2024 EVALUATIONS

STRONGMINDS®

StrongMinds 2024 Evaluation Report

November 26, 2024

Summary

StrongMinds commissioned independent psychologists to interview representative samples of clients pre-treatment and two-weeks post-treatment, for the second treatment cycle in 2024. A total of 575 clients were interviewed pre- and post-therapy. In Uganda the sample includes adults, out-of-school adolescents, in-school adolescents, and NGO partner clients, while in Zambia the sample is mainly adults.

Clients in both countries experienced large reductions in depressive symptoms on average, with results similar to previous years' evaluations. Across both countries, clients' average PHQ-9 scores were reduced from 15.3 before therapy to 3.2 two weeks after therapy had ended. In Uganda clients' PHQ-9 scores were went from 15.4 to 2.9 on average, and in Zambia from 15.1 to 3.5. Changes in functioning difficulty, subjective wellbeing, and several secondary indicators related to labor supply, nutrition, school attendance, and social support are also presented.

We recommend a follow up survey to measure depressive symptoms and other outcomes after 6-12 months. Additionally, since this was evaluation was not controlled, we recommend proceeding with one or more randomized controlled trials to rigorously assess causal impact.

Methodology

Evaluation method

This evaluation measures changes in the outcomes of StrongMinds clients in Uganda and Zambia, using a simple pre-treatment post-treatment design.

Sample

In each country, the evaluation design called for interviewing 400 clients prior to treatment as well as following up with each client approximately two weeks after treatment.

In Uganda, the plan was to randomly sample 132 adult and out-of-school clients from 4 districts (Kampala, Wakiso, Kotido and Mbale), 132 students from the same districts, and 136 NGO partner clients from two districts (Kamuli and Kanungu). Districts were purposively sampled to ensure 1) that the mean PHQ-9 outcomes from the sampled districts did not differ from the overall population of treated clients using 2023 data, 2) geographic coverage of all regions where StrongMinds operated, and 3) no newly opened districts were included. The selected districts included 47% of all of StrongMinds adult and out-of-school clients, 33% of all students, and 26% of NGO partner clients. Within each district, StrongMinds facilitators were randomly sampled first, and then 4 of each facilitator's potential clients (based on PHQ-4 screening) were randomly sampled.

In Zambia, the original plan was to sample 132 clients in the Ministry of Health (MoH) program Kabwe district (the only district where this program exists), 148 clients from the Peer facilitator program in Lusaka and Kabwe districts (small NGO partner programs also exist in Southern and Western districts and were excluded from this study), and 120 clients from an NGO partner program. In the Peer program, sampling was stratified by 7 Areas within Lusaka. In all programs, facilitators were randomly sampled first, and then 4 of each facilitator's potential clients (based on PHQ-4 screening) were randomly sampled.

In both countries, if potential clients scored below 10 on the PHQ-9 pre-treatment, they were ineligible for treatment and for participation in this study, so they were replaced with randomly sampled clients from the same facilitator.

Data collection

<u>Client journey</u>
StrongMinds programming and monitoring proceeds is as follows: facilitators and StrongMinds staff conduct community mobilization and sensitization on mental health, mental illness and depression, prior to any therapy sessions. People that identify themselves with depressive symptoms are then pre-screened / pre-assessed using the PHQ-4. A second screening using the PHQ-9 tool is done after 2 weeks to confirm depression. People found depressed (i.e. scoring 10 points and above) are assigned groups to start the six weeks therapy sessions. Upon completing therapy, the clients are assessed with the PHQ-9 tool again to measure whether there has been any reduction in the depression symptoms. This evaluation builds on top of that timeline.

Data was collected by independent psychologists who were trained be the StrongMinds country M&E and Program staff. Pre-treatment, the psychologists were introduced to the potential clients by the StrongMinds facilitators who had pre-screened them with the PHQ-4. The psychologists then would conduct their interviews in private, instead of the facilitator, and if the potential client scored 10 or higher on the PHQ-9 they were eligible for treatment. Potentially suicidal clients were immediately referred to StrongMinds staff for management, following StrongMinds' suicide protocol.

Data was collected using KoboToolbox. Extensive data quality procedures were used, including 100% audio auditing, daily data checks using Stata, and unannounced spot checks by M&E staff. Errors were corrected during daily debriefs based on the data quality procedures' findings.

The most important data collected is based on the Patient Health Questionnaire-9 (PHQ-9), which is a 9-item scale used to assess the severity of depressive symptoms. The PHQ-9 is a list of 9 questions asking how often respondents have been bothered by different depressive symptoms in the past 2 weeks. Responses are rated on a 4-point Likert-type scale, ranging from 0 (not at all) to 3 (nearly every day). Total score can range from 0 to 27, with high scores meaning high depression.

Therapy
In StrongMinds therapy groups, facilitators try to create a safe space for clients to open up with peers about their worries and struggles. Over six sessions, these lay counselors guide structured discussions to help participants identify their underlying triggers of depression and examine how their interpersonal relationships and depression symptoms are linked. Group members work together to strategize solutions to their problems, learn coping mechanisms, practice interpersonal skills, and identify support structures that they can continue to lean on after therapy has ended.

Two-week follow-up
The independent psychologists used the same protocols to interview the same clients two-weeks after they had completed therapy.

Analysis

The main outcome variables are means or proportions of levels and pre-post differences, analyzed using Stata. The analysis applies sampling weights to account for district sampling and stratification by program. Standard errors are clustered at the facilitator level. The results are representative of the programs and districts included in the study.

Results

Sample

Uganda: In the adult and out-of-school sample, 118 out of 133 (88.7%) completed pre- and post-treatment interviews, as did 124 of 133 (93.2%) NGO clients and 123 of 139 (88.5%) students.[1] Of these, 1 partner client, 1 adult client and 3 students did not attend any sessions and were excluded from the analysis. The final sample size is 360 for Uganda.

Zambia: The evaluation for the NGO program had to be canceled during pre-treatment data collection for practical reasons.[2] Following this, the sample size for the other two programs was slightly increased to make use of available data collection time. For the MoH program, 149 out of 160 clients (93.1%) completed pre- and post-treatment interviews, as did 158 of 170 Peer program clients (92.4%). Of these, 5 MoH clients and 1 peer client did not attend any sessions and were excluded from the analysis. In addition, 49 MoH clients and 42 Peer clients were removed from analysis, as they were accidentally included in the study without being part of the sampling plan.[3] The final sample is 215 for Zambia.

Results by country

Table 1 gives the main PHQ-9 related outcomes. Uganda and Zambia have similar PHQ-9 outcomes, which are also in line with the 2023 results. The differences between countries and over time are not statistically significant at $p=0.05$ for the PHQ-9 scores post-treatment or for symptom reductions scores. Across both countries, clients' average PHQ-9 scores were reduced from 15.3 to 3.2.

Table 1: PHQ-9 based indicators

	Uganda (n = 360)				**Zambia** (n = 215)			
	Mean	95% CI			Mean	95% CI		
Pre-treatment avg.	15.4	14.7	−	16.2	15.1	14.0	−	16.3
2-weeks post-treatment avg.	2.9	2.3	−	3.5	3.5	2.2	−	4.8
Symptom reduction avg.	12.5*	11.6	−	13.3	11.6*	10.0	−	13.2
5 pt. reduction	92.8%*	88.5%	−	97.1%	84.6%*	75.0%	−	94.2%
10 pt. reduction	77.0%*	70.7%	−	83.4%	64.7%*	52.8%	−	76.6%
Depression free, 0-4	74.1%	67.1%	−	81.0%	72.5%	58.2%	−	86.9%
Mild, 5-9	18.4%	13.2%	−	23.6%	17.9%	8.3%	−	27.4%
Moderate, 10-14	5.9%	2.0%	−	9.8%	4.8%	-0.1%	−	9.7%
Mod-sev, 15-19	1.6%	0.1%	−	3.2%	3.7%	0.0%	−	8.4%
Severe, 20+	0.0%	0.0%	−	0.0%	1.1%	0.0%	−	3.3%

* $p < 0.05$. Note: hypothesis tests were conducted for all outcomes that measured changes, compared to 0 change.

---

[1] The pre-treatment sample slightly exceeded the sample set in the study design

[2] The psychologists that conducted the interviews needed to ask students to stay late after school in order to complete their interviews before treatment began. This was unacceptable to parents and the NGO partner, so this portion of the evaluation was terminated.

[3] Specifically, the plan was to sample only 4 eligible clients per facilitator, but some psychologists included extra clients for some non-random facilitators

The two graphs below compare the average pre-treatment and two-weeks post-treatment PHQ-9 of the 2023 and 2024 evaluations with historical results from Uganda in 2014 – 2019. The first graph shows the pre-treatment results have been fairly constant since 2016, always hovering around 15. And for post-treatment, the results from the two most recent years in Uganda are slightly higher than the 2014–2019 average of 2.5, though not outside the typical range. The results in Zambia are slightly higher than in Uganda, though the difference in means is not statistically significant, and are at the upper edge of the historical distribution. These slightly higher numbers in both countries *could* reflect a real increase in depressive symptoms, perhaps due to marginally reduced quality as StrongMinds expands rapidly. They could also be due to expansion into more challenging contexts. Alternatively, it could reflect stricter data quality standards, as we have moved to having independent psychologists collect data, along with other quality improvements. Unfortunately, we cannot be sure from this data alone. Despite this small increase, the results still compare very favorably to other depression treatments, with 73-74% depression free.

Pre-treatment average PHQ-9
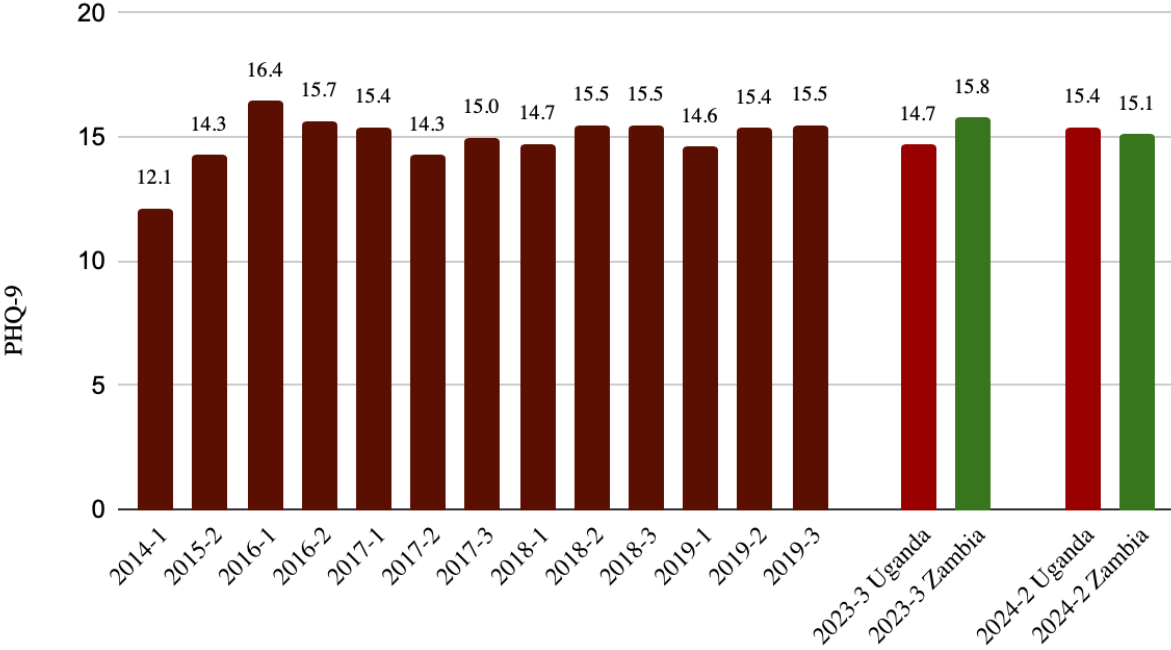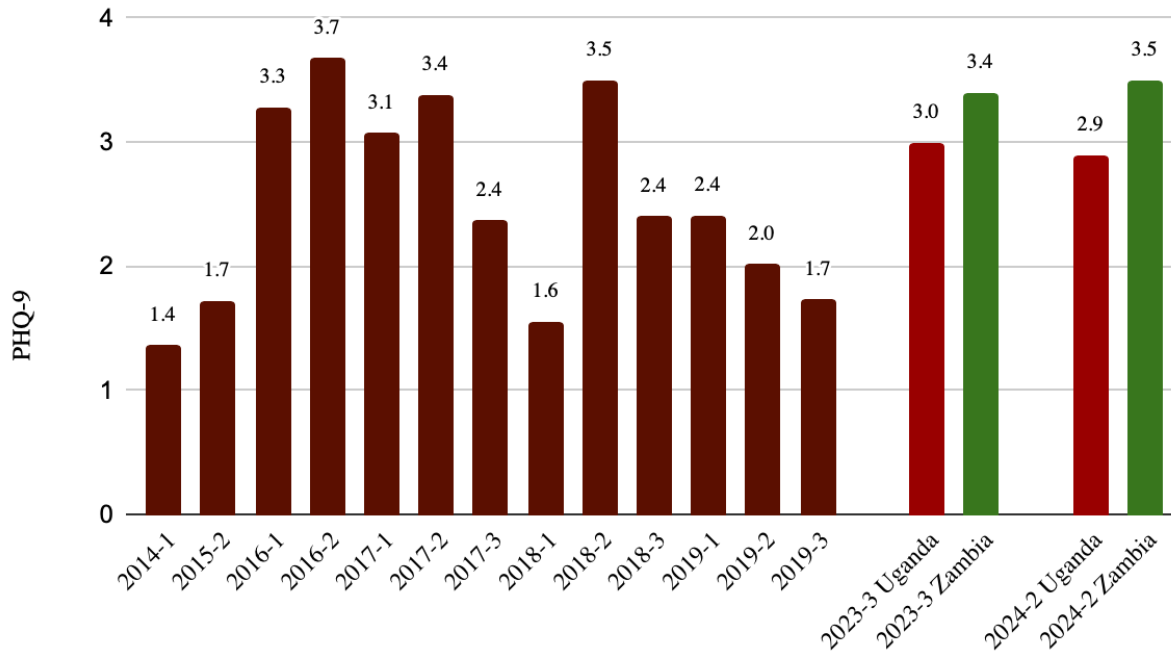
Two-weeks post-treatment average PHQ-9

Table 2 gives the results for responses to the functional impairment that is included with the PHQ-9: "How difficult have your symptoms made it for you to do your work, take care of things at home, or get along with other people?" This question aligns with the WHO's International Classification of Functioning, Disability, and Health,[4] which includes physiological functions, activities such as work, and participation in society. The functional impairment question is a quick way to assess the practical effects of depressive symptoms on key domains of life. In clinical settings, responses of "very difficult" or "extremely difficult" can indicate a need for more intensive interventions, such as psychotherapy, medication, or combined approaches. These responses are predictive of slower recovery and poorer treatment outcomes compared to "not difficult" or "somewhat difficult" responses.

On average, clients in both countries experienced large reductions in difficulty functioning. In Uganda, 72% of clients responded "very" or "extremely" difficult prior to treatment, and by two-weeks after treatment only 10% said the same. In Zambia, these proportions fell from 66% to 6%. The 62-percentage point reduction in Uganda is almost identical to the 63-percentage point reduction in 2023 (this was not measured in Zambia before 2024).

---

[4] WHO, accessed Nov. 2024: https://www.who.int/standards/classifications/international-classification-of-functioning-disability-and-health

6

Table 2: Functioning difficulty

	Uganda (n = 360)			Zambia (n = 215)		
	Mean	95% CI		Mean	95% CI	
Pre-treatment						
Not difficult at all	1.3%	0.0% – 3.1%		1.9%	0.0% – 3.9%	
Somewhat difficult	26.6%	21.0% – 32.3%		32.3%	22.2% – 42.4%	
Very or extremely difficult	72.1%	66.0% – 78.1%		65.9%	55.2% – 76.5%	
Post-treatment						
No symptoms	36.6%	28.8% – 44.3%		31.1%	20.6% – 41.5%	
Not difficult at all	26.3%	19.5% – 33.1%		44.8%	35.7% – 53.9%	
Somewhat difficult	26.8%	21.2% – 32.4%		17.8%	9.0% – 26.7%	
Very or extremely difficult	10.3%	5.3% – 15.4%		6.3%	0.8% – 11.8%	
Reduction in very or extremely difficult	61.7%*	54.4% – 69.1%		59.6%*	45.3% – 63.9%	

* p < 0.05. Note: hypothesis tests were conducted for all outcomes that measured changes, compared to 0 change.

Table 3 presents the results from a subjective well-being (SWB) question used in the United Nations World Happiness Report: "Please imagine a ladder, with steps numbered from 0 at the bottom to 10 at the top. The top of the ladder represents the best possible life for you and the bottom of the ladder represents the worst possible life for you. On which step of the ladder would you say you personally feel you stand at this time?" Prior to treatment, clients in both countries evaluated their lives to be below the average values in their countries, not surprisingly. Clients experienced large improvements in SWB in both countries and were far ahead of the national averages two-weeks post-treatment.

Table 3: Subjective well-being

	Uganda (n = 367)			Zambia (n = 292)		
	Mean	95% CI		Mean	95% CI	
Pre-treatment	3.2	2.9 – 3.4		2.8	2.1 – 3.4	
Post-treatment	7.1	6.7 – 7.6		7.1	6.3 – 7.8	
Change	4.0*	3.4 – 4.5		4.3*	3.4 – 5.3	
National average (UN)	4.4			3.5		

* p < 0.05. Note: hypothesis tests were conducted for all outcomes that measured changes, compared to 0 change.

Table 4: Secondary indicators

		Uganda	Zambia
1. Did not miss work in the past week		(n = 237)	(n = 197)
	Pre	21%	26%
	Post	41%	61%
	Change	20%*	36%*
2. Meals children had in the past 24 hours		(n = 194)	(n = 166)
	3 meals		
	Pre	26%	20%
	Post	50%	35%
	Change	24%*	15%
	2+ meals		
	Pre	73%	54%
	Post	90%	77%
	Change	17%*	24%*
	# of meals		
	Pre	2.0	1.6
	Post	2.4	2.1
	Change	0.4*	0.4*
3. Children did not miss school in the past week		(n = 155)	(n = 120)
	Pre	52%	57%
	Post	72%	67%
	Change	20%*	10%
4. Have someone for support		(n = 237)	(n = 208)
	Pre	61%	66%
	Post	80%	68%
	Change	19%*	2%

* $p < 0.05$. Note: hypothesis tests were conducted for all outcomes that measured changes, compared to 0 change.

Table 4 above shows the four secondary indicators StrongMinds uses for adults and out of school adolescents. There were large gains across most indicators for both countries, though in Zambia the gains were smaller for school attendance and for having someone for support, and larger for not missing any work in the past week.

The sample sizes vary somewhat question to question. The first and fourth questions are asked to everyone, the second is only asked to adults with children living at home, and the third question is only asked for adults with children enrolled in school.

The most notable difference from this year to last year is the dramatic decrease in meals eaten in Zambia. Last year, on average Zambians ate 2.6 meals per day pre-treatment and 2.8 post, but these number have fallen to 1.6 pre and 2.0 post. This might be due to the drought and its effects that led to crop shortages and much higher food prices.

Table 5: SIT indicators for students

		Uganda
1. Did not miss school in the past week		(n = 120)
	Pre	63.7%
	Post	73.3%
	Change, p.p.	9.5%
	Change, %	14.9%
2. Grades: "good", "very good", and "excellent"		
	Pre	35.0%
	Post	85.5%
	Change, p.p.	50.5%*
	Change, %	144.3%
3. Hope: "always" or "often" feel hopeful about your future		
	Pre	50.1%
	Post	52.2%
	Change, p.p.	2.0%
	Change, %	4.0%

* $p < 0.05$. Note: hypothesis tests were conducted for all outcomes
that measured changes, compared to 0 change.

Table 5 shows students had a 10 percentage point increase in perfect attendance for the prior week, and a dramatic 51 percentage point increase in the percentage reporting that their grades are good, very good, or excellent, as opposed to fair or poor.[5] However, the share that always or often felt hopeful about their future remained essentially unchanged. Zambia's school program in partnership with the Ministry of Education was in early stages as we planned this evaluation, and will be included in future studies.

---

[5] We are in the process of validating self-reported grades with schools' records, to be published later.

Results by program

Table 6: PHQ-9 related indicators by program

	Uganda			
	MoH Adult (n = 71)	Out of school (n = 46)	In-school (n = 120)	NGOs (n = 123)
Pre-treatment avg.	15.3	16.3	14.5	15.8
2-weeks post-treatment avg.	3.1	2.6	2.8	2.8
Symptom reduction avg.	12.2*	13.6*	11.6*	13.0*
5 pt. reduction	90.7%*	94.8%*	95.8%*	92.9%*
10 pt. reduction	78.7%*	79.1%*	69.1%*	79.4%*
Depression free, 0-4	72.3%	71.9%	75.9%	77.9%
Mild, 5-9	19.7%	22.1%	18.5%	12.3%
Moderate, 10-14	5.6%	6.1%	5.5%	7.0%
Mod-sev, 15-19	2.3%	0.0%	0.0%	2.8%
Severe, 20+	0.0%	0.0%	0.0%	0.0%

* $p < 0.05$. Note: hypothesis tests were conducted for all outcomes that measured changes, compared to 0 change.

	Zambia			
	MoH Kabwe (n = 98)	PTG (combined) (n = 117)	PTG Kabwe (n = 67)	PTG Lusaka (n = 50)
Pre-treatment avg.	15.8	14.9	15.0	14.9
2-weeks post-treatment avg.	3.1	3.6	3.9	3.6
Symptom reduction avg.	12.7*	11.3*	11.1*	11.4*
5 pt. reduction	91.1%*	82.8%*	83.1%*	82.7%*
10 pt. reduction	72.3%*	62.6%*	64.7%*	62.1%*
Depression free, 0-4	74.3%	72.1%	78.0%	70.7%
Mild, 5-9	17.8%	17.9%	13.3%	19.0%
Moderate, 10-14	4.9%	4.7%	0.9%	5.6%
Mod-sev, 15-19	3.0%	3.9%	0.0%	4.7%
Severe, 20+	0.0%	1.4%	7.8%	0.0%

* $p < 0.05$. Note: hypothesis tests were conducted for all outcomes that measured changes, compared to 0 change.

Table 6 above breaks down the country averages into different programs, and in Zambia the Peer Therapy program is divided by two districts.

Perhaps the best single indicator to compare performance is the "depression free" rate: the share of clients who scored below 5 on the PHQ-9 post therapy, which is also called "minimal depression." In the graph below we see a fairly narrow range of depression free status, ranging from 70.7% to 77.9%, with an average across both countries of 74.0%.

Proportion of clients "depression free" (PHQ-9 < 5) 2-weeks post
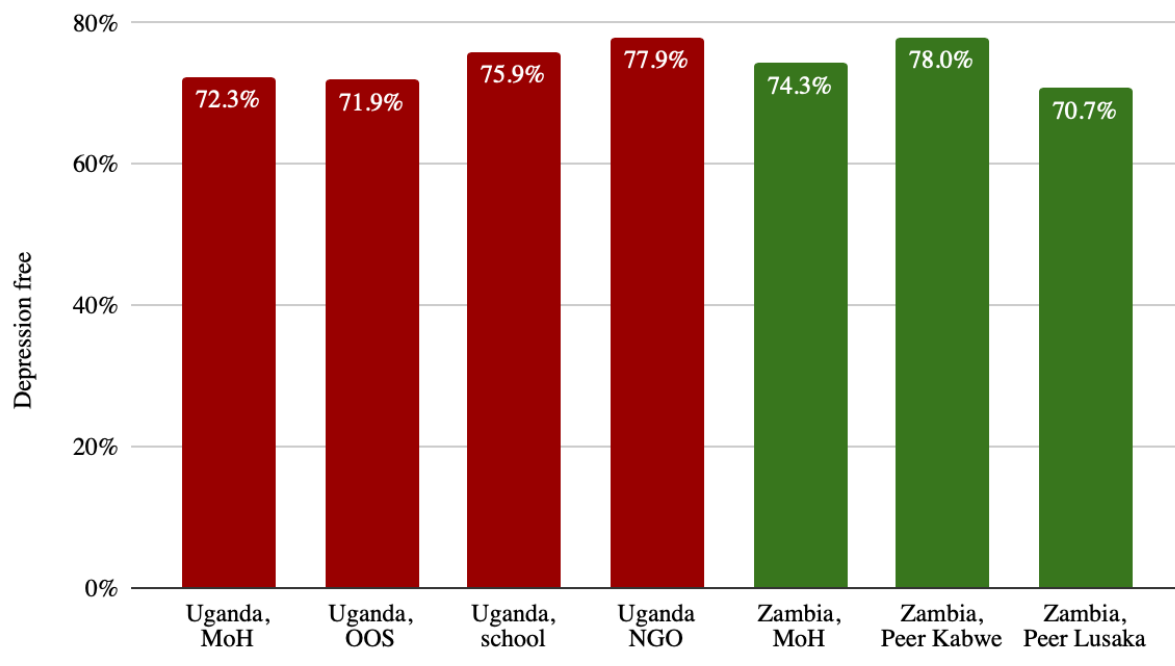
Table 7: Functioning difficulty by program

	Uganda			
	MoH Adult (n = 71)	Out of school (n = 46)	In-school (n = 120)	NGOs (n = 123)
Pre-treatment				
Not difficult at all	1.9%	1.4%	0.3%	1.0%
Somewhat difficult	23.4%	24.4%	43.8%	18.7%
Very or extremely diff.	74.7%	74.2%	55.9%	80.2%
Post-treatment				
No symptoms	30.9%	49.1%	35.7%	40.2%
Not difficult at all	26.9%	19.3%	22.4%	34.5%
Somewhat difficult	27.6%	26.5%	35.1%	17.3%
Very or extremely diff.	14.6%	5.2%	6.9%	8.1%
Reduction in very or extremely difficult, p.p.	60.1%*	69.1%*	49.0%*	72.1%*

* $p < 0.05$. Note: hypothesis tests were conducted for all outcomes that measured changes, compared to 0 change.

	Zambia			
	MoH Kabwe (n = 98)	PTG (combined) (n = 117)	PTG Kabwe (n = 67)	PTG Lusaka (n = 50)
Pre-treatment				
Not difficult at all	1.0%	2.1%	0.3%	2.5%
Somewhat difficult	32.6%	32.2%	50.5%	28.1%
Very or extremely difficult	66.4%	65.7%	49.2%	69.4%
Post-treatment				
No symptoms	38.1%	29.1%	30.9%	28.7%
Not difficult at all	35.2%	47.5%	50.7%	46.8%
Somewhat difficult	21.8%	16.7%	6.7%	19.0%
Very or extremely difficult	4.9%	6.7%	11.7%	5.6%
Reduction in very or extremely difficult, p.p	61.4%*	59.0%*	37.5%*	63.8%*

* $p < 0.05$. Note: hypothesis tests were conducted for all outcomes that measured changes, compared to 0 change.

Table 7 above displays fairly similar trends in functioning as compared to PHQ-9 outcomes. In Uganda, students attending school do not see as much improvement as one might expect based on changes in depressive symptoms. In Zambia, the two PTG districts ranking reversed: Kabwe had better PHQ-9 outcomes but a lower reduction in functioning difficulty, which cannot be explained by this data alone. Notably, Kabwe had the lowest share reporting "very" or "extremely" difficult functioning pre-treatment, so had less room to improve.

Table 8: Secondary Indicators by program

	Uganda		
	MoH Adult	Out of school	NGOs
1. Did not miss work in the past week	(n = 71)	(n = 43)	(n = 123)
Pre	22.2%	11.7%	25.6%
Post	34.7%	31.5%	61.2%
Change	12.5%	19.8%*	35.6%*
2. Meals children had in the past 24 hours	(n = 59)	(n = 32)	(n = 103)
3 meals			
Pre	22.8%	26.8%	33.3%
Post	51.6%	53.0%	45.7%
Change	28.8%*	26.2%*	12.4%*
2+ meals			
Pre	64.8%	76.7%	88.2%
Post	87.5%	95.8%	93.1%
Change	22.7%*	19.0%	4.8%
# of meals			
Pre	1.8	2.0	2.2
Post	2.4	2.5	2.4
Change	0.6*	0.5*	0.2*
3. Children did not miss school in the past week	(n = 51)	(n = 20)	(n = 84)
Pre	47.5%	53.8%	60.6%
Post	67.2%	54.2%	90.9%
Change	19.7%	0.4%	30.3%*
4. Have someone for support	(n = 71)	(n = 43)	(n = 123)
Pre	57.9%	65.3%	64.5%
Post	77.9%	90.3%	75.3%
Change	19.9%*	25.0%*	10.8%*

* p < 0.05. Note: hypothesis tests were conducted for all outcomes that measured changes, compared to 0 change.

	Zambia			
	MoH Kabwe	PTG (combined)	PTG Kabwe	PTG Lusaka
1. Did not miss work in the past week	(n = 88)	(n = 109)	(n = 63)	(n = 46)
Pre	32.4%	23.7%	28.0%	22.8%
Post	61.4%	61.5%	69.6%	59.7%
Change	29.0%*	37.8%*	41.6%*	36.9%*
2. Meals children had in the past 24 hours	(n = 77)	(n = 89)	(n = 52)	(n = 37)
3 meals				
Pre	21.6%	19.8%	1.1%	24.8%
Post	30.6%	36.2%	12.6%	42.5%
Change	8.9%	16.4%	11.5%	17.7%
2+ meals				
Pre	53.5%	53.8%	19.4%	63.0%
Post	68.2%	80.1%	68.9%	83.1%
Change	14.7%	26.3%*	49.6%*	20.1%
# of meals				
Pre	1.7	1.6	1.0	1.8
Post	1.9	2.1	1.7	2.2
Change	0.2	0.5	0.7	0.4
3. Children did not miss school in the past week	(n = 60)	(n = 60)	(n = 36)	(n = 24)
Pre	79.6%	49.6%	47.9%	50.0%
Post	75.6%	64.0%	76.7%	61.0%
Change	-4.0%	14.4%	28.8%	11.0%
4. Have someone for support	(n = 94)	(n = 114)	(n = 67)	(n = 47)
Pre	66.0%	66.4%	66.4%	66.4%
Post	74.3%	66.0%	64.4%	66.4%
Change	8.2%	-0.4%	-2.0%	0.0%

* $p < 0.05$. Note: hypothesis tests were conducted for all outcomes that measured changes, compared to 0 change.

Table 8 above shows some interesting variation by program. In Uganda, work attendance improved significantly more for the NGO programs than the MoH adult or out of school adolescent programs. Zambia had improvements similar to Uganda's NGO programs.

For meals, in Uganda there were larger improvements for the MoH adult and out of school adolescent programs compared to NGO programs. In Zambia the PTG programs saw larger gains in meals compared to the MoH program, but as noted earlier, the total number of meals consumed per day were significantly lower than in the previous years' evaluation. This is likely due to the drought and drought-related food price inflation, and is a serious concern.

There are three instances in Zambia where outcomes appear to have worsened, however these are not statistically significant: the share of children with perfect school attendance in the past week in the MoH program went down 4 percentage points, and the share of people who say they have someone for support decreased by 0.4 and 2.0 percentage points for PTG overall and for PTG Kabwe, while it remained unchanged in Lusaka. Technically, none of these results are significantly different than 0. In 2023 86% of Zambian clients reported having someone for support six months after therapy had ended, notably more than this year, so it is worth exploring more deeply what might have changed.

This secondary outcome appears to be strongly linked to depression outcomes, though it is impossible to determine causality based on the data we have. Table 9 below shows that the group with the highest share of "depression-free" clients were those who had no support pre-treatment but had gained support by the end of treatment: 88%. Whereas only 45% of those who felt they did not have support before or after therapy were "depression-free".

Table 9: Depression-free status by Pre- and Post-Treatment Support, Zambia

		Post-treatment	
		No support	Have support
Pre-treatment	No support	45.5% (n = 33)	88.4% (n = 43)
	Have support	66.7% (n = 30)	82.4% (n = 102)

Gender analysis

Below is the share of clients in the evaluated programs that are female, by country, when looking at all treated clients (i.e. not just the evaluation sample):

Uganda
1. MoH: 88%
2. Out of school: 84%
3. School: 58%:
4. NGO partners: 77%

Zambia
1. MoH: 81%
2. Peer: 89%

Since the vast majority of clients are female, and because we did not oversample males, we only have sufficient power for a gender analysis in the school program in Uganda. Our student sample includes 65 girls and 55 boys, and the results are almost identical across genders. Girls have a mean two-week post-treatment PHQ-9 of 2.83, and boys 2.82, and the difference is not statistically significantly significant.

Validating program monitoring data with evaluation data

The evaluation clients were sampled in a manner that allows us to validate the accuracy of our program monitoring data. They were randomly sampled from a list of all clients their facilitators treated, and the facilitators were randomly sampled from a larger group of all the facilitators in selected districts. The sampled clients had their pre-treatment and post-treatment data collected in ideal conditions with strong data quality controls, as mentioned earlier in this report. The non-sampled clients had their data collected by their facilitators, who are not as highly trained and monitored as the psychologists that collected the evaluation data. In theory, this data collected by facilitators could be biased: for example, we could speculate that the facilitators artificially inflate the pre-treatment scores and artificially reduce the post-treatment scores. Since we have the random sample surveyed by independent psychologists, we can directly compare the data and run two types of tests:

1. Compare sampled clients to non-sampled clients of the same facilitators and see whether the mean and variance of these two groups are statistically different.

2. Compare the clients of sampled facilitators to non-sampled facilitators and see whether the means are statistically different. This checks whether our random sampling produced a representative sample. If it did, then we can validate all of our program monitoring data from all facilitators, not just the sampled facilitators

The analysis focuses on pre-treatment PHQ-9 scores since that was collected at the same time by facilitators and independent psychologists. Whereas we also compare post-treatment scores, we do not expect those to be the same because of different timing: facilitators collect PHQ-9 immediately at the end of the final therapy session (what we call termination), whereas the evaluation happened two weeks later. We include this analysis to quantify the difference from termination to two-weeks post therapy. The difference may be explained by a mix of social desirability bias, facilitator bias, and actual symptom changes over two weeks. Therefore, this analysis is not meant to validate the termination outcomes directly.

Validating pre-treatment PHQ-9 scores

Table 10 below presents the results in which we regressed the pre-treatment score on a variable that indicated sampled status. In this and all following regressions standard errors were clustered at the facilitator level. The sampled clients had slightly higher pre-treatment PHQ-9 scores on average in both countries, though the difference was not statistically significant. In Uganda the sampled clients were 0.35 points above the mean for non-sampled clients (p=0.21), and in Zambia the difference was 0.37 (p=0.43).

Table 10: Comparing sampled to non-sampled clients of the same facilitators, pre-treatment PHQ-9

Uganda

Predictor	Coefficient	Robust S. E.	p-value	Observations
Sampled clients	0.35	0.28	0.21	383
Constant (not sampled)	14.77	0.23	0.00	2,735

Zambia

Predictor	Coefficient	Robust S. E.	p-value	Observations
Sampled clients	0.37	0.43	0.39	226
Constant (not sampled)	14.82	0.32	0.00	2,024

Table 11 below compares all the clients of sampled facilitators to non-sampled facilitators in the same programs and districts. In both Uganda and Zambia, the clients of sampled facilitators had mean pre-group PHQ-9 scores that were slightly lower than for the clients of non-sampled facilitators, but these differences was not statistically significant at conventional levels (Uganda: $p=0.13$; Zambia: $p=0.66$).

Table 11: Comparing clients of sampled facilitators to clients of non-sampled facilitators in the same programs and districts, PHQ-9

Uganda

Predictor	Coefficient	Robust S. E.	p-value	Observations
Sampled facilitators	-0.35	0.23	0.13	3,118
Constant (not sampled)	15.17	0.10	0.00	19,566

Zambia

Predictor	Coefficient	Robust S. E.	p-value	Observations
Sampled facilitators	-0.14	0.31	0.66	2,202
Constant (not sampled)	14.97	0.09	0.00	29,814

Table 12 below extends the comparison to all clients in all programs and all districts. Once again, the differences were small and not statistically significant.

Table 12: Comparing clients of sampled facilitators to clients of non-sampled facilitators in all programs and districts

Uganda

Predictor	Coefficient	Robust S. E.	p-value	Observations
Sampled facilitators	-0.35	0.22	0.11	3,118
Constant (not sampled)	15.16	0.06	0.00	47,561

Zambia

Predictor	Coefficient	Robust S. E.	p-value	Observations
Sampled facilitators	0.08	0.31	0.79	2,250
Constant (not sampled)	14.78	0.09	0.00	36,365

We also explored whether the standard deviations were different for the same comparisons. When comparing sampled clients to non-sampled clients of the same facilitators, the standard deviations were 3.80 and 3.12 respectively (p=0.00) in Uganda, and 4.17 vs. 2.90 in Zambia (p=0.00). These significant differences seem to be due to a higher proportion of scores greater than 20 among sampled clients (the skewness was 0.68 vs. 0.43 in Uganda and 0.60 vs. 0.26). The differences in standard deviations for the other two comparisons - clients of sampled facilitators vs. non-sampled facilitators in the same programs and districts, or vs. all non-sampled facilitators - were smaller and not statistically significant in Uganda, and small but statistically significant in Zambia (2.86 vs. 2.75).

Comparing termination and two-week post-treatment scores

When assessed on the last day of therapy (termination) by their facilitators in Uganda, the StrongMinds clients who were part of the evaluation sampled scored an average of 1.82 on the PHQ-9. Below 5 is considered no or minimal depression. When assessed by independent psychologists two weeks later, the same clients scored an average of 3.07, for a difference of 1.25 points, which was statistically significant (p=0.00). The majority (56.6%) had the same or lower score two weeks after therapy. The difference between termination and two-weeks post-treatment is small from a clinical perspective, and small relative to the average score change from pre- to post-treatment of 12.5 points. As mentioned above, we cannot be sure whether this represents a real increase in depressive symptoms or reflects bias on the side of the clients or facilitator during termination.

In Zambia, the mean termination score was 3.19, and two weeks later, when assessed by the psychologists, it had slightly increased to 3.25, though this change was not statistically significant (p=0.87). While these averages are surprisingly close, there is still a possibility that the termination scores include some bias on the part of clients or facilitators.

Validation summary

Taking these results together, we consider this to be a strong indication that the program monitoring data is accurate. The independent data has pre-treatment means that were similar and not statistically significantly different from the program monitoring data. And while we did not directly validate the termination data, the mean of the post-two week data was not statistically different in Zambia, and only slightly higher in Uganda, giving us some confidence that the termination data is likely to be accurate.

In future research we plan to investigate two issues in more depth. First, we will investigate why the standard deviations are higher in the psychologist-collected data than in the program monitoring data – we know that psychologists observe a relatively higher proportion of people with scores greater than 20, but does this reflect a higher depressive symptom burden or is it measurement error? Second, we will consider validating termination data directly, instead of comparing it to two-week post-treatment data. If validated, this would make us more comfortable to use program monitoring data for research purposes, which among other uses, could enable lean A/B testing.

Limitations and further research

This study demonstrates that StrongMinds clients experience significant symptom reduction during the six-week therapy. They also report large improvements in functional impairment, subjective well-being, and many secondary indicators. However, as there is no control group in this study, our ability to attribute these changes directly to the StrongMinds program is limited. We recommend proceeding with one or more randomized controlled trials to rigorously measure the causal impact and cost-effectiveness of StrongMinds' programs.

Appendix: Combined averages

The following tables combine the outcomes for Uganda and Zambia, taking a weighted average based on how many clients were treated in cycle 2, 2024.

PHQ-9 based indicators

Pre-treatment avg.	15.3
2-weeks post-treatment avg.	3.2
Symptom reduction avg.	12.1
5 pt. reduction	89.3%
10 pt. reduction	71.8%
Depression free, 0-4	73.4%
Mild, 5-9	18.2%
Moderate, 10-14	5.4%
Mod-sev, 15-19	2.5%
Severe, 20+	0.5%

Functioning difficulty

Pre-treatment	
Not difficult at all	1.5%
Somewhat difficult	29.0%
Very or extremely difficult	69.4%
Post-treatment	
No symptoms	34.2%
Not difficult at all	34.1%
Somewhat difficult	23.0%
Very or extremely difficult	8.6%
Reduction in very or extremely difficult, percentage point	60.8%
Reduction in very or extremely difficult, %	87.6%

Subjective well-being

Pre-treatment	3.0
Post-treatment	7.1
Change	4.1

Secondary indicators

1. Did not miss work in the past week

Pre	23.0%
Post	49.5%
Change, percentage point	26.5%
Change, %	115.2%

2. Meals children had in the past 24 hours

3 meals	
Pre	23.7%
Post	43.8%
Change, percentage point	20.1%
Change, %	84.9%
2+ meals	
Pre	64.8%
Post	84.9%
Change, percentage point	20.0%
Change, %	30.9%
# of meals	
Pre	1.8
Post	2.3
Change, meals	0.4
Change, %	23.7%

3. Children did not miss school in the past week

Pre	54.0%
Post	69.7%
Change, percentage point	15.7%
Change, %	29.0%

4. Have someone for support

Pre	63.2%
Post	74.6%
Change, percentage point	11.4%
Change, %	18.0%