

Themed Section: Health Economics and Outcomes Research of Mental Health

Estimating the Lifetime Costs and Benefits of the Incredible Years Teacher Classroom Management Intervention Using Data From 30 Months Follow-Up of the Supporting Teachers and childRen in Schools Trial

Sarah Bates, PhD, Yekta Saidi, MPhys, Richard Cookson, PhD, Ieva Skarda, PhD, Tamsin Ford, PhD, Rachel Hayes, PhD, Poushali Ganguli, MSc, Sarah Byford, PhD, Alan Brennan, PhD

ABSTRACT

Objectives: The Incredible Years Teacher® Classroom Management (IY-TCM) intervention is associated with short-term improvements in mental health difficulties in young people. The aim was to estimate the long-term impact and cost-effectiveness of the IY-TCM intervention compared with no intervention.

Methods: An existing health economic model (LifeSim 1.0) was used to translate short-term changes in the Strength and Difficulties Questionnaire (SDQ), based on the Supporting Teachers and childRen in Schools cluster randomized controlled trial of the IY-TCM intervention in schools, into estimated medium- and long-term effects using multiple longitudinal data sets. LifeSim 1.0 was adapted to incorporate teacher-reported SDQ and account for individual heterogeneity. Cost-effectiveness analyses were conducted using the trial-based intervention cost with subgroup analyses on deprivation, conduct scores and parental depression in the simulated baseline population.

Results: Regression analyses show significant predictor variables for intervention effectiveness, including deprivation and baseline SDQ. LifeSim results indicate small gains in long-term outcomes, and cost-effective analyses estimated that the IY-TCM intervention could be cost-effective, but there was a large amount of uncertainty (net monetary benefit = £10, Estimated CI = -£134, £156). Benefits and certainty of cost-effectiveness were greater for some subgroups, such as those with high conduct scores at baseline (net monetary benefit = £206, Estimated CI = £26, £318).

Conclusions: IY-TCM could be cost-effective, but there was a large amount of uncertainty around costs and benefits. Greater benefits for pupils with difficulties at baseline suggest that the intervention may be more cost-effective for schools in more deprived areas with high levels of conduct problems.

Keywords: health economic modeling, mental health, school-based interventions.

VALUE HEALTH. 2025; 28(3):326–335

Highlights

- The Supporting Teachers and childRen in Schools trial of the incredible years teacher classroom intervention indicated short-term improvements in mental health and some evidence of within-trial cost-effectiveness.
- Using a microsimulation model (LifeSim), we estimated modest simulated reductions in depression, unemployment, and mortality. Estimates indicated that the intervention could be cost-effective over a lifetime time horizon, but there was a lot of uncertainty. Estimated benefits and probability of cost-effectiveness were greater for children in deprived areas or with conduct issues.
- The findings provide useful information for healthcare-related decision making; the intervention could be cost-effective, and there is evidence to suggest that it might be particularly effective in schools in deprived areas with a high prevalence of children with conduct disorder.

Introduction

Childhood mental health issues are prevalent and persistent, leading to adverse outcomes¹ and prompting a growing focus on children's mental health and the role of schools.² Conduct disorder affecting 4% to 8% of school age children predicts various adult mental disorders,^{3,4} and children living in deprived circumstances are at higher risk.^{5,6} Whereas parent-targeted interventions may face barriers related to attendance and engagement, school-based approaches can reach all children attending and so have the potential to reduce mental health inequalities.⁷

The Incredible Years Teacher Classroom Management (IY-TCM) aims to prevent and reduce conduct problems and improve socio-emotional skills in children aged 3 to 8.⁸ A meta-analysis indicated small positive effects on children.⁹ The Supporting Teachers and

childRen in Schools (STARS) trial⁷ of the IY-TCM was a 2-arm, pragmatic, parallel group, superiority, cluster randomized controlled trial in 80 primary schools from Reception to Year 4 (2075 children aged 4–9 years) with a 30 month follow-up. The intervention consisted of 6 whole-day sessions, based on theories of how coercive cycles of interaction between adults and children can reinforce disruptive behavior and the importance of modeling and self-efficacy. There was a reduction in the teacher-reported Strength and Difficulties Questionnaire (SDQ) Total Difficulties score at 9 months, but this did not persist at 18 or 30 months. A within-trial cost-effectiveness analysis generated a “cost-per-SDQ total difficulties score point improvement” achieved by the intervention (see Table 5 in Ford et al⁷). Interpreting this metric was difficult because, unlike the quality-adjusted life year (QALY) for which there is National Institute for Health and Care Excellence (NICE) or Department

of Health and Social Care guidance on the financial economic value, there is no guidance on a national policy maker's "willingness to pay" for a 1-point improvement in SDQ total difficulties score. The authors conclude that IY-TCM might be cost-effective over the short-term but highlight considerable uncertainty and the need to conduct long-term health economic modeling.

The aim of the study is to estimate the long-term health impact, economic impact, and cost-effectiveness of the IY-TCM intervention versus control "treatment as usual" to establish whether the intervention is cost-effective and inform decisions regarding wider implementation in schools. The specific objectives were as follows:

1. Reanalyze the STARS trial data set to understand individual and subgroup effects on SDQ subscores.
2. Adapt a microsimulation model (LifeSim version 1.0) model to incorporate the evidence on both teacher- and parent-reported SDQ and to model individual level heterogeneity of effect.
3. Quantify the long-term impact, cost-effectiveness, and uncertainty around these estimates, for a whole-class approach (all children exposed to the intervention).
4. Apply the estimated intervention effect from the STARS trial, to a simulated population in the LifeSim version 1.0 model to quantify the long-term effects and cost-effectiveness for specific subgroups of the simulated baseline population defined by (1) being in each of the 5 Index of Multiple Deprivation (IMD) quintile categories, (2) a parent-reported SDQ conduct subscore at baseline ≥ 4 , (3) a parent having depression at baseline (measured by a Kessler scale score ≥ 13), and (4) pupils who have BOTH parent-reported baseline SDQ conduct subscore of 4+ AND a parent with baseline Kessler scale of 13+.

Methods

We examined the impact of the IY-TCM intervention using the STARS trial⁷; a 2-arm, pragmatic, parallel group, superiority, cluster randomized controlled trial in 80 primary schools from Reception to Year 4 (2075 children aged 4-9 years) with a 30 month follow-up (full results in [Supplementary Materials, Appendix 1](https://doi.org/10.1016/j.jval.2024.05.002) found as <https://doi.org/10.1016/j.jval.2024.05.002>). The direct cost per pupil participating in the intervention, based on the cost of training and staff time costs was estimated at £11.52.⁷ To estimate the long-term costs and benefits of the intervention in comparison with the control group of no intervention we first reanalyzed the individual participant data from the STARS trial to examine the impact of the intervention on SDQ sub scores. We then used the LifeSim microsimulation model to translate the short-term change in SDQ to long-term outcomes and calculate the incremental cost-effectiveness ratio (ICER) and compared it with the willingness-to-pay (WTP) threshold, determined by NICE. Although the intervention was school based, the main outcome of the intervention was change in mental health, and the main cost-savings are within health and social care sectors; therefore, the intervention is still in the remit of NICE. However; we also calculated wellbeing-adjusted life years (WELBYs) using discounted good life years (a good life is equal to 1 for someone who is receiving the national average income and is also in full health¹⁰). This analysis was then used to calculate a return-on-investment ratio per £ invested in the IY-TCM intervention.

Statistical Analyses of Trial Data

Obtaining and processing data

The individual level data were obtained from the STARS trial⁷ via secure information governance approaches. [Appendix](https://doi.org/10.1016/j.jval.2024.05.002)

[Table 1 in Supplemental Materials](https://doi.org/10.1016/j.jval.2024.05.002) found at <https://doi.org/10.1016/j.jval.2024.05.002>, shows key baseline characteristics and result. The arms were balanced in terms of age, gender, Income Deprivation Affecting Children Index (IDACI), and baseline measures of SDQ subscores. There were missing data on parent-reported outcomes (26% missing at 9 months), including parent-reported SDQ and measures of deprivation. The original STARS trial reported a difference in intervention effect size between those with (−1.6, 95% CI −2.8 to −0.4) and without (−0.8, 95% CI −1.7 to 0.1) missing data indicating that those with missing data might benefit the most. Therefore, multiple imputation was performed on the original STARS data set using the mice package¹¹ in R.¹²

Estimating change in SDQ

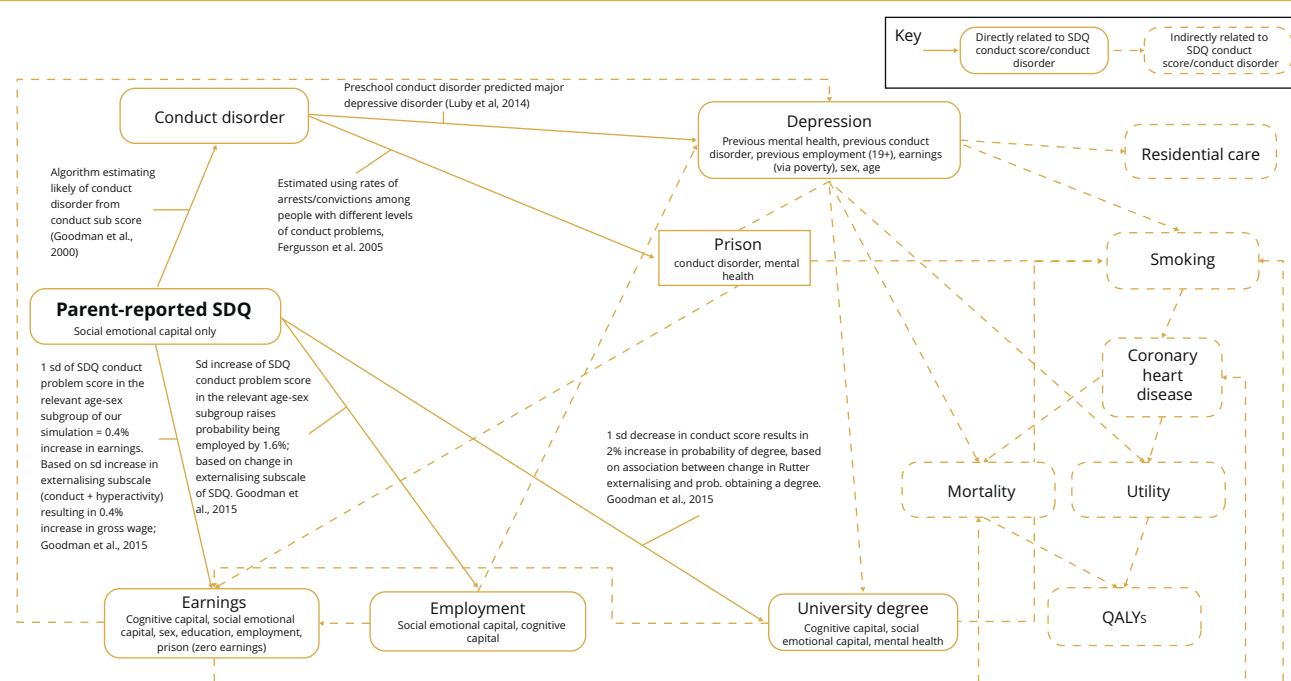
We undertook 12 regression analyses to estimate the expected change in 4 SDQ subscores (teacher-reported SDQ conduct subscore, teacher-reported SDQ impact subscore, parent-reported SDQ conduct subscore, parent-reported SDQ impact subscore) at the 3 follow-up time points (9, 18, and 30 months) in the trial. A random effects linear regression model using Generalized Estimating Equations to allow for the correlation of children's outcome scores within schools was used (also the method used in the original trial⁷). Our final regression models used covariates as follows: baseline SDQ subscore (either conduct [0-10] or impact [0-2] depending on the regression), year group, gender, intervention allocation, deprivation quintile categorically (using the IDACI deprivation quintile), and 2 interaction effects; between intervention allocation and both baseline SDQ subscore and IDACI deprivation quintile. This means that each subgroup defined by baseline SDQ, gender, and IDACI quintile had a different expected effect on SDQ subscores.

Representing individual level heterogeneity in effects

The regressions estimated the expected effect on SDQ conduct and impact subscores (for IY-TCM and control) at each time point adjusted for covariates, but there was heterogeneity around the actual effect observed. Some pupils had substantially larger improvements than the regression expected, and others had lower improvement or even worsening of SDQ subscores. This heterogeneity affects long-term outcomes and cost-effectiveness. To represent this in the model, we examined the residual difference between the actual observed and the predicted change in SDQ subscore and generated empirical distributions, which were then utilized in the adapted LifeSim model. See [Appendix 2.2 in Supplemental Materials](https://doi.org/10.1016/j.jval.2024.05.002) found at <https://doi.org/10.1016/j.jval.2024.05.002> for more detail.

Long-term Modeling With the Lifesim Microsimulation Model

LifeSim 1.0 is an existing health economic microsimulation model¹³ developed to estimate the lifetime health and economic impacts of policies affecting children (available open access: <https://github.com/ievask/lifesim-simulator>). The development¹³ and previous application of LifeSim¹⁴ have been described in detail previously and details of model structure, outcome, and input parameters can be found in [Appendix 3.1 and 3.2, Appendix Figures 1 and 2, and Appendix Tables 2-5 in Supplemental Materials](https://doi.org/10.1016/j.jval.2024.05.002) found at <https://doi.org/10.1016/j.jval.2024.05.002>. [Figure 1^{13,15-18}](https://doi.org/10.1016/j.jval.2024.05.002) shows the relationships that were coded into the LifeSim health economic microsimulation model¹³ and how the SDQ scores (detail in [Appendix 3.3 in Supplemental Materials](https://doi.org/10.1016/j.jval.2024.05.002) found at <https://doi.org/10.1016/j.jval.2024.05.002>) translate through to other factors and affect other longer-term outcomes.

Figure 1. Key Relationships between SDQ and longer-term outcomes in LifeSim model version 1.0.^{13,25-28}

SDQ indicates Strength and Difficulties Questionnaire; QALYs, quality-adjusted life year.

This model enabled us to estimate the long-term impact of an intervention, translating the short-term changes in a child's SDQ conduct and impact subscores, into estimated medium- and long-term outcomes. LifeSim 1.0 modeled ages 0 to 18 primarily by sampling observed life histories from the UK birth cohort study, the Millennium Cohort Study (MCS).¹⁹ It then modeled adulthood using a life-stage-specific network of equations parameterized using effect estimates from earlier longitudinal studies and target data from surveys and administrative records. These equations used research on the relationship between early life circumstances and skills formation with long-term outcomes^{20,21} to model causal pathways. Outcomes included the probability of graduating from university, time in (un)employment, life years spent in poverty, and living with health conditions, such as coronary heart disease and mental illness. LifeSim measured health gains including life expectancy, QALYs, and a well-being measure called years of good life, which represents both income and health.²² The model takes a broad lifetime societal perspective to capture the wider, long-term benefits of improving the mental health of young people. Cost and benefits were both discounted at 3.5% in line with NICE recommendations.

The research team worked with the LifeSim model developers to undertake adaptations. The existing LifeSim model used the parent-reported conduct problems and impact SDQ subscores as 2 key variables, which then were used to model longer-term consequences. For this analysis, the LifeSim model was adapted to simulate each year, in turn, for all 4 SDQ subscore measures (parent/teacher-reported conduct/impact). The algorithm used to estimate probability of conduct disorder was adapted so that it was based on either parent or teacher-reported subscore meeting certain criteria as in the original development of the algorithm (see Appendix 3.4 in Supplemental Materials found at <https://doi.org/10.1016/j.jval.2024.05.002>). Finally, the model was adapted to represent heterogeneity of intervention effect. The research team confirmed functionality in a series of model code verification tests.

The impact of the intervention on SDQ on ages 7 to 9 and ages 10 to 18 are shown in Eq. (1)-(3) and (3), respectively.

$$SDQ_{i,IYTCM,a=7} = SDQ_{i,MCS,a=7} + \Delta meanSDQ_{i,g,9months} + SampledErrorSDQ_{i,g,9month} \quad \text{Eq. (1)}$$

$$SDQ_{i,IYTCM,a=8} = SDQ_{i,MCS,a=8} + \Delta meanSDQ_{i,g,18month} + SampledErrorSDQ_{i,g,18month} \quad \text{Eq. (2)}$$

$$SDQ_{i,IYTCM,a=9} = SDQ_{i,MCS,a=9} + \Delta meanSDQ_{i,g,30month} + SampledErrorSDQ_{i,g,30month} \quad \text{Eq. (3)}$$

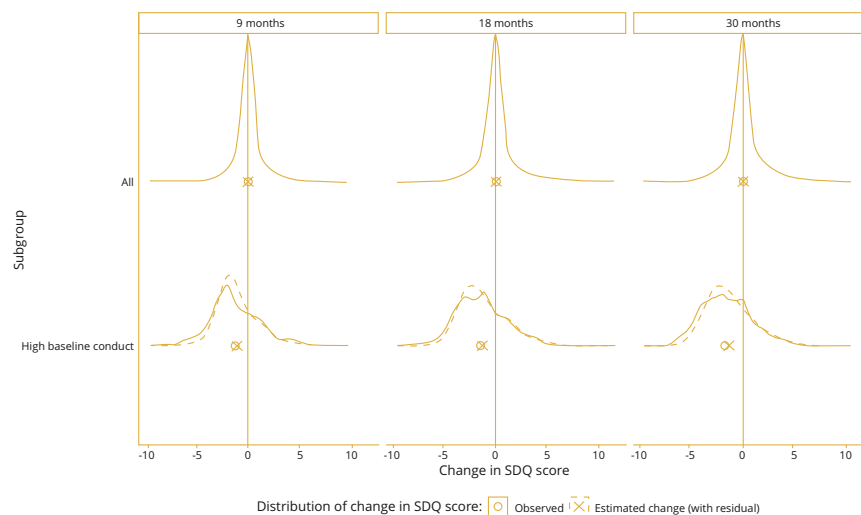
$$SDQ_{i,IYTCM,a=10} = SDQ_{i,IYTCM,a=9} + SDQ_{i,MCS,a=10} - SDQ_{i,MCS,a=9} \quad \text{Eq. (4)}$$

$$SDQ_{i,IYTCM,a=11} = SDQ_{i,IYTCM,a=10} + SDQ_{i,MCS,a=11} - SDQ_{i,MCS,a=10} \quad \text{Eq. (5)}$$

$$SDQ_{i,IYTCM,a=18} = SDQ_{i,IYTCM,a=17} + SDQ_{i,MCS,a=18} - SDQ_{i,MCS,a=17} \quad \text{Eq. (6)}$$

We assumed that each simulated participant starts with baseline SDQ subscores defined by the MCS at age 7. Change in SDQ subscore for the simulated individual receiving IY-TCM for intervention and the same simulated individual receiving control after 9, 18, and 30 months was applied at age 7, 8, and 9, respectively in the model. From age 10 to 18, we added the subscore for the previous year to the change in subscore estimated for the year in the LifeSim model so that the difference between the simulated SDQ score and the original SDQ predicted in the model (based on the MCS data) at 30 months was continued through to age 18. In practice, this means that we

Figure 2. Observed and estimated change in teacher-reported SDQ (conduct subscore) for all participants and a subgroup with a high score (4+) on the conduct subscale at baseline.



SDQ indicates Strength and Difficulties Questionnaire.

assumed that any difference between the simulated SDQ subscore for the IY-TCM and the control at 30 months on each subscore persisted through to age 18.

Base-Case, Subgroup, and Scenario Analysis

For the base-case analysis, we applied the intervention effect as described previously to the simulated individuals in the LifeSim model (which were based on participants of the MCS) and assumed an intervention cost per pupil of £11.52. Twenty thousand pupils were simulated (by bootstrapping from the 7 year olds in wave 4 of the MCS) with a lifetime time horizon. The comparator was the simulation of the control group from the trial. To examine uncertainty in treatment effects promulgated through the simulation model, we generated approximate upper and lower confidence interval scenarios for the effect of IY-TCM. More details on approximated confidence intervals is in [Appendix 3.3 in Supplemental Materials](https://doi.org/10.1016/j.jval.2024.05.002) found at <https://doi.org/10.1016/j.jval.2024.05.002>. We examined the cost-effectiveness of the intervention in 8 subgroups; pupils in each of the IMD quintile groups (5 analyses), pupils with baseline parent-reported SDQ conduct subscore of 4+ (high conduct subscore), pupils with a parent having baseline Kessler scale score of 13+ (indicating possible parental depression), and finally, the combination of these 2: pupils with both SDQ conduct subscore 4+ and a parent with Kessler 13+. The subgroups were based on categories of the simulated population within the LifeSim model. The original STARS study did not collect data on parental depression. Deprivation was measured using the IDACI quintiles in the STARS trial, and the IMD in the LifeSim model. Because of the high correlation between the IDACI and IMD²³ we assumed that quintiles of IDACI translates equivalently across to those of IMD. To reflect full implementation, we estimated the incremental effects and cost-benefit on a typical size year group in England ($n = 775\,000$).²⁴

Results

Regression Analyses

Full regression analyses are presented in [Appendix 4](#), [Appendix Tables 6 and 7](#) and [Appendix Figure 4 in Supplemental Materials](#)

found at <https://doi.org/10.1016/j.jval.2024.05.002>. Across all regression analyses, baseline score on the subscore (either conduct or impact depending on the regression) predicted change in score; poorer conduct or greater negative impact was significantly associated with a greater change in SDQ subscore. Gender consistently predicted change in SDQ scores with boys having a greater change over time (statistically significant at $P < .05$ for all outcomes other than parent-reported SDQ conduct score at 9 months). Deprivation was a statistically significant predictor of the outcome in 5 of the 12 regressions with greater deprivation predicting greater conduct problems.

[Figure 2](#) compares the observed change in SDQ and the predicted change in SDQ for each time point, first for the whole trial population and then for those with high teacher-reported baseline SDQ conduct sub score. The predicted change included the residual score which is a random value from an empirical distribution based on the difference between actual and predicted change in SDQ subscore. See [Appendix 3.2 in Supplemental Materials](#) found at <https://doi.org/10.1016/j.jval.2024.05.002> for more detail on how the residual was implemented. Equivalent figures for the remaining subgroups, parent-reported SDQ subscore and the teacher- and parent-reported impact score are in the [Appendix Figures 5 to 9 in Supplemental Materials](#) found at <https://doi.org/10.1016/j.jval.2024.05.002>. Estimated and observed results were very similar even when distributions were skewed. Change in SDQ was greater for those with high baseline levels of conduct, and there was a trend toward a more positive intervention effect for those living in areas of high deprivation.

Health Economic Modeling

Intervention effectiveness

[Table 1](#) shows the estimated long-term effectiveness and cost-effectiveness per pupil. The central estimate results show IY-TCM delivering a difference in the percentage of pupils with conduct disorder at age 7 (9.11% vs 9.72%). This reduction was much smaller by the time the simulation age was 18 (9.83% vs 9.85%). The mean teacher-reported SDQ subscores at age 7 also show an estimated improvement, that is, a lower score delivered by IY-TCM (mean difference = -0.13 on conduct subscore, and -0.23 on impact subscore). Examining long-term outcomes, there were

Table 1. Summary results of LifeSim estimated lifetime effects of incredible years teacher intervention per recipient.

Outcome	No intervention (LifeSim)	Control arm of STARS trial	Intervention arm of STARS trial	Difference between intervention and control		
				Central estimate	Approximated lower CI	Approximated upper CI
Pupils aged 7 with conduct disorder (%)	10.46	9.72	9.11	−0.61*	−0.20*	−0.87*
Pupils aged 18 with conduct disorder (%)	9.09	9.85	9.83	−0.02*	0.41†	−0.50*
SDQ conduct score aged 7 (parent)	1.33	1.41	1.42	0.01†	0.07†	−0.06*
SDQ conduct score aged 18 (self)	1.43	1.78	1.76	−0.02*	0.05†	−0.09*
SDQ impact score aged 7 (parent)	0.26	1.02	0.98	−0.04*	0.07†	−0.12*
SDQ impact score aged 18 (self)	0.26	1.71	1.51	−0.20*	−0.05*	−0.32*
SDQ conduct score aged 7 (teacher)	0.75	0.94	0.81	−0.13*	−0.09*	−0.16*
SDQ Impact score aged 7 (teacher)	0.35	0.91	0.68	−0.23*	−0.18*	−0.28*
University Graduates (% population)	38.23	37.77	37.77	0	−0.10†	0.07*
Unemployed during working years (%)	5.67	6.13	6.11	−0.02*	0.12†	−0.15*
Years spent in poverty (%)	28.05	28.59	28.58	−0.01*	0.14†	−0.16*
Years spend in prison (%)	1.61	1.73	1.73	0	0.07†	−0.08*
Smoking (%)	5.22	5.34	5.34	0	0.01†	−0.02*
Coronary heart disease (%)	6.59	6.6	6.6	0	−0.01*	0
Depression (%)	6.73	6.84	6.83	−0.01*	0.01†	−0.03*
Life years	79.77	79.89	79.89	0.00*	−0.01†	0.01*
Premature mortality (%)	25.86	25.56	25.55	−0.01*	0.02†	−0.04*
Average annual earnings	28 223	27 991	27 995	4*	−61†	69*
Average annual savings	2844	2823	2823	0	−4†	4*
Average annual consumption discounted	14 594	14 522	14 523	1*	−13†	15*
Healthy years	68.3019	68.3605	68.3621	0.0016*	−0.0144†	0.0177*
Healthy years discounted	40.9551	40.9769	40.9780	0.001*1	−0.0061†	0.0084*
Good years	66.4072	66.3648	66.3684	0.0036*	−0.0356†	0.0424*
Good years discounted	40.1140	40.0868	40.0887	0.0019*	−0.0165†	0.0202*
NICE reference case						
Intervention cost				£11.52		
ICER: Incremental cost per discounted QALY				£10 473		

continued on next page

Table 1. Continued

Outcome	No intervention (LifeSim)	Control arm of STARS trial	Intervention arm of STARS trial	Difference between intervention and control		
				Central estimate	Approximated lower CI	Approximated upper CI
Incremental net monetary health benefit (£20 000)				£10 [†]	–£134 [†]	£156*
Incremental net monetary health benefit (£30 000)				£21*	–£195 [†]	£240*
UK treasury reference case						
Intervention cost per discounted WELBY				£6063		
Value of WELBYs [‡]				£173*	–£1501 [†]	£1838*
Total net social benefit				£161*	–£1513 [†]	£1827*
Return on Investment				15.0		

ICER indicates incremental cost-effectiveness ratio; SDQ, Strength and Difficulties Questionnaire; QALYs, quality-adjusted life year; WELBYs, wellbeing-adjusted life years.

*Improvement in favor of the intervention.

[†]Change in favor of usual care

[‡]Good life years $\times 7 \times £13\ 000$.

small, simulated differences between IY-TCM and control in the percentage of working years spent unemployed (–0.02%), percentage of life years spent in poverty (–0.01%), and percentage of adult years living with a mental health problem (–0.01%). Premature mortality was also slightly reduced (–0.01%). There was a very small increase in estimated average annual earnings (+£4) and annual economic consumption (+£1).

Cost-effectiveness

These differences translated into a small gain in QALYs of +0.0011 per pupil. The traditional health economic analysis generated an estimated incremental cost per discounted QALY gained of £10 473 which is below the typical £20 000 to £30 000 per QALY WTP threshold used by NICE. This implied that an NHS or government decision maker could potentially consider the IY-TCM intervention to be cost-effective. However, there was considerable uncertainty. The approximated 95% confidence interval for discounted QALYs gained ranged from –0.0061 to +0.0084, and the incremental cost-per-QALY was almost equally likely to be above £30 000 per QALY as below the threshold (Appendix Fig. 9 in Supplemental Materials found at <https://doi.org/10.1016/j.jval.2024.05.002>). The IY-TCM intervention was estimated to result in a gain in the discounted good years of life (WELBYs) of +0.0019, which resulted in a cost per WELBY of £6063, overall financial value for the benefit per pupil at £173, and a net social benefit of £161 (Table 1).

Subgroup and Scenario Analysis

Intervention effectiveness

The full results on all outcomes are in Appendix Table 8 in Supplemental Materials found at <https://doi.org/10.1016/j.jval.2024.05.002>. The difference between IY-TCM intervention and the control group on percentage of pupils with conduct disorder at age 7 was much larger for the most deprived subgroup (IDACI quintile group 1; –1.14%) compared with the base case of all pupils (–0.61%). There was also larger difference for pupils with baseline parent-reported SDQ conduct 4+ (–1.65%), pupils with parent having baseline Kessler 13+ (–1.11%), and pupils with both SDQ conduct 4+ & parent having Kessler 13+ (–3.19%).

Cost-effectiveness

Pupils in the most deprived IDACI quintile group had around 8 times larger QALY gain (0.0081) compared with those not (0.0011), and pupils with a baseline SDQ conduct 4+ and parent having Kessler 13+ had around 30 times larger QALY gain (0.031; Fig. 3). IY-TCM was estimated to be more cost-effective in the subgroups in which large effects were seen (Table 2). For example, the overall incremental cost per discounted QALY gained was £10 473, but the equivalent figure for the most deprived Income Deprivation Affecting Children Index (IDACI) quintile group was £1422 per QALY, and for pupils with both SDQ conduct 4+ and parent having Kessler 13+, it was just £372 per QALY gained. The analysis using a years of good life approach also shows much larger net social benefits for the subgroups estimated to experience larger effects. Although uncertainty for the overall cost-effectiveness per pupil was substantial, the subgroup analyses can show much more certainty. For example, for pupils with both SDQ conduct 4+ and parent having Kessler 13+, the net benefit using the adjusted cost per pupil was positive (green in Table 2) no matter which end of the confidence interval was used for either QALYs or costs.

Finally, in Table 2, for a typical size year group in England, the total incremental cost was estimated at £23.46m, which would produce an estimated 853 QALYs, and an overall social benefit using WELBYs of £133.99 m, which was a net social benefit of £110 m and a return-on-investment ratio of 15. The subgroup analysis (Fig. 3) indicates that most of the net social benefit would be in the most deprived IDACI quintile and that estimated return on investment ratios per £1 invested for this subgroup and for the 3 subgroups defined SDQ conduct 4+ and parent having Kessler 13+, respectively, were 46.6, 62.2, 58.7, and 192.9.

Discussion

This was the first study to investigate the long-term effects and cost-effectiveness of the IY-TCM intervention. The results indicate that the intervention could be cost-effective; the analysis suggests a small gain in QALYs and WELBYs per pupil and an estimated incremental cost per discounted QALY (£10 473) that falls below

Table 2. Summary of subgroup results: estimated long-term effects of incredible years teacher intervention per recipient.

Cost-effectiveness outcome	All Pupils	IDACI Quintile 1	IDACI Quintile 2	IDACI Quintile 3	IDACI Quintile 4	IDACI Quintile 5	Parent-reported SDQ conduct 4+	Parent having baseline Kessler 13+	Pupils with both SDQ conduct 4+ and parent Kessler 13+
NICE reference case									
Intervention cost	£11.52	£11.52	£11.52	£11.52	£11.52	£11.52	£11.52	£11.52	£11.52
ICER: Incremental cost-per-QALY gained	£10 473	£1422	£5760	£19 200	–£3491*	–£16 457*	£1057	£993	£372
Incremental net monetary health benefit (£20 000)									
Central estimate	£10 [†]	£150 [‡]	£28 [‡]	£0 [‡]	–£78*	–£26*	£206 [‡]	£220 [‡]	£608 [‡]
Approximated upper CI (QALYs)	£156 [‡]	£284 [‡]	£146 [‡]	£192 [‡]	£90 [‡]	£92 [‡]	£318 [‡]	£340 [‡]	£718 [‡]
Approximated lower CI (QALYs)	–£134*	–£70*	–£76*	–£140*	–£210*	–£164*	£26 [‡]	–£156*	£318 [‡]
Incremental net monetary health benefit (£30 000)									
Central estimate	£21 [‡]	£231 [‡]	£48 [‡]	£6 [‡]	–£111*	–£33*	£315 [‡]	£336 [‡]	£918 [‡]
Approximated upper CI (QALYs)	£240 [‡]	£432 [‡]	£225 [‡]	£294 [‡]	£141 [‡]	£144 [‡]	£483 [‡]	£516 [‡]	£1083 [‡]
Approximated lower CI (QALYs)	–£195*	–£99*	–£108*	–£204*	–£309*	–£240*	£45 [‡]	–£228*	£483 [‡]
UK treasury reference case									
Intervention cost per discounted WELBY	£6063	£743	£19 200	£1920	–£1422*	–£6400*	£554	£591	£180
Value of WELBYs [‡]	£173 [‡]	£1411 [‡]	£55 [‡]	£546 [‡]	–£737*	–£164*	£1893 [‡]	£1775 [‡]	£5833 [‡]
Total net social benefit	£161 [‡]	£1399 [‡]	£43 [‡]	£534 [‡]	–£749*	–£175*	£1881 [‡]	£1763 [‡]	£5822 [‡]
Return on investment	15.0	122.4	4.7	47.4	–64.0*	–14.2*	164.3	154.0	506.3
Population analysis for a typical sized year group in England	775 000	132 551	163 767	162 059	163 379	153 245	66 004	27 877	6096
Total adjusted cost	£8 928 000	£4 008 341	£4 952 309	£4 900 649	£4 940 569	£4 634 131	£1 995 952	£842 996	£184 332
Total incremental QALYs gained (discounted)	853 [‡]	1074 [‡]	328 [‡]	97 [‡]	–539*	–107*	719 [‡]	323 [‡]	189 [‡]
Incremental net monetary health benefit [‡]	£16 647 000 [‡]	£30 682 900 [‡]	£7 939 417 [‡]	£1 050 139 [‡]	–£18 056 602*	–£4 983 530*	£20 822 850	£9 380 006 [‡]	£5 598 720 [‡]
Total net social benefit	£110 561 500 [‡]	£182 954 808 [‡]	£3 989 360 [‡]	£83 583 299 [‡]	–£125 366 926*	–£29 735 675*	£122 935 865 [‡]	£48 624 492 [‡]	£35 372 125 [‡]
Return on investment (WELBYs per £)	15.0 [‡]	46.6 [‡]	1.8 [‡]	18.1 [‡]	–24.4*	–5.4*	62.6 [‡]	58.7 [‡]	192.9 [‡]

ICER indicates incremental cost-effectiveness ratio; IDACI, Income Deprivation Affecting Children Index; SDQ, Strength and Difficulties Questionnaire; QALYs, quality-adjusted life year; WELBYs, wellbeing-adjusted life years.

*Xxx.

[†]Improvement in favor of the intervention.

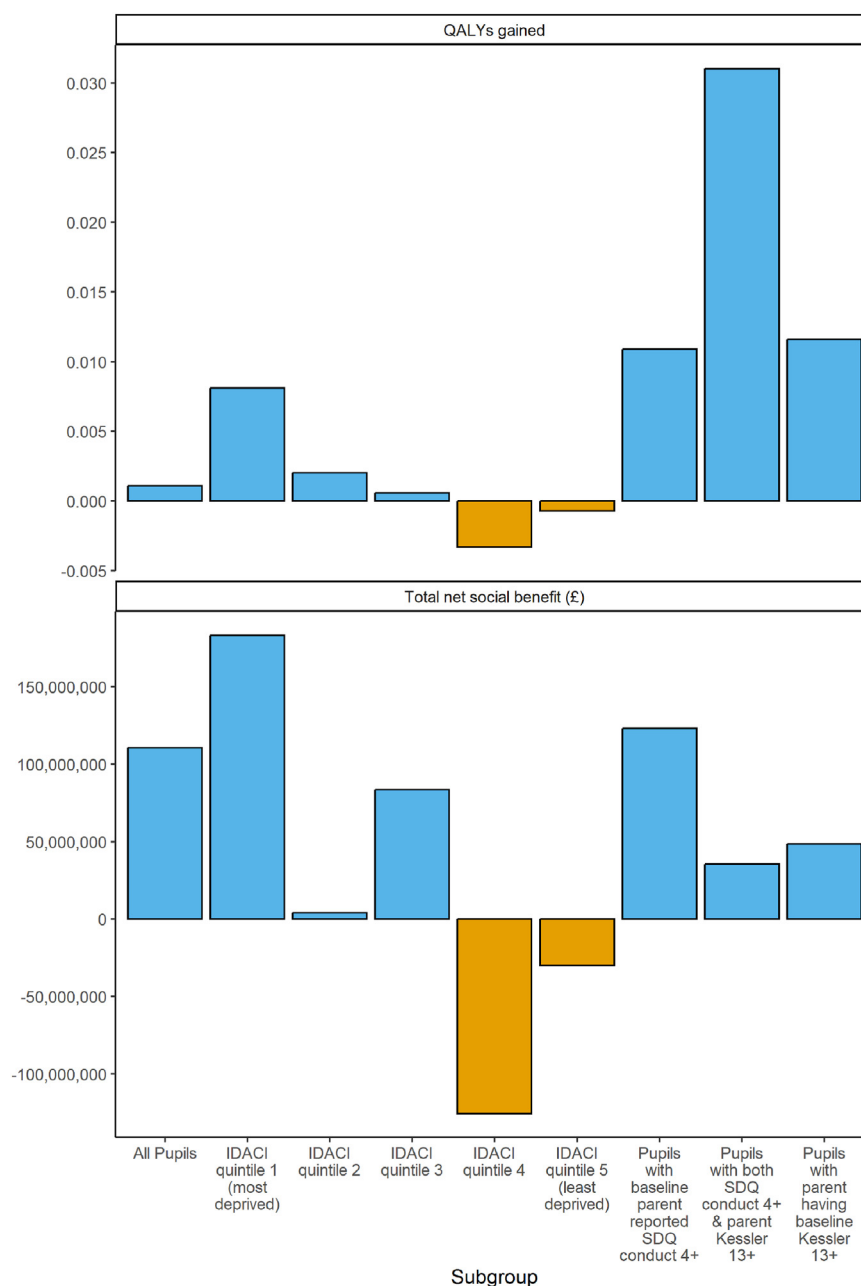
[‡]Xxx.

[§]Xxx.

the £20 000 NICE maximum acceptable ICER. Estimated effects and cost-effectiveness were greater for specific subgroups, including pupils in the most deprived IDACI quintile group and those with higher baseline SDQ conduct scores and a parent with depression. Although overall cost-effectiveness was subject to considerable uncertainty, there was more certainty that the intervention was cost-effective in these subgroups. When applied to a full year group, estimated net social benefits and return on investment were substantial.

This analysis built on the STARS trial by estimating effects beyond the end of the trial and measuring the benefits in QALYs and WELBYs, whereas the original analysis was only able to explore cost per change in SDQ for which there is not a known WTP threshold. By estimating a cost-per-QALY, we determined that the intervention could be cost-effectiveness based on a

maximum acceptable ICER for incremental QALY of £20 000. There was a large amount of uncertainty around the results because of the small effect size and wide confidence intervals around the effect size of the intervention. We can compare our results with the health economic analysis of the Incredible Years parent intervention,¹⁴ targeting parents with children at-risk of developing a conduct disorder. The parent version led to a greater increase in lifetime well-being (0.69 good years) compared with the IY-TCM (0.0019 good years). However, cost per good life year gained were more similar than difference in effect size might suggest (Parent program: £3212, Teacher program: £6063) due to large differences in cost per recipient (Parent program: £1773-£2660, Teacher program: £11.52). These findings suggest the potential for combined approach in which the classroom management intervention targets all children, whereas the parenting

Figure 3. Cost-benefit analysis by subgroup for England assuming a single year group with 750 000 pupils.

IDACI indicates Income Deprivation Affecting Children Index; QALYs, quality-adjusted life year.

intervention focuses on those at high risk. This approach overcomes attendance and engagement barriers associated with parent interventions and therefore children that may not have the opportunity to benefit from the parent intervention still could do so from IY-TCM.

The subgroup analysis provided further insight into cases in which estimated benefits were originating from and suggests that the intervention may be more cost-effective if applied in schools resembling the characteristics of the subgroups that particularly benefited in the analysis, which were living in deprived areas and with poor baseline conduct. Although our analysis indicates that IY-TCM may be more effective and cost-effective in these subgroups, factors such as implementation challenges (eg, staff turnover) and

role of school environment^{25,26} necessitate further research on how school characteristics affect intervention effectiveness.

The analysis has several strengths. The LifeSim model allowed us to represent the trajectories of physical and mental health for pupils with different circumstances and explore how these were differentially affected by the intervention in the subgroup analysis. Furthermore, by combining the effect size estimated through regression with the residual treatment effect, we simulated change in SDQ score that reflected the heterogeneity of the actual SDQ change observed in the trial. We represent both QALYs and WELBYs, which enable the effectiveness and cost-effectiveness to be evaluated from a well-being perspective, as well as using a traditional health economic approach.

There were some limitations. First, LifeSim does not include school-level variables. Effectiveness of classroom interventions are likely to be affected by the school-level factors, such as overall deprivation and eligibility for free school meals.²⁷ Future developments of the LifeSim model will explore how school-level variables can be represented. Second, LifeSim also focuses on conduct disorder. Benefits of the intervention on other aspects of mental health were not represented; therefore, the intervention effect may be underestimated. Third, Trajectories within the LifeSim model were based on observational data; therefore, unobserved confounders could potentially be generating part of the effects measured in the resulting coefficients. However, relationships between outcomes were based on plausible causal relationships for which there is a large body of theoretical and empirical evidence, and experts in childhood development and childhood policy, demography, epidemiology, human capital economics, and labor economics were consulted. Finally, because of limited MCS data available at the time of development, the outcomes between 15 and 18 were approximated based on the data for age 14. Future iterations of the model will be updated based on the more recent MCS data.

This analysis has highlighted some areas of further research. First, given the significant uncertainty, information about the participants beyond the trial would be beneficial, perhaps through linking trial data with the national pupil database. Additionally, this study highlights the potential to develop the LifeSim model to better assess school-based programs by including school-level factors. Finally, the model was not set up to conduct full probabilistic sensitivity analyses, which means that we could not take into account the distribution in costs and the relationship between costs and QALYs. In the future, methods such as analysis of variance could be applied.²⁸

In conclusion, this study suggests that the intervention could be cost-effective when looking at both cost per QALYs and WELBYs. There was a large amount of uncertainty in cost and outcomes, but both cost-effectiveness and certainty were greater when examining subgroups of pupils with higher deprivation, high baseline conduct score, or parents with higher depression scores. The intervention may be particularly cost-effective when implemented in a school with high levels of deprivation and conduct problems. Ongoing longer-term follow-up of the trial participants will give further insights into impacts on academic attainment and inform future modeling.

Author Disclosures

Author disclosure forms can be accessed below in the [Supplemental Material](#) section.

Supplemental Material

Supplementary data associated with this article can be found in the online version at <https://doi.org/10.1016/j.jval.2024.05.002>.

Article and Author Information

Accepted for Publication: May 10, 2024

Published Online: June 27, 2024

doi: <https://doi.org/10.1016/j.jval.2024.05.002>

Author Affiliations: Sheffield Centre for Health and Related Research, University of Sheffield, Sheffield, South Yorkshire, England, UK (Bates, Saidi, Brennan); Centre for Health Economics, University of York, York, North Yorkshire, England, UK (Cookson, Skarda); Developmental

Psychiatry, University of Cambridge, Cambridge, Cambridgeshire, England, UK (Ford); Department of Psychology, University of Exeter, Exeter, Devon, England, UK (Hayes); King's Health Economics, Health Service and Population Research, King's College London, London, Greater London, England, UK (Ganguli, Byford).

Correspondence: Sarah Bates, PhD, Sheffield Centre for Health and Related Research (SCHARR), University of Sheffield, Regent Court, 30 Regent Street, Sheffield, South Yorkshire, England, United Kingdom. Email: s.e.bates@sheffield.ac.uk

Author Contributions: *Concept and design:* Bates, Brennan, Ganguli
Acquisition of data: Bates, Brennan, Hayes, Ford, Saidi
Analysis and interpretation of data: Bates, Saidi, Brennan, Cookson, Byford, Ganguli
Drafting of the manuscript: Bates, Brennan, Ford, Skarda, Saidi
Critical revision of the paper for important intellectual content: Bates, Saidi, Brennan, Skarda, Cookson, Hayes, Ford, Byford, Ganguli
Statistical analysis: Bates, Brennan, Byford, Saidi
Provision of study material or patients: Byford, Hayes
Obtaining funding: Brennan, Hayes
Administrative, technical, or logistic support: Cookson, Hayes, Skarda
Supervision: Brennan, Cookson
Design and support with the long-term microsimulation modeling: Skarda

Funding/Support: This project is funded by the National Institute for Health Research (NIHR) School for Public Health Research Programme (grant reference number PD-SPH-2015). The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care. Ieva Skarda was supported by the Wellcome Trust (Grant No. 205427/Z/16/Z). The STARS trial was funded by the National Institute for Health and Care Research, Public Health Research Programme (NIHR PHR project number 10/3006/07) and the National Institute for Health and Care Research Applied Research Collaboration South West Peninsula (PenARC).

Role of the Funder/Sponsor: The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Acknowledgment: Thanks to the original STARS trial research team (Vanessa Edwards, Malcolm Fletcher, Stuart Logan, Brahm Norwich, Will Pritchard, Kate Allen, Matthew Allwood, Katie Grimes, Lorraine Hansford, Bryony Longdon, Shelley Norman, Anna Price, and Obioha C. Ukoumunne) for their work on the original trial and permitting the authors to use the individual participants data from the trial. Special thanks to Sarah Byford, Kings College London, for answering queries on cost analyses.

REFERENCES

1. Ford T, Macdiarmid F, Russell A, Racey D, Goodman R. The predictors of persistent DSM-IV disorders in 3-year follow-ups of the British Child and Adolescent Mental Health Surveys 1999 and 2004. *Psychol Med*. 2017;47(6):1126–1137.
2. Kim-Cohen J, Caspi A, Moffitt TE, Harrington H, Milne BJ, Poulton R. Prior juvenile diagnoses in adults with mental disorder: developmental follow-back of a prospective-longitudinal cohort. *Arch Gen Psychiatry*. 2003;60(7):709–717.
3. Polanczyk GV, Salum GA, Sugaya LS, Caye A, Rohde LA. Annual research review: a meta-analysis of the worldwide prevalence of mental disorders in children and adolescents. *J Child Psychol Psychiatry*. 2015;56(3):345–365.
4. Costello EJ, Egger H, Angold A. 10-year research update review: the epidemiology of child and adolescent psychiatric disorders: I. Methods and public health burden. *J Am Acad Child Adolesc Psychiatry*. 2005;44(10):972–986.
5. Ford T, Goodman R, Meltzer H. The relative importance of child, family, school and neighbourhood correlates of childhood psychiatric disorder. *Soc Psychiatry Psychiatr Epidemiol*. 2004;39(6):487–496.
6. Anderson LR. Adolescent mental health and behavioural problems, and intergenerational social mobility: a decomposition of health selection effects. *Soc Sci Med*. 2018;197:153–160.
7. Ford T, Hayes R, Byford S, et al. The effectiveness and cost-effectiveness of the Incredible Years Teacher Classroom Management programme in primary school children: results of the STARS cluster randomised controlled trial. *Psychol Med*. 2019;49(5):828–842.
8. Webster-Stratton C. *The incredible years: parents, teachers, and children training series*. Innov Ment Health Interv Child. Routledge; 2014:31–45.
9. Korest R, Carlson JS. A meta-analysis of the current state of evidence of the Incredible Years Teacher-Classroom Management program. *Children*. 2021;9(1):24.

10. HM Treasury, Social Impact Task Force. Wellbeing guidance for appraisal: supplementary green book guidance. https://assets.publishing.service.gov.uk/media/60fa9169d3bf7f0448719daf/Wellbeing_guidance_for_appraisal_-_supplementary_Green_Book_guidance.pdf; Published 2021. Accessed March 1, 2024.
11. Zhang Z. Multiple imputation with multivariate imputation by chained equation (MICE) package. *Ann Transl Med*. 2016;4(2):30.
12. R Core Team. *R: a language and environment for statistical computing*. <http://www.R-project.org/>; Published 2013. Accessed March 1, 2024.
13. Skarda I, Asaria M, Cookson R. LifeSim: a lifecourse dynamic microsimulation model of the Millennium birth cohort in England. *medRxiv*. 2021; 14(1):2–42.
14. Skarda I, Asaria M, Cookson R. Evaluating childhood policy impacts on lifetime health, wellbeing and inequality: lifecourse distributional economic evaluation. *Soc Sci Med*. 2022;302:114960.
15. Goodman R, Ford T, Simmons H, Gatward R, Meltzer H. Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. *Br J Psychiatry*. 2000;177(6):534–539.
16. Goodman A, Joshi H, Nasim B, Tyler C. Social and emotional skills in childhood and their long-term effects on adult life. Early Intervention Foundation. What Works for Children's Social Care. <https://www.eif.org.uk/report/social-and-emotional-skills-in-childhood-and-their-long-term-effects-on-adult-life>; Published 2015. Accessed March 1, 2024.
17. Fergusson DM, John Horwood L, Ridder EM. Show me the child at seven: the consequences of conduct problems in childhood for psychosocial functioning in adulthood. *J Child Psychol Psychiatry*. 2005;46(8):837–849.
18. Luby JL, Gaffrey MS, Tillman R, April LM, Belden AC. Trajectories of pre-school disorders to full DSM depression at school age and early adolescence: continuity of preschool depression. *Am J Psychiatry*. 2014; 171(7):768–776.
19. Connelly R, Platt L. Cohort profile: UK millennium Cohort study (MCS). *Int J Epidemiol*. 2014;43(6):1719–1725.
20. Almond D, Currie J, Duque V. Childhood circumstances and adult outcomes: act II. *J Econ Lit*. 2018;56(4):1360–1446.
21. García JL, Heckman JJ, Leaf DE, Prados MJ. Quantifying the life-cycle benefits of an influential early-childhood program. *J Pol Econ*. 2020;128(7):2502–2541.
22. Cookson R, Skarda I, Cotton-Barratt O, Adler M, Asaria M, Ord T. Quality adjusted life years based on health and consumption: a summary wellbeing measure for cross-sectoral economic evaluation. *Health Econ*. 2021;30(1):70–85.
23. Crawford C, Greaves E. A comparison of commonly used socioeconomic indicators: their relationship to educational disadvantage and relevance to teach first. <https://ifs.org.uk/publications/comparison-commonly-used-socio-economic-indicators-their-relationship-educational>; Published 2013. Accessed March 1, 2024.
24. UK Government. Academic year 2022/23 Schools, pupils and their characteristics. <https://explore-education-statistics.service.gov.uk/find-statistics/school-pupils-and-their-characteristics>; Published 2023. Accessed March 1, 2024.
25. Oberle E, Guhn M, Gadermann AM, Thomson K, Schonert-Reichl KA. Positive mental health and supportive school environments: a population-level longitudinal study of dispositional optimism and school relationships in early adolescence. *Soc Sci Med*. 2018;214:154–161.
26. Melendez-Torres G, Warren E, Viner R, Allen E, Bonell C. Moderated mediation analyses to assess intervention mechanisms for impacts on victimisation, psycho-social problems and mental wellbeing: evidence from the INCLUSIVE realist randomized trial. *Soc Sci Med*. 2021;279:113984.
27. Moore GF, Littlecott HJ, Turley R, Waters E, Murphy S. Socioeconomic gradients in the effects of universal school-based health behaviour interventions: a systematic review of intervention studies. *BMC Public Health*. 2015;15(1):1–15.
28. O'Hagan A, Stevenson M, Madan J. Monte Carlo probabilistic sensitivity analysis for patient level simulation models: efficient estimation of mean and variance using ANOVA. *Health Econ*. 2007;16(10):1009–1023.