Online Supplementary Appendices for "Evaluating Childhood Policy Impacts on Lifetime Health, Wellbeing and Inequality: Framework and Illustrative Application"

Contents

List of Tables

List of Figures

Appendix A Wellbeing Summary Measures

The three types of summary measures that we use are consumption (in £ per year), lifetime health (in healthy years) and lifetime wellbeing (in good years).

Consumption. Average individual consumption is calculated for each individual i as the lifetime average of their consumption in each life year,
i.e. $consumption_i = \frac{1}{(age.at.death+1)} \sum_{age=0}^{age.at.death} consumption_{i,age}$ where $age.at.death$ is the age at which individual i dies. We then obtain the average aggregate consumption in the cohort or specific group by averaging across the individuals:

$$Consumption = \frac{1}{N} \sum_{i=1}^{N} consumption_i \tag{1}$$

where N is the total number of people in the group we analyse.

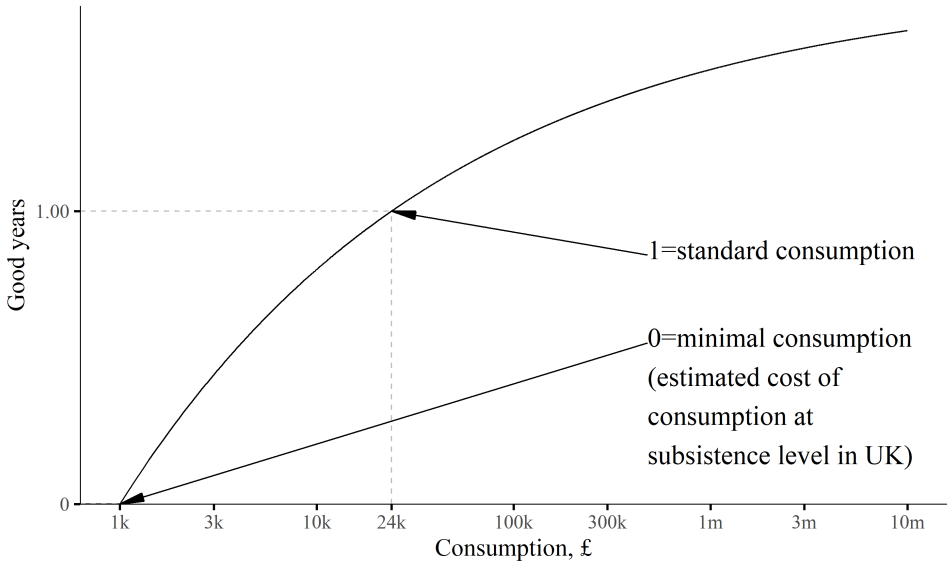
Lifetime Health. Individual lifetime health is calculated using a standard health QALY approach:
$health_i = \sum_{age=0}^{age.at.death} health_{i,age}$ where $health_{i,age}$ is the health-related quality of life of individual at a particular age, measured in healthy life years. We then obtain the aggregate lifetime health in the cohort or specific group that we analyse by aggregating across the individuals:

$$Health = \frac{1}{N} \sum_{i=1}^{N} health_i \tag{2}$$

Health-related quality of life depends on the two health outcomes that we model – mental illness (depression) and physical illness (CHD) – as well as the aggregate health quality in England. More specifically, $health_{i,age} = h(chd_{i,age}, depressed_{i,age})$, where $h(.)$ is a function decreasing in negative health experiences, and with a maximum of 1 when individual is in full health and anchored at 0 when individual is dead or in a health state as bad as death. More specifically, we assume $h(..) = \min[1, \max[0, \overline{health}[age, sex, sep] - (d(chd) \times chd_{i,age} + d(depressed) \times depressed_{i,age})]]$, where $\overline{health}[age, sex, sep]$ is the average health quality in England by age, sex and English IMD quintile group (Love-Koh et al., 2015), $d(x)$ represents the excess reduced health quality from the health condition x (we use data for health quality with affective disorders and coronary atherosclerosis from Sullivan et al. (2011)).

Lifetime Wellbeing. We follow Cookson et al. (2020), who quantify lifetime wellbeing in good (life) years by adjusting for both consumption and health-related quality of life. A good year is a year spent in full health and consuming the average consumption level in the UK. Figure A.1 plots the relationship between annual consumption and the 'good year' measure.

Figure A.1: Good Life Years



First, to quantify individual lifetime wellbeing for each individual i, we calculate $wellbeing_i = \sum_{age=0}^{age.at.death} w(health_{i,age}, consumption_{i,age})$ where $w(.)$ is a monotonically increasing function in health quality and consumption. We use the particular form $w(..) = health_{i,age} + u(consumption_{i,age})$ where $u(.)$ is a standard isoelastic utility of income function defined as $u(.) = A - B \times consumption_{i,age}^{1-\eta}$. The parameter $\eta > 1$ captures diminishing marginal value of income, and A and B are constants. We then obtain the aggregate wellbeing in the cohort or specific group that we analyse, by aggregating across individuals:

$$Wellbeing = \frac{1}{N} \sum_{i=1}^{N} wellbeing_i \tag{3}$$

where N is the total number of people in the group we analyse.

The constants A and B anchor the scale of the measure, such that 'one good year' represents a year lived in full health and by enjoying high living standards (consuming the income level of an average individual in the UK, c_{std}); and 'zero good years' represent a year either being dead or lived in a state that a person considers 'no better than death', given their level of health and income. If denote by c_{min} the lowest possible level of income at which life is considered

worth living for an individual in full health, then: $A = \frac{c_{min}^{(1-\eta)}}{c_{min}^{(1-\eta)} - c_{std}^{(1-\eta)}}$ and $B = \frac{1}{c_{min}^{(1-\eta)} - c_{std}^{(1-\eta)}}$. In our illustrative application we set c_{min} =£1,000 (estimated amount required to buy basic food supplies in the UK for a year) and c_{std} =£24,000 (the mean consumption in our simulated cohort), and $\eta = 1.26$ (see Cookson et al. (2020)).

The Production Cost of a Good Life-Year. Following Frijters and Krekel (2021), we estimate the production cost of a good life-year, or 'wellbeing QALY', as defined by Cookson et al. (2020), simply by reusing existing estimates of the production cost of a standard health QALY. If individual income is fixed at the average level of income then a health QALY and a wellbeing QALY are interchangeable - a score of 1 means the same thing. By construction, therefore, one wellbeing QALY has approximately the same value as one health QALY for someone with an average level of income. We use an estimate of £13,724 from Claxton et al. (2015).

This estimate of the production cost of a good life-year rests on the strong assumption that, at the margin, the wellbeing impacts of changes in government expenditure outside the health sector - e.g. in education, welfare, social care and justice - are the same as the wellbeing impacts of changes in health sector expenditure.

The production cost of a good life-year differs from the consumption value of a good life-year. The former concept is about the wellbeing impact of increasing or reducing public expenditure at the margin. The latter concept is about the wellbeing impact of increasing or reducing private consumption at the margin, via taxation. Standard estimates of the production cost of a QALY are much lower than standard estimates of the consumption value of a QALY. For example, the UK Treasury Green Book estimates that the consumption value of a health QALY is approximately £60,000 (HM Treasury, 2020), and UK Treasury guidance on wellbeing in economic appraisal estimates a consumption value of a WELLBY of £13,000 that implies a consumption value of a wellbeing QALY of £91,000 (7 times £13,000).

As with the health QALY, there is plenty of room for debate about whether the production cost of a good life-year or the consumption value of a good life-year is the appropriate "decision threshold" to use for assessing the cost-effectiveness of public expenditure. In this study we use the production cost. A helpful guide to this debate is at this blog: `https://aheblog.com/2020/09/23/opportunity-costs-marginal-productivity-and-cost-effectiveness-thresholds-`

what-are-they-and-how-are-they-related/.

Relationship Between WELLBY and Wellbeing QALY. UK Treasury wellbeing guidance for appraisal proposes that wellbeing should be quantified in terms of WELLBYs, where a WELLBY represents a one point improvement in life satisfaction, measured on 0-10 scale for one year (HM Treasury, 2021). As explained in the main manuscript, this implies that one WELLBY is worth one seventh of one wellbeing QALY.

This conversion rate is based on the following assumptions, in line with UK Treasury wellbeing guidance for appraisal:

- A year of life in full health on average income is equivalent to a year of life at the national average level of life satisfaction, of 8.
- A year of life as bad as death is equivalent to a year of life at a life satisfaction level of 1.
- A one point improvement in life satisfaction is equally valuable at any existing level of life satisfaction.

The WELLBY is sometimes described as a "wellbeing-adjusted life year", or a "wellbeing-year". These descriptions are somewhat misleading, however, since the WELLBY is actually only worth one seventh of a life-year - i.e. just under 2 months of life extension - for someone with an average life satisfaction score of 8. A WELLBY is only worth a full year of life extension for someone at an extremely low level of life satisfaction of 2. A more appropriate concise description for the WELLBY might be a "wellbeing-year-increment".

Since both WELLBYs and wellbeing QALYs have the same theoretical basis and the same convenient additive properties, the conversation process is straightforward - WELLBYs can be converted into wellbeing QALYs simply by multiplying by seven, and vice versa. This implies that lifetime wellbeing could also be measured using WELLBYs. For example, lifetime wellbeing for someone who lives for 80 years at a good level of wellbeing would be 560 WELLBYs (80 times 7). We think that the wellbeing QALY scale is more intuitive for presenting information about lifetime wellbeing, since it is on the same scale as length of life, and so that is the scale used in this study. However, it would be a straightforward matter to rescale things onto the WELLBY scale if desired.

Appendix B Modelling Public Service Costs

Table B.1 summarises the public service costs that we model.

Table B.1: Public Service Costs

Cost type	Components of the cost	Annual cost per person, £	Source
Coronary heart disease[1]	Direct health care cost;	840;	Liu et al. (2002);
	Informal care cost;	1,173;	
Mental illness	Costs to the National Health Service, the Accident and Emergency department, other support services (average);	5,260;	McCrone et al. (2008);
Conduct disorder	Cost to the National Health Service;	1,243 (age 5-10), 113 (age 11+);	Edwards et al. (2007); Scott et al. (2001), cited by Bonin et al. (2011);
	Cost to the Social Services Department;	175 (age 5-10), 70 (age 11+);	Edwards et al. (2007); Romeo, Knapp and Scott (2006), cited by Bonin et al. (2011);
	Cost to the Department for Education;	985 (age 5-10), 1,3402 (age 11-16), 0 (age 17+);	Edwards et al. (2007); Scott et al. (2001), cited by Bonin et al. (2011);
	Cost to the voluntary Sector;	26;	Edwards et al. (2007), cited by Bonin et al. (2011);
Prison	Unit annual costs of custody (per year);	31,925;	
	Unit costs of police (per record crime);	553;	Dubourg et al. (2005);
	Unit costs of courts (per court event);	7,103;	
Residential care	Cost of residential home;	29,934;	Curtis and Burns (2017);
Other healthcare	Average English National Health Service healthcare spending in the financial year 2011/12 by age, sex and English neighbourhood deprivation quintile group;	n/a;	Asaria (2017).

Note: We uprate all the costs to year 2015/16 prices.

Appendix C Technical Details on Modelling the Policy Effect

Gardner et al. (2017) measure child conduct problems using Eyberg Child Behaviour Inventory Intensity Scale (ECBIIS) and analyse a pooled dataset pooling data from several studies (further, pooled dataset – P). We use their estimates and model the policy effect which then is added to the corresponding age-specific baseline problem score of recipient child i at any age≥5, as:

$$effect.score_{i,age} = SD(\overline{score}_{L,5}) \left\{ E + \sum_{j=1}^{J} M_j \left(\frac{moderator_{j,i,5}}{SD(\overline{moderator}_{j,L,5})} - \frac{\overline{moderator}_{j,P}}{SD(\overline{moderator}_{j,P})} \right) \right\} \quad (4)$$

where $effect.score$ denotes the effect on our problem score (SDQ conduct problem score or impact score), E is the average effect size of IY on conduct problems reported by Gardner et al. (2017) and expressed in terms of the number of standard deviations of ECBIIS in the pooled dataset; and M_j is the effect size of the jth moderator variable $moderator^j$ (where $j = 1..J$), expressed in terms of the number of standard deviations of ECBIIS associated with a a standard deviation increase in the moderator variable above it's mean (in the pooled dataset). As mentioned, we use two moderators that Gardner et al. (2017) found significant (i.e., $J = 2$), and their estimated effect sizes: baseline conduct problems (measured using ECBIIS in the pooled dataset, and represented by SDQ conduct problem score at age 5 in LifeSim) and parental depression (measured using Beck Depression Inventory in the pooled dataset, and represented by parental 6-item Kessler psychological distress scale score at age 5 in LifeSim). More specifically, we use the following parameter values: $E = 0.46$, $M_1 = 0.05$ (moderator effect size for the baseline conduct problems), $M_2 = 0.09$ (moderator effect size for the parental depression).

Also, in equation (4), we need to convert the effect to the original problem score units, so we multiply the expression with the standard deviation of the problem score in the simulated LifeSim recipient group at age 5 (denoted $SD(\overline{score}_{L,5})$). We also need to standardise the moderator effects, so that an increase in the moderator variable above mean is asociated with a positive effects, and a decrease in the moderator variable below mean is asociated with a negative moderating effect. Therefore, we divide the age-5 simulated moderator effect variable, $moderator^j_{i,5}$, by its standard deviation in the simulated LifeSim recipient group at age 5 $SD(\overline{moderator}_{j,L,5})$, and subtract the standardised mean of the moderator variable i the pooled dataset $\frac{\overline{moderator}_{j,P}}{SD(\overline{moderator}_{j,P})}$.

Table C.2: LifeSim Causal Pathways Activated by the Parent Training Programme (Public Cost Savings)

Pathway to Conduct Disorder
SDQ conduct problem and impact scores → ***conduct disorder***
Pathway to Prison
SDQ conduct problem and impact scores → conduct disorder → ***prison***
Pathways to Mental Illness
SDQ conduct problem and impact scores → conduct disorder → ***childhood mental illness***
... childhood mental illness → ***adulthood mental illness***
Pathways to Taxes and Benefits
SDQ conduct problem and impact scores → conduct disorder → childhood mental illness → education → earnings → ***taxes and benefits***
SDQ conduct problem and impact scores → conduct disorder → prison → employment status → earnings → ***taxes and benefits***
SDQ conduct problem score → earnings → ***taxes and benefits***
SDQ conduct problem score → employment status → earnings → ***taxes and benefits***
... ***taxes and benefits*** → **consumption → mental illness → prison → employment status → earnings →** ***taxes and benefits***

Note: Outcomes in bold are the outcomes directly associated with public cost savings. Feedback loops between final benefits are written with the prefix "...". For example, mental illness can increase the risk of future mental illness.

Appendix D Detailed Evaluation Results

Table D.3: Average Policy Benefits for the Recipient Children

Outcome	Without policy (baseline)	With policy	Gain	Standard error
Child outcomes				
Conduct disorder at age 5, %	34.18	18.01	-16.17	0.38
Conduct disorder at age 18, %	17.67	12.48	-5.19	0.23
SDQ conduct problem score at age 5	4.73	4.03	-0.70	0.002
SDQ conduct problem score at age 18	2.96	2.35	-0.61	0.003
SDQ impact score at age 5	0.72	0.63	-0.09	0.002
SDQ impact score at age 18	1.05	0.94	-0.12	0.002
Cognitive skills at age 5 (standardised)	0.98	0.98	0.00	0.000
Cognitive skills at age 18 (standardised)	0.98	0.98	0.00	0.000
Adult outcomes				
University graduates, %	33.24	33.95	0.71	0.09
Working years in unemployment, %	9.04	7.66	-1.38	0.04
Life years in poverty, %	33.85	32.52	-1.32	0.04
Working years in prison, %	3.14	2.26	-0.88	0.04
Retirement years in residential care, %	4.22	3.59	-0.63	0.04
Adult years as a smoker, %	15.28	13.53	-1.75	0.08
Adult years with CHD, %	6.37	6.40	0.04	0.01
Life years with mental illness, %	13.00	11.41	-1.59	0.06
Total life years (life span)	78.52	78.69	0.17	0.02
Premature mortality rate (before age 75), %	29.01	28.54	-0.47	0.07
Annual earnings (lifetime average), £	29,511	29,800	291.34	10.09
Annual savings (lifetime average), £	2,807	2,848	41.23	1.82
Annual interest (lifetime average), £	327	355	27.93	0.78
Final wellbeing outcomes				
Annual consumption (lifetime average), £	21,589	21,876	286.62	8.44
Healthy years	66.03	66.46	0.43	0.02
Healthy years (discounted)	39.90	40.09	0.19	0.01
Good years	61.86	62.55	0.69	0.02
Good years (discounted)	37.54	37.85	0.31	0.01

Note: The effects are calculated on average per child recipient (9,228 child recipients in total). The gain is the average difference between the level of an outcome with and without the programme, where the averages are calculated by aggregating the individual outcomes over time and then averaging them across child-recipients. CHD – coronary heart disease; SDQ conduct problem score ranges 0-10 with a higher value representing more conduct problems; cognitive skills measure is a common factor extracted from the cognitive skills measures available in MCS, with a higher value representing better skills, standardised with a mean of 1.00 and standard deviation of 0.15. We use year 2015/16 prices and the annual discount rate of 1.5% (Paulden and Claxton, 2012).

Table D.4: Cumulative Cost Savings Over Various Time Periods

Public cost savings (per recipient), £	5 years	10 years	15 years	20 years	Lifetime
Conduct disorder	1,123	1,542	1,688	1,688	1,688
Healthcare: coronary heart disease	0	0	0	0	-9
Healthcare: mental illness	114	183	243	300	2,902
Healthcare: other	0	0	0	8	87
Prison	0	0	271	1,583	8,837
Residential care	0	0	0	0	643
Benefit payments	0	0	101	511	2,493
Tax revenues	0	0	8	109	2,818
Total savings	1,237	1,725	2,311	4,200	19,457

Note: Savings as a result of the parent-training programme per young child at risk of conduct disorder at age 5 in year 2015/16 prices, and discounted at 1.5% annual rate. See details on cost sources in table 1 in the main text.

Table D.5: Summary Statistics of the Simulated Outcomes

	Entire cohort (n=100,000)				Group of recipients (n=9,228)			
Outcome	Mean	SD	Min	Max	Mean	SD	Min	Max
General characteristics								
Sex: male (indicator)	0.49	0.50	0	1	0.60	0.49	0	1
Age at death	79	13	1	100	79	13	6	100
Parental characteristics								
Parental depression (indicator if parental 9-item Rutter malaise inventory score 4+)	0.14	0.35	0	1	0.27	0.45	0	1
Parental education (indicator if parental NVQ 4+)	0.31	0.46	0	1	0.21	0.41	0	1
Parental socio-economic position (income quintile group of household at birth)	3.06	1.37	1	5	2.34	1.26	1	5
Skills and education								
Social skills: SDQ conduct problem score (up to age 18)	1.79	1.83	0	10	3.82	2.17	0	10
Cognitive skills (up to age 18)	1.02	0.14	0.39	1.50	0.97	0.14	0.60	1.43
Conduct disorder (indicator, age 5-18)	0.09	0.28	0	1	0.21	0.41	0	1
Education: university degree (indicator, age 19)	0.39	0.49	0	1	0.33	0.47	0	1
Health								
Unhealthy behaviour: smoking (indicator, age 19+)	0.10	0.30	0	1	0.15	0.36	0	1
Mental illness: emotional disorder between ages 5-18, depression age 19+ (indicator, age 5+)	0.10	0.30	0	1	0.14	0.35	0	1
Physical illness: coronary heart disease (indicator, age 19+)	0.07	0.25	0	1	0.07	0.25	0	1
Financial								
Earnings from employment, £ (age 19-69)	28,199	12,703	0	112,411	27,201	13,654	0	95,171
Earnings from interest, £ (age 19+)	525	504	0	4,843 433	463	0	3,485	
Disposable income, £ (age 19+)	25,640	9,218	10,000	81,059	25,009	9,605	10,000	68,632
Individual socio-economic position (income quintile group of householdage 0-69)	2.80	1.24	1	5 2.55	1.23	1	5	
Savings, £ (age 19-69)	2,837	2,006	0	12,969	2,814	2,038	0	10,981
Family wealth, £	40,657	49,383	0	484,274	33,236	44,410	0	348,517
Individual taxes paid, £ (age 19+)	2,290	2,768	0	35,206	2,215	2,769	0	26,921
Individual benefits received, £	369	1,931	0	31,110	537	2,456	0	31,110
Pension, £ (age 70+)	8,160	0	8,160	8,160	8,160	0	8,160	8,160
Bad life outcomes								
Prison (indicator, age 19-69)	0.02	0.13	0	1	0.03	0.17	0	1
Residental care (indicator, age 70+)	0.03	0.16	0	1	0.04	0.20	0	1
Unemployed (indicator, age 19-69)	0.06	0.23	0	1	0.09	0.28	0	1
Poverty: below 60% median income (indicator, age 0-69, see note below)	0.21	0.41	0	1	0.28	0.45	0	1
Wellbeing measures								
Yearly consumption, £	23,962	13,010	10,000	128,246	21,479	10,914	10,000	110,480
Yearly consumption, £ (discounted)	14,742	12,052	2,256	128,246	12,806	9,115	2,256	110,480
Lifetime health, healthy years	67.53	10.06	0.87	87.87	66.08	10.17	5.58	86.41
Lifetime health, healthy years (discounted)	40.62	4.00	0.87	47.86	39.93	4.02	5.37	47.21
Lifetime wellbeing, good years	64.94	10.36	0.69	91.65	61.91	10.34	4.89	89.41
Lifetime wellbeing, good years (discounted)	39.49	4.96	0.69	52.04	37.57	4.87	4.68	50.86

Table continues on the next page.

Table B.2: Summary Statistics of the Simulated Outcomes (*Continued*)

	Top gainers (n=354)			
Outcome	Mean	SD	Min	Max
General characteristics				
Sex: male (indicator)	0.76	0.43	0	1
Age at death	79	12	42	100
Parental characteristics				
Parental depression (indicator if parental 9-item Rutter malaise inventory score 4+)	0.25	0.44	0	1
Parental education (indicator if parental NVQ 4+)	0.19	0.40	0	1
Parental socio-economic position (income quintile group of household at birth)	2.10	1.22	1	5
Skills and education				
Social skills: SDQ conduct problem score (up to age 18)	4.39	1.85	0	10
Cognitive skills (up to age 18)	0.95	0.14	0.62	1.36
Conduct disorder (indicator, age 5-18)	0.35	0.48	0	1
Education: university degree (indicator, age 19)	0.27	0.44	0	1
Health				
Unhealthy behaviour: smoking (indicator, age 19+)	0.46	0.50	0	1
Mental illness: emotional disorder between ages 5-18, depression age 19+ (indicator, age 5+)	0.35	0.48	0	1
Physical illness: coronary heart disease (indicator, age 19+)	0.07	0.26	0	1
Financial				
Earnings from employment, £ (age 19-69)	22,088	15,415	0	79,628
Earnings from interest, £ (age 19+)	125	194	0	1,632
Disposable income, £ (age 19+)	21,802	9,867	10,000	59,631
Individual socio-economic position (income quintile group of householdage 0-69)	2.22	1.22	1	5
Savings, £ (age 19-69)	2,537	2,098	0	9,541
Family wealth, £	9,592	17,779	0	163,236
Individual taxes paid, £ (age 19+)	1,718	2,443	0	20,920
Individual benefits received, £	1,396	4,364	0	31,110
Pension, £ (age 70+)	8,160	0	8,160	8,160
Bad life outcomes				
Prison (indicator, age 19-69)	0.16	0.37	0	1
Residental care (indicator, age 70+)	0.12	0.33	0	1
Unemployed (indicator, age 19-69)	0.23	0.42	0	1
Poverty: below 60% median income (indicator, age 0-69, see note below)	0.41	0.49	0	1
Wellbeing measures				
Yearly consumption, £	18,663	10,433	10,000	78,324
Yearly consumption, £ (discounted)	11,273	8,726	2,256	78,324
Lifetime health, healthy years	62.92	8.21	36.19	80.23
Lifetime health, healthy years (discounted)	38.58	2.93	27.19	44.06
Lifetime wellbeing, good years	56.15	7.61	30.01	72.72
Lifetime wellbeing, good years (discounted)	34.78	3.73	22.17	43.53

Note: In year 2015/16 prices. The discounted variables are discounted at 1.5% annual rate. Top gainers are the recipients who gain at least five good years from the parent-training programme.

Appendix E Comparison with Long-Term Trial Follow-Up Data

This section compares our estimates with data from two long-term trial follow-up studies with mean follow-up of 7 years, by conducting subgroup analysis by different levels of baseline conduct problem scores (Scott, Briskman and O'Connor, 2014). These authors analyse the effect of "Incredible Years" on conduct disorder in an indicated sample based on GP or other professional indications of conduct disorder, and a selectively screened sample which generally had lower conduct problem scores at baseline. They find a significant positive effect on conduct disorder in the indicated sample but no effect in the lower-risk sample.

Figure E.2 shows our estimates of the policy effect on conduct disorder, SDQ conduct problem score and SDQ impact score over time, when delivering parent training to groups of recipients screened using different SDQ conduct problem score cut-offs. In particular, the plots with "2+" and "5+" correspond to child recipients above the 75th and 97th percentile respectively of the conduct problem severity distribution.[2] These groups can be compared to the groups analysed by Scott, Briskman and O'Connor (2014), with children above the 82th and 97th percentile of the Parent Account of Child Symptoms measure. Figure E.2 shows that in the "2+" group the children experience a negligible policy effect in terms of conduct disorder (around 2% reduction); in the "5+" group – a larger effect (around 5-9%). Our general patterns of sub-group estimates are therefore broadly comparable with the findings of Scott, Briskman and O'Connor (2014).

[2]SDQ frequency for British 5-10 year olds, both sexes, available at `https://www.sdqinfo.com/norms/UKNorm8.pdf`; accessed on 2020-02-09.

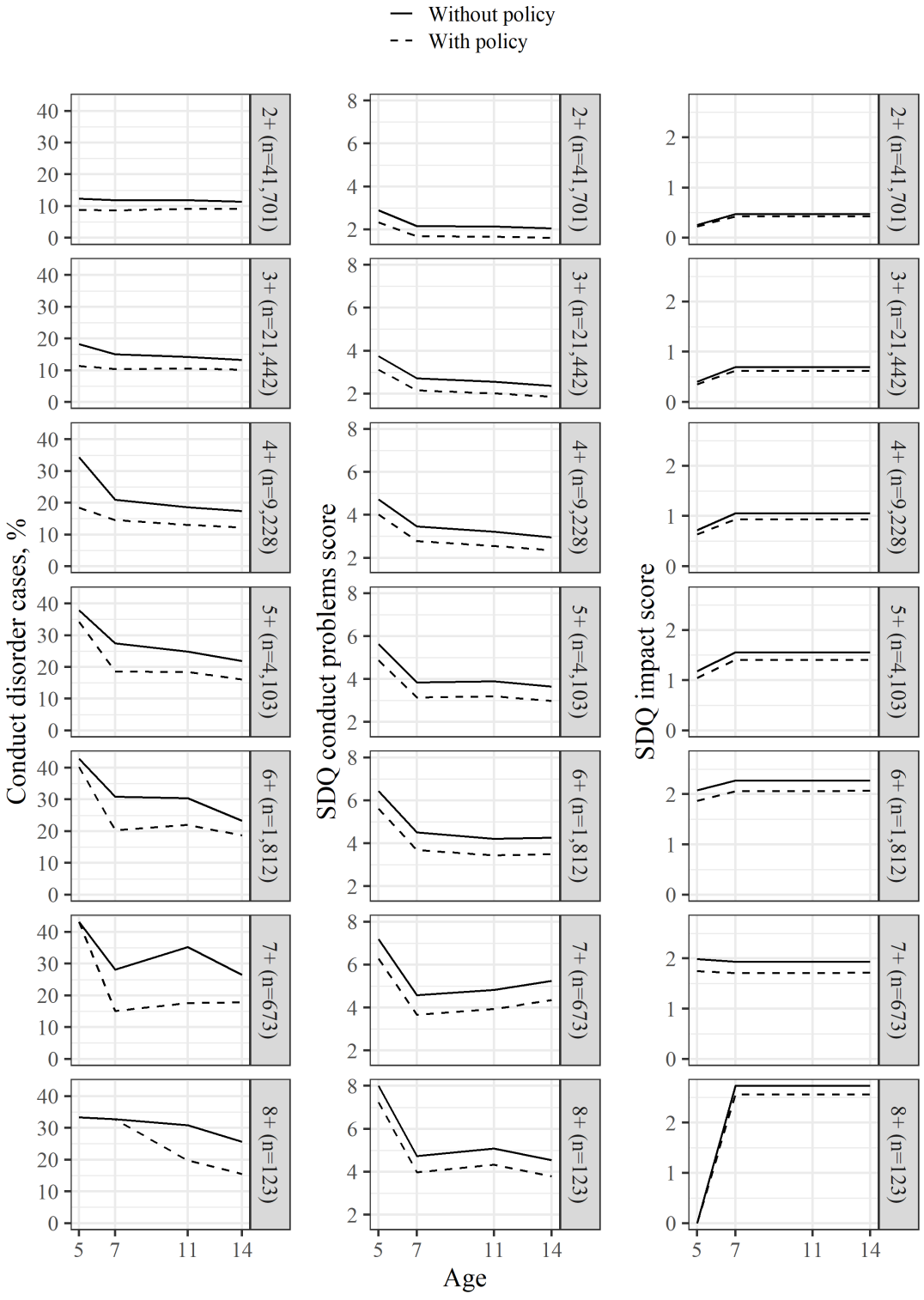


Figure E.2: The Policy Effect on Conduct Disorder, Conduct Problem and Impact Scores Over Time for Children with Various Levels of Conduct Problem Severity

Note: The first number in the shaded box to the right of each plot marks the SDQ conduct problem score cut-off value above which children are included in the sample, i.e. the plot marked with "2+" corresponds to the sample screened based on SDQ conduct problem score greater or equal than 2.

Appendix F Sensitivity Analysis

F.1 Alternative Policy Effects

In this section we compare different ways of modelling the policy effect: a simple version assuming a homogeneous effect for everyone, and more sophisticated versions that allow for group-level and individual-level heterogeneity in short-term effects (see Table F.6 and Figure F.4(a)). We also vary the policy effect by plus or minus 50% of the initially assumed level (see Figure F.4(b)).

Homogeneous Effect. We assume a homogeneous decrease in the child's conduct problem and impact scores at age 5, which persists throughout the subsequent years, i.e. for age ≥ 5:

$$score_{i,age}^{withpolicy} = \max[0, \quad score_{i,age}^{withoutpolicy} - E \times SD(\overline{score}_{L,5})] \tag{5}$$

where $score_{i,age}^{withoutpolicy}$ denotes either the child's SDQ conduct problem score or impact score without policy and $score_{i,age}^{withpolicy}$ is either the predicted conduct problem score or impact score with policy; E is the policy effect on child's score, assumed to be constant over time and across the child recipients (see section 3.2 in the main paper, in particular the equation (5) in that section for which we define E and $SD(\overline{score}_{L,5})$.

Group-Level Heterogeneity. This is the way policy effect is modelled in the main paper. A higher effect is modelled for children with higher baseline conduct problems and whose parents have worse mental health (Gardner et al., 2017):

$$score_{i,age}^{withpolicy} = \max[0, \quad score_{i,age}^{withoutpolicy} - effect.score_{i,age}] \tag{6}$$

where $effect.score_{i,age}$ is governed by equation (4).

Group and Individual-Level Heterogeneity. In addition to group-level heterogeneity, the effect also has an individual random error which represents random individual heterogeneity:

$$score_{i,age}^{withpolicy} = \max[0, \quad score_{i,age}^{withoutpolicy} - effect.score_{i,age} + error_{i,age}] \tag{7}$$

where $effect.score_{i,age}$ is governed by equation equation (4) and $error_{i,age}$ is a random error with mean zero and standard deviation equal to 1/4 of the effect size.

Table F.6: Average Policy Benefits: Different Ways of Modelling the Policy Effect

		Homogeneous effect		Group-level heterogeneity		Group and individual-level heterogeneity	
Outcome	Baseline	Gain	SE	Gain	SE	Gain	SE
Child outcomes							
Conduct disorder at age 5, %	34.18	-15.34	0.3752	-16.17	0.3833	-16.48	0.3863
Conduct disorder at age 18, %	17.67	-4.90	0.2249	-5.19	0.2312	-5.20	0.2314
SDQ conduct problem score at age 5	4.73	-0.67	0.0000	-0.70	0.0022	-0.70	0.0028
SDQ conduct problem score at age 18	2.96	-0.58	0.0024	-0.61	0.0032	-0.61	0.0036
SDQ impact score at age 5	0.72	-0.08	0.0013	-0.09	0.0015	-0.09	0.0016
SDQ impact score at age 18	1.05	-0.10	0.0014	-0.12	0.0017	-0.12	0.0018
Cognitive skills at age 5 (standardised)	0.98	0.00	0.0000	0.00	0.0000	0.00	0.0000
Cognitive skills at age 18 (standardised)	0.98	0.00	0.0000	0.00	0.0000	0.00	0.0000
Adult outcomes							
University graduates, %	33.24	0.64	0.08	0.71	0.09	0.76	0.09
Working years in unemployment, %	9.04	-1.31	0.04	-1.38	0.04	-1.38	0.04
Life years in poverty, %	33.85	-1.26	0.04	-1.32	0.04	-1.33	0.04
Working years in prison, %	3.14	-0.82	0.04	-0.88	0.04	-0.88	0.04
Retirement years in residential care, %	4.22	-0.60	0.04	-0.63	0.04	-0.63	0.04
Adult years as a smoker, %	15.28	-1.67	0.08	-1.75	0.08	-1.75	0.08
Adult years with CHD, %	6.37	0.04	0.01	0.04	0.01	0.04	0.01
Life years with mental illness, %	13.00	-1.54	0.05	-1.59	0.06	-1.59	0.06
Years of life	78.52	0.15	0.02	0.17	0.02	0.17	0.02
Premature mortality rate (before age 75), %	29.01	-0.43	0.07	-0.47	0.07	-0.45	0.07
Annual earnings (lifetime average), £	29,510.56	273.57	9.74	291.34	10.09	289.24	10.09
Annual savings (lifetime average), £	2,806.63	39.25	1.77	41.23	1.82	40.75	1.81
Annual interest (lifetime average), £	326.60	27.02	0.78	27.93	0.78	27.92	0.78
Final wellbeing outcomes							
Annual consumption (lifetime average), £	21,589	273.61	8.31	286.62	8.44	286.89	8.42
Healthy years	66.03	0.41	0.02	0.43	0.02	0.43	0.02
Healthy years (discounted)	39.90	0.18	0.01	0.19	0.01	0.19	0.01
Good years	61.86	0.67	0.02	0.69	0.02	0.70	0.02
Good years (discounted)	37.54	0.30	0.01	0.31	0.01	0.32	0.01

Note: The effects are calculated on average per child recipient (9,228 child recipients in total). The gain is the average difference between the level of an outcome with and without the programme, where the averages are calculated by aggregating the individual outcomes over time and then averaging them across child-recipients. SE – standard error; CHD – coronary heart disease; SDQ conduct problem score ranges 0-10 with a higher value representing more conduct problems; cognitive skills measure is a common factor extracted from the cognitive skills measures disseminated by MCS, with a higher value representing better skills, standardised with a mean of 1.00 and standard deviation of 0.15. We use year 2015/16 prices and the annual discount rate of 1.5%.

F.2 Sleeper and Fadeout Effects

While Feinstein, Chowdry and Asmussen (2017) highlight the importance of policy effect fadeout within the context of childhood policy evaluation, Van Aar et al. (2017) find considerable heterogeneity in the sustainability of the effects of interventions for child conduct problems. They find that some trials exhibit fadeout effects (effect becomes weaker over time) and others – sleeper effects (effect becomes stronger over time), and the type of effects could not be explained by any of the moderators tested. The change in effect size over time, between the immediate post-test and long term follow-up ranged between -0.65 and 0.65, with an insignificant 0.01 (p=0.78) on average; 16% of the 91 effect sizes (from 40 trials) indicated significant fadeout effects and 12% indicated significant sleeper effects.

To investigate the possible maximum impact of potential fadeout and sleeper effects, we model the most extreme cases, i.e. a reduction in the initial effect on the probability of getting a conduct disorder by 0.65 effect size units within a year post delivery (extreme fadeout effect) and an increase in the initial effect by 0.65 effect size units (extreme sleeper effect). The analysis assumes the policy with group-level heterogeneity, and the results are presented in Figure F.3, and Figure F.4(c) and Table F.7.

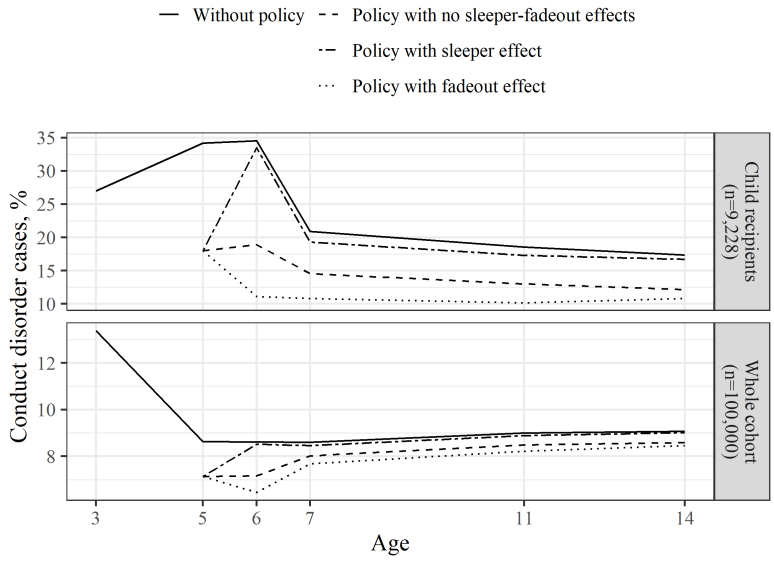
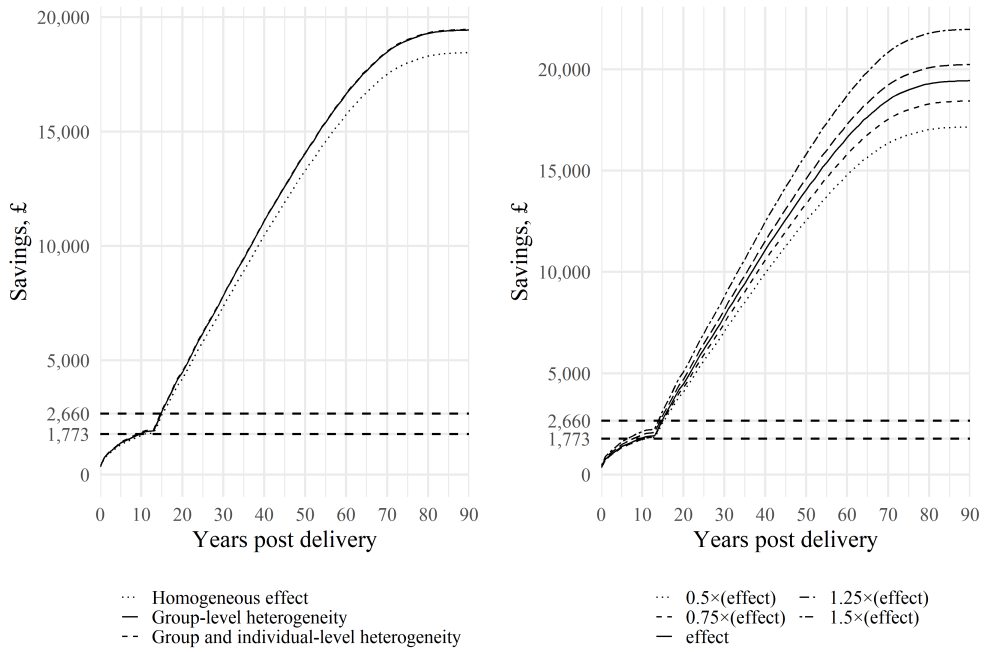


Figure F.3: Prevalence of Conduct Disorder Over Time: Sleeper and Fadeout Effects



(a) Different Ways of Modelling the Effect

(b) Varying the Effect

(c) Sleeper and Fadeout Effects

Figure F.4: Cumulative Cost Savings Over Time: Sensitivity Analysis

Note: The dashed lines represent the range of estimated unit costs of the "Incredible Years"; 'effect' denotes the policy effect. All plots in panels (b) and (c) assume the policy with group-level heterogeneity.

Table F.7: Average Policy Benefits: Sleeper and Fadeout Effects

		Fadeout effect		No fadeout-sleeper effects		Sleeper effect	
Outcome	Baseline	Gain	SE	Gain	SE	Gain	SE
Child outcomes							
Conduct disorder at age 5, %	34.18	-16.17	0.3833	-16.17	0.3833	-16.17	0.3833
Conduct disorder at age 18, %	17.67	-0.59	0.0796	-5.19	0.2312	-6.63	0.2594
SDQ conduct problem score at age 5	4.73	-0.70	0.0022	-0.70	0.0022	-0.70	0.0022
SDQ conduct problem score at age 18	2.96	-0.02	0.0007	-0.61	0.0032	-1.33	0.0063
SDQ impact score at age 5	0.72	-0.09	0.0015	-0.09	0.0015	-0.09	0.0015
SDQ impact score at age 18	1.05	-0.01	0.0003	-0.12	0.0017	-0.26	0.0035
Cognitive skills at age 5 (standardised)	0.98	0.00	0.0000	0.00	0.0000	0.00	0.0000
Cognitive skills at age 18 (standardised)	0.98	0.00	0.0000	0.00	0.0000	0.00	0.0000
Adult outcomes							
University graduates, %	33.24	0.00	0.00	0.71	0.09	1.68	0.13
Working years in unemployment, %	9.04	-0.11	0.01	-1.38	0.04	-2.26	0.05
Life years in poverty, %	33.85	-0.10	0.01	-1.32	0.04	-2.33	0.05
Working years in prison, %	3.14	-0.09	0.01	-0.88	0.04	-1.14	0.05
Retirement years in residential care, %	4.22	-0.06	0.01	-0.63	0.04	-0.93	0.04
Adult years as a smoker, %	15.28	-0.18	0.03	-1.75	0.08	-2.37	0.09
Adult years with CHD, %	6.37	0.00	0.00	0.04	0.01	0.06	0.01
Life years with mental illness, %	13.00	-0.17	0.02	-1.59	0.06	-2.32	0.06
Years of life	78.52	0.01	0.01	0.17	0.02	0.27	0.03
Premature mortality rate (before age 75), %	29.01	-0.02	0.02	-0.47	0.07	-0.75	0.09
Annual earnings (lifetime average), £	29,510.56	19.11	2.86	291.34	10.09	494.34	12.54
Annual savings (lifetime average), £	2,806.63	3.30	0.52	41.23	1.82	63.74	2.24
Annual interest (lifetime average), £	326.60	2.04	0.27	27.93	0.78	50.47	0.98
Final wellbeing outcomes							
Annual consumption (lifetime average), £	21,589.23	17.56	2.74	286.62	8.44	498.70	10.21
Healthy years	66.03	0.04	0.01	0.43	0.02	0.67	0.02
Healthy years (discounted)	39.90	0.02	0.00	0.19	0.01	0.29	0.01
Good years	61.86	0.06	0.01	0.69	0.02	1.14	0.03
Good years (discounted)	37.54	0.03	0.00	0.31	0.01	0.51	0.01

Note: The effects are calculated on average per child recipient (9,228 child recipients in total). The gain is the average difference between the level of an outcome with and without the programme, where the averages are calculated by aggregating the individual outcomes over time and then averaging them across child-recipients. The model with fadeout affects assumes a decrease and the model with sleeper effects – an increase in the initial policy effect by 0.65 effect size units the subsequent year post delivery (between ages 5-6). The effects are for the policy with group-level heterogeneity. SE – standard error; CHD – coronary heart disease; SDQ conduct problem score ranges 0-10 with a higher value representing more conduct problems; cognitive skills measure is a common factor extracted from the cognitive skills measures disseminated by MCS, with a higher value representing better skills, standardised with a mean of 1.00 and standard deviation of 0.15. We use year 2015/16 prices and the annual discount rate of 1.5%.

Appendix G Further Details of Policy Targeting Analysis

Table 3 in the main paper demonstrates that some individuals – the top gainers – benefit substantially from the parent training programme.

The top gainers are predominantly individuals who experience a cluster of multiple bad life outcomes at baseline, and for whom the policy is beneficial in preventing the clusters of bad life outcomes. Figure G.5 compares the top gainers with other recipients in terms of bad life outcomes with and without the policy in four twenty-five-year periods. Bad life outcomes are defined as any of the following: having a conduct disorder at least once, being in prison at least once, smoking for at least five years, CHD for at least five years, depression for at least five years, being unemployed for at least five years, living in poverty for at least five years, living in residential care home at least once, and death.[3]

[3]Because many people smoke, experience unemployment, depression, CHD and poverty at some point in their lives, in Figure G.5 a person is marked as having a bad life outcome only if these outcomes occurred for at least five years in total.

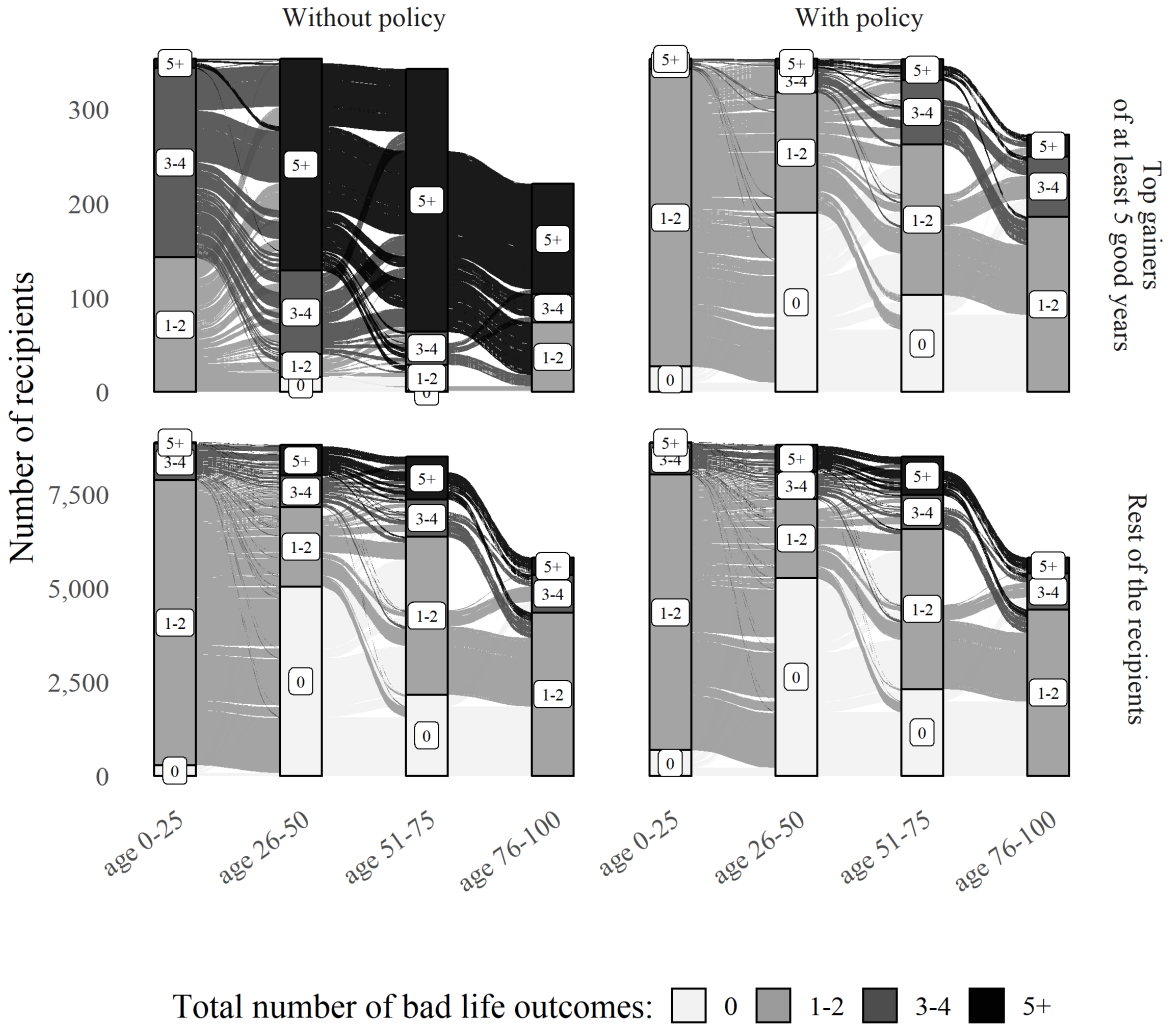


Figure G.5: Clustering and Compounding of Bad Life Outcomes Among the Top Gainers and Rest of the Child Recipients, With and Without the Policy

Note: Each individual is represented by a path – alluvium – that goes through the bar-regions representing the number of bad life outcomes in each twenty-five-year period.

Without the policy, bad life outcomes cluster more densely among the top gainers (top panel) compared to the rest of recipients (bottom panel). Most top gainers have experienced at least three bad life outcomes up to age 25, and at least five bad life outcomes after age 25. In comparison, most other recipients experience two bad life outcomes at most during any of the life periods, and most experience no bad life outcomes during ages 26-50. With the policy, the degree of clustering among the top gainers is substantially reduced, nearly to the level of other recipients.

To find out what childhood characteristics best predict lifetime wellbeing policy gains, we regress the various childhood circumstance and their interactions on the good years gained. We include

the following childhood circumstances: SDQ conduct problem and impact scores, indicators for male, high conduct problems (SDQ conduct problem score 7+), childhood poverty (household income below 60% median income), parent with a degree (NVQ level 4+), parental mental illness (parental depression assessed using Rutter malaise inventory modified 9 item score), and all possible 3-way and 2-way interactions of high conduct problems, parental degree and poverty; and also 2-way interactions between high conduct problems and poverty only. The results of these regressions can be found in Table G.8. The estimated policy benefits in the two scenarios after re-targeting the parent training programme can be found in Table G.9.

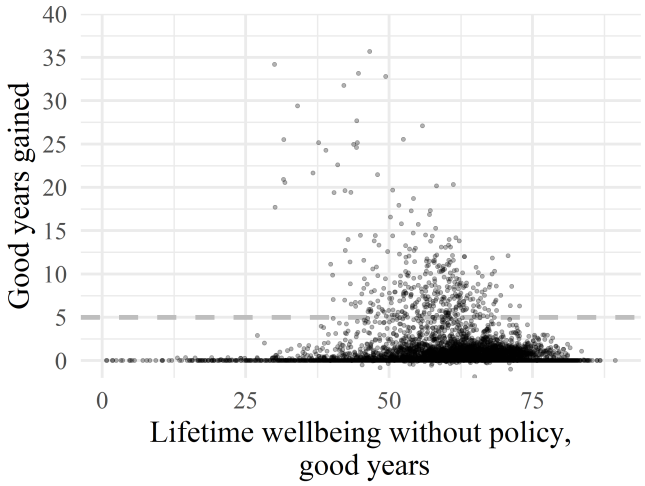


Figure G.6: Good Years Gained Versus Baseline Lifetime Wellbeing

Note: The gains represent the individual policy gains among the 9,228 child recipients.

Table G.8: The Effect of Early Conditions on Good Years Gained

	(i)		(ii)		(iii)		(iv)	
Male: M	0.198	(0.049)	0.200	(0.048)	0.197	(0.048)	0.190	(0.048)
In poverty: POV	0.118	(0.050)	0.102	(0.049)	0.107	(0.054)	0.082	(0.050)
Parental degree: PDGR	-0.009	(0.062)	-0.006	(0.062)	-0.036	(0.071)	-0.002	(0.062)
Parental mental illness: PMI	-0.028	(0.055)	-0.009	(0.053)	-0.133	(0.054)	0.010	(0.054)
SDQ conduct problem score: 5	0.146	(0.058)						
SDQ conduct problem score: 6	-0.198	(0.081)						
SDQ conduct problem score: 7	0.588	(0.116)						
SDQ conduct problem score: 8	0.663	(0.203)						
SDQ conduct problem score			0.021	(0.035)	0.048	(0.035)	0.023	(0.035)
SDQ impact score: 1	-0.078	(0.069)						
SDQ impact score: 2	0.196	(0.127)						
SDQ impact score: 3	0.193	(0.124)						
SDQ impact score: 4	-0.684	(0.219)						
SDQ impact score: 5	0.084	(0.126)						
SDQ impact score: 6	-1.177	(0.246)						
SDQ impact score: 7	-0.521	(0.306)						
SDQ impact score: 8	0.018	(0.441)						
SDQ impact score			-0.053	(0.017)	-0.091	(0.017)	-0.054	(0.017)
Cognitive skills	-0.324	(0.202)	-0.203	(0.194)	-0.421	(0.193)	-0.196	(0.195)
High conduct problems: HCP			0.455	(0.125)	0.516	(0.230)	0.194	(0.193)
POV × PDGR					-0.439	(0.142)		
POV × HCP					-0.476	(0.241)	0.358	(0.203)
PDGR × HCP					-0.807	(0.348)		
POV × PDGR × HCP					6.278	(0.516)		
Observations	9,228		9,228		9,228		9,228	

Note: Estimated coefficients and standard errors in parentheses, from a linear OLS regression of good years gained as dependent variable and the following independent variables: SDQ conduct problem and impact scores (indicators for levels in (i) and total score in (ii)-(iv)), indicators for male (M), high conduct problems (HCP), childhood poverty (POV), parent with a degree (PDGR), parental mental illness (PMI), and in (iii) all 3-way and 2-way interactions of high conduct problems, parental degree and poverty; in (iv) the 2-way interactions between high conduct problems and poverty only. The variables dropped out due to collinearity are not included in the table.

Table G.9: Average Policy Benefits for the Recipient Children After Re-Targeting

	Scenario 2 (n=494)			Scenario 3 (n=42)		
Outcome	Baseline	Gain	SE	Baseline	Gain	SE
Child outcomes						
Conduct disorder at age 5, %	44.53	0.00	0.00	57.14	0.00	0.00
Conduct disorder at age 18, %	32.11	-11.38	1.43	50.00	-40.48	7.67
SDQ conduct problem score at age 5	7.25	-0.82	0.01	7.00	-1.13	0.00
SDQ conduct problem score at age 18	6.14	-0.82	0.01	5.00	-1.13	0.00
SDQ impact score at age 5	2.23	-0.18	0.01	5.00	-0.45	0.00
SDQ impact score at age 18	1.61	-0.16	0.01	9.00	-0.45	0.00
Cognitive skills at age 5 (standardised)	0.89	0.00	0.00	1.05	0.00	0.00
Cognitive skills at age 18 (standardised)	0.84	0.00	0.00	0.95	0.00	0.00
Adult outcomes						
University graduates, %	17.68	0.20	0.20	28.57	0.00	0.00
Working years in unemployment, %	16.40	-2.57	0.29	15.62	-7.48	1.32
Life years in poverty, %	56.83	-2.22	0.24	51.74	-7.27	1.30
Working years in prison, %	5.82	-2.11	0.30	8.86	-7.10	1.41
Retirement years in residential care, %	7.99	-1.37	0.22	8.47	-5.47	1.48
Adult years as a smoker, %	27.15	-3.69	0.48	23.56	-12.78	2.68
Adult years with CHD, %	6.57	0.07	0.03	7.34	0.51	0.26
Life years with mental illness, %	20.34	-3.45	0.36	24.96	-13.66	2.16
Years of life	77.77	0.37	0.14	76.93	2.12	1.20
Premature mortality rate (before age 75), %	30.57	-1.21	0.49	28.57	-4.76	3.33
Annual earnings (lifetime average), £	27,396	548.92	67.58	30,516	2,033.84	446.88
Annual savings (lifetime average), £	2,516	89.46	11.04	2,847	319.59	69.40
Annual interest (lifetime average), £	163	31.62	3.68	215	139.54	25.98
Final wellbeing outcomes						
Annual consumption (lifetime average), £	17,512	433.71	48.13	19,044	1,681.12	302.46
Healthy years	63.90	0.89	0.13	63.51	4.15	0.98
Healthy years (discounted)	38.89	0.38	0.05	39.00	1.75	0.37
Good years	56.91	1.27	0.14	57.97	5.45	1.00
Good years (discounted)	34.67	0.57	0.06	35.48	2.41	0.41

Note: The effects are calculated on average per child recipient. The gain is the average difference between the level of an outcome with and without the programme, where the averages are calculated by aggregating the individual outcomes over time and then averaging them across child-recipients. In scenario 2 parent training is delivered to parents of all 5 year old children with SDQ conduct problem score$\geq$ 7, who are also in poor households; in scenario 3 it is further restricted to parents of such children who also have a university degree (NVQ level 4 or above). SE – standard error; CHD – coronary heart disease; SDQ conduct problem score ranges 0-10 with a higher value representing more conduct problems; cognitive skills measure is a common factor extracted from the cognitive skills measures disseminated by MCS, with a higher value representing better skills, standardised with a mean of 1.00 and standard deviation of 0.15. We use year 2015/16 prices and the annual discount rate of 1.5%.

Appendix H Further Distributional Analyses

We look at the gap in expected lifetime wellbeing between the best off and worst off groups of children, following an *ex ante* approach to evaluating inequality based on the distribution of expected outcomes predicted on the basis of early childhood circumstances.[4] Our "extreme best off group" focuses on individuals in the top category of all four main markers of social disadvantage in early life (top 20% parental income, high parental education, no parental mental illness, high baseline conduct problems). Our "best off 20% group" focuses on the best off 20% of individuals in terms of predicted lifetime wellbeing based on all four main markers of social disadvantage in early life.

Table H.10 summarises the differences in lifetime expected wellbeing between these best off and worst off groups. Figure H.7 shows the baseline good years and policy gains for all of these groups. The intervention reduces inequality between the best off 20% and worst off 20% children by 0.12 good years per child, and extreme worst off and best off children by 2.52 good years per child.[5]

Table H.10: Whole Cohort Lifetime Inequality Impacts by Childhood Circumstance

Childhood circumstance	Number of children	Annual consumption, £		Lifetime health, healthy years		Lifetime wellbeing, good years	
		Baseline	Gain	Baseline	Gain	Baseline	Gain
Best off 20%	20,000	32,559	3	68.71	0.02	69.59	0.03
Worst off 20%	20,000	18,471	62	66.31	0.10	59.84	0.15
Difference		14,088	-58	2.40[6]	-0.08	9.76	-0.12
Extreme best off	12,149	32,909	3	68.81	0.02	69.83	0.02
Extreme worst off	26	16,808	914	62.16	1.78	54.51	2.55
Difference		16,101	-910	6.66	-1.76	15.32	-2.52

Note: The average policy gains per cohort member for the subgroups of the simulated cohort of 100,000 individuals.

[4] It would be possible to follow an *ex post* approach based on the distribution of realised lifetime outcomes after death, but we leave that as a future exercise.

[5] The number of children in the extreme worst off group is very small, so this number should be interpreted with caution.

[6] Standard estimates of gaps in healthy life expectancy by current socioeconomic status are substantially larger than our estimate of gaps by childhood circumstance, due to dynamic interdependence between health and social status over the lifecourse.

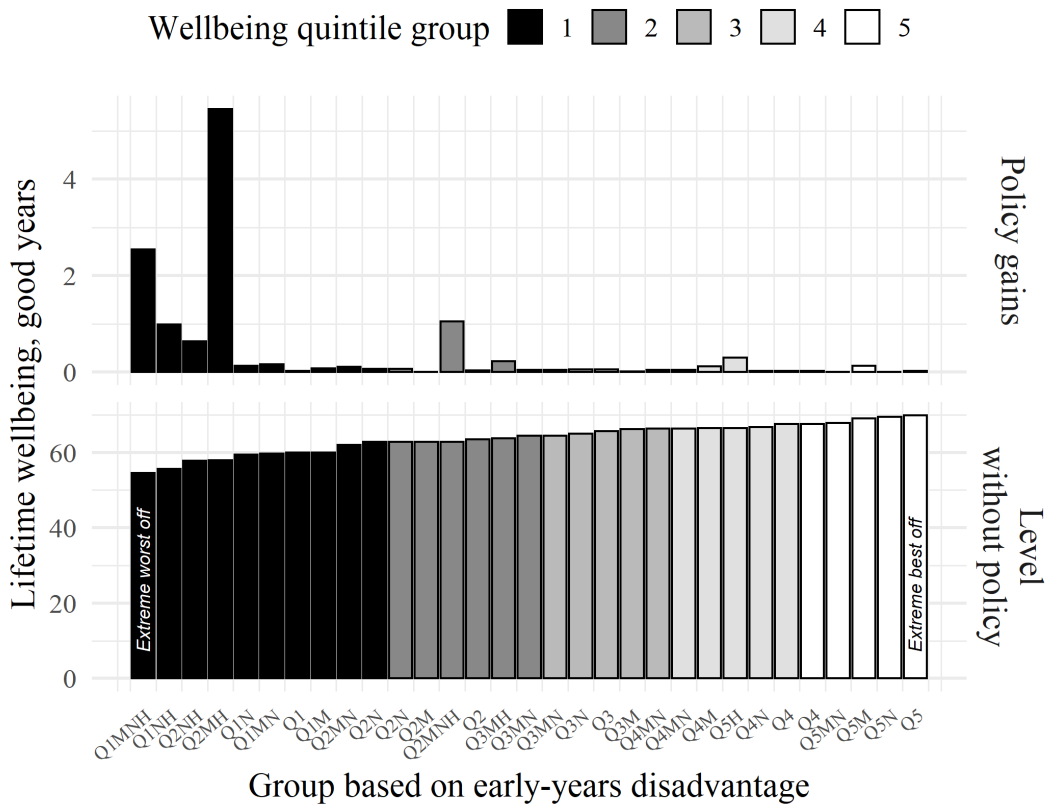


Figure H.7: Change in Wellbeing Distribution for Groups Based on Early Life Social Disadvantage

Note: The groups are based on combinations of markers of social disadvantage in early life, as indicated on the horizontal axis: parental income quintile group – Q(1-5), parental mental illness – M, no parental university degree – N, high baseline conduct problems – H; the groups are then ranked from worst to best in terms of lifetime wellbeing. The "extreme worst off group" are individuals in the bottom category of all four main markers of social disadvantage (bottom 20% parental income, no parental degree, parental mental illness, high baseline conduct problems). "The worst off 20% group" or the wellbeing quintile group 1 (in the darkest shade) are the worst off 20% of individuals in terms of predicted lifetime wellbeing, based on all four markers of social disadvantage. Because different idividuals in the the groups Q2N, Q3MN, Q4MN and Q4 can fall into different lifetime wellbeing quintiles, these groups are represented by two bars.

We also illustrate the use of Prioritarian analysis using a summary measure of inequality in life chances, based on the Atkinson index of inequality in lifetime wellbeing. This index represents the proportion of total good years that decision maker would be willing to give up, in order to have more equal shares of wellbeing in the society. The inequality aversion parameter, denoted by ε, is a normative parameter for which a higher value implies that the decision maker is willing to give up more in exchange for greater equality. We calculate the Atkinson inequality index using two values of ε, 1 and 10 (Robson et al., 2017). For each of these quantities, we

also calculate the equally distributed equivalent, which is the (lower) average level of wellbeing considered acceptable to achieve full equality. Higher inequality aversion implies a lower equally distributed equivalent.

Table H.11 shows the calculated Atkinson indices and equally distributed equivalents of wellbeing, first at baseline, then assuming the three differently targeted policies; the table also shows the change (difference) in the measure as result of each policy compared to the baseline. At baseline the Atkinson index is around 0.15% with ε equal to 1 and around 1.50% with ε equal to 10.[7] The latter coefficient implies that decision maker is willing to exchange 1.5% of good years (i.e. reduce average lifetime wellbeing from 64.94 good years to 63.97 good years) for a fully equal distribution of lifetime wellbeing.

Table H.11: Inequality Impact Assessment Using the Atkinson Inequality Index

	Atkinson index, %		Lifetime wellbeing, good years	Equally distributed equivalent of lifetime wellbeing, good years	
	$\varepsilon = 1$	$\varepsilon = 10$		$\varepsilon = 1$	$\varepsilon = 10$
			Without policy (baseline)		
	0.147	1.504	64.943	64.848	63.967
			With policy		
Scenario 1	0.143	1.460	65.005	64.912	64.056
Difference	-0.004	-0.044	0.062	0.064	0.090
Scenario 2	0.146	1.487	64.950	64.855	63.984
Difference	-0.001	-0.017	0.006	0.007	0.017
Scenario 3	0.147	1.501	64.946	64.850	63.971
Difference	0.000	-0.003	0.002	0.002	0.004

Note: ε denotes the inequality aversion parameter. The equally distributed equivalent of lifetime wellbeing is calculated as Wellbeing$\times$(1$-$Atkinson index). In scenario 1, parent training is delivered to parents of all 5 year old children with SDQ conduct problem score$\geq$ 4; in scenario 2, it is restricted to parents in poor households of all 5 year old children with SDQ conduct problem score$\geq$ 7; in scenario 3, it is further restricted to parents of such children who also have a university degree.

The basic full-scale parent-training programme (scenario 1) clearly reduces inequality: the Atkinson index decreases by 0.004 percentage-points when assuming ε equal to 1 and 0.044

[7] As expected, the Atkinson index is always higher when assuming $\varepsilon = 10$ as opposed to $\varepsilon = 1$, as a higher inequality aversion parameter suggests that decision maker cares more about inequality and are therefore is willing to give up more of the wellbeing for its sake.

percentage-points when assuming ε equal to 10.

The two re-targeted policies also reduce inequality, but by a smaller amount. More specifically, with the re-targeted policy in scenario 2, the reduction in the Atkinson index is 0.001 percentage-points with ε equal to 1 and 0.017 percentage-points with ε equal to 10. In scenario 3, the effect is even smaller.

Appendix I Comparison With Other Economic Evaluations of Parent Training

Table I.12: Short-Term Cost-Effectiveness Analyses

Study (Type, setting and duration)	Method	Findings
Edwards et al. (2007) (IY, UK, 6 months)	Trial-based CEA using a single RCT, primary effect Eyberg Child Inventory intensity (ECBI-I) score, public costs included primary care, hospital, special education, social services.	CE ratio of £92 per point improvement in the ECBI-I score; at a WTP of £127 per point increase in ECBI-I score, the probability of cost-effectiveness is 84%.
O'Neill et al. (2013) (IY, Ireland, 6 months)	Trial-based CEA using a single RCT, primary effect Eyberg Child Inventory intensity (ECBI-I) score, public costs included primary care, hospital, special education, social services.	CE ratio of £69 per one point improvement in ECBI-I scores; at a WTP of £110 per point increase in ECBI-I score, the probability of cost-effectiveness is 90%.
Edwards et al. (2016) (IY, UK, 18 months)	Trial-based CEA using a single RCT, primary effects Eyberg Child Inventory intensity (ECBI-I) score and SDQ score, costs included primary care, hospital, special education, social services.	CE ratio of £300 per point improvement in the ECBI-I score and £1,423 per point improvement in the total SDQ score; at a WTP of £2,747 per point increase in SDQ score, the probability of cost-effectiveness is 82%.
Gardner et al. (2017) (IY, UK and Ireland, 3-12 months)	Trial-based CEA, using seemingly unrelated regression to pool cost data from 5 RCTs, costs included primary care, hospital, mental health, special education, social services, accommodation, voluntary sector.	At a WTP of £109 per point improvement on the ECBI-I, the probability of cost-effectiveness is 50%; this increases to 99% at a WTP of £145; higher probability of effectiveness for subgroups with high baseline conduct problems and boys.
Nystrand et al. (2019) (IY, Sweden, up to age 18)	Markov-model based CEA based on a single RCT, primary effects for estimating DALYs were reduction in conduct and ADHD problems, cost savings restricted to health care and education sector.	CE ratio of £9,814 per DALY (CE threshold in Sweden £52,339 per averted DALY).
Scavenius et al. (2020) (Caring in Chaos volunteer-delivered parental training programme, Denmark, 4 months)	Trial-based CEA based on a single RCT, primary effects Parenting Sense of Competence Scale (PSOC) and child functioning (Home Situations Questionnaire, HSQ); no cost savings as only a 4-month time horizon.	CE ratio of £1,117 per effect size improvement in the PSOC; at a WTP of £1,459 or £3,650 per SD gain in PSOC or HSQ, respectively, the probability of cost-effectiveness is 90%.

Note: All monetary values inflated up to 2015/16 prices. IY – Incredible Years, CE – cost effectiveness, CEA – cost effectiveness analysis, RCT – randomised controlled trial, WTP – willingness to pay.

Table I.13: Long-Term Cost-Benefit Analyses

Study (Type, setting and duration)	Method	Findings
Bonin et al. (2011) (Generic parenting programme, England, up to age 25)	Markov-model based CBA estimating reduction in CD incidence and public cost savings (NHS, social services, education, voluntary sector, criminal justice system, health impacts of crime, benefit payments).	Payback period 5-8 years in base-case scenario (up to 12 years in the worst case scenario); total savings £18,345 per family, with intervention cost range of £1,063-£2,321.
O'Neill et al. (2013) (IY, Ireland, up to age 30)	CBA study based on simple linear aggregate-level extrapolation and looking at public cost savings in education, crime and unemployment.	Net present value (public cost savings minus the upfront costs, and discounted) £2,346/child.
Gardner et al. (2017) (IY, UK and Ireland, up to age 30)	Markov-model based cost saving analysis based on the same model as Bonin et al. (2011) and using a meta-analysis of multiple RCTs. Public cost savings included NHS, social services, education, voluntary sector, criminal justice system, health impacts of crime and benefit payments.	In the worst-case scenario, total savings of £1023-£7565. In the best-case scenario, total savings of £1254 - £9408. The highest savings were from crime and education.
Washington State Institute for Public Policy (2019) (IY, USA, up to age 50)	Model-based CBA based on meta-analysis of multiple RCTs and aggregate-level modelling of multiple costs and benefits (labour market earnings, criminal justice system, education and special education, healthcare).	Payback period 20 years; benefit-cost ratio 5.65; benefits minus cost per child £4,751 (largest benefits were labour market earnings).

Note: All monetary values inflated up to 2015/16 prices. IY - Incredible Years, CBA - cost benefit analysis, RCT - randomised controlled trial.

References

Asaria, Miqdad. 2017. "Health Care Costs in the English NHS: Reference Tables for Average Annual NHS Spend by Age, Sex and Deprivation Group." Centre for Health Economics at the University of York Research Paper 147.

Bonin, Eva-Maria, Madeleine Stevens, Jennifer Beecham, Sarah Byford, and Michael Parsonage. 2011. "Costs and Longer-Term Savings of Parenting Programmes for the Prevention of Persistent Conduct Disorder: A Modelling Study." *BMC Public Health*, 11(1): 803.

Claxton, Karl, Steve Martin, Marta Soares, Nigel Rice, Eldon Spackman, Sebastian Hinde, Nancy Devlin, Peter C Smith, and Mark Sculpher. 2015. "Methods for the Estimation of the National Institute for Health and Care Excellence Cost-Effectiveness Threshold." *Health technology assessment (Winchester, England)*, 19(14): 1.

Cookson, Richard, Ieva Skarda, Owen Cotton-Barrett, Matthew D Adler, Miqdad Asaria, and Toby Ord. 2020. "Quality Adjusted Life Years Based on Health and Consumption: A Summary Wellbeing Measure for Cross-Sectoral Economic Evaluation." *Health Economics*, 30(1): 70–85.

Curtis, Lesley A, and Amanda Burns. 2017. "Unit Costs of Health and Social Care 2017." Canterbury, United Kingdom: Personal Social Services Research Unit, University of Kent. Retrieved 26 February 2020, from `https://www.pssru.ac.uk/project-pages/unit-costs/unit-costs-2017/`.

Dubourg, Richard, Joe Hamed, Jamie Thorns, et al. 2005. "The Economic and Social Costs of Crime Against Individuals and Households 2003/04." Home Office Online Report 30/05. Retrieved 26 February 2020, from =http://library.college.police.uk/docs/hordsolr/rdsolr3005.pdf.

Edwards, Rhiannon T, Alan Céilleachair, Tracey Bywater, Dyfrig A Hughes, and Judy Hutchings. 2007. "Parenting Programme for Parents of Children at Risk of Developing Conduct Cisorder: Cost Effectiveness Analysis." *British Medical Journal*, 334(7595): 682.

Edwards, Rhiannon Tudor, Carys Jones, Vashti Berry, Joanna Charles, Pat Linck, Tracey Bywater, and Judy Hutchings. 2016. "Incredible Years Parenting Programme: Cost-Effectiveness and Implementation." *Journal of Children's Services*, 11(1): 54–72.

Feinstein, Leon, Haroon Chowdry, and Kirsten Asmussen. 2017. "On Estimating the Fiscal Benefits of Early Intervention." *National Institute Economic Review*, 240(1): R15–R29.

Frijters, Paul, and Christian Krekel. 2021. *A Handbook of Well Being Decision-Making in the UK: History, Measurement, Theory, Implementation, and Examples.* Oxford University Press.

Gardner, Frances, Patty Leijten, Joanna Mann, Sabine Landau, Victoria Harris, Jennifer Beecham, Eva-Maria Bonin, Judy Hutchings, and Stephen Scott. 2017. "Could Scale-Up of Parenting Programmes Improve Child Disruptive Behaviour and Reduce Social Inequalities? Using Individual Participant Data Meta-Analysis to Establish for Whom Programmes Are Effective and Cost-Effective." *Public Health Research*, 5(10). Retrieved 11 October 2021, from `https://doi.org/10.3310/phr05100`.

HM Treasury. 2020. "The Green Book: Appraisal and Evaluation in Central Government." Retrieved 25 February 2022, from `https://www.gov.uk/government/publications/the-green-book-appraisal-and-evaluation-in-central-governent`.

HM Treasury. 2021. "Green Book Supplementary Guidance: Wellbeing." Retrieved 25 February 2022, from `https://www.gov.uk/government/publications/green-book-supplementary-guidance-wellbeing`.

Liu, Joseph L Y, Nikolaos Maniadakis, Andrew Gray, and Mike Rayner. 2002. "The Economic Burden of Coronary Heart Disease in the UK." *Heart*, 88(6): 597–603.

Love-Koh, James, Miqdad Asaria, Richard Cookson, and Susan Griffin. 2015. "The Social Distribution of Health: Estimating Quality-Adjusted Life Expectancy in England." *Value in Health*, 18(5): 655–662.

McCrone, Paul R, Sujith Dhanasiri, Anita Patel, Martin Knapp, and Simon Lawton-Smith. 2008. *Paying the Price: The Cost of Mental Health Care in England to 2026.* London, United Kingdom: King's Fund.

Nystrand, Camilla, Inna Feldman, Pia Enebrink, and Filipa Sampaio. 2019. "Cost-Effectiveness Analysis of Parenting Interventions for the Prevention of Behaviour Problems in Children." *PLOS One*, 14(12): e0225503.

O'Neill, Donal, Sinéad McGilloway, Michael Donnelly, Tracey Bywater, and Paul Kelly. 2013. "A Cost-Effectiveness Analysis of the Incredible Years Parenting Programme in Reducing Childhood Health Inequalities." *The European Journal of Health Economics*, 14(1): 85–94.

Paulden, Mike, and Karl Claxton. 2012. "Budget Allocation and the Revealed Social Rate of Time Preference for Health." *Health Economics*, 5(21): 612–618.

Robson, Matthew, Miqdad Asaria, Richard Cookson, Aki Tsuchiya, and Shehzad Ali. 2017. "Eliciting the Level of Health Inequality Aversion in England." *Health Economics*, 26(10): 1328–1334.

Romeo, Renee, Martin Knapp, and Stephen Scott. 2006. "Economic Cost of Severe Antisocial Behaviour in Children-And Who Pays It." *The British Journal of Psychiatry*, 188(6): 547–553.

Scott, Stephen, Jackie Briskman, and Thomas G O'Connor. 2014. "Early Prevention of Antisocial Personality: Long-Term Follow-Up of Two Randomized Controlled Trials Comparing Indicated and Selective Approaches." *American Journal of Psychiatry*, 171(6): 649–657.

Scott, Stephen, Martin Knapp, Juliet Henderson, and Barbara Maughan. 2001. "Financial Cost of Social Exclusion: Follow up Study of Antisocial Children into Adulthood." *British Medical Journal*, 323(7306): 191.

Sullivan, Patrick W, Julia F Slejko, Mark J Sculpher, and Vahram Ghushchyan. 2011. "Catalogue of EQ-5D Scores for the United Kingdom." *Medical Decision Making*, 31(6): 800–804.

Van Aar, Jolien, Patty Leijten, Bram Orobio de Castro, and Geertjan Overbeek. 2017. "Sustained, Fade-Out or Sleeper Effects? A Systematic Review and Meta-Analysis of Parenting Interventions for Disruptive Child Behavior." *Clinical Psychology Review*, 51: 153–163.

Washington State Institute for Public Policy. 2019. "Benefit-Cost Results. Incredible Years Parent Training, Children's Mental Health: Disruptive Behavior." Retrieved 25 February 2022, from `http://www.wsipp.wa.gov/BenefitCost/ProgramPdf/158/Incredible-Years-Parent-Training`.