Cost-effectiveness analysis: StrongMinds

Joel McGuire

October 2021

Happier Lives Institute

Contents

Executive summary

This analysis estimates the effectiveness of StrongMinds' core programme of group interpersonal therapy (g-IPT) by combining the evidence of g-IPT's effectiveness with the broader evidence of lay-delivered psychotherapy in LMICs. We then expand our analysis to include StrongMinds' other psychotherapy programmes. Finally, using StrongMinds' average cost to treat an individual's depression, we estimate the total effect a $1,000 donation to StrongMinds will have on depression. We then compare this to a $1,000 donation to GiveDirectly, an organisation which provides cash transfers. Outcomes are assessed in terms of standard deviation changes in measures of affective mental health disorders (MHa). Currently, we estimate that $1,000 donated to StrongMinds would improve MHa by 12 SDs (95% CI: 7.2, 20) and be about 12x (95% CI: 4, 24) more cost-effective than GiveDirectly.

1. Aim of report

The aim of this report is to accomplish two goals. The first is to estimate the cost-effectiveness of StrongMinds' psychotherapy programmes in terms of standard deviations of improvement in MHa scores. The second goal is to compare this cost-effectiveness to other global development interventions such as cash transfers. We hope that this report will inform decision-makers about the potential cost-effectiveness of StrongMinds and motivate further research into mental health treatments.

This is the first HLI report that specifically reviews a charity, StrongMinds. Strongminds implements psychotherapy. We review psychotherapy in LMICs more broadly and mental health as a cause area in our reports on the topics (HLI, 2020a; HLI 2020b). To find out more about the wider project, see Area 2.3 of our Research Agenda and Context.

2. What is StrongMinds?

StrongMinds is an NGO that treats women's depression using several programmes that deliver group interpersonal psychotherapy[1] (g-IPT) in person or over the phone. Interpersonal therapy aims to treat people's depression by increasing social support, decreasing the stress of social interactions, and improving communication skills (Lipsitz & Markowitz, 2013). StrongMinds deploys g-IPT over 12 weeks in roughly 90 minute sessions. The 12 weeks are broken into three official phases that roughly correspond to the three aforementioned paths through which IPT aims to treat depression. These three official phases are sometimes followed by a longer unofficial phase where the groups continue to meet and support one another without the presence of an official facilitator.

StrongMinds delivers g-IPT in several programmes deployed in Uganda and Zambia. These programmes vary by:

- the type of training the facilitators who lead each group receive (formal training and certification or apprenticeship),
- the population they target (adult women or adolescents)
- the format of the group discussion (either in person or over the phone)
- the role StrongMinds plays in delivering the therapy (directly or through partners)

---

[1] Psychotherapy is a relatively broad class of interventions delivered by a trained individual who intends to directly and primarily benefit the subject person's mental health (the "therapy" part) through discussion (the psych part). Cuijper et al. wrote a brief and helpful summary of the various types of psychotherapy for their database of studies on psychotherapies effects on depression (2020).

In the core programme, formally-trained mental health facilitators deliver face-to-face g-IPT to adult women, but StrongMinds began testing providing g-IPT over the telephone in response to the COVID-19 pandemic. We discuss the programmes in more detail at the beginning of the cost-effectiveness analysis (CEA) section.

3. What did we do?

This report builds on previous work by Founders Pledge to estimate the cost-effectiveness of StrongMinds (Halstead et al., 2019) in three ways. First, we combine evidence from the broader literature of psychotherapy with the direct evidence of StrongMinds' effectiveness to increase the robustness of our estimates. Secondly, we update the direct evidence of StrongMinds' costs and effectiveness to reflect the most recent information. Lastly, we design our analysis to include the cost-effectiveness of all StrongMinds' programmes. Fourth, we use Monte Carlo simulations which should better account for and convey uncertainty. This allows us to compare the impact of a donation to StrongMinds to other interventions such as cash transfers delivered by GiveDirectly.

We focus on estimating the cost-effectiveness of the core programme, which is face-to-face g-IPT, because of its greater body of evidence, but then expand this analysis to include other programmes. We estimate the effectiveness of StrongMind's core programme by combining estimates of StrongMinds' effects from the direct evidence with broader evidence of similar interventions to g-IPT. We consider evidence to be "direct" if it was generated by StrongMinds itself, or was an RCT of g-IPT in a similar context i.e., Bolton et al., (2003; 2007) or Thurman et al., (2017). We consider interventions indirect evidence (but still incorporate them) if they included psychotherapeutic elements and were delivered by non-specialists to groups of people living in low- and middle-income countries.

Although we are less certain of StrongMinds' non-core programmes, we perform a back-of-the-envelope calculation of the overall cost-effectiveness of a donation to StrongMinds, not just their core program. StrongMinds collects data on participants' changes in depression over the course of their programmes (normally 8-12 weeks) and information on the cost per person for each of those programmes. We use this information to estimate the relative cost-effectiveness of their various programmes, and compare this to the core programme. More information on how we did this is given in Section 6.2.

Our general method for performing CEAs is shown in Figure 1 below, followed by an explanation of the main components.

Figure 1: Summary of our cost-effectiveness process
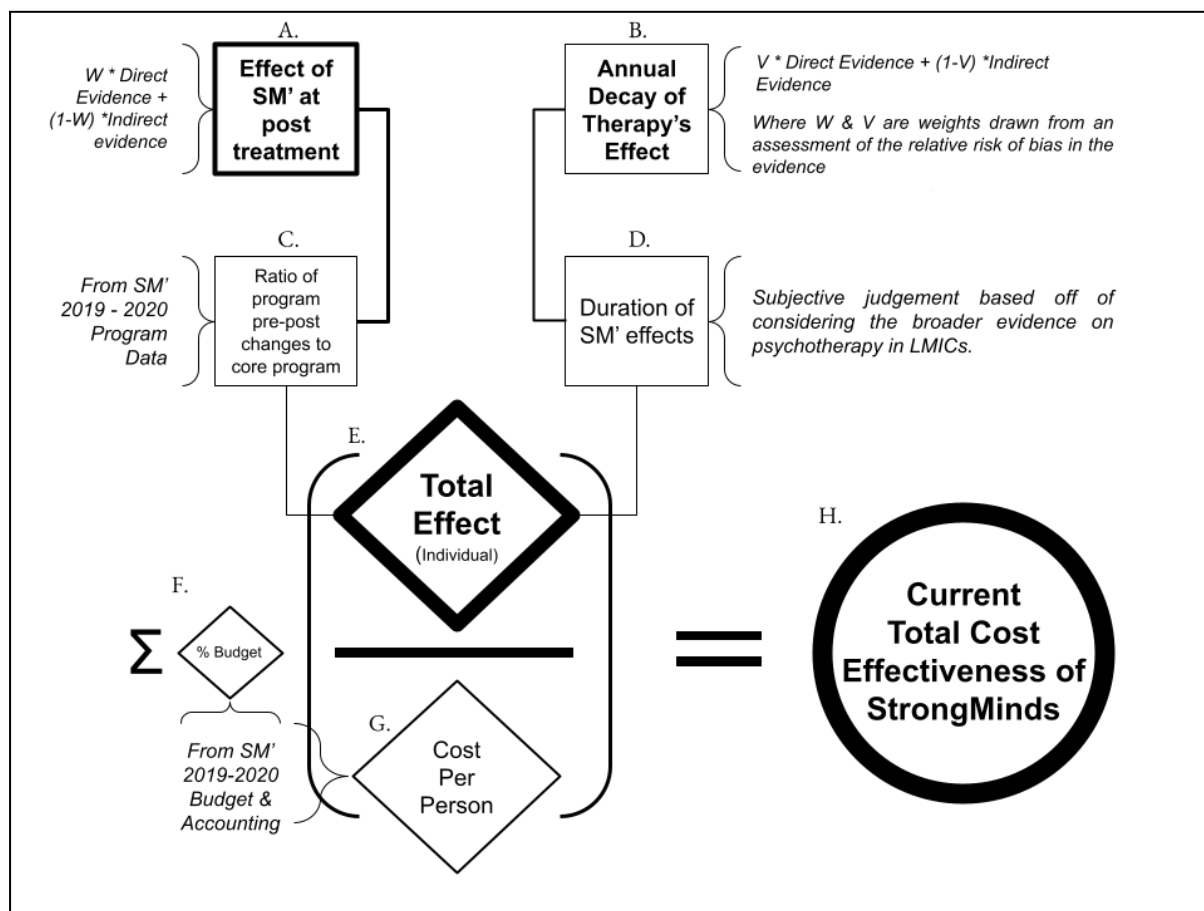
Note: SM = StrongMinds. We estimate the parameters A-D for several of StrongMinds' programmes and combine these to produce an overall estimate of the cost-effectiveness of a donation to the organisation.

A. Effect of StrongMinds' core programme (g-IPT) immediately following treatment

This parameter (as well as B & D) draws from the direct and indirect evidence of g-IPT's effectiveness at reducing psychological distress of women in Africa.

Direct evidence of effectiveness at post-treatment comes from StrongMinds' preliminary results of its recent RCT (2020), pre-post data on its programmes (2019-2020), its second pilot trial (2017), the RCTs that StrongMinds is based on: Bolton et al., (2003; 2007), and an additional RCT of g-IPT deployed in Africa (Thurman et al., 2017).

Indirect evidence of effectiveness at post-treatment comes from two meta-analyses that we performed using regression techniques on a sample of 39 RCTs of group or task-shifted therapy performed in LMICs. See our report on the effectiveness of psychotherapy in LMICs for more details on the source of this data (HLI, 2020b).

We currently combine the direct and indirect evidence of StrongMinds' effectiveness by assigning weights to each source of evidence, then taking the weighted average. These weights represent our evaluation of each source's relevance to StrongMinds and their relative risk of bias. We discuss this more in Section 4.1.1.

B & C. Annual decay of effects and duration

We chose to use an exponential decay model to represent the trajectory of psychotherapy's effectiveness through time, as we did in our review of psychotherapy in LMICs ([HLI, 2020b](#)). We discuss our choice of an exponential model further in Section 4.2 and in the psychotherapy report. The key intuition of an exponential decay model is that it assumes that the benefits dissipate quickly after therapy ends, then more slowly through time.

This model requires specifying the annual decay rate at which the benefits degrade[2]. This input also combines evidence from StrongMinds' related studies and the broader literature. For the duration, we provide a subjective judgement of the range of years the effects of the programmes last based on evidence from all psychotherapy trials we have reviewed.

### D. Relative effectiveness compared to the core programme

We have less evidence on StrongMinds' non-core programmes, such as its peer-delivered g-IPT or tele-therapy. To estimate their effectiveness, we make the assumption that the relative effectiveness of the programmes can be approximated by differences in StrongMinds recorded impact relative to the core programme. The impact they've recorded is the pre- to post-treatment changes in PHQ-9 scores collected in 2019-2020 from the women who completed therapy (n = 21,955). In other words, we assume that the differences in the average depression changes between programmes approximate the relative differences in each programme's causal impact. This assumption allows us to frame every non-core programme as a multiple of the core programme's effectiveness, and thereby expand our cost-effectiveness analysis.

E. Total effect of a StrongMinds programme

Inputs A-D all feed into an estimate of the total effectiveness. The total effectiveness of StrongMinds on the individual recipient is the sum of the benefits over the duration that the recipient experiences any benefit. We illustrate the total effectiveness of an intervention or programme in Figure 2.

---

[2] To find the [exponential decay rate,](#) we solved the following equation for k: $d_t = d_0 * e^{-kt}$. Where d_t is the effect size (Cohen's d), at the time of the follow-up, t, while d_0 is the effect size at post-treatment and e is Euler's constant (approximately 2.718) to the power of the decay rate, -k multiplied by t.

Figure 2: Illustration of total effectiveness of an intervention or programme
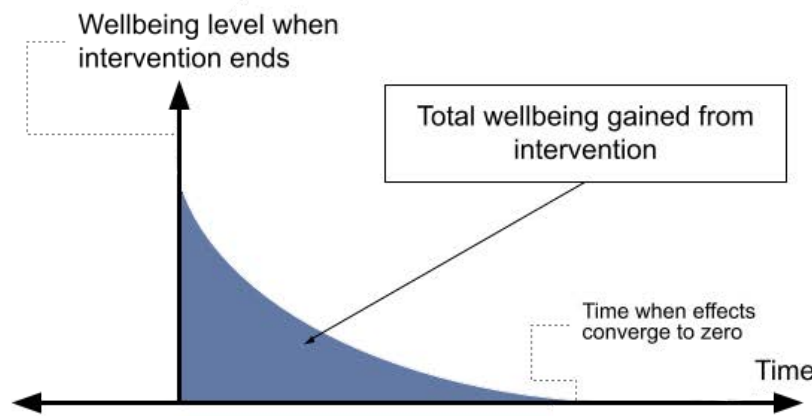
G: Cost and cost-effectiveness of a programme

After we calculate the total effect of a programme, we estimate the cost per person for a programme using the cost information from StrongMinds' accounting. The cost-effectiveness of a programme is given by taking the average value of the total effect and dividing it by the cost StrongMinds incurs to treat a person, C_{pp}, or: $\frac{mean(T_p)}{mean(C_{pp})}$.[3]

H. Cost-effectiveness of StrongMinds as an organisation

Finally, the total cost-effectiveness of a donation to StrongMinds is estimated as the weighted average cost-effectiveness of each programme weighted by the proportion of the budget each programme receives, w_b.

$$CE_{total} = (\Sigma w_b * \frac{E(T_p)}{E(C_{pp})}) \qquad (1)$$

This a conservative way of thinking about an organisation's effectiveness: the idea is that anything not spent directly on a programme must be paid for, but has no value in itself, hence an organisation will be less cost-effective the greater proportion of its budget is spent on non-programme items. This is conservative because an organisation might be doing highly effective things with its non-programme budget, such as lobbying governments, doing research, or providing training, which this calculation ignores.

This estimate also assumes that StrongMinds will continue to operate at its current level of cost-effectiveness. A further question is how much funding can be absorbed and still spent

---

[3] Not to be confused with the expected value of the total value divided by the cost per person which would overestimate the cost-effectiveness according to Jensen's inequality.

productively (discussed further in Section 6.5). Next, we discuss in more detail the source of the parameters discussed in the CEA diagram, starting with evidence of StrongMinds' effectiveness.

4. Effectiveness of StrongMinds' core programme

The previous section gave a broad overview of the CEA we performed, and its inputs. Next, we discuss the relevant evidence for the effects of g-IPT on depression at post-intervention. Then we discuss the evidence of how quickly those effects decay. Finally, we explain how we combined the different pieces of evidence to arrive at an estimate of the total benefit g-IPT provides to StrongMinds' participants.

4.1 What is the relevant evidence to inform the effectiveness of StrongMinds' core programme?

We consider a study 'relevant' if it differs by very few features that plausibly change its effectiveness. To reiterate the features of StrongMinds core programme: it is face-to-face group interpersonal therapy delivered by non-specialists to women in low-income countries. Which of these elements are critical to its effectiveness? Options include:

- the form of psychotherapy (IPT versus CBT or another type of psychotherapy)
- the format (group versus individual)
- the expertise of the deliverer
- the characteristics of the target population (women in low-income countries).

We assume that all studies of group or task-shifted (delivered by non-specialists) therapies in LMICs are relevant and we consider any study by StrongMinds or Bolton et al., (2003; 2007) as the most relevant. We explain this reasoning in a footnote[4]. This, in practice, means using the studies included in our review of psychotherapy's effectiveness in LMICs (HLI, 2021). However, we ran a

---

[4] There is a relatively quick way to inform our view due to MetaPsy, an unusually accessible database of RCTs that study psychotherapy's effect on depression. Therefore, we defer primarily to the extant literature by running a regression analysis on these 393 RCTs and adding variables to explore whether the features change the effect size. Unfortunately, the database does not contain any information on the speciality of the deliverer so we cannot include that variable. We ran the regression of the form $MHa = \beta_0 + \beta_1 Group + \beta_2 IPT + \beta_3 LMIC + \beta_3 percent\ women$, where MHa is the standard deviations of depression improved. The equation resolves its parameters to these estimates:
$MHa = 0.403^{***} + 0.132 Group^{***} + 0.05 IPT + 0.152 LMIC^{***} + 0.251 Women^{***}$. This means that studies report higher effects when psychotherapy is delivered to a group, or in a LMIC, or to women -- so every category matters *except* its form. In other words, IPT does not differ significantly in its effect from other psychotherapies. However, this is not conclusive as such broad categorizations obscure the complex interactions possible that describe the particular circumstances in which one psychotherapy is superior. Investigating this heterogeneity is outside the scope of this report.

robustness check using studies in high income countries and found similar results (see footnote 4). We call the most relevant evidence 'direct' evidence if it studies a programme similar to StrongMinds. We distinguish between the direct and indirect evidence because all else equal we give more weight to the direct evidence. We discuss the sources of evidence in more depth next.

Sources of direct evidence

The direct effect of StrongMinds' g-IPT immediately after treatment has ended comes from three sources. The sources will be explained in greater detail below: (1) Bolton et al., (2003; 2007), (2) StrongMinds' preliminary findings from their recent RCT (StrongMinds, forthcoming), (3) and the two non-randomized impact evaluations StrongMinds piloted in 2014 and 2015 to measure its effectiveness. Next, we discuss the relevance and quality of these sources of evidence, which we summarize in Table 1.A.

Table 1.A: Sources of direct evidence and their effects

Programme	Sample Size	Effect on depression at t=0 in Cohen's d	Relevance	Quality
Bolton et al., 2003	248	1.13	Moderate	Moderate
SM's RCT, 2020	250	1.72	High	Unknown
SM's Phase 2 trial	296	1.09	High	Low
Thurman et al., 2017	482	0.092	Moderate-low	Moderate
Bolton et al., 2007	31	1.79	Moderate -low	Low

Note: SM = StrongMinds

Bolton et al., (2003) and its six-month follow-up (Bass et al., 2006) were studies of an RCT deployed in Uganda (where StrongMinds primarily operates). StrongMinds based its core programme on the form, format, and facilitator training[5] of this RCT, which makes it highly relevant as a piece of evidence. StrongMinds initially used the same number of sessions (16) but later reduced its number of sessions to 12. They did this because the extra sessions did not appear to confer much additional benefit (StrongMinds, 2015, p.18), so it did not seem worth the cost to maintain it. Bolton et al., (2003) also included men and women which could make it less relevant insofar as g-IPT differs in efficacy between men and women. However, there is no solid empirical evidence that gender is a moderator of IPT's effectiveness in HICs (Bernecker et al., 2017).

Overall Bolton et al., (2003) appears to be a moderate-to-high quality RCT for several reasons. First, it is clustered at the village level which should reduce any spillovers on the controls. Second, the authors address, somewhat convincingly, the concerns they raise about the vulnerability of the

---

[5] In personal communication StrongMinds says that their mental health facilitators receive slightly less training than those in the Bolton et al., (2003) RCT.

study to selection bias and inadequate blinding of interviewers. Cuijpers et al. (2020) assigned a risk of bias score of 3 out of 4, where 4 out of 4 is the lowest at risk of bias, in their database of psychotherapy studies[6], and we think this is a suitable score for the aforementioned reasons.

The control group was allowed to pursue whatever treatment was usual, although the authors were not sure what that entailed. This somewhat defrays concerns that the counterfactual impact of StrongMinds is small if what participants would otherwise do was effective. However, the study is around two decades old, so the usual treatment might have changed (and become more effective), which would decrease Bolton et al.'s (2003) relevance.

Attrition or non-response was relatively high and slightly differential between control and treatment groups, with 24% attriting in the control group and 30% in the treatment group. If people are more likely to drop out of the study because they are relatively more depressed, then the effect will be overestimated. The evidence is mixed regarding whether this appears to be the general case in psychotherapy[7].

StrongMinds recently conducted a geographically-clustered RCT (n = 394 at 12 months) but we were only given the results and some supporting details of the RCT. The weight we currently assign to it assumes that it improves on StrongMinds' impact evaluation and is more relevant than Bolton et al., (2003). We will update our evaluation once we have read the full study.

StrongMinds' impact evaluation for treating depression at scale in Africa (2017).
This study is highly relevant since it is extremely similar to the current core programme but it has a relatively higher risk of bias for reasons we discuss next, in order of importance:

The control group was not randomized, "The control arm was formed by those who declined to join the IPT groups during screening; although they were also diagnosed with depression, they preferred individual therapy which StrongMinds does not provide." (p. 9). The formation of this control group could be problematic. The participants in the control might have selected out of group therapy based on a preference for individual therapy. However, this may have also been because they were predisposed to dissatisfaction with all available treatment options. If this was the case, then selection in the control group could be related to a higher likelihood that depressive symptoms persist. That would be a way that the trial overestimated its effectiveness at follow-ups. The control group was also put on a waitlist to be offered an opportunity they had previously declined (group therapy), which could overestimate the treatment's benefits if this waitlist makes people feel worse, a phenomenon there is some evidence for (Furukawa et al., 2014).

---

[6] MetaPsy contains no explanation for why a certain risk of bias score was assigned to a study.

[7] Stubbs et al., (2016) finds that, for exercise therapy, higher baseline depressive symptoms predict greater dropout. Karyotaki et al., (2015) find that self-guided web-based interventions for depression that have higher baseline depression severity predicts higher dropout rates, but not significantly. Gersh et al., (2017) finds a positive but not significant relationship between severity of anxiety and dropout rate in psychotherapy for generalized anxiety disorder.

Participants with mild cases were dropped from the analysis in order to avoid artificially inflating the impact evaluation, but the evidence on whether case severity is related to treatment efficacy is mixed. Driessen et al., (2010) and Bower et al., (2013) find that the most distressed patients benefit the most from therapy but Furukawa et al., (2017) finds a null effect. If more distressed individuals benefit more from psychotherapy, then excluding these mild cases would inflate the estimated effect size.

Thurman et al., (2017) is a moderate to low relevance RCT since its sample is younger, of mixed gender, and was not selected based on symptoms of psychological distress. It has a relatively large sample size (n = 482), uses an intention to treat design, and carries out a relatively long follow-up (one year). These factors signal a moderate quality but the study lacks a detailed description of the intervention's implementation beyond the amount who attended at least one session, 77%, which is relatively low.

Bolton et al., (2007) is the piece of direct evidence least relevant to StrongMinds core programme as its population is younger and of mixed gender[8]. Although StrongMinds has a programme that targets young women, it is not their core programme. My main concern with this study is that the sample size is very small (n = 31), which is generally associated with larger effect sizes in meta-analyses (Vivalt, 2020; Cheung & Slavin, 2016; Pietschnig et al., 2019).

A final note on direct evidence is that there appears to be a large RCT (n = 1,914) comparing StrongMinds to UCTs currently underway (scheduled to finish September, 2022). The RCT seems affiliated with Berk Ozler and Sarah Baird, who are both respected economists.

Sources of indirect evidence

We include evidence from psychotherapy that isn't directly related to StrongMinds (i.e., not based on IPT or delivered to groups of women). We draw upon a wider evidence base to increase our confidence in the robustness of our results. We recently reviewed **any form** of **face-to-face** modes of psychotherapy delivered **to groups** or **by non-specialists**, deployed in **LMICs** (HLI, 2020b)[9]. At the time of writing, we have extracted data from 39 studies that appeared to be delivered by non-specialists and or to groups from five meta-analytic sources[10] and any additional studies we found in our search for the costs of psychotherapy.

---

[8] Psychotherapy appears to have different and inconsistent effects across age groups (Cuijpers et al., 2020), and sexes (see footnote 3).

[9] See footnote 3 for why we do not restrict the sample based on the form of the therapy.

[10] Rahman et al., (2012), Morina et al., (2017), Vally and Abrahams (2016), Singla et al., (2017) and the meta-analytic database MetaPsy.

These studies are not exhaustive. We stopped collecting new studies due to time constraints (after 10 hours), and the perception that we had found most of the large and easily accessible studies from the extant literature[11]. The studies we include and their features can be viewed in this spreadsheet.

Using these studies, we specified two regression models on the study outcomes we collected (which makes them "meta-regressions"). In the first, we simply explain the variation in the effect on MHa that a study found with a variable that tracked the time since the therapy ended for each study. This is also the primary model we used in estimating the average total effect of psychotherapy in LMICs (HLI, 2020b).

$$MHa \ = \ intercept \ + \ time \ since \ end + \text{error} \ (2)$$

The second model attempts to predict the effect of a study that looks like StrongMinds by adding variables to indicate that a study was primarily psychotherapy[12], delivered to groups, women, and by non-specialists (sometimes referred to as 'task-shifting' to lay practitioners). This model is underpowered so the moderating variables are not significant (recall there are only 39 studies), but we think it's worth including as a source of evidence because it incorporates more StrongMinds relevant information than equation (2).

$$MHa \ = \ intercept \ + \ time \ since \ end \ + \ form \ + \ group \ + \ women \ + \ lay \ + \ error \ (3)$$

The results of these models are shown in Table 1.B, which is similar to Table 1.A except the evidence is less relevant and the effects are lower.

Table 1.B: Indirect evidence of StrongMinds efficacy

Programme	Sample Size	Effect on depression at t=0 in Cohen's d	Relevance	Quality
Meta-regression (1): MHa ~ time	38,663	0.46	Low	High
Meta-regression (2): MHa ~ time + SM-like traits	38,031	0.80	Low-Moderate	Moderate

---

[11] We searched for studies on Google Scholar using the terms "meta-analysis depression satisfaction happiness psychotherapy low income". Once we found a study, we would search its citations and the studies that cited it using Google Scholar and the website www.connectedpapers.com.

[12] In our review, we include therapies that are not exclusively psychotherapy and may also include educational elements unrelated to managing distress.

What outcome measures do we use?

We tend to consider measures of subjective well-being - overall rating of experiences and evaluations of life as our preferred measures of benefits. However, we only find measures of affective mental health (depression, distress and anxiety) in the evidence we consider. Measures of affective mental health capture people's moods and thoughts about their lives, but also ask questions about how well an individual functions[13].

We assume that treatment improves the 'subjective well-being' factors to the same extent as the 'functioning' factors, and therefore we could unproblematically compare depression measures to 'pure' SWB measures using changes in standard deviations (Cohen's d). If, however, there is a disparity, that would bias our comparison between interventions. To push the point with an implausible example, if therapy only improved functioning, but not evaluation and mood, it would be wrong to say it raises SWB and compare it to interventions that did. We assume such a scenario is unrealistic, but further work could consider whether (and if so, to what extent) this problem exists.

How do we aggregate the evidence of StrongMinds' effectiveness?

At this point, we have a range of different pieces of evidence that are relevant. What we want to do next is somehow sensibly aggregate so that we end up with a prediction about how impactful we expect StrongMind's programmes to be in the future. What we do is assign weights to each piece of evidence, using an appraisal[14] of its risk of bias and relevance to StrongMinds' present core programme. In the previous section, we discussed some factors which may affect the risk of bias and relevance of each study. Table 2 presented below gives the sample size, effect size in Cohen's *d* (with 95% confidence intervals), and the credence we place in each source of evidence representing the true effect of StrongMinds' core programme. On the bottom row, we provide the estimated total effect size, which is a weighted average. We discuss the reasoning behind the weights assigned in more detail in Appendix A.

---

[13] For example, the PHQ-9 asks about someone's appetite, sleep quality, concentration, and movement in addition to whether they feel pleasure, depressed, tired, bad about oneself, or think they would be better off dead.

[14] We considered objective methods for weighing the pieces of evidence but they all lead to intuitively high reliance on some groups of evidence. For instance, taking the naive average or the inverse variance weighted average would give far too much weight to the indirect evidence because we include a summary of studies in the estimate. Including all of the studies themselves would lead to confusion and the opposite problem. Weighting the studies by their sample size would give far too little weight on the direct evidence.

Table 2: Evidence of direct and indirect evidence of StrongMinds effectiveness

Row	Programme	Sample Size	Effect size at t=0 in Cohen's d	Credence / Weight in ES
1	Meta-regression (1): MHa ~ time	38,663	0.46	0.30
2	Meta-regression (2): MHa ~ time + SM-like traits	38,031	0.80	0.28
3	SM's RCT 2020	250	1.72	0.13
4	Bolton et al., 2003 RCT	248	1.13	0.13
5	SM's Phase 2 trial	296	1.09	0.08
6	Thurman et al., 2017 RCT	482	0.092	0.05
7	Bolton et al., 2007 RCT	31	1.79	0.03
	Estimated ES of SM's Core Programme		0.880	

Note: The term t=0 refers to the time at which treatment ends. The estimated effect size of StrongMinds' core programme is the average of the effect sizes from the other sources of evidence weighted by our belief that they best represent the true effect of the programme.

Our current estimate is that StrongMinds has a higher effect at post intervention (0.88) than predicted for a study similar to StrongMinds (0.80 SDs). This effect is highly sensitive to the weights we assigned but recourse to a sensitivity analysis is not necessary. Since this is a weighted average, and the weights are judgements, the effect must lie between the smallest and largest points: 0.092 and 1.79 SDs.

To get the total effect, we need to understand how long these effects last. We turn to that question in the next section and explain how we combine the pieces of evidence to estimate how long the effects of StrongMinds last.

4.2 Trajectory of efficacy through time

How long do the effects last for StrongMinds core programme of face-to-face g-IPT? We base our answer on the estimates we find in our broader psychotherapy results, then adjust them (downwards in our case), based on the direct evidence.

We use the same model for decay as we do in our psychotherapy report (HLI, 2020b). To recapitulate, we believe that the effects of psychotherapy decline nonlinearly: quickly at first, then more slowly as time goes on. We discuss this topic in more detail in sections 4.1 of our psychotherapy report (HLI, 2020b), however it remains a source of moderate uncertainty.

Does StrongMinds' g-IPT decay faster or slower than the average psychotherapy in LMICs? We compile the evidence we use to consider this question in Table 3 below. Table 3 contains the benefits retained annually in each piece of evidence and how much we weigh them when estimating the decay rate for StrongMinds' core programme of g-IPT.

Here, we give considerably more weight to the indirect evidence i.e., the meta-analytic sources (68%) than we did when looking at only the effects at post-intervention (58%). The reason for this is that estimating exponential decay based on only two timepoints is not very informative and there were no more than two timepoints in the direct evidence for StrongMinds[15]. Both meta-regression specifications estimate a very similar decay weight, so we divide the weight given to the wider evidence between them.

Similar to Table 2, we give more weight to StrongMinds' recent RCT. Their recent RCT showed an unusually rapid decay rate of the benefit. Between the 6-month follow-up and the 12-month follow-up the effects more than halve.

The other two pieces of direct evidence we consider are Bolton et al.'s (2003) six-month follow-up (Bass et al., 2006) and the change in StrongMinds' participants' PHQ-9 scores which they also collected at a six-month follow-up (Appendix E). These sources found relatively slower rates of decay than the average psychotherapy programme. We put little weight on the StrongMinds data because it did not compare to a control group[16].

We should note an additional reason for assigning more weight to sources that find slower decay. At the six-month follow-up of the 2003 RCT, 14 out of 15 groups continued to meet on their own without a mental health facilitator organizing the groups (Bass et al., 2006). This was also found in StrongMinds' pilot study eighteen months after the formal sessions ceased, where 78% of the group participants continued to meet informally. If groups continuing to meet leads to a higher retention of benefits over time this may indicate an advantage unique to certain kinds of group-psychotherapy[17].

---

[15] StrongMinds followed-up its initial impact evaluations at 18 and 24 months, but did not report a continuous score for changes in depression or any error term, making standardization impossible.

[16] Merely looking at 'within individual' changes overestimates the effectiveness of an intervention if the sample is selected based on unusually high depressive symptoms (which nearly all psychotherapy RCTs are). Additionally, people who feel extremely bad often improve at least somewhat with or without treatment i.e. they regress to the mean.

[17] One can test this empirically by introducing an interaction term between time and format when explaining variation in effect size between studies, such as: MHa ~ time*Group. Surprisingly, when we run this regression it indicates that group psychotherapies have a quicker decay rate than individual psychotherapies (see model 1.5 in Table 2 of HLI, 2020b). This may be in part because group psychotherapies tend to have higher effects to begin with and a higher initial effect is related to a faster decay. That would be an issue with our regression as it implies simultaneity, and endogeneity, and that the results of our model are biased. We can test this with a two stage regression where first we estimate the

Table 3: The evidence of how the effect of StrongMinds decays

Row	Programme	Sample Size	Percent of Benefits Retained Annually	Decay Credence
1	Meta-regression (1): MHa ~ time	38,031	0.713	0.34
2	Meta-regression (2): MHa ~ time + SM-like traits	38,663	0.712	0.34
3	SM's RCT 2020	250	0.37	0.14
4	Bolton et al 2003 RCT	248	0.8	0.09
5	Thurman et al., 2017 RCT	482	0.55	0.05
6	SM's 2019-2020 pre-post data	298	0.9	0.04
7	Bolton et al., 2007 RCT	31	-	-
8	SM's Phase 2 trial	296	-	-
	Estimated Decay Rate of SM's Core Programme		**0.671**	

Overall, we estimate that StrongMinds has a slightly quicker decay in its benefits than studies pulled from a broader evidence base (67% of the effects remain after a year compared to 71%). This is also slightly faster than the decay rate Founders Pledge used in their cost-effectiveness analysis of psychotherapy. Specifically, they cite Reay et al. (2012) which estimated on a small sample (n = 50) that interpersonal therapy had a half-life of about 2 years, or it decayed by about 30% annually[18].

4.3 Total effect of StrongMinds

In the previous sections we discussed how we estimated the effectiveness of StrongMinds' at treating depression at post-intervention and how quickly those effects decline over time. These two parameters provide us almost everything we need to estimate the total effects of StrongMinds core programme. The only remaining parameter we need is how long the effects last, since in an exponential model the effects never reach zero. We assume the effects have entirely dissipated in five years[19] (95% CI: 2, 10). However, the effects get close to zero for each year after five years, so

---

slope of each study's change in effects through time, then we use those results as the dependent variable in the second regression and see if the effect size of the study is a significant predictor of its change in efficacy.

[18] This is also a faster rate than what we estimate for CTs (96% of the benefit was retained every year, HLI, 2020c).

[19] We think it's implausible that the benefits last the rest of someone's life, so the question is when do effects dissipate? The longest follow-up of Baranov et al., (2020) provides some evidence that the effects can last seven years. Other

choosing a larger value would not increase the effects much (extending the duration from 5 to 10 years only increases the effects by 15%).

We plug the post-intervention effect, decay rate and time effects end into equation 4 to find the total impact of a StrongMinds programme T_p on an individual. Where d_0 decays at a rate of k for t years.

$$T_p = \int_0^t d_0 * e^{-kt} \qquad (4)$$

We estimate the total effect of StrongMinds on the individual recipient as **1.92 (95% CI: 1.1, 2.8) SDs of improvement in depression**. See Figure 3 for a visualization of the predicted trajectory of the effects of StrongMinds core programme compared to the average psychotherapy through time.

Figure 3: Trajectory of StrongMinds compared to lay psychotherapy in LMICs
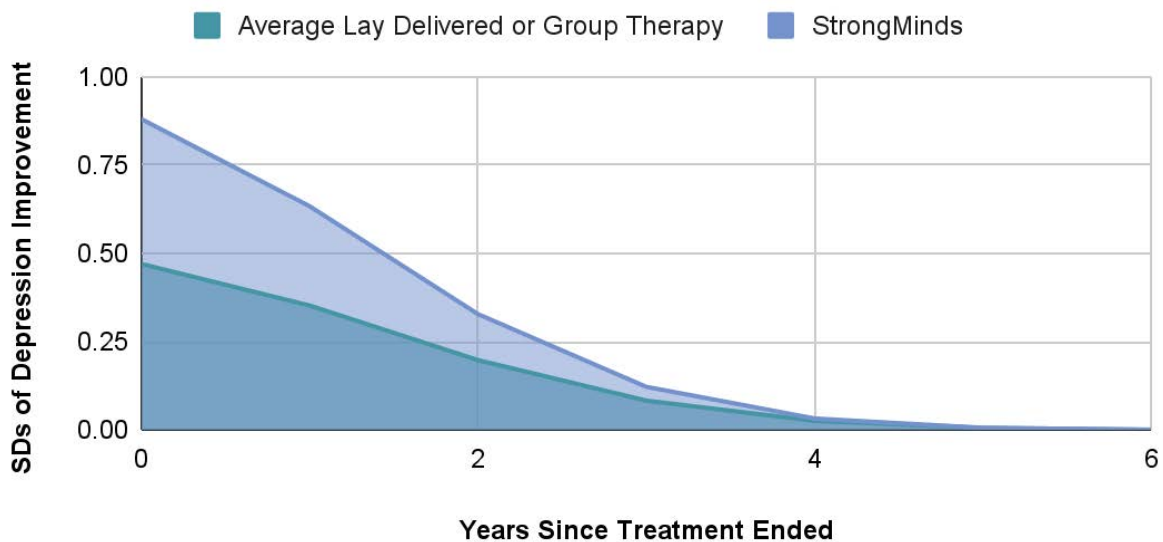
5. Cost of StrongMind's core programme

We are interested in the cost-effectiveness of a donation to StrongMinds. For donations that would make up a small portion of StrongMinds' budget, we think that using their recent cost figures is a good approximation of their near-term future costs. We discuss their room for more funding and how the cost-effectiveness may change in the future in Section 6.5.

---

long-run follow-ups in HICs find significant results at five and a half years, (Bockting et al., 2009), six years (Fava et al., 2014), but not ten years (Bockting et al., 2015). Two studies with 14 and 15 year follow-ups find the effects of drug prevention and a social development intervention have effects of 0.13 and 0.27 SDs on adult mental health service use and likelihood of a clinical disorder (Riggs and Pentz, 2009; Hawkins et al., 2009). But it's unclear how relevant these are to psychotherapy.

StrongMinds records the average cost of providing treatment to an additional person (i.e. total annual expenses / no. treated)and has shared the most recent figures for each programme with us.

Additionally, we assume that StrongMinds can continue to treat depression at a cost comparable to previous years. A further note is that they defined treatment as attending more than six sessions (out of 12) for face-to-face modes and more than four (out of 8) for teletherapy. If we used the cost per person reached (attended at least one session), then the cost would decrease substantially[20].

Our estimates of the average cost for treating a person in each programme are taken directly from StrongMinds' accounting of its costs from 2019. We use this figure because we expect it to be closer to the expected value in the future than in 2020 where costs increased substantially to $372 per person due to COVID-19-related adjustments in programming. This is supported by the initial figures published for the first quarter of 2021, which have dropped considerably to $180 per person although COVID-19 was ongoing.

StrongMinds surveyed other organisations that are in the mental health space and found that their cost per person treated ranged from $3 to $222. However, they lack details on how these costs were calculated. From our own survey of cost assessments of psychotherapy in LMICs, we estimate that the average costs per person treated for such programmes vary widely from $50 to $1,189. StrongMinds' core programme appears relatively inexpensive at an average cost per person of $128, while their peer programme is nearly half as expensive at $72.

We take StrongMinds' cost figures at face value in the point estimate cost-effectiveness calculations. However, we set the upper and lower bounds in the cost-effectiveness simulations (discussed in section 6.1 and Appendix D) according to high- and low-cost years reported in the past five years (barring the year of the pandemic). We expect StrongMinds to truthfully report these cost figures.

Table 4: Average cost of StrongMinds to treat one person

	StrongMinds	**Lay or group psychotherapy**
Avg. cost	$128	$359.29
Range lower	$72	$50.48
Range upper	$288	$1,189.00

---

[20] The pre-post data for StrongMinds core programme supports this, where the effect on those who were reached but not treated is 10% lower. Even if we assume that the effect for non-completers is 99% lower, the decrease in cost still offsets the decrease in efficacy.
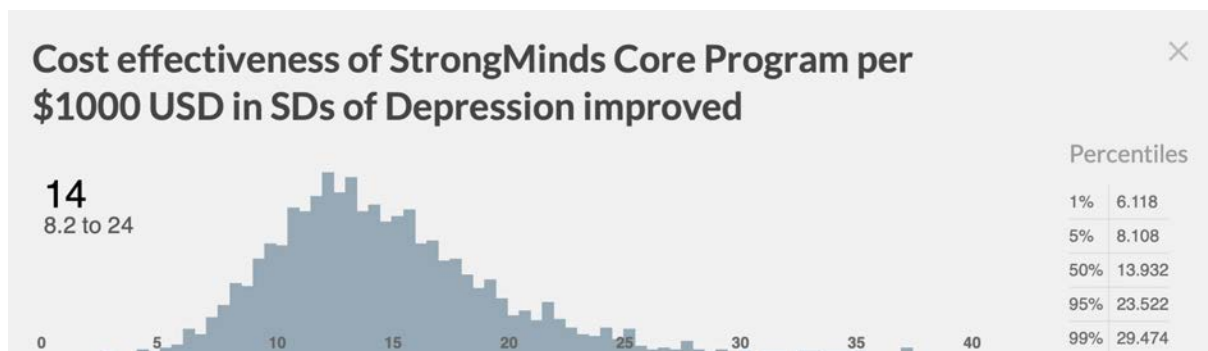
6. Cost-effectiveness analysis

We have discussed our estimates for the core programme costs and the total effect of the core programme. Now, we use this information to estimate the core programme's cost-effectiveness.

6.1 Guesstimate CEA of StrongMinds core programme

We use Guesstimate to perform a probabilistic estimate of the core programme's cost-effectiveness. This has several benefits[21]. We specified normal distributions for nearly all parameters. The simulation estimates that the cost-effectiveness ranges from 8.2 to 24 SDs (mean is 14) of improvement in MHa per $1,000 spent.

Figure 3: Cost-effectiveness of StrongMinds' core programme

6.2 Cost-effectiveness of other programmes

Not all of StrongMinds budget goes to its core programme.To estimate the cost-effectiveness of a donation to StrongMinds we need to estimate the cost-effectiveness of StrongMinds other programmes as well. In Table 4A, we show the cost-effectiveness estimate for each of StrongMinds' programmes. The core programme in Uganda is the one we have discussed extensively throughout this document. We estimate the cost-effectiveness of other programmes by comparing the estimated ratio of pre to post g-IPT improvements in depression relative to the core programme. We assume these changes represent the relative efficacy of the non-core programme compared to the core programme. So if the changes in the peer programme are 80% what they are in the core programme, we assume this represents the true difference in effectiveness at post-treatment.

---

[21] Monte Carlo simulations allow us to treat inputs in a CEA, often merely stated as point estimates, as randomly drawn from a probability distribution. In the Monte Carlo simulation each element is drawn from its stipulated probability distribution thousands of times. This allows us to not only specify the magnitude of our uncertainty for each element of the CEA but propagate it through our calculations. Using only point estimates for every element of our model obscures our uncertainty and compounds any error present in our estimate. The end result of using a Monte Carlo simulation for our CEA is a cost-effectiveness estimate that has its own distribution so we can think of the estimated cost-effectiveness probabilistically, e.g., "Distributing hats with plastic helicopter blades on top of them has a 50% chance of having an effect between 10 and 22 units of well-being per thousand dollars spent."

Table 4A: Point estimate of cost-effectiveness for several StrongMinds programmes

Type	Country	Percent of Program Budget	Avg. Cost Per Person	Estimated Effect after g-IPT ends	Decay of Effect (yearly)	Duration of Effect In Years	Total Effect	Effect / Cost	CE in D SD per $1k to programme	SD impact from share of $1k donation
Core	Uganda	20.85%	$ 128.49	**0.88**	0.67	5.10	1.92	0.01	**14.93**	3.11
Core	Zambia	12.97%	$ 101.70	0.81	0.67	5.10	1.76	0.02	**17.33**	2.25
Peer	Uganda	5.70%	$ 72.00	0.81	0.67	5.10	1.77	0.02	**24.57**	1.40
Youth	Uganda	14.07%	$ 197.15	0.75	0.67	5.10	1.63	0.01	**8.25**	1.16
Tele	Uganda	23.85%	$ 248.06	0.73	0.67	5.10	1.59	0.01	**6.41**	1.53
Tele	Zambia	7.47%	$ 439.44	0.79	0.67	5.10	1.72	0.00	**3.91**	0.29
Youth Tele	Uganda	6.94%	$ 215.64	0.71	0.67	5.10	1.55	0.01	**7.18**	0.50
Covid, Refugee & Partner (Peer)	Uganda	8.15%	$ 86.35	0.75	0.67	5.10	1.64	0.02	**19.03**	1.55

Total impact (SDs) from an additional $1,000 (sum of programmes)   **11.79**

Table 4B: Uncertainty on parameters for programme Guesstimate

Type	% Budget 2021	% Budget 2022	% Budget 2023	Upper Change	Lower Change	Cost Lower	Cost Upper
Core	33.82%	30.00%	25.00%	100.00%	100.00%	100	160
Peer	5.70%	7.00%	9.00%	105.00%	72.00%	60	90
Youth	14.07%	12.00%	10.00%	95.00%	73.00%	130	200
Tele	38.26%	30.00%	30.00%	95.00%	76.00%	150	400
Covid, Refugee & Partner	8.15%	21.00%	26.00%	91.67%	60.00%	60	90

We also assume that the programmes decay at the same rate as the core programme, which may be more reasonable for the peer and youth programme than tele-therapy. The tele-therapy could decay more quickly because for tele-therapy it does not seem as likely that groups will continue meeting informally[22].

Now we will explain the details of how each of these programmes is implemented, starting with the programmes given the largest share of the 2021 budget to the least. The share of budget we observe and estimate each programme will receive is shown in Figure 4.

Figure 4: Present and projected budget allocations between StrongMinds' programmes
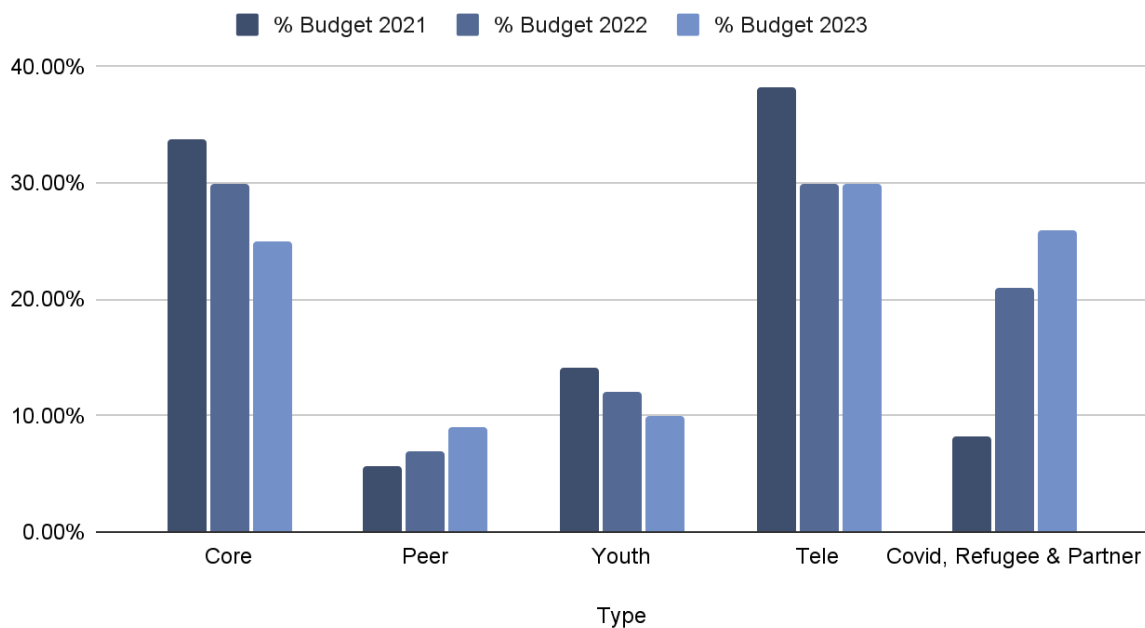
Group tele-therapy (38.27% of 2021 budget) is delivered over the phone by trained mental health facilitators and volunteers (peers) to groups of 5 (mostly women) for 8 weeks. We expect the share of the budget this programme receives to decline as the threat of COVID diminishes.

Teletherapy's cost-effectiveness does not compare favorably to the core programme. This is mostly due to higher costs that we assume are due to the recent startup of the programmes. It's also probably because the groups are smaller (with 5 instead of 12 women) due to the greater difficulty of facilitating group therapy over the telephone.

All **face-to-face formats of StrongMinds g-IPT** are based on 12 weekly sessions of 60-90 minutes (10 weeks in Zambia and with youth). Reportedly, the groups often continue to meet for longer than 10-12 weeks, sometimes forming village saving and loan associations.

---

[22] If we assume that teletherapy decays at a rate of 40% per year, then that decreases the estimated overall impact of an additional $1k donation by a 0.77 SD improvement in depression. This decreases the cost-effectiveness of teletherapy in particular by 3.21 SDs per $1k spent on teletherapy.

The original programme, which we refer to as the "core programme" **(33.82 % of budget)** is delivered by trained mental health facilitators to groups of 12 women. These mental health facilitators are required to have at least a high school diploma, and be members of the community they work in.

According to StrongMinds, "They receive two weeks of training by a certified IPT-G expert [certified at Columbia University] and receive on-going supervision and guidance by a mental health professional[23]" ([StrongMinds 2017, p.30](#)). This programme is deployed both in Uganda and Zambia. This programme is closest to those studied academically, and is what we base our effectiveness estimate on.

In addition to the core programme, StrongMinds also implements face-to-face g-IPT directed to young women and has begun a volunteer-run model. The **youth programme (14.07%)** is delivered by trained mental health facilitators to groups of adolescent girls.

StrongMinds' peer programme (5.70%) is described as "self-replicating, volunteer-led talk therapy groups of eighteen people led by individuals trained in IPT-G. For this programme component, mental health facilitators recruit [core programme] graduates eager to give back to their communities and train them to be volunteer peer mental health facilitators. They train by co-facilitating courses of the core programme for half a year" ([StrongMinds.org, 2021](#)). The peer groups are smaller than the core programme groups [(6-8) instead of 12-14.](#)

StrongMinds' peer programmes appear quite promising due to their relative cheapness but comparable effectiveness. We do not think this is very surprising, if true. The peer facilitators are former participants in the core programme, and they engage in half a year of less formal but more practical training than the mental health facilitators. The peer facilitators understand the other women in the group and have the personal experience of improving their own condition. For these reasons we do not think that using peer facilitators will decrease the efficacy more than they decrease the cost of the programme.

StrongMinds partner, COVID and refugee programmes are experimental programmes (8.15% of budget). In their partner programme, StrongMinds' trained mental health facilitators travel to new locations to train other organisations on how to deliver the StrongMinds model, stay to perform quality control and then leave (Personal communication, 2020). In their COVID programme they deploy socially distanced IPT-g. In their refugee programme they specifically treat refugees using an adaptation of their core programme.

---

[23] An interesting note: "This mental health supervisor was actually a member of the 2002 RCT in Uganda and has over ten years of IPT-G experience"

Their partner programme strikes us as potentially more cost-effective than their other interventions (assuming that the partner organizations were not doing something more cost-effective than g-IPT), but there is still limited evidence of its effectiveness.

6.3 Overall expected cost-effectiveness of a future donation to StrongMinds

Estimating the cost-effectiveness of each programme allows us to calculate the expected cost-effectiveness of an additional donation to StrongMinds under the further assumptions that a) an additional donation will be distributed across programmes according to the current budget allocations and b) StrongMinds will implement its programmes in the future at the same cost-effectiveness as it has in the past. We discuss relaxing this last assumption in the next section, 6.4, on "room for more funding".

Using the average cost StrongMinds incurs to provide psychotherapy with each programme, we estimate that the overall cost-effectiveness of StrongMinds (**12 SDs improvement of depression per $1,000, 95% CI: 8.2 to 24 SDs**) is less than the cost-effectiveness of its core programme (**14 SDs per $1,000**).

All of the information relevant for calculating the cost effectiveness of each programme and an additional information donation can be found in Table 4A. To get the confidence intervals we run a Monte Carlo simulation of the total cost-effectiveness of a $1000 donation to StrongMinds. We discuss it in detail in Appendix D.

6.4 Room for more funding

In the previous section we assessed the retrospective cost-effectiveness of StrongMinds′ programmes. But how much more funding can they absorb before the cost-effectiveness declines substantially?[24] In Appendix B, we propose how a simple model of economies of scale can keep the discussion of 'room for more funding' in terms of cost-effectiveness.

In our case, viewing Figure 6 below gives us the impression that StrongMinds′ costs are no longer rapidly diminishing, which appeared to happen when they were treating between one and seven thousand people. StrongMinds expects its cost per person to increase slightly in the near future as they plan on investing in scaling their programmes. They expect this to lead to a lower cost per

---

[24] GiveWell defines the size of the funding gap as "The total amount that each top charity has told us it could use productively in the next three years, less the total amount of funding it has or we project it will have in that period."

person treated in the long run as they plan on shifting their focus to training other NGOs and governments to deliver g-IPT.

StrongMinds' funding gap has grown. Founders Pledge noted that the gap would be $2.6M in 2020 (Halstead et al., 2019). In our communication, StrongMinds said they could reasonably absorb about $6M in the last seven months of 2021. For 2022, they have a funding gap of $9.5M USD.

Figure 6: Patients and cost for StrongMinds' Core and Peer programme in Uganda
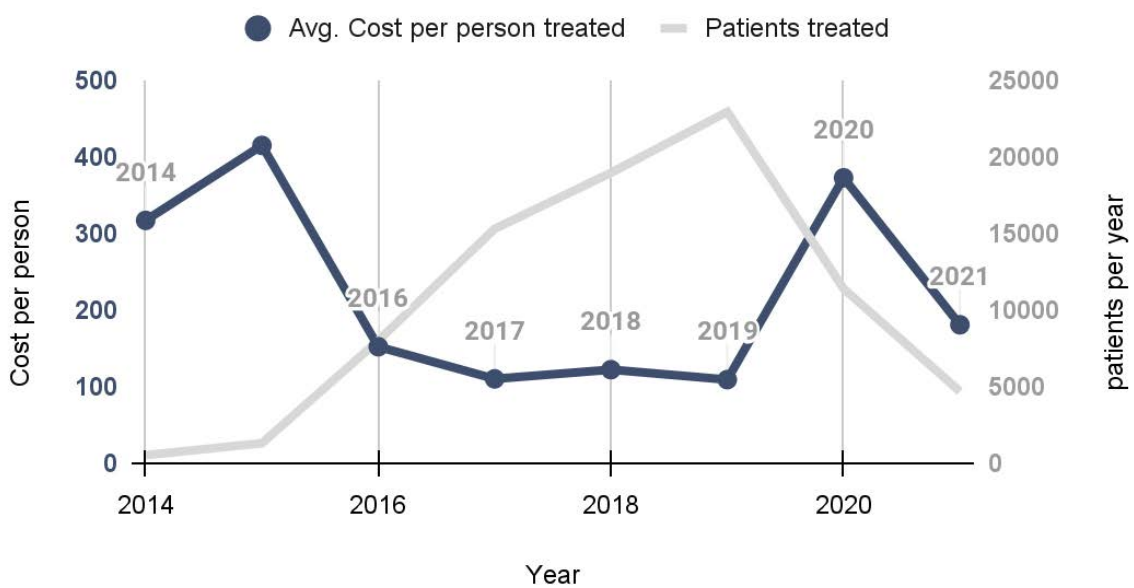
Organisation strength

We agree with the assessment of StrongMinds given by Founders Pledge "StrongMinds appears to be a transparent and self-improving organisation which is contributing to the global evidence base for cost-effective treatment for mental health." (Halstead et al., 2019, p.29)

StrongMinds has been unusually transparent and consistently helpful throughout the writing of this report. They responded promptly and in depth to at least six rounds of questions throughout the course of a nine month period.

6.5 Considerations & limitations

In this section we discuss the limitations of our analysis, how we could be wrong, and what information we need to address these concerns. We conclude with proposing further research questions.

Is social desirability bias a unique risk to psychotherapy?

One further concern you may have is whether there is an 'social desirability bias' for this intervention, where recipients artificially inflate their answers because they think this is what the experimenters want to hear. In conversations with GiveWell staff, this has been raised as a serious worry that applies particularly to mental health interventions and raises doubts about their efficacy.

In the second phase of StrongMinds' impact evaluation StrongMinds appeared particularly concerned with social desirability bias because "These IPT group members were aware that the post-assessment gathering was being conducted to collect final depression scores only" (p.16). Due to this, StrongMinds concluded that their impact evaluation overestimated the effectiveness of their programme. StrongMinds has made it clear that they're concerned about social desirability bias and have taken steps to limit its influence such as hiring a third party to collect all follow-up information to make it less likely that responses are inflated by patients reporting their depression scores to their facilitator.

As far as we can tell, this is not a problem. Haushofer et al., (2020), a trial of both psychotherapy and cash transfers in a LMIC, perform a test 'experimenter demand effect', where they explicitly state to the participants whether they expect the research to have a positive or negative effect on the outcome in question. We take it this would generate the maximum effect, as participants would know (rather than have to guess) what the experimenter would like to hear. Haushofer et al., (2020), found no impact of explicitly stating that they expected the intervention to increase (or decrease) self-reports of depression. The results were non-significant and close to zero (n = 1,545). We take this research to suggest social desirability bias is not a major issue with psychotherapy. Moreover, it's unclear why, if there were a social desirability bias, it would be proportionally *more* acute for psychotherapy than other interventions. Further tests of experimenter demand effects would be welcome.

Other less relevant evidence of experimenter demand effects finds that it results in effects that are small or close to zero. Bandiera et al., (n =5966; 2020) studied a trial that attempted to improve the human capital of women in Uganda. They found that experimenter demand effects were close to zero. In an online experiment Mummolo & Peterson, (2019) found that "Even financial incentives to respond in line with researcher expectations fail to consistently induce demand effects." Finally, in de Quidt et al., (2018) while they find experimenter demand effects they conclude by saying

"Across eleven canonical experimental tasks we ... find modest responses to demand manipulations that explicitly signal the researcher's hypothesis... We argue that these treatments reasonably bound the magnitude of demand in typical experiments, so our ... findings give cause for optimism."

Will StrongMinds scaling strategy positively impact its cost-effectiveness?

From personal correspondence we've learned that StrongMinds is planning to scale its delivery of psychotherapy to reach more people more cheaply by relying more heavily on training governments and other NGOs to deliver IPT-g. For instance, StrongMinds partnered with the Ministry of Education and Mental Health in Uganda and the Ministry of Health in Zambia to run programmes (personal communication, 2020). This has the promise of potentially lowering costs considerably for StrongMinds as it leverages other organizations which likely would have not been using their resources as effectively.

The pitfall of this transition is that it will move more of its budget farther away from an already slim evidence base. Updated evidence on the efficacy of the programmes delivered by partners would be helpful in reducing uncertainty of this transition's impact on StrongMinds efficacy and cost-effectiveness.

Our assessment from reviewing the impacts of lay-delivered psychotherapy is that even a moderate reduction in the fidelity of psychotherapy, if tied to an equally large decrease in cost, will generally benefit the cost-effectiveness as effectiveness is somewhat harder to degrade than costs.

Do we underestimate StrongMinds cost-effectiveness by excluding non-completers?

StrongMinds calculates its number treated based on how many who started treatment attended a majority of the sessions of a programme. For instance, in 2019 StrongMinds treated around 22,000 but around 6,000 started but did not attend a majority of sessions. One concern is that these individuals who began but did not finish experienced significantly fewer benefits, and this decline in benefits would offset the decrease in costs and lead to a lower cost-effectiveness. However, this is only possible if dropouts experience negative effects.

The difference we observe between dropouts and non-dropouts indicates that dropouts receive 91% the benefit of completers. If this was in fact the case, then the cost-effectiveness of a StrongMinds donation would increase from 12 to 17 SDs-years improvement in depression per $1,000 donation. Such an increase would be non-trivially more cost-effective than our current estimate. This seems plausible if benefits accrue at a decreasing rate, which appears in line with a review of the session-response literature in psychotherapy (Robinson et al., 2019). However, we do not have good evidence on this, so we do not currently include this in our main analysis.

We discuss more general concerns with psychotherapy such as spillover effects, using SD changes as our standardized measure of effect size, and equating MHa and SWB measures in section 7 of our psychotherapy report (HLI, 2020b).

**What is the counterfactual impact of StrongMinds?**

We do not assess StrongMinds to be at high risk of doing work that would otherwise be done by the government since mental health care appears neglected in all LMICs. This neglect appears to be the case in Uganda and Zambia as well (which we discuss in Appendix C).

6.6 Comparison of StrongMinds to GiveDirectly

We currently estimate, based on a meta-analysis of studies that use GiveDirectly, that $1,000 in GiveDirectly cash transfers would cost $1,170 in total and lead to a decrease in depression symptoms of 0.92 SDs-years. In this report, we estimate StrongMinds' intervention would cost $128 and provide 1.92 (95% CI: 1.1, 2.8) SDs-years of improvement in depression.

We illustrate a simulation of the comparison between psychotherapy, StrongMinds, monthly unconditional cash transfers, and GiveDirectly in Figure 7 below. In the figure, the total effects are given on the y-axis and the average costs are given on the x-axis. In this comparison, a $1,000 donation to StrongMinds is around 12x (95% CI: 4, 24) more cost-effective than a comparable donation would be to GiveDirectly. Each point is an estimate given by a single run of a Monte Carlo simulation. Lines with a steeper slope reflect a higher cost-effectiveness in terms of depression reduction. The bold lines reflect the interventions cost-effectiveness and the grey lines are for reference.

Figure 7: Cost-effectiveness of StrongMinds compared to GiveDirectly
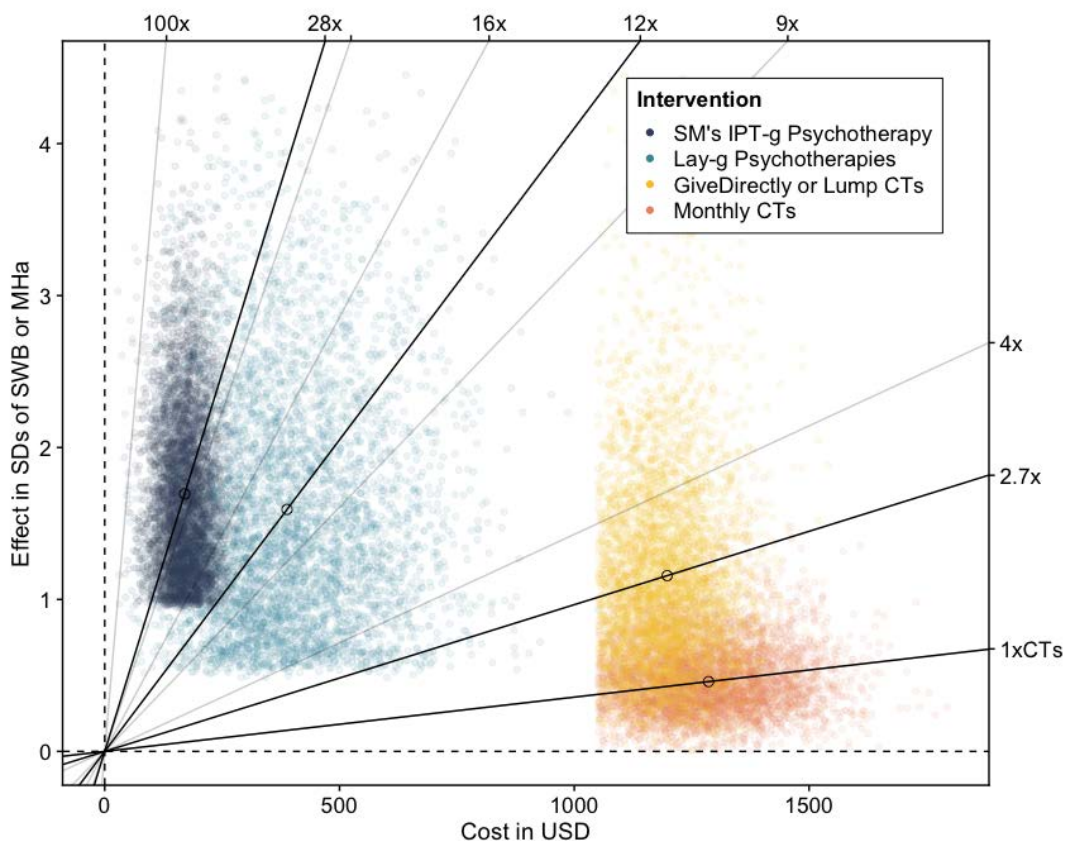
A word of caution is warranted when interpreting these results. Our comparison to cash transfers is only based on depression. We were unable to collect any outcomes using direct measures of SWB for psychotherapy, so whether the impacts are substantially different for those measures remains to be seen. Finally, this comparison only includes the effects on the individual and it is possible that the spillover effects differ considerably between interventions.

That being said, even if we take the upper range of GiveDirectly's total effect on the household of the recipient (8 SDs), psychotherapy is still around twice as cost-effective.

7. Conclusions

We updated Founders Pledge's CEA of StrongMinds with a more thorough consideration of the evidence. We estimate that StrongMinds is more cost-effective relative to GiveDirectly when considering the recipient. We do not currently have enough evidence to include spillover effects on the household and community in estimating the total effects. StrongMinds remains an organisation with strong norms of transparency, collaboration, basing its methods on existing evidence, generating new evidence, and a commitment to cost-effectiveness.

Appendix A: Explanation of weights assigned to evidence

The evidence directly related to StrongMinds, except for Thurman et al., (2017), estimates a considerably higher improvement in depression than the average group or lay-delivered psychotherapy. If we judged that we completely trusted the direct StrongMinds evidence, we would assign no weight to the effects implied by the wider evidence base, or vice versa.

When considering how much weight to assign the direct versus indirect evidence: the question is whether the larger effects of StrongMinds' direct evidence are because g-IPT is better than any other form of lay or group-delivered psychotherapy studied in our sample?

For the meta-analyses of the wider evidence, we jointly assign a credence of 58%. While by definition less relevant, we think that giving the majority of the weight to the effects from the wider evidence is warranted because the evidence base for StrongMinds is still quite small. If the new RCT is truly randomized then there will be two RCTs of the programme with a sample size greater than 100, compared to 38 in the wider evidence. The total sample size of the wider evidence (n = 36,663) is also much larger than that of the direct evidence (n = 1,123). In general, we expect that when an exceptional programme is held to more scrutiny its efficacy will shrink.

We divide the credence of 58% roughly evenly between the meta-regression described in equation (2) and equation (1). The second equation includes more variables, allowing us to predict how a study that looks like StrongMinds on observable characteristics would perform. We give slightly less weight to the second meta-regression because while it is more relevant, it is underpowered so we think the decrease in the quality of its estimate offsets the increase in its relevance.

A reader may wonder why we do not include an even broader set of evidence when estimating the effects of StrongMinds g-IPT. When we use data from 393 RCTs in mostly HICs (which mostly do not overlap with the 39 studies we use in my psychotherapy report) we estimate that effects of a g-IPT delivered to women in LMICs would have 0.988 SDs improvement in depression. Which is surprisingly close to the estimate we arrive at of 0.964. If we included it as a source of evidence, we would put more weight on the indirect evidence. If the weight on the indirect evidence was 59% We would change it to 64-67% but probably not much more than 70%. We use the equation

$MHa = 0.403^{***} + 0.132Group^{***} + 0.05IPT + 0.152LMIC^{***} + 0.251Women^*.$

Where *** is significant at the 0.001 level, ** is significant at the 0.01 level and * is significant at the 0.05 level.

The practical implication of assigning more weight to the broader evidence is to estimate that StrongMinds is less effective than it would be if the direct evidence alone was considered.

The piece of direct evidence we weigh the most (at 13% of my credence) is StrongMinds' recent RCT. We are still awaiting further details about the study, but if it significantly improves on the methods used in the Phase 2 evaluation, we will give it more weight than Bolton et al., (2003). However, its effects are so high we find them nearly too good to be true. However, all of StrongMinds' direct evidence reports high results.

We assign Bolton et al. (2003) a similar weight, 13%, to the new RCT because while it's less relevant than the Phase 2 evaluation or the pre-post data, it is a decent RCT that StrongMinds appears to have hewn relatively close to. We assign Thurman et al., (2017) a low weight of 0.05 because its sample is less relevant to StrongMinds. Its sample was composed of youth drawn from the general population, that is they were not selected based on pre-existing signs of experiencing mental distress. We assign the smallest weight to Bolton et al., (2007) of 0.03 due to its small sample size (n = 31) and the fact its sample is composed of internally displaced youths (half of each traditional gender) which is not the targeted population of StrongMinds core programme (although female youth and refugees are separately the target of other programmes we will discuss later).

For StrongMinds' Phase 2 trial it is highly relevant as it reflects StrongMinds' current core programme However, we weigh it less because its effects can not be interpreted as causal. However, this is a bit of a moot point since the effects are so similar to Bolton et al., (2003) so a vote for one is in practice a vote for the other.

We do not collect an effect size for, and place no weight on, StrongMinds' change in depression scores since it does not include a comparison to a control condition.

Appendix B: Economies of scale and room for more funding

Economies of scale represent the common principle that there's a period for an organisation where producing more reduces the average cost of production (illustrated as the blue area in Figure D1), which may stabilize for a period before inevitably increasing (illustrated as the yellow area in Figure D1). We assume that the effectiveness of a programme is unrelated to the quantity that the organisation produces, which seems reasonable in this case, although this assumption could conceivably weaken if very high levels of the total population are treated.

How does organisational capacity fit in? A weaker organisation with the same budget will face diseconomies of scale sooner, so it will have less room for funding.

Figure B.1: Economies of scale and room for more funding
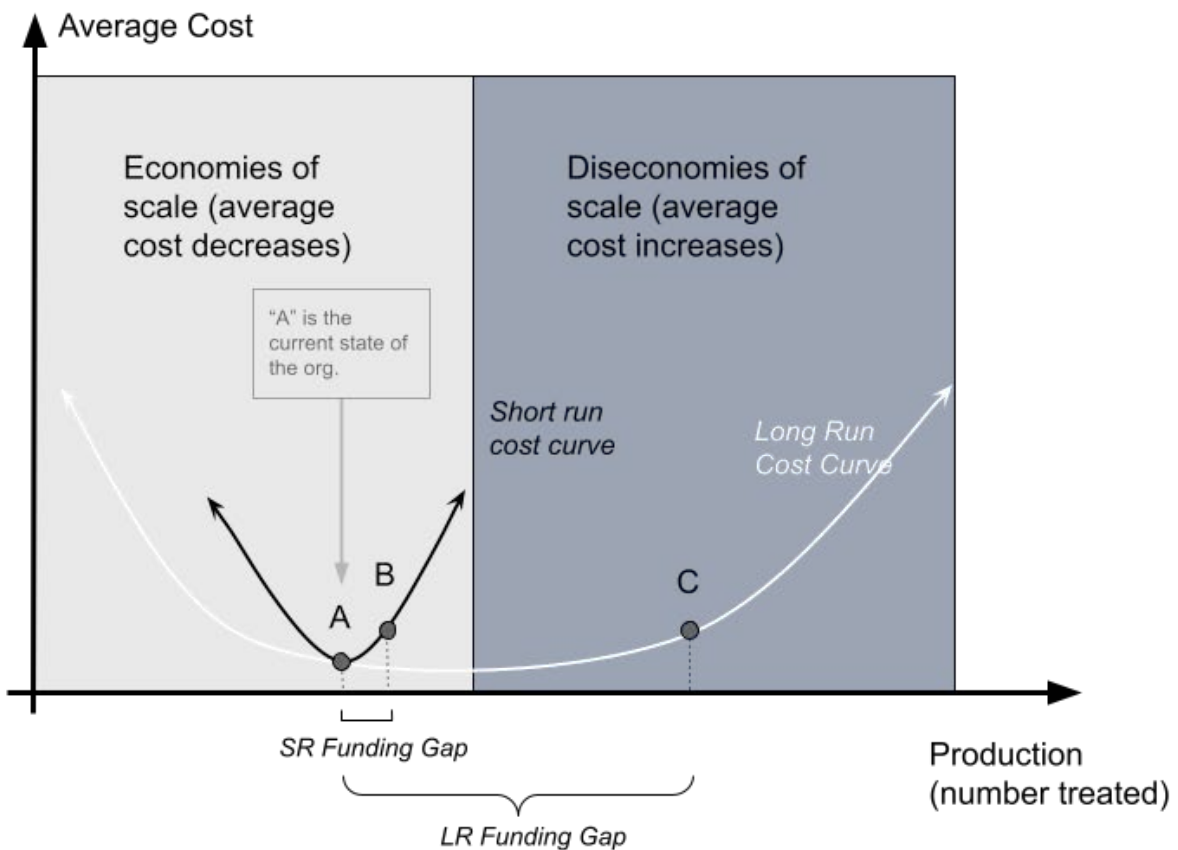
After observing this model, the questions become: Where is the organization (point A)? What's the short run funding gap (distance from A to B)? And what's the long run funding gap (distance from A to C)?

Appendix C: Counterfactual total impact of StrongMinds

What would have happened without StrongMinds' work? Would recipients of StrongMinds have received effective treatment anyway? We can think of adjusting for the counterfactual treatment a participant would receive based on the likelihood of them receiving that treatment and its relative effectiveness. Ideally, the counterfactual treatment should be assigned to the control group in RCTs, and we think that most of the evidence of StrongMinds compares treatment to care as usual, whatever that may be. This reduces my concern that we are neglecting the counterfactual treatment a participant in StrongMinds may receive. The counterfactual (c.f.) treatment effect can be thought of as the following.

$$C.f.\ Effect\ =\ StrongMinds'\ effect\ -\ pr.\ alternative\ *\ effect\ of\ alternative$$

If a participant would have certainly received a treatment that was equally effective (from the government, or elsewhere), the effect of StrongMinds may be nil. However, that seems unlikely for reasons we will explain.

We assume the likelihood of being treated outside of StrongMinds for participants to be around 10%. We take this from the proportions of people with a mental disorder who receive treatment (Rathod et al., 2017), and they note these treatments are not necessarily evidence-based.

Murray et al. (2015) write, "With a population of 33 million people, Uganda has only 28 psychiatrists and approximately 230 mental health nurses, most of whom work at Butabika National Referral Mental Hospital [which is in the capital]." They also comment that "Additionally, over 60% of mental health services are located in urban areas, requiring traveling long distances that families cannot afford. An estimated 88% of Uganda's population lives in rural areas of the country (Kigozi, Ssebunnya, Kizza, Cooper, & Ndyanabangi, 2010)."

If that is true, and allocation of mental health sources is proportional to allocation of effective treatment,then someone living in a city is 5x more likely to receive treatment than someone in a rural area (60% / 12%). So if we assume that an average of 10% get treatment across the whole country then that means that 34% coverage, but 6.8% get coverage in rural areas which is where I believe StrongMinds operates.

Finally, as hinted at previously, the treatment they receive may not be effective. There is a chance that the alternative treatment may be SSRIs which could be as effective as g-IPT (Cuijpers et al., 2020) However, this passage in an article about the human rights abuses in Ugandan mental health facilities (Molodynski et al., 2017) suggests that the treatment may be worth less than nothing: "The minority of people who do receive state mental healthcare will generally do so at Butabika

Hospital and be subject to overcrowding, poor conditions and coercion, as outlined in the MDAC report and other publications (Cooper et al, 2010). While the individual may finally receive some form of treatment, *the trauma involved may have additional long-term consequences.*" (emphasis mine).

Given this, I assume there's a chance of 6.8% for a would-be StrongMinds participant to find an alternative treatment. This could be less for women. We assume an alternative treatment has a small chance (30%) of making things as bad as StrongMinds' is good and a high chance (70%) of something as effective as StrongMinds. These values can be inputted into the previous equation.

$$Counterfactual\ Effect\ =\ 2.4\ -\ (0.068)\ *\ (0.3\ *-\ 2.4\ +\ 0.7\ *\ 2.4)\ =\ 2.33$$

In this back-of-the-envelope calculation, we adjusted the effectiveness of StrongMinds according to the counterfactual impact and found that it makes little difference. In this case the benefit reduces from 2.4 SD improvement in depression, to 2.33 SDs.

Appendix D: Guesstimate CEA of a future $1,000 donation to StrongMinds

To incorporate our uncertainty on the cost-effectiveness of an additional donation to StrongMinds we expand our Guesstimate model of the core programme to include each programme weighted by its estimated budget share over the next three years. We allow the budget share, effectiveness, and cost to vary. We explain my reasoning for the upper and lower bounds for each parameter next, the ranges which are summarized in Table 5B below. These bounds include forecasts and subjective estimates thus they are speculative in nature.

We estimate the range of budget share according to an assessment of how the budget allocation will change in the next three years according to correspondence with StrongMinds. They stated that they intend to shift more money into partner and peer programmes and away from delivering psychotherapy services directly.

We assume the pace of their shift will linearly relate to their previous diversification away from their core programme such that they will allocate 26% of their budget to partner and refugee programmes by the end of 2023. We also assume that they will allocate less to tele-therapy in the future because pressure to avoid meeting in person will ease as COVID-19 precautions are relaxed.

The calculated difference in pre-post changes of a programme relative to the core programme we treat as the best guess and define the upper bound of the 95% CI as ranging between 105% (slightly better than the core programme in the case of the peer programme) to 92% as effective as the core programme in the case of the StrongMinds refugee and partner programmes, with the youth and tele-therapy programmes lying between that at 95%.

We assume that every non-peer programme is at best not as effective as the core programme because the programmes have lower intensity and fidelity to the core programme which we assume leads to lower effectiveness. The lower bound is then defined automatically as the lowest decline in efficacy that one can input and maintain the best guess value given by the data and a normal distribution. This bound ranges from 60% to 76%, meaning we would be very surprised if the true effectiveness of StrongMinds programmes fell below 60-76% the efficacy of the core programme.

Finally, we specify the costs as taking on the values they were reported to in 2019 for peer or core programmes or 2020 for teletherapy, youth and partner programmes. However, for programmes that contain past values such as the core programme we let the possible costs range from the highest and lowest figures reported. For other programmes we give them a similarly wide range. The exception is tele-therapy which we assume will get cheaper to provide in the next few years, so we assume its costs will range from $150-400.

Once we include the uncertainty surrounding these parameters in our guesstimate simulation, the result is an estimate that a $1,000 donation to StrongMinds will result in a 7.2 to 20 SD improvement in depression (95% CI). This range of uncertainty is similar to that of the core programme (95% CI: 8.2 to 24 SDs)[25]. We show the distribution of this estimate in Figure 6 below.

Figure 6: Distribution of overall cost-effectiveness of a donation to StrongMinds

Percentiles	
1%	5.467
5%	7.046
50%	11.877
95%	19.514
99%	25.113

---
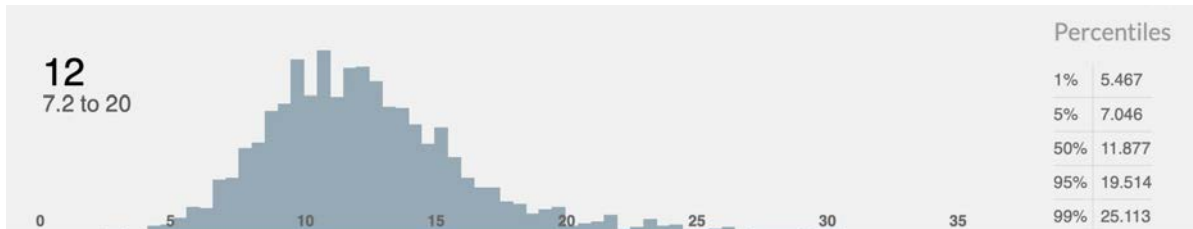
[25] One may expect a greater range of uncertainty because we are combining programmes which individually have more speculative and uncertain parameters, but since we are taking the weighted average of each programmes' cost-effectiveness, this has the property of shrinking our uncertainty by smoothing out the extremes. One can think of this as closely related to the property of scaling the variance of a random variable where if a variable is scaled by a constant, its variance is scaled by the square of that constant. This has the consequence for averages, which are scaled by $1/m$ will have their variances scaled by $(1/m)^2$.

Appendix E: StrongMinds pre-post data on programmes for participants reached (attended >= 1 Session)

| Programme | Type | Country | Year | Pre-Therapy | | | Post-Therapy | | | Mean change | Change ratio (rel. to core UG) |
				N	Mean	SD	N	Mean	SD		
STG-Community	Core	Uganda	2019	16,292	15.60	4.17	12,367.00	1.41	3.08	14.19	100.00%
STG-Community	Core	Zambia	2019	1,904	16.28	3.51	1,665.00	3.25	2.66	13.03	91.83%
PTG-Community	Peer	Uganda	2019	6,228	16.67	9.04	5,627.00	3.59	3.82	13.08	92.18%
ATG-Community	Youth	Uganda	2019	3,870	14.77	6.55	3,203.00	2.75	2.35	12.02	84.71%
STG-Teletherapy	Tele	Uganda	2020	3981	13.69	4.12	3,136.00	1.94	2.67	11.75	82.80%
STG-Teletherapy	Tele	Zambia	2020	1,025	16.18	6.60	722.00	3.49	2.66	12.69	89.43%
Uganda ATG-Teletherapy	Seed Tele	Uganda	2020	2297	13.87	3.27	1,893.00	2.42	2.05	11.45	80.69%
PTG-COVID-19 Model	Peer Covid	Uganda	2020	673	14.59	2.83	532.00	2.44	2.61	12.15	85.62%
Refugee-Hybrid	Refugee- Hybrid	Uganda	2020	993	15.13	3.20	692.00	3.86	2.55	11.27	79.42%

Appendix F: StrongMinds compared to GiveDirectly

Table 4: StrongMinds compared to GiveDirectly

	GiveDirectly lump-sum CTs	StrongMinds psychotherapy	Multiple of StrongMinds to GiveDirectly
Initial effect	0.244	0.79	
	(0.11, 0.43)	(0.52, 1.2)	
Duration	8.7	5	
	(4, 19)	(3, 10)	
Total effect on SWB & MHa	1.09	1.7	Explains 13% of differences in c-e
	(0.329, 2.08)	(1.1, 2.8)	
Cost per intervention	$1,185.16	$128	Explains 87% of differences in c-e
	($1,098, $1,261)	($70, $300)	
Cost-effectiveness (per $1,000 USD)	**0.916**	**11.79**	**12**
	(0.278, 1.77)	**(7, 21)**	**(4, 24)**