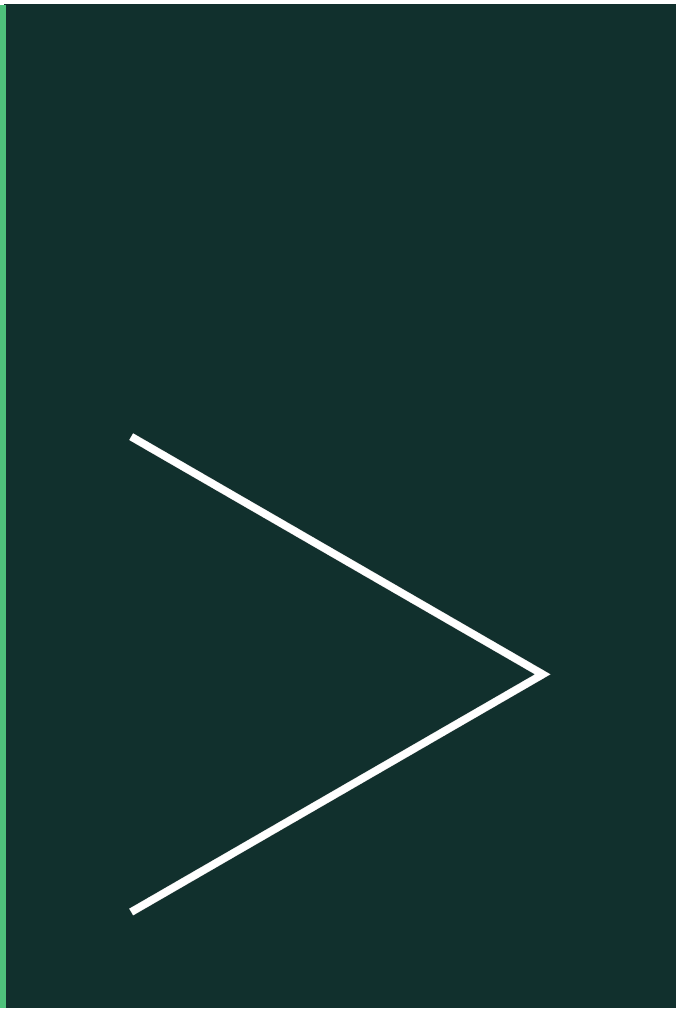
Founders Pledge

# Charity Recommendation Report: Action for Happiness

AUTHORS: AIDAN GOTH, MMATHPHIL AND SJIR HOEIJMAKERS, MSC

LAST UPDATED 01/2020



# Executive Summary

## 1. Money, health, and subjective well-being

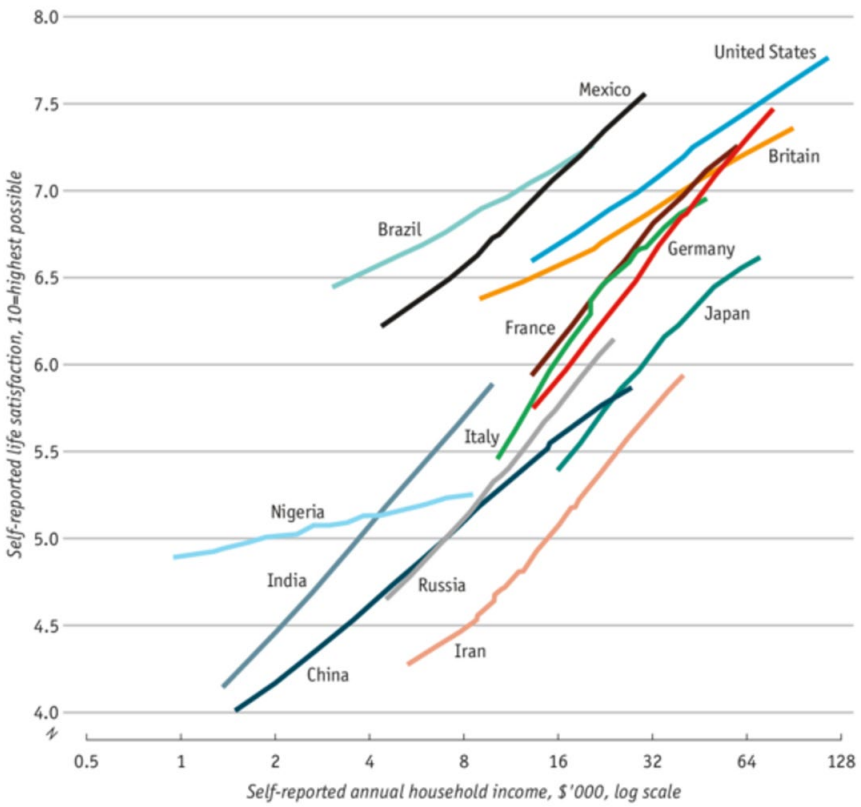
Almost everyone cares about experiencing positive well-being: to be happy and satisfied with life, and free from negative emotions and depression. Often, when we try to improve the world, we try to increase people's economic status or their health, but it is often unclear how well these things translate into subjective well-being.

For instance, as Figure 1 shows, increasing income only has a weak effect on increasing subjective well-being.[1] Here, household income on the x-axis is shown on a logarithmic scale: the gap between $1,000 and $2,000 is the same as the gap between $32,000 and $64,000. The data suggest that income has a rapidly declining effect on subjective well-being the richer you get: increasing your income from $1,000 to $2,000 has roughly the same effect as increasing your income from $32,000 to $64,000. There is evidence that the effect declines to zero once equivalised household income reaches around $95,000.[2]

[1] Betsey Stevenson and Justin Wolfers, "Subjective Well-Being and Income: Is There Any Evidence of Satiation?," American Economic Review 103, no. 3 (May 2013): 598–604, https://doi.org/10.1257/aer.103.3.598.

[2] Andrew T. Jebb et al., "Happiness, Income Satiation and Turning Points around the World," Nature Human Behaviour 2, no. 1 (January 2018): 33–38, https://doi.org/10.1038/s41562-017-0277-0.

Figure 1: Self-reported life satisfaction and self-reported annual household Income.




Source: "Money Can Buy Happiness, Money Can Buy Happiness," The Economist, accessed January 27, 2020, https://www.economist.com/graphic-detail/2013/05/02/money-can-buy-happiness.

There is also evidence that health problems have a much smaller effect on subjective well-being than one might imagine.[3]

[3] Paul Dolan and Daniel Kahneman, "Interpretations Of Utility And Their Implications For The Valuation Of Health*," The Economic Journal 118, no. 525 (2008): 215–34, https://doi.org/10.1111/j.1468-0297.2007.02110.x.

## 2. The case for focusing on subjective well-being

Because income fails to track subjective well-being accurately in some cases, it is important to look outside the typical realm of economic analysis when identifying the best opportunities for improving people's lives. By failing to consider subjective well-being directly, it is possible that many philanthropists and governments miss out on some outstanding opportunities to do good.

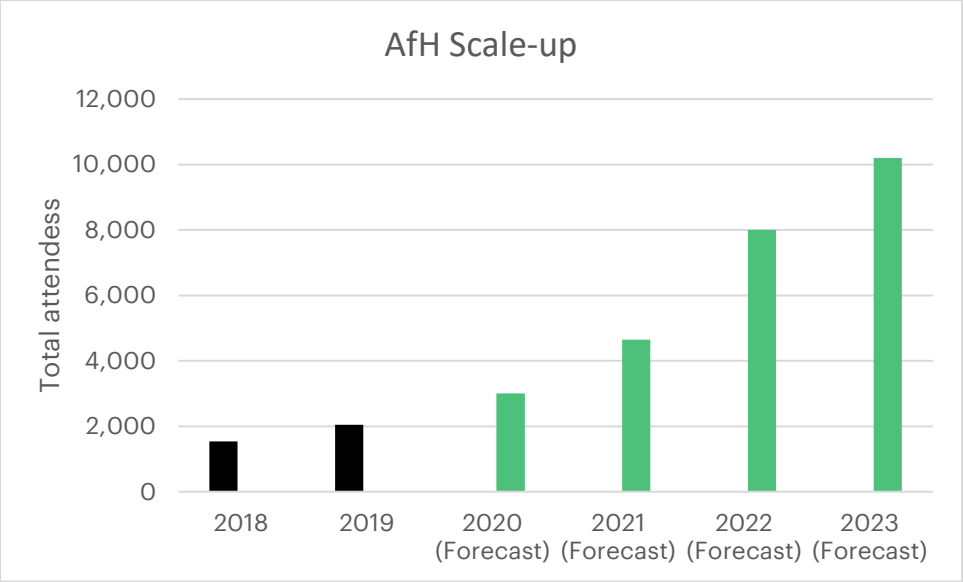
Over the last year, the Founders Pledge research team has explored ways to increase subjective well-being directly. During the course of this research, we came across the charity Action for Happiness. Their programme seemed promising in improving participants' subjective well-being, and their scale-up seemed like a highly leveraged funding opportunity. This prompted us to carry out a more in-depth evaluation, resulting in our recommendation and this report. We plan to expand our work on how best to improve subjective well-being in the future.

## 3. Charity recommendation: Action for Happiness

Action for Happiness (AfH) is a UK-based charity that brings people together in small, face-to-face groups to explore what really matters for a happy and meaningful life. AfH is trying to build a community of people transforming their own lives to be happier and to help those around them. AfH provides 8-week courses, called Exploring What Matters (EWM), run by volunteers in their local community. The course aims to help people to tune in to what really matters for a happy and meaningful life, connect with others in meaningful face-to-face conversations, and to take action to boost happiness for participants and for others. Most courses to date have taken place in the UK, but courses have also been run in 20 countries around the world, including the US, Australia, Germany and Italy.

AfH is planning a five-fold scale-up over the next three years. In 2018, AfH provided 108 courses for a total of 1,537 attendees, and provided 148 courses for 2,198 attendees in 2019.[4] The scale-up aims to reach 600 courses for 10,200 attendees per year from 2023 onwards, with a projected cost of £1 million ($1.3 million).[5] The majority of scale-up funding would be spent hiring additional staff to facilitate the scale-up.

Figure 2: AfH scale-up




Source: Action for Happiness Growth Model 2020-2023 v2.

On average, interventions in high-income countries are less cost-effective than interventions in low- and middle-income countries. This is because high-income countries have more resources to spend

[4] Personal communication with Dr Mark Williamson.

[5] Action for Happiness Growth Model 2020-2023 v2.

on improving lives than low- and middle-income countries, so many of the best opportunities have already been taken. In part as a consequence of this, people in high-income countries also tend to be happier and healthier, so it is more difficult to improve their lives. However, we believe that facilitating AfH's scale-up is an unusually cost-effective donation opportunity for a high-income country intervention. This is because (i) participants make voluntary donations, which are estimated to provide more than 50% of revenue after scale-up and (ii) AfH generates revenue from some of its other activities, such as its educational services and events. As a result, donations are not used to directly pay for EWM courses but rather cover the scale-up costs that would enable AfH to sustainably reach far more people through EWM courses.

## Summary

**What do they do?** Action for Happiness helps people to live a happy and meaningful life, predominantly through its 8-week Exploring What Matters course, which is run by volunteers in their local communities. Action for Happiness is seeking funding to scale-up to reach 5 times more people through the EWM course and to run sustainably at this larger scale.

**Is there evidence the intervention works?** The main evidence for the efficacy of the Exploring What Matters course comes from a recent randomised controlled trial (RCT) of the programme. We also considered data routinely collected by AfH to measure the effect of the course, as well as less direct evidence, in the form of another RCT of a similar course designed to improve subjective well-being.

**Is the intervention cost-effective?** We estimated the cost-effectiveness of Action for Happiness in terms of reduction in depression and gains in happiness and life satisfaction. Reductions in depression are given in terms of DALY-equivalents averted and years of severe major depressive disorder prevented. DALYs measure the burden of disease by accounting for the premature death (mortality) that it causes and for the years lived with illness (morbidity) it causes: a DALY burden can stem from

premature death or from short-term or long-term ill health. The *disability weights* of different diseases range from 0 to 1 (no disability to 100% disabled). One DALY can be thought of as one lost year of healthy life.

Happiness and life satisfaction points are measured on a 0-10 scale. One happiness point year gain is one year of life with an additional happiness point on the 0-10 scale. Life satisfaction point year gains can be understood similarly.

We estimated cost-effectiveness as follows:

| Metric | Conservative estimate | Best guess estimate | Optimistic estimate |
|---|---|---|---|
| Cost per DALY-equivalent averted (USD) | 30,778 | 2,770 | 177 |
| Cost per equivalent of a year of severe major depressive disorder prevented (USD) | 20,252 | 1,823 | 116 |
| Cost per happiness point year gain (USD) | 2,325 | 197 | 10 |
| Cost per life satisfaction point year gain (USD) | 1,897 | 188 | 10 |

**What are the wider benefits?** The Exploring What Matters course also improves other subjective well-being measures, such as compassion, worthwhileness and anxiety, and increases in self-reported measures of social trust and pro-social behaviour. By running these courses and other related activities, Action for Happiness is building a movement for happiness and prosociality, the benefits of which could be large but are not taken into account in our cost-effectiveness model.

**Is it a strong organisation?** Action for Happiness has a good track record and takes a keen interest in measuring its effects on participants through a recent RCT and ongoing measurements of its effects on course participants. The organisation has been transparent in its communication with us.

**Is there room for funding?** Action for Happiness is seeking £1 million ($1.3 million) over the next three years to facilitate its scale-up.

# Acknowledgements

For discussion, advice and expertise, we are grateful to:

- Professor Ed Diener, University of Utah and University of Virginia
- Dr Michael Plant, Happier Lives Institute and Wellbeing Research Centre, Oxford
- Josh Rosenberg, GiveWell

# Table of Contents

# 1 Charity Recommendation: Action for Happiness

## 1.1 What do they do?

Action for Happiness (AfH) is a UK-based charity that brings people together in small, face-to-face groups to explore what really matters for a happy and meaningful life. AfH is trying to build a community of people transforming their own lives to be happier and to help those around them. AfH provides 8-week courses, called Exploring What Matters (EWM), run by volunteers in their local community. The course aims to help people to tune in to what really matters for a happy and meaningful life, connect with others in meaningful face-to-face conversations and to take action (beyond the course) to boost happiness for participants and for others. Most courses to date have taken place in the UK, but courses have taken place in 20 countries around the world, including the US, Australia, Germany and Italy.

The EWM course brings groups of 8-24 people together for 8 sessions which each last 2-2.5 hours. Each course is led by two unpaid volunteer co-leaders. Potential leaders are screened for appropriate motivation and skills, as well as the ability to bring enough people together in the community. Leaders are intended to act as facilitators of peer-to-peer group discussions, not as expert teachers, which allows leaders to come from a wide range of backgrounds without the need for prior knowledge or extensive training. AfH provides leaders with structured resources and support to run the course.

Participants are encouraged to make donations to AfH to help cover the costs of venue hire, refreshments and resources. This is done on a “donation by choice basis” so that the course is affordable to everyone.

AfH is planning a five-fold scale-up over the next three years. In 2018, AfH provided 108 courses for a total of 1,537 attendees, and has provided 148 courses for 2,198 attendees in 2019.[6] The scale-up aims to reach 600 courses for 10,200 attendees per year from 2023 onwards, with a projected cost of £1 million ($1.3 million).[7] The majority of scale-up funding would be spent hiring additional staff to facilitate the scale-up.

By 2023, AfH aims to be financially self-sufficient at the new, larger scale. AfH plans to achieve self-sufficiency by:

- Increasing course revenue
- Decreasing course costs
- Continuing to earn revenue from other activities, such as events and educational services
- Continuing to receive other donations and support, though on a lower scale than previously

AfH aims to increase course revenue by increasing the revenue per attendee at their courses (while maintaining the voluntary donation model) by encouraging those who can afford to pay to give more generously, as well as by increasing the number of attendees per course and the total number of courses run. AfH aims to decrease the course costs primarily by using more free and low cost venues. Currently, venues require video projectors to show video-based content but AfH is developing

[6] Personal communication with Dr Mark Williamson.

[7] Action for Happiness Growth Model 2020-2023 v2.

improved audio-based content to replace the video-based content. This would open a wider range of possible venues as video projectors would no longer be required.

Besides providing the EWM course, AfH offers published books for adults and children, a downloadable guidebook, posters and postcards centred on the Ten Keys to Happier Living framework, as well as monthly action calendars with daily action ideas and supporting smartphone app providing daily behavioural nudges. It has over 170,000 signed up members and an online community of 1.2m+ followers on social media.

## 1.2 Is there evidence the intervention works?

In evaluating the efficacy of the intervention, we focused on measures of depression, happiness and life satisfaction. We considered depression to allow for comparisons with other charities, such as StrongMinds, a charity we recommend in our Mental Health and Women's Empowerment reports. We considered happiness and life satisfaction scores because these are what the EWM course is intended to promote primarily, and because improvements in happiness and life satisfaction are not completely captured by improvements in depression. Measures of depression focus on negative emotions and mostly neglect positive emotions, so do not measure all the well-being benefits expected from the EWM course. The randomised controlled trial (RCT) of the EWM course also shows evidence of improvements in various other subjective well-being measures and self-reported behaviours, such as compassion and social trust, but we did not evaluate these due to time constraints.

We carried a form of statistical analysis known as Bayesian analysis to estimate the effect of this programme. We believe that Bayesian analysis is the best way to form judgements under uncertainty.[8] This involves specifying a prior probability distribution (or 'prior' for short) for each of depression, happiness and life satisfaction, which represents our beliefs about how effective the intervention is in improving these measures, before taking into account the direct evidence for the intervention. This is based on the effectiveness of similar interventions (external evidence), and the fact that any 'average' intervention is unlikely to have a large effect. We then took direct evidence into account which, combined with our prior probability distributions, results in a posterior probability distribution (or 'posterior' for short) for the effect size of the course for each of depression, happiness and life satisfaction.

Further details of our Bayesian analysis can be found in the Appendix. This is our first time carrying out a Bayesian analysis, so this process was experimental and can no doubt be improved going forwards. We think that this methodology is an improvement on our previous methodology because it allows us to incorporate more information, including the strength of evidence, into our cost-effectiveness estimate. We aim to continue to improve our methodology as we conduct further research and evaluations.

### External evidence

We based our prior probability distributions, in part, on external evidence of the effectiveness of similar interventions. The highest quality relevant external evidence we found is an RCT of a similar

[8] For an introduction to the idea of Bayesian analysis, see Aaron Kramer, "Introduction to Bayesian Inference," accessed December 3, 2019, https://blogs.oracle.com/datascience/introduction-to-bayesian-inference.

programme, called ENHANCE.[9] This programme lasts for 12-weeks and aims to teach participants about the principles of happiness, engage in activities that apply these principles and develop habits that integrate these principles into daily life.

The study measured the effect of the ENHANCE programme on 155 people, of which 133 completed the posttest assessment and 127 completed follow-up 6 months after the start date. The study found standardised effects of -0.193 for depression, 0.193 for life satisfaction, 0.205 for positive affect and -0.088 for negative affect at the end of treatment.[10] Depression was measured with the PHQ-9 (discussed below),[11] life satisfaction was measured by the 5-item Satisfaction with Life Scale,[12] positive and negative affect were measured by the 12-item Scale of Positive and Negative Experience.[13]

Similar programmes designed to increase happiness and life satisfaction, focusing on "keeping busy, spending more time socializing, developing positive thinking, and working on a healthy personality"

[9] S. J. Heintzelman et al., "ENHANCE: Evidence for the Efficacy of a Comprehensive Intervention Program to Promote Subjective Well-Being.," Journal of Experimental Psychology. Applied, October 2019, https://doi.org/10.1037/xap0000254.

[10] This means, for instance, that the depression of participants in the treatment group decreased by 0.193 standard deviations compared to the control group.

[11] Kurt Kroenke, Robert L. Spitzer, and Janet B. W. Williams, "The PHQ-9," Journal of General Internal Medicine 16, no. 9 (2001): 606–13, https://doi.org/10.1046/j.1525-1497.2001.016009606.x.

[12] Ed Diener et al., "The Satisfaction With Life Scale," Journal of Personality Assessment 49, no. 1 (February 1, 1985): 71–75, https://doi.org/10.1207/s15327752jpa4901_13.

[13] Ed Diener et al., "New Well-Being Measures: Short Scales to Assess Flourishing and Positive and Negative Feelings," Social Indicators Research 97, no. 2 (June 1, 2010): 143–56, https://doi.org/10.1007/s11205-009-9493-y.

have been shown to significantly increase happiness compared to controls.[14,15] There is also evidence that interpersonal therapy effectively reduces depression, with reported effect sizes of -0.63 (Cohen's d)[16] and -0.26 (Hedge's g).[17]

There is preliminary evidence that encouraging pro-social behaviour is tractable. Group sessions that encourage pro-social behaviour in school children have been successful in increasing pro-social behaviour in the short-term, as measured by self-reported answers, by a game designed to measure

[14] Michael W. Fordyce, "Development of a Program to Increase Personal Happiness," Journal of Counseling Psychology 24, no. 6 (1977): 511–21, https://doi.org/10.1037/0022-0167.24.6.511.

[15] Michael W. Fordyce, "A Program to Increase Happiness: Further Studies," Journal of Counseling Psychology 30, no. 4 (1983): 483–98, https://doi.org/10.1037/0022-0167.30.4.483.

[16] Pim Cuijpers et al., "Interpersonal Psychotherapy for Depression: A Meta-Analysis," The American Journal of Psychiatry 168, no. 6 (June 2011): 581–92, https://doi.org/10.1176/appi.ajp.2010.10101411.

[17] Laura E. Sockol, "A Systematic Review and Meta-Analysis of Interpersonal Psychotherapy for Perinatal Women," Journal of Affective Disorders 232 (May 2018): 316–28, https://doi.org/10.1016/j.jad.2018.01.018.

student's capacity for perspective-taking and by questionnaires answered by teachers and parents or carers.[18,19,20]

However, social psychology studies are often misleading. About 40% of such studies fail to replicate and those that do often have effect sizes reduced by approximately 50%.[21] This is accounted for in our prior probability distributions and discounts we made to our posterior probability distributions. This concern led us to be somewhat sceptical of the external evidence considered when choosing priors and of the direct evidence used to update to posteriors.

Overall, the external evidence made us somewhat optimistic that interventions like EWM would have a positive effect but we did not expect the effect size to be extremely large. This is reflected in our prior probability distributions.

Our priors for the mean effect on future participants are as follows:

- Depression: we chose a normal distribution with mean -0.092 and standard deviation 0.09

[18] Helen Street et al., "The Game Factory: Using Cooperative Games to Promote Pro-Social Behaviour among Children," Australian Journal of Educational & Developmental Psychology 4 (2004): 97–109.

[19] I. Etxebarria et al., "Design and Evaluation of a Programme to Promote Prosocial-Altruistic Behaviour in the School," Journal of Moral Education 23, no. 4 (January 1, 1994): 409–25, https://doi.org/10.1080/0305724940230403.

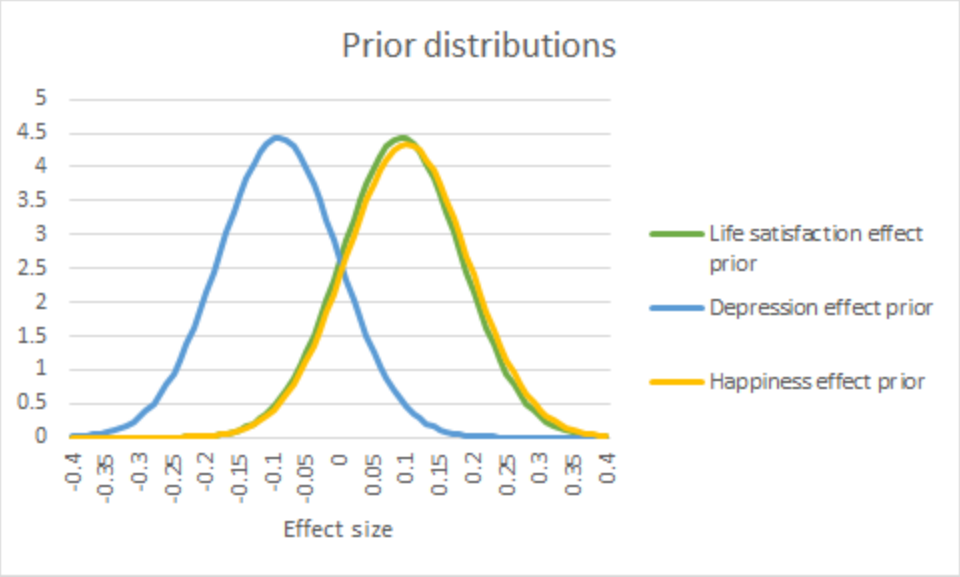
[20] Lisa Flook et al., "Promoting Prosocial Behavior and Self-Regulatory Skills in Preschool Children through a Mindfulness-Based Kindness Curriculum," Developmental Psychology 51, no. 1 (2015): 44–51, https://doi.org/10.1037/a0038256.

[21] Colin F. Camerer et al., "Evaluating the Replicability of Social Science Experiments in Nature and Science between 2010 and 2015," Nature Human Behaviour 2, no. 9 (September 2018): 637–44, https://doi.org/10.1038/s41562-018-0399-z.

- Happiness: we chose a normal distribution with mean 0.1 and standard deviation 0.092
- Life satisfaction: we chose a normal distribution with mean 0.092 and standard deviation 0.09

Units are standardised, meaning that we expected a drop in depression of 0.092 standard deviations, an increase in happiness of 0.1 standard deviations and an increase in life satisfaction of 0.092 standard deviations, as a result of participating in the course. For more details on how we chose our priors, please see the Appendix on our Bayesian analysis.

Figure 3: Prior distributions




## Direct evidence

The direct evidence for the effectiveness of the EWM course is an RCT of the course.[22] This study measured the effect of EWM on 146 people. The study used waitlist randomisation: participants were

[22] Krekel, De Neve, Fancourt, and Layard, 'A local community course that raises mental wellbeing and pro-sociality'.

randomly allocated to either the treatment or control group. The control group was put on a waitlist and was treated after the treatment group. Baseline data for both groups was collected together without the participants knowing which group they were in. We also considered data routinely collected by AfH to measure the effect of the course through surveys of course participants, but we did not take this into account in the Bayesian analysis. We see this as a positive sign of AfH's commitment to monitoring its effectiveness but this evidence is naturally not of the same quality and rigour as the RCT.

### Results

The study reported some very promising results. Standardised effect sizes at the end of the course were -0.497 for depression, 0.596 for happiness and 0.633 for life satisfaction. Follow-up 8 weeks post-treatment looked promising but was speculative because there was no control (since the original control group had been treated by this time). Depression and life satisfaction did not change much between the end of the course and follow-up, but it is possible that they improved slightly.

Depression was measured using the 9-item Patient-Health Questionnaire (PHQ-9).[23] The PHQ-9 is a valid and reliable measure of depression in many settings.[24] It consists of 9 questions, scored 0-3, resulting in a score between 0 and 27. The overall score can be interpreted as follows:

[23] Kroenke, Spitzer, and Williams, "The PHQ-9."

[24] Karen L. Smarr and Autumn L. Keefer, "Measures of Depression and Depressive Symptoms: Beck Depression Inventory-II (BDI-II), Center for Epidemiologic Studies Depression Scale (CES-D), Geriatric Depression Scale (GDS), Hospital Anxiety and Depression Scale (HADS), and Patient Health Questionnaire-9 (PHQ-9)," Arthritis Care & Research 63, no. S11 (2011): S454–66, https://doi.org/10.1002/acr.20556.

| Overall PHQ-9 score | Interpretation |
|---|---|
| 0 | No depression |
| 1-4 | Minimal depression |
| 5-9 | Mild depression |
| 10-14 | Medium depression |
| 15-27 | Strong depression |

Source: Krekel, De Neve, Fancourt, and Layard, 'A local community course that raises mental wellbeing and pro-sociality'.

Happiness and life satisfaction were measured using the following single-item scales:

> "Overall, how satisfied are you with your life nowadays?": (0) "Not at all" to (10) "Completely"
>
> "Overall, how happy did you feel yesterday?": (0) "Not at all" to (10) "Completely"

These are the happiness and life satisfaction measures used by the UK's Office for National Statistics.[25] There is evidence that single-item happiness and life satisfaction scales give very similar results to multiple-item measures.[26,27,28]

### Limitations

However, there are some limitations to the study, which led us to discount the estimated effect sizes. We emphasise that our aim is to predict the effect sizes of the course as accurately as possible. Some of the following limitations are not direct criticisms of the study quality, but inevitable limitations of social psychology studies.

- Study not yet peer reviewed
    - The study has not yet been peer reviewed nor published in an academic journal.
- Potential bias of study authors

[25] "Surveys Using Our Four Personal Well-Being Questions - Office for National Statistics," accessed November 6, 2019, https://www.ons.gov.uk/peoplepopulationandcommunity/wellbeing/methodologies/surveysusingthe4officefornationalstatisticspersonalwellbeingquestions.

[26] Christina P. Moldovan, "AM Happy Scale: Reliability and Validity of a Single-Item Measure of Happiness," 2017.

[27] Ahmed Abdel-Khalek, "Measuring Happiness with a Single-Item Scale," Social Behavior and Personality: An International Journal 34 (January 1, 2006): 139–50, https://doi.org/10.2224/sbp.2006.34.2.139.

[28] Felix Cheung and Richard E. Lucas, "Assessing the Validity of Single-Item Life Satisfaction Measures: Results from Three Large Samples," Quality of Life Research : An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation 23, no. 10 (December 2014): 2809–18, https://doi.org/10.1007/s11136-014-0726-4.

    - There is potential for bias in the study because one of the co-authors of the study is Lord Richard Layard, who is a co-founder of AfH. We do not see strong evidence of bias, however, avoiding bias altogether is very challenging. The tone of the writing in the study is overly positive at times, with some questionable interpretations of results (e.g. effects on biological markers), which is some evidence of bias in the study.

- Course was run to a higher standard in the study than normal
    - We think it likely that the course was run to a higher standard in the study than it will be in the future. For instance, there were incentives to choose experienced course leaders who have performed well in previous courses.
- Course is different to the average course because of lower attrition
    - Attrition rates in the study were unusually low. 9% of participants dropped out and on average, participants attended 7 sessions out of 8. However, typically, the attrition rate for EWM courses is about 20-30%, with participants attending about 6 sessions out of 8 on average.[29] This makes it difficult to know what the average effect on participants who complete the course will be outside of the study because attendance will likely tend to be lower, which could plausibly change the dynamics of group interactions. This mainly increases our uncertainty about the results rather than evidencing bias.
- Close to zero effect on all biological markers

[29] Personal communication with Dr Mark Williamson and Lord Richard Layard.

- The study also measured biological markers of well-being, such as levels of cortisol and cytokines, because, as the study notes, these markers are responsive to short-term and long-term psychosocial interventions.[30] However, the study found close to zero effect on all biological markers. We took the absence of an effect here as evidence against the effectiveness of EWM.

- Participants knew that they were part of a study
  - We expect the reported effect sizes to be overestimates because the participants knew they were part of the study, which probably led some of them to overreport the benefits of the course.

- Risk of replication failure
  - As noted above, social psychology studies have notoriously low replication rates and even those that do replicate tend to have lower effect sizes.[31] It is plausible that many studies fail to replicate for similar reasons to those already given above, but we applied a further discount to account for considerations not already taken into account. AfH's surveys suggest that the effect size for life satisfaction is similar to that reported in the RCT.[32] However, these surveys naturally have no control group to compare participants

[30] Daisy Fancourt et al., "Group Drumming Modulates Cytokine Response in Mental Health Services Users: A Preliminary Study," Psychotherapy and Psychosomatics 85, no. 1 (November 27, 2015): 53–55, https://doi.org/10.1159/000431257.

[31] Camerer et al., "Evaluating the Replicability of Social Science Experiments in Nature and Science between 2010 and 2015."

[32] Personal communication with Dr Mark Williamson.

with and do not feature the same level of rigorous analysis as the RCT. As a result, we made a fairly substantial discount due to this concern, but we discounted slightly less than we would have in the absence of AfH's surveys.

Each of these limitations led us to discount the reported effect sizes in a way that brought our final posterior estimate of effect size closer to our prior estimate of effect size. The total discount was 33%, with details on the "Discounting update" sheet of our cost-effectiveness analysis. See the Appendix on Bayesian analysis for the details and justification of our method of discounting.

### Overall judgement

Finally, we note that since we have just one RCT, with a medium-sized sample, our Bayesian update is relatively small. Overall, based on our Bayesian analysis, we expect the EWM course to have effect sizes directly after the course of -0.176 for depression, 0.174 for happiness and 0.186 for life satisfaction. This corresponds to a point difference of -0.774 on the PHQ-9 (depression), 0.324 on the above happiness scale and 0.308 on the above life satisfaction scale. These effect sizes are 35%, 29% and 29% as large as the reported effect sizes for depression, happiness and life satisfaction, respectively.[33] We do not have longer-term data on the effects of the course, so this analysis only estimates short-term effects.

For comparison, being partnered as opposed to being single has been reported to have an effect on happiness of about 0.6 for men and 0.45 for women, corresponding to about 0.5 points on an 11-point

[33] Note that given the 60% replication rate in social psychology and the fact that the studies that do replicate tend to report effect sizes about half as large as originally reported, one might expect effect sizes to be about 30% as large as the originally reported effect sizes on average.

(0-10) scale.[34] We previously estimated StrongMinds's intervention to reduce depression, as measured by the PHQ-9, by 3.13 points for the average participant. However, note that these results have been subjected to neither a Bayesian analysis nor critical discounting due to risks of bias, unlike our estimate of the effect of AfH's courses, so a straightforward comparison is not possible. As a result, the actual effects of being partnered and StrongMinds's courses are probably lower than these reported effects. Note that the reported effect of AfH's courses is actually very similar to the reported effect of being partnered.

### 1.3 Is the intervention cost-effective?

In estimating cost-effectiveness, it is important to construct models that are as accurate as possible, but that are also constructed in such a way that they allow for fair comparisons with other funding opportunities. We work to continually improve our methodology, but this can make it harder to make fair comparisons between different interventions, as cost-effectiveness models across interventions will rely on different assumptions and take different factors into account. In this case, we aimed to estimate the cost-effectiveness of funding AfH's scale-up as accurately as possible but to also allow for a fair comparison with StrongMinds, one of our other recommended charities. In addition to our usual conservative, best guess and optimistic estimates in our cost-effectiveness model, we made estimates that are comparable to our best guess cost-effectiveness estimate of StrongMinds.

We estimated the number of extra participants on AfH's courses, based on data provided by AfH and some subjective and uncertain estimates. These subjective estimates are difficult to get right but also

[34] Shuai Chen and Jan C. van Ours, "Subjective Well-Being and Partnership Dynamics: Are Same-Sex Relationships Different?," Demography 55, no. 6 (December 2018): 2299–2320, https://doi.org/10.1007/s13524-018-0725-0.

important to the resulting cost-effectiveness estimate. Accordingly, our cost-effectiveness estimates have very wide bounds, with the conservative and optimistic estimates differing by more than a factor of 100. Additionally, given our time constraints, we were unable to account for all factors that could bear on impact. For instance, we have not accounted for fungeing with other donors, and how the number of participants eventually reached by AfH might depend on when the scale-up happens.

The nature of this funding opportunity makes it higher-risk than many of our other recommended charities. For instance, donors are not directly funding an intervention, but rather enabling AfH to scale-up over three years and to then be self-sufficient at the new, larger scale. Moreover, there is significant uncertainty about the future. Getting ambitious future plans right is challenging, so we discounted AfH's impact to account for uncertainty about their plans being successful. Additionally, most of the impact of the scale-up is realised years ahead but the future is inherently uncertain, so lots of things could happen in the meantime that might disrupt the future success. We are very uncertain about the extent to which we should discount AfH's impact, resulting in a wide range of cost-effectiveness estimates. Furthermore, as with most of our recommendations, we do not have any evidence on longer-term effects, which also increases the uncertainty of our cost-effectiveness models.

We found that facilitating AfH's scale-up seems to be roughly as cost-effective as donating to StrongMinds in expectation, in terms of depression alone. Using similar methodology to our cost-effectiveness estimate of StrongMinds, we estimated that AfH costs $373 per DALY-equivalent averted due to depression, compared to $377 for StrongMinds. DALYs measure the burden of disease by accounting for the premature death (mortality) that it causes and for the years lived with illness (morbidity) it causes: a DALY burden can stem from premature death or from short-term or long-term

ill health. The disability weights of different diseases range from 0 to 1 (no disability to 100% disabled). One DALY can be thought of as one lost year of healthy life.

We estimated cost-effectiveness in terms of reduction in depression, to allow for direct comparisons with other charities that improve mental health, such as StrongMinds, but we think that many of the benefits of AfH come from improvements in happiness and life satisfaction, which are not completely captured by the depression metric.

Our rough model suggests that facilitating the AfH scale-up would have the following effects:

| Metric | Conservative estimate | Best guess estimate | Optimistic estimate | Estimate comparable to StrongMinds CEA best guess |
|---|---|---|---|---|
| Cost per DALY-equivalent averted (USD) | 30,778 | 2,770 | 177 | 373 |
| Cost per equivalent of a year of severe major depressive disorder prevented (USD) | 20,252 | 1,823 | 116 | 245 |
| Cost per life happiness point year gain (USD) | 2,325 | 197 | 10 | 19 |
| Cost per life satisfaction point year gain (USD) | 1,897 | 188 | 10 | 18 |

On average, interventions in high-income countries are much less cost-effective than interventions in low- and middle-income countries. This is because high-income countries have more resources to spend on improving lives than low- and middle-income countries, so many of the best opportunities have already been taken in such countries. Relatedly, people in high-income countries tend to be happier and healthier, so it is more difficult to improve their lives. However, we believe that facilitating AfH's scale-up is an unusually cost-effective donation opportunity for a high-income country intervention. This is because (i) participants make voluntary donations, which are estimated to provide more than 50% of revenue after scale-up and (ii) AfH generates revenue from some of its other activities, such as its educational services and events. As a result, donations are not used to directly pay for EWM courses but rather cover the scale-up costs that would enable AfH to sustainably reach far more people through EWM courses. We ignored costs to participants in taking part in the course because we expect them to be very small compared to the benefits, since the participants' donations would probably otherwise be spent on regular consumption, which would yield comparatively small benefits.

If successful, AfH's scale-up could be very high-impact. According to our optimistic estimate, AfH could avert a year of depression for $116. Additionally, we expect AfH to have wider benefits that are not accounted for in our cost-effectiveness models but which could be large.

## 1.4 What are the wider benefits?

The cost-effectiveness analysis ignores some of the benefits of the EWM course. These include other subjective well-being measures, such as compassion, worthwhileness and anxiety, and increases in self-reported measures of social trust and pro-social behaviour. Participants also feel more knowledgeable about what matters to them personally in life and about what contributes to a happy and meaningful life. They feel more able to take action to improve their own well-being, and to some

extent, the well-being of others too. Other improvements include self-reported increases in meditation and mindfulness practice and participants treating themselves in a kinder way.

In addition to running the EWM courses, AfH is building a movement for happiness and prosociality, and funding AfH would facilitate this as well. AfH will likely be able to carry out movement building on a much larger scale with the scale-up funding, so the wider benefits of these activities will probably be increased by the scale-up but this increase is not accounted for in the cost-effectiveness models. In this way, our cost-effectiveness model is conservative.

### 1.5 Is it a strong organisation?

Action for Happiness has a strong track record. Co-founder Dr Mark Williamson has been Director of AfH since it began in 2011, and has grown the movement to reach hundreds of thousands of members. He has considerable corporate, non-profit, and start-up experience, having worked in the corporate world for over 10 years, mainly as a management consultant, before working as director of the Innovation unit at the Carbon Trust, and non-executive Director of clean technology start-up Solar Press. He has a PhD in Electronics and Communications. Co-founder Lord Richard Layard is a Professor of Economics at the LSE. He is also Founder-Director of the Centre for Economic Performance and the head of its Well-Being programme. He has been very active in happiness and well-being research, for instance, co-editing the World Happiness Report, an annual publication of the United Nations Sustainable Development Solutions Network that contains articles about happiness and rankings of national happiness based on self-reported happiness measures.

AfH has been run sustainably and has steadily grown since it was founded in 2011. The organisation has been run efficiently: for instance, it has made extensive use of automation for administrative tasks and promotion. It is also a very lean organisation. Dr Williamson has led AfH since its launch, gradually growing the team from one full time equivalent staff member in 2011 to 2.5 in 2015, when the first

course was launched, and to 4 full time equivalent staff members in 2019. The number of attendees at AfH's courses has grown steadily from 943 in 2016 to 1,454 in 2018, with 1,435 already in the first two thirds of 2019.[35] AfH has a stable financial base, with good cash reserves and low liabilities.

AfH appears to be very transparent. Dr Williamson and Lord Layard have been very open in their communication with us. For instance, they openly shared that they are in touch with other potential funders and voluntarily brought up some of the limitations of the study of the EWM course.

Finally, AfH has a strong commitment to monitoring and evaluation. It has carried out an RCT of the EWM course and continually tracks well-being outcomes of participants, with about 2,500 pairs of surveys from participants filled in before and after taking the course. AfH also plans to evaluate the longer term impact of the course over multiple years.

### 1.6 Is there room for funding?

AfH estimates that it needs £1 million (about $1.3 million USD) over the next three years to facilitate the planned scale-up. The break-down of funding over the next three years is as follows:

- 2020: £300,000 ($390,000)
- 2021: £400,000 ($520,000)
- 2022: £300,000 ($390,000)

About 70% of the scale-up funding is budgeted to be spent on team growth. AfH plans to hire a Chief Operating Officer, a Director of Communications, a Director of Community Building, and an Office

[35] Personal communication with Dr Mark Williamson.

Manager. The rest of the scale-up funding will be spent primarily on additional promotion of courses (e.g. via Facebook ads), additional community development costs during scale-up period (e.g. training and events), increasing in office space from 2020 onwards to support team growth, and investing in a new website and user experience to support the scale-up. AfH aims to be financially self-sufficient at their greater scale from 2023 onwards.

AfH is also seeking funding from other donors, so it is possible that some of this room for more funding will be filled by other donors.

### 1.7 What are the main uncertainties?

Evaluating the impact of facilitating the scale-up is challenging, and we have a number of important uncertainties:

- The extent to which we can expect AfH's plans to be successful
- How likely it is that unforeseen future events will reduce (or increase) impact
- How likely it is that another donor would otherwise fund the scale-up anyway
- How large the short-term benefits of the course are
- How well the benefits of the course will be retained

These uncertainties are accounted for very roughly in our cost-effectiveness model, and resulted in a wide range of cost-effectiveness estimates.

In addition, there are some uncertainties that we have not accounted for in our cost-effectiveness model, but which could make a large difference to the cost-effectiveness. We have not accounted for

fungeing with other donors and have not accounted for the effect that an early scale-up could have on AfH's overall trajectory.

# 2 Appendix

## 2.1 Bayesian analysis

We used Bayesian inference to estimate the effect sizes of the EWM course in terms of depression, happiness and life satisfaction. This enabled us to take all relevant information into account, without relying too heavily on the RCT on the EWM course.

### A brief overview of Bayesian inference

In Bayesian inference, we have a model, which depends on parameters, at least some of which we are uncertain about. In this case, we aim to predict the effect of the course on future participants. For each of depression, happiness, and life satisfaction, the effect on a future participant is given by the effect size plus random noise, which we treat as a random variable, with unknown mean. The effect on a future participant $x$ is a random variable that depends on parameter $\mu$. We write $x \sim p(x|\mu,\sigma)$ to mean that $x$ has a probability density given by $p(x|\mu,\sigma)$, which depends on given values of $\mu$ and $\sigma$.

The aim of Bayesian inference is to estimate the parameters $\mu$ and $\sigma$, though in this case, we assumed that $\sigma$ was known. We begin with a prior for $\mu$, which is a probability distribution $p(\mu|\alpha)$, which depends on the hyperparameter $\alpha$. Given new data $D$, we revise our prior $p(\mu|\alpha)$, in light of the new evidence via Bayes' rule: the posterior probability distribution for  given $D$ is given by conditioning on $D$: $p(\mu|D,\alpha)$. We calculate this using Bayes' Theorem:

$$p(\mu|D,\alpha) = \frac{p(D|\mu)p(\mu|\alpha)}{p(D|\alpha)} \propto p(D|\mu)p(\mu|\alpha)$$

The term $p(D|\mu)$ is called the likelihood, which gives the probability of observing the data $D$, given the value of $\mu$. Note that $D$ is fixed and that $\mu$ is the variable here, so the likelihood need not be a probability distribution.

The posterior probability distribution for μ can be used to predict the effect on a future participant $\tilde{x}$:

$$p(\tilde{x}|D,\alpha) = \int p(\tilde{x}|\mu)p(\mu|D,\alpha)\ d\mu$$

This distribution is known as the posterior predictive distribution.

### Selecting priors

We choose a prior $p(\mu|\alpha)$ for the mean effect for each of depression, happiness and life satisfaction, informed by external evidence. We assumed that $\mu$ is normally distributed with mean $\mu_0$ and standard deviation $\sigma_0$, i.e. $\mu \sim N(\mu_0, \sigma_0^2)$, $\alpha = (\mu_0, \sigma_0)$. While we think that a normal distribution is a reasonable approximation, this assumption is mainly pragmatically motivated. Ideally, we would have used a positively skewed distribution, with most probability density close to zero and a positive tail for happiness and life satisfaction (and the reverse for depression, since the effect is negative) as a small effect seems most likely, with some chance of a large effect. We decided to opt for a normal distribution for ease and simplicity. We doubt that this will have significantly affected our analysis.

Our priors for the mean (standardised) effect on future participants and are as follows (the standardised means from the ENHANCE study are included for comparisons):

| **Metric** | $\mu_0$ | $\sigma_0$ | **ENHANCE mean** |
|---|---|---|---|
| Depression | -0.092 | 0.090 | -0.193 |
| Happiness | 0.100 | 0.092 | 0.205 |
| Life satisfaction | 0.092 | 0.090 | 0.193 |

Choosing priors is often subjective and difficult. Our choices are based primarily on intuitive judgements based on the external evidence we found. We used two different methods to choose the hyperparameters of our priors and for each metric and each hyperparameter, we took the median of the resulting four values from these two methods and two authors. Each author tried to carry out this process independently to avoid biasing each other. However, since we were continually developing our methodology, the process was iterative, so complete independence was not feasible.

### Method 1

We used the reported effect sizes on depression, positive affect (which is similar, though not identical, to happiness) and life satisfaction of the ENHANCE study as a benchmark. This study reported unstandardised effects and standard deviations, so we standardised the effects by dividing by the standard deviations. We considered the probability that the future effect size of the EWM course is various proportions of the ENHANCE effect size (as far as possible, we ignored all information from the RCT on the EWM course so that we could update on this as new evidence later in the analysis).

For simplicity, we assumed that the only possible outcomes were that the effect sizes could be -⅓, 0, ⅓, ⅔, 1, or $^4/_3$ times the corresponding ENHANCE effect sizes. Both authors gave subjective estimates of the probability distribution over these events (i.e. we each gave subjective estimates of the probability that the effect sizes of the EWM course are -⅓, 0, ⅓, ⅔, 1, $^4/_3$ times the corresponding ENHANCE effect sizes). We then calculated the means and standard deviations of these distributions.

### Method 2

We chose hyperparameters by subjectively estimating the 0.25, 0.5 and 0.75 quantiles of the distribution of effect sizes (without imposing normality). This approach offers a nice way to pick standard deviations because quantifying beliefs about median values is easier than quantifying beliefs about spread, but the latter can be inferred from the former. We first estimated the 0.5 quantile

(median), denoted by '$q_1$' in our spreadsheet. We then selected $q_2$ and $q_3$, the 0.25 and 0.75 quantiles, respectively.

We took $q_1$, the 0.5 quantile estimate, as the mean. Since $q_2$ and $q_3$ need not be distributed symmetrically about $q_1$, but we were ultimately aiming to use a normal distribution (which is symmetric), we selected values $q_2$' and $q_3$' which are distributed symmetrically about the mean $q_1$. These were selected by calculating the average of the difference between $q_1$ and $q_2$ and the difference between $q_1$ and $q_3$, and placing $q_2$' and $q_3$' symmetrically about $q_1$, at this average difference. That is, $q_2' = q_1 - \frac{(q_1 - q_2) + (q_3 - q_1)}{2} = q_1 - \frac{q_3 - q_2}{2}$ and $q_3' = q_1 + \frac{q_3 - q_2}{2}$.

This provides enough information to determine the standard deviation hyperparameter. We have $F(q_2', q_1, \sigma_0) = 0.25$, where $F$ is the cumulative distribution function of a normal distribution with mean $q_1$ and standard deviation $\sigma_0$, evaluated at $q_2$'. This equation can be solved for $\sigma_0$. Both authors provided values for $q_1$, $q_2$ and $q_3$, which determined values for $q_2$' and $q_3$', which determined values for $\sigma_0$. We carried out this process for each of depression, happiness and life satisfaction.

### Final prior

Given values for $\mu_0$ and $\sigma_0$ for each of depression, happiness and life satisfaction from both authors via two methods, we selected final values by taking the median over these four values (2 authors x 2 methods) for each hyperparameter.

The standard errors from the regression in the ENHANCE study could have been used to find appropriate $\sigma_0$ values (by dividing the standard errors by the sample standard deviation). We did not think that these values would be appropriate though, because (i) we are generally more uncertain about the results of studies in social psychology than the studies claim we should be but (ii) we think that most effects in social psychology are overstated, and hence closer to zero than claimed. All else

being equal, (i) pushes for a higher value of $\sigma_0$ than the study would suggest but (ii) pushes for a lower value. The $\sigma_0$ values we would have obtained with this method were 0.085, 0.090 and 0.091 for depression, happiness and life satisfaction, respectively. As it happens, this is very slightly lower than our final values for depression and happiness, and very slightly higher than our final value for life satisfaction.

### Inference

The RCT is represented by data $D$. We assumed that the data is normally distributed around the true effect size, with mean $\mu$ and standard deviation $\sigma$. Given the sufficiently large sample of 146 participants and 279 observations, the assumption of normality is justified by the Central Limit Theorem. This determines the likelihood $p(D|\mu)$, which, as a function of $\mu$, is a normal distribution with mean $D$ and standard deviation $\sigma$.[36] The mean $D$ is taken as the standardised coefficient from the regression in the RCT on the EWM course and the standard deviation $\sigma$ is the corresponding standard error. The table below contains the parameters of the likelihood:

[36] Note that in this case, our likelihood is conveniently a probability distribution as a function of μ, due to the symmetry in normal probability density function with respect to μ and D.

| Metric | $\sigma$ | $D$ |
|---|---|---|
| Depression | 0.135 | -0.497 |
| Happiness | 0.173 | 0.596 |
| Life satisfaction | 0.152 | 0.633 |

We then computed the Bayesian update. This can be done analytically for our chosen prior and likelihood, using the theory of conjugate priors.[37] This resulted in the following posterior for $\mu$:

| Metric | Posterior mean |
|---|---|
| Depression | -0.216 |
| Happiness | 0.209 |
| Life satisfaction | 0.232 |

More generally, this analytic approach is not always available because it requires particular prior and likelihood combinations. We also computed the posterior in Python, which can be adapted to

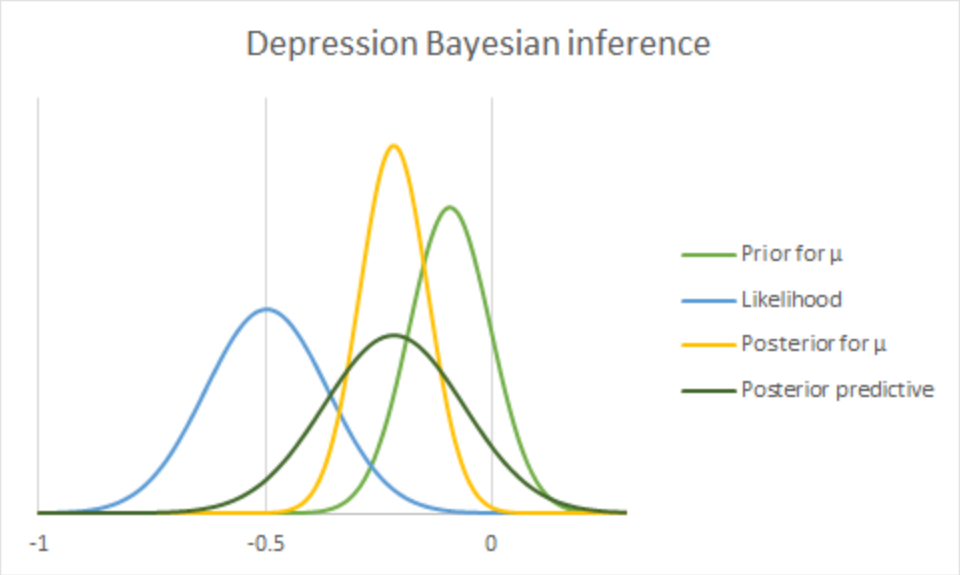
[37] "Conjugate Prior," in Wikipedia, November 23, 2019, https://en.wikipedia.org/w/index.php?title=Conjugate_prior&oldid=927609654.

compute posteriors more generally, with greater flexibility in the choice of prior. In the future, we would like to use more sophisticated priors when appropriate (such as skewed distributions), which will be easier to update computationally than analytically.

The mean (which is also the median and mode) of the posterior for the mean effect on future participants is the same as the mean of the posterior predictive distribution. That is, $E[\mu|D, \mu_0, \sigma_0] = E[\tilde{x}|D, \mu_0, \sigma_0]$.

Note that μ is the mean effect we predict on future participants (i.e. the effect size) so the posterior for μ gives our probability distribution for the effect size. The posterior predictive gives our probability distribution for the effect on a single future participant, which is less certain than the mean effect. The Bayesian inference for depression is pictured below:

Figure 4: Depression Bayesian inference



## Discounting for limitations

Typical Bayesian inference treats evidence as perfect and certainly correct but in practice, reported results rarely reflect reality perfectly. In this particular case, as noted above, the results of the RCT were limited in a few ways. We think that there are possibilities of bias, which means that the true effect sizes of the course on the study participants might have been lower than the reported effect sizes. When evidence is uncertain, one can use Jeffrey's rule to update while accounting for uncertainty about one's evidence.[38,39] Rather than updating via Bayes' rule, we used a modification based on Jeffrey's rule, as described below:

- The posterior $\bar{p}$ is given by $\bar{p}(\mu) = p(\mu|D)(1-\delta) + p(\mu)\delta$, where $\delta$ is a discount factor that can be interpreted roughly as the probability that we should stick with our prior rather than update on the evidence
- This assumes that the only possible outcomes are (i) the study is perfect and (ii) the study provides no new information (in which case we should stick with our prior)
    - This is a simplification, but given our time constraints, we think that it is a reasonable way to discount the study findings
- For each limitation *l*, both authors gave a subjective estimate of the discount $\delta_l$ that should be applied due to this limitation, and we took the median value

[38] Richard Jeffrey, "Bayesianism with a Human Face," n.d., 24.

[39] Persi Diaconis and Sandy L. Zabell, "Updating Subjective Probability," Journal of the American Statistical Association 77, no. 380 (December 1, 1982): 822–30, https://doi.org/10.1080/01621459.1982.10477893.

- We then calculated the total discount $\delta$, treating each limitation as independent from all the others
- We tried to account only for limitations that were not taken into account in choosing our priors, to avoid double counting
- This process resulted in a discount of $\delta=0.33$, i.e. as the output of our Bayesian analysis, we took the weighted average of our Bayesian posterior (assuming the study is perfect) and our prior, in which our prior had a weight of 0.33 and our Bayesian posterior had a weight of 0.67.

This discounting resulted in the following posterior means:

| Metric | Discounted posterior mean |
|---|---|
| Depression | -0.176 |
| Happiness | 0.174 |
| Life satisfaction | 0.186 |

We also considered altering the Bayesian update by increasing the variance of the likelihood. The rationale for this is that some limitations increase uncertainty about the result instead of, or as well as, biasing the result upwards. This can be accounted for by increasing the variance of the likelihood. However, due to time-constraints, we decided not to opt for this route this time.

### Output of Bayesian analysis

We used point changes (on the PHQ-9, Happiness and Life satisfaction scales) as the output of our Bayesian analysis that is used as inputs of our cost-effectiveness model. Since our effect sizes are

standardised (i.e. tell us the effect as a proportion of the standard deviation), we multiplied the effect sizes by the pre-intervention standard deviation, to estimate the point differences for each metric:[40]

| Metric | Point difference |
|---|---|
| Depression (0-27) | -0.774 |
| Happiness (0-10) | 0.324 |
| Life satisfaction (0-10) | 0.308 |

[40] Minimum-maximum points in parentheses.