[1]	0.17**	0.01	0.34	9	4	0.24
Functioning at work	0.74***	0.49	0.99	1	1	1.00
Panel C: Combined interventions targeting severe mental disorders						
Work aggregate	0.28***	0.07	0.49	22	10	0.73
In employment	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[1]	0.18***	0.06	0.30	8	3	0.00
Days unable to work[1]	0.11**	0.00	0.22	1	1	1.00
Functioning at work	0.20	-0.26	0.67	8	5	0.85


A: Psychosocial - CMD
B: Combined - CMD
C: Combined - SMD
Work aggregate
In employment*
Time in work*
Unable to work^
Days unable to work^*
Functioning at work*
Job search
-1
0
1
Method
Bayesian
Frequentist

Robustness and publication bias

Our work aggregate findings are broadly robust to:

- Controlling for intervention and study characteristics ▸ Details
  - We see evidence that effects are largest in the 6 months following programme exit
- Re-analysis using a Bayesian hierarchical model ▸ Details
- Successively excluding the three largest studies ▸ Details
- Looking at sub-samples with similar control conditions ▸ Details

Our study sample shows no evidence of publication bias

- A histogram of study effect sizes shows little evidence of bunching around significance thresholds ▸ Details
- We perform the Egger (1997) test for funnel plot assymetry and find no evidence of small sample effects $(p = 0.27)$ ▸ Details
- We implement the Andrews & Kasy (2019) model, finding no evidence of differential publication probabilities by effect size ▸ Details

Non-work-related outcomes: tentative evidence

Studies reported non-work-related economic outcomes sporadically, so it's difficult to present conclusive findings. No non-work economic outcomes were reported by combined interventions targeting CMD, and only 2 by those targeting SMD

However, we find some interesting indicative evidence that psychosocial interventions targeting common mental disorders increase an aggregate of non-work-related outcomes (SD=0.09, $p < 0.05$)

This appears to be driven by effects on:

- education (SD=0.21, $p < 0.05$)
  (e.g. effects on school attendance or expenditure on education), and
- subjective poverty (SD=0.16, $p < 0.1$).

What mechanisms account for economic effects of treatments?

1. Mental health treatments in this sample reduce mental health symptoms and improve "functioning"
   - Mental health symptoms: e.g. PHQ9: "How often have you been bothered by the following over the last two weeks: little interest in doing things, feeling down, poor appetite, feeling bad about yourself, trouble concentrating"
   - Functioning: e.g. BDQ: Difficulties in doing usual activities or in social interactions
2. Interventions with larger effects on these mechanisms also have larger effects on economic outcomes
3. Use individual-level data to quantify magnitude of the association between changes in mental health and economic outcomes. Among these studies we see midline effects on depression but not the exononic outcome, indicsting againt reverse causality

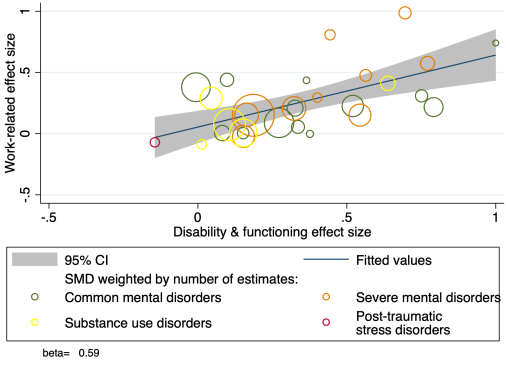
This sample of interventions reduces mental ill health symptoms

Functioning

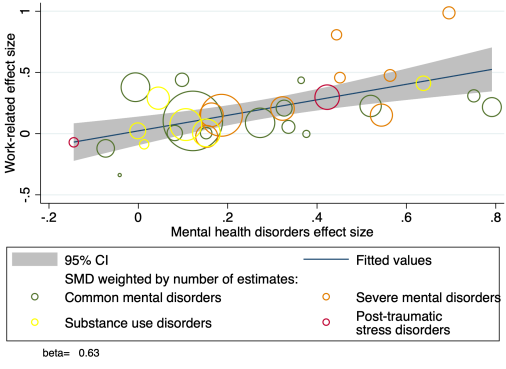
	estimate	95% CI lower	95% CI upper	# of effect sizes	# of interventions	I^2
Panel A: Psychosocial interventions targeting common mental disorders						
Mental health disorder symptoms	0.22***	0.11	0.34	48	9	0.65
Suicide attempts or at risk of suicide	0.08	-0.23	0.40	4	2	0.82
Relapse (dummy)	0.24***	0.08	0.41	7	3	0.58
Recovery (dummy)	0.23**	0.02	0.45	1	1	1.00
Diagnosed with mental disorder	0.31**	0.04	0.58	7	4	0.51
CMD symptoms	0.31***	0.12	0.49	21	8	0.81
Overall assessment of mental disorder	0.65***	0.22	1.07	2	1	1.00
Panel B: Combined interventions targeting common mental disorders						
Mental health disorder symptoms	0.27***	0.07	0.48	28	5	0.00
Suicide attempts or at risk of suicide	0.27**	0.05	0.49	8	4	0.00
Rehospitalisation	0.31***	0.08	0.54	1	1	1.00
CMD symptoms	0.24**	0.00	0.49	17	5	0.21
Panel C: Combined interventions targeting severe mental disorders						
Mental health disorder symptoms	0.43***	0.24	0.63	24	9	0.67
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	0.13	-0.08	0.34	5	3	0.36
SMD symptoms	0.44**	0.06	0.83	6	5	0.74

Correlations between mechanism effect sizes and work-related outcomes

(a) Larger effects on functioning are correlated with larger effects on work-related outcomes



(b) Larger effects on mental health are correlated with larger effects on work-related outcomes



At individual level, change in depression associated with treatment is associated with change in days worked

- Five studies which measure days able to work, treat using a psychosocial intervention and have publicly available data
- Pool individual observations, including study fixed effects, and clustering at original study cluster variable (robust to clustering at study-level)
- Treatment reduces depression by $\sim$0.265 SD ▸ First stage
- Treatment reduces days unable to work by 0.4 days in last 30 days ▸ Econ outcome
- 1 SD decrease in depression associated with reduction in days unable to work of 1.5 days ▸ IV results

Costs overview: interventions are cheap but with a lot of heterogeneity

	Mean	(10^{th} pct.)	(90^{th} pct.)
Intervention type- target condition combination			
Psych. + CMD	290.27	1.43	1226.41
Pharm.+psych. + CMD	104.98	104.98	104.98
Pharm.+psych. + SMD	316.18	55.69	570.96
Psych. + PTSD	1599.49	1456.28	1742.69
Psych. + SUD	149.34	35.14	370.37
Other interventions	233.69	227.95	239.43
Region			
East Asia & Pacific	288.69	55.69	570.96
Europe & Central Asia	1456.28	1456.28	1456.28
South Asia	94.21	12.79	180.09
Sub-Saharan Africa	949.36	370.37	1742.69

Notes: This table summarises the distribution of per participant intervention costs in our study sample in 2011 US-Dollars.

Conclusions

Systematic search: studies in LMIC that screen for a mental health condition, test a mental health intervention and measure an economic outcome until Aug 2022

1. Positive, significant effects of mental health treatments on work-related outcomes 0.3 SD impacts for combined psychosocial and pharmacological interventions, modest effects (SD=0.05) for psychosocial interventions targeting CMD
2. We see concurrent effects on mental health, functioning and economic outcomes, indicating that mental health and functioning are plausible mechanisms for economic outcomes
3. Intervention costs are low, making mental health interventions plausibly cost-effective to improve economic outcomes in populations with a high incidence of mental ill health
4. Measurement of standardised economic outcomes in well-powered mental health studies would enable clearer conclusions

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

Crick Lund[*], *Kate Orkin*[*,**], *Marc Witte*[*], *Thandi Davies*, **John Walker**[**], *Johannes Haushofer*, *Judy Bass*, *Paul Bolton*, *Sarah Murray*, *Laura Murray*, *Wietse Tol*, *Graham Thornicroft*, *Vikram Patel*

[*] Joint first author, [**]University of Oxford

April 14, 2023

Work-related economic outcomes

	(1) Number of observations	(2) Share of observations	(3) Number of interventions	(4) Share of interventions
Work-related outcomes	81	0.55	35	0.90
Employment dummy	9	0.06	5	0.13
Time in work	9	0.06	5	0.13
Unable to work	17	0.12	4	0.10
Days unable to work	28	0.19	14	0.36
Functioning at work	18	0.12	11	0.28

Notes: There are 39 interventions and 147 economic effect sizes. "Functioning at work" measures are qualitative measures of functioning on the job. For example, the IDEAS scale is a rating by intervention staff that evaluates a patient's disability in work on a 5 point scale. We split out any outcomes where outcomes are linked to illness e.g. WHODAS 2.0: disability "because of any health condition".

Other economic outcomes

	(1) Number of observations	(2) Share of observations	(3) Number of interventions	(4) Share of interventions
Other economic outcomes	66	0.45	12	0.31
Education	19	0.13	2	0.05
Assets	7	0.05	2	0.05
Income, consumption and input expenditure	15	0.10	5	0.13
Subjective poverty measures	18	0.12	5	0.13
Social networks	7	0.05	2	0.05

Notes: There are 39 interventions and 147 economic effect sizes. Income and expenditure are made up of income, consumption and input expenditure.

Mental health and functioning outcomes

	(1) Number of observations	(2) Share of observations	(3) Number of interventions	(4) Share of interventions
Suicide attempts or at risk of suicide	18	0.05	10	0.26
Relapse rate	7	0.02	5	0.13
Recovery rate	3	0.01	3	0.08
Rehospitalisation	6	0.02	4	0.10
Substance use	35	0.10	6	0.15
CMD symptoms	72	0.21	24	0.62
PTSD symptoms	13	0.04	6	0.15
SMD symptoms	4	0.01	4	0.10
Overall assessment of mental disorder	10	0.03	7	0.18
Disability	57	0.17	27	0.69
Functioning in social interactions	2	0.01	1	0.03
Social functioning	7	0.02	4	0.10
Self-esteem/self-efficacy	21	0.06	7	0.18
Cognition	18	0.05	4	0.10
Physical health	10	0.03	4	0.10

Notes: There are 39 interventions and 336 mental health effect sizes.

Functioning and physical health results

▸ MHeffects

	estimate	95% CI lower	95% CI upper	# of effect sizes	# of interventions	I^2
Panel A: Psychosocial interventions targeting common mental disorders						
All disability and functioning	0.26***	0.11	0.42	27	6	0.78
Overall measures of functioning	0.05**	0.00	0.10	12	5	0.88
Social support	0.25***	0.09	0.41	15	4	0.54
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: Combined interventions targeting common mental disorders						
All disability and functioning	0.33	-0.09	0.75	10	6	0.87
Overall measures of functioning	0.28	-0.22	0.77	9	5	0.89
Functioning in social interactions	0.60***	0.15	1.04	1	1	1.00
Panel C: Combined interventions targeting severe mental disorders						
All disability and functioning	0.46***	0.30	0.63	8	7	0.21
Overall measures of functioning	0.35***	0.20	0.50	6	5	0.00
Functioning in social interactions	0.66***	0.38	0.93	2	2	0.00

Detailed empirical strategy

Within group of outcomes:

1. Standardise outcomes
   - Reverse code items so "good" effects are coded in the same direction
   - Standardised mean difference (SMD) has a bias and overestimates the absolute value of the effect in small samples (Hedges, 1981)
   - Small sample adjustment (where n_t and n_c are numbers in treatment and control):

     Hedges' g: $g = SMD * \left(1 - \frac{3}{4(n_t+n_c)-9}\right)$
2. Aggregate across outcomes: Average over multiple effect sizes from one study and adjust standard error of estimate of the average (Borenstein et al. 2009)

Back

Aggregate across interventions

Inverse variance weighted random effects analysis to aggregate many estimates of a similar effect into a single effect

- Inverse variance weighted: more precise estimates get more weight
- Random effects
  - Must make an assumption about if studies are estimating the same underlying intervention effect (fixed effects) or estimating different intervention effects (random effects)
  - Random effects incorporates some heterogeneity between studies
    - Assumes that the different studies are estimating different, yet related, intervention effects
    - Assumes we don't know why studies differ i.e. differences are random and effects follow a normal distribution across studies
    - Adjust standard errors of study-specific estimates to incorporate measure of the extent of heterogeneity, among the intervention effects in different studies (Tau-squared)

Back

Effects are qualitatively similar to estimates in HIC

	Study	Countries	Their Effect Size	Our Effect Size
Aggregate				
Labour Supply	Timbie et al (2006)	USA	$d = 0.12$ [95% CI 0.00 to 0.24]	$g = 0.16$ [95% CI 0.10 to 0.23]
Days Unable to Work				
Sick Leave	Nieuwenhuijsen et al (2020)	Europe, USA Australia and Canada	-0.3 days in 2 weeks [95% CI -0.08 to -0.22]	-0.82 days in 30 days [95% CI -1.89 to 0.25]
Sick Leave[12]	Nieuwenhuijsen et al (2020)	Europe, USA Australia, Canada	SMD $= 0.15$ [95% CI 0.03 to 0.28]	$g = 0.04$ [95% CI 0.00 to 0.08]
Sick Days[12]	Nigatu et al (2016)	Netherlands, Scandanavia	SMD $= 0.14$ [95% CI 0.01 to 0.26]	$g = 0.04$ [95% CI 0.00 to 0.08]
Sick Leave[1]	Salomonsson et al. (2018)	Europe, USA, India	$g = 0.15$ [95% CI 0.08 to 0.22]	$g = 0.04$ [95% CI 0.00 to 0.08]
Functioning at Work				
Functioning (Pysch)	Kamenov et al (2017)	USA, UK, Netherlands	$g = 0.43$ [95% CI 0.33 to 0.54]	$g = 0.34$ [95% CI 0.13 to 0.55].
Functioning (Pharm)	Kamenov et al (2017)	USA, UK, Netherlands	$g = 0.31$ [95% CI 0.26 to 0.36]	$g = 0.34$ [95% CI 0.13 to 0.55]

Notes: g= Hedges' g; d = Cohen's d; and, SME = Standardised Mean Difference. Outcomes marked [1] have been reverse-coded, so that for all measures higher values can be interpreted as indicating better employment outcomes. [2] Nigatu et al (2006) and Nieuwenhuijsen et al (2020) did not reverse code Sick Days, so we have given the absolute value of their effect size for comparison.

At individual level, change in depression associated with treatment is associated with change in days worked

- Focus on studies of psychosocial interventions targeting depression, measuring days able to work
- Find individual level data where available

	Study name	Study country	Study region	Intervention
1	Fuhr et al. (2019)	India	Goa	CBT
2	Sikander et al. (2019)	Pakistan	Rawalpindi	CBT
3	Baranov et al. (2020)	Pakistan	Punjab	CBT
4	Barker et al. (2022)	Ghana	Northern and Middle Belt	CBT
5	Weobong et al. (2017)	India	Goa	

First stage: treatment correlated with depression

Back

	Combined measure			DSM-IV	PHQ-9	BDI	Kessler
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treatment	-0.265***	-0.257***	-0.164***	-0.662***	-0.391***	-0.606***	-0.186***
	(0.031)	(0.041)	(0.044)	(0.129)	(0.057)	(0.102)	(0.037)
Above median age		0.079***					
		(0.029)					
Treatment=1 × Above median age		-0.015					
		(0.054)					
Mild depr.			0.316***				
			(0.032)				
Moderate depr.			0.470***				
			(0.037)				
Severe depr.			0.668***				
			(0.045)				
Treatment=1 × Mild depr.			0.005				
			(0.065)				
Treatment=1 × Moderate depr.			-0.200***				
			(0.061)				
Treatment=1 × Severe depr.			-0.209***				
			(0.076)				
Constant	-0.007	-0.047	-0.302***	-0.000	0.002	0.000	-0.000
	(0.025)	(0.030)	(0.031)	(0.116)	(0.042)	(0.066)	(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
Obs.	8240	8240	8240	429	1094	447	6717
Studies	5	5	5	1	3	1	1

Individual-level: treatments increase days able to work

Back

	Combined days able to work measure			Healthy days	Days unable to work
	(1)	(2)	(3)	(4)	(5)
Treatment	0.407**	0.270	0.524**	0.288	-0.404**
	(0.171)	(0.206)	(0.236)	(0.661)	(0.185)
Above median age		-0.593***			
		(0.179)			
Treatment=1 × Above median age		0.286			
		(0.330)			
Mild depr.			-0.675***		
			(0.207)		
Moderate depr.			-1.011***		
			(0.225)		
Severe depr.			-1.569***		
			(0.260)		
Treatment=1 × Mild depr.			-0.505		
			(0.454)		
Treatment=1 × Moderate depr.			-0.326		
			(0.412)		
Treatment=1 × Severe depr.			0.375		
			(0.479)		
Constant	25.003***	25.305***	25.668***	26.155***	3.777***
	(0.100)	(0.124)	(0.131)	(0.476)	(0.112)
Study FE	Yes	Yes	Yes	Yes	Yes
Control mean	24.99	24.99	24.99	26.16	3.81
Obs.	8240	8240	8240	429	7811
Studies	5	5	5	1	4

Individual-level regressions of work-related outcomes on depression, instrumented by random treatment allocation

Back

	Combined days able to work measure		Healthy days		Days unable to work	
	(1) OLS	(2) IV	(3) OLS	(4) IV	(5) OLS	(6) IV
Depression reduction	1.588***	1.534**	0.0762	0.434	-1.854***	-1.747**
	(0.0856)	(0.645)	(0.363)	(1.041)	(0.0887)	(0.784)
Constant	24.99***		26.27***		3.791***	
	(0.0816)		(0.362)		(0.0884)	
Study FE	Yes	Yes	Yes	Yes	Yes	Yes
Control mean	24.99	24.99	26.16	26.16	3.81	3.81
Obs.	8240	8240	429	429	7811	7811
Studies	5	5	1	1	4	4
Underidentification		0.00		0.00		0.00
Weak identification		71.37		26.41		53.16

Study sample characteristics by control condition

Back

	(1) Number of interventions	(2) Share of interventions
All interventions	39	1.00
Control condition		
Enhanced Usual Care	7	0.18
No Treatment	19	0.49
Treatment As Usual (Pharmacological)	13	0.33

Notes: There are 30 interventions, 115 economic effect sizes and 256 mental health effect sizes. Categories are mutually exclusive.

Robustness checks (Combined-CMD) Back

	Dep. var.: work-related outcomes (Hedges' g)						
Constant (τ estimate)	0.28	0.30	0.29	0.45	0.30	0.29	0.29
Standard deviation of $\hat{\tau}$	(0.11)	(0.11)	(0.11)	(0.33)	(0.12)	(0.11)	(0.11)
Measurement heterogeneity							
Variance of error term	No	Yes	No	No	Yes	No	No
Control conditions	No	No	Yes	No	Yes	No	No
Measurement timing	No	No	No	Yes	No	No	No
Intervention and context heterogeneity							
Implementing party	No	No	No	No	No	Yes	No
Sample characteristics	No	No	No	No	No	No	Yes
Residual heterogeneity (p-value)	0.79	0.73	0.87	0.74	0.81	0.87	0.84
Modifier relevance (p-value)	NA	0.72	0.19	0.58	0.41	0.19	0.36
Degrees of freedom	10	9	9	9	8	9	9
Number of interventions	6	6	6	6	6	6	6

Robustness checks (Combined-SMD)

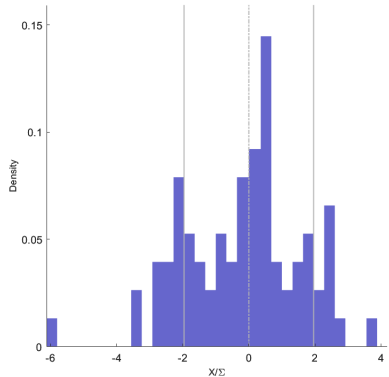
Back

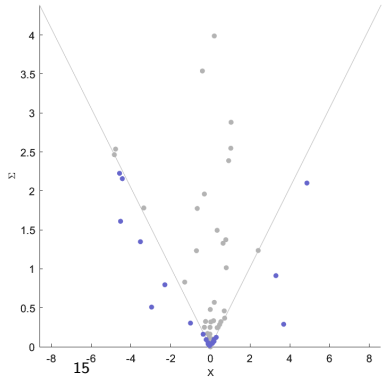
	Dep. var.: work-related outcomes (Hedges' g)						
Constant (τ estimate)	0.27	0.30	0.28	0.27	0.30	0.30	0.30
Standard deviation of $\hat{\tau}$	(0.10)	(0.09)	(0.13)	(0.12)	(0.14)	(0.13)	(0.12)
Measurement heterogeneity							
Variance of error term	No	Yes	No	No	Yes	No	No
Control conditions	No	No	Yes	No	Yes	No	No
Measurement timing	No	No	No	Yes	No	No	No
Intervention and context heterogeneity							
Implementing party	No	No	No	No	No	Yes	No
Sample characteristics	No	No	No	No	No	No	Yes
Residual heterogeneity (p-value)	0.80	0.82	0.71	0.75	0.72	0.76	0.74
Modifier relevance (p-value)	NA	0.27	0.91	0.80	0.88	0.73	0.72
Degrees of freedom	20	19	19	19	18	19	19
Number of interventions	10	10	10	10	10	10	10

Limited graphical evidence of publication bias in our sample

Back

- A histogram of standardised effect sizes shows only limited bunching at key thresholds (e.g. $Z = -1.96$) indicating against substantial publication bias
- A funnel plot shows little evidence of asymmetry, indicating against small-study effects. The Egger (1997) test supports this: $(p = 0.27)$





Modelling publication bias to allow for bias corrections

Back

We follow Andrews & Kasy (2019), modelling differential probability of publication conditional on reported standardised effect size $P(publication|Z = z)$ using MLE

- Under no small sample effects and no selectivity, we model the distribution of reported high variance effect sizes as that for low variance effects plus noise
- We assume that all effects with $Z \geq 1.96$ are published, i.e. $P(pub|Z \geq 1.96 = 1)$ and that the distribution of latent (true) treatment effects follows a t-distribution
- Finding: the relative probabilities are close to one with moderately large standard errors. We have some evidence against large publication bias, and if anything, point estimates indicate relatively higher publication of insignificant effects

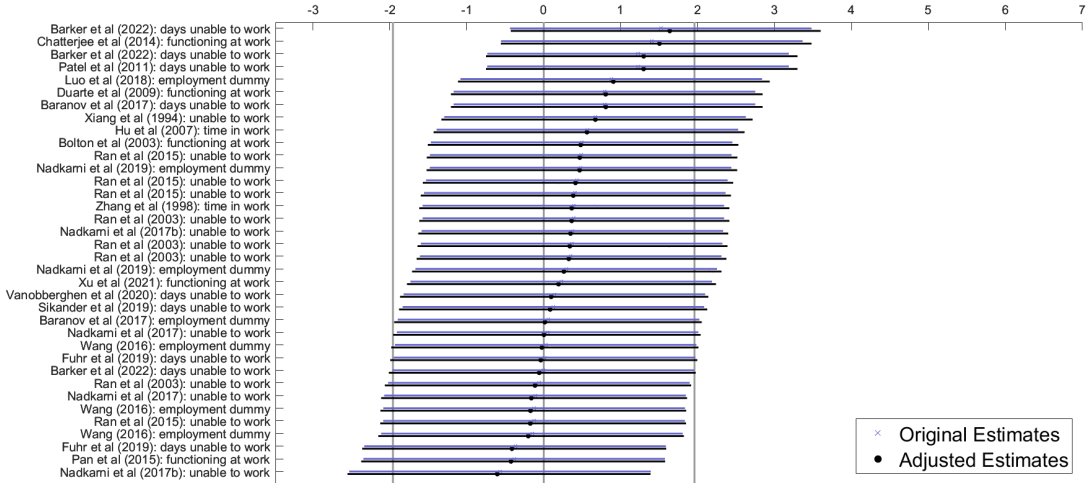
Relative publication probabilities

	$Z < -1.96$	$Z \in [-1.96, 0)$	$Z \in [0, 1.96)$	$Z \geq 1.96$
P(Z)	1.18	1.23	1.14	1
SE(P)	(0.30)	(0.57)	(0.52)	.

Point-wise adjusted estimates of work-related outcomes

Back

There is no evidence of publication bias, so adjusting for differential publication probabilities has little impact



Bayesian hierarchical analysis

Back

We estimate the Rubin (1981) model via the Bayesian hierarchical approach popularised by Meager (2019)

- This allows us to simultaneously identify between-study heterogeneity and study-level shrinkage estimates
- We expect little power improvement from the Bayesian hierarchical model because we have a large study sample and expect significant heterogeneity between sites
- We retrieve the following estimates of the average effect τ and heterogeneity parameter σ, finding an essentially unchanged mean posterior τ
- Results are robust to the choice of prior, and our preferred specification is $\tau \sim N(0,1), \sigma \sim HalfCauchy(1)$

	All economic outcomes	Employment	Non-employment
τ [CI]	0.17 [0.10, 0.25]	0.17 [0.08, 0.26]	0.09 [-0.03, 0.23]
σ [CI]	0.17 [0.10, 0.24]	0.21 [0.13, 0.3]	0.11 [0, 0.43]