January 2024

Mental Health Report

Primary Author: Stanley Pinsent
Supported By: Joel Tan

Centre for Exploratory Altruism Research

This research project was commissioned by Flourishing Minds Fund. It consisted of a review of the available literature on mental health and consultation with over a dozen experts in the field.

Key Findings

- ***Cost* is generally the most important variable** in assessing mental health interventions, and it can vary widely
- **Interpersonal group therapy** is a promising way of improving mental health provision in poor countries, although the scale of the effect of psychotherapy is unclear. Technology may allow psychotherapy to be delivered more cheaply and at scale.
- **Tackling risk factors** can *prevent* mental illness. Policy advocacy could be a good lever for doing this.
- **Suicide prevention** appears to be highly neglected in poor countries.
- Mental illness will increase in relative importance over time, so we should **place a premium on any work that increases our understanding** of how to mitigate it

Structure

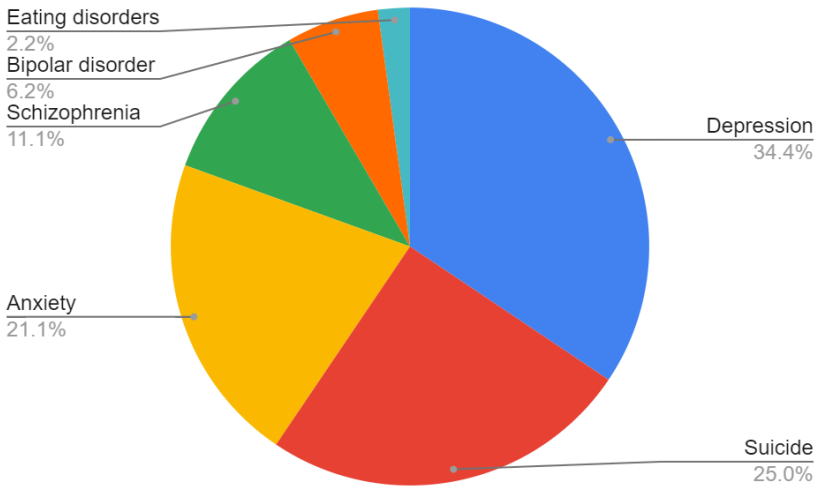
Executive Summary

Burden

Data on the burden of mental health is unreliable, especially in low-middle income countries (LMICs). The IHME's Global Burden of Disease (GBD) study simply has *no data* from many poor countries. When they do have data from a low-income country, it is more likely to come from small, unrepresentative samples and less likely to come from expert diagnosis.

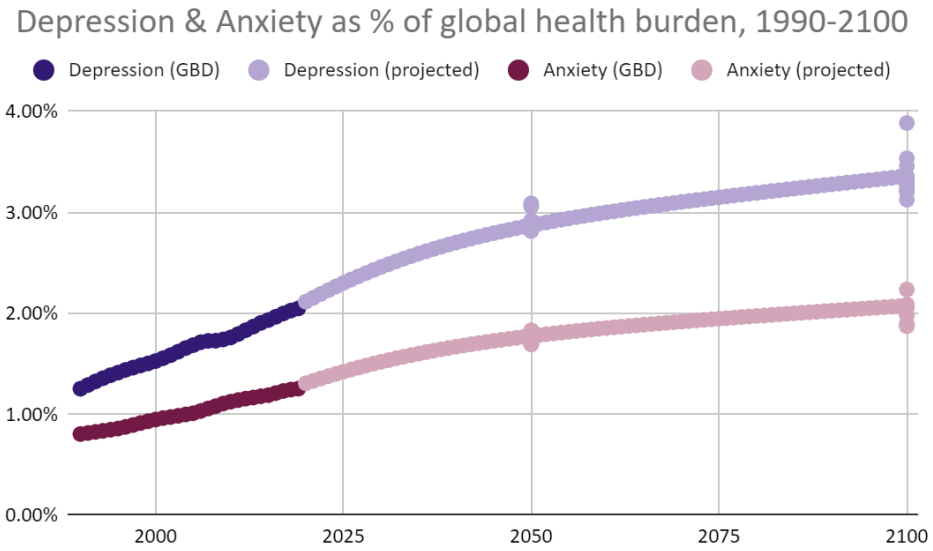
However, we can be confident that mental illness is a big problem. Depending on how "mental illness" is defined, it represents between 3% and 16% of the total disease burden. This report focuses on the conditions in the table below, which account for 5.4% of the global disease burden.

Condition	**Burden** **DALYs per 100,000 (GBD, 2019)**	**Prevalence** **per 100,000 (GBD, 2019)**
Depression	606	3,614
Suicide	441	10
Anxiety	371	3,895
Schizophrenia	195	305
Bipolar disorder	110	511
Eating disorders	38	176



Relative burden attributable to six key conditions (GBD, 2019)

If rapid progress against communicable diseases and malnutrition continues, we can expect mental illness to rise up the agenda. We project that depression and anxiety, the most prevalent mental disorders, will increase 50% in relative importance by the end of the century.



The extra points at 2050 and 2100 are the outputs of various linear and exponential models based on 1990-2019 data.

How bad is mental illness?

DALYs, QALYs and WELLBYs are among the units most often used to measure health. But they may not be well-suited for capturing mental illness.

DALY and QALY weightings come from the assessments of the general public, many of whom have never experienced the health state in question. We suspect that because of its "invisibility", this leads to a relative *underweighting* of mental illness. Our analysis of how *sufferers of depression* rate depression suggests that the condition is around 20% more severe than IHME's DALY weighting.

The WELLBY is derived from individuals' assessments of their own lives. This *may* allow us to capture important effects beyond health or quality of life. But it also introduces the potential for response bias that clouds results. The same goes for other survey methods that gauge mental health by checking for symptoms.

Most measures, including the DALY, QALY and WELLBY, are bounded and therefore cannot capture extreme positive and negative states of being. If we believe that extreme suffering is possible and widespread, we probably ought to weigh severe pain, depression and psychosis more heavily than these measures allow.

Comparing mental illness to other illnesses is inherently difficult. We think that *converting between measures* should be avoided when possible, because of the highly contested assumptions required. Overall, decision-makers should be aware of the limitations of the measures they are using, and should be clear about any philosophical assumptions they are making.

Treatment interventions

In general, we find that the most important consideration for treatment interventions is the *cost per person treated*. For this reason, and because of the lack of existing mental health provision, LMICs are usually the best target. The weakness of existing mental health infrastructure in these regions means that organizations should be prepared to overcome major logistical challenges.

Task-shifted interpersonal group therapy is a proven way of scaling up mental health provision in LMICs. However, a controversy over StrongMinds exposed key weaknesses in the literature on psychotherapy and its sensitivity to the methods used to account for bias. Furthermore, estimates for the scale of household spillover effects and the persistence of benefits, which are crucial to the overall effect, rely on few data points. The internet, smartphones and AI offer the potential to deliver psychotherapy more cheaply and to scale up more easily, although this field is in its early stages.

There are huge gaps in access to antidepressants and pain medication in poor countries. Improving access could be highly effective at reducing suffering, but we suspect that it would be a greater challenge than simply providing medicines. One option is to simply improve health workers' pain-management by providing targeted training.

As mental health in LMICs increases in importance, we should aim to fail fast and improve quickly. We need better evidence on the full effects of interventions over time and across the community. Ambitious programs that garner useful feedback should be valued highly.

Prevention Interventions

This report devotes more space to *prevention* interventions than *treatment* interventions, but this should not be seen as an indication that prevention interventions are more promising overall.

We investigate a number of risk factors associated with a higher risk of mental illness. Overall we are surprised at the widespread acceptance of simple *correlational* evidence and the lack of evidence for *causation*. Genetic susceptibility and poverty seem likely to explain some of the correlation observed between, say, alcohol abuse and suicide, but studies often do a poor job at isolating these effects.

Interventions targeting risk factors typically have only a small effect on mental health. Cash transfers appear to have a very modest effect on mental health for the spending required. The effect of contraceptive access on averted perinatal depression is likely to be small. Extremely potent risk factors like intimate partner violence may be an exception to this rule.

When the effect size is small, only very cheap and scalable interventions show promise. Health policy "nudges" like sugar taxes or promoting physical activity may operate efficiently enough to avert a year of depression for tens or hundreds of dollars.

We were surprised by the degree to which many suicides appear to be preventable, especially in societies where suicide is taboo. Advocating for bans on deadly means of

suicide could have an outsize impact, even though major progress has already been made on this front. It's even possible that supplementing drinking water with small amounts of lithium could slash the suicide rate.

Much of our work on preventing mental illness relies on speculation. More research is needed on the counterfactual impact of targeting mental health risk factors.

Intervention BOTECs

Type	Intervention	Cost-effectiveness compared to cash transfers[1]	Proportion of benefits from mental health
Psychotherapy	Interpersonal group therapy, LMIC	5×	100%
	Internet-delivered psychotherapy	10×	100%
Social support	Child poverty grant advocacy	30×	3%
Mass media	Radio campaign to combat Intimate Partner Violence (IPV)	5×	30%
Health nudges	Physical activity campaign advocacy	17×	17%
	Sugar-sweetened beverage tax advocacy	79×	4%
Suicide prevention	Pesticide ban advocacy	7×	100%
	Lithium supplementation study	38×	100%

We have constructed BOTECs for the cost-effectiveness of various mental health calculations. *Most of the inputs are subjective estimates*, and all results should be interpreted with caution. We find that there is generally more uncertainty around *policy* interventions due to the unknown probability of successfully influencing policy and the unknown scale of the effects. Psychotherapy is a relatively "safe" option, but it is likely to fall short of the best interventions for tackling malaria and other highly preventable diseases.

We hope the calculations can help users determine the *most important unknowns* in an intervention's effectiveness. With more context-specific information, the model can be adapted to give more accurate estimates.

Conclusion

Mental illness is a huge problem. We are still refining ways to cost-effectively treat it, and the field of prevention is in its infancy. This means that even the best-understood mental health treatments such as psychotherapy are riskier than the most robust *physical* health interventions. Mental health work will appeal most to people who prioritize *reducing suffering* over *extending life*. Due to the state of the field, we suggest that a premium be placed on

[1] Based on CEARCH's analysis of GiveWell-recommended charities, which estimates that donations to GiveDirectly avert 37 DALYs per $100,000. This implies that the average GiveWell top charity is approximately 20× cash.

interventions that improve our understanding by addressing key questions and filling important gaps in the scientific literature.

Burden

Key Findings

- Mental illness forms 3% to 16% of the global disease burden, depending on how mental health is categorized
- The data on mental health prevalence is unreliable, especially in LMICs. But mental illness appears to be a major problem everywhere in the world. We suspect that geographical variation in mental illness is usually not one of the top considerations when deciding where to implement an intervention.
- As we continue to make progress against communicable diseases and poverty, mental illness will become a larger proportion of the total disease burden over time. We estimate that the relative importance of depression and anxiety will increase by 50% by the end of the century
- The nature of the burden will change over time. Conditions affecting the elderly will become a bigger problem, and we may see an increase in mental illness among adolescents.
- We don't know much about how to address mental illness in resource-poor societies. Further research is likely to pay dividends in the coming decades.

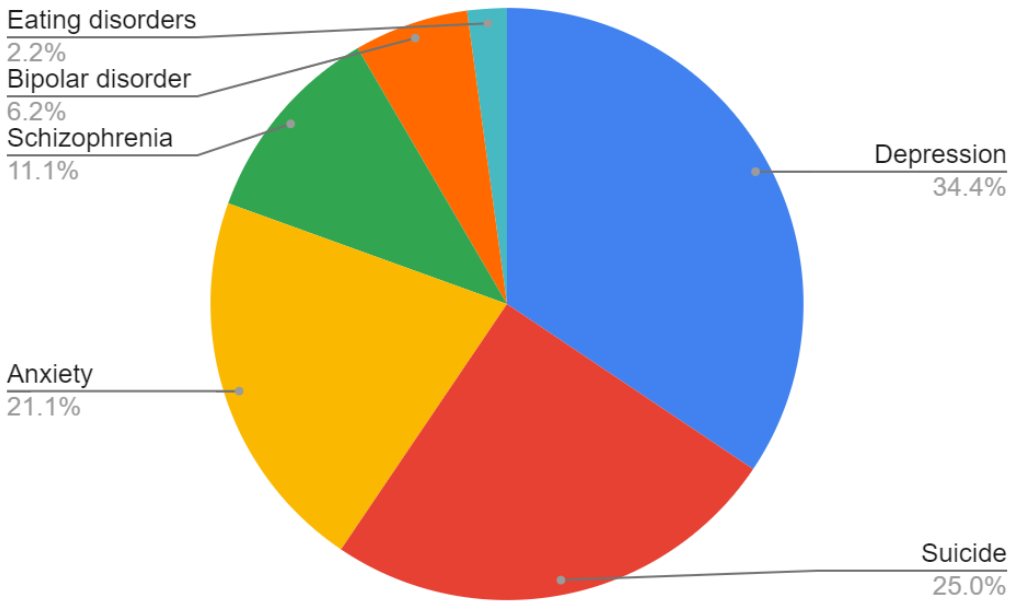
What counts as mental illness?

Mental illness can be categorized in a number of ways. In general this report focuses on mental health conditions *that are less likely to fall under the remit of typical "Global Health and Development" work*. Hence we exclude conditions that are tightly linked with infectious disease, exposure to toxic chemicals, and malnutrition (such as IDID, which IHME class as a mental disorder). We also choose to exclude neurodevelopmental and childhood-behavioral disorders such as Autism and Conduct Disorder on the basis that they are slippery conditions which are very poorly understood. We focus more heavily on conditions with a lower degree of heritability.

We include *suicide* on the assumption that suicide and self-harm is primarily caused by mental illness, unhappiness and despair.

Most of this report focuses on the conditions in the table below (all figures extracted from GBD, 2019).

Condition	Burden DALYs per 100,000 (% of total in this table)	Prevalence per 100,000	Average Severity DALY burden / Prevalence	Heritability[2]
Depression	606 (34%)	3,614	0.17	Moderate
Suicide[3]	441 (25%)	10	44.96	Moderate
Anxiety	371 (21%)	3,895	0.10	Moderate
Schizophrenia	195 (11%)	305	0.64	High
Bipolar disorder	110 (6%)	511	0.22	High
Eating disorders	38 (2%)	176	0.21	Moderate



[2] Heritability is a measure of how strongly the prevalence of the condition is explained by genetic factors. Broadly, less-heritable illnesses are better targets for preventative interventions, and may be more easily "cured" by treatment interventions. This is largely why this report does not explore interventions aimed at Schizophrenia and Bipolar disorder.
[3] Suicide is not classed as a mental disorder in the GBD study. I have used the GBD figures for "self-harm", the burden of which is almost all attributable to premature death.

Scale of the Problem

Summary:

- The burden of mental illness is huge
- The economic effects are also huge, but poorly understood
- Healthcare spending does not match the size of the problem

Mental illnesses impose an enormous health and economic burden worldwide. The conditions in the table above account for 5.4% of the total DALY burden of all causes of death or injury in IHME's Global Burden of Disease (GBD) study[4], which is comparable to the combined burden of HIV/AIDs, Malaria and Tuberculosis. A study by Arias et al. (2022) cast the net wider by including substance abuse, neurological and neurodevelopmental disorders, and by considering the increase in premature mortality and chronic pain. Together this accounted for 16% of the total DALY burden.

The economic impact is also enormous. Mental illness can impact education, employment, absenteeism and productivity. It imposes healthcare costs, too. Since mental illness is more likely than most conditions to affect people of working age, and much mental illness goes untreated, we suspect that the economic costs of stem more from lost productivity than healthcare.

Beck et al. (2011) identify a strong link between depressive symptoms and productivity, and a 2011 report by the World Economic Forum identifies mental illness as the biggest source of lost productivity due to disability. The report estimates that global economic loss due to mental illness was $2.5 trillion in 2010 (3.8% of global GDP), and projects that by 2030 the economic toll of mental illness will be "greater than that of cancer, diabetes, and respiratory ailments combined".

However, healthcare spending does not correspond to the size of the problem. "High-income countries allocate about 5.4% of their total government health expenditure to mental health, compared to 1.7% in LMICs — and only 0.02% in the nine low-income countries covered by the WHO mental health Atlas" (WHO, 2021, quote from Banerjee et al., 2022). India allocated just 0.8% of its 2022 healthcare budget to mental health, representing about USD 0.06 per citizen. This means that mental healthcare costs are often borne out-of-pocket or by private insurance, and that those who cannot afford it are simply left untreated.

Mental health conditions may be seen (with some justification) as more intractable than key physical conditions, especially in poor countries where preventable infectious diseases are common. This may partly explain the gap between funding and burden. It also underlines the urgent need to find *cost-effective* ways of tackling mental illness so that philanthropic, private and public funding can be better put to use.

[4] All GBD data in this report are 2019 figures unless stated otherwise, and are extracted from vizhub.healthdata.org/gbd-results/.

The burden is uncertain, and may be underrated

The global burden may be underrated for a number of reasons.

- Mental illness is often less visible than physical illness, which makes underdiagnosis and misdiagnosis more likely
- Medical mental health provision, and the quality of diagnosis and record-keeping, varies widely between countries. Statistical methods are used to estimate the burden in regions with scarce data, which could lead to large errors (see *Quality of GBD study data* below)
- The QALY and DALY weightings of health conditions are derived from the judgments of ordinary people, who may systematically underrate mental illnesses. Bounded measures such as QALYs and DALYs may fail to capture states of extreme suffering associated with severe mental illness (see *How bad is mental illness?*)

Geographical Distribution

Summary:

- The burden of mental illnesses appears to be high everywhere, with relatively modest differences between countries
- Overall we suspect that geographical distribution should not be the main concern when determining *where* to do an intervention. The ease and cost of operating are probably much more important concerns.
- Data from the GBD study and the WHO should be seen as unreliable, especially data on LMICs

Results from the Global Burden of Disease (GBD) study suggest that the total burden of mental disorders does not vary widely between countries. The vast majority of countries have an age-standardized burden of between 1250 and 2500 DALYs per 100,000.

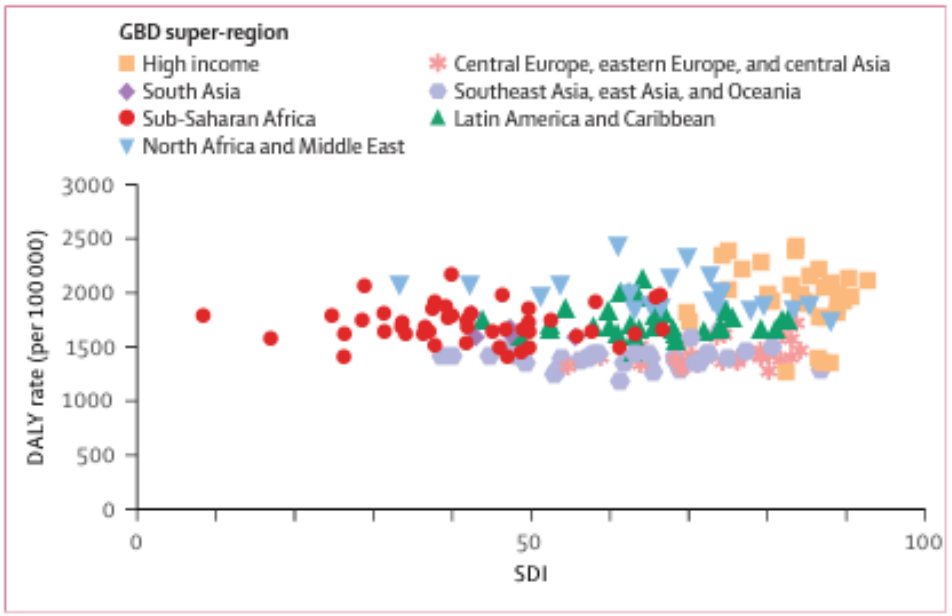
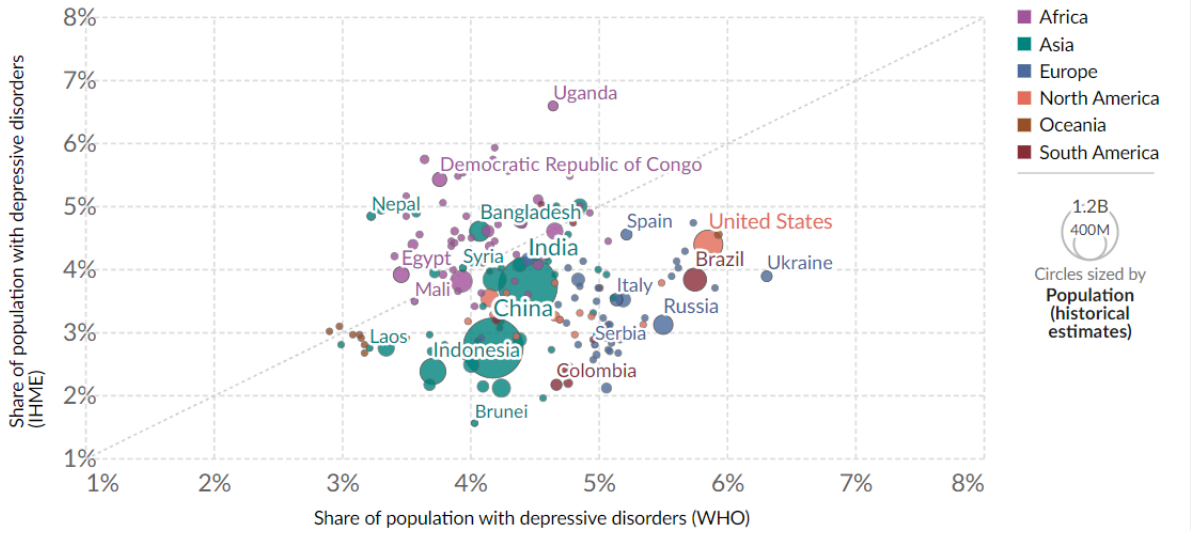


Figure 3: Age-standardised DALY rates for each location by SDI, both sexes combined, 2019

The chart above displays the burden from mental disorders in the GDB study, with each country represented by a dot. Note that this report uses a different categorization of mental illness.

The chart above suggests that the burden is lower than average is South Asia, and higher than average in most HICs. However, the extrapolation methods used in the GBD - demographically similar countries are assumed to have similar disease profiles - probably exaggerates similarities within categories. In fact, we think the unreliability of the results is sufficient to potentially negate or even reverse the trends we see in the data (see Data quality).

As an indication of the unreliability of the data, we can compare the GBD data against that of the WHO's Global Health Observatory (see chart below). The two sources disagree on the prevalence of depressive disorders (probably the most-studied mental disorder), with several estimates diverging by more than a factor of two. The GBD estimates higher depression rates than the WHO does in Africa, and lower rates in Asia and HICs (OWID, 2022).



Taking the GBD results at face value, we find that anxiety has moderate *positive* correlation with log(GDP per capita) (OWID, 2019a), while depression has moderate *negative* correlation with log(GDP per capita) (OWID, 2019b). In other words, weather countries have *more* anxiety and *less* depression than poor countries. However, the scale of the effect of wealth is modest in both cases: the expected difference in anxiety/depression between the wealthiest and poorest countries is no more than 2×.

Several experts told us that *finding people in need of mental health treatment* is not a major challenge in LMICs. With a good screening process, it should be possible to target sufferers of depression and other mental illnesses in any region. Other considerations such as cost, quality of existing healthcare, infrastructure and availability of labor are likely to be far more important than local disease prevalence.

There are some exceptions to this rule:

- Population-level interventions such as policy advocacy. These primarily have fixed costs, and are more promising when the number of potential beneficiaries is high

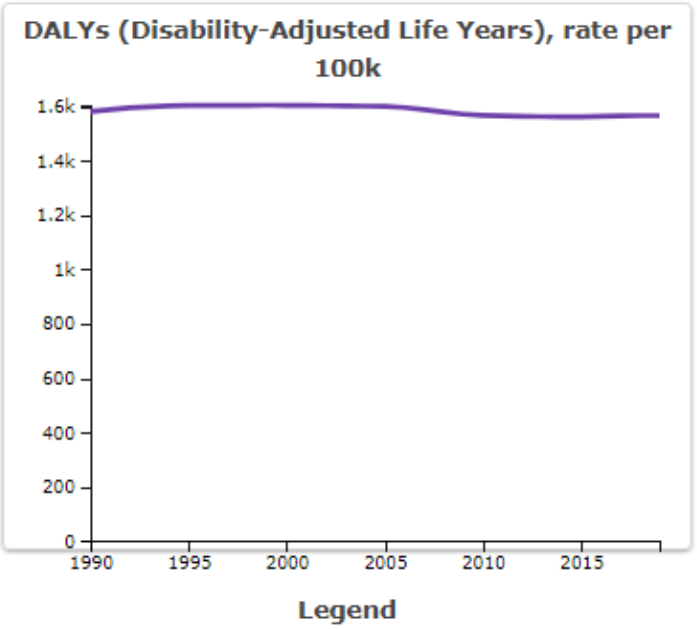
- Instances when the mental disorder is likely to be highly localized, such as PTSD in the aftermath of a civil war.

Change over time

Summary

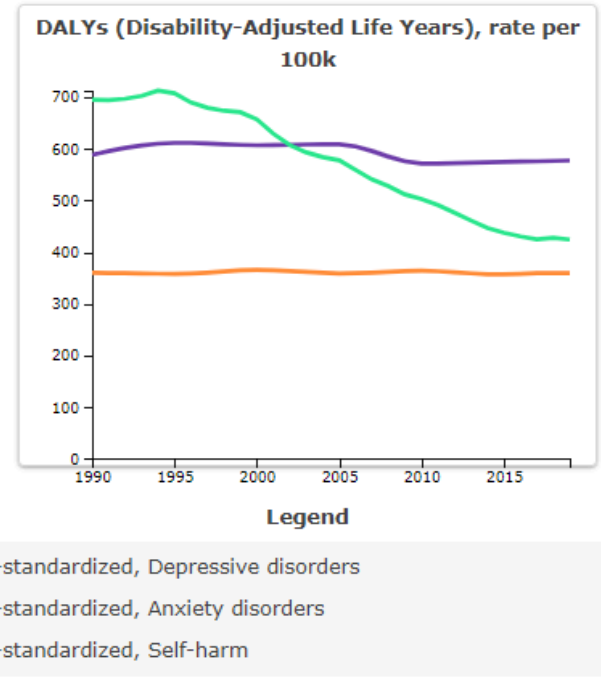
- There are signs that the per-capita mental health burden is increasing
- As we make progress in other areas of global health, mental illness' share of the total burden will grow
- We estimate that the relative importance of depression and anxiety will grow 50% by the end of the century
- Mental illnesses associated with the elderly will increase in importance as populations age
- It is possible that adolescent mental health will deteriorate as countries develop

The GBD study suggests that the burden of mental illness remained stable between 1990 and 2019. The graph below shows the age-adjusted global burden of mental disorders.



This result should be taken with a heavy pinch of salt. We can see from the chart below that the prevalence of suicide[5] (not classified as a mental disorder in the GBD) varied far more than anxiety or depression, which are harder to diagnose. We strongly suspect the apparent stability of the mental health burden reflects modeling choices rather than underlying trends.

[5] The GBD measures "self-harm", not suicide. Since the vast majority of the self-harm burden comes from deaths, we treat the GBD's self-harm figures as equivalent to suicide in this report.



We suspect that the GBD is doing a poor job of detecting changes in the mental health burden over time. However we are more confident in their assessments of *other* health conditions, and this data suggests that mental health will increase in relative importance over time.

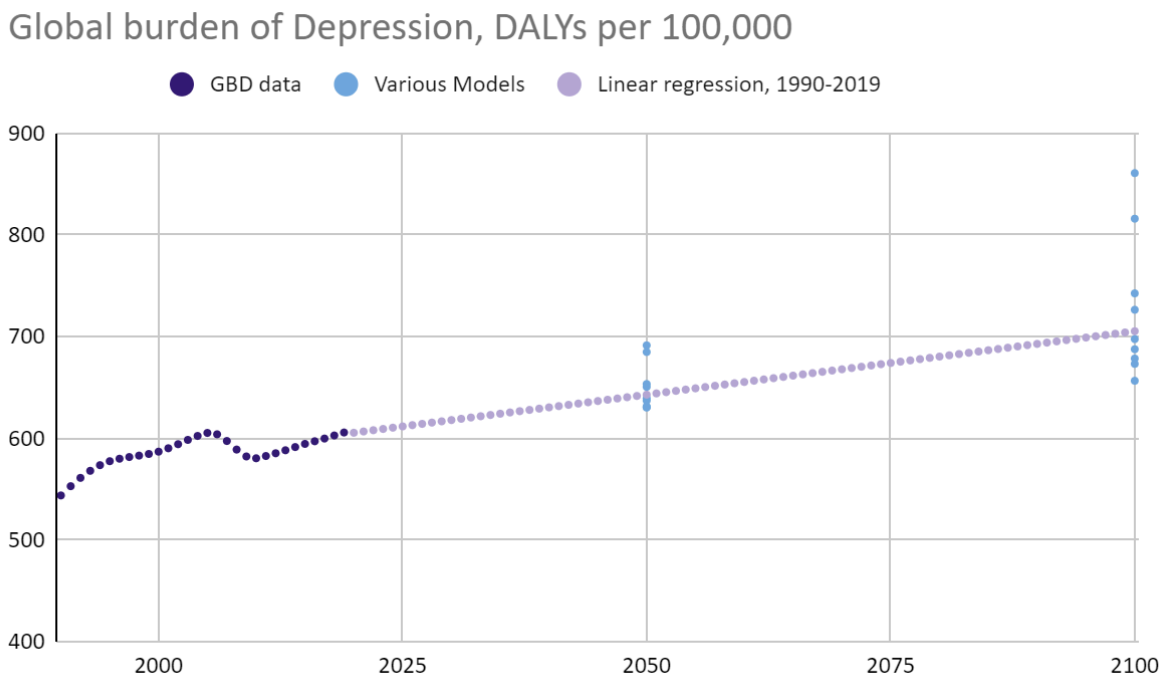
Depression and Anxiety will probably become more important

Summary

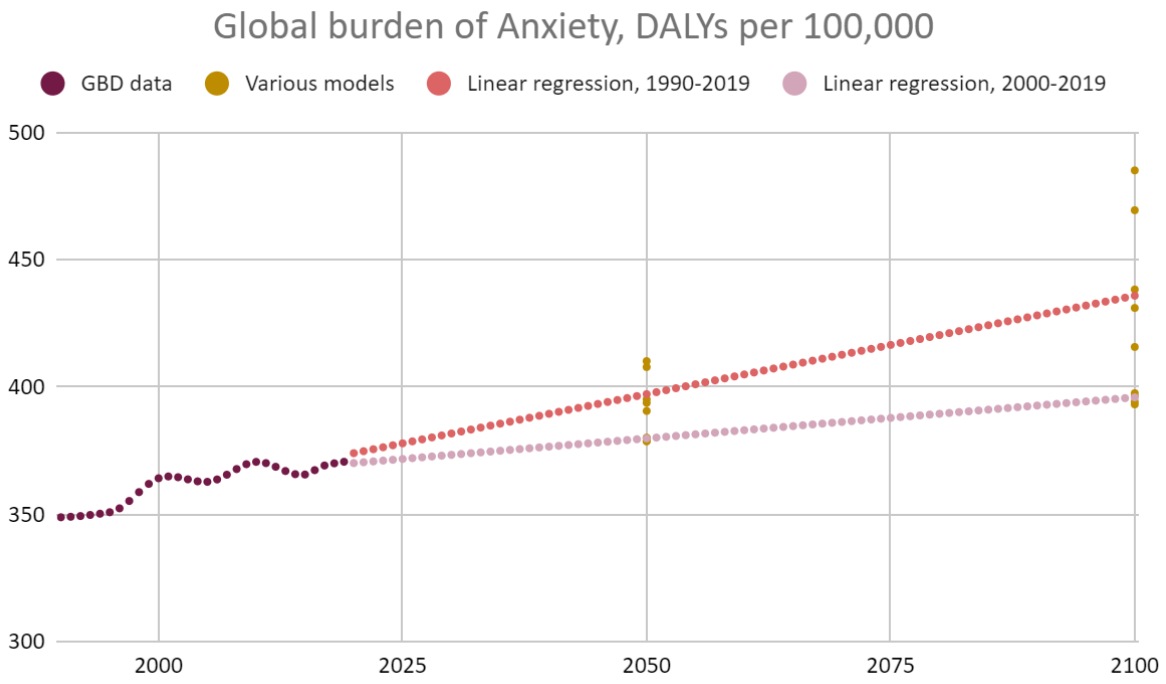
- The poor quality of available data makes future projection *very uncertain*
- As we make progress against poverty and communicable diseases, mental health will increase in importance
- Depression and anxiety alone may account for 5% of the global disease burden by the end of the century

According to the GBD survey, the period 1990-2019 saw depression rise, then fall, then rise again. Given the sparsity of data from many countries, we are very skeptical that this data represents real fluctuation (see Data quality below). It is more likely the result of tweaks and improvements in the way GBD handles data, and of the increase in data coming in.

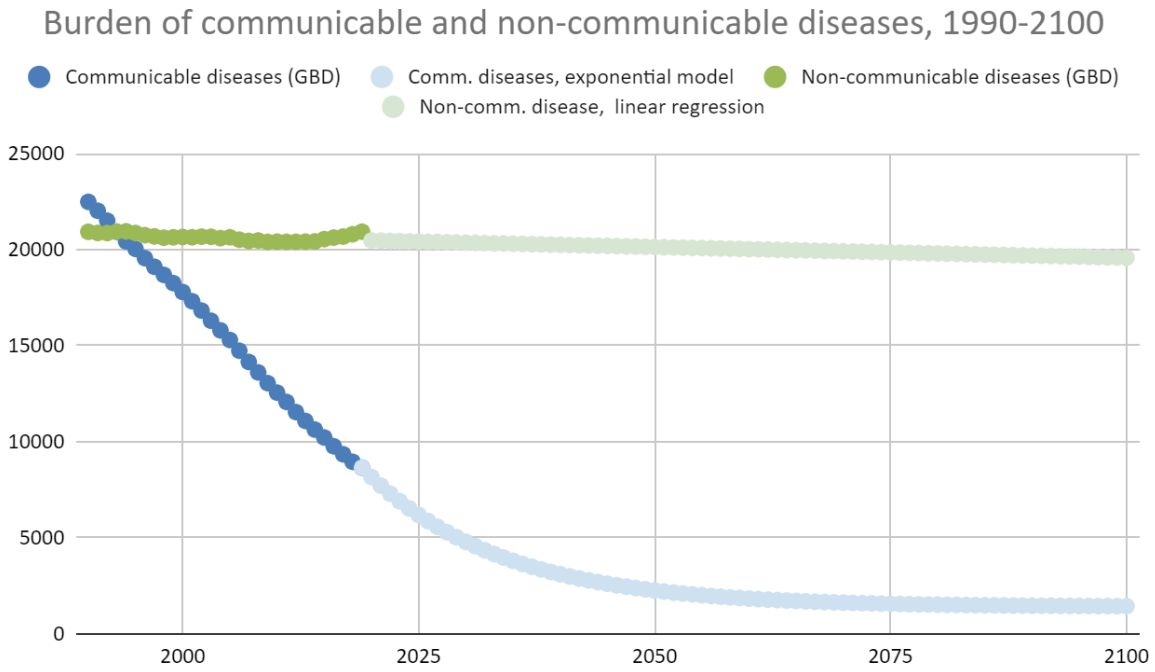
Nevertheless, we can attempt to extrapolate into the future. A linear regression model suggests that the burden of depression will increase from 600 to 700 DALYs per 100,000 by the end of the century. The blue dots in the graph below are products from various linear and exponential models based on the 1990-2019 data.



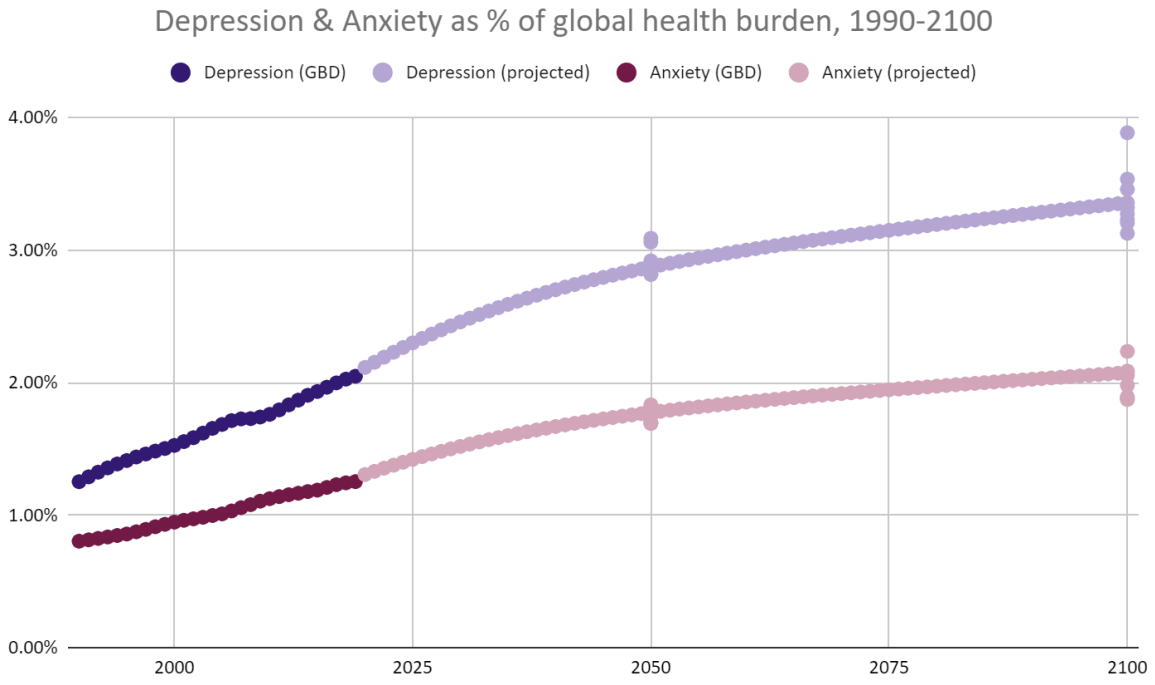
Applying a similar method to the anxiety data, we predict that the burden will increase from 370 to over 400 DALYs per 100,000 by 2100. The pink data points represent a linear regression based on the 2000-2019 data only.



While the projected increases in the burdens are modest, they should be understood in the context of the decreasing global disease burden. The burden from communicable, maternal, neonatal, and nutritional diseases plummeted in the period 1990-2019, while the burden from non-communicable diseases remained stable.



By projecting the burden of communicable and non-communicable diseases into the future[6], we can estimate what *proportion* of the disease burden will be attributable to depression and anxiety.



[6] We use a linear regression model with the non-communicable disease data. This is not possible with the communicable disease data, as it would soon lead to impossible negative values. Instead, we assume a "constant + exponential" model, where the constant part is 1400 DALYs per 100,000, which is the rate in modern-day high SDI countries, according to the GBD survey.

We find that depression is projected to go from 2% to over 3% of the total disease burden by 2100, while anxiety will go from 1.3% to approximately 2%. Thus **we expect the relative importance of depression and anxiety to increase by about 50%.**

The raw data from the graphs in this section can be found here.

This phenomenon - an increase in the relative importance of mental illness as we make progress against communicable diseases - may be even more pronounced in poor countries, where the communicable disease burden is high and (usually) falling fast. A case can be made for funding research that increases our ability to tackle mental illness in the future, *even if the interventions in question are not yet cost-effective*. As explored in *Treatment*, we don't know how best to tackle mental illness in poor countries. Increasing the evidence base will allow us to work more effectively when mental illness moves up the agenda.

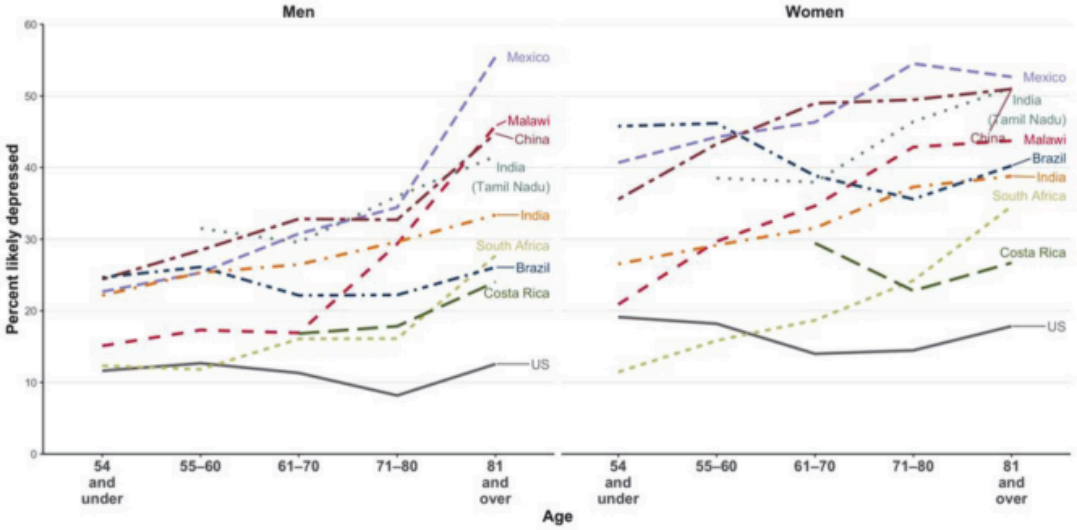
One notable exception may be suicide. The suicide rate dropped by 29% between 1990 and 2019. Rapidly industrializing countries that once had above-average suicide rates have seen significant drops, as in China (53% drop) and India (22% drop). This is probably partly due to stricter controls on the most toxic pesticides. Although many countries still have insufficient restrictions, most of the available gains may already have been realized (see *Prevention*).

Conditions affecting the elderly will become more important

We can expect the nature of the mental illness burden to evolve as the population ages. Against popular perception, there are signs that loneliness and depression among the elderly is more prevalent in LMICs than in HICs. Banerjee et al, (2022) compared results from a group of surveys in seven LMICs and the US. They found that over-55s in the LMICs were much more likely to be depressed than their US counterparts, although the survey types and diagnosis thresholds differed between countries, making comparison fraught[7]. What we *can* say with more certainty is that depression appears to become more prevalent above the age of 70 in LMICs, but not in the US.

[7] The study drew from the 2016 wave of the US Health and Retirement Study (HRS) and conducted "HRS-style" surveys in the LMICs. Cutoff points differed by country based on what has been found to maximize sensitivity and specificity in different contexts.We suspect this partly explains why prevalence of depression seems to be far higher that the GBD estimate (3.2% across the global adult population).

Figure 1: Prevalence of depression symptoms by gender, age, and country



In all countries, functional impairment, loneliness and living alone were strongly correlated with depression. Higher prevalence of these indicators seems to account for some of the depression gap between the US and the LMICs, but probably not all of it. It seems plausible that poverty, untreated pain and lack of access to transport and communication could also be causal factors.

Old social norms of caring for elderly neighbors and relatives are breaking down at a time when the elderly population is steadily growing. It seems plausible that government policy should step in to plug the gap. Long-term mental health policy would benefit from a better understanding of the determinants of mental illness among the elderly and the interventions that can mitigate it. Banjeree et al. say their evidence "suggests that policies to lower poverty and financial strain might go a long way toward reducing depression and improving psychological well-being more broadly (for example, via pensions)". Technology may give the elderly relief from loneliness, and we may benefit from new interventions aimed at an increasingly tech-literate elderly population.

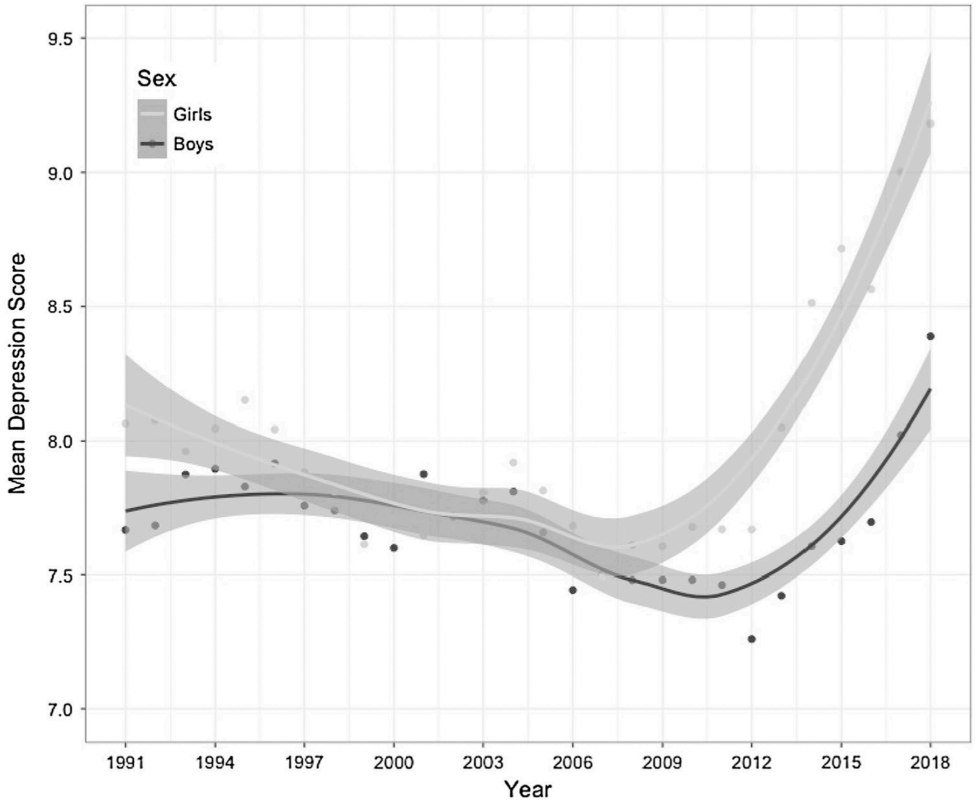
Adolescent mental health may deteriorate

If the population is aging, can we expect interventions aimed at children and mothers to decline in relative importance? It's possible that early-life mental health interventions are underrated (see *Prevention* for more on this). It's also possible that youth mental health is worsening over time. This is a very slippery concept: many people believe that there is a youth mental health crisis of depression, anxiety and eating disorders, perhaps fueled by social media. However, any observed increase could be the result of the increasing awareness of mental health over time.

Birth cohort studies attempt to cut through changes in culture and awareness by surveying successive generations using the same methods. The results from two US studies imply a significant deterioration in youth mental health in the second half of the 20th century. Twenge (2000) saw anxiety and neuroticism among children and college students increase by about one standard deviation between 1953 and 1993. In a more expansive study spanning 1938

to 2007, Twenge et al. (2010) found that “[t]he current generation of young people scores about a standard deviation higher (average d = 1.05) on the clinical scales, including Psychopathic Deviation, Paranoia, Hypomania, and Depression.”

A more recent cohort study by Keyes et al (2019) observed an uptick in depressive scores among US adolescents in the 2010, especially among girls.



Mean depressive affect symptoms by year among US adolescent girls and boys, 1991–2018. The standard deviation of scores is approximately 4.

We simply don’t have birth cohort data of this quality from LMICs. If the factors contributing to youth mental illness are tied in with development, we can expect youth mental health to deteriorate in the coming century.

Data quality

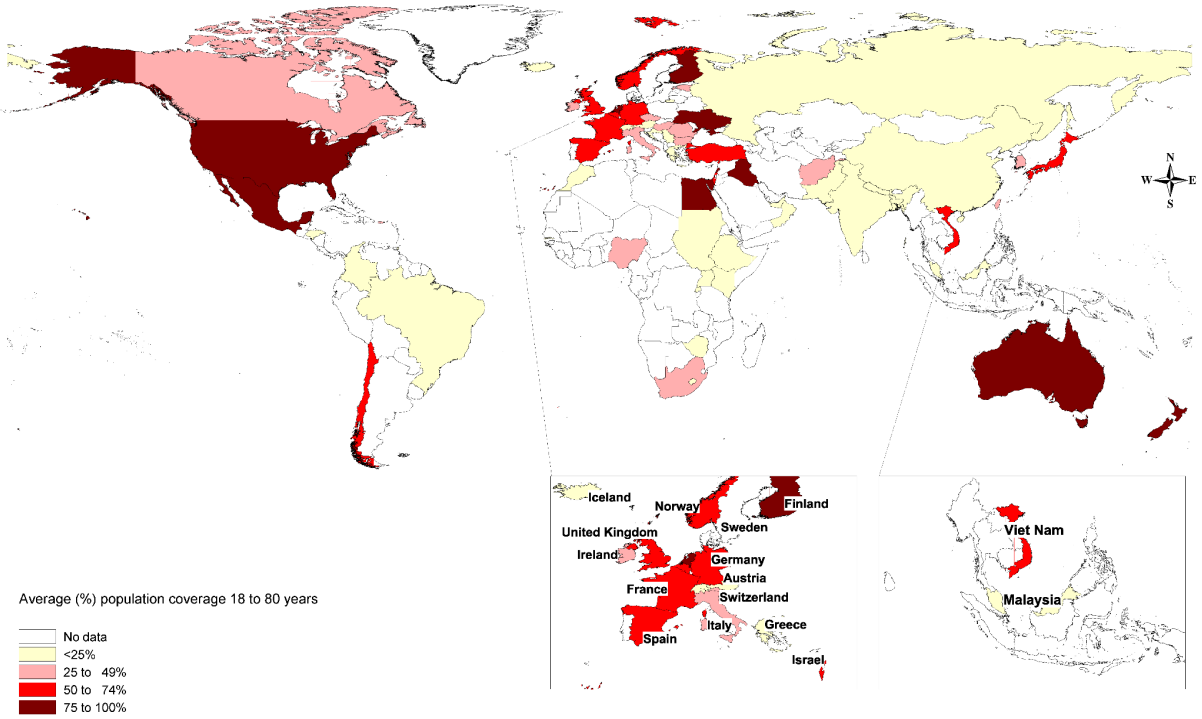
GBD study data

There are a number of reasons why estimates of the mental health burden in LMICs could be highly inaccurate, especially in LMICs:

- Mental illnesses are difficult to diagnose consistently between countries
- Mental health data is scarce in LMICs
- Studies in LMICs are often smaller and have less representative samples
- Diagnosis tools are usually developed in the West, and thus cultural differences may make them less powerful in other regions

The GBD attempts to piece together evidence to estimate disease burdens in every country, but the lack of comprehensive data means that IHME uses statistical techniques to fill the gap. For example, depression is the mental disorder with the *broadest* evidence base, but the GBD only has primary sources from 111 countries (IHME, 2019). This means that dozens of countries are not represented at all.

From the map below we can see that depression and anxiety data from most LMICs is sparse or non-existent. Coverage of East, Southern and Central Sub-Saharan Africa, and South Asia is pitifully low[8] (IHME, 2008).



Population coverage of prevalence data for common mental disorders: averaged across major depressive disorder, dysthymia and anxiety disorders (Baxter et al., 2013).

In regions where the data is sparse or non-existent, IHME extrapolates based on other data such as age sex and risk factors.

Adding to the difficulty is the fact that the available sources are often not easily comparable. From some countries the data may come from clinical diagnoses and health insurance claims, in others they may come from layperson-administered surveys (OWID, 2023). In general, LMIC data is less likely to involve clinical diagnosis, and is more likely to involve smaller, less representative samples. Between countries, different definitions of mental health conditions and different clinical practices make cross-comparison even more difficult.

IHME attempts to account for all of these factors with statistical adjustments. This is a huge challenge, and it seems probable that IHME's stated confidence intervals are too optimistic.

[8] The data in the map is dated and the GBD appears to draw from more sources now. However, the data gap between rich and poor countries remains.

For example, the IHME estimates the age-adjusted prevalence of Schizophrenia at 241 and 434 per 100,000 in Denmark and the USA respectively. This seems implausible, given that the condition is believed to be ~79% heritable (Hilker et al., 2018).

It is possible that the prevalence of depression and anxiety in LMICs is *radically different* from what is found in the GBD results. Experts have suggested that diagnosis methods and surveys developed in the West may have limited utility in other cultures, where MH conditions manifest differently. For example, a doctor in Sub-Saharan Africa suggested to us that patients who seek medical help for physical aches and pains are often found to be suffering from depression. People largely engaged in domestic or manual labor may be less likely to notice "struggling to concentrate" (often used as a key symptom of depression) than white-collar workers in HICs. Hence cultural differences may be leading to under-diagnosis in non-Western societies.

Depression data

In general, the data on depression prevalence is based on clinical diagnoses in HICs, and questionnaire data in LMICs. Questionnaires are usually more *sensitive*, so the prevalence rates that spring from the two methods are not directly comparable. The IHME attempts to use statistical adjustments to account for the differing sensitivity levels, but it is unclear how reliable the results are.

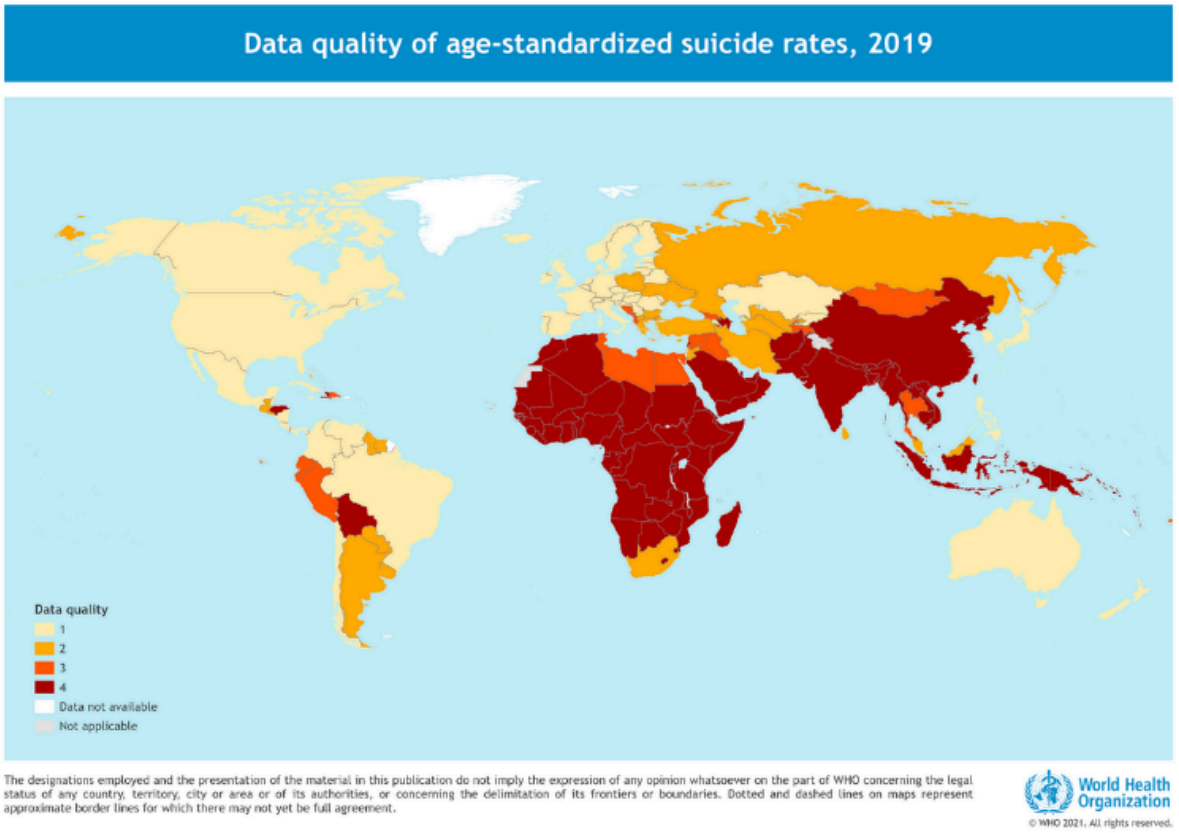
Layperson-administered survey methods have been applied in both HIC and LMIC settings, which can in theory enable us to compare depression rates more directly. However, these studies usually examine small, unrepresentative samples. A meta–analysis by Bello et al. (2022) of depression studies conducted in Africa during the Covid pandemic detected depression prevalence ranging from 8% to more than 80%. Two studies conducted *in the same country* found prevalence levels of 13% and 83% respectively. Data from the United States National Health and Nutrition Examination Survey (NHANES) is probably far more representative than the studies from Africa. Tomitaka et al. (2018) find that PHQ-9 scores from the NHANES "exhibited an exponential pattern, except at the lower end of the distribution" (Tomitaka et al., 2018). Most scores were under 5 (out of 27), and there was a long right tail of higher scores. Unfortunately, without large, representative studies in poor countries, we are unable to "directly" compare PHQ-9 scores between rich and poor countries.

Even if we had better studies, it is possible that our diagnostic tools are less reliable in non-Western societies. One expert suggested to us that depression manifests differently in different cultures, and hence Western-developed diagnosis tools can under-diagnose when applied in other settings. Scorza et al. (2018) examined data from the World Mental Health Surveys (WMHS), a series of studies aimed at predicting prevalence of mental disorders across 27 countries. They found that the estimated depression prevalence in the Nigeria sample was *several times higher* when using a Latent Transition model instead of the algorithm used by the WMHS. Prevalence estimates in the US and New Zealand barely changed under the different models. This illustrates just how sensitive measurement methods our estimates are.

Suicide data

We are doubtful that the IHME will be able to model suicide in LMICs accurately, given the complex factors that affect suicide rates and the reliability of available data. Suicide appears to have a strong cultural component which may not be captured by the demographic factors used in extrapolation. For example, suicide rates appear to be lower in religious societies, but it is unclear how much of this is attributable to underreporting vs. actual lower prevalence. Many LMICs have criminalized suicide, which makes accurate measurement near-impossible. Experts told us that many suicides are not reported, and that police and medical authorities often keep separate suicide records without cross-referencing.

The WHO provide a map (below) which illustrates the suicide data gap in LMICs (A score of 1 denotes high-quality data).



Anyone considering a suicide intervention in an LMIC should not complacently rely on official data. The Centre for Pesticide Suicide Prevention know this: they combine data from police reports, health services and local surveys in order to build their own best guess.

Is the LMIC burden underrated?

For anyone who wants to combat mental illness, the precise scale of the mental health burden in LMICs is not particularly important. We are confident that people suffer from mental illness all over the world, and that people in poor countries are both less likely to have access to treatment, and are cheaper to treat. Hence as a general rule interventions targeting mental illness in LMICs will be more cost-effective than those in HICs.

However, scale *is* important if we wish to determine how much of the limited resources available in LMICs should be allocated to mental health. Unfortunately, we feel that there is not enough data to determine with confidence whether the mental health burden in LMICs is underrated.

Reasons why the burden of mental health may be **underestimated** in LMICS:

- Diagnostic criteria and tools are largely developed in the West. When applied in culturally different contexts, they may be less sensitive to mental illness
- Awareness of the importance of mental health is considered to be worse in LMICs. This may cause people to ignore mental health problems, or to ascribe symptoms to physical or spiritual causes.
- Our health metrics (such as the QALY and DALY) may generally underrate the badness of mental illness compared to other conditions, as explored in *How bad is mental illness?*.

Reasons why the burden of mental health may be **overestimated** in LMICS:

- Researchers may be motivated to exaggerate the prevalence of mental illness in LMICs in order to raise the profile of this neglected field
- Many symptoms that in wealthy countries are signs of mental illness may in LMICs be the product of poverty. Poor sleep and trouble concentrating, for example, are symptoms of depression but could also be caused by malnutrition.

Critical research gaps

We need research to improve our understanding of the burden of mental illness, especially in LMICs.

- Studies that use identical (and more culturally-neutral) methods **and** representative samples to estimate the prevalence of mental illnesses in both wealthy and poor countries
- Investigations into the different manifestations of depression and other mental illnesses in specific LMIC settings

Conclusion

Mental illness imposes an enormous burden. It inflicts misery and reduces productivity around the world. However, a lack of reliable data makes it difficult to pinpoint precisely *how bad* the toll is, and makes it almost impossible to trace how it is changing over time. Improving the quality of this data will be instrumental at improving our powers of tackling mental illness.

We can be confident that depression, suicide and anxiety are the greatest sources of mental health burden. These conditions are correlated with each other and contribute to each other, and work on one condition is likely to mitigate the others.

As we make progress against communicable diseases, and as the average age rises, mental health is projected to become a larger share of the disease burden. As the

easiest-to-prevent physical illnesses fade away, mental illness interventions will become more cost-competitive.

We can expect the nature of the burden to evolve, and the problems of loneliness and depression among the elderly to become more important. But there are also indications that youth mental health has been deteriorating in wealthy countries. If this trend holds true in developing countries, we can expect disorders of adolescence to become a larger problem.

How bad is mental illness?

Key Findings

- The importance of mental illness is highly contingent on one's philosophical assumptions
- Health measures such as QALYs and DALYs probably fail to fully capture the badness of severe mental illness. The WELLBY may be better at capturing internal states but has many of the same weaknesses
- We think that even with no change in philosophical assumptions, the DALY value of averting depression should be 20-45% higher than the IHME weighting suggests
- *Converting between* measures of health, wellbeing and life satisfaction introduces significant uncertainty and should be avoided when possible.
- Decision-making on mental health interventions should be clear about any philosophical and measurement assumptions that are being made

How can we measure mental health outcomes?

Philosophical Assumptions

Averting mental illness looks good from all major philosophical perspectives. Mental illness increases suffering, reduces happiness, is a state of ill-health, prevents people from fulfilling their desires, and so on.

Most of the benefits of mental health interventions come from *improving*, not *extending* lives. Thus the importance of averting mental illness relative to other causes largely depends on philosophical assumptions.

- **The value of improving lives vs bringing new ones into being.** Improving wellbeing and extending life are fundamentally different. Prioritizing between mental health interventions and physical health interventions depends on moral positions, such as the relative value of averting an infant death, which vary widely between people.
- **Trade-offs between preventing suffering and increasing happiness.** Some measurement frameworks impose limits on the worst and best possible states of health or wellbeing. Those who believe that severe suffering outweighs extreme pleasure are more likely to favor interventions that address suffering.
- **The value of wellbeing versus freedom.** Those who place a high value on personal freedom are less likely to support interventions that restrict freedom in the name of public health, such as bans on commodities used to commit suicide.

In general, mental health work becomes *more valuable* for those who:

- **prioritize improving lives over bringing new ones into being**

- **prioritize preventing suffering over increasing happiness**[9][10]
- **(in the case of preventing suicide) believe that curtailment of freedom is justified**

Inevitably, the units of value that we use to determine the effectiveness of various interventions will not precisely reflect our philosophical assumptions. Hence it is important to select the best measures available, and to understand whether they are under- or over-estimating value according to our philosophical position.

HALYs and their limitations

Health-Adjusted Life Years (HALYs) attempt to put a value on the burden of disability and early death. QALYs and DALYs are the most widely used. They are derived from ordinary people's evaluations of how bad it would be to live with the condition as described.

Foster (2020) gives an excellent overview of QALYs and DALYs, which informs most of the summary below and is the source of the images of health scales.

Quality-Adjusted Life Years (QALYs) aim to measure health-related quality of life (HRQoL). Karimi & Brazier (2016) describes HRQoL as "the way health (as measured by health status questionnaires) affects QoL (as measured by QoL questionnaires) as empirically estimated using statistical techniques". Notice the narrowness of the definition: QALYs are *not* attempting to capture the full spectrum of QoL, only the portion determined by health.

In surveys of the general population, respondents are asked about a number of health states[11]. Each survey produces a "value set" which assigns a QALY score to each of many possible health states.

1 QALY is equal to a year of life in full health, while 0 QALYs is a health state equivalent to death.

The QALY scale

[9] I intend a very vague meaning of "prioritize" here: the *more heavily you weigh* preventing suffering over increasing happiness, the stronger the case for tackling severe depression over, say, saving lives.
[10] Mental illness was found to be the biggest global cause of misery (defined as the bottom 10% on life-satisfaction ratings).
[11] There are several ways of eliciting QALY scores for each health state, which we will not go into here.

The QALY scale admits scores below zero, which represent states worse than death. In practice, QALY value sets have few health states with negative scores, and the worse states never seem to be lower than -1.

Disability-Adjusted Life-years (DALYs) attempt to quantify how *healthy* different states are. Survey respondents are given short descriptions of two people experiencing different health states, and asked “who do you think is healthier overall, the first person or the second person?”.

For example, a respondent might be asked to who is healthier out of:

- **Person 1:** “has overwhelming, constant sadness and cannot function in daily life. The person sometimes loses touch with reality and wants to harm or kill himself (or herself)” [Major depressive disorder, severe episode]
- **Person 2: “**drinks a lot of alcohol and sometimes has difficulty controlling the urge to drink. While intoxicated, the person has difficulty performing daily activities” [Alcohol use disorder, mild]

(WHO, 2020)

From thousands of such comparisons, researchers derive a weighting for each health state. In contrast to QALYs, 1 DALY is equal to a year of lost healthy life and 0 is a year of full health.

The DALY scale

Importantly, there are no health states worse than death.

Since it is based on judgments of which states are *healthier* than others, the DALY may be particularly bad at capturing the unpleasantness of mental illness. Respondents may not see unhappiness, anxiety, pain, etc. as *unhealthy* relative to "obviously unhealthy" injuries and infections.

QALYs and DALYs share a host of serious **limitations**.

- They rely on people’s assessments of health states that they may never have actually experienced. Arguably, mental health states are harder to imagine and easier to trivialize than physical health states, and so the survey process may undervalue mental illness
- There are signs that people incorporate the mitigating effects of treatments and care in their assessment of health states, even though they are not supposed to (Feng et al., 2020, Patenaude & Bärnighausen, 2019), which may lead to some illnesses being undervalued.
- This also means that we weigh illnesses equally across cultures, even though it is likely that suffering is more acute among those who are living in poverty. For

example, the HALY weightings of arthritis may "assume" easy access to painkillers, which is often not the case in poor countries.
- They cannot capture states of extreme happiness and extreme suffering
- They appear to weight pain very lightly. For example, *terminal illness with constant, untreated pain* has a disability (DALY) weight of 0.569, which is only 0.029 more than the weight for the same condition with pain medication.
- They only aim to measure the impact of the health state, not its comorbidities.

How should we interpret HALYs?

In practice, mental health data often comes in the form of QALYs and DALYs. Some key questions for determining whether they may over- or under-estimate the things we care about:

- Is this health condition likely to be well understood by the general population?
- Does the health condition involve extreme suffering that may not be captured by the HALY?
- Is pain central to the "badness" of the health condition? Such conditions are likely to be underestimated, especially if respondents assumed access to pain relief
- Is the health condition a risk factor for other conditions? Many health conditions raise the risk of other illnesses, but HALYs do not account for this.

Case study: Is the DALY weight of depression an underestimate?

Most people have never experienced depression, and yet the DALY weightings for mild, moderate and severe depression are derived from surveys of the general population. Furthermore, the IHME's disability weights do not capture the effects of depression as a risk factor for other conditions, including suicidality. In our own analysis, we attempt to correct the DALY weightings of depression to account for how sufferers weight the condition, and for the excess suicide burden associated with depression.

In general, sufferers of a given illness are found to rate it similarly to the general population, or even to rate it as less severe. To quote heavily from Pyne et al. (2009):

> *A number of studies have compared health state preference scores generated by different groups. Some of these studies have found differences based on health experience (Gabriel et al. 1999; Lenert, Treadwell, and Schwartz 1999; De Wit, Busschbach, and De Charro 2000; Postulart and Adang 2000; Insinga and Fryback 2003; Rashidi, Anis, and Marra 2006;) while others have not (Balaban et al. 1986; Revicki, Shakespeare, and Kind 1996; Dolders et al. 2006;). In general, studies that compare patient and general population health state preferences find that patients assign preference scores [...] that are equal to or greater than the preference scores assigned by members of the general population (Sackett and Torrance 1978; Balaban et al. 1986; Froberg and Kane 1989b; De Wit, Busschbach, and De Charro 2000; Dolders et al. 2006;)*

Pyne et al. found that depressed patients were found to rate depression as *worse* than members of the general population did:

> *Depressed patients report lower preference scores for depression health states than the general population. In effect, they perceived depression to be worse than the general public perceived it to be.*

In our analysis, we find that the ratio between disability weights derived from depressed people and those derived from the general population in Pyne et al. is **1.20**. In essence, depressed people perceive depression to be 20% more severe than non-sufferers.

We also seek to estimate the extra health burden from the increased risk of suicide associated with depression. Depression is widely quoted to be responsible for at least half of suicides (CPSP, JED, CSP) even though only ~3% of the population is believed to suffer from depression in a given year (GBD study). There does not appear to be reliable data to back up the "at least half" figure, and it is presumably highly contingent on the depression threshold used for diagnosis[12]. But the link between depression and suicide is well-established. A 2001 Swedish study found that hospital diagnosis for depression increased suicide risk by 20x (Osby et al., 2001).

We examine males and females separately, since they have very different rates of depression and suicide, and find that

- In males, depression increases suicide risk by ~40x, or an extra 0.4% per year of depression
- In females, depression increases suicide risk by ~25x, or an extra 0.1% per year of depression
- Overall, the increased suicide risk adds 0.066 to the DALY weighting of depression

These figures are *highly contingent* on several factors, including the prevalence of depression (which we suspect is higher than the GBD figure) and the background suicide rate. We should expect the burden of increased suicide risk to vary widely based on the context.

In the end, we determine the DALY weighting of a year of depression to be 0.39, which is 44% higher than the IHME figure [link to calculations].

[12] We suspect that the suicide risk for severe depression is higher than for mild depression. If this is true, the stricter your diagnostic criteria for depression, the stronger the relative risk of suicide.

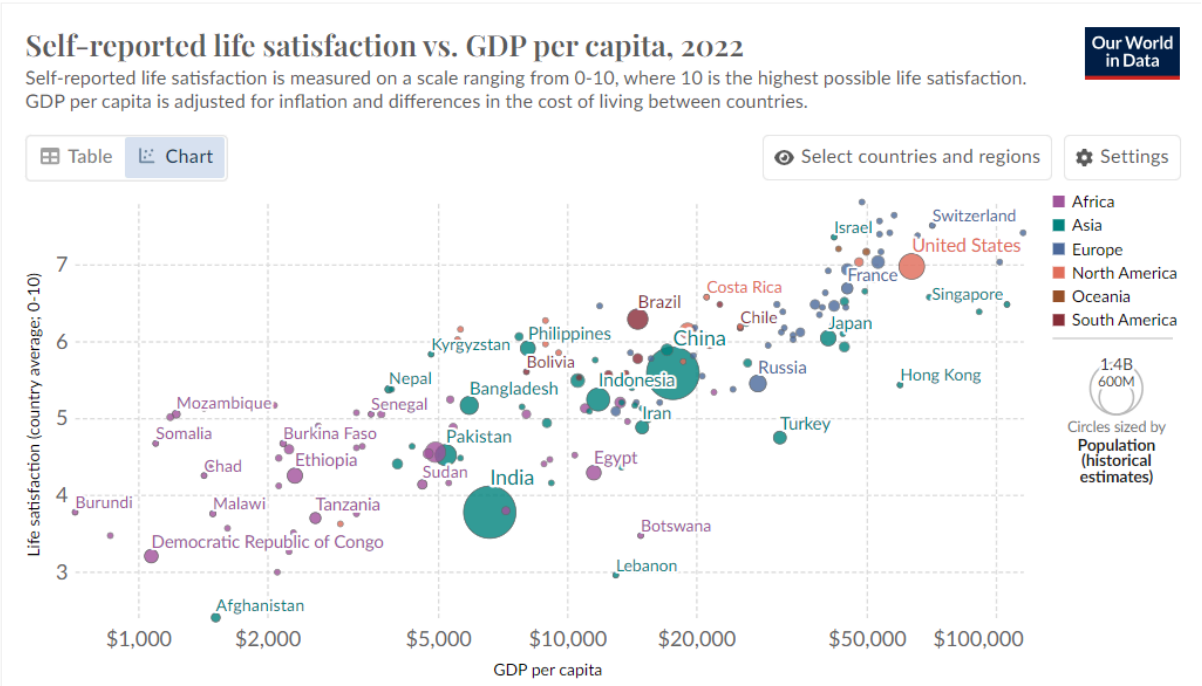
	Mild depression	Moderate depression	Severe depression	Average *(weighted according to prevalence in population)*
DALY weighting according to IHME	0.16	0.41	0.66	0.27
Updated DALY weights, adjusted for sufferers' view of depression.	0.19	0.51	0.76	0.33
Updated DALY weight, incorporating increased risk of suicide and sufferers' view of depression	-	-	-	0.39

Wellbeing approach

The Happier Lives Institute (HLI) has proposed subjective wellbeing (SWB) as an alternative measure of value that avoids some of the pitfalls of health and income measures. We feel that SWB does have a number of strengths, but ultimately suffers from other serious limitations.

The best-known way of measuring SWB is with life satisfaction (LS) scores. Respondents are asked to rate their satisfaction with their life on a scale of 0 to 10 (although the wording of the questioning varies). An increase in LS lasting for one year is known as a WELLBY.

LS has been found to be significantly linked to income (OWID, 2017), which suggests that LS scores are capturing some objective level of wellbeing and are not merely an indication of wellbeing *relative to others in the community*.



Life-satisfaction and GDP per capita, from Our World in Data, 2017

HLI argues that by measuring SWB before and after an intervention, we can determine improvements in wellbeing that are comparable across physical health, mental health and anti-poverty interventions. LS scores may capture *actual* changes in quality of life, rather than merely inferring them by observing indicators like health scores. They can also lay claim to being more free from the subjective judgment of the scientist: respondents are able to evaluate their own lives based on what is most important to them.

However, SWB has a number of its own limitations:

- People have been found to report widely differing levels of LS when asked at different times (Krueger & Schkade, 2008). These variations may "average out" on a population level, but they imply that respondents' self-reported life satisfaction may be heavily influenced by short-term influences. This raises the risk of reporting bias.
- Evidence suggests that introverts and extroverts may report the same internal state differently (Fabian, 2021).
- SWB data is often not available, which makes it difficult to evaluate interventions on this basis. In response to the absence of SWB data, HLI make the very questionable assumption that one SD improvement in affective mental health (MHa) is roughly equivalent to one SD improvement in SWB[13] (HLI, 2021) (UPDATE: HLI have recently adopted their stance on this). This kind of conversion adds further uncertainty.
- Taken literally, a SWB approach would imply that the lives of those in wealthy countries are up to twice as valuable as those in poor countries[14]. For many people, this violates the reasonable assumption that all lives are broadly equal in value.
- The life satisfaction scale may not be "truly" linear. For example, response data suggests that scores of 0, 5 and 10 are given "more often" than the distribution of other scores would suggest. It seems plausible that the difference between people at 1 and 3 on the scale is on average far greater than those at 6 and 8, and yet they are assumed to be equal by the linearity of the scale. This would lead us to underweighting interventions that help the most miserable.

Just like HALYs, life satisfaction (LS) and the WELLBY do not capture extreme positive or negative states, since scores are bounded by 0 and 10. It is possible that the full toll of severe pain, depression and psychosis simply cannot be measured by the WELLBY, QALY or DALY.

Overall, we expect that taking a wellbeing approach probably leads to valuing mental health work more highly[15]. Mental illnesses are hard to observe from the outside but are often acutely felt by the sufferer. In contrast, physical conditions like blindness or amputation might be highly feared by non-sufferers (so they look severe according to QALY/DALY weights) but may be "not so bad" according to sufferers who become accustomed to their health state.

[13] They justify this by examining correlations in improvement of MHa, depression symptoms and SWB in studies that measure several of these factors (link to HLI analysis)
[14] Or even more, if the "neutral point" is set above zero. In Kenya the average LS score is just 4.4, so a neutral score of 5 would imply that the average Kenyan person's life is net-negative.
[15] Mental illness was found to be the biggest global cause of misery (defined as the bottom 10% on life-satisfaction ratings).

On the other hand, health scores may fail to capture the secondary mental health effects of various health states, and a wellbeing approach may make some health conditions look worse.

Many of the limitations of SWB stem from the measure of life satisfaction. It is possible that better wellbeing measures will be invented in the future. Derek Foster has proposed the HALY+ and sHALY, health metrics which incorporate wellbeing weights, and the WELBY (with one "L"), which is similar to the WELLBY but can measure states worse than death. He is now exploring this further as a PhD candidate at Oxford University.

How should we interpret wellbeing measures?

The credence you give to wellbeing measures depends upon your moral priorities. You should give more weight to wellbeing measures if you think that an individual's subjective experience of life is more important than their objective health condition.

As with HALYs, it is worth considering whether the health condition you are examining involves extreme happiness or extreme suffering that cannot be captured on a 0-10 scale.

Perhaps the most important heuristic is to check for signs of response bias in the study. Is there a blinded control group? Does the control group show an increase in SWB? If non-SWB measures were reported, do they also improve with the intervention?

Relative mental health

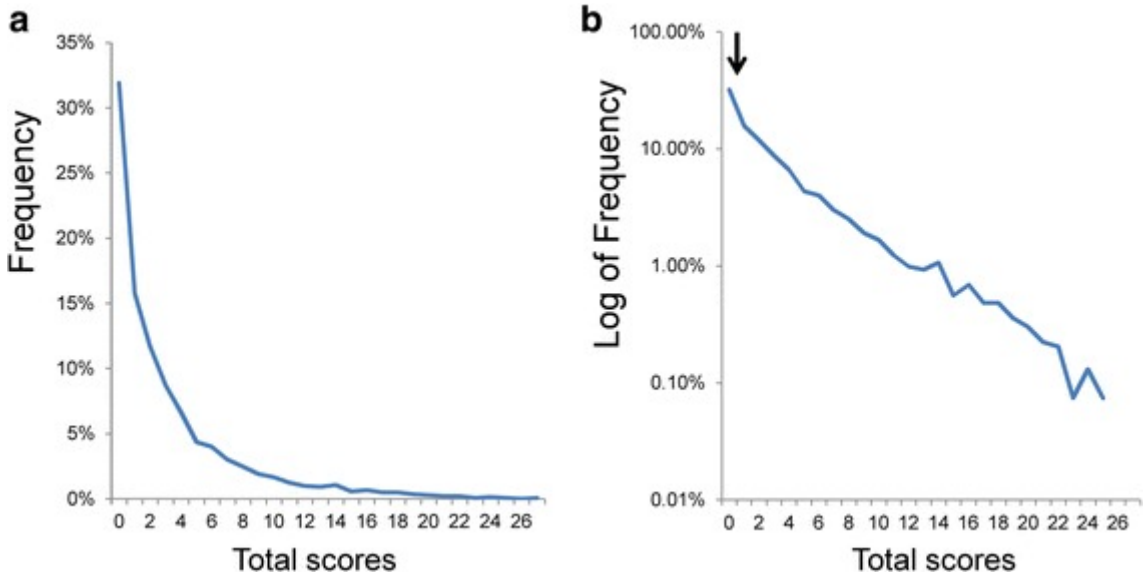
Interventions targeting mental illness may want to measure mental health before and after the treatment in order to track improvements. This is possible with HALY and Wellbeing approaches, but not always ideal.

HALYs assign fixed weights to health states: at best, there are different weights for mild, moderate and severe cases of the condition. This makes HALYs quite insensitive to the *magnitude* of improvement. Furthermore, judgements about *what counts* as mild, moderate or severe will inevitably introduce noise.

Wellbeing approaches, as explained above, capture many things other than the condition being targeted.

One response to these problems is to determine the severity of mental illness by measuring *symptoms*. A symptom questionnaire assigns a score to each patient. The population standard deviation (SD) of these scores is often used as a unit of value. A one SD improvement of symptoms for one year is known as an SD-year.

The PHQ-9, for example, is a 9-item questionnaire which asks *how often* the patient has experienced each of nine depression symptoms in the past two weeks. This provides a score of 0 to 3 on each question, for a total score out of 27. A population survey in the US by Tomitaka et al. (2018ii) found that the mean PHQ-9 score was just 3, with an SD of around 4. Thus a person going from a score of 20 to a score of 8 would experience a 3-SD improvement in depressive symptoms.



The distribution in PHQ-9 scores among a sample of the US population (Tomitaka et al., 2018i)

The distribution of scores was heavily right-tailed Most people reported very few depressive symptoms, and a minority reported far more than average.

There are questions about the validity of measuring progress in terms of SDs. Is a four-point improvement *really* equally valuable for people with scores of 5 and 20?

Straining credulity further is the practice of *converting between* SD measures, which we examine below.

Converting between measures

Converting between measures introduces a lot of uncertainty. Past efforts at conversion have rested on highly contestable assumptions - including the assumption that it is valid to linearly map one measure onto another.

As seen above, there is evidence that depressive symptom scores (as measured by the PHQ-9) are positively skewed with a very long right tail, while life satisfaction scores in LMICs are negatively skewed or symmetric (OWID, 2017). Even if we use correlational evidence to derive a conversion rate between SDs of PHQ-9 score and SDs of life satisfaction scores (as HLI have attempted) (UPDATE: HLI have recently adopted their stance on this), it may not be valid to use this conversion.

For example: if a severely depressed person (PHQ-9 score 24) experienced a 3 SD improvement in depression symptoms (reducing their score to 12), *they would probably still be among the most depressed 5% of the population* (Tomitaka et al., 2018ii). But a 3 SD improvement in life satisfaction is an *increase* of around 6.5[16], which is on par with average

[16] HLI assume the standard deviation of LS to be 2.17 (see "input" tab here). It is generally thought to be higher than this in HICs and lower in LMICs.

wellbeing in Italy. It seems implausible that someone within the most depressed 5% of the population would have higher-than-average life satisfaction.

Converting between SD-years of depression and DALYS

Although fraught, we do use conversions in order to compare mental health interventions. Psychotherapy intervention effect sizes are often given in terms SD-years, where one SD-year is a one-standard-deviation improvement in symptoms for one year, and we attempt to convert this to DALYs. Much rests on this conversion rate[17].

A survey of a general US population (Tomitaka et al, 2018ii) found the SD of PHQ-9 scores (a 27-point scale) to be slightly more than 4 points. We estimate that depression on the cusp of moderately-severe and severe has a DALY weight at 0.65[18], and corresponds to a PHQ-9 score of 19.5. Thus we estimate an exchange rate of around $0.65 \times \frac{4.34}{19.5} = 0.14\ DALYs$ per SD of depressive symptoms. However, it is unclear whether this logic can be extended to smaller improvements, or to people with milder depression, or the (non-depressed) family members who we assume are benefiting from spillover effects.

Furthermore, effect sizes are usually calculated using the SD of the *study population,* which may not be equal to the value of 4.3 found in Tomitaka et al, 2018ii. This means that a high effect size could simply be the product of low variance in symptoms in the study population (which has, after all, been screened for acute mental illness). HLI considers this "range restriction" effect, but they determine that a 10% discount is enough to account for it.

Conclusion

It is fundamentally difficult to compare the badness of mental illness to the badness of physical illness or death. Measures like the DALY, QALY or the WELLBY will lead to different valuations of mental illness, and we should be aware of the philosophical assumptions implicit in these measures.

It's possible that the QALY and DALY, which are common currencies in the field of global health, systematically underestimate the badness of depression and of severe mental illness. Life satisfaction has a better chance of capturing internal welfare states, but comes with its own problems. We can partially correct for limitations in these measures by making our own adjustments. Some things, like extreme pain, depression and psychosis, may never be fully captured by bounded scales.

[17] Or, for that matter, the WELLBY weighting of an SD-year of depression.
[18] DALY weight of depression

Treatment

Key Findings

- LMICs are generally the best target for treatment interventions due to the lower costs of operating and higher rates of unmet needs
- The quality of psychotherapy studies is highly variable, and many of the observed effects may be the product of bias.
- Interpersonal Group Therapy (IGT) appears to be a cheap way of scaling up mental health provision in LMICs. The persistence and spillover effects are important but not well understood
- Psychotherapy delivered online, by AI or through apps has the potential to be far cheaper than in person, although this is a very young field
- Expanding antidepressant access could be highly effective in some countries, but we have not found examples of past success to learn from
- Severe or chronic pain is often left untreated in LMICs, but increasing opiate access comes with significant risks

In-person psychotherapy

Summary

- No single type of psychotherapy appears to be superior, and effect sizes are relatively independent from the quality of the therapist or the number of sessions
- There are concerning signs of publication bias, response bias and undetected placebo effects in the literature on psychotherapy, which are difficult to adjust for
- Interpersonal group therapy (IGT) in poor countries is cheaper and likely to be more cost-effective than other types of in-person psychotherapy.
- The persistence of the benefits and the scale of spillovers are not well-known, and would benefit from further study
- Driving down cost is probably the best lever for improving the cost-effectiveness of psychotherapy

Effectiveness

There are various types of psychotherapy, including psychoanalysis, cognitive behavioral therapy (CBT) and humanistic therapy. Psychotherapies are used to tackle many mental health disorders, especially depression and anxiety, which together account for over half of the global burden from mental disorders (IHME).

The evidence does not point to a single “best” form of psychotherapy (SoGive, 2023), and so we suggest that the *style* of psychotherapy is less important than other factors like the cost of delivery.

A number of attempts have been made to determine the average effect size of psychotherapy in general. In a meta-analytic evaluation of meta-analyses (2022),

Leichsenring et al. found a modest standardized mean difference (SMD) of 0.34 (95% CI: 0.26-0.42) for psychotherapies compared with placebo or treatment as usual (across disorders). A recent meta-analysis by SoGive (2023) found an effect size of g=0.32 for 6-session psychotherapy interventions for depression.

It is important that psychotherapies are compared with adequate controls and not a theoretical "no change" group, given that control groups tend to show considerable improvement from baseline[19]. Leichsenring et al. found that among patients suffering from depression, subjects in control groups achieved remission in 23% and 33% of cases when they were assigned no treatment or treatment-as-usual, respectively. The figure for those receiving psychotherapy was not much higher, at 43%.

The effectiveness of mental health treatments is often given in "SD-years" of the metric being measured. Hence psychotherapy effect sizes are often given in terms of SD-years of depressive symptoms, where one SD-year is equivalent to reducing someone's depressive symptoms by one standard deviation for one year. Hence it is not just the initial effect size that matters, but also the persistence of the effects. Mitigating mental illness is thought to have positive effects on people close to the beneficiary, known as spillover effects. The overall effect size of a psychotherapy is thus highly contingent on three factors:

- **Initial effect size:** the improvement in symptoms at the end of the course of treatment.
- **Persistence:** the duration of the effects, which is determined by the rate at which the effects decay over time
- **Spillover effects:** the benefits experienced by those close to the recipient (usually those in the same household)

All studies attempt to measure initial effect size, although these results should be interpreted with caution. Persistence and spillover effects are far less-well understood. If at all, follow-up data is usually collected just a few months after the intervention. Estimates of the long-term persistence of the effects must either extrapolate from short follow-ups, or else rely on a small number of studies with multi-year follow-ups. Happier Lives Institute (2023) found that the exclusion of the *five* long-term follow-up data points (out of 222 available) *decreased the expected persistence of the effects by 58%.* Spillover effects are perhaps even more difficult to estimate: there is a small evidence base, and the results are highly heterogeneous.

This suggests that the *overall* effect size of psychotherapy is not well-understood, and that extra research could be valuable in reducing this uncertainty.

Interpersonal group therapy

Although pioneered in the West, psychotherapy has been shown to be effective in LMIC contexts (Cuijpers et al., 2018). Yet mental health provision in LMICs is generally poor, and very few are able to access psychotherapy. A third of LMICs do not have a mental health budget, and those that do have one spend an average 0.5% of the national health budget on mental health (Founders Pledge, 2019). There are nowhere enough mental health

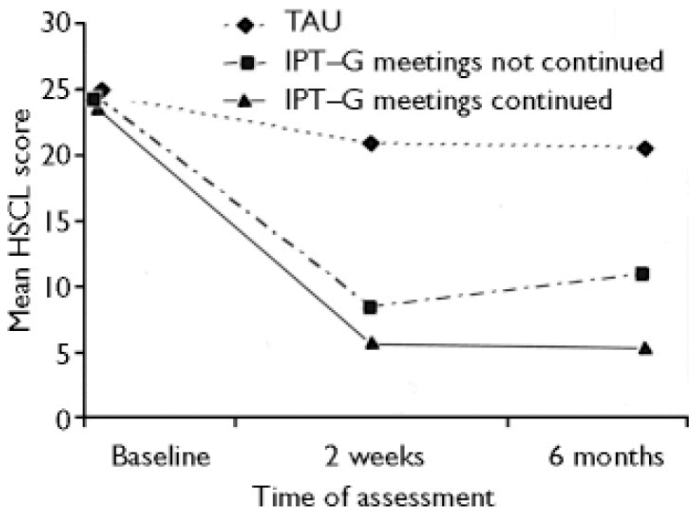
[19] This is probably due to regression to the mean, since participants are screened for serious illness, and may also be caused by response bias and placebo effect.

professionals to meet demand: HICs have 50 times more mental health workers per capita than LICs (Founders Pledge, 2019). This makes expert-led psychotherapy an inefficient and insufficient intervention in LMIC contexts.

Interpersonal Group Psychotherapy is a model of therapy that focuses on the individual's relationships with others (in contrast to CBT, which is more about fostering healthier thought patterns). When delivered by laypeople, IGT can be a scalable and more affordable way of filling the treatment gap - it can cost just $35-650 per person treated (HLI, 2021 i, ii). Local people are given training and run a series of group therapy sessions to people who have been screened for depression or another mental illness. When mental health experts *are* available, they can be used to train lay-practitioners and to lead projects, in a process known as *task-shifting* (Patel, 2009). IGT is recommended by the WHO, which has created an open manual for delivering an 8-session program.

Group therapy has been shown to be similar or even greater in effectiveness to individual psychotherapy in treating anxiety and depression (Barkowski et al., 2020, Cuijpers et al., 2019) and was found in one meta-analysis to reduce depression symptoms with a Cohen's d of 0.63 (95% CI: (0.36, 0.90)) (Cuijpers et al., 2011, mostly HICs). Another meta analysis (2020, mostly HICs) by Janis et al. found improvements in depression scores with Cohen's d of 0.66 (95% CI: (0.29, 1.03)). This fell to 0.60 in short-term follow-ups and 0.24 after long term follow-ups, suggesting diminishing counterfactual effect over time.

Promising results from a trial in Uganda by Bolton et al. (2003) kick-started interest in IGT in LMICs: " [a]fter intervention, 6.5% and 54.7% of the intervention and control groups, respectively, met the criteria for major depression compared with 86% and 94%, respectively, prior to intervention". The benefits appeared to be persistent, with signs of only slight attrition at 6-month follow-up (Bass et al., 2006). However it should be noted that *most recipients continued to meet up informally after the intervention* (triangle group in the chart below), and only 15% (n=15) did not continue to meet up (square group), which makes the figures on persistence unreliable and perhaps unrepresentative of what would happen in other contexts.



Results from Bass et al. (2006) suggest that IPT-G provided distinct counterfactual benefits compared to treatment as usual (TAU) in Uganda.

It has been argued that “a lot—perhaps too much—hinges on this RCT”.

Concerns about the evidence base

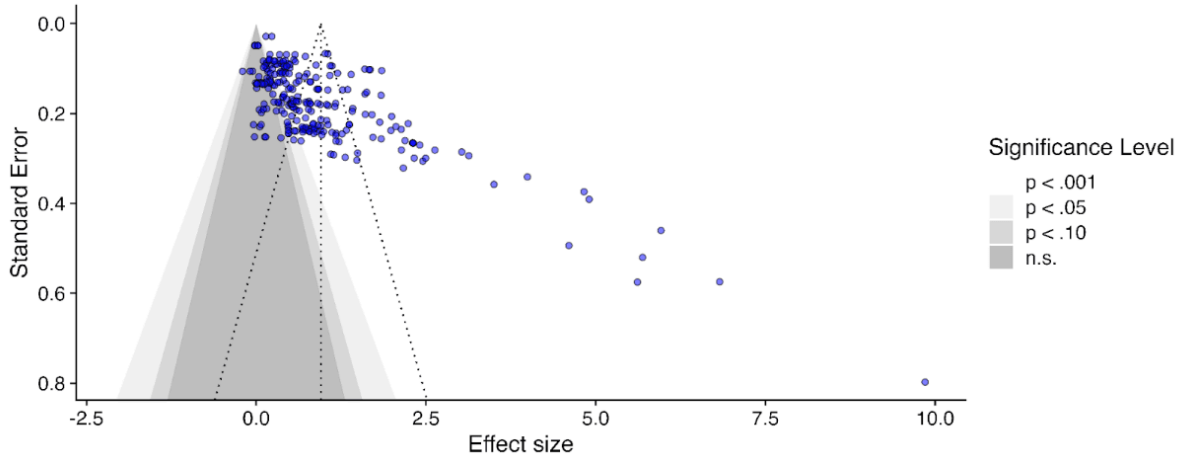
One surprising finding about psychotherapy effect sizes is that they do not seem to be significantly affected by seemingly important factors like the number of sessions, the type of therapy , the mode of delivery and the quality of the therapist. This may reflect an important truth about psychotherapy, or it may suggest that large variations in study quality introduce enough noise to drown out underlying differences in effect size.

Specifically, concerns have been raised that there are signs of publication bias in the literature on psychotherapy, and that studies are often susceptible to response bias and undetected placebo effect. There are established methods for accounting for these effects, but endline results are sensitive to the exact methods chosen. This means that we should interpret all results with caution.

Publication bias

Funnel plots suggest a strong tendency for smaller psychotherapy studies to show larger effect sizes – a sign of publication bias. The extreme nature of the outliers means that publication bias adjustments are *extremely sensitive* to subjective decisions about which outliers to exclude from the analysis, if any.

The funnel plot below comes from Happier Lives Institute’s psychotherapy meta-analysis, part of their 2023 CEA of StrongMinds. They decided to eliminate all results with an effect size greater than g=2 on the basis that these outliers were probably the result of “poor study quality or statistical noise”. However, as Gregory Lewis points out in a detailed comment, this reduces the publication bias discount enough to lead to a substantially higher effect size than would have been reached had the outliers been included.



However, it is unclear whether *keeping* the outliers is a better move, since this would make the final results extremely sensitive to the results of a few small, suspicious studies. Overall we urge that readers carefully check publication bias adjustment processes when interpreting estimates for the effect size of psychotherapy.

Response bias

Psychotherapy studies detect mental illness symptoms before and after the intervention by questioning the patient. This exposes them to possible response bias, especially:

- Exaggerating symptoms during screening in order to qualify for treatment
- Exaggerating improvements in their symptoms in order to please the surveyor (social desirability bias)
- Exaggerating improvements in their symptoms in the belief that they may get "cash or material goods" (future hope bias).

In theory, a well-conducted RCT would experience equal amounts of response bias in its control and intervention groups, and hence underlying *differences* in improvement between the two groups would not be affected. In their 2023 investigation of StrongMinds, SoGive concluded that response bias was probably not a major concern in their meta-analysis, which screens for study quality and accounts for publication bias. GiveWell, on the other hand, discounted effect sizes in HLI's meta-analysis by 20% to account for response bias.

To us it seems likely that studies with a "waiting list" or "treatment as usual" control group are significantly affected by response bias.

Placebo effect

SoGive (2023) provides a good summary for the evidence of placebo effects in depression studies. The effect size of psychotherapy interventions is generally found to be lower when the control group receives a pill placebo, and placebos perform quite well when compared with passive controls. The placebo effect could explain why the mode of delivery of psychotherapy appears to have little impact on the effect size.

The StrongMinds controversy

StrongMinds is a non-profit founded in 2013 that provides group interpersonal group therapy to impoverished women in Uganda and Zambia, which they now claim to be able to deliver for just $63 per person. StrongMinds has perhaps been assessed more than any other organization providing psychotherapy in LMICs: they were recommended as a top charity by Founders Pledge and Happier Lives Institute (HLI), and these recommendations in turn led to further scrutiny.

We will provide only a brief timeline of the controversy here, but would highlight that the gradual downgrading of StrongMinds' cost-effectiveness over time appears to be the product of both Optimizer's Curse and some subconscious bias among those analyzing the organization.

Overall we conclude that StrongMinds is unlikely to be the "holy grail" mental health organization that could rival GiveWell's top charities. But it remains possible that a cheap and effective psychotherapy intervention can alleviate mental illness for $100s per DALY.

2017

The Oxford Prioritization Project performed a cost-effectiveness analysis of StrongMinds (no longer public), estimating the cost-effectiveness at 154 DALYs per $100,000

2019
Founders Pledge estimated that StrongMinds could avert a DALY for $377 (naively, 265 DALYs per $100,000), and began recommending them as a top mental health charity.

2021
Happier Lives Institute created a cost-effectiveness analysis of StrongMinds that drew upon both general research on psychotherapy and specific data on StrongMinds. They converted mental health improvements into wellbeing improvements, and concluded that StrongMinds could provide 6200 WELLBYs per $100,000. Under certain philosophical assumptions, they argued, this made StrongMinds comparably cost-effective to the best malaria-prevention interventions.

The report was criticized for rushing the meta-analysis process and thereby drawing upon data from interventions that had little in common with StrongMinds. Some commenters accused them of using favorable data on persistence and spillover effects, leading to overestimates of these critical quantities.

2022
Giving What We Can and Founders Pledge were criticized for downplaying the uncertainty around the effectiveness of psychotherapy in general and StrongMinds in particular.

Early 2023
GiveWell released a CEA which applied a similar structure to HLI's. They used some different inputs and a number of extra discounts to the effect size used by HLI. In a "death by a thousand cuts", they estimated that StrongMinds was around 4x less cost-effective than HLI had concluded (although still likely more cost-effective than GiveDirectly, a GiveWell-recommended cash transfer charity).

Late 2023
HLI released an updated CEA of StrongMinds that addressed many of the previous criticisms. Lower estimates for effect size and spillover effects decreased the estimated cost-effectiveness. This was partly offset by the decreasing cost per patient of StrongMind's group therapy, but the final estimate was 3000 WELLBYs per $100,000, around half of their previous estimate.

Although it was a significant improvement on the previous CEA, Gregory Lewis pointed out in a detailed comment that:

1. The publication bias discount is extremely sensitive to how outliers are eliminated. HLI appears to have removed outliers in a way that favors StrongMinds. *Not* removing them would have led to an effect size more than 3x smaller
2. The Bayesian approach taken by HLI places too much confidence in the general effectiveness of psychotherapy, to the extent that it is virtually impossible for StrongMinds-specific data to change the conclusion that the intervention is highly cost-effective.

An RCT on StrongMinds by Baird & Ozler should release its results soon, and the expectation is that it will find a very modest effect size – smaller than the pooled effect size of psychotherapy according to the literature.

Cost-effectiveness of interpersonal group therapy

All of the BOTECs in this report aim to estimate the cost-effectiveness of a *hypothetical* intervention, using inputs informed by scientific studies and existing organizations. The calculations are found in this spreadsheet which can be run through Carlo to perform a Monte Carlo simulation that provides a *distributional* cost-effectiveness estimate.

Our interpersonal group therapy BOTEC assumes that an SD-year of depression is worth approximately 0.14 DALYs and finds that there is a 2% chance that IPT-G (modeled on StrongMinds) is more cost-effective than a GiveWell top charity (which we estimate at 737 DALYs per $100,000).

It should be noted that the cost-effectiveness is highly contingent on cost per treatment. We assume that the effect size of the therapy is broadly in line with what has been found in psychotherapy meta-analyses, and that the cost is in line with some IGT-in-LMIC interventions. However, due to the scarcity of high-quality studies of IGT in LMICs, it is possible that these relatively cheap interventions do not yield effect sizes similar to those found in better-studied contexts.

Type	Intervention	Cost-effectiveness compared to cash transfers	Proportion of benefits from mental health
Psychotherapy	Interpersonal group therapy, LMIC	5x	100%

Sensitivity analysis estimates how much of the variance in the final output is attributable to each of the inputs. It suggests that *initial effect size, persistence* and *cost per treatment* are the most important unknowns. Note that we are estimating the cost-effectiveness of a *hypothetical* IGT intervention, not StrongMinds.

The cost of an intervention is probably far more controllable than the persistence of its effects. For this reason we advise that *reducing the cost per treatment* is the single best way to increase the cost-effectiveness of psychotherapy interventions. The experts we spoke to suggested that the best ways of reducing costs were 1) avoiding high salaries by headquartering in the target country, and 2) economies of scale. One complication, however, is that monitoring and evaluation add significant costs, and so “cheaper” interventions are more difficult to assess. We feel that funding an RCT on an already-planned program could be a very good way of building the evidence base.

Heuristics for in-person psychotherapy

- Interpersonal group therapy is likely to be cheaper and more scalable than other forms of psychotherapy, especially in LMICs
- Experimental results should be checked for publication bias, response bias and placebo effect

- The best interventions are likely to be longer-lasting than average, with larger positive spillover effects.
- Reducing the cost per treatment is probably the best way of improving psychotherapy interventions. This can be done by avoiding high salaries and by operating efficiently at scale

Self-administered & tech-based psychotherapy

Summary:

- In principle, self-led and tech-delivered psychotherapies can provide most or all of the benefit of in-person therapies at a fraction of the cost
- They come with unique challenges such as low retention rates
- The evidence base is slim but growing: new organizations in the space will provide valuable data

Technology now allows us to deliver psychotherapy remotely. This has the opportunity to be cheaper and more scalable than in-person therapy. Options include:

- Psychotherapy delivered online by video call
- Self-led interventions, usually through a smartphone app
- Chatbot-delivered therapy, which can use conventional or AI chatbot tools

There is evidence that internet-delivered CBT can be effective, and that the effect size is comparable to in-person CBT (Etzelmueller et al., 2020, Ebert et al., 2016, Carlbring et al., 2018). Self-guided internet-based CBT has been found to have a significant effect on depressive symptoms, although the effect is much weaker for those who engage less with the treatment (Karyotaki et al., 2017). There are signs that the gains are still significant 6 months later (Ebert et al., 2016).

High attrition rates are a challenge with online interventions. Among internet-based treatments for psychological disorders “involving minimal therapist contact”, Melville et al. (2010) found dropout rates varying from 2% to 83%, with an average dropout rate of 31% (compared to an estimated 20% in conventional psychotherapy (Swift & Greenberg, 2012)). We suspect that dropout rates are much more variable among online interventions, and that providers of online interventions should have a robust system for minimizing attrition.

The problem of attrition is likely to be worse for self-led interventions. An expert estimated that as few of 1% who download a self-help app end up completing a course of treatment, since the app is competing for attention with all the distractions of a smartphone, and there is no therapist to provide accountability. Self-help interventions have been shown to be effective, but less so than guided interventions (Andersson & Cuijpers, 2009, Baumeister et al., 2014), and we suspect that higher attrition rates account for much of the difference. The expert we spoke to suggested that even if app rights could be acquired for free, high attrition rates can increase marketing costs enough to prevent the intervention being cost-effective. We suspect that self-help apps are high-risk, high-reward options, since apps tend to either attain significant market share or fade into complete obscurity. The counterfactual is important here: self-help apps are a crowded market, so new products must be more effective than the alternatives in order to add value.

There is not much research on chatbot-delivered psychotherapy. An initial study with a non-AI chatbot by Fitzpatrick et al. (2017) suggests a small mitigating effect on depression and anxiety symptoms. AI chatbots such as ChatGPT represent a major step up from earlier chatbots, and we think it is likely that the effectiveness of AI psychotherapy will only increase over time.

There is very little data on the effectiveness of internet-based psychotherapies in LMIC contexts. A meta-analysis by Carter et al. (2020) found mixed results, but the heterogeneity of the studies means that we can not infer much from it. A meta-analysis by Fu et al. (2020) evaluated digital psychological interventions for mental health problems among young people with depression or substance misuse and found “moderate” effect sizes with Hedges’s g 0.60 (vs. control intervention) and 0.54 (vs. treatment as usual).

Poor access to technology is likely to be a barrier in LMICs, especially for targeting older adults. We see stronger potential for targeting young people in LMICs, who are more likely to have a smartphone. For example, it is claimed that about half of children in rural India have access to one.

Online interventions may be less persistent than in-person IGT, where there is potential for lasting, supportive relationships to form, and even the continuation of informal meetings as observed in Bass et al. (2006). However, we think that the constraints on in-person psychotherapy impose a much lower cost-effectiveness ceiling than applies to online interventions. Internet-delivered psychotherapy can rely on cheaper labor from overseas, can bring together geographically disparate people for group interventions, and avoids the cost of hiring venues. Self-help and chatbot interventions do away with many labor costs completely, and have the potential to scale up very quickly.

Kaya Guides is a new organization pioneering a Whatsapp-based guided self-help intervention in India. Their model is to provide minimal levels of human interaction in order to hit a sweet spot between cost and attrition. We think this kind of work is risky, because of the difficulty of gaining share in a competitive app marketplace, but has the potential to be highly cost-effective at scale.

Overcome provides internet-based psychotherapy for mild and moderate mental illness and is free to recipients. It drives down costs by harnessing volunteer labor, chiefly among recent graduates looking for experience to prepare them for a career in clinical psychology. Their main costs are administration and marketing, and the latter is largely covered by Google’s offer of $10k per month in free ads to non-profits.

Overall, we conclude that although it is possible that internet-based psychotherapies have a smaller effect size than in-person alternatives, the opportunities to cut costs mean that they have the potential to be far more cost-effective. Minimizing attrition rates should be a top priority, and internet-based interventions need a sustainable strategy for finding new patients using word-of-mouth or marketing. We have heard anecdotally that staff retention can be a challenge for internet-based psychotherapy, so a sustainable supply of therapists may also be a necessity.

The evidence base is relatively weak, and we think that funding to increase the available evidence could be very useful for the long-term improvement and expansion of internet-based interventions. Funding an existing organization to conduct better M&E or even an RCT could be highly impactful. Important knowledge gaps include the *persistence* of these interventions, RCTs that directly compare them with in-person interventions, the effectiveness of AI chatbot therapy, and the opportunities for counterfactual impact in crowded markets such as self-help apps.

Cost-effectiveness of internet-delivered psychotherapy

Our BOTEC assesses an intervention along the lines of Overcome, described above.

We estimate that a fully scaled intervention of this type has a 12% chance of being more cost-effective than a GiveWell top charity. Although the evidence base in this field is thin, and our calculations rely on some shaky subjective inputs.

Type	Intervention	Cost-effectiveness compared to cash transfers	Proportion of benefits from mental health
Psychotherapy	Internet-delivered psychotherapy	10x	100%

Heuristics for self-administered & tech-based psychotherapy

- Psychotherapy can still be effective when delivered online, by a chatbot, or as a self-administered course
- These modes have the potential to offer treatment at a far lower cost
- Low retention rates are a major threat to effectiveness
- This also increases the cost of finding new patient through marketing
- It is harder to reach people without reliable internet access, so these modes are better aimed at the young and those with a smartphone.

Antidepressants

Summary:

- SSRIs can have mild-to-moderate effects on depression
- Access is very poor in LMICs
- Poor infrastructure could be a barrier to extending reliable access to SSRIs within LMICs

Antidepressants have been shown to reduce symptoms of depression. In a meta-analytic evaluation of meta-analyses (2022), Leichsenring et al. found that the effect size of antidepressants on mental disorders was modest and broadly comparable to psychotherapy (standardized mean difference (SMD) of 0.34 (95% CI: 0.26-0.42) for psychotherapies and 0.36 (95% CI: 0.32-0.41) for pharmacotherapies compared with placebo or TAU).

The validity of research on antidepressants has been called into question. Many trials are conducted by scientists with financial conflicts of interest (Perlis et al., 2005). Publication

bias is considered to be rampant (Turner et al., 2008). Drugs appear to me more effective when newly released, and become less promising over time, indicative of "novelty bias". Antidepressants' strong side effects can undermine blinding in an RCT, and there may be a link between severity of side-effects and estimated effect size (Moncrieff & Middleton, 2018).

Antidepressant use is thought to be 2-4 times higher in HICs compared to LMICs, indicating an unmet need in poor countries (Kazdin et al., 2021). A 2022 study by Rahman et al. spanning 8000 "representative samples" of health facilities in Bangladesh, the Democratic Republic of Congo, Haiti, Nepal, Malawi, Senegal, and Tanzania found that an estimated 8.2% of facilities had amitriptyline and 46.1% had diazepam on the day of assessment (Amitriptyline, an antidepressant, and diazepam, an anxiolytic drug, are classified as essential medications by the WHO).

There are a number of trials exploring the effectiveness of antidepressants in LMICs, but many combine the treatment with psychotherapy and other interventions, which clouds the picture. One such 2010 study in Goa, India, by Patel et al. found that a "psychosocial intervention [plus] antidepressant drugs" was not superior to treatment as usual among private facility attenders, and increased the chance of recovery by 55% among public facility attenders. This is a reminder that the counterfactual quality of treatment is important when considering the benefits of an intervention. A cost-effectiveness study in Nigeria found that older, cheaper antidepressants were significantly more cost-effective than newer, more expensive ones.

There are a number of barriers to successfully improving antidepressant access in poor countries. Shortages of mental health professionals will make diagnoses less accurate, increasing the chances of underprescription for those with severe depression and overprescription for those who are not depressed, as was observed in India (Pillai et al., 2021). Unreliable supplies of medicines at pharmacies and health centers increase the chance that patients will run out of medicine, which can have negative consequences. Poor awareness of mental health will prevent people from seeking help, and people may not identify depression symptoms as a medical issue. Costs can be prohibitive, and the burden of paying for antidepressants may cause more people to cease treatment prematurely.

We suspect that the greatest opportunities for change are in societies where medical infrastructure is relatively strong, and where the causes of antidepressant undersupply can be easily fixed. There may be opportunities to improve government policy or health worker training to ensure that antidepressants are issued when they are needed.

Pain mitigation

Summary:

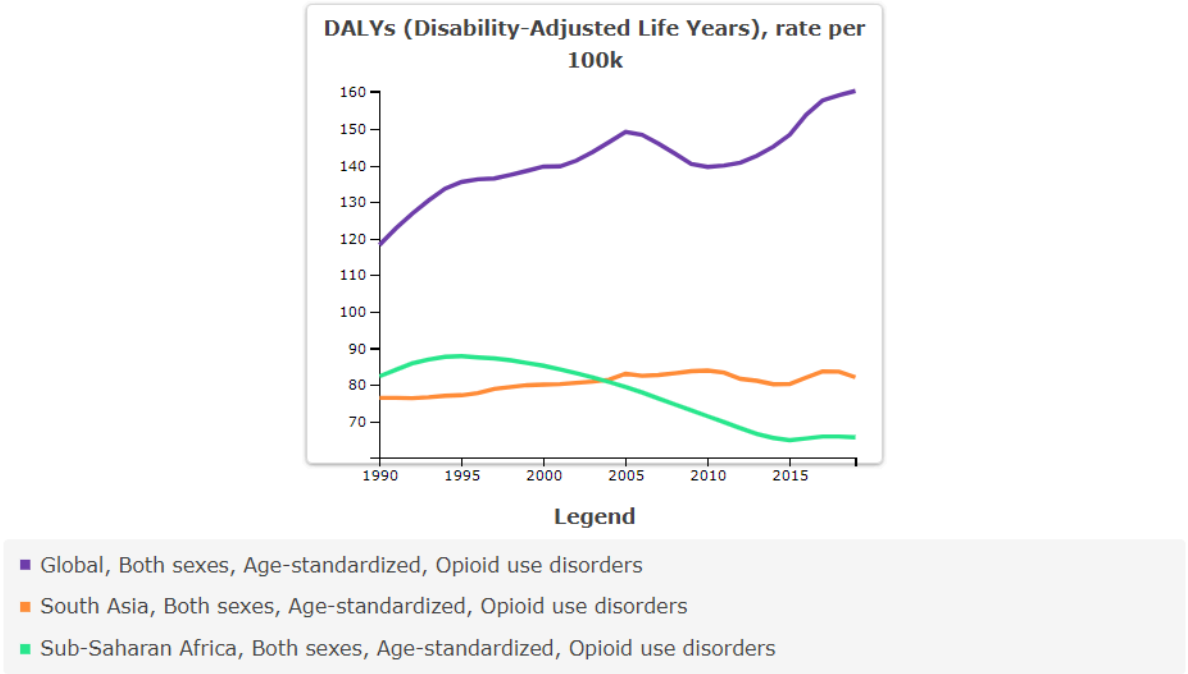
- Lack of awareness and access may be preventing many in LMICs from getting relief from chronic low-level pain
- Lack of opiate access leads to a huge amount of avoidable severe pain
- Increasing access to pain relief drugs comes with major downsides such as the risk of addiction

Pain is bad, and this badness is probably not fully captured by DALYs and QALYs. Pain may also increase the severity of mental illness: a WHO study found that those experiencing persistent pain were three times likelier to suffer from anxiety and depression than others (AOR: 4.14) (Gureje et al., 1998).

The prevalence of chronic pain in developing countries has been estimated at 18% (Sá et al., 2019), and yet the WHO estimates that over 80% of the global population do not have access to treatments for moderate to severe pain (Morriss & Roques, 2018). 50% of the world's poorest people receive just 1% of the available opioids (Yao et al., 2023).

The treatment gap for mild pain relief is less extreme. A 2021 scoping review on over-the-counter analgesics (painkillers) in various Sub-Saharan African countries by Kawuma et al. (2021) found that analgesic use "among pregnant women was between 26–78%, students 55– 82%, general community members 38–97% and hospital patients 45–89%". However, a medical profession working in Uganda suggested to us that arthritis pain, which is very common among the elderly in the country, is severely undertreated.

Increased access to pain medications comes with risk. The GBD survey indicates that the burden of opioid abuse in South Asia and Sub-Saharan Africa is about half the global rate (see graph below, generated at vizhub.healthdata.org/gbd-results/). This is partly thanks to a lack of opiate availability in the medical system: opiate prescription is one of the main pathways to addiction, and medical opiates, when they are available, sometimes end up on the black market. A number of experts we spoke to urged caution when expanding access to opiates because of the danger of addiction and abuse.



Low-level painkillers come with health risks. Even for persistent arthritis pain, the NHS and Mayo Clinic advise only occasional use of painkillers due to the health risks. Nonsteroidal anti-inflammatory drugs (NSAIDs) have been associated with cardiovascular risk (Reddy &

Roy, 2013) and peptic ulcers, and are dangerous for people who already have peptic ulcers, which are more prevalent in poor countries (IHME).

The International Association for the Study of Pain (IASP) recommend "[e]ducation, clinical training, and advocacy initiatives" to improve the mitigation of pain in poor countries. The 1-day Essential Pain Management program "is a simple 1-day workshop aimed at teaching a multidisciplinary group to better recognise, assess and treat pain of all types" and has been delivered in many countries since 2011 (Goucke & Morris, 2012).

We think the best opportunities are lower-risk ways of expanding access to pain relief. Better education and training can increase the safe and responsible allocation of pain relief drugs by healthcare professionals. Increasing opiate access for terminal illness carries less risk of addiction and abuse. Increasing painkiller availability is less dangerous in countries that have robust prescription and restriction systems, and it is possible that helping countries to develop better medicine safeguarding controls will make expanding painkiller access safer and more feasible in the future.

Critical research gaps

As explored in Burden, mental illness is expected to grow in relative importance this century. As the cost–effectiveness of the best physical health interventions drops, the best mental health interventions will become more appealing. We need:

- Studies with long-follow ups to track the persistence of the benefits from the intervention
- Better evidence on spillover effects on the whole household, including data on wellbeing
- Better ways of minimizing response bias when administering symptom-based questionnaires like the PHQ-9
- Data on the counterfactual impact of pain interventions such as increased medication access and training programs

Conclusion

The vast majority of people in LMICs suffering from mental illness have no access to treatment. This makes the counterfactual impact of providing treatment very strong, but it also throws up barriers. A lack of existing infrastructure means that organizations must plan carefully to ensure that their intervention is feasible in the target country. Interpersonal group therapy has been delivered in a way that is mostly independent from existing mental health provision. Improving antidepressant coverage, however, is probably very difficult to do without a deep understanding of existing health structures.

It remains unknown to what extent mental health problems in poor countries are caused by poverty itself. However, research suggests that patients in LMICs *do* respond positively to mental health interventions. Much more research is needed to determine just how powerful these interventions are.

Overall we suspect that *reducing the cost of treatment* is usually the best lever for improving the cost-effectiveness of mental health treatments.

Prevention

Key Findings

- We can clearly identify risk factors associated with future mental illness, but it is very difficult to establish the strength of causation. This makes prevention interventions more uncertain.
- It is possible that policy advocacy promoting healthier lifestyles could cost-effectively improve mental health by preventing diabetes, obesity and physical inactivity
- It is widely believed that early-life mental health interventions are preferable, but we are not sure whether the preventative power of early interventions outweighs the ease of targeting mental illness in adulthood
- We can reduce the burden of suicide by restricting access to the deadliest methods. Pesticide bans have been highly effective, but the marginal impact of extra funding is unclear
- More research is needed on the counterfactual effects of targeting mental health risks factors

Mental illness risk factors

Mental illness correlates with a number of economic, social and health factors. By targeting these factors we can hope to prevent mental illness before it occurs.

It can be very difficult to determine causation, however. All mental health disorders appear to have genetic components which are difficult to isolate from the effects of parenting and the home environment. Mental illness is linked with self-destructive behavior, traumatic life events, loneliness, unemployment and many other things. But we can't say for certain how much an improvement in one of these indicators will improve mental health. Furthermore, these indicators tend to be linked with other factors like ancestry, poverty and poor emotional literacy which may be the underlying cause.

Cash transfers to alleviate poverty

Summary:

- Cash transfers appear to have a small but significant positive effect on mental health
- The effects are strongest when people in poverty are the target
- The mental health benefits of cash transfers do not radically alter the cost-effectiveness of CT programs
- We are confident that directly addressing mental health can mitigate mental illness more cost-effectively than cash transfers

Cash transfer programs vary in the scale and style of the payments they provide, but findings generally suggest that "there may be a positive, albeit modest, relationship between cash transfer receipt and improved mental health" (Evans-Lacko et al., 2023). The effect size

(for SWB or depression/anxiety) seems to be around 0.1-0.15SD (Ohrnberger et al., 2020, McGuire et al., 2022, Wollburg et al., 2023).

The effect is far weaker when the recipients are not living in poverty (Thompson et al., 2022). Single payments may be preferable to regular payments, and the best CT programs communicate with recipients so that they know how much money to expect, and when. The persistence of mental health improvements is poorly understood and, given the small size of the improvement, difficult to measure. Of four studies that had follow-ups after 2-9 years, effects appear to have approximately halved (Wollburg et al., 2023).

In their analysis of cash transfer programs, Happier Lives Institute (HLI) estimate that a CT program costing $1,274 improves recipients' mental health (which they equate to SWB) by 0.4 SD-years, or a total of 1.7 SD-years among the whole household (confidence interval (0.40, 5.94)) (HLI, 2022). As a rough comparison, Givewell estimated that just $105 spent on interpersonal group therapy would improve SWB by 1.1 SD-years across the household.

It seems likely that CT programs have positive spillover effects within the household, but it is also possible that there could be *negative* spillover effects for those who do not receive a cash transfer, who become poorer in relative terms and may feel a sense of injustice. There is little data on this. McGuire et al., 2022 assess four trials that explore non-household spillover effects, of which "[t]wo found negative spillovers but the average effect is not statistically significant and is close to zero."

Overall it seems very unlikely that cash transfer programs can rival interpersonal group therapy for cost-effectiveness on mental health or wellbeing alone. However, proponents of cash transfers are usually interested in the poverty-alleviating effects. The small positive effect on mental health may make cash transfer programs slightly more appealing than they previously were.

Mental health effects should also be accounted for when considering introducing maternity pay, pensions or UBI for populations in poverty.

Our BOTEC estimates that a campaign to persuade a poor country to implement a child poverty grant (in the style of South Africa's CSG) has a 46% chance of being more cost-effective than a GiveWell top charity. However, only around 3% of the expected benefits would be from mental health.

Type	Intervention	Cost-effectiveness compared to cash transfer	Proportion of benefits from mental health
Social support	Child poverty grant advocacy	30x	3%

Our results are highly contingent on the counterfactual value of government spending, of which we are highly uncertain. In short, it is unclear to us how impactful child poverty grants would be in a low-income country *compared to the things the government currently spends money on*.

Alcohol

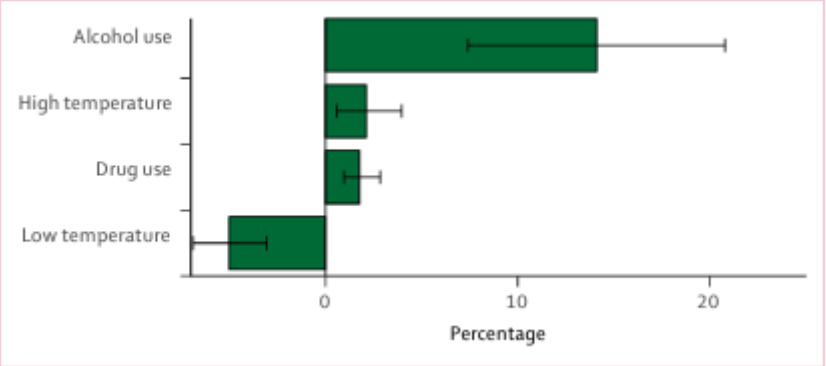
Summary:

- The association between alcohol abuse and mental illness is strong
- However, the strength of the causal link is unclear

Alcohol use and abuse is associated with poor mental health. “The prevalence of anxiety, depression, and other psychiatric disorders is much higher among persons with Alcohol Use Disorder (AUD) compared to the general population” (NIH).

Negative spillover effects of alcohol abuse are likely to add to the burden. Dawson et al. (2007) found that “women whose partners had alcohol problems were more likely to experience victimization, injury, mood disorders, anxiety disorders, and being in fair or poor health than women whose partners did not have alcohol problems”. Even after accounting for the women’s increased rate of substance use, odds ratios for the above indicators were 2.1-3.4.

According to the GBD, alcohol use is the largest risk factor for suicide (which forms the vast majority of the burden of “self-harm”), accounting for 15% of the burden.



Percentage of Self-Harm DALYs attributable to top risk factors for both sexes combined (GBD, 2019).

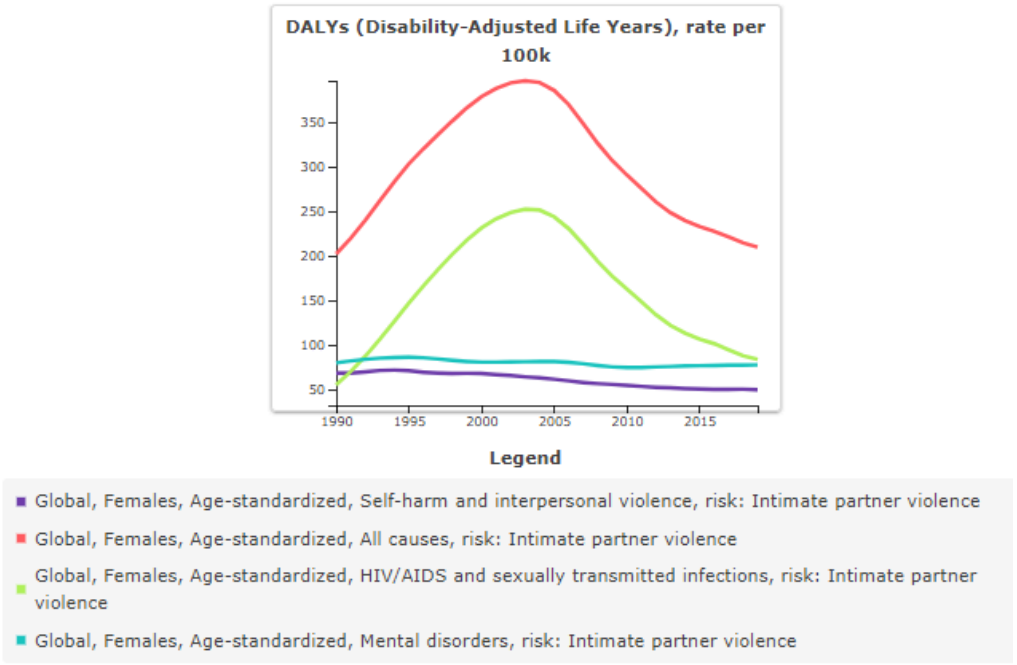
The strength of the causal relationship is difficult to determine, as alcohol is often used as a “coping mechanism” by those suffering from mental illness. We believe that policies that reduce alcohol consumption, and especially abuse of or addiction to alcohol, will improve mental health outcomes.

Intimate Partner Violence (IPV)

- IPV is strongly associated with poor mental health
- Depression will soon overtake HIV/AIDs as the primary health burden associated with IPV
- However, the strength of the causal link is unclear, as there are other possible factors at play
- IPV is notoriously difficult to study and to tackle, although a recent mass media intervention shows early promise

Intimate Partner Violence (IPV) “is defined as any lifetime experience of physical or sexual violence perpetrated by a current or former intimate partner, and is estimated among females only” (IHME).

The data is extremely unreliable due to the difficulty of measuring something that is usually hidden in the home and is a taboo topic. The GBD survey measures IPV as a binary score, so we are unable to differentiate between different types of IPV. According to the GBD, mental disorders are becoming the main source of health burden associated with IPV, as the global burden of HIV/AIDS declines.



Data from GBD study (2019)

IPV is estimated to be responsible for 11.0% (95% UI 0.1–24.0) of DALYs due to depressive disorders among females, increasing each individual’s risk of major depression by an estimated 54%. However, the GBD data implies an extremely low level of certainty on the mental health burden of IPV. The Lower and Upper estimates for the global depression burden attributable to IPV *differ by a factor of 500x*, which is the highest we have seen for any cause.

IPV is associated with 0.53 suicides (deaths from self-harm) and 1.58 other deaths (presumably at the hands of their partner) for every 100,000 women per year. It has been associated with increased rates of postpartum depression (Beydoun et al., 2012), and may have persistent effects on the children of victims.

Links between IPV and mental health outcomes should be seen as tentative. There are a host of potential confounding factors associated with IPV which may also influence mental health outcomes.

IPV is difficult to prevent. Education programs may be effective, but studies typically measure improvements in knowledge and self-proclaimed beliefs on the subject, not

changes in the incidence of IPV, so the effectiveness is unknown. Bettle (2022) suggests that mass media interventions are particularly effective "where unhelpful social norms (such as the acceptability of intimate partner violence) persist", hypothesizing that funding an IPV mass media campaign could be significantly more cost-effective than donating to GiveWell. Fairless (2023) recommended that Charity Entrepreneurship incubate a charity pioneering an "entertainment-based" mass media intervention to reduce Intimate Partner Violence. They caution that monitoring and evaluation will be a challenge for a sensitive subject "that presents additional challenges for outcome verification."

Cost-effectiveness of a radio campaign to prevent IPV

Charity Entrepreneurship have identified a radio campaign to combat IPV as a promising idea. Similar campaigns have been delivered by Development Media International and Family Empowerment Media to encourage breastfeeding and contraceptive use in Sub-Saharan Africa. Radio campaigns may be most helpful "where unhelpful social norms (such as the acceptability of intimate partner violence) persist" (Bettle, 2022).

Our BOTEC estimates that a radio campaign to prevent IPV would have a 4% chance of being more cost-effective than a GiveWell top charity. However, it is likely that the GBD, our source for the health impacts of IPV, severely underestimates the full harm associated with IPV. The GBD attempts to capture the increased risk of depression, violent death and injury, and increased HIV transmission associated with IPV, and not stress, shame, fear and pain.

We are also highly uncertain about the cost of the campaign and the power it would have to change perpetrators' decision-making. We do not attempt to model positive or negative secondary effects. Mitigation of physical impacts constitutes around 70% of the benefit.

Type	Intervention	Cost-effectiveness compared to cash transfers	Proportion of benefits from mental health
Mass media	Radio campaign to combat Intimate Partner Violence (IPV)	5x	30%

Religion

Summary:

- Religious belief is associated with better mental health, although it is difficult to convert this insight into intervention recommendations
- Religious organizations have several properties which make them well-placed to implement interventions in LMICs

Religious people seem to be more resilient to mental illness (Dein, 2010, LondonGal, 2023). This could be because of strong community support and spiritual practices that encourage positive emotions (Mueller et al, 2001).

Although we don't endorse encouraging people to be more religious, there are signs that a religiously-aligned organization could be well-placed to improve mental health in deeply religious societies.

- Members of the religious community hold a lot of sway which could be used to increase awareness of mental health. Experts have suggested to us that people in poor countries are less likely to consider mental illness to be a major health problem, and may equate mental illness with conditions like severe schizophrenia, but not depression and anxiety.
- Marketing costs are a potential barrier to online or tech-based psychotherapy interventions. Religious groups may have strong networks which provide "free marketing" and could increase the chances of reaching a wide audience
- People often give to causes within their faith, which could make fundraising easier for a religious organization
- Religious groups often have a tradition of volunteering, which can drive down labor costs.

Many of these points stemmed from a conversation with the founder of Wailing Women, a Christian charity. WW grew a global audience during lockdown through social media, and produced uplifting live events, group events and courses of therapy.

There are a number of reasons to be wary about supporting religious organizations. They may only help people with the same religious affiliation, or they may allocate resources to religious activities with no obvious health or wellbeing benefits. It's also possible that people giving time and money to religious charities may be manipulated into doing so.

Physical activity

- Physical activity is associated with lower rates of developing depression and less severe symptoms of depression and anxiety, although the direction of causality is unclear
- Government health campaigns may be able to prevent cases of depression by encouraging physical activity

There is an established correlation between physical incidence and both incidence and severity of mental disorders. Schuh et al., 2018 examined 1.8m person-years of data and found that compared with people with low levels of physical activity, those with high levels had lower odds of developing depression (adjusted odds ratio=0.83, 95% CI=0.79). A Systematic Review and Meta-analysis by Pearce et al. (2022) found that "relative to adults not reporting any activity, those accumulating half the recommended volume of physical activity had 18% (95% CI, 13%-23%) lower risk of depression. Adults accumulating the recommended volume had 25% (95% CI, 18%-32%) lower risk with diminishing potential benefits and higher uncertainty observed beyond that exposure level".

Rebar et al., 2015 is a "meta-meta-analysis of the effect of physical activity on depression and anxiety in non-clinical adult populations". Physical activity had a moderate effect on depression [standardized mean difference (SMD) = -0.50; 95% CI: -0.93 to -0.06] and a small effect on anxiety (SMD = -0.38; 95% CI: -0.66 to -0.11). An overview of systematic reviews by Singh et al. (2022) found that physical activity had medium effects on depression

(median effect size=−0.43, IQR=−0.66 to –0.27) and anxiety (median effect size=−0.42, IQR=−0.66 to –0.26).

These results should be taken with a large grain of salt. Depression is associated with low energy, poor sleep and diminished motivation. It is likely that those experiencing *mild or undiagnosed* depression are less physically active. When they later cross the threshold to diagnosed depression, this may be attributed to low physical activity rather than their previous state of depression. Exercise is known to improve sleep, and poor sleep is known as a symptom of depression. It could be that some of the observed effects of exercise on depression symptoms stem from sleep improvements rather than improvements in "underlying" depression.

Our BOTEC, which relies on mostly subjective estimates for inputs, estimates that advocacy for a physical activity campaign has a 21% chance of rivaling GiveWell top charities for expected cost-effectiveness. We are very uncertain about the cost-per-person and the effect size of physical activity campaigns, and most of the expected value lies in low-cost, high-effect scenarios.

Type	Intervention	Cost-effectiveness compared to cash transfers	Proportion of benefits from mental health
Health nudges	Physical activity campaign advocacy	17x	17%

Obesity

A systematic review by Blasco et al. (2020) "confirmed that there is indeed a link between depression and obesity, although there are doubts as to the significance of this relationship." A A systematic review and meta-analysis of longitudinal studies by Lippino et al. (2010) found that obesity increased the risk of onset of depression (unadjusted OR, 1.55; 95% confidence interval [CI], 1.22-1.98; P < .001), and that the link was stronger for obesity than overweight, suggesting a dose-response relationship. They also found that depression increased the odds of developing obesity (OR, 1.58; 95% CI, 1.33-1.87; P < .001), indicating a two-way relationship.

Diabetes

Diabetes is associated with an increased risk of depression. CEARCH's cost-effectiveness analysis on policy advocacy for sugar-sweetened beverage taxes (Report, CEA) found that 2-3% of cases of depression globally are linked to diabetes. It estimates that 1.4% of the total burden associated with diabetes comes from death and disability linked with depression and suicide. Even though this is a small fraction of the overall burden, the high expected cost-effectiveness of the intervention may make it an appealing candidate on mental health grounds alone.

These results are highly uncertain because of the difficulties in modeling the tractability and effectiveness of government health policy.

Another weakness is that the causal link between diabetes and depression is not fully established. Two meta-analyses have found significant correlation between diabetes and depression (Chireh et al., 2019, Elamoshy et al., 2018). Diabetes was found to increase the risk of suicide by Wang et al. (2017). The suicide link failed to reach significance in two other studies (Wang et al., 2016, Elamoshy et al., 2018), but with risk ratio of 1.61 and odds ratio of 1.85 respectively.

Our BOTEC estimates a 63% chance that the expected cost-effectiveness of SSB tax advocacy exceeds GiveWell top charities. Only 4% of the benefits would come from improved mental health.

Type	Intervention	Cost-effectiveness compared to cash transfers	Proportion of benefits from mental health
Health nudges	Sugar-sweetened beverage tax advocacy	79x	4%

Unwanted pregnancy

Increased contraceptive access and family planning education may improve mental health by preventing the negative effects of unplanned pregnancies.

According to Stevenson et al. (2023), “The prevalence of perinatal depression is estimated to be 12–17% and the prevalence of perinatal anxiety is estimated to be 15% globally among the general population”. In a South Korean study, unintended pregnancy was associated with 20-22% greater odds of maternal depression (compared with intended pregnancy) and with higher depression and stress scores (Bahk et al., 2015).

The counterfactual impact of pregnancy on depression may be exaggerated. Pregnant women and mothers have more contact with the healthcare system and may be more likely to be diagnosed with depression. Many studies on perinatal depression rely on survey methods that tend to find higher prevalence rates[20]. It is not clear to what extent pregnancy increases the risk and severity of depression.

Furthermore, studies probably fail to separate the direct effects of unwanted pregnancy from those of other, correlated, factors. In Bahk et al. (2015), much of the increased risk and severity of depression linked with unplanned pregnancy was explained by higher rates of marital conflict and lower rates of father participation.

Our best guess is that preventing unwanted pregnancies *would* prevent mental illness, but that these benefits would pale in comparison to others (such as reductions in maternal mortality, unsafe abortions and stillbirths (Sully et al., 2019)) and would probably not be cost-effective on their own.

[20] For illustration, a 2022 meta-analysis by Bello et al. found a staggering 48% prevalence of depression in Africa. It relied entirely on studies that used survey data.

Early-life intervention

Summary:

- Interventions aimed at mothers, babies, children and young adults are believed to have preventative mental health benefits
- However, factors like genetics and poverty cloud the picture, undermining the strength of this claim
- Many mental health conditions emerge in adulthood, making targeting children less efficient
- There are well-known childhood risk factors associated with later mental illness, but evidence-based interventions are lacking

Adult mental illness correlates with perinatal depression, childhood trauma and childhood mental illness. There is a widespread conviction that early-life interventions are therefore preferable to adult interventions, but we believe that this is not always true.

Most mental disorders begin in adolescence or adulthood (Kessler et al., 2007). Mental disorders are typically not diagnosed or treated until years after the symptoms begin, and earlier diagnosis could lead to better outcomes. Kessler et al. (2007) suggest that "[a]s many mental disorders begin in childhood or adolescence, interventions aimed at early detection and treatment might help reduce the persistence or severity of primary disorders and prevent the subsequent onset of secondary disorders", although "research is needed on treatments of early cases [...] to determine whether this is true."

It's possible that tackling childhood risk factors associated with adult mental illness could have long-lasting benefits. A number of key risk factors have strong correlation with adult mental illness, but most studies fail to completely control for the effects of genetics.

Childhood trauma, childhood adversities and childhood sexual abuse are all correlated with poor mental health outcomes in childhood. (McKay et al., 2018) found that exposure to bullying, emotional abuse, maltreatment and parental loss were associated with adult mental disorder, with odds ratios varying from 1.24 to 2.09. Data from the GBD survey suggests that bullying victimization "accounted for 7.1% (2.2–14.4) and 4.6% (1.1–9.6) of YLDs for anxiety disorders and major depressive disorder (MDD), respectively". Childhood sexual abuse was found to be "responsible for 4.4% (95% UI 2.4–6.8) of global YLDs due to depressive disorders and 9.6% (1.3–22.1) of global YLDs due to alcohol use disorder." Kessler et al (2010) estimate that eliminating childhood adversities would cause a 30% drop in mental disorders.

We see reasons to be skeptical of these estimates. Child abuse is difficult to measure. When asking parents if a child has been abused, they may have reason to lie to protect the perpetrator. When asking the child, they may be dishonest or they may not know that they have been abused. Even when asking adults whether they were abused as a child, they may be more likely to remember abuse if they believe it affects their mental health as an adult. This not only undermines the correlation data, but hampers feedback loops for any interventions aimed at reducing the incidence of childhood adverse experiences. Furthermore, adverse childhood experiences could be associated with factors such as genetics and poverty which may themselves have causal effects on adult mental illness.

Interestingly, there appears to be a genetic component in the effect that maltreatment has on a child. One study found that holders of the MAOA gene had a much stronger correlation between childhood maltreatment and antisocial behavior (Taylor, et al., 2007).

The earliest interventions target pregnant mothers and mothers of babies. A number of educational programs aimed at mothers and young children appear to have effects on behavior and child development & mental health (Klasen & Crombag, 2012). But data on long-term mental health outcomes is simply not there.

Early adolescence may be a good time to tackle behavioral problems and teach emotional regulation. The evidence is nicely summarized in Javier, 2022, but there is no convincing confirmation that interventions in adolescence have long-term effects.

The key weakness of early-life interventions is that *most mental disorders have not yet emerged*, so it is difficult to target children accurately. Risk factors can improve targeting, and it is well understood that prioritizing children with risk factors can yield better results (Yu et al., 2023). But targeting before symptoms have emerged can only be so good.

On the other hand, the link between childhood and adult mental health is inherently difficult to study, and we remind readers that absence of evidence is not evidence of absence.

Early-life interventions have a number of strengths. Mothers are relatively easy to reach through medical appointments during pregnancy. Schools provide a convenient platform for educational interventions, and can be a useful ally in early diagnosis of risk factors and mental disorders. It seems possible that policy advocacy could unlock and direct government funding towards youth-targeted mental health prevention work. Javier, 2022 suggests prioritizing “interventions focusing on stress, strengths, values, problem-solving etc.--more familiar terms than mental health.” It seems plausible that such interventions could be integrated into school curricula.

It is not clear whether the pros of pre-emptive early-life interventions outweigh the cons. Early diagnosis, however, would be entirely positive. There may be scope for school or online services that can identify signs of mental illness sooner, and provide treatment several years earlier than it would have been.

Preventing suicide

We think that the top ways of preventing suicides are to treat and avert cases of depression (see Treatment) or to restrict the means of suicide. It is accepted in the field that blocking highly-reliable methods of suicide forces people to turn to less reliable methods and increases their chance of surviving the suicide attempt. The effectiveness of means-restriction depends on how determined to commit suicide the person is. An expert we spoke to split suicide attempts into two categories: those that are the product of persistent, severe depression; and those that result from a passing moment of despair. Means restriction is considered to be effective at preventing the latter, “low-intent” type of suicide.

Famous examples of this effect include the drop in suicides after carbonoxide was removed from the UK gas supply (Kreitman, 1976) and the drop in suicides in response to various pesticide bans in Sri Lanka (Gunnell et al. 2007 & Knipe et al., 2017).

It is uncertain how highly to value an averted suicide, given that suicide is associated with depression and suffering. But we don't think this uncertainty is one of the main concerns when assessing suicide-prevention measures.

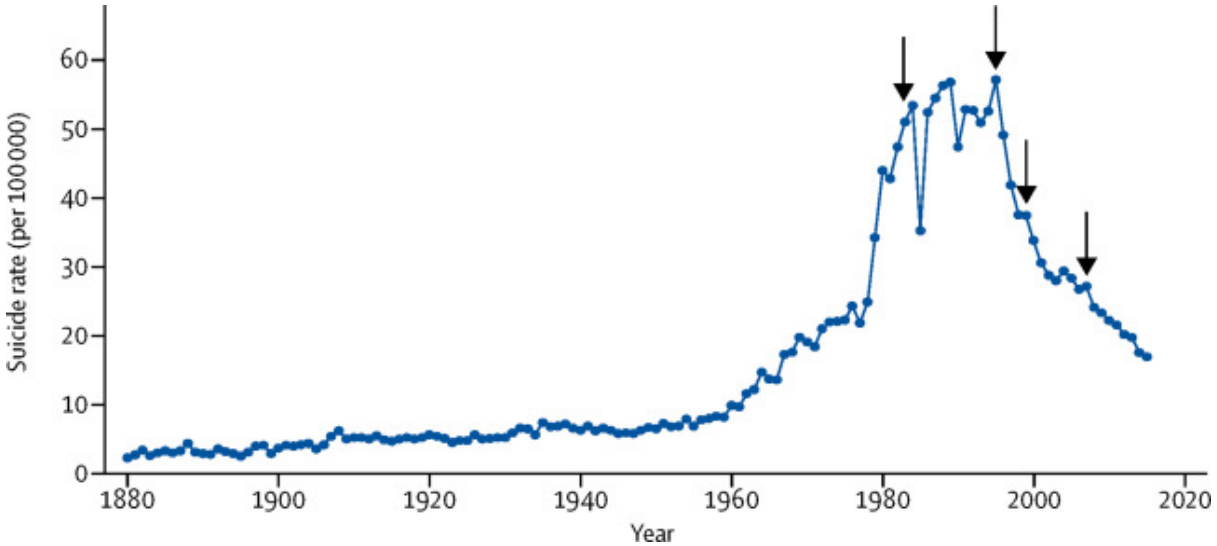
Pesticide bans

Summary

- Millions of people have committed suicide with pesticides since 1960
- Banning the most toxic substances is associated with drops in overall suicide rates
- Progress has been achieved by a small organization in the space
- Most of the best opportunities have been taken, but there is still scope for highly cost-effective gains by targeting smaller countries or by aiding enforcement efforts

During the Green Revolution, highly toxic pesticides became widely used across much of the world. They became a popular means of suicide in poor countries, and are thought to have been used in between 9 and 17 million suicides since 1960[21] (Karunarathne et al. 2019).

In Sri Lanka, for example, suicide rates exploded as pesticide use increased across the country. Prof Michael Eddleston of Edinburgh University claims that turning points in the national suicide rate can be attributed to new substance bans[22] (Knipe et al., 2017).



Similar but less drastic effects have been observed in Nepal (Utyasheva et al., 2021), India, Taiwan and South Korea (Bonvoisin et al., 2021). In China, pesticide suicides dropped by 60.5% between 2006 and 2018, during which time 12 deadly pesticides were banned or restricted.

[21] Although suicide record-keeping has been, and continues to be, poor across much of the world.
[22] The reader should be aware that the increase and decrease of the suicide rate also coincides with the beginning and end of the Sri Lanka civil war. Aida et al. found that suicide rates were not higher in the districts most affected by war, suggesting that the war does not account for the rise and fall of the national suicide rate.

One challenge is that most pesticides are somewhat toxic, and so removing one substance causes people to move to another, still toxic, pesticide. Leah Utyasheva and Michael Eddleston of The Centre for Pesticide Suicide Prevention see this as an iterative process where each round of bans helps to identify the substances that should be banned in the next round. They insist this does not continue indefinitely, since controls on new substances have improved over time and hence once all of the worst chemicals from the early days of the Green Revolution have been restricted, the risk becomes very low.

They also believe that most pesticide suicides come from moments of temporary despair and that preventing them often leads to the beneficiary living a normal life.

It has been suggested that banning pesticides can have serious downsides by reducing the agricultural output of poor, food-insecure countries. In a BOTEC, GiveWell subjectively discounted the effectiveness of pesticide bans by 30% to account for decreased agricultural output. Sethi et al. (of which Michael Eddleston was a co-author) claim to have found that the 2011 pesticide bans in India seemed to have "negligible impact" on crop yields on the 2012 harvest, although it seems likely that the study would only have detected very large effects.

The Centre for Pesticide Suicide Prevention (CPSP) have received $8millon from Open Philanthropy. They focus on advocacy for pesticide bans, plus help with finding safer substitute pesticides and with enforcement of new bans. They argue that they have been instrumental in a number of bans. Although they have a team of 20 working in four regions, they say they are funding-constrained, and would like to build a larger team that could operate in more countries. Open Philanthropy recently established Global Health Policy as a new program, naming "suicide prevention through means restriction" as one of the four focus areas. This suggests to us that funding may become easier to acquire in the space, and additional philanthropic funding may become less impactful.

We think that pesticide suicides have probably peaked and begun to decline globally. India and China accounted for most pesticide suicides globally and have seen large declines in recent decades. Targeting smaller countries is likely to be less cost-effective, and yet may be an extremely powerful intervention. It is not clear whether CPSP would use extra funds efficiently: they "are guided by the human rights-based approach" and make no explicit commitment to maximizing the good they do.

It may be possible for other organizations to complement the work of CPSP. Ideas include:

- Helping governments with enforcement[23] of pesticide bans
- Identifying countries with a pesticide suicide problem. Record-keeping is notoriously bad in many African countries, where suicide is taboo and often criminalized. CPSP often cross-reference police, hospital and survey data in an effort to estimate true rates

[23] Poor enforcement can undermine the effectiveness of a ban. In Nepal, which has a porous border with India, methyl-parathion was found to be the most common pesticide used for suicide despite having been banned years earlier.

Our BOTEC attempts to account for the fact that much of the most important advocacy has been done or is being done, and we are highly uncertain about the marginal value of extra work in this space. We attempt to account for this by examining the effect of advocating a small country (population 10 million), as these are less likely to have already been targeted, and assuming a relatively low baseline rate of pesticide suicides (as may be expected from a country that has already banned some pesticides). We estimate that the expected cost-effectiveness of pesticide restriction advocacy has an 8% chance of exceeding GiveWell top charities' cost-effectiveness. It seems plausible that advocacy in a well-chosen target country could outperform GiveWell top charities.

Type	Intervention	Cost-effectiveness compared to cash transfers	Proportion of benefits from mental health
Suicide prevention	Pesticide ban advocacy	7x	100%

Lithium supplements

Summary:

- Lithium is widely prescribed as a mood stabilizer and is associated with lower rates of suicide
- There is evidence that populations with naturally-occurring lithium in their drinking water have lower suicide rates
- If successful, policy advocacy for adding lithium to the water supply could cut suicide rates across the population
- There are a number of positive/negative effects that are difficult to predict

Lithium is well-known as a mood stabilizer that is widely prescribed for Bipolar disorder, and is associated with lower rates of suicide. A meta analysis by Cipriani et al. in 2013 found that patients given lithium had significantly lower rates of suicide than those given a placebo (odds ratio 0.13, 95% confidence interval 0.03 to 0.66). Lithium treatment comes with the risk of damage to the thyroid gland and the kidneys, and patients must have their serum lithium levels monitored in order to minimize the danger (Young & Hammond, 2007).

A number of studies have explored the link between levels of naturally-occurring lithium in drinking water and suicide rates in the population. These levels of lithium exposure are far lower than among those taking lithium medication, and it is unknown whether lithium has a physiological effect at such low doses.

Generally there seems to be a trend that areas with "high enough" lithium levels often do display significantly lower suicide rates. Knudsen et al. (2017), a 22-year study in Denmark, where lithium levels range from 0.6 to 30.7 µg/L, found "no significant indication of an association" between lithium and reduced suicide rates. However, they point out that "[p]revious studies that found a significant association with suicide consistently reported the highest lithium exposure levels with up to 59 µg/L in the Oita prefecture in Japan [16], 121 µg/L in Greece [13], and 219 µg/L in Texas [7]. Conversely, studies with the lowest levels up to a maximum of 12.9 µg/L in the Aomori prefecture in Japan [6] and 21 µg/L in the east of England [15] did not find an association, like in our present study in Denmark."

The two existing meta-analyses on the subject find that lithium in drinking water reduced suicide risk. Barjasteh-Askari et al. (2020) find an odds ratio of 0.42 (95% CI: 0.27 to 0.67; p-value <0.01), while Memom et al. (2020) get a pooled β = −0.27 (95% CI −0.47 to −0.08; P = 0.006, I2 = 83.3%). Both studies found a stronger effect among males. It should be noted that all studies apart from Knudsen et al. (2017) rely on regional suicide rates and therefore fail to account for individual exposure to lithium and for underlying differences between populations.

Psychiatrist Dr Moosajee Bhamjee openly called for lithium to be added to Irish drinking water in 2011, but no country has yet begun to do so. On the release of Memom et al. (2020), Professor Anjum Memom suggested that "[n]ext steps might include testing [the] hypothesis by randomised community trials of lithium supplementation of the water supply, particularly in communities (or settings) with demonstrated high prevalence of mental health conditions, violent criminal behaviour, chronic substance abuse and risk of suicide. This may provide further evidence to support the hypothesis that lithium could be used at the community level to reduce or combat the risk of these conditions."

Araya et al. (2022) assesses the feasibility of introducing a policy to add lithium to drinking water to decrease suicide risk. They provide an excellent summary of pros and cons, including the possible negative effects of increased suspicion about government motives (as we have seen with the water floridation "controversy") and increased use of bottled water.

Our opinion is that regional trials could provide evidence that may motivate future country-level interventions to add lithium to the water supply. It would be best to trial this in a country with high trust in the government to minimize downside risks. The relative rarity of suicide and the possible slow-acting effects of low-level lithium means that such trials would take years to gather results.

This is the most speculative intervention we evaluate, so any calculations should be interpreted accordingly. Our BOTEC finds that advocating for a small country to pilot lithium supplementation could have large benefits. The expected benefits increase further when we consider that promising experimental results may prompt other countries to begin supplementing their water supplies with lithium. We estimate that the expected cost-effectiveness exceeds GiveWell top charities in 40% of cases.

Type	Intervention	Cost-effectiveness compared to cash transfers	Proportion of benefits from mental health
Suicide prevention	Lithium supplementation study	38x	100%

Attitudes to suicide

Many LMICs have no strategies for preventing suicide. Although the benefits would be hard to predict, it seems highly likely that "helping" governments to strategize could be highly impactful:

- Decriminalizing suicide would make it much easier for suicidal people to seek help
- Centralized record-keeping would illustrate the scale of the problem and would provide a gauge of the effects of suicide-reduction efforts

- Countries with a suicide reduction strategy would be more likely to endorse restrictions on deadly pesticides and other means of suicide.

Critical research gaps

- The *direct* mental health effects of tackling risk factors. Too much of our understanding is correlational, and we may be overestimating the effects that tackling risk factors would have. It would be good to have RCTs that evaluate the mental health impact of interventions which target physical health, or to simply start including mental health metrics among the variables measured in future poverty and physical health work
- The effect of mother/child educational interventions on mental health in adolescence and adulthood (most studies only measure change in knowledge and attitudes, and have short follow-ups (see Klasen & Crombag, 2013). This would require studies spanning many years, which are expensive and difficult to conduct.

Conclusion

Mental illness is clearly linked with adversity in childhood, poor physical health, and being the victim of abuse. These correlations can be valuable for targeting groups who are more likely to be present or future sufferers. Especially with children, risk factors can help us help people before mental illness manifests. However, preventative interventions will always be hits-based: some efforts will be spent on people who were never going to suffer mental illness.

There is an implicit assumption that preventative interventions targeting mothers and children have an edge over those that target adult sufferers. In the absence of long-term studies that track the impact of early-life interventions, we are unable to endorse this assumption.

It would certainly help to diagnose mental illness sooner. But we suspect that *truly preventative* interventions are best when they are cheap and scalable. For example: policy changes that encourage healthy lifestyles, or new laws that restrict access to deadly means of suicide.

We often don't know how much of the correlation between risk factors and mental health is driven by genetics or other hidden variables. In order to understand which "nudges" best promote mental health, we need long-term studies to evaluate the counterfactual impact of tackling risk factors.

Intervention BOTECs

[Link to full distributional model]

Here we provide estimates of the cost-effectiveness of various mental health interventions. These should be considered as preliminary “best guesses”, as most of the inputs are either subjective estimates or are taken from others’ analyses.

- We estimate the costs of a *hypothetical* intervention, and generally assume that it is well-run and operating at scale.
- We *discount* costs borne by governments, on the basis that these funds are redirected from other types of government expenditure that are relatively inefficient
- We give cost-effectiveness in terms of DALYs. For psychotherapy and child poverty grant advocacy, we had to convert the effect size from measures of mental health into DALYs, which introduces considerable uncertainty
- As always, users should consider how their moral standpoint affects these results, especially when comparing mental health interventions against GiveWell top charities.

How it works

This spreadsheet model is put through UseCarlo.com, which performs a Monte Carlo simulation on Carlo with 1500 runs. Each input (eg. the effect size of an intervention, or the cost per person treated by an intervention) is modeled as a random variable. For each run, the model takes a sample of each random variable and computes the corresponding cost-effectiveness. These 1500 cost-effectiveness estimates can be used to understand the uncertainty in the results. To adapt the model, simply make a copy of the spreadsheet, change the inputs, then re-run the simulation. The outputs used to produce the table below are available in spreadsheet form here.

The next page provides a summary of the results. After that is an analysis of the key trends.

The figures below are mostly derived from subjective estimates and should be used as a guide only!

See the next page for an explanation of each column.

Type	Intervention	Cost-effectiveness, DALYs per $100,000						Proportion of benefits from mental health
		Mental health benefits only				All health benefits		
		Expected	Quantiles			Expected (DALYs per $100,000)	Expected (multiple of cash transfers)	
			0.1	median	0.9			
Psycho-therapy	Interpersonal group therapy, LMIC[24]	183	33	138	379	183	5×	100%
	Internet-delivered psychotherapy[25]	383	22	153	852	383	10×	100%
Social support	Child poverty grant advocacy[25] *(Country of population 40 million)*	30	4	16	67	1110	30×	3%
Mass media	Radio campaign to combat Intimate Partner Violence (IPV)	55	1.7	16	119	186	5×	30%
Health nudges	Physical activity campaign advocacy *(Country of population 40 million)*	106	0	18	251	622	17×	17%
	Sugar-sweetened beverage tax advocacy *(Country of population 40 million)*	114	6	44	271	2906	79×	4%
Suicide prevention	Pesticide ban advocacy *(Country of population 10 million)*	263	6	64	554	263	7×	100%
	Lithium supplementation study	1413	24	499	3648	1413	38×	100%

[24] Psychotherapy and social support interventions were initially measured in SD-years of depressive/affective mental health symptoms, and converted to DALYs with the approximate conversion rate 1 SD-year = 0.14 DALYs. The conversion rate was uncertain, with 90% confidence interval (0.093,0.203). Roughly, this equates curing severe depression for one year with averting 0.65 DALYs.

How to read the table:

- **We model the expected ex-ante cost-effectiveness of each intervention.** This means we are estimating the *average* cost-effectiveness of a future action. For example, the ex-ante cost-effectiveness of a game of roulette is $0.97 per $1 invested, while the ex-post cost effectiveness can only be $0 or $36 per $1 invested. It should be noted that some of the interventions in the table above (those that depend upon binary policy decisions) are likely to have an ex-post cost-effectiveness of zero.
- **Mean** cost-effectiveness is the mean of all cost-effectiveness outputs. The **Median** cost-effectiveness is the middle value among the outputs. 10% of outputs are lower than the **0.1** quantile, and 90% of outputs are lower than the **0.9** quantile.
- According to our estimates, the average cost-effectiveness among GiveWell top charities is 737 DALYs per $100,000.
    - **Mean (multiple of GW top charity)** simply expresses the mean cost-effectiveness as a multiple of our GW estimate.
    - **Probability of exceeding GW top charity** gives the proportion of Monte Carlo outputs that exceed 737 DALYs per $100,000. We are not certain what probability counts as "good", but we would point out that GiveWell top charities are themselves uncertain, and that we would expect most of them to score well below 50% on this metric[25].

Analysis

- **Mean cost-effectiveness is driven by a minority of highly cost-effective possible scenarios.** Mental health interventions are uncertain, so funders should be discerning when considering grant applications. Even for promising-looking proposals there will be a high degree of uncertainty, so funders should be willing to take a "hits-based" approach[26].

- **Cost is critical.** For all interventions, *cost* was among the most important variables[27]. Reducing cost is critical for building a better intervention, and reducing *uncertainty* about costs is key to making better funding decisions.

- **Favorable conditions are required in order for interventions to rival GiveWell top charities.** Reducing the cost of interventions is critical for direct work. For policy interventions, working with the most tractable governments and/or the largest countries will boost the expected cost-effectiveness considerably. Lithium supplementation seems to be highly promising, but it is the least-understood intervention on the list.

- **Policy interventions hinge on the likelihood of advocacy success.** This is a very difficult thing to measure. We estimate that the probability of success is 11% (90% confidence interval: (1%,30%)) for all policy interventions. Opportunities that are more

[25]This is for two reasons. Firstly, the expected cost-effectiveness of an intervention is usually driven by the right tail, and so *most outcomes* are below the expected cost-effectiveness. Secondly, the cost-effectiveness of GiveWell top charities varies, and most have expected cost-effectiveness *below* the spending-weighted average across all top charities.
[26] We would expect most interventions to be heavy-tailed in this way, however, so this should not be taken as a sign that mental health interventions are inherently more uncertain than others.
[27] We find that the results of the uncertainty analysis in Carlo change considerably between runs. However, cost variables are consistently among those responsible for the most uncertainty.

promising than this should be valued more highly. In reality, successful lithium supplementation advocacy (for which there is no historical precedent) is probably less likely than successful SSB tax advocacy.

- **The best mental health interventions may not directly target mental illness.** Health "nudge" policies may be so powerful that their net effects on mental health are greater than targeted interventions.