



LifeSim: A Lifecourse Dynamic Microsimulation Model of the Millennium Birth Cohort in England

Ieva Skarda^{1*}, Miqdad Asaria², Richard Cookson¹

¹Centre for Health Economics, University of York, Heslington, York, YO10 5DD, UK;

²Department of Health Policy, Cowdray House, London School of Economics and Political Science, Houghton Street, London WC2A 2AE, UK

Abstract We present a dynamic microsimulation model for childhood policy analysis that models developmental, economic, social and health outcomes from birth to death for each child in the Millennium Birth Cohort (MCS) in England, together with public costs and a summary wellbeing measure. The model is a discrete event simulation in discrete time (annual periods), implemented in R, which progresses 100,000 individuals through each year of their lives from birth in the year 2000 to death. From age 0 to 18 the model draws observational data from the MCS, with explicit modelling of only a few derived outcomes (mental health, conduct disorder, mortality, health-related quality of life, public costs and a general wellbeing metric). During adulthood, all outcomes are modelled dynamically using explicit networks of stochastic process equations, with separate networks for working age and retirement. Our equations are parameterised using effect estimates from existing studies combined with target outcome levels from up-to-date administrative and survey data. We present our baseline projections and a simple validation check against external data from the British Cohort Study 1970 and Understanding Society survey.

DOI: <https://doi.org/10.34196/ijm.00228>

1. Introduction

Recent scientific advances have established beyond reasonable doubt that childhood programmes can have important effects on health and wellbeing many decades in the future, during working years and retirement (*Conti et al., 2019; Heckman, 2012*). Policy makers want quantitative information about these long-term effects, and they also want information about distributional impacts on inequality in lifetime health and wellbeing, as well as inequality in current period health and wellbeing. While childhood policy analysis using randomised control trials and quasi-experiments is the gold standard in establishing cause and effect relationships, this is rarely possible when quantifying lifetime policy effects over many decades. Even when long-term follow-up data is available, such analysis yields insights about historical cohorts born many decades ago with questionable relevance to the current childhood policy context. Microsimulation offers a forward-looking alternative for childhood policy analysis, as it can extrapolate long-term outcomes for cohorts living in the present and project the effects of the policies that policymakers are considering today.

In this paper we introduce a dynamic childhood policy microsimulation model “LifeSim” which models the co-evolution of economic, social and health outcomes from birth to death for each child in a general population birth cohort of 100,000 English children born in year 2000-1. In addition to modelling the individual outcomes, LifeSim also models the associated costs and savings to the public budget.

*For correspondence:
ieva.skarda@york.ac.uk

©This article is distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use and redistribution provided that the original author and source are credited.

Author Keywords: childhood, conduct problems, inequality, lifecourse, policy evaluation, simulation, skills, well-being
© 2021, Skarda et al.

The chosen life outcomes and the structure of our model are designed to address cross-sectoral childhood policy concerns and to align with the large body of theory and knowledge about human capital formation in childhood and later life economic and health outcomes. From age 0 to 18 we heavily rely on observed life outcomes from the Millennium Cohort Study (MCS), and only explicitly model three specific childhood outcomes - mental health, conduct disorder and mortality - which are then combined with MCS data to estimate public costs and a general wellbeing metric. During adulthood, however, we specify explicit networks of stochastic processes, with different networks for working years and retirement, and parameterise these using estimates from published studies of longitudinal data on earlier cohorts.

LifeSim has the following distinct features:

1. It jointly models the co-evolution of many economic, social and health outcomes, capturing how outcomes in multiple domains interact, compound and cluster over time, emphasising how early-life disadvantages can compound over life creating a spiral of multiple disadvantage;
2. It simulates long-run outcomes for a whole general population cohort of children, not just one specific subpopulation of trial participants, which allows the model to serve as a platform for many different kinds of informative policy analysis, including optimal policy targeting analysis, population-wide distributional impact analysis and assessment of the opportunity costs falling on the individuals not directly affected by the intervention;
3. It simulates individual-level outcomes for each heterogeneous child in the cohort, instead of only producing average-level outcomes, allowing us to produce multidimensional individual wellbeing measures, which have been discussed in the literature and have well-known advantages over unweighted cost-benefit analysis (**Adler and Fleurbaey, 2016**);
4. It simulates outcomes over the whole lifecourse from birth to death, enabling policy analysis to adopt a broad lifetime perspective.

We capture all of these features by combining many different sources of data, which requires strong assumptions. We make all of our assumptions explicit and subject to scrutiny by providing carefully labelled and fully referenced details of all modelling equations, parameters and data sources in the appendix, and by publishing our complete programming code. We use longitudinal data on children born in 2000 as our primary data source but supplement this with other sources of data including more up-to-date cross-sectional administrative and survey data as well as older sources of longitudinal data on children born in earlier decades. In choosing how many assumptions to make and how many sources of data to use, there are trade-offs between internal and external validity.¹ Using a single source of experimental data with long-term follow-up over many decades would maximise internal validity, but is only possible for backward-looking evaluation of policy experiments many decades ago. Using assumptions and multiple sources of data is necessary to achieve external validity for forward-looking economic appraisal of current policy options in the current policy environment.

To our knowledge, LifeSim is the first microsimulation model that provides information on many developmental, economic, social, health, and public cost outcomes from birth to death for each individual in a birth cohort. In the economics literature, there are dynamic microsimulation models of many co-evolving economic and social outcomes across the life-cycle (e.g. LINDA, a rational agent dynamic microsimulation based on dynamic programming, **Van der Ven (2016)**) and dynamic microsimulation models of childhood development (e.g. MELC, a discrete event simulation from age 0 to 13, **Milne et al. (2015)**). And in the health literature there are dynamic microsimulation models of multiple co-evolving health and public cost outcomes (e.g. HealthPaths, **Wolfson and Rowe (2014)**, POHEM, **Hennessy et al. (2015)** and IMPACT NCD, **Kypridemos et al. (2016)**). However, none of these cover developmental, economic, social, and health outcome domains and few provide information on the whole lifecourse from birth to death.² Modelling the entire lifecourse allows us to examine how child-

1. Internal validity relates to claims about cause and effect within the study population, whereas external validity relates to how applicable the findings are to real world policy settings.

2. According to Statistics Canada, they developed a dynamic microsimulation model in the 1990s with a rich set of co-evolving economic, social and health outcomes, called LifePaths (**Spelauer et al., 2013**), which has subsequently been discontinued. However, this model seems to have had limited detail on developmental outcomes in childhood and we could not find detailed technical information or any published economic evaluations based on it.

hood outcomes can lead to spirals of advantage and disadvantage in later life, whereby economic, social, and health outcomes interact, compound and cluster over time. For example, a young child with poor cognitive and social skills is at heightened risk of multiple adverse outcomes as they grow older – including unhealthy behaviour, mental illness, unemployment, low earnings imprisonment and physical illness – all of which can interact and compound in a spiral of disadvantage (**Zucchelli et al., 2012; Layard et al., 2014; Frijters et al., 2017**). Modelling this also provides a platform for more informative long-term economic evaluation, targeting analysis and distributional analysis of childhood policies from a lifetime perspective, as we illustrate in a companion paper under review elsewhere (**Skarda et al., 2021**).

2. Methods

2.1. Model Structure

Our microsimulation model is a discrete event simulation in discrete time (annual periods), which progresses 100,000 individuals through each year of their lives from birth in the year 2000 to death. From ages 0 to 18 it closely follows observed Millennium Cohort Study (MCS) data, and thereafter predicts the annual evolution of each life outcome based on the current values of relevant

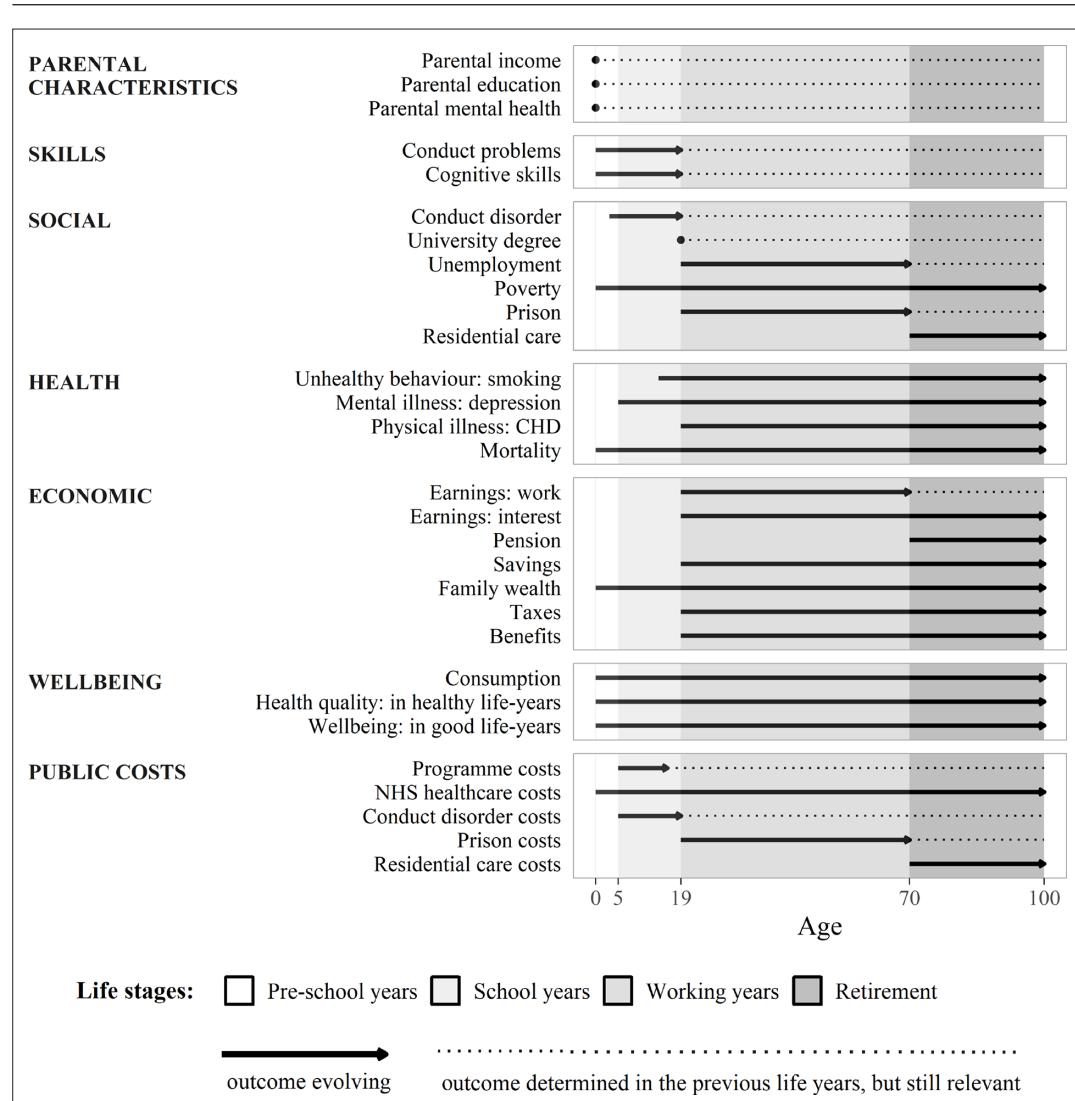


Figure 1. Overview of Key Outcomes Over the Lifecourse

characteristics and outcomes, which in turn depend on lagged values.³ This kind of model can be seen as a pragmatic compromise between a simpler Markov model structure, which has no "memory" or dependence upon lagged values, and a more complicated agent-based model structure, which explicitly models interactions between individuals and how individual behaviour may depend upon the macro-level policy environment as well as the behaviour of others. Allowing dependence upon lagged values allows a rich analysis of the dynamic clustering and compounding of multiple outcomes over time, while setting aside agent-based interactions keeps the model tractable, even when modelling a relatively large number of outcomes.

The model links together a diverse set of individual-level life outcomes of interest to policymakers (**Figure 1**). By using rich observational data from the MCS, our model provides information on various aspects of human capital development in childhood - including social skills, cognitive skills, and health behaviour (teenage smoking) - and then extrapolates later life outcomes across economic, social, and health domains for the rest of the lifecourse. For simplicity and concreteness we focus on one important and readily measurable dimension of social skills - conduct problems - as proxied by two separate parent reported measures. Child conduct is related to self-control and regulation, which have been shown to matter in many aspects of life, including wellbeing, income, employment, crime and health outcomes (**Goodman et al., 2015**). We also model mental illness and health-related quality of life during childhood, using external datasets (Mental Health of Children and Young People Great Britain, and a dataset by **Love-Koh et al. (2015)**).

Let $i = 1,..100,000$ index the individual children in the cohort. Let yearly time periods also corresponding to the age of children be indexed as $age = 0, 1,..18..T$ where $age = 18$ marks the end of childhood, and T is the last time period in which there are any cohort members still alive (which we assume to be 100, since small number problems make predictions decreasingly reliable at older ages). Let X_i be the vector of initial conditions assumed to be constant for child (e.g. individual and family characteristics at birth or other early time period – if data at birth on the condition is not available); let $\theta_{i,age}$ be an age-specific vector of child Strengths and Difficulties Questionnaire (SDQ) scores – multi-dimensional parent-reported score on child's difficulties, $cd_{i,age}$ – an age-specific outcome of whether child develops a conduct disorder, and $cog_{i,age}$ – an age-specific child's cognitive skills measure. Finally, let $Y_{i,age}$ be the age-specific vector of life-cycle outcomes (further, outcomes) for child i . These outcomes can be further classified as social, health and economic outcomes, i.e. $Y_{i,age} \equiv \{S_{i,age}, H_{i,age}, E_{i,age}\}$, where $S_{i,age}, H_{i,age}, E_{i,age}$ are the vectors of social, health and economic outcomes respectively. It is allowed for the vector X_i to also contain elements of $\{S_i, H_i, E_i\}$.

At each age the individual probability of dying $pr.dead_{i,age}$ is modelled and defined over the closed interval from zero to one, i.e. $pr.dead_{i,age}(E_{i,age}, S_{i,age}, H_{i,age}) \in [0, 1]$, which then determines the discrete outcome $dead_{i,age}$ – whether the individual at a certain age is dead ($dead_{i,age} = 1$) or alive ($dead_{i,age} = 0$). More specifically, we can represent the outcome 'dead or alive' by a function $l(\cdot)$ such that if in the previous year individual was alive then they can be either alive or dead in the following year, i.e. $dead_{i,age} = l(pr.dead_{i,age}, \zeta_{i,age} | dead_{i,age-1} = 0) \in \{0, 1\}$ (where $\zeta_{i,age}$ represents stochasticity); and if in the previous year individual was dead then, because death is an 'absorbing state', they can be only dead in the following year, i.e. $dead_{i,age} = l(. | dead_{i,age-1} = 1) = 1$. Individual life span is then $T_i = \sum_{age=0}^T (1 - dead_{i,age})$.

To describe the initial conditions (in vector X_i), we draw observations on each child, i.e. 100,000 times in total to represent a cohort of 100,000 individuals, using re-sampling with replacement from the initial sweeps of MCS – a longitudinal survey of English children born in 2000-2001 (see panels A-C of **Table 1**).

Similarly, we use MCS data from all of the sweeps (up to age 14) to collect data for the vector $\theta_{i,age}$ – information on child SDQ conduct problem subscale score and a further parent-reported "behavioural impact" score (see panel D of **Table 1**). Both of these scores range from 0-10, with a higher score representing more conduct problems and a more severe impact of difficulties in child's life. MCS data are reported at sweeps every 2 to 4 years, so we use the most recent MCS sweep data available to fill in the missing values in the time gaps, and for age 15-18.

3. Because the MCS data was only available up to age 14 when our model was developed, we approximate the outcomes between ages 15-18 using the MSC data for age 14; this can now be updated with the data from the latest MCS wave which has recently become available and at which children are 17 years old.

Child's characteristic / family condition	Data used in modelling	Source	Mean	SD	Min	Max
A: BASIC CHILD'S CHARACTERISTICS						
Sex	Indicator if child is male	MCS1	0.49	0.50	0	1
Teenage smoking	Indicator if child smokes at 14	MCS6	0.16	0.37	0	1
B: SOCIAL CONDITIONS						
Parental income	OECD equivalised household income after taxes and benefits, £	MCS1	32,004	19,972	1,445	128,246
Parental wealth	Parental assets, £	MCS5	3,068	19,937	0	600,000
C: PARENTAL CHARACTERISTICS						
Parental education	Indicator if parent has a university degree (NVQ 4 or above)	MCS1	0.31	0.46	0	1
Parental depression at child's birth	Indicator if Rutter malaise inventory score is 4 or above	MCS1	0.14	0.35	0	1
Parental depression severity at child's birth	9-item Rutter malaise inventory score	MCS1	1.66	1.68	0	9
Parental depression when child is 5 years old	Indicator if Kessler psychological distress scale score is 13 or above	MCS3	0.03	0.18	0	1
Parental depression severity when child is 5 years old	6-item Kessler psychological distress scale score	MCS3	3.17	3.72	0	24
D: CHILD'S SOCIAL SKILLS						
Conduct problems up to age 4	SDQ conduct problem score (see the note)	MCS2	2.87	2.02	0	10
Conduct problems, ages 5-6	SDQ conduct problem score	MCS3	1.46	1.47	0	8
Conduct problems, ages 7-10	SDQ conduct problem score	MCS4	1.32	1.50	0	9
Conduct problems, ages 11-13	SDQ conduct problem score	MCS5	1.38	1.58	0	10
Conduct problems, age 14+	SDQ conduct problem score	MCS6	1.42	1.68	0	10
Impact of problems up to age 4	SDQ impact supplement score	MCS2	0.11	0.58	0	8
Impact of problems, ages 5-6	SDQ impact supplement score	MCS3	0.13	0.63	0	8

Continued

Child's characteristic / family condition		Data used in modelling				
		Source	Mean	SD	Min	Max
Impact of problems, age 7+	SDQ impact supplement score	MCS4	0.25	0.91	0	9
E: CHILD'S COGNITIVE SKILLS						
Cognitive skills up to age 4	Various measures (see the note)	MCS2	1.02	0.14	0.58	1.44
Cognitive skills, ages 5-6	Various measures	MCS3	1.03	0.13	0.57	1.45
Cognitive skills, ages 7-10	Various measures	MCS4	1.03	0.14	0.56	1.41
Cognitive skills, ages 11-13	Various measures	MCS5	1.03	0.15	0.39	1.49
Cognitive skills, age 14+	Various measures	MCS6	1.02	0.14	0.43	1.50

Note: The analysis is for 100,000 individuals in the LifeSim cohort. MCS denotes MCS sweep 1 (6 sweeps in total). Children were 9 months old in MCS1, 3 years old in MCS2, 5 years old in MCS3, 7 years old in MCS4, 11 years old in MCS5 and 14 years old in MCS6. SDQ conduct problem and impacts scores have a scale 0-10, with a higher value representing more problems/higher impact of problems. The cognitive skills measure is an age-specific common factor extracted from the various cognitive skills measures in MCS, including the British Ability Scales II, Bracken School Readiness Assessment (for age 3), National Foundation for Educational Research Progress in Maths (for age 7), Cambridge Neuropsychological Test Automated Battery tests (for ages 11 and 14) and Applied Psychology Unit (for age 14). We extract a common factor for each age where test results are available using principal component analysis, and standardise it to be with a mean of 1.00 and standard deviation of 0.15 (following **Jones and Schoon (2008)**). Similar to the SDQ score data, we use the most recent MCS sweep data available to fill in the missing values in the time gaps, and for age 15-18.

We then use the reported SDQ score components to model whether or not a child develops conduct disorder, using a previously developed algorithm which predicts a child's probability of developing conduct disorder as a function of SDQ score components – the SDQ conduct problem score and the "behavioural impact" score. This provides a specific probability of conduct disorder based on a classification as either "possible" or "probable" (**Goodman et al., 2003; Goodman et al., 2000**).⁴ This modelled probability is then combined with a random draw from a uniform distribution over 0-1, which allows us to simulate the discrete outcome of whether or not a child develops conduct disorder. Formally, the age-specific conduct disorder outcome $cd_{i,age}$ can be represented using a function $g(\cdot)$ as:

$$cd_{i,age} = g(\theta_{i,age}, \xi_{i,age}) \quad (1)$$

where $\xi_{i,age}$ represents stochasticity.

We also use the MCS data from later sweeps (up to age 14) to build $cog_{i,age}$ – a single measure of each child's cognitive skills at each age throughout their childhood up to age 18 (see panel E of **Table 1**). More specifically, our cognitive skills measure is an age-specific common factor extracted from the cognitive skills measures available in MCS, including the British Ability Scales II (for ages 3, 5, 7, 11), Bracken School Readiness Assessment (for age 3), National Foundation for Educational Research Progress in Maths (for age 7), Cambridge Neuropsychological Test Automated Battery tests (for ages 11 and 14) and Applied Psychology Unit (for age 14). We extract a common factor for each age where test results are available using principal component analysis, and standardise it to be with a mean of 1.00 and standard deviation of 0.15 (following **Jones and Schoon (2008)**). Similar to the SDQ score data, we use the most recent MCS sweep data available to fill in the missing values in the time gaps, and for age 15-18.

During adulthood, child's SDQ scores, conduct disorder outcomes and cognitive skills are assumed to stay fixed at the level achieved by the end of childhood, i.e. $\theta_{i,age} = \theta_{i,18}$, $cd_{i,age} = cd_{i,18}$ and $cog_{i,age} = cog_{i,18}$ for age = 19.. T_i .

Over the life-cycle (age = 0..18.. T_i), the vector of other life-cycle outcomes $Y_{i,age}$ evolves as:

$$Y_{i,age} = k_{age}(Y_{i,age}, Y_{i,age-1}, \theta_{i,18}, cd_{i,18}, cog_{i,18}, X_i, \eta_{i,age}) \quad (2)$$

where $\eta_{i,age}$ represents stochasticity. It should be noted that separate outcomes in the vector $Y_{i,age}$ can depend on a subset (and not necessarily all) of the outcomes in the vectors $Y_{i,age-1}, Y_{i,age}, \theta_{i,18}, cd_{i,18}, cog_{i,18}, X_i$, which can be achieved by restricting coefficients. Also, a period-specific outcome in the vector $Y_{i,age}$ will generally not depend on itself, but can depend on other outcomes at that time period included in the vector $Y_{i,age}$.

4. More specifically, the algorithm allocates a probability of 0.61 for children with the SDQ conduct problem score of at least 5 combined with the impact score of at least 2; a probability of 0.31 for children with the conduct problem score equal to 4 (irrespective of the impact score) and a probability of 0.06 for all the other children with conduct problem scores below 4.

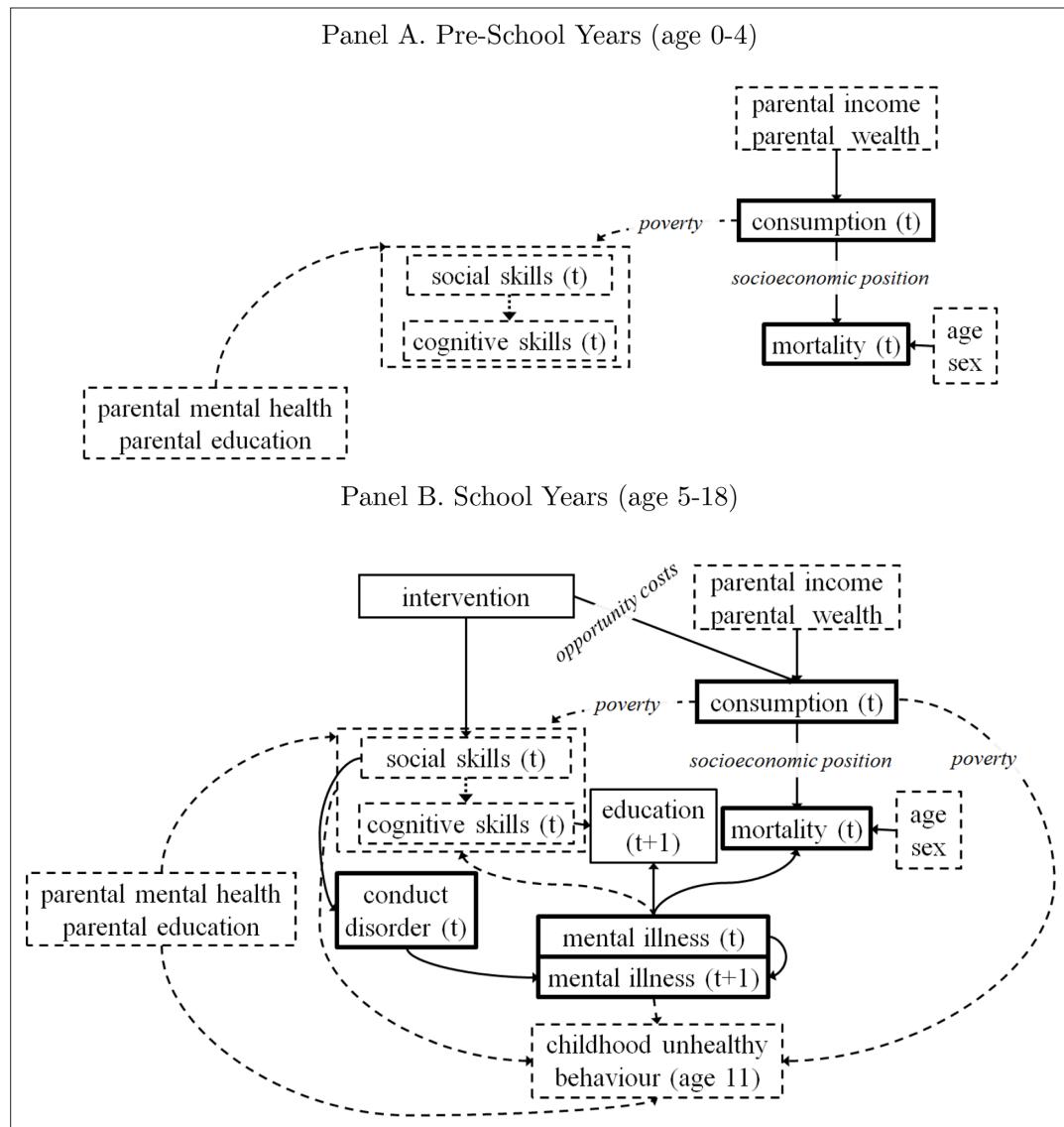


Figure 2. Model Structure for Key Life Stages in Childhood

Note: The solid arrows represent the stochastic processes that we model using modelling equations as described in Section 2.3; the dashed arrows represent implicit processes that we do not explicitly model, but that exist in the childhood dataset. The boxes represent life outcomes: the dashed boxes represent exogenous inputs into our model that are taken as given either from the childhood dataset or the previous life-stage, the thick boxes represent final outcomes that directly influence wellbeing or impose a cost to the public budget.

The model structure specified by $k_{age}(\cdot)$ changes as individuals progress through key life stages. In each life stage, the dependencies between the initial conditions and the life-course outcomes are represented by model structure diagrams in **Figures 2 and 3**, and are also summarised in **Table 2**. In the model structure diagram each solid arrow is modelled using equations (as we will explain in more detail in Section 2.3).

In choosing the model outcomes and formulating the model structure, we consulted with experts in childhood development and childhood policy, demography, epidemiology, human capital economics, and labour economics (see list of advisory group members in the acknowledgements) and were also guided by inter-disciplinary theory on human capital formation in childhood and how this influences educational attainment, earnings, physical illness, mental illness, mortality, and other outcomes with important impacts on individual wellbeing and public cost (**Almond et al., 2018; Goodman et al.,**

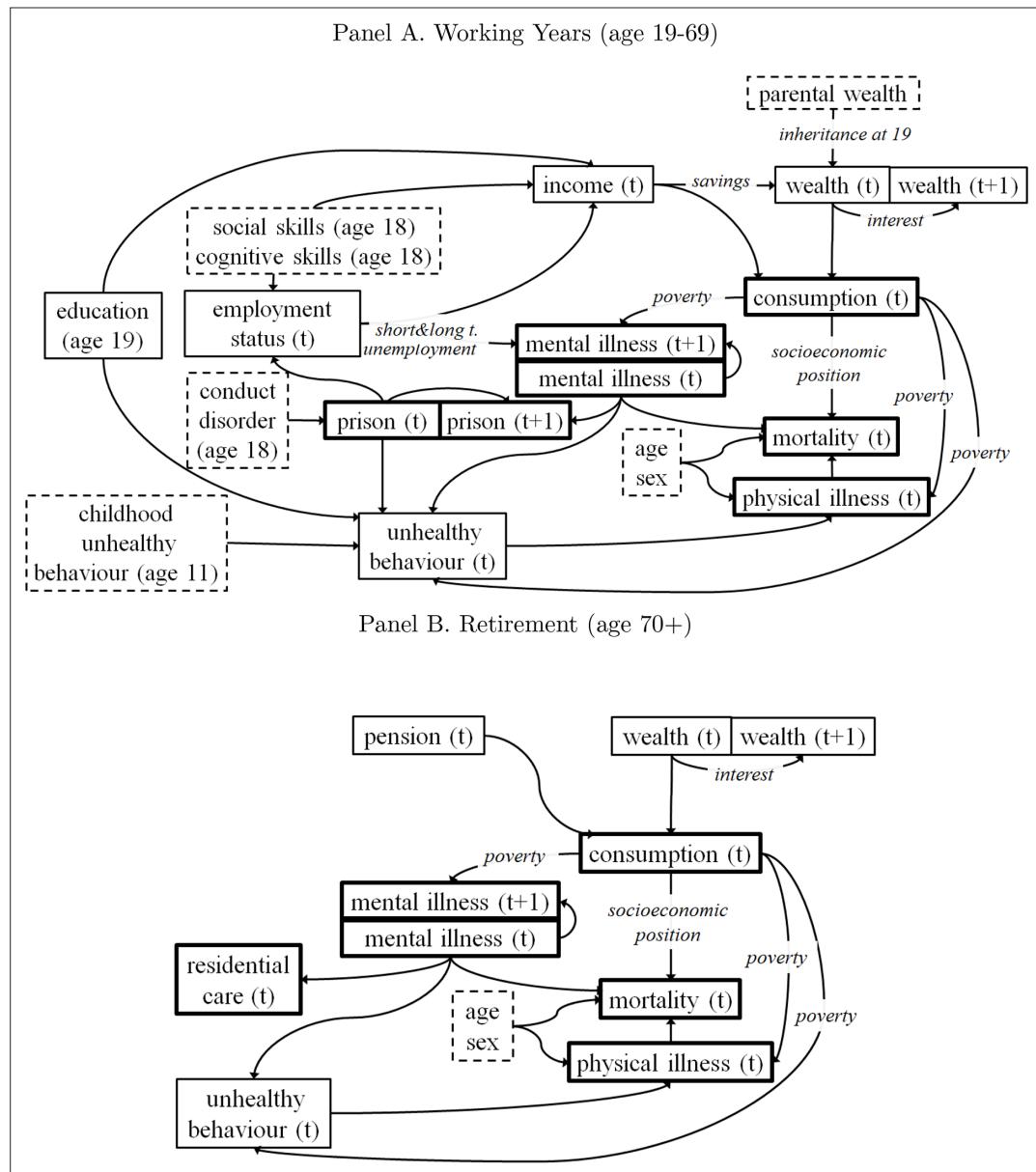


Figure 3. Model Structure for Key Life Stages in Adulthood

Note: The solid arrows represent the stochastic processes that we model using modelling equations as described in Section 2.3. The boxes represent life outcomes: the dashed boxes represent exogenous inputs into our model that are taken as given either from the childhood dataset or the previous life-stage, the thick boxes represent final outcomes that directly influence wellbeing or impose a cost to the public budget.

2015; Nelson et al., 2020; Cunha and Heckman, 2010; Adler et al., 2010; O'Donnell et al., 2015;; Layard et al., 2014; Shonkoff, 2010; Black et al., 2017).

LifeSim also models variables relevant to the public budget (**Figure 4**). This includes modelling the public costs over time associated with certain life outcomes, such as conduct disorder, being in prison, mental illness, coronary heart disease, as well as cash benefits paid to people who are in poverty and/or unemployed. This also includes modelling the taxes paid over time on individual earnings and financial gains. These can be aggregated, to assess the overall impact on the public budget as well as cost savings under different policy scenarios and over various time spans. Details of the evidence and assumptions about the unit costs of public services and our simple approach to modelling long-run taxes and benefits are found in Appendix A.

Table 2. Determinants of the Modelled Outcomes

Outcomes (Y)		Determinants (X)		Exogenous variables	
Name	Type	Other modelled outcomes (parameter source in brackets)	From childhood dataset (MCS)	From target datasets	
SOCIAL					
Conduct disorder	P, I		SDQ conduct problem score, SDQ impact score (both Goodman et al. (2015));		
Education (university degree)	P, I	Poverty at age 18 (Goodman et al., 2015; Fletcher, 2010; Farahati et al., 2003);	Cognitive skills (age 14), SDQ conduct problem score (age 14) (both Goodman et al. (2015));	Estimated likelihood of a person participating in Higher Education by age 30 (DE, 2016);	
Unemployment (employment)	P, I	L.prob. of employment; prison; age;	Cognitive skills (age 14) (Goodman et al. (2015)), SDQ conduct problem score (age 14) (both Goodman et al. (2015));	Employment rate in UK by age and sex (ONS, 2018);	
Poverty	I	Consumption;	60% equivalised household income in UK (ONS, 2011);		
Prison	P, I	L.prob. of prison; conduct disorder at age 18 (Fergusson et al., 2005); L.depression (Anderson et al., 2015);		Prison rates in England by age and sex (MJ, 2017 and ONS, 2018);	
Residential care	P, I	L.prob. of residential care; L.depression (McDougall et al., 2007; Stewart et al., 2014);		Rates of people aged 65+ in care home by sex in England (ONS, 2011).	
HEALTH					
Smoking	P, I	L.prob. of smoking; education, poverty (both Jefferis et al. (2003)); depression (Lasser et al., 2000), prison (Singleton et al., 2003);		Teenage smoking rates by sex (MCS, age 14) (param. from Jefferis et al. (2003)), smoking rate in England by age, sex and IMD quintile group (HSE, 2006);	
Coronary heart disease (CHD)	P, I	L.prob. of CHD; L.smoking (Bazzano et al., 2003; Critchley and Capewell, 2003); L.poverty (Marmot et al., 1997);		CHD rates in England by age, sex and IMD quintile group (HSE, 2006);	
Depression	P, I	L.prob. of depression; L.conduct disorder (Luby et al., 2014); employment (Thomas et al., 2005); poverty (Weich and Lewis, 1998);		Emotional disorder rates in England by age, sex and IMD quintile group (MHCPGB, 2004), depression rates in England by age, sex and IMD quintile group (HSE, 2014);	
Mortality	P, I	Depression (Chang et al., 2010); CHD (estimated using HSE (2006) and ONS data);		Mortality rates in England by age, sex and IMD quintile group (ONS, 2011);	

Continued

Table 2. Continued

Outcomes (Y)		Determinants (X)		Exogenous variables	
Name	Type	Other modelled outcomes (parameter source in brackets)		From childhood dataset (MCS)	From target datasets
ECONOMIC					
Earnings from employment (gross), £	C	Education (Blundell et al., 2000); employment; age;	Cognitive skills at age 14, SDQ conduct problem score at age 14 (both Goodman et al. (2015));	Full time annual gross pay in UK by age and sex (ONS, 2015);	
Interest, £	C	L.Wealth;		Interest rates in UK;	
Pension, £	C	Years in employment; age;		Level of state pension in UK;	
Savings, £	C	Earnings from employment; interest; taxes; L.consumption;			
Wealth, £	C	L.Wealth; Earnings from employment; interest; pension; tax; consumption; residential care; age;	Parental wealth; parental income;		
Taxes, £	C	Earnings from employment; interest; pension;		Income tax brackets in UK;	
Benefits, £	C	Earnings from employment; interest; pension; wealth; residential care;	Parental income;	Conditions for claiming benefits in UK;	
WELLBEING					
Consumption, £	C	Earnings from employment; interest, pension; taxes; savings; residential care; L.consumption, L.wealth;	Parental income;		
Health quality, QALYs	C	CHD, depression (both Sullivan et al. (2011));		Average health quality in England by age, sex and IMD quintile group (Love-Koh et al., 2015);	
Wellbeing, wellbeing-QALYs	C	Health quality, consumption (both Cookson et al. (2021)).			

Note: Outcome types: P -- probability, I -- indicator, C -- continuous. Other abbreviations: prob. - probability, L. -- lagged (previous year), MCS -- Millennium Cohort Study, ONS -- Office for National Statistics, HSE -- Health Survey for England, MHCYPGB -- Mental Health of Children and Young People in Great Britain, IMD -- Index of Multiple Deprivation, DE -- Department of Education. Full references for the parameter sources are in [the Appendix B](#).

2.2. Parameters

To model later life outcomes, we use equations parameterised using (i) cross-sectional target data which describe expected levels of and associations between variables at a point in time, usually based on up-to-date survey or administrative data, and (ii) effect estimates which attempt to draw inferences about the effect of one variable on another variable, either at the same time or a future point in time, usually based on statistical analysis of longitudinal data on historical cohorts.

Our target data comes from recent and nationally representative available surveys and administrative records in England. Our effect estimates come from studies based on longitudinal data in a UK context, unless robust estimates are only available from other high-income countries.

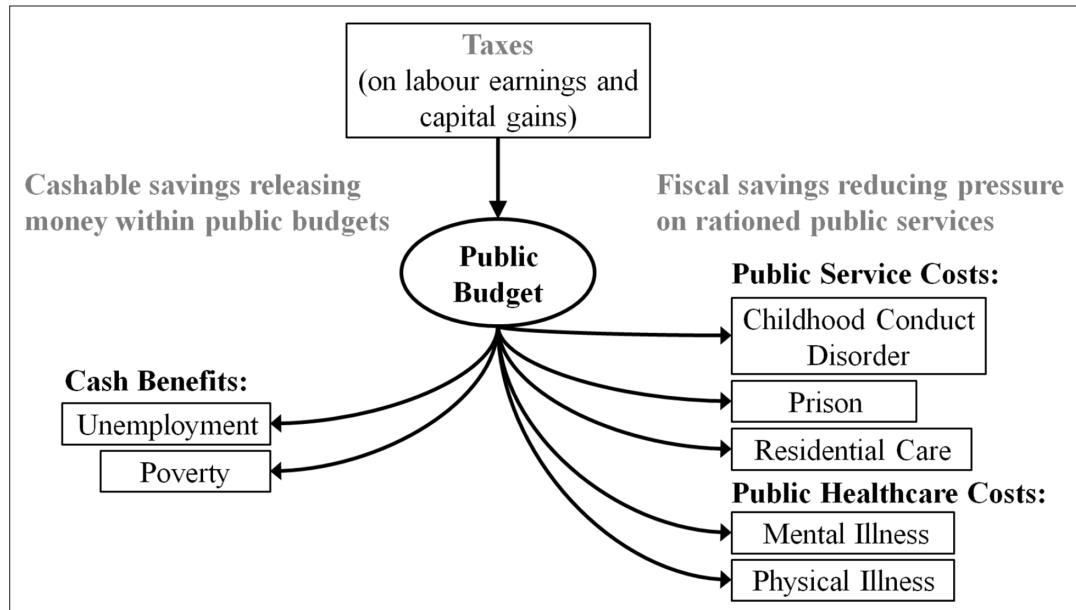


Figure 4. Model Structure for Public Costs

Our effect estimates come from studies of longitudinal data which control for observed confounding factors and focus on plausible causal relationships for which there is a large body of theoretical and empirical evidence. Nevertheless, our estimates are subject to potential omitted variable bias and cohort bias. For example, we take the estimated effect of childhood SDQ score on earnings in young adulthood from a study of longitudinal data on children born in 1970, which controls for observed child-level, family-level, and neighbourhood-level factors. We interpret this as a causal estimate i.e. if you increase SDQ score you will increase adult earnings by this amount. However, this estimate may be too low or too high if there are unobserved variables which influence both SDQ score and earnings ("omitted variable bias"). It may also be biased if the underlying stochastic processes have changed since 1970, such that SDQ score is now a more or less powerful determinant of adult earnings ("cohort bias"). Using estimates based on past cohorts of individuals thus relies on the assumption that micro-level causal effects do not change much over many decades (e.g. the proportional effect of social skills on earnings for an individual), even though the macro-level prevalence of each outcome within society may change dramatically (e.g. the average levels of social skills and earnings).

Table 2 summarises the dependencies between the modelled outcomes together with parameter sources for effects estimates, if applicable, as well as the dependencies of the modelled outcomes on the target datasets and the variables from the MCS childhood dataset. More details, as well as the full description of the target datasets are found in Appendix A.

2.3. Modelling Equations

Most of the equations modelling the outcomes can be described as one of the following: (i) simple level equations based on target data only; (ii) complex level equations based on target data supplemented with effect estimates; (iii) simple difference equations based on age associations observed in cross-sectional target data; (iv) complex difference equations based on age associations observed in cross-sectional target data supplemented with effect estimates. We illustrate each below in turn with a simple example. We also use equations that do not fit this taxonomy to model specific variables, such as savings behaviour and wealth accumulation over time, as well as public costs (more on this can be found in Appendix A).

2.3.1. Level Equations

To model the individual probability of dying, the simplest approach is to use historical mortality rates:⁵

$$pr.\overline{dead}_{i,age} = \overline{dead}[age_i, sex_i, imd_i] \quad (3)$$

where $\overline{dead}[age_i, sex_i, imd_i]$ is the mean probability of dying conditional on age, sex and English index of multiple deprivation (IMD) quintile group, calculated using a target dataset such as the Office for National Statistics mortality data (see Table A2). We denote means from a target dataset using an overline.

We can also supplement equation (3) with effects estimates. For example, we may wish to model that coronary heart disease (CHD) increases one's probability of dying by a certain proportion (denoted by $\beta_{chd}^{pr.\overline{dead}}$). In this case, we use:

$$\begin{aligned} pr.\overline{dead}_{i,age} &= f(\overline{dead}[\dots], \overline{chd}[\dots], chd_{i,age}, \beta_{chd}^{pr.\overline{dead}}) = \\ &= \max[0, \min[1, \overline{dead}[\dots](1 + \beta_{chd}^{pr.\overline{dead}}(chd_{i,age} - \overline{chd}[\dots]))]] \end{aligned} \quad (4)$$

where $chd_{i,age}$ is the simulated binary outcome of individual having a CHD at a certain age, $\overline{chd}[\dots] \equiv \overline{chd}[age_i, sex_i, imd_i]$ is the mean CHD prevalence given age, sex and IMD quintile group from a target dataset and min, max restrict the probability to have a value from 0 to 1. Notice that we subtract the mean CHD prevalence from the simulated CHD outcome to avoid double counting, as the term $\overline{dead}[\dots] \equiv \overline{dead}[age_i, sex_i, imd_i]$ is not independent from CHD, but the variable CHD is not observable in the ONS mortality target dataset, so we cannot directly condition the target mortality mean on the CHD status. After multiplying each term in the brackets by the beta coefficient, it can be seen that our approach is equivalent to subtracting the 'population attributable risk' from the risk of the simulated individual (**Webb et al., 2016**).

2.3.2. Difference Equations

If a level of a variable is already known, we can proceed by modelling the evolution of a variable as a difference from a previous time period. For example, when the level of earnings has been established at age 19 (the start of 'working years' life stage), we can model the change in individual earnings during the subsequent periods as:

$$\Delta earnings_{i,age} = trend.\overline{earnings}[age_i, sex_i] \quad (5)$$

where $\Delta earnings_{i,age} = earnings_{i,age} - earnings_{i,age-1}$ is the change in earnings from the previous year, and $trend.\overline{earnings}[age_i, sex_i]$ is a trend that governs the changes in earnings over time, calculated from a target dataset on earnings by age and sex.

Similar to level-equations, we can supplement equation (6) with an effect estimate. For example, to model that developing depression reduces earnings by a certain level represented by $\beta_{depressed}^{earnings}$ we use:

$$\Delta earnings_{i,age} = trend.\overline{earnings}[age_i, sex_i] + \beta_{depressed}^{earnings} \Delta depressed_{i,age} \quad (6)$$

where $depressed_{i,age}$ is an indicator of an individual having a depression at a given age and $\Delta depressed_{i,age} = depressed_{i,age} - depressed_{i,age-1}$.

2.4. Wellbeing Summary Measure

Conventional methods of unweighted benefit-cost analysis do not provide direct information about impacts on wellbeing and can be criticised on two important grounds. First, by focusing on unweighted consumption they ignore the well-established concept in economics of diminishing marginal value of consumption; second, they provide no information about the social distribution of costs and benefits and their impact on inequalities (see discussion in **Cookson et al. (2021)**). There is a large literature on

5. This equation and other equations in this section are simplified examples of the actual equations that we use; see Appendix A for the full mortality equation and the other equations that we use.

the theoretical and practical shortcomings of unweighted cost-benefit analysis and the advantages of alternative utilitarian and prioritarian approaches to economic evaluation based on explicit individual wellbeing and social welfare functions (**Adler and Fleurbaey, 2016**).

Our framework generates individual-level outcomes that could be used in many different ways to create summary indices of wellbeing for use in economic evaluation. In our illustrative evaluation we follow **Cookson et al. (2021)**, who propose a simple approach based on the quality-adjusted life year (QALY) concept in health economics, but adjusting for consumption as well as health-related quality of life. Our approach could be used to construct many other multidimensional measures of wellbeing that have been proposed in the literature, including equivalent income measures and measured based on life satisfaction (**Adler and Fleurbaey, 2016**). **Cookson et al. (2021)** refer to their approach as an "equivalent life" approach (**Canning, 2013**), and the resulting wellbeing metric as "years of good life" or "wellbeing QALYs". Following them, we represent individual wellbeing in year t by a function $w(\cdot)$ increasing in both consumption and health. More specifically, $w(\cdot) = \text{health}_{i,age} + u(\text{consumption}_{i,age})^{1-\eta}$ where $u(\cdot)$ is a standard isoelastic utility of income function defined as $u(\cdot) = A - B \times \text{consumption}_{i,age}^{1-\eta}$. The parameter $\eta > 1$ captures diminishing marginal value of income, and A and B are constants which depend on normative parameters: η (already mentioned), minimal consumption for a life worth living and standard consumption for a good life. In the current application we set minimal consumption at £1,000 (estimated amount required to buy basic food supplies in the UK for a year) and standard consumption at £24,000 (the mean consumption in the LifeSim simulated cohort), and $\eta = 1.26$ (see **Cookson et al. (2021)**).

The interpretation is that a good year is a year lived enjoying full health and consuming the equivalent of the average consumption in a rich country. The good-years measure is more informative than conventional monetary measures because it takes into account the notion that one pound of additional consumption is worth substantially more to a poor individual than a rich individual.

2.5. Computing Methods

LifeSim is implemented in software R (tested on R version 3.6.2) using object-oriented programming for R (requires R6 and tidyverse packages). The code and related data files are available on GitHub (<https://github.com/ievask/lifesim-simulator>) and can be run on a high performance computing (HPC) cluster (Slurm Workload Manager).

When we split the simulation into 500 partitions, it takes 28 minutes to run it on the HPC cluster. The simulation can also be run on a standard PC, for any chosen number of individuals.

The code is written using an object-oriented approach built around individuals, capturing their initial endowments and the skills and assets they acquire through life as they undergo various experiences, the probability of which are influenced by their past histories. This allows us to simulate individual life histories in an intuitive manner and easily communicate and validate our modelling assumptions in discussion with domain experts in various stages of the life-course. The code is currently written in R allowing us to elegantly incorporate advanced statistical methods into our modelling. However, R being an interpreted language can be slow to run and if performance was a concern our code could easily be translated into a compiled object oriented programming language such as C++. There are also ways of re-writing the original R code in more compact ways, known as "vectorisation", which are harder for non-specialists to follow but faster to run because they avoid conventional programming loops that require the same time-consuming interpretation operations to be applied repeatedly.

3. Baseline Results

In this section we show our baseline simulation results, and demonstrate some formats in which they can be analysed.

Table 3 provides key summary statistics for the simulated outcomes, including child outcomes, adult outcomes and final wellbeing outcomes. We show means, standard deviations, and the minimum and maximum value of an outcome in the total distribution of the simulated individuals in the baseline simulation, as well as means and standard errors for a bootstrap simulation, i.e. after running the simulation 100 times with a different random seed each time. **Table 3** does not present the summary statistics of the the initial conditions, as well as the child's cognitive skills and SDQ scores that we obtain from the childhood survey dataset (MCS), as these variables have already been summarised in **Table 1**.

Table 3. Summary Statistics of the Simulated Outcomes

Outcome	Baseline simulation				Bootstrap simulation	
	Mean	SD	Min	Max	Mean	SE
CHILD OUTCOMES						
Conduct disorder at age 5, %	8.63	28.08	0.00	100.00	8.59	0.10
Conduct disorder at age 18, %	9.09	28.74	0.00	100.00	9.00	0.08
ADULT OUTCOMES						
Proportion of university graduates, %	38.51	48.66	0.00	100.00	38.51	0.15
Proportion of working years in unemployment, %	5.73	6.63	0.00	100.00	5.70	0.02
Proportion of lifetime in poverty, %	26.85	18.30	0.00	100.00	26.80	0.05
Proportion of working years in prison, %	1.62	5.12	0.00	75.00	1.60	0.01
Proportion of retirement in residential care, %	1.28	3.95	0.00	100.00	1.28	0.02
Proportion of adult years as a smoker, %	5.32	5.12	0.00	100.00	5.31	0.02
Proportion of adult years with CHD, %	6.16	4.07	0.00	33.33	6.15	0.01
Proportion of life years with mental illness, %	5.98	2.81	0.00	25.81	5.99	0.01
Years of life	78.78	13.04	0.00	100.00	78.80	0.04
Premature mortality rate (before age 75), %	28.13	44.96			28.04	0.14
Annual earnings (lifetime average), £	29,655	7,638	4,792	67,879	29,659	23
Annual savings (lifetime average), £	2,833	942	0	7,803	2,832	3
Annual interest (lifetime average), £	402	234	0	3,321	402	1
FINAL WELLBEING OUTCOMES						
Annual consumption (lifetime average), £	24,114	6,648	10,000	113,817	24,115	22
Healthy years	68.28	9.99	0.87	88.16	68.30	0.03
Healthy years (discounted)	40.94	3.96	0.87	48.01	40.96	0.01
Good years	65.67	10.21	0.69	91.93	65.69	0.03
Good years (discounted)	39.80	4.89	0.69	52.18	39.82	0.01

Note: The baseline simulation mean, standard deviation (SD), minimum value (Min) and maximum value (Max) are calculated for the simulated population of 100,000 (for the lifetime aggregates, or yearly – for the annual variables). The bootstrap simulation mean and standard error (SE) are calculated for the distribution of the means of the 100 bootstrap simulations. The time periods for calculating life-stage proportions are as follows: ‘working years’ refer to the period between ages 19–69; ‘retirement’ refers to the time period from age 70 up to death; adult years refer to the time period from age 19 up to death; lifetime refers to the entire period from birth to death. CHD – coronary heart disease. We use year 2015/16 prices and the annual discount rate of 1.5% (Paulten and Claxton, 2012).

The baseline simulation means do not differ much from the bootstrap means, and the bootstrap standard errors are small, implying that changing the random seed has a negligible effect on the simulated outcome means with the simulation size that we use.

Approximately 9% of 18 year-old adults develop conduct disorder in the LifeSim simulation. This estimate fits within the range of 1–10%, commonly reported in the epidemiology literature on conduct disorder (see a review in **Hinshaw and Lee (2003)**, also **Patel et al. (2018)**). Our estimate, however, slightly exceeds the 8% of young men and 5% of young women with conduct disorder estimated by Mental Health of Children and Young People in England survey in year 2017. This small difference may be caused by the fact that the algorithm that we use to simulate conduct disorder incidence is based and validated on child samples attending child mental health clinics (**Goodman et al., 2000**; **Goodman et al., 2003**), and therefore it may overestimate the actual conduct disorder prevalence in the general population. On the other hand, conduct disorder diagnosis in the clinic sample can be argued to be more precise and sensitive than in the survey data sample, because in the clinic sample

diagnosis was made by mental health specialists using detailed information on symptoms and resultant impairments gathered from multiple informants, whereas in the specific survey sample diagnosis was based on a single specific tool – Development and Well-Being Assessment.

Figure 5 shows the simulated distributions of some core outcomes, which also include the distribution of lifetime wellbeing (measured using the approach by *Cookson et al. (2021)* described in section 2.4.)

Table 4 shows the average costs to the public budget associated with certain outcomes, cash benefits paid to people who are in poverty or unemployed, as well as taxes on earnings and financial gains. These are calculated over various time intervals over the life-course, and separately for the general population, and then for people born in the lowest and top income quintile groups at birth.⁶

Table 5 provides two summary measures of inequality, based on differences in lifetime expected wellbeing between best off and worst off groups on the basis of the following early childhood circumstances – sex, parental income quintile group (poorest vs. richest 20%), parental mental health, parental education, and high baseline conduct problems (SDQ conduct problem score at age 5 equal to 7 or above). 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.

4. Comparison With Other Datasets

4.1. Comparison With 1970 Birth Cohort Study

Table 6 compares the LifeSim predictions with data from the 1970 Birth Cohort Study (BCS70) at ages 26, 29, 42 and 46, as a simple validation check. We list the number of observations, means and standard-deviations of the LifeSim variables for children born in the year 2000 and the BCS70 variables for children born in the year 1970, representing the same outcomes. For each outcome, we quantify the difference between the LifeSim distribution and BCS70 distribution in terms of the absolute difference in their means and standard deviations.

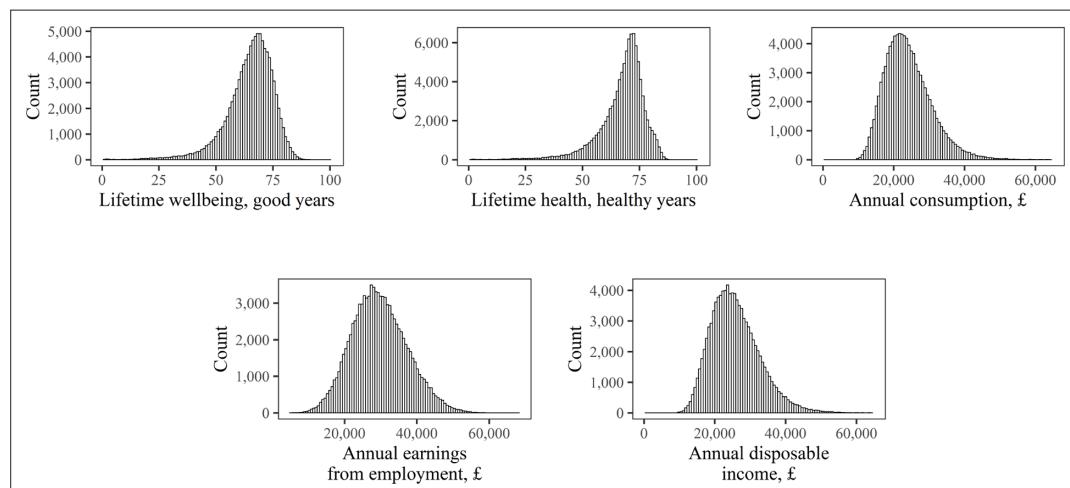


Figure 5. Distributions of Core Outcomes

Note: Disposable income calculated net of taxes, and include earnings from employment, earnings from interest, earnings from pension during retirement years, and cash benefits.

6. The total residential care cost figure does not include the substantial private costs of residential care, which we assume fall on individuals if they have sufficient savings, nor the public costs of residential care before the age of 60. It may be an underestimate of public costs, because we make simple and conservative assumptions about the need for residential care and eligibility for public funding - for example, we use simple sex-specific rates of care home use in people aged 65 and over (2% for men and 4% for women) but do not model the rapid age-related increase in risk which results in much higher rates for people surviving into their 80s and beyond.

Table 4. Cumulative Costs Over Various Time Periods

Costs (Per Capita), £	Age 0-10	Age 0-15	Age 0-20	Age 0-25	Lifetime
GENERAL POPULATION COHORT					
PUBLIC SERVICES					
Conduct disorder	1,100	1,700	1,800	1,800	1,800
Prison	0	0	930	3,200	17,000
Residential care	0	0	0	0	1,200
HEALTHCARE					
CHD	0	0	0	0	1,400
Depression	740	1,900	3,600	4,900	14,000
Other	9,300	14,000	20,000	27,000	88,000
Benefit payments	1,400	2,100	4,100	6,300	12,000
LOWEST INCOME QUINTILE GROUP AT BIRTH					
PUBLIC SERVICES					
Conduct disorder	1,400	2,100	2,200	2,200	2,200
Prison	0	0	1,100	4,000	20,000
Residential care	0	0	0	0	1,200
HEALTHCARE					
CHD	0	0	0	0	1,400
Depression	770	2,000	3,600	5,000	14,000
Other	11,000	16,000	22,000	29,000	91,000
Benefit payments	8,600	12,000	17,000	19,000	22,000
TOP INCOME QUINTILE GROUP AT BIRTH					
PUBLIC SERVICES					
Conduct disorder	880	1,300	1,500	1,500	1,500
Prison	0	0	800	2,700	14,000
Residential care	0	0	0	0	1,100
HEALTHCARE					
CHD	0	0	0	0	1,400
Depression	710	1,900	3,500	4,900	14,000
Other	8,400	13,000	18,000	25,000	85,000
Benefit payments	0	0	1,300	3,000	7,900

Note: All values are calculated per simulated individual in year 2015/16 prices, and discounted at 1.5% annual rate, and rounded to 2 significant figures. See details on cost sources in Table A6 in Appendix A.

Table 5. Whole Cohort Lifetime Inequality by Childhood Circumstance

Childhood circumstance	Number of children	Annual consumption, £	Lifetime health, healthy years	Lifetime wellbeing, good years
Best off 20%	20,000	32,559	68.71	69.59
Worst off 20%	20,000	18,471	66.31	59.84
Difference		14,088	2.40 ⁷	9.76
Extreme best off 20%	12,149	32,909	68.81	69.83
Extreme worst off 20%	26	16,808	62.16	54.51
Difference		16,101	6.66	15.32

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

7. Standard period estimates of gaps in healthy life expectancy by current socioeconomic status are substantially larger than our cohort estimate of gaps by early childhood circumstance, due to dynamic interdependence

Table 6. Comparison with the British Cohort 1970

Outcome	N		Mean		Difference in means	SD		Difference in SDs
	LifeSim	BCS70	LifeSim	BCS70		LifeSim	BCS70	
AGE 26								
Male (indicator)	99,402	9,003	0.48	0.46	0.03	0.50	0.50	0.00
University Degree (indicator)	99,402	8,399	0.39	0.25	0.13	0.49	0.43	0.05
Employed (indicator)	99,402	9003	0.95	0.96	-0.01	0.23	0.20	0.02
Earnings (in year 2015 £)	99,402	6,642	19,736	14,279	5,457	5,882	7,110	-1,228
Depression (indicator)	99,402	9003	0.07	0.10	-0.03	0.25	0.30	-0.05
Smoking (indicator)	99,402	8,892	0.06	0.27	-0.20	0.25	0.44	-0.20
AGE 29								
Male (indicator)	99,267	11,261	0.48	0.49	-0.00	0.50	0.50	-0.00
University Degree (indicator)	99,267	11,211	0.39	0.27	0.11	0.49	0.44	0.04
Employed (indicator)	99,267	9,506	0.94	0.96	-0.02	0.23	0.19	0.04
Earnings (in year 2015 £)	99,267	8,102	22,400	20,796	1,604	6,682	68,798	-62,115
Depression (indicator)	99,267	11,261	0.07	0.10	-0.03	0.25	0.30	-0.05
Smoking (indicator)	99,267	11,205	0.06	0.29	-0.23	0.24	0.45	-0.22
AGE 42								
Male (indicator)	98,149	9,841	0.48	0.48	0.00	0.50	0.50	0.00
University Degree (indicator)	98,149	9,841	0.39	0.34	0.05	0.49	0.47	0.01
Employed (indicator)	98,149	8,594	0.95	0.97	-0.02	0.21	0.16	0.05
Earnings (in year 2015 £)	98,149	2,158	29,327	22,107	7,220	9,604	15,567	-5,963
Depression (indicator)	98,149	9,756	0.07	0.11	-0.04	0.26	0.31	-0.05
Smoking (indicator)	98,149	9,801	0.06	0.20	-0.14	0.24	0.40	-0.17
AGE 46								
Male (indicator)	97,532	8,581	0.48	0.48	-0.00	0.50	0.50	-0.00
University Degree (indicator)	97,532	8,444	0.39	0.34	0.04	0.49	0.47	0.01
Employed (indicator)	97,532	5,038	0.95	0.99	-0.03	0.21	0.12	0.09
Earnings (in year 2015 £)	97,532	358	28,558	22,538	6,020	9,978	31,559	-21,581
CHD (indicator)	97,532	8,353	0.02	0.00	0.02	0.15	0.05	0.10
Depression (indicator)	97,532	8,486	0.06	0.14	-0.08	0.23	0.35	-0.12
Smoking (indicator)	97,532	8,578	0.05	0.15	-0.10	0.23	0.36	-0.13

Note: N – number of observations, SD – standard-deviation. We quantify the difference between the LifeSim distribution and BCS70 distribution in terms of the absolute difference in their means and standard deviations. Earnings is the net pay from employment.

We would expect some adult outcomes to be similar (e.g. health) but others to be substantially different (e.g. earnings, rates of smoking and university education), and so this can be seen as a simple validation check to ensure that our model provides broadly similar findings in the same ballpark where appropriate, and substantially different findings where we know different generations had very different experiences e.g. smoking. Nevertheless, most variables do not deviate substantially from the same quantities characterising the cohort born in 1970.

between health and social status over the lifecourse. Adult-onset illness that is unrelated to early childhood circumstances may cause downward social mobility, and deterioration of social and economic outcomes that is unrelated to early childhood circumstances may cause deteriorating health.

One exception already mentioned is smoking, which is expected and can be explained by the change in smoking rates over time. Another exception is education – the proportion of people with a degree under 30 years old – which is much higher in the LifeSim cohort. This can be explained by the change in higher education participation rates over time, and increased equality between the genders in the cohort born in 2000. Over time the 1970s cohort partially catches up with the LifeSim cohort by obtaining qualifications at a later age – at the age 46 the proportion of people with a university degree is more similar in both samples than at the age 26. Finally, the LifeSim earnings at all ages on average exceed the 1970s cohort earnings. This can be explained by cohort effects, such as general differences in economy, society, culture and politics experienced by the two cohorts.

4.2. Comparison With Recent Cross-Sectional Data

To avoid such general cohort effects which arise when comparing two generations born 30 years apart, we also carry out a simple validity check using more recent cross-sectional datasets. More specifically, we compare our age-specific LifeSim outcomes with age-specific outcomes in cross-sectional data.

Figure 6 compares the age-earnings profile for males and females in the LifeSim simulation with our target dataset – ONS Annual Survey of Hours and Earnings in year 2015, and in the Understanding Society survey in year 2015. The concave trend with age, initially increasing and then – decreasing earnings, is very similar in the three datasets.

Figure 7 compares the earnings distributions by sex and different age groups in the LifeSim cohort and the Understanding Society data. Both distributions have similar medians for the different sex-age groups, and also become more uniform with increasing age. One issue left to be addressed as part of future work is modelling of the relatively longer right hand side tail which can be observed for the Understanding Society data and not for the LifeSim data. This tail represents the highest-earning people in the distribution. The LifeSim earnings output does not have this tail, as we do not model the outcome of being employed in extremely-high earning jobs. Addressing this feature in LifeSim would require modelling the link with variables in early life that would lead to such extremely-high earning states.

In **Figure 8**, we compare the prevalence of the different discrete outcomes in LifeSim cohort, and in our corresponding target datasets, which include Health Survey for England for the health-related

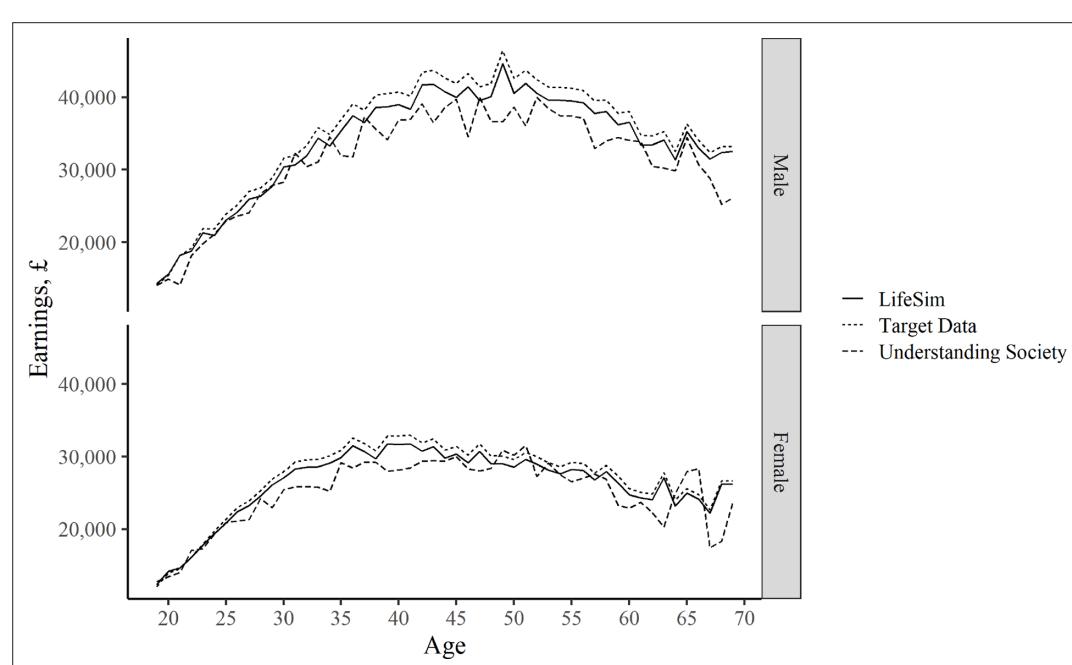


Figure 6. Earnings: Comparison with Other Data Sources

Note: Target data is the annual full-time gross pay data from Annual Survey of Hours and Earnings, ONS, 2015. See more details in Table A2. The Understanding Society data is the estimated annual full-time gross pay data from wave 7 (started in year 2015), calculated for individuals in full-time employment as usual gross pay times the frequency of pay per year. All earnings are in year 2015 £.

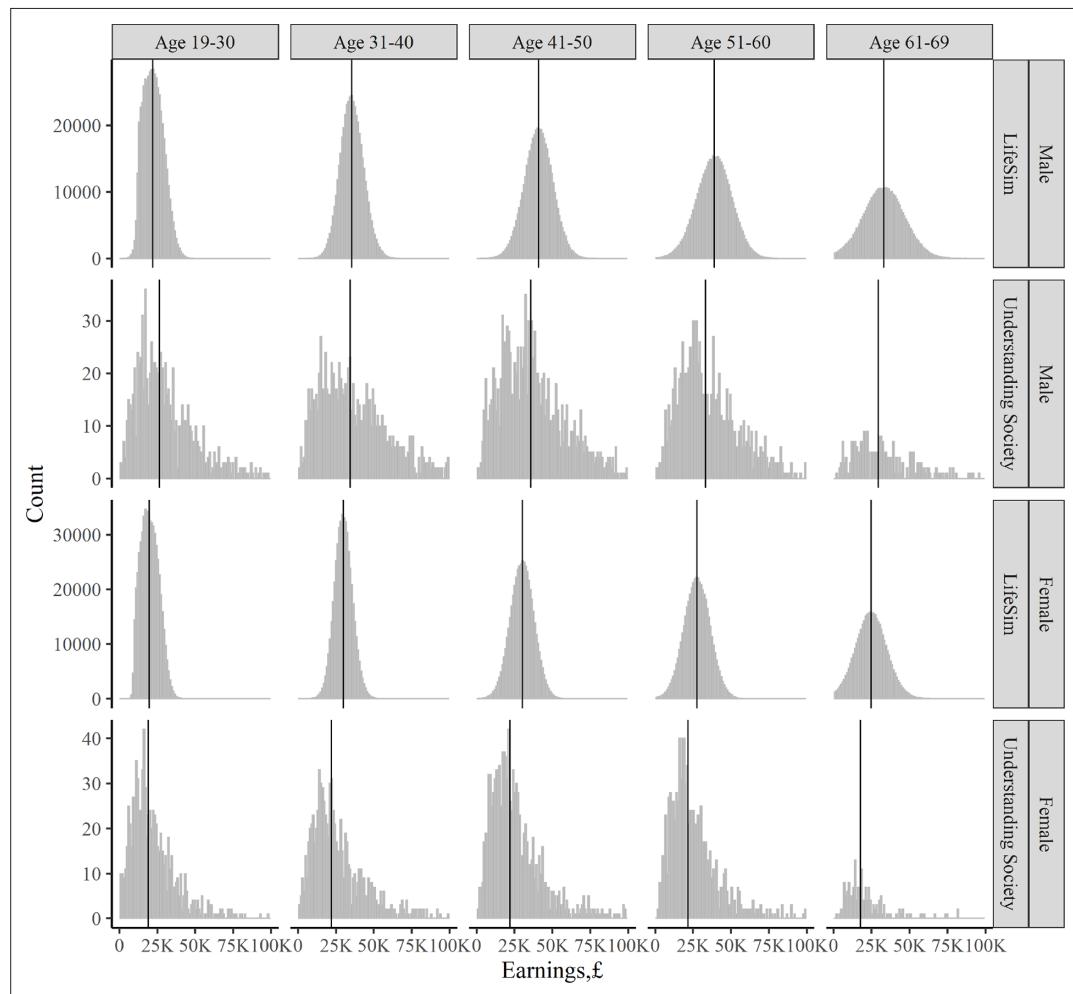


Figure 7. Other Outcomes: Comparison with Understanding Society

Note: Data is the estimated annual full-time gross pay data from Understanding Society Jwave 7 (started in year 2015), calculated for individuals in full-time employment as usual gross pay times the frequency of pay per year. All earnings in year 2015 £. The black lines denote medians.

outcomes, ONS Labour Force Survey for unemployment and Department for Education estimates for participation in higher education.

The simulated outcomes matches the target data well, but there is some small discrepancy with the Understanding Society data, which can be explained by differences how data on similar outcomes is collected across different surveys.

5. Discussion

Microsimulation offers a forward-looking alternative to conventional approaches to long-term childhood policy analysis, which have often relied on long-term follow-up of high-profile historical childhood policy experiments that took place decades ago – such as the US Abecedarian experiment (*Garcia et al., 2020*) – and are of questionable relevance to current policy decisions. We present LifeSim – a proof of concept microsimulation model for analysing the full long-term consequences of childhood policies from a lifetime perspective. LifeSim is capable of modelling a rich set of developmental, social, economic, and health outcomes from birth to death for each child in a general population birth cohort of 100,000 English children born in the year 2000-1, together with public costs and summary wellbeing metrics.

Since our model is designed for the purpose of partial equilibrium policy analysis rather than forecasting of macro-level trends, the most important criteria for model credibility arguably relate to the quality of the underlying conceptual framework and data sources rather than ability to predict

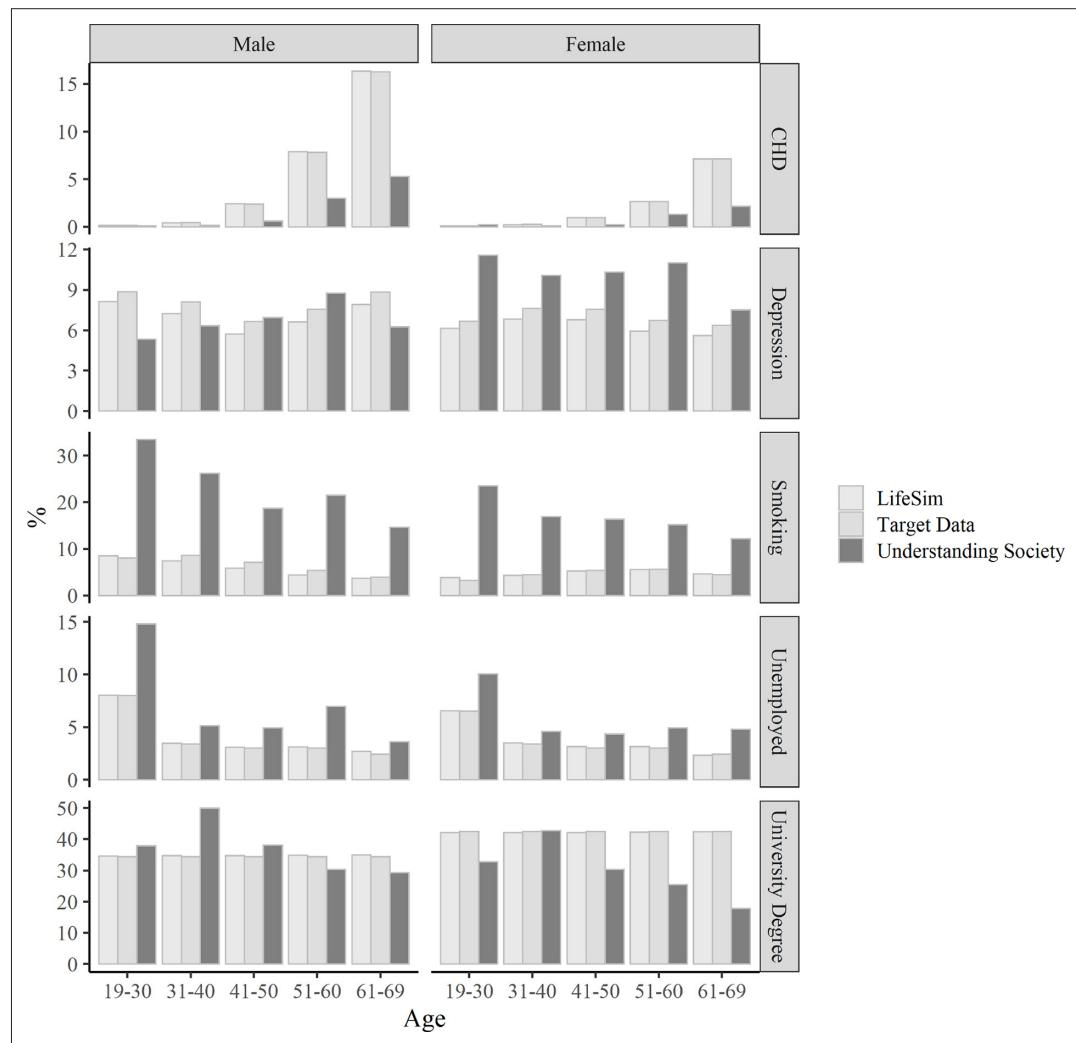


Figure 8. Earnings: Comparison with Other Data Sources

Note: Target data for the outcomes is as follows. Depression – Health Survey for England, 2014; CHD – Health Survey for England, 2006; smoking – Health Survey for England, 2006; employment – Labour Force Survey, ONS, 2018; university degree – Department for Education, 2016.

population-level trends (**Kopeck et al., 2010**). Nevertheless, we provide a simple comparison of our simulation with external data on population-level trends. First, we provide a comparison with data from the 1970 Birth Cohort Study up to age 46. We find that our simulation is broadly consistent with the external data and substantially divergent when appropriate – for example, our simulation for people born in 2000 has a much lower proportion of people smoking than the 1970 cohort, reflecting the reduction in smoking rates in the UK since the 1970s. Also, our simulation for people born in 2000 has a much larger proportion with young people having obtained a university degree at age 26 than the 1970 cohort at that age, reflecting the massive expansion in university provision in the UK since the 1970s.

We also provide a comparison with a recent external cross-sectional dataset – Understanding Society (in the year 2016). Our simulated earnings outcome replicates reasonably well the sex-age specific distributions observed in the Understanding Society data. Also, for our simulated key discrete outcomes – including health-related outcomes and unemployment – the sex specific prevalence trends against age are not too deviant from the trends observed in the Understanding Society data. Any minor discrepancies can be explained by differences in data collection methods for Understanding Society and our target datasets.

Finally, we provide an additional check of LifeSim output against the various target datasets that we directly use to calibrate our equations, such as Health Survey for England, and Office for National Statistics datasets. As expected, our simulated outcomes match very well the trends and patterns

observed in the target data. Because our model is flexible and can be used together with many data sources, if needed, one can easily substitute our target datasets with alternative datasets, to match the trends and patterns observed in these alternative sources.

The main strength of our model is that it captures the dynamic individual-level interaction between many outcomes across the social, economic and health domains over the entire lifecourse. Previous models have modelled either two or three of these domains or only a part of the lifecourse. Simultaneously analysing many outcomes allows us to capture how many early life disadvantages can compound over the lifecourse creating a spiral of multiple disadvantage.

Another strength of LifeSim is that it simulates the long-run outcomes for a whole general population cohort of children, not just analysing the outcomes of a narrow group of trial participants. This allows carrying out more complex and policy-relevant analysis, including assessment of the distributional impacts and policy opportunity costs on the general population, and exploring options for targeting the policy to different subgroups of the population.

LifeSim also generates long-term individual-level data, which makes it compatible with applying new multidimensional summary indices of wellbeing recently proposed in the theoretical literature (*Cookson et al., 2021; O'Donnell et al., 2014; Fleurbaey et al., 2013; Fleurbaey and Schokkaert, 2013*). These indices are more informative than conventional monetary valuation based on aggregate outcomes, as they allow to account for the diminishing marginal value of consumption and other sources of heterogeneity in the marginal value of different life outcomes to different individuals. However, application of these indices in practice requires individual level long-term time series data on many outcomes across the health, social and economic outcome domains. Such rich long-term data is difficult to obtain from existing datasets, especially if we are interested in analysing cohorts living in present rather than historical cohorts of people born decades ago. Models such as LifeSim can compile the many data sources together to extrapolate the required individual-level long term outcomes.

Perhaps the most important limitation of our modelling approach is the assumption that micro-level causal pathways are invariant to social trends and policy intervention. LifeSim can readily accommodate macro-level social trends, such as changes in average earnings and educational attainment, by using up-to-date target data. However, some social trends do raise potential threats to our fundamental assumption of causal pathway invariance. For example, the massive expansion in higher education participation since the early 1990s means that the "signalling" value of a university degree has diminished as a way of helping employers to identify job candidates with exceptional ability. The proportional effect of obtaining a university degree on adult earnings for the 2000 cohort graduating in the early 2020s is thus likely to be smaller than the effect estimated in the 1970 cohort graduating in the early 1990s. We do not take into account potentially measurable biases of this kind in the effect estimates used in the present version of LifeSim, which are based on existing published analysis of longitudinal data on historical cohorts of children. However, this limitation could be addressed in future work by identifying the most important potential biases in effect estimates from longitudinal data on historical cohorts and making appropriate adjustments through careful synthesis of evidence and elicitation of expert opinion. Relatedly, specific transition pathway estimates could also be modified in specific cases to strengthen external validity for specific populations. For example, estimates based on long-term outcomes for mostly white children born in the 1970s may not be applicable to Asian British populations. Using external data sources to estimate long-run health effects for Asian British populations would produce more applicable estimates for those populations.

Further, in principle, using up-to-date cross-sectional target data as well as effect estimates is a methodological strength of our approach to parameterisation, which can help to improve the external validity of the model by correcting for omitted variable and cohort biases in our effect estimates and ensure that our predictions are calibrated in line with current population-level outcomes. However, the current version of our model relies more heavily on cross-sectional target data than effect estimates, which is a limitation from the perspective of internal validity and causal inference. This is an important limitation, given the intended use of the model for the purpose of policy analysis and evaluation. Future work in developing a version of the model for routine policy analysis could aim to reverse this balance, placing more emphasis on credible effect estimates to improve internal validity while maintaining a role for target data to ensure external validity.

LifeSim can also be extended to incorporate additional features. One extension would be to incorporate more outcomes. Our model includes many different categories of human capital (e.g. cognitive skills, social skills, educational attainment, health, employment) but within each category, more

nuanced distinctions could be made. Health outcomes are modelled using just three binary variables – mental illness (depression), physical illness (CHD), and mortality – educational outcomes focus only on gaining a university degree; employment outcomes focus only on unemployment not precarious employment; and our modelling of the tax and benefit system and retirement savings is extremely stylised. Similarly, more individual-level factors could be included (e.g. ethnicity), more family-level factors (e.g. child abuse) and it would also be possible to add neighbourhood-level factors (e.g. neighbourhood-level deprivation in childhood, as well as family-level income). Also, our tax benefit modelling is highly stylised and could be improved by incorporating a standard static tax benefit calculator, such as Euromod (*Sutherland and Figari, 2013*).

LifeSim also currently does not model many important outcomes during childhood, such as cognitive skills, but rather takes them as given from the MCS. Future work could undertake formal dynamic modelling of all the relevant outcomes during childhood and adolescence, based on structural equation modelling and mediation analysis of MCS data that estimates all the relevant parameters in a single, integrated longitudinal data analysis.

Next, LifeSim parameters that govern the evolution of lifecourse outcomes during different stages of childhood and adulthood are currently estimated separately using different studies based on different datasets, specifications and estimation methods. This increases the parameter uncertainty, which compounds over time when modelling lifecourse trajectories. Future childhood policy modelling could adopt a more joined-up and systematic approach to estimate these parameters simultaneously by linking together data on different stages of the lifecourse from successive cohort studies (*Hughes et al., 2021*). This would make the model more ready for prime time policy analysis, and also allow a formal analysis of parameter uncertainty by bootstrapping parameters using an estimated variance-covariance matrix.

Another extension would be to re-calibrate our model to other populations – e.g. the UK in 2025, or England or Scotland, or a sub-national area of England – by updating the initial conditions of the birth population and the external macro target data on average population level outcomes and associations within that birth population in subsequent years.

Furthermore, LifeSim currently does not model interactions between individuals, and an important extension would be to model interactions, such as the dynamics of family formation and dissolution and spillover effects on other family members. Building an interactive model would also allow modeling the effects of infectious disease transmission, as well as the non-communicable mental and physical illnesses that are currently the focus of the model.

Our model structure could also be extended in more fundamental ways – for example, to model the all-age population rather than just a birth cohort, and to model parental investment choices and other behavioural responses that may depend on social trends, changes in the policy environment, and/or the behaviour of other individuals. It should be acknowledged that considering any extensions involves making trade-offs between model complexity and tractability, and in some cases it may be preferable to use other more specialist models and combine the findings from different models, rather than expand an existing model. For example, as already mentioned – our model could be combined with Euromod (*Sutherland and Figari, 2013*) – the tax and benefit microsimulation model, to generate more comprehensive output on taxes and benefits for the assessment of the consequences to the public budget.

Overall, LifeSim is a flexible childhood policy model which serves as proof of concept in demonstrating the potential added value of lifecourse microsimulation in long-term childhood policy analysis. It sets a foundation for the development of a long-term childhood policy model which can be routinely used to carry out prime time policy analysis.

ORCID iDs

- Ieva Skarda  <https://orcid.org/0000-0002-0866-2936>
Miqdad Asaria  <https://orcid.org/0000-0002-3538-4417>
Richard Cookson  <https://orcid.org/0000-0003-0052-996X>

Acknowledgements

We would first like to thank the members of our advisory group: Annalisa Belloni, Sarah Cattan, Leon Feinstein, Paul Frijters, Peter Goldblatt, Heather Joshi, Catherine Law, Lara McClure and Christine Power. For useful comments we also are grateful to Shehzad Ali, Mark Ashworth, Karen Bloor, Laura Bojke, Eva Maria Bonin, Jonathan Bradshaw, Penny Breeze, Alan Brennan, Eric Brunner, Tracey Bywater,

Simon Capewell, Maria Guzman Castillo, Bette Chambers, Brendan Collins, Gabriella Conti, Peter Diggle, Tim Doran, Susan Griffin, Nils Gutacker, James Heckman, Nathan Hendron, Bruce Hollingsworth, Andrew Jones, Noemi Kreif, Christodoulos Kypridemos, Richard Mattock, Cheti Nicoletti, Owen O'Donnell, Martin O'Flaherty, Kate Pickett, George Ploubidis, Gerry Richardson, Jemimah Ride, Matthew Robson, Tracey Sach, Filipa Sampaio, Trevor Sheldon, Tushar Srivastava, Mark Strong, David Taylor-Robinson, Valentina Tonei, Aki Tsuchiya, Simon Walker, Margaret Whitehead and Mark Mon Williams, and anonymous reviewers of previous versions of the manuscript.

We would also like to thank Matteo Richiardi and two anonymous reviewers for detailed and constructive comments on our original submission to the International Journal of Microsimulation.

The errors and opinions expressed in this paper are our own.

Funding

This is independent research supported by the National Institute for Health Research (SRF-2013-06-015), the Wellcome Trust (Grant No. 205427/Z/16/Z), and the Prevention Research Programme (ActEarly Programme, MR/S037527/1). The authors have no other conflicts of interest to report. The views expressed in this publication are those of the authors and not necessarily those of the National Institute for Health Research, the Wellcome Trust, the NHS, the Department of Health and Social Care, or the Prevention Research Programme.

Conflict of Interest

No competing interests reported.

Data and Code Availability

The code is available from GitHub: <https://github.com/ievask/lifesim-simulator>. Further details on code availability can also be found in Section 2.5 and on computing methods on page 13.

References

- Adler MD**, Fleurbaey M. 2016. *The Oxford Handbook of Well-being and Public Policy*. Oxford, United Kingdom: Oxford University Press.
- Adler NE**, Stewart J, Adler N, Stewart J. 2010. Preface to the biology of disadvantage: Socioeconomic status and health. *Annals of the New York Academy of Sciences* **1186**: 1–4. DOI: <https://doi.org/10.1111/j.1749-6632.2009.05385.x>
- Almond D**, Currie J, Duque V. 2018. Childhood circumstances and adult outcomes: Act II. *Journal of Economic Literature* **56**: 1360–1446. DOI: <https://doi.org/10.1257/jel.20171164>
- Anderson DM**, Cesur R, Tekin E. 2015. Youth depression and future criminal behavior. *Economic Inquiry* **53**: 294–317. DOI: <https://doi.org/10.1111/eicn.12145>
- Asaria M**. 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. https://eprints.whiterose.ac.uk/135407/1/CHERP147_health_care_costs_NHS.pdf
- Bazzano LA**, He J, Muntner P, Vupputuri S, Whelton PK. 2003. Relationship between cigarette smoking and novel risk factors for cardiovascular disease in the United States. *Annals of Internal Medicine* **138**: 891–897. DOI: <https://doi.org/10.7326/0003-4819-138-11-200306030-00010>
- Black MM**, Walker SP, Fernald LCH, Andersen CT, DiGirolamo AM, Lu C, McCoy DC, Fink G, Shawar YR, Shiffman J, Devercelli AE, Wodon QT, Vargas-Barón E, Grantham-McGregor S, Lancet Early Childhood Development Series Steering Committee. 2017. Early childhood development coming of age: Science through the life course. *Lancet* **389**: 77–90. DOI: [https://doi.org/10.1016/S0140-6736\(16\)31389-7](https://doi.org/10.1016/S0140-6736(16)31389-7)
- Blundell R**, Dearden L, Goodman A, Reed H. 2000. The returns to higher education in Britain: Evidence from a british cohort. *The Economic Journal* **110**: F82–F99. DOI: <https://doi.org/10.1111/1468-0297.00508>
- Bonin E-M**, Stevens M, Beecham J, Byford S, Parsonage M. 2011. Costs and longer-term savings of parenting programmes for the prevention of persistent conduct disorder: A modelling study. *BMC Public Health* **11**: 803. DOI: <https://doi.org/10.1186/1471-2458-11-803>
- Canning D**. 2013. Axiomatic foundations for cost-effectiveness analysis. *Health Economics* **22**: 1405–1416. DOI: <https://doi.org/10.1002/hec.2889>
- Chang C-K**, Hayes RD, Broadbent M, Fernandes AC, Lee W, Hotopf M, Stewart R. 2010. All-cause mortality among people with serious mental illness (SMI), substance use disorders, and depressive disorders in southeast London: A cohort study. *BMC Psychiatry* **10**: 77. DOI: <https://doi.org/10.1186/1471-244X-10-77>
- Conti G**, Mason G, Poupanis S. 2019. Developmental origins of health inequality. *Oxford Research Encyclopedia of Economics and Finance*: 1. <https://oxfordre.com/economics/view/10.1093/acrefore/9780190625979.001.0001/acrefore-9780190625979-e-4>
- Cookson R**, Skarda I, Cotton-Barratt O, Adler M, Asaria M, Ord T. 2021. Quality adjusted life years based on health and consumption: A summary wellbeing measure for cross-sectoral economic evaluation. *Health Economics* **30**: 70–85. DOI: <https://doi.org/10.1002/hec.4177>

- Critchley JA**, Capewell S. 2003. Mortality risk reduction associated with smoking cessation in patients with coronary heart disease: A systematic review. *JAMA* **290**: 86–97. DOI: <https://doi.org/10.1001/jama.290.1.86>
- Cunha F**, Heckman J. 2010. Working paper 16201. Investing in our young people. National Bureau of Economic Research. <https://www.nber.org/papers/w16201>
- Curtis L**, Burns A. 2017. Unit Costs of Health and Social Care 2017. Canterbury, United Kingdom: Personal Social Services Research Unit, University of Kent. <https://www.pssru.ac.uk/project-pages/unit-costs/unit-costs-2017/>
- Dubourg R**, Hamed J, Thorns J. 2005. The Economic and Social Costs of Crime Against Individuals and Households 2003/04. <http://library.college.police.uk/docs/hordsolr/rdsolr3005.pdf>
- Edwards RT**, Céilleachair A, Bywater T, Hughes DA, Hutchings J. 2007. Parenting programme for parents of children at risk of developing conduct disorder: Cost effectiveness analysis. *BMJ* **334**: 682. DOI: <https://doi.org/10.1136/bmj.39126.699421.55>
- Farahati F**, Marcotte DE, Wilcox-Gök V. 2003. the effects of parents' psychiatric disorders on children's high school dropout. *Economics of Education Review* **22**: 167–178. DOI: [https://doi.org/10.1016/S0272-7757\(02\)00031-6](https://doi.org/10.1016/S0272-7757(02)00031-6)
- Fergusson DM**, Horwood LJ, Ridder EM. 2005. Show me the child at seven: The consequences of conduct problems in childhood for psychosocial functioning in adulthood. *Journal of Child Psychology and Psychiatry, and Allied Disciplines* **46**: 837–849. DOI: <https://doi.org/10.1111/j.1469-7610.2004.00387.x>
- Fletcher JM**. 2010. Adolescent depression and educational attainment: Results using sibling fixed effects. *Health Economics* **19**: 855–871. DOI: <https://doi.org/10.1002/hec.1526>
- Fleurbaey M**, Luchini S, Muller C, Schokkaert E. 2013. Equivalent income and fair evaluation of health care. *Health Economics* **22**: 711–729. DOI: <https://doi.org/10.1002/hec.2859>
- Fleurbaey M**, Schokkaert E. 2013. Behavioral welfare economics and redistribution. *American Economic Journal* **5**: 180–205. DOI: <https://doi.org/10.1257/mic.5.3.180>
- Frijters P**, Bellet C, Krekel C. 2017. Micro-Macro simulations for wellbeing, version 1.0. <https://www.york.ac.uk/media/healthsciences/documents/research/Micro-macro%20simulations%20for%20wellbeing.pdf>
- Garcia JL**, Heckman JJ, Leaf DE, Prados MJ. 2020. Quantifying the life-cycle benefits of an influential early-childhood program. *The Journal of Political Economy* **128**: 2502–2541. DOI: <https://doi.org/10.1086/70718>
- Goodman A**, Joshi H, Nasim B, Tyler C. 2015. Social and Emotional Skills in Childhood and Their Long-Term Effects on Adult Life. London, United Kingdom: Institute of Education. <https://www.nuffieldfoundation.org/news/active-policy-required-avoid-covid19-crisis-exacerbating-inequalities>
- Goodman R**, Ford T, Simmons H, Gatward R, Meltzer H. 2003. Using the strengths and difficulties questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. *International Review of Psychiatry* **15**: 166–172. DOI: <https://doi.org/10.1080/0954026021000046128>
- Goodman R**, Renfrew D, Mullick M. 2000. Predicting type of psychiatric disorder from strengths and difficulties questionnaire (SDQ) scores in child mental health clinics in London and Dhaka. *European Child & Adolescent Psychiatry* **9**: 129–134. DOI: <https://doi.org/10.1007/s007870050008>
- Heckman JJ**. 2012. The developmental origins of health. *Health Economics* **21**: 24–29. DOI: <https://doi.org/10.1002/hec.1802>
- Hennessy DA**, Flanagan WM, Tanuseputro P, Bennett C, Tuna M, Kopec J, Wolfson MC, Manuel DG. 2015. The population health model (POHEM): An overview of rationale, methods and applications. *Population Health Metrics* **13**: 24. DOI: <https://doi.org/10.1186/s12963-015-0057-x>
- Hinshaw S**, Lee S. 2003. Conduct and Oppositional Defiant Disorders. Mash EJ, Barkley RA (Eds). *Child Psychopathology*. p. 144–198.
- Hughes RA**, Tilling K, Lawlor DA. 2021. Combining longitudinal data from different cohorts to examine the life-course trajectory. *American Journal of Epidemiology*: kwab190. DOI: <https://doi.org/10.1093/aje/kwab190>
- Jeffeiris B**, Graham H, Manor O, Power C. 2003. Cigarette consumption and socio-economic circumstances in adolescence as predictors of adult smoking. *Addiction* **98**: 1765–1772. DOI: <https://doi.org/10.1111/j.1360-0443.2003.00552.x>
- Jones E**, Schoon I. 2008. Child cognition and behaviour. London, United Kingdom: Centre for Longitudinal studies. Hansen K, Joshi H (Eds). *Millennium Cohort Study Third Survey: A User's Guide to Initial Findings*. p. 118–144.
- Kopec JA**, Finès P, Manuel DG, Buckeridge DL, Flanagan WM, Oderkirk J, Abrahamowicz M, Harper S, Sharif B, Okhmatovskaia A, Sayre EC, Rahman MM, Wolfson MC. 2010. Validation of population-based disease simulation models: A review of concepts and methods. *BMC Public Health* **10**: 710. DOI: <https://doi.org/10.1186/1471-2458-10-710>
- Kypridemos C**, Allen K, Hickey GL, Guzman-Castillo M, Bandosz P, Buchan I, Capewell S, O'Flaherty M. 2016. Cardiovascular screening to reduce the burden from cardiovascular disease: Microsimulation study to quantify policy options. *BMJ* **353**: i2793. DOI: <https://doi.org/10.1136/bmj.i2793>
- Lasser K**, Boyd JW, Woolhandler S, Himmelstein DU, McCormick D, Bor DH. 2000. Smoking and mental illness: A population-based prevalence study. *JAMA* **284**: 2606–2610. DOI: <https://doi.org/10.1001/jama.284.20.2606>
- Layard R**, Clark AE, Cornaglia F, Powdthavee N, Vernoit J. 2014. What predicts a successful life? A life-course model of well-being. *Economic Journal* **124**: F720–F738: F720. DOI: <https://doi.org/10.1111/eco.12170>
- Liu JLY**, Maniadakis N, Gray A, Rayner M. 2002. The economic burden of coronary heart disease in the UK. *Heart* **88**: 597–603. DOI: <https://doi.org/10.1136/heart.88.6.597>
- Love-Koh J**, Asaria M, Cookson R, Griffin S. 2015. The social distribution of health: Estimating quality-adjusted life expectancy in England. *Value in Health* **18**: 655–662: S1098-3015(15)01847-1. DOI: <https://doi.org/10.1016/j.jval.2015.03.1784>

- Luby JL**, Gaffrey MS, Tillman R, April LM, Belden AC. 2014. Trajectories of preschool disorders to full DSM depression at school age and early adolescence: Continuity of preschool depression. *The American Journal of Psychiatry* **171**: 768–776. DOI: <https://doi.org/10.1176/appi.ajp.2014.13091198>
- Marmot MG**, Bosma H, Hemingway H, Brunner E, Stansfeld S. 1997. Contribution of job control and other risk factors to social variations in coronary heart disease incidence. *Lancet* **350**: 235–239. DOI: [https://doi.org/10.1016/S0140-6736\(97\)04244-X](https://doi.org/10.1016/S0140-6736(97)04244-X)
- McCrone P**, Dhanasiri S, Patel A, Knapp M, Lawton-Smith S. 2008. Paying the Price: The Cost of Mental Health Care in England to 2026. London, United Kingdom: King's Fund. https://www.kingsfund.org.uk/sites/default/files/Paying-the-Price-the-cost-of-mental-health-care-England-2026-McCrone-Dhanasiri-Patel-Knapp-Lawton-Smith-Kings-Fund-May-2008_0.pdf
- McDougall FA**, Kvala K, Matthews FE, Paykel E, Jones PB, Dewey ME, Brayne C, Medical Research Council Cognitive Function and Ageing Study. 2007. Prevalence of depression in older people in England and Wales: The MRC CFA study. *Psychological Medicine* **37**: 1787–1795. DOI: <https://doi.org/10.1017/S0033291707000372>
- Milne BJ**, Milne B, Lay-Yee R, Lay J, Pearson J, Randow M, Davis P. 2015. Modelling the early life-course (MELC): A microsimulation model of child development in New Zealand. *International Journal of Microsimulation* **8**: 28–60. DOI: <https://doi.org/10.34196/ijm.00116>
- Nelson CA**, Scott RD, Bhutta ZA, Harris NB, Danese A, Samara M. 2020. Adversity in childhood is linked to mental and physical health throughout life. *BMJ* **371**: m3048. DOI: <https://doi.org/10.1136/bmj.m3048>
- O'Donnell G**, Deaton A, Durand M, Halpern D, Layard R. 2014. . <https://li.com/wp-content/uploads/2019/03/commission-on-wellbeing-and-policy-report-march-2014-pdf.pdf> [Accessed January 24, 2020].
- O'Donnell O**, Doorslaer EV, Ourti TV. 2015. Health and inequality. Atkinson A, Bourguignon F (Eds). *Handbook of Income Distribution*. Vol. Volume 2A-2B Oxford, United Kingdom: Elsevier B.V. p. 1419–1533.
- Patel RS**, Amaravadi N, Bhullar H, Lekireddy J, Win H. 2018. Understanding the demographic predictors and associated comorbidities in children hospitalized with conduct disorder. *Behavioral Sciences* **8**: E80. DOI: <https://doi.org/10.3390/bs8090080>
- Paulden M**, Claxton K. 2012. Budget allocation and the revealed social rate of time preference for health. *Health Economics* **21**: 612–618. DOI: <https://doi.org/10.1002/hec.1730>
- Romeo R**, Knapp M, Scott S. 2006. Economic cost of severe antisocial behaviour in children--and who pays it. *The British Journal of Psychiatry* **188**: 547–553. DOI: <https://doi.org/10.1192/bjp.bp.104.007625>
- Scott S**, Knapp M, Henderson J, Maughan B. 2001. Financial cost of social exclusion: Follow up study of antisocial children into adulthood. *BMJ* **323**: 191. DOI: <https://doi.org/10.1136/bmj.323.7306.191>
- Shonkoff JP**. 2010. Building a new biodevelopmental framework to guide the future of early childhood policy. *Child Development* **81**: 357–367. DOI: <https://doi.org/10.1111/j.1467-8624.2009.01399.x>
- Singleton N**, Farrell M, Meltzer H. 2003. Substance misuse among prisoners in England and Wales. *International Review of Psychiatry* **15**: 150–152. DOI: <https://doi.org/10.1080/0954026021000046092>
- Skarda I**, Asaria M, Cookson R. 2021. Full Lifecourse Economic Evaluation of Childhood Policies. *EQUIPOL Working Paper*. <https://www.york.ac.uk/media/healthsciences/documents/research/Full%20Lifecourse%20Economic%20Evaluation%20of%20Childhood%20Policies.pdf>
- Spelauer M**, Hicks C, Gribble S, Rowe G, Lin X, Moore K, Plager L. 2013. The Lifepaths Microsimulation Model: An Overview. Ottawa, Canada: Statistics Canada – Modelling Division. <https://www.statcan.gc.ca/eng/microsimulation/lifepaths/overview>
- Stewart R**, Hotopf M, Dewey M, Ballard C, Bisla J, Calem M, Fahmy V, Hockley J, Kinley J, Pearce H, Saraf A, Begum A. 2014. Current prevalence of dementia, depression and behavioural problems in the older adult care home sector: The South East London care home survey. *Age and Ageing* **43**: 562–567. DOI: <https://doi.org/10.1093/ageing/afu062>
- Sullivan PW**, Slejko JF, Sculpher MJ, Ghushchyan V. 2011. Catalogue of EQ-5D scores for the United Kingdom. *Medical Decision Making* **31**: 800–804. DOI: <https://doi.org/10.1177/0272989X11401031>
- Sutherland H**, Figari F. 2013. EUROMOD: The European Union tax-benefit microsimulation model. *International Journal of Microsimulation* **6**: 4–26. DOI: <https://doi.org/10.34196/ijm.00075>
- Thomas C**, Benzeval M, Stansfeld SA, Stansfeld S. 2005. Employment transitions and mental health: An analysis from the British household panel survey. *Journal of Epidemiology and Community Health* **59**: 243–249. DOI: <https://doi.org/10.1136/jech.2004.019778>
- Van der Ven J**. 2016. LINDA: A dynamic microsimulation model for analysing policy effects on the evolving population cross-section. *National Institute of Economic and Social Research. National Institute of Economic and Social Research Paper* **459**: 1. <https://www.niesr.ac.uk/sites/default/files/publications/dp459.pdf>
- Webb P**, Bain C, Page A. 2016. *Essential Epidemiology: An Introduction for Students and Health Professionals*. Cambridge, United Kingdom: Cambridge University Press.
- Weich S**, Lewis G. 1998. Material standard of living, social class, and the prevalence of the common mental disorders in Great Britain. *Journal of Epidemiology and Community Health* **52**: 8–14. DOI: <https://doi.org/10.1136/jech.52.1.8>
- Wolfson M**, Rowe G. 2014. Healthpaths: Using functional health trajectories to quantify the relative importance of selected health determinants. *Demographic Research* **31**: 941–974. DOI: <https://doi.org/10.4054/DemRes.2014.31.31>
- Zucchelli E**, Jones AM, Rice N. 2012. The evaluation of health policies through dynamic microsimulation methods. *International Journal of Microsimulation* **5**(1): 2–20.

Appendix A Modelling Equations

This appendix summarises the principles that we follow and data that we use to build our modelling equations.

A.1. Functional Forms

The functional form of each modelling equation is chosen depending on (a) the type of variable that we model (continuous quantity vs. indicator); and (b) the format in which the parameter estimate is reported (e.g. coefficient estimates from a linear regression, odds ratios from a logistic regression, percentage changes, etc.).

A.1.1. Modelling A Continuous Quantity

Literature on continuous quantities (e.g. earnings, IQ scores, age at death) most often report parameter estimates in the form of beta coefficients from a linear regression, that represent either (i) absolute change in the dependent variable (Y) as a result of a unit change in the independent variable (X); or (ii) relative (or percentage) change in Y , as a result of a unit change in X . Following this, we model the effects on Y either as absolute or relative changes.

Absolute Change. Assume that we want to model β_{cog}^{earn} – the linear effect of individual cognitive ability at age 18 (denoted $cog_{i,18}$) on individual earnings at age 19 (denoted $earn_{i,19}$). The linear specification is:

$$earn_{i,19} = \alpha_{i,19} + \beta_{cog}^{earn} cog_{i,18} + error_{i,19} \quad (1)$$

where $\alpha_{i,19}$ captures the constant, as well as the effects of other observable and unobservable variables not explicitly specified in equation (1); $error_{i,19}$ is random noise with a zero mean.

Equation (1) does not explicitly account for all the possible variables which may drive the term $\alpha_{i,19}$, for example, it does not explicitly model economic conditions, social networks and many other characteristics of the modelled individual i . To overcome this problem, we assume that the modelled individual i is ‘average’ in terms of all of the outcomes that we do not explicitly account for, i.e. for all $i = 1,..N$ it is assumed that $\alpha_{i,19} = \frac{1}{N} \sum_{i=1}^N \alpha_{i,19} \equiv \bar{\alpha}[19]$, where N is the number of 19 year-old individuals in the representative population. The term $\bar{\alpha}[19]$ can also be expressed from an averaged equation (1) as $\bar{\alpha}[19] = \bar{earn}[19] - \beta_{cog}^{earn} \bar{cog}[18]$ and then substituted for $\alpha_{i,19}$ in equation (1), to get:

$$earn_{i,19} = \bar{earn}[19] + \beta_{cog}^{earn} (cog_{i,18} - \bar{cog}[18]) + error_{i,19} \quad (2)$$

We approximate the average values, such as $earn[age]$, $cog[age]$ by averages of survey data representative of the cohort that we model (i.e. what we call ‘target data’, see Section 2.3 in the main paper and **Table A2** in this Appendix).

Relative Change. Assume that a standard deviation increase in the cognitive skills at age 14 is estimated to cause an $k\%$ increase in the earnings at age 19. It can be shown that in this case using the procedure described above will yield:

$$earn_{i,19} = \bar{earn}[19] \exp \left((1 + \frac{k}{100}) (cog_{i,19} - \bar{cog}[14]) + error_{i,19} \right) \quad (3)$$

A.1.2. Modelling A Discrete Event

Sometimes we wish to model a discrete event – e.g. whether a person obtains a degree, smokes or not, is employed or not, etc. In this case, we first model the individual age-specific probability of event occurring, and then – determine whether the event actually occurs by comparing the modelled probability with a random draw from a uniform distribution over a closed interval from zero to one. Literature researching discrete events most often reports

estimates from a probabilistic regression, that represent either (i) average absolute change (percentage-point change) in the probability of the event occurring as a result of a unit change in the independent variable; or (ii) odds ratio.

Percentage Point Change. Assume that we wish to model the effect of cognitive ability at age 18 (denoted $cog_{i,18}$) on whether an individual obtains university degree (denoted edu_i). Also, assume that it is known that a standard deviation change in the cognitive ability at age 18 increases the probability of obtaining a degree by g percentage points. For example, **Goodman et al. (2015)** reports such estimates as average marginal effects from a probit regression model.¹ In a crude way, we can model the probability of obtaining a degree ($pr.edu_i$) as:

$$pr.edu_i = \overline{edu}[sex] + \frac{g}{100} (cog_{i,18} - \overline{cog}[18, sex]) + error_{i,19} \quad (4)$$

Odds-Ratio. When the effect estimates are obtained from a logistic probability regression model, they are often reported as odds ratios. For example, to assess the effect of depression on smoking, literature may report estimates of the following equation:

$$\ln\left(\frac{pr.sm_{i,age}}{1-pr.sm_{i,age}}\right) = \alpha_{i,age} + \beta_{dep}^{pr.sm} dep_{i,age} + error \quad (5)$$

where $dep_{i,age}$ is an indicator of individual-depression; $\beta_{dep}^{pr.sm}$ is the natural logarithm of the reported odds ratio. Again, we can average equation (5), and as long as $\ln\left(\frac{\overline{pr.sm}[\dots]}{1-\overline{pr.sm}[\dots]}\right)$ is a good approximation of $\frac{1}{N}\sum_i^N \{\ln\left(\frac{pr.sm_{i,age}}{1-pr.sm_{i,age}}\right)\}$, we can assume that $\ln\left(\frac{\overline{pr.sm}[\dots]}{1-\overline{pr.sm}[\dots]}\right) = \overline{\alpha}[age, sex] + \beta_{dep}^{pr.sm} \overline{dep}[\dots]$, where $\overline{pr.sm}[\dots] \equiv pr.sm[age, sex]$, and $\overline{dep}[\dots] \equiv dep[age, sex]$. We can express $\overline{\alpha}[age, sex]$ from this expression, and again assume that $\alpha_{age,i} = \overline{\alpha}[age, sex]$ and substitute $\overline{\alpha}[age, sex]$ in (5), then rearrange to get:

$$pr.sm_{i,age} = \left(\frac{1}{\frac{\overline{pr.sm}[\dots]}{1-\overline{pr.sm}[\dots]} \exp\left(\beta_{dep}^{pr.sm} (dep_{i,age} - \overline{dep}[\dots])\right)} + 1 \right)^{-1} \quad (6)$$

A.2. Parameter Sources

Table A1 explains the notation that we use to specify the modelling equations throughout the rest of the **Table A2** summarises the target data; **Table A3** lists the literature sources of the parameter estimates used in parameterising the modelling equations; **Table A4** summarises what other variables these literature sources control for. We then provide full detailed specifications of the modelling equations to model each of the lifecourse outcomes, as well as full details on modelling taxes, cash benefits and costs associated with costly outcomes, in the next subsection.

Table A1. Notation

Notation	Explanation
SIMULATED VARIABLES	
<i>rec</i>	Recipient for the parent-training programme (indicator);
<i>cp</i>	Conduct problem measure;
<i>ip</i>	Impact of problems;

Continued

1. It should be noted that the average marginal effect is not always a good approximation of the true effect, as the actual individual marginal effect is not constant across individuals. So this method is a crude way of modelling the effect.

Table A1. Continued

Notation	Explanation
<i>cd</i>	Childhood conduct disorder (indicator);
<i>cog</i>	Cognitive skills;
<i>edu</i>	University degree (indicator);
<i>sm</i>	Smokes (indicator);
<i>dep</i>	Mental illness (indicator);
<i>chd</i>	Coronary heart disease (CHD) (indicator);
<i>dead</i>	Dead (indicator);
<i>pris</i>	In prison (indicator);
<i>care</i>	In residential care (indicator);
<i>empl</i>	Employed (indicator);
<i>earn</i>	Annual earnings, £;
<i>wealth</i>	Lifetime accumulated wealth;
<i>cons</i>	Annual consumption level, £;
<i>pov</i>	In poverty (indicator);
<i>tax</i>	Annual amount of taxes paid;
<i>benef</i>	Annual amount of benefits received;
<i>sav</i>	Savings rate;
<i>min.cons</i>	Minimum consumption level, which government subsidises if it cannot be sustained by an individual;
<i>sex</i>	Male (indicator);
<i>sep</i>	Socio-economic position (quintile group);
<i>sdq.cp.MCSj</i>	SDQ conduct problem score reported in MCS sweep <i>j</i> (<i>j</i> = 2; 3; 4; 5; 6);
<i>sdq.ip.MCSj</i>	SDQ impact score reported in MCS sweep <i>j</i> ;
<i>cog.MCSj</i>	Extracted factor using principal component analysis based on cognitive skills tests reported in MCS sweep <i>j</i> , standardised with a mean of 1.00 and standard deviation of 0.15 following Jones and Schoon (2008);
OTHER NOTATION	
prefix <i>pr.</i>	Probability, i.e. <i>pr.sm</i> denotes probability of smoking;
line over variable (—)	Mean calculated from a target dataset, i.e. $\overline{sm}[age, sex]$ is proportion of people smoking in a particular age and sex group;
prefix <i>trend.</i>	Modelled time trend, i.e. the mean increase in variable over time, estimated from a target dataset, i.e. during working years expected earnings increase as people get past their youth, as they gain work experience, climb the career ladder, etc.;
prefix <i>sd.</i>	Modelled variation in some variable, i.e. standard deviation in the variable, estimated from a target dataset;
β_x^y	Parameter representing the effect of some outcome $\beta_{sm}^{pr.chd}$ denotes the effect of smoking on CHD risk. Depending on the context, we use it to represent coefficients from a linear regression, odds-ratios, etc. See full list of parameters, and their sources in Table A3 ;
<i>SMR_x</i>	Standard mortality ratio given condition <i>x</i> , i.e. the probability of dying from condition <i>x</i> divided by the probability of dying in the general population.

Note: MCS – Millennium Cohort Study.

Table A2. Target Data

Parameter	Description	Source
$\overline{dead}[age, sex, sep]$	Mortality rates (by age, sex and the English IMD quintile group);	ONS, 2011;
$\overline{dep}[age, sex, sep]$	Ages 5-18: proportion of children with any emotional disorder (by age, sex, and IMD quintile group); age 18+: depression diagnosed by a doctor and present or being treated within the past 12 months in England (by age, sex, and English IMD quintile group);	For age 5-18: Mental Health of Children and Young People Great Britain, 2004; age 18+: Health Survey for England, 2014;
$\overline{chd}[age, sex, sep]$	Proportion of people with CHD in England (by age, sex, and English IMD quintile group);	Health Survey for England, 2006;
$\overline{earn}[age, sex]$	Mean full time annual gross pay in UK (by age and sex);	Annual Survey of Hours and Earnings, ONS, 2015;
$\overline{empl}[age, sex]$	Seasonally adjusted employment rate, expressed as a proportion of the economically active population (by age and sex);	Labour Force Survey, ONS, 2018;
$\overline{sdq.cp}[age, sex]$	Mean SDQ conduct problem score (by age and sex);	MCS, 2000–2014;
$\overline{cog}[age, sex]$	Mean cognitive measure (by age and sex);	MCS, 2000–2014;
$\overline{edu}[19]$	Higher Education Initial Participation Rate in 2015/2016 (estimate of the likelihood of a person participating in Higher Education by age 30, based on current participation rates, adjusted by the probability of dropping out);	Department for Education, 2016;
$\overline{sm}[14, sex]$	Proportion of 14-year-old children smoking (by sex);	MCS, 2014;
$\overline{sm}[19, sex]$	Proportion of daily smokers in England (by age, sex and English IMD quintile group) in England;	Health Survey for England, 2006;
$\overline{pov}[sex]$	Proportion of households below 60% median income by sex in UK;	Family Resources Survey, Department for Work & Pensions, 2016/2017;
$\overline{cd}[4, sex]$	Proportion of 4-year-old children with conduct disorder (by age, sex);	Mental Health of Children and Young People Great Britain, 2004;
$\overline{pris}[age, sex]$	Average proportion of people in prison (by age and sex) in England and Wales over 31 March 2017 - 31 March 2018 (calculated using population estimates in mid-2017);	Offender Management Statistics, Ministry of Justice, 2017–2018; Population Estimates for UK, England and Wales, Scotland and Northern Ireland Mid-2017, ONS;
$\overline{care}[70, sex]$	Proportion of people aged 65+ in resident care homes (by sex) in England and Wales, 2011;	"Changes in the older resident care home population between 2001 and 2011" 2014, ONS.

Note: MCS – Millennium Cohort Study, ONS – Office for National Statistics, IMD – Index of Multiple Deprivation. Our notation uses an overline to denote averages from a target dataset.

Table A3. Parameters

Parameter	Value	Source	Notes
SMR_{dep}	3.21 among 15-44 year olds, 1.75 – 45-64 year olds and 1.18 for 65+	Chang et al. (2010)	Age standardised mortality ratios in southeast London 2007–2009, for people with depressive episode against the general population of England and Wales in 2008;
$\beta_{chd}^{pr.dead}$	See Table A5	Health survey for England (2006); the 20th Century Mortality Files, ONS; Mid-year population estimates for England and Wales, ONS	Estimated probability of dying from CHD among those who have a CHD, in England and Wales, 2008 using CHD prevalence rates of 2006;

Continued

Table A3. Continued

Parameter	Value	Source	Notes
$\beta_{sdq.cond}^{earn}$	$\ln 1.004/SD_{sdq.cp}$	Goodman et al. (2015) social	0.4% increase in gross wage with standard deviation increase in externalising subscale (conduct+peer); $SD_{sdq.cp}$ – standard deviation of SDQ conduct problem score in the relevant age-sex subgroup of our simulation.
β_{cog}^{earn}	$\ln 1.072/SD_{cog}$	Goodman et al. (2015) social	7.2% increase in gross wage with standard deviation increase in IQ score; SD_{cog} – standard deviation of cognitive skills in the relevant age-sex subgroup of our simulation.
β_{edu}^{earn}	$\ln 1.17$ if male; $\ln 1.37$ if female	Blundell et al. (2000)	17% increase in hourly wage from having undergraduate degree for males, 37% for females;
$\beta_{cog}^{pr.edu}$	$0.12/SD_{cog(19)}$	Goodman et al. (2015) social	standard deviation increase in cognitive ability associated with 12% point increase in prob. obtaining a degree;
$\beta_{sdq.cp}^{pr.edu}$	$0.02/SD_{sdq.cp(19)}$	Goodman et al. (2015) social	standard deviation decrease in Rutter externalising score associated with 2.2% point increase in prob. obtaining a degree;
$\beta_{dep}^{pr.edu}$	-0.04	Goodman et al. (2015) social; Fletcher (2010) adolescent; Farahati et al. (2003) effects	goodman2015social fletcher2010adolescent find no statistically significant effect; but fletcher2010adolescent finds that being depressed increases the probability of dropping out of high school by around 2.4% points, and decreases the probability of college enrolment by 2.7–7.2 percentage points. farahati2003effects find that parent's depression increases child's probability of dropout by over 3% points for females. In the light of these findings, the current model specification sets the parameter at 4% points;
$\beta_{teen.sm}^{pr.sm}$	$\ln 3.38$ if male; $\ln 3.68$ if female	Jefferis et al. (2003) cigarette	Estimates obtained using logistic regression;
$\beta_{pov}^{pr.sm}$	$\ln 1.91$ if male; $\ln 1.81$ if female	Jefferis et al. (2003) cigarette	Estimates obtained using logistic regression;
$\beta_{edu}^{pr.sm}$	$\ln 3.32$ if male; $\ln 3.26$ if female	Jefferis et al. (2003) cigarette	Estimates obtained using logistic regression;
$\beta_{dep}^{pr.sm}$	$\ln 2.7$	Lasser et al. (2000) smoking	Estimates obtained using logistic regression;
$\beta_{pris}^{pr.sm}$	0.07 if male and 0.06 if female	Singleton et al. (2003) substance	Calculated using the prevalence rates in a population before and after imprisonment, does not take into account the contribution of this increase because of mental illness, poverty and potentially other variables;
$\beta_{cd}^{pr.dep}$	$\ln 3.63$	Luby et al. (2014) trajectories	Including the effect that occurs via non-supportive parenting (see discussion below); estimated using logistic regression;

Continued

Table A3. Continued

Parameter	Value	Source	Notes
$\beta_{unempl}^{pr.dep}$	In 2.05 if male; In 1.72 if female	Thomas et al. (2005) employment	Estimated using logistic regression; the effect on psychological problems measured by general health questionnaire;
$\beta_{empl}^{pr.dep}$	In 0.87 if male; In 0.79 if female	Thomas et al. (2005) employment	Estimated using logistic regression; the effect on psychological problems measured by general health questionnaire;
$\beta_{pov}^{pr.dep}$	In 1.24	Weich and Lewis (1998) material	Estimated using logistic regression; the effect on psychological problems measured by general health questionnaire;
$\beta_{pov}^{pr.chd}$	In 1.49 if male; In 1.18 if female	Marmot et al. (1997) contribution	Calculated using logistic regression controlling for age and CHD risk factors (incl. smoking), social support and job control. Using the parameters depends on assuming poverty correlates with low employment grade;
$\beta_{sm}^{pr.chd}$	In 2	Bazzano et al. (2003) relationship, Critchley and Capewell (2003) mortality	Based on estimates of odds ratios reported in the cited sources (see discussion in section A.1.2);
$\beta_{cd}^{pr.pris}$	0.18	Fergusson et al. (2005) show	Estimated using rates of arrests/ convictions among people with different levels of conduct problems;
$\beta_{dep}^{pr.pris}$	0.015	Anderson et al. (2015) youth	
$\beta_{dep}^{pr.care}$	0.18	McDougall et al. (2007) prevalence; Stewart et al. (2014) current	Calculated using depression prevalence rates;
$\beta_{sdq.cp}^{pr.empl}$	$0.016 \beta_1 / SD_{sdq.cp}$	Goodman et al. (2015) social	Standard deviation increase in externalising subscale of SDQ raises probability being employed by 1.6%; $SD_{sdq.cp}$ – standard deviation of SDQ conduct problem score in the relevant age-sex subgroup of our simulation;
$\beta_{cog}^{pr.empl}$	$0.021 \beta_1 / SD_{cog}$	Goodman et al. (2015) social	Standard deviation increase in IQ test score raises probability being employed by 2.1%; SD_{cog} – standard deviation of the cognitive skills measure in the relevant age-sex subgroup of our simulation.

Table A4. Modelled Variables and Controls

Continued

Table A4. Continued

Y	Effect parameter	Method	Parameter reported	Explanatory variables (X) in the modelling equation												
				Cond. prob.	CD	Cog. skills	Education	Smoking	smoking	Depression	CHD	Employment	Prison care	Poverty	Income	Age
CD	Logit;	OR;		✓												
Depressed	Unemployment	Logit;	OR;	*										*	✓	✓
	Poverty	Logit;	OR;	*		(✓)										
	Smoking	Logit;	OR;				✓						*			
CHD	Poverty	Logit;	OR;					✓								
	Depression	Estimation of standardised mortality ratios;	Age standardised mortality ratio;													
Mortality	CHD	Estimation of dying probability from CHD;	Probability of dying from CHD;													
ECONOMIC																
	Conduct problems	Probit;	AME;	✓				✓								
	Cognitive skills	Probit;	AME;	✓												
	Education	Regression based linear matching;	Regression coefficient;	*				✓	*							
	Earnings															

Note: ✓ – variable X is included in the modelling equation for Y, as well as was controlled for in the literature; (✓) – variable X is not included in the modelling equation for Y, but indirectly influences Y through the other LifeSim equations, as well as was controlled for in the literature; * – variable X is included in the modelling equation for Y, but was not controlled for in the literature; AME – average marginal effects; OLS – ordinary least squares. Other abbreviations: AME – average marginal effects, OR – odds ratio, cond. prob. – conduct problems, CD – conduct disorder, teen. smoking – teenage smoking, CHD – coronary heart disease, res. care – residential care.

Table A5. Mortality from Coronary Heart Disease

	Sex	Age band					
		16–24	25–34	35–44	45–54	55–64	65–74
Mortality, %	male	0.19	1.24	2.81	1.83	1.61	2.07
	female	0.06	0.49	1.26	1.13	1.43	1.95
							8.82

Note: Estimated mortality from coronary heart disease (CHD) among people diagnosed with CHD. These estimates are used to model the parameter $\beta_{chd}^{pr,dead}$ in equation (18) and **Table A3**.

A.3. Specification

We present the full specification of the modelling equations which follows the structure outlined in **Table 2** in the main text. This material should be used together with **Table A1**, which clarifies the notation, as well as **Table A2** in the main text, which specifies the target data sources, and **Table A3** and **Table A4**, which specify the parameters, and details about their sources.

A.3.1. Skills Outcomes

Conduct Problems. Modelled using SDQ conduct problems scale data from the MCS.

$$\left\{ \begin{array}{ll} cp_{i,age} = sdq.cp.MCS2_i & \text{if } age \leq 4; \\ cp_{i,age} = sdq.cp.MCS3_i & \text{if } age \in [5, 6]; \\ cp_{i,age} = sdq.cp.MCS4_i & \text{if } age \in [7, 10]; \\ cp_{i,age} = sdq.cp.MCS5_i & \text{if } age \in [11, 13]; \\ cp_{i,age} = sdq.cp.MCS6_i & \text{if } age \in [14, 18]; \\ cp_{i,age} = n/a & \text{if } age \geq 19. \end{array} \right. \quad (7)$$

Impact of Problems. Modelled using SDQ impact supplement data from the MCS.

$$\left\{ \begin{array}{ll} ip_{i,age} = sdq.ip.MCS2_i & \text{if } age \leq 4; \\ ip_{i,age} = sdq.ip.MCS3_i & \text{if } age \in [5, 6]; \\ ip_{i,age} = sdq.ip.MCS4_i & \text{if } age \in [7, 18]; \\ ip_{i,age} = n/a & \text{if } age \geq 19. \end{array} \right. \quad (8)$$

Cognitive Skills. Modelled using principal component analysis to extract a common factor from the various cognitive skills measures disseminated by the MCS, following **Jones and Schoon (2008)** standardised with a mean of 1.00 and standard deviation of 0.15.

$$\left\{ \begin{array}{ll} cog_{i,age} = cog.MCS2_i & \text{if } age \leq 4; \\ cog_{i,age} = cog.MCS3_i & \text{if } age \in [5, 6]; \\ cog_{i,age} = cog.MCS4_i & \text{if } age \in [7, 10]; \\ cog_{i,age} = cog.MCS5_i & \text{if } age \in [11, 13]; \\ cog_{i,age} = cog.MCS6_i & \text{if } age \in [14, 18]; \\ cog_{i,age} = n/a & \text{if } age \geq 19. \end{array} \right. \quad (9)$$

A.3.2. Social Outcomes

Childhood Conduct Disorder. Modelled using the predictive algorithm by **Goodman et al. (2003)** and **Goodman et al. (2000)**.

$$\begin{cases} pr.cd_{i,age} = 0.61 & \text{if } age \in [5, 18] \& cp_{i,age} \geq 5 \& ip_{i,age} \geq 2 \\ pr.cd_{i,age} = 0.31 & \text{if } age \in [5, 18] \& cp_{i,age} \geq 4; \\ pr.cd_{i,age} = 0.06 & \text{if } age \in [5, 18] \& cp_{i,age} < 4; \\ pr.cd_{i,age} = n/a & \text{if } age \in [0, 4] \text{ or } age > 18. \end{cases} \quad (10)$$

Education (University Degree). We model the probability of obtaining a university degree at age 19:²

$$\begin{cases} pr.edu_{i,age} = \max[0, \min[1, \overline{edu}[\cdot] + \beta_{cog}^{pr.edu} (cog_{i,age-1} - \overline{cog}[\cdot]) + \\ + \beta_{cp}^{pr.edu} (10 - cp_{i,age-1} + \overline{cp}[\cdot]) + \\ + \beta_{dep}^{pr.edu} (dep_{i,age-1} - \overline{dep}[\cdot])]] & \text{if } age = 19; \\ pr.edu_{i,age} = n/a & \text{if } age \neq 19. \end{cases} \quad (11)$$

where $\overline{edu}[\cdot] \equiv \overline{edu}[age_i, sex]$, $\overline{cog}[\cdot] \equiv \overline{cog}[age_i - 1, sex]$, $\overline{cp}[\cdot] \equiv \overline{cp}[age_i - 1, sex]$, $\overline{dep}[\cdot] \equiv \overline{dep}[age_i - 1, sex]$

Unemployment/Employment. During 'working years' we model the individual probability of being employed; if individual is in prison, he/she is not employed by definition and this probability is zero.

$$\begin{cases} pr.empl_{i,age} = n/a & \text{if } age \in [0, 18] \text{ or } age \geq 70; \\ pr.empl_{i,age} = 0 & \text{if } pris_{i,age} = 1; \\ pr.empl_{i,age} = \max[0, \min[1, \overline{empl}[\cdot] + \\ + \beta_{cp}^{pr.employ} (cp_{i,age-1} - \overline{cp}[\cdot]) + \\ + \beta_{cog}^{pr.employ} (cog_{i,age-1} - \overline{cog}[\cdot])]] & \text{if } age = 19; \\ pr.empl_{i,age} = \max[0, \min[1, pr.empl_{i,age-1} + trend.empl[\cdot]]] & \text{if } age \in [20, 69]. \end{cases} \quad (12)$$

where $\overline{empl}[\cdot] \equiv \overline{empl}[age_i, sex]$, $\overline{cp}[\cdot] \equiv \overline{cp}[age_i - 1, sex]$, $\overline{cog}[\cdot] \equiv \overline{cog}[age_i - 1, sex]$.

Poverty. We model poverty as an indicator when individual consumption level falls below the absolute poverty line, 60% median equivalised household income in the UK in year 2011, which we set at £14,637 (Office for National Statistics).

Prison. During 'working years', individuals can go to prison, so we model the probability of being in prison. Imprisoned individuals are assumed to be unemployed and do not receive any salary; they are assumed to consume at a level equivalent to the state-subsidised minimum, which is subsidised by their own wealth (if sufficiently wealthy) or the state.

2. We assume that whether an individual obtains a university degree is determined at age 19.

$$\left\{ \begin{array}{ll} pr.pris_{i,age} = n/a & \text{if } age \in [0, 18] \text{ or } age \geq 70; \\ pr.pris_{i,age} = \max[0, \min[1, \overline{pris}[.] + \\ + \beta_{cd}^{pr.pris} (cd_{i,age-1} - \overline{cd}[.]) + \beta_{dep}^{pr.pris} (dep_{i,age-1} - \overline{dep}[.])] \\ \quad \text{if } age = 19; \\ pr.pris_{i,age} = \max[0, \min[1, \\ pr.pris_{i,age-1} + \beta_{dep}^{pr.pris} \Delta dep_{i,age-1}]] & \text{if } age \in [20, 69]. \end{array} \right. \quad (13)$$

where $\overline{pris}[.] \equiv \overline{pris}[age_i, sex_i]$, $\overline{cd}[.] \equiv \overline{cd}[age_i - 1, sex_i]$ and $\overline{dep}[.] \equiv \overline{dep}[age_i - 1, sex_i]$.

Residential Care. During 'retirement', individuals can live in residential care home, so we model the probability of living in a care home. We assume that individuals cover their care home cost (denoted *care.cost*, see **Table A6**), if they have sufficient resources to do so; otherwise, the state subsidises their care home cost.

$$\left\{ \begin{array}{ll} pr.care_{i,age} = n/a & \text{if } age \leq 69; \\ pr.care_{i,age} = \max[0, \min[1, \overline{care}[.] + \\ + \beta_{dep}^{pr.care} (dep_{i,age} - \overline{dep}[.])] & \text{if } age = 70; \\ pr.care_{i,age} = \max[0, \min[1, pr.care_{i,age-1} + \beta_{dep}^{pr.care}] \\ \quad \text{if } age > 70. \end{array} \right. \quad (14)$$

where $\overline{care}[.] \equiv \overline{care}[age_i, sex_i]$ and $\overline{dep}[.] \equiv \overline{dep}[age_i, sex_i]$.

A.3.3. Health Outcomes

Smoking.

$$\left\{ \begin{array}{ll} pr.sm_{i,age} = n/a & \text{if } age \in [0, 18]; \\ pr.sm_{i,age} = \max[0, \min[1, \left(\left(\frac{\overline{sm}[.]}{(1-\overline{sm}[.])} \exp \Phi \right)^{-1} + 1 \right)^{-1} + \\ + \beta_{pris}^{pr.sm} (pris_{i,age} - \overline{pris}[.])] & \text{if } age = 19; \\ pr.sm_{i,age} = \max[0, \min[1, \\ \left(\frac{1}{\frac{pr.sm_{i,age-1}}{(1-pr.sm_{i,age-1})} \exp (\beta_{pov}^{pr.sm} \Delta pov_{i,age} + \beta_{dep}^{pr.sm} \Delta dep_{i,age})} + 1 \right)^{-1} + \\ + \beta_{pris}^{pr.sm} \Delta pris_{i,age} + trend.\overline{sm}[.]] & \text{if } age \in [19, 69]; \\ pr.sm_{i,age} = \max[0, \min[1, \\ \left(\frac{1}{\frac{pr.sm_{i,age-1}}{(1-pr.sm_{i,age-1})} \exp (\beta_{pov}^{pr.sm} \Delta pov_{i,age} + \beta_{dep}^{pr.sm} \Delta dep_{i,age})} + 1 \right)^{-1} + \\ + trend.\overline{sm}[.]] & \text{if } age \geq 70. \end{array} \right. \quad (15)$$

where

$\Phi = \beta_{teen.sm}^{pr.sm} (sm_{i,14} - \overline{sm}[14, sex]) + \beta_{pov}^{pr.sm} (pov_{i,age-1} - \overline{pov}[.]) + \beta_{edu}^{pr.sm} (edu_{i,age} - \overline{edu}[.]) + \beta_{dep}^{pr.sm} (dep_{i,age-1} - \overline{dep}[.])$,
and $\overline{sm}[.] \equiv \overline{sm}[age, sex]$, $\overline{pris}[.] \equiv \overline{pris}[age, sex]$, $\overline{pov}[.] \equiv \overline{pov}[age - 1, sex]$, $\overline{edu}[.] \equiv \overline{edu}[age, sex]$,
 $\overline{dep}[.] \equiv \overline{dep}[age - 1, sex]$

Depression.

$$\left\{
 \begin{aligned}
 pr.dep_{i,age} &= n/a \quad \text{if } age \leq 4; \\
 pr.dep_{i,age} &= \max[0, \min[1, \\
 &\quad \left(\frac{1}{\frac{\overline{dep}[age, sex]}{1 - dep[age, sex]} \exp(\beta_{cd}^{pr.dep} (cd_{i,age-1} - \overline{cd}[age-1, sex])) + 1} \right)^{-1}] \\
 &\quad \text{if } age = 5; \\
 pr.dep_{i,age} &= \max[0, \min[1, \\
 &\quad \left(\frac{1}{\frac{pr.dep_{i,age-1}}{1 - pr.dep_{i,age-1}} \beta_{cd}^{pr.dep} \exp(\beta_{cd}^{pr.dep} \Delta cd_{i,age-1}) + 1} \right)^{-1} + trend.\overline{dep}[age, sex]]] \\
 &\quad \text{if } age \in [6, 18]; \\
 pr.dep_{i,age} &= \max[0, \min[1, \\
 &\quad \left(\frac{1}{\frac{pr.dep_{i,age-1}}{1 - pr.dep_{i,age-1}} \exp(-\beta_{unempl}^{pr.dep} \Delta emp_{i,age-1} + \beta_{pov}^{pr.dep} \Delta pov_{i,age-1}) + 1} \right)^{-1} + \\
 &\quad + trend.\overline{dep}[age, sex]]] \\
 &\quad \text{if } age \in [19, 69] \\
 pr.dep_{i,age} &= pr.dep_{i,age-1} \quad \text{if } age \geq 70.
 \end{aligned}
 \right. \tag{16}$$

Coronary Heart Disease.

$$\left\{
 \begin{aligned}
 pr.chd_{i,age} &= n/a \quad \text{if } age \in [0, 18]; \\
 pr.chd_{i,age} &= \max[0, \min[1, \\
 &\quad \left(\frac{1}{\frac{\overline{chd}[\cdot]}{1 - chd[\cdot]} \exp(\beta_{sm}^{pr.chd} (sm_{i,age-1} - \overline{sm}[\cdot]) + \beta_{pov}^{pr.chd} (pov_{i,age-1} - \overline{pov}[\cdot])) + 1} \right)^{-1}] \\
 &\quad \text{if } age = 19; \\
 pr.chd_{i,age} &= \max[0, \min[1, \\
 &\quad \left(\frac{1}{\frac{pr.chd_{i,age-1}}{1 - pr.chd_{i,age-1}} \exp(\beta_{sm}^{pr.chd} \Delta sm_{i,age} + \beta_{pov}^{pr.chd} \Delta pov_{i,age}) + 1} \right)^{-1}] + \\
 &\quad + trend.\overline{chd}[\cdot] \quad \text{if } age \geq 20.
 \end{aligned}
 \right. \tag{17}$$

where $\overline{chd}[\cdot] \equiv \overline{chd}[age_i, sex_i]$,

Mortality.

$$\left\{
 \begin{aligned}
 pr.dead_{i,age} &= \overline{dead}[\cdot] \quad \text{if } age \in [0, 4]; \\
 pr.dead_{i,age} &= \max[0, \min[1, \overline{dead}[\cdot] (1 + (smr_{dep} - 1) dep_{i,age})]] \\
 &\quad \text{if } age \in [5, 18]; \\
 pr.dead_{i,age} &= \max[0, \min[1, \overline{dead}[\cdot] (1 + (smr_{dep} - 1) dep_{i,age}) + \\
 &\quad + \beta_{chd}^{pr.dead} (chd_{i,age} - \overline{chd}[\cdot])] \quad \text{if } age > 18.
 \end{aligned}
 \right. \tag{18}$$

where $\overline{dead}[\cdot] \equiv \overline{dead}[age_i, sex_i, sep_i]$ and $\overline{chd}[\cdot] \equiv \overline{chd}[age_i, sex_i, sep_i]$.

A.3.4. Economic Outcomes

Earnings from Employment. We model the gross annual salary for people who are employed.

$$\left\{ \begin{array}{ll} earn_{i,age} = n/a & \text{if } age \in [0, 18] \text{ or } age \geq 70; \\ earn_{i,age} = 0 & \text{if } age \in [19, 70] \& empl_{i,age} = 0; \\ earn_{i,19} = \max[0, & \\ & (\bar{earn}[\cdot] + sd.\bar{earn}[\cdot]) \exp(\beta_{cp}^{earn} (10 - cp_{i,age-1} + \bar{cp}[\cdot]) + \\ & + \beta_{cog}^{earn} (cog_{i,age-1} - \bar{cog}[\cdot]) + \beta_{edu}^{earn} (edu_{i,age} - \bar{edu}[\cdot]))] \\ & \text{if } age = 19 \& employ_{i,19} = 1; \\ earn_{i,age} = \max[0, & trend.\bar{earn}[\cdot]] \\ & \text{if } age \in [20, 69] \& empl_{i,age} = 1. \end{array} \right. \quad (19)$$

where ' h ' is years since individual was last employed, or $age - 19$, if individual was never employed (in this case we use the value of individual's potential earnings at 19); $\bar{earn}[\cdot] \equiv \bar{earn}[age_i, sex_i]$, $\bar{cp}[\cdot] \equiv \bar{cp}[age_i - 1, sex_i]$, $\bar{cog}[\cdot] \equiv \bar{cog}[age_i - 1, sex_i]$ and $\bar{edu}[\cdot] \equiv \bar{edu}[age_i, sex_i]$.

Earnings from Interest (Interest). See details in the section below on modelling "Wealth".

Pension. Individuals are assumed to receive equivalent to the new basic UK state pension, if they have been employed for at least 10 years.³

Savings. During 'working-years', some individuals save part of their annual net income (i.e. earnings from employment and interest, net of tax). It is assumed that individuals save as much as they can, given that (a) they can sustain at least the previous year's consumption level and (b) their total annual savings do not exceed 16% of their annual net income.⁴ This means that individuals who experience a negative income shock, reduce their savings relative to previous year to sustain their consumption level; individuals who experience a positive income shock and can afford to consume at least the level of government subsidised minimum 'min.cons' – save a positive amount up to 16% net income;⁵ finally, individuals who experience a high positive income shock, and who could potentially afford saving above 16% while still sustaining previous year's consumption – consume anything excess of 16% of their net disposable income (instead of saving it).

$$\left\{ \begin{array}{ll} sav_{i,age} = 0 & \text{if } age \in [0, 18] \text{ or } age \geq 70; \\ sav_{i,age} = \max[0, \min[& 0.16 \times (earn_{i,age} + int_{i,age} - tax_{i,age}), \\ & earn_{i,age} + int_{i,age} - tax_{i,age} - cons_{i,age-1}]] \\ & \text{if } age \in [19, 69]. \end{array} \right. \quad (20)$$

Family Wealth. At the age 19, individuals are assumed to inherit wealth from parents. Wealth generates annual interest, i.e. $int_{i,age} = wealth_{i,age-1} \times (1 + r)$, where r is the annual interest rate, which we set at 1%. During their 'working years' individuals accumulate additional wealth by saving, if they can afford to save. If individuals experience a negative income shock, they use their wealth to subsidise their consumption.

3. See details on UK new State Pension at <https://www.gov.uk/new-state-pension>.

4. This threshold is chosen as a maximum, informing from historical UK households savings ratios reported by ONS: <https://www.ons.gov.uk/economy/grossdomesticproductgdp/timeseries/dgd8/ukea>.

5. See the paragraph below about "Consumption" in Section A.3.5.

$$\left\{ \begin{array}{l} wealth_{i,age} = \max[0, par.wealth_{i,age-1} + int_{i,age} + \\ \quad + \min[0, par.inc_i - cons_{i,age}]] \quad \text{if } age \leq 18; \\ wealth_{i,age} = \max[0, wealth_{i,age-1} + \\ \quad + \min[sav_{i,age}, earn_{i,age} + int_{i,age} - tax_{i,age} - cons_{i,age}]] \\ \quad \text{if } age \in [19, 69]; \\ wealth_{i,age} = \max[0, wealth_{i,age-1} + int_{i,age} + pens_{i,age} - \\ \quad - cons_{i,age} - tax_{i,age} - care_{i,age} \times care.cost] \\ \quad \text{if } age \geq 70. \end{array} \right. \quad (21)$$

where $par.wealth$ – parental wealth and $par.inc$ – parental income, as given in the childhood dataset.

Taxes. Individuals pay annual taxes on their income, i.e. earnings from employment and interest, as well as pension. The individual tax rate is set according to the corresponding UK tax bracket.⁶

Benefits. Individuals receive benefits subsidised by the public budget ($benef_{i,age}$) to sustain the minimum consumption level of £10,000, whenever they cannot afford it from their own net income (parental income and interest during 'pre-school years' and 'school years', salary and interest during 'working years', and pension and interest during 'retirement') and wealth. During 'retirement', individuals also receive benefits when in care to cover the care home costs, when they do not have sufficient own resources to cover them.

A.3.5. Wellbeing Outcomes

Consumption. It is assumed that government subsidises consumption level of at least ' $min.cons$ ' (the state-subsidised minimum), in the case when individual cannot afford it given their income or wealth. We set $min.cons = £ 10,000$.

Up to age 18, individuals are assumed to consume the level of their household equivalised income ($par.inc$) as given in the childhood dataset, or the state-subsidised minimum.

During 'working years' individuals consume what is left of their income from employment and interest after tax and savings, or an amount equal to the state-subsidised minimum (this may be subsidised by state or own wealth, depending on whether individual has positive wealth). For more details, read about the savings equation above.

During 'retirement', individuals try to sustain their previous year's consumption level if they can afford it given their resources (i.e. net income from interest, state pension, their wealth and minus residential care home cost, if in care); if individuals cannot afford sustaining previous year's consumption level, then they consume the maximum amount that they can afford given their resources, or the state-subsidised minimum.

$$\left\{ \begin{array}{l} cons_{i,age} = \max[min.cons, par.inc_i] \quad \text{if } age \leq 18; \\ cons_{i,age} = \max[min.cons, earn_{i,age} + int_{i,age} - tax_{i,age} - \\ \quad - sav_{i,age}] \quad \text{if } age \in [19, 69]; \\ cons_{i,age} = \max[min.cons, \min[cons_{i,age-1}, wealth_{i,age-1} + \\ \quad + int_{i,age} - tax_{i,age} + pens_{i,age} - care_{i,age} \times care.cost]] \\ \quad \text{if } age \geq 70. \end{array} \right. \quad (22)$$

6. See UK income tax rates at <https://www.gov.uk/income-tax-rates>. We use the year 2018/19 rates, converted to year 2015/16 prices.

Health Quality. Health quality 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, $health_{i,age} = h(chd_{i,age}, dep_{i,age})$, where $h(\cdot)$ 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

$$h(\cdot) = \min[1, \max[0, \overline{health}[age, sex, sep] - (d(chd) \times chd_{i,age} + d(dep) \times dep_{i,age})]]$$

where $\overline{health}[age, sex, sep]$ is the average health quality in England by age, sex, and English IMD quantile 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)**).

A.3.6. Public Costs and Revenues

We model the costs associated with different outcomes, as summarised in **table A6**. We assume that the following outcomes incur costs to the public service: CHD, depression, other healthcare, conduct disorder, prison, residential care.

Table A6. Public Service Costs

Cost type	Components of the cost	Annual cost per person, £	Source
Healthcare: coronary heart disease ¹⁴	Direct health care cost; Informal care cost;	840; 1,173;	Liu et al. (2002) economic;
Healthcare: depression	Costs to the National Health Service, the Accident and Emergency department, other support services (average);	5,260;	McCrone et al. (2008) paying;
Other healthcare	Average English National Health Service healthcare spending in the financial year 2011/12 by age, sex and English neighbourhood deprivation quintile group;	see Asaria (2017);	Asaria (2017);
Conduct disorder	Cost to the National Health Service;	1,243 (age 5-10), 113 (age 11+);	Edwards et al. (2007) parenting; Scott et al. (2001) financial, cited by Bonin et al. (2011);
	Cost to the Social Services Department;	175 (age 5-10), 70 (age 11+);	Edwards et al. (2007) parenting; Romeo et al. (2006) economic, 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) parenting; Scott et al. (2001) financial, cited by Bonin et al. (2011);
	Cost to the voluntary Sector;	26;	Edwards et al. (2007) parenting, 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);

Continued

7. We use the average of the direct and informal care cost, to quantify the annual cost of a person with CHD.

Table A6. Continued

Cost type	Components of the cost	Annual cost per person, £	Source
	Unit costs of courts (per court event);	7,103;	
Residential care	Cost of residential home;	29,934;	Curtis and Burns (2017).

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