Human Capital Indicators consultation

Additional Information

3rd September 2019

Contents

Overview

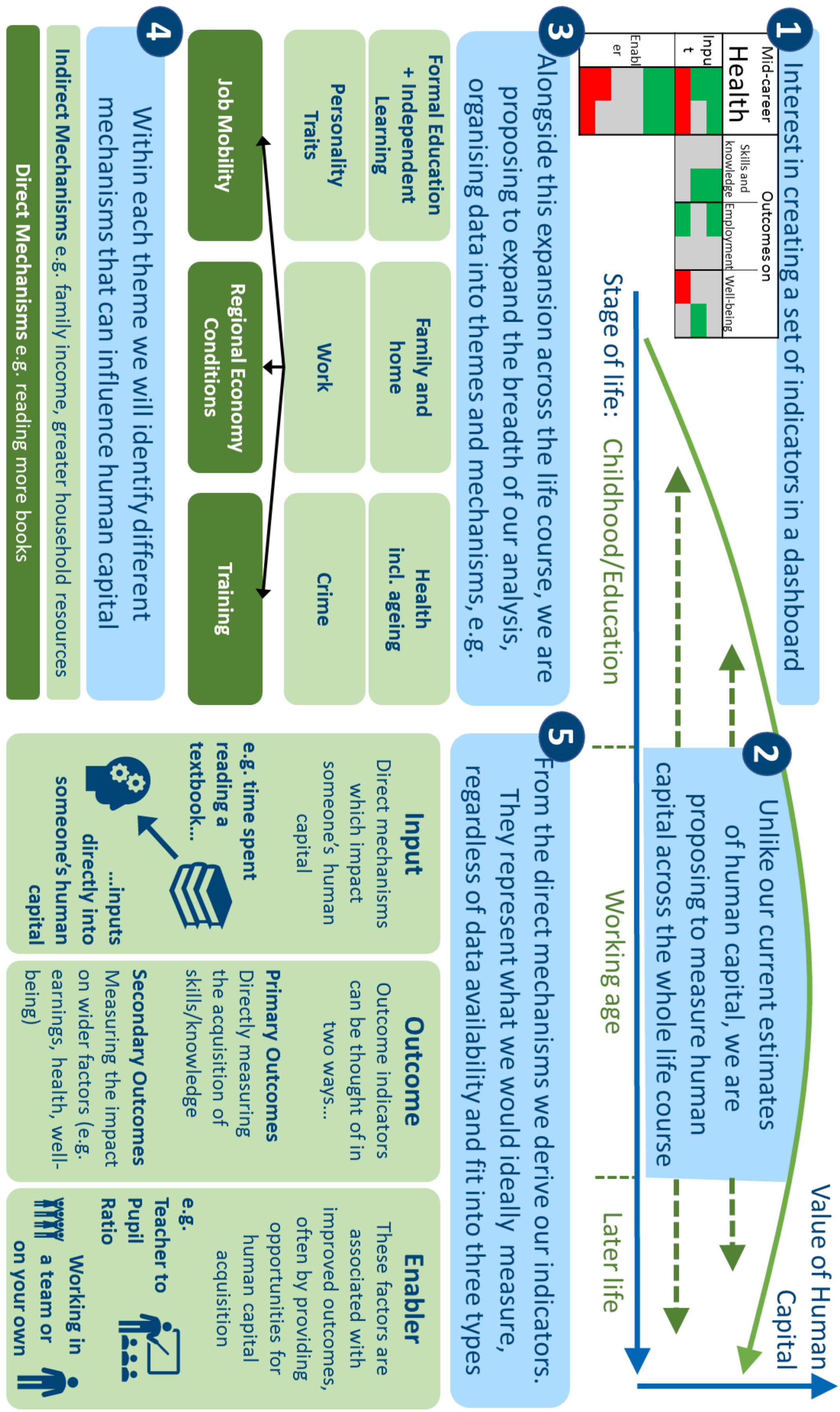
This annex has been published alongside the Human Capital Indicators Consultation. It provides additional information around our plans to review the way we measure human capital in the UK.

A copy of the accompanying consultation document can be found here:

https://consultations.ons.gov.uk/well-being-inequalities-sustainability-and-environment/indicator-based-approach-to-measuring-human-capita/supporting_documents/Human_capital_consultation_final.doc

The proposals we are looking to consult on can be seen in the diagram on the next page.

Figure 1 Summary of proposals



Phases of wider review of our measures of human capital

Table 1 Different phases of wider human capital developments

Phase stage (each stage would **likely take several months**)	Workstream 1 – consultations and engagement	Workstream 2 – sourcing data	Workstream 3 – deriving indicators	Workstream 4 – non-indicator-based workstreams relating to workplan
1	Consult on list of indicators Engagement day workshop discussing some of the approaches in more detail			
2	Start engagement on full list of skills, knowledge, competencies and attributes.	Source initial set of data for indicators (from available sources)		Publish latest human capital stock estimates up to year ending 2018, assess feasibility of making various improvements
3			Derive initial set, which may include proxy indicators. More focused on inputs and enablers	Investigate progression assumption of human capital stock estimates Start investigating conceptual research to incorporating human capital

				into a National Accounts framework, including consulting internationally on the topic
4		Categorically define data gaps + continue sourcing wider set of data, including outside of ONS		Process skills data from existing sources (e.g. vacancy data) to get a) skills stocks estimates b) skills supply and demand estimates
5			Derive larger set, including with new administrative data. More focused on outcome indicators	Incorporate some effects of health into human capital stocks model Investigate publishing more granular regional human capital stock data
6		Collect new data relating to skills and knowledge of individuals – new survey and/or linkage of existing datasets		
7			Deriving new set of indicators from new survey sources	

Interest in an indicator-based approach (additional information)

Who may be interested in an indicator-based approach to Human Capital?

It is envisaged that an indicator-based approach could be used by **central and local government** in reviewing the latest evidence for a specific policy area, or when appraising different policy options. A dashboard of indicators could feed into the estimates of costs and benefits resulting from different policy options. Outside of government, it is expected that an indicator-based approach would also be useful to **academics and thinktanks**. The greater breadth and depth of information proposed in this framework could support more detailed and innovative research and analysis, not only in relation to the value of education but on other topics too. The indicator dashboard could be used to monitor the different themes alongside each other, highlighting areas which might warrant further research and analysis.

Finally, a human capital dashboard could also be used to better inform decisions faced by **citizens**. This could be in the context of the family and parenting, looking at the outcomes of increased parental involvement with children. Similarly, citizens are also likely to be interested in how career decisions, such as moving jobs, could impact on things such as their earnings and well-being levels.

Where has an indicator approach been used elsewhere?

Other statistical releases at the ONS use a dashboard approach to visualise a key set of indicators which people can use to monitor progress and sustainability over-time.

Measures of National well-being

The Measures of National Well-being Dashboard presents the well-being of the UK in a set of 10 headline domains. Given the multi-dimensional aspect of well-being in the UK, these are presented alongside each other to allow comparisons. These measures include:

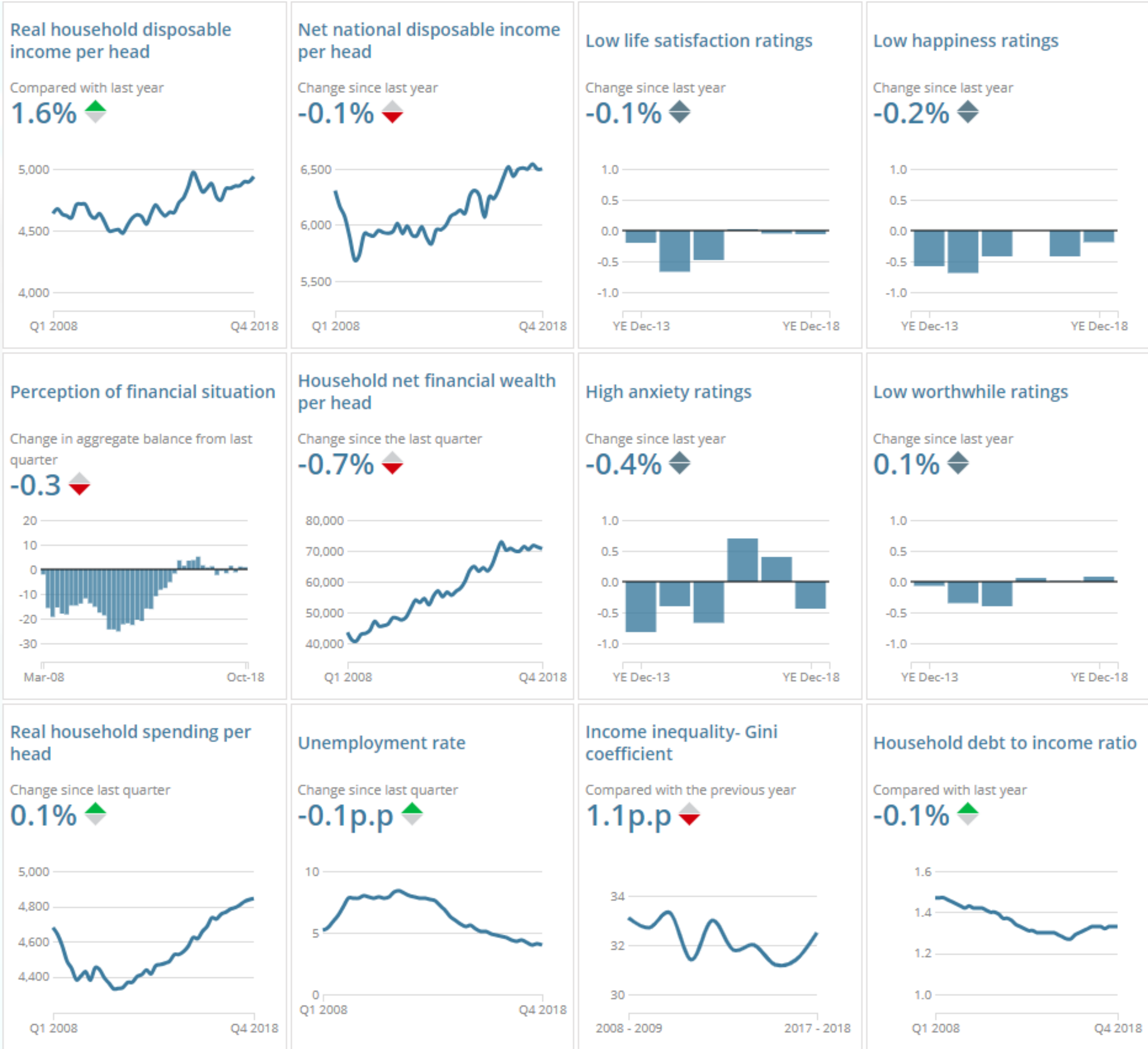
- Personal Well-being
- Our Relationships
- Health
- What we do
- Where we live
- Personal Finance
- Economy

- Education and Skills
- Governance
- Environment

The domains naturally lend themselves to analysis of short, medium and long-term trends in well-being. The personal well-being domain is shown below, as an example:

Personal and Economic well-being

Alongside the National Measures of well-being, ONS also publishes a quarterly dashboard of personal and economic well-being in the UK. This dashboard is designed to give a wider picture of recent trends in well-being than can be shown by traditional measures like GDP. For instance, GDP doesn't consider the distribution of income, meaning that changes in income inequality do not impact on the headline measure of economic growth. The dashboard provides supplementary information, such as on people's happiness and life satisfaction that can help policymakers to assess the wider impacts from economic and social policy. The dashboard is shown below:



Sustainable Development Goals

ONS also takes an indicator-based approach to [measuring sustainable development goals](#) (SDGs) in the UK. These SDGs are a set of ambitious goals and targets which forms part of agenda 2030 on sustainable development[1]. At present, 17 goals are presented, with 244 proposed indicators. Some of these indicators are already reported online, whilst

[1] https://sustainabledevelopment.un.org/post2015/transformingourworld

others are still in their inception.

Other Human Capital indicator approaches

Two other organisations' approaches to measuring and presenting human capital are worth highlighting here – the World Bank and the World Economic Forum.

The World Bank report has derived composite indices for most countries in the world in a comparable way. The data is from a range of sources and is designed to capture the amount of human capital a child born today could expect to attain by age 18. There are three components considered – survival, quality and quantity of education, and health. Given the focus is on developing countries, however, some of the indicators are less relevant for a UK context. Additionally, the breadth of each of the components is not as broad as the proposal laid out in this consultation. Finally, various important themes, such as family background and lifelong learning, are not considered by this methodology.

The World Economic Forum also published a global analysis of countries' human capital, in 2017. It presents a set of indicators relating to people's skills and knowledge, through 4 'elements' – capacity, development, deployment and know-how. Each of these is focused on outcomes throughout life, rather than considering the inputs into human capital, and is presented as a "distance to the ideal" or gap in human capital optimisation. The 4 elements each have a 25% weighting to get to an aggregate index to rank the different countries. The indicators focus on high-level education and labour market outcomes, not considering other important factors such as health and family background. Additionally, there is not consideration of wider factors (which we label as enabling indicators below), as well as the lack of input indicators to track.

What are the pros and cons to an Indicator-based approach to measuring Human Capital?

The indicator approach aims to present a range of information relevant to human capital in one place. The approach means that not all information presented needs to be able to be monetised or fully independent from other indicators.

The UNECE (2016) describe the approach as rich in data but lacking a common money metric. Table 1 summarises the pros and cons of this approach to measuring human capital.

Table 2 Pros and Cons of an indicator-based approach

Concept	Pros	Cons
Data detail	Rich data – allows a diverse range of indicators to be displayed in one place. Indicators may complement each other and guide researchers to explore relationships. If one indicator falls and another rises, this may show one subgroup of the population is at risk of falling behind in terms of the development of their human capital.	More difficult to track overall progress over time, or to measure sustainability.
Unit of measure	Not restricted to having monetary metrics only.	In some cases, policy makers may want to understand monetary benefits associated with human capital and that's harder to achieve using an indicator approach.

Coverage	Does not have to represent the entire population – indicators can be developed for population subgroups.	Some indicators may lack impact at a total population level as they may only reflect population subgroups.
Compatibility with other measures	Can complement the cost-based or income-based approaches to measuring human capital.	Harder to communicate a headline message to an audience than a single measure metric.
Breadth	May suit a wide range of user needs – people can select only those indicators relevant for them.	Interactions between different indicators difficult to portray
Data sources	If data sources cease to exist for certain indicators it does not have a detrimental impact on the rest of the indicators. This is less true if the indicators were presented as a composite index.	

What other ways can human capital be measured?

The UNECE guide to measuring human capital (2016) outlines three distinct methods for its measurement:

- cost-based approach,
- life-time earnings approach
- indicators approach.

The cost-based approach

According to guidance from the UNECE (2016) the cost-based approach looks to add up all the costs associated with providing education, whether formally or informally, and then apply a rate of depreciation to that value. This scope is more limited than the set of themes proposed within this consultation.

The lifetime income-based approach

The UNECE (2016) guidance goes on to describe the lifetime income-based approach, which is the approach already used by the ONS to value human capital. The approach works by summing discounted values of future income received by individuals over their lifetime. We are looking to expand this current methodology, as outlined in our workplan published last year.

Other non-UNECE methods

Alongside the three suggested methods, there are other ways of measuring human capital, which we also have plans to research and develop alongside the indicator-based approach, though at different time scales. These include, but are not limited to:

- experimental statistics integrating human capital into the national accounts framework, through a human capital satellite account;
- direct skills, knowledge, competencies and attributes measurement to work out human capital supply and demand. Here, supply and demand could be defined to relate to the labour market, or the education sector in terms of provision, or wider societal needs;
- through a series of articles highlighting relevant areas of user interest e.g. training, automation, the role of organisational structure and management, job mobility etc.

How the indicators may be visualised

How do we plan to present these indicators?

We would use a dashboard to provide a visual overview of a set of indicators that have a direct or indirect relationship with human capital. The human capital dashboard will allow users to monitor the progress of the knowledge, skills, competencies, and attributes of individuals, over different life-stages. Users would be able to see a range of inputs that have an impact on human capital as well as the outcomes of higher human capital. Further information on direct mechanisms and how we plan to measure these can be found in the *Types of Indicators (additional information)* section.

Given the breadth and categorisation of indicators, we are considering similar ways to visualise the data as the indicator approaches shown above.

Some examples are presented below, though the final approach will be determined by which indicators are calculated first:

Table 3 An example dashboard highlighting changes in input indicators of the primary-school life stage

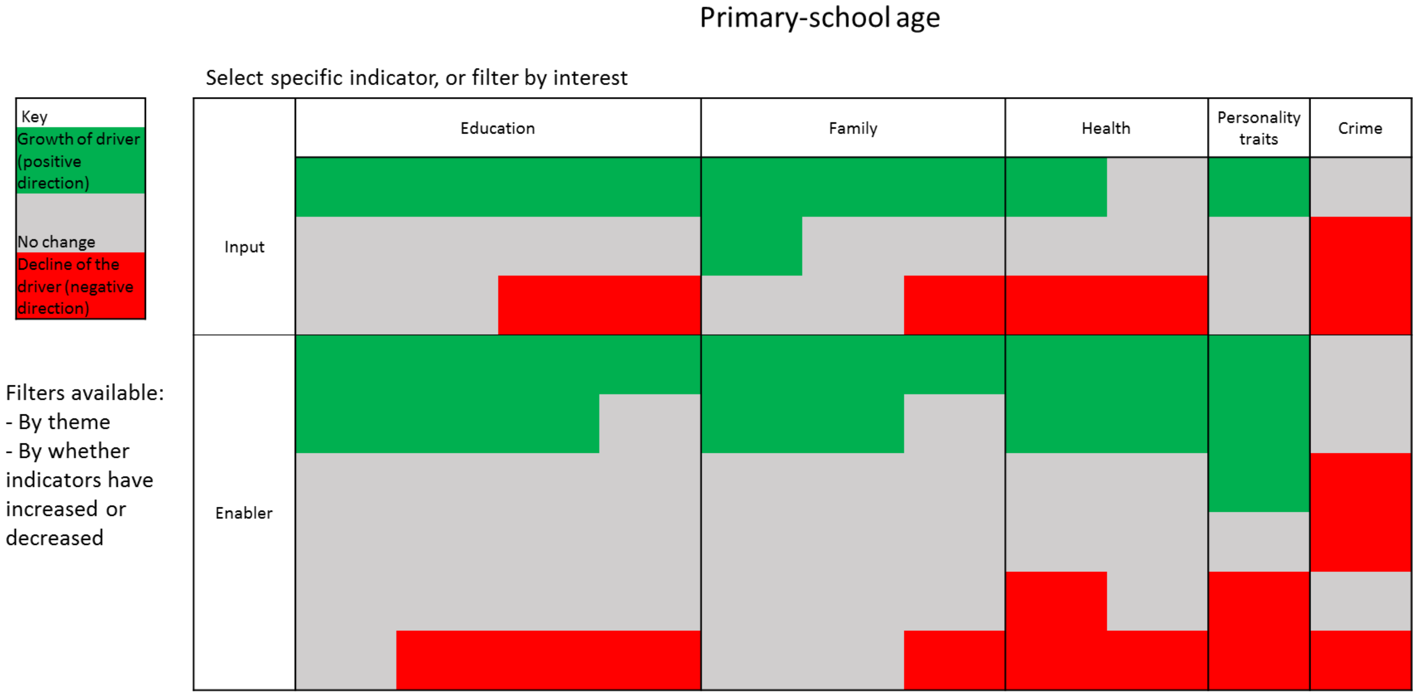
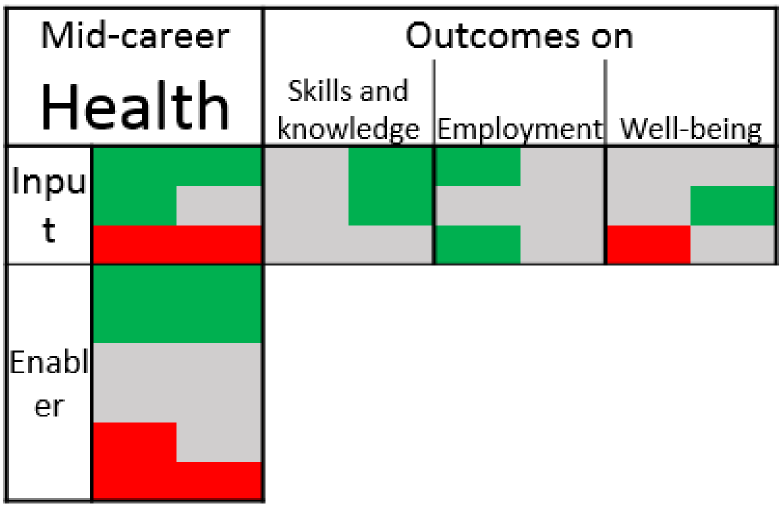


Table 4 An example dashboard for a specific theme and specific life stage



Beneath these higher-level dashboards, individual indicators can be monitored as a time series graph and can be split by different characteristics of the individuals, such as sex, socio-economic status etc, where the data allows.

Themes to measure human capital (additional information)

How have we identified the themes and factors to measure human capital?

To understand the broad factors that affect individual-level human capital, as well as data sources used and methods applied to make such inferences, we publish an evidence review alongside this annex. This was built upon from an initial literature review, commissioned by ONS to consider the latest evidence relevant to the UK, focusing where possible over the last 10 years.

The National Institute of Economic and Social Research (NIESR) provided a literature review in early 2019, focused on two aspects of human capital, namely the factors that affect human capital accumulation, and the determinants of individual earnings. The former focuses on factors that affect human capital as embodied more widely, representing educational attainment and skills acquisition, and the latter focuses on human capital as measured through labour market outcomes, particularly earnings.

The factors that affect human capital accumulation were initially found to be:

1. **Family background:** Among the many factors shown to contribute to human capital accumulation and adulthood labour market

outcomes, family background has been found to be the most important characteristic. Higher parental education, parental cognitive skills, family structure, financial constraints, cultural background, paternal/ maternal influences on education and ethnicity among others all have an impact on human capital accumulation. Also, there is some evidence of genetic differences having substantial impacts on resulting educational and labour outcomes.

2. **Health:** The role of health in human capital accumulation has been known to increase the productivity of workers and as a result, lead to an increase in earnings. Ageing is also likely to reduce productive human capital, along with people with poor self-reported health, which may lead to an increase in absenteeism (sick leave) or presenteeism (at work but not productive). This section mainly captures the important reverse causal effect, where those with higher human capital are more likely to engage in healthy behaviour, while healthy individuals also tend to have better outcomes.
3. **Job mobility:** Moving industry, occupation, or losing your job may lead to a loss of human capital. However, occupational change may have a positive impact on human capital accumulation if it’s a response to a job mismatch. Providing education and training to workers moving from one occupation to another may help to maintain existing skills.
4. **Crime:** this is seen as an enabling, or preventative factor, as it can reduce the potential for human capital accumulation through providing less opportunities and lower life-time earnings for those that commit crime. Those with a higher education and higher wages are less likely to commit crime as the opportunity cost of committing a crime is higher. In addition, family life, including parental management, family size and marital status, are strong predictors for children’s offending, and criminal participation.

The determinants of earnings, and their use in valuing human capital were found to be:

1. **Education:** as a determinant of earnings, degree subject and educational institution quality are all found to be important determinants of earnings. In addition, non-cognitive skills such as social skills and certain personality traits can also have an impact on earnings.
2. **Health:** there are negative effects on earnings for those that experience a health shock, which may prevent people from

working, reduce their productivity at work, or force them to leave the labour market.

3. **Job mobility**: those that change occupation tend to benefit from a change in wages, relative to those that do not experience such change. However, this depends on factors such as whether the move is to a similar or different occupation or industry, and different factors are implicated in the causes of this.
4. **Family background:** often measured as parental income, there is a strong relationship found between this and children's economic success. It is often considered to be an indicator of equality of opportunity, as higher parental income can be associated with higher investment in a children's human capital.
5. **On-the-job training:** the relationship with this and earnings is positive, where undertaking on-the-job training has been known to increase productivity and wages.

After this literature review, various sections were expanded and further investigated through engagements with several government departments, think tanks and academics, to expand the scope which is presented in the accompanying evidence review. Some themes were then grouped together or split out in a way we believe makes the indicators more relatable to each other, but we would be interested to hear if you think these can be categorised differently, or if any of the themes should further be expanded upon.

Using the themes to choose indicators

The evidence review captured sub-themes for every theme. For example, within the theme of health, there is a consideration of the effect ageing has on human capital development. Sometimes these sub-themes reflect mechanisms for how an individual's human capital can change. For example, moving jobs has been shown in some cases to improve people's skills and knowledge, while in others, such as if it is involuntary, this may lead to lower human capital. We term this an indirect mechanism, as there are more specific ways that ultimately increase or decrease someone's level of human capital.

From this, two methods were applied to derive indicators. Firstly, digging further into the literature and the references provided, we have proposed sets of direct mechanisms relating to each sub-theme, which are the drivers of changing someone's human capital. These we then convert to individual indicators. As outlined below, there are different types of indicators, and multiple ones can be associated with a direct mechanism.

The second method was to categorise direct mechanisms already explicit in the literature, and abstract to the more general indirect mechanism (or sub-theme). From this, new direct mechanisms can then be suggested, from which we derived other indicators.

For example, one of the indicators suggested in the evidence review within the education theme is mean faculty salary. Evidence shows that this could impact upon the educational outcomes of students. Hence, this was derived directly from the evidence review, and indirect mechanism was considered to belong to the "wider learning environment". This wider learning environment mechanism was then subsequently explored and led to the inclusion of indicators such as aggregate student intake and whether or not the child has undertaken a school move. The learning environment is seen to be an indirect mechanism which can affect the ability of an individual to accumulate skills and knowledge, which impacts more indirectly through the wider education environment.

Types of Indicators (additional information)

Input Indicators: Direct and Indirect Mechanisms

In considering the input indicators to human capital, we propose focusing on mechanisms which evidence shows can impact directly on individuals' human capital, rather than wider indirect mechanisms. Looking only at direct mechanisms for our input indicators allows a focus on the key aspects that matter for an individual's human capital development. It will allow us to interpret a change in a certain input as either directly improving or decreasing the value of someone's human capital or the human capital of people in the UK in a certain life stage.

Another key reason why direct mechanisms are considered is because various indirect factors can influence human capital acquisition through the same direct mechanism. For example, higher parental income may encourage parents to invest more resources in the child's education through hiring tutors, which can lead to an improvement in the child's development. This could also occur from more parental time, more time with grandparents, as well as other mechanisms. Separately, longer hours at school could also impact on the amount of time children spend talking to adults, as would various extra-curricular activities they could get involved in. Hence, measuring the amount of time children spend in conversations with adults as a direct mechanism, would remove the need to specify the means by which this may occur through multiple indirect mechanisms.

Outcome indicators (Primary and Secondary)

We propose to split human capital outcomes into two levels: primary and secondary. Primary outcomes capture the change in skills and knowledge whereas secondary outcomes look at the wider individual impacts of such a change in skills and knowledge. This will give a framework to capture the benefit to skills and knowledge development, as well as wider impacts of value to the individual, such as on their earnings, health or well-being. It is important to state that, at the secondary level, these outcomes may also interact with each other throughout the life-course. For example, better health may also lead to better well-being).

We also recognise there are other wider outcomes, discussed below, though we propose to keep them out of scope for now.

Primary outcomes (human capital skills and knowledge)

ONS plan to carry out further research to ensure the scope of skills, knowledge competencies and attributes are diverse and robust while, at the same time, driven by user needs and previous work to define human capital. Ahead of this we will use a working definition of skills and knowledge from previous work in the area. For example, the skills and knowledge developed as a primary outcome can initially be defined in accordance with work carried out by the OECD in their report 'The Well-being of Nations: The Role of Human and Social Capital' (2001). The following skills are listed:

- Communication (including foreign language competence in each of the items)
    - Listening
    - Speaking
    - Reading
    - Writing
- Numeracy
- Intra-personal skills – Motivation/perseverance – "Learning to learn" and self-discipline (including self-directed learning strategies) – Capacity to make judgements based on a relevant set of ethical values and goals in life
- Inter-personal skills – Teamwork – Leadership
- Other skills and attributes (relevant to many areas above) – Facility in using information and communications technology – Tacit knowledge – Problem-solving (also embedded in other types of skills) – Physical attributes and dexterity

Although wide ranging, the definition of skills and knowledge is still noted by OECD in the same report as ‘some key skills and personal attributes relevant to human capital’, which suggests that this list could be built upon in future. For example, this can look to incorporate work done by organisations such as O-Net and Nesta on skills, competencies and knowledge classifications for jobs.

Secondary outcomes (impact from higher/lower skills and knowledge)

Based on user input to date and in line with the OECD (2001) guidance, we plan to focus on four types of secondary outcomes: earnings/labour remuneration; education; health; subjective well-being.

Impact on earnings/total labour remuneration

Skills and knowledge are often considered as a significant determinant of labour market outcomes. For instance, learning at work, either through courses or other means, may build an individual’s skills and knowledge, which could allow them to produce either better or an improved range of goods and services. This may then be reflected in the individual’s earnings. A secondary outcome indicator would try to measure the change in future total labour remuneration brought about by a change in skills and knowledge.

This secondary impact will also consider wider labour market opportunities, from hourly wage effects to employment aspects such as getting a job after improving your skills and knowledge while unemployed or retaining a job in the future once employed.

Impact on educational attainment

During childhood, measuring labour market outcomes would not be a suitable measure of the impacts of changes in skills and knowledge. This is because most individuals in this life stage do not participate in the labour market, and so either earnings would have to be projected, based on current labour market returns who have gone through a different schooling system, or can only relate to drivers further in the past, when the individuals were children.

As a child, the main indicator of your skills and knowledge are the qualifications awarded by schools and other educational institutions.

Impact on individual health

As captured in the accompanying evidence review, the relationship between health and human capital goes both ways. As there is a potential reverse causality issue, we will try and capture the impact higher skills have on people's general health. There are, broadly speaking, three areas through which human capital and health interact.

Perhaps the most obvious of these is where individuals learn to have a healthier lifestyle, for example by learning about the human body, nutrition or exercise. However, this must be acted upon to get the benefit of this extra knowledge.

Secondly, an individual's occupation will vary along with their human capital which could in-turn impact on their health. If you are a sports coach for example, you may have a healthier average day than an office worker who is mostly sedentary during their working time. Hazardous working environments may also lead to health implications should someone inhale dangerous particles or have a serious accident during their working time.

Finally, depending on your level of human capital, you are likely to have a different income, which can affect your health in multiple ways. This could be from your neighbourhood, your lifestyle, and cultural and consumption habits.

Impact on subjective well-being

Being satisfied with life, being happy, feeling that life is worthwhile and feeling less anxious are all aspects of what might define a higher state of subjective well-being. Reporting higher subjective well-being may happen for many reasons, but often those factors may be related to the skills and knowledge shared within society. An individual may directly improve their own well-being, for example through having better problem-solving skills enabling them to be more organised in their daily life, and hence less stressed.

They can also improve someone else's well-being. We will look to focus on individual impacts but recognise there can be wider effects. For example, the knowledge and skills possessed by entertainers, actors, musicians and celebrities can impact on the well-being of their audience.

Wider outcomes

As highlighted above, our proposed focus is on outcomes for individuals, whether their skills and knowledge, or secondary impacts such as on their health. However, there is evidence that human capital has an even wider impact. We propose to keep this out of scope, but still want to recognise it.

- Individually, human capital has been found to be associated with improved civic participation and citizen engagement.
- At the community level, increased human capital is associated with improved social capital such as higher trust, as well as more volunteering and participating in the community.
- At the firm and industrial sector level, employees' human capital has been associated with productivity gains, though there is a strong relationship between these and how firms are organised.
- At the national level, human capital is also associated with higher economic growth.

Enabling factors

As outlined in the consultation document, certain factors are known to be associated with improved outcomes, but their impact is on the opportunities they provide for individuals to improve rather than directly causing an improvement. An example can be wider environmental factors, or local economy labour market conditions, which would direct the availability people have in developing and using their skills to their full potential. These aspects still matter and contextualise the benefits of certain factors. It is clear these are still important, but benefits cannot be as easily attributed. Hence, we propose measuring these as enabling factors.

Focussing on the individual rather than group level

The proposed indicators largely focus on impacts and factors at the individual level, as opposed to at a group level.

Individual Impacts

For example, someone may attend a training course through work which improves their knowledge in that topic. This is the mechanism which we are attempting to capture in our proposed indicator on in-work training. However, there may also be wider effects on other people the individual knows, but these would only be transmitted through other mechanisms (such as sharing best practice or collaborating at work) so the benefit of improved skills should be attributed to these mechanisms.

Additionally, there may be broader impacts, such as wider societal benefits from improved human capital, or on the productivity of the organisation employing the individual. These and other spill-over effects tend to be more difficult to attribute to the initial cause, as lots of other factors also influence the outcome. Hence, we have generally kept the scope in this consultation to only capturing the direct benefits on individuals, rather than wider spill-over effects, as much as these can be distinguished from each other.

What indicators are we planning to include?

The indicators we plan to include in our indicator-based approach can be found in the attached spreadsheet.

How we plan to derive these indicators (additional information)

Data Sources

Our focus in considering relevant indicators has been to develop a conceptually complete and coherent set of measures relevant to human capital development. In practice, this means that not all of the proposed indicators are necessarily measurable by data that are currently available to ONS. However, we believe it is important to focus on an optimal set of indicators, rather than necessarily being limited to what can currently be measured. This is to provide a more complete picture of what impacts on individuals' human capital, as well as to identify data gaps that may be important for our users to fill with new sources.

With this in mind, we have considered which datasets could be relevant in deriving the different indicators. This is not limited to social and business surveys, but also extends to administrative data, which can provide broader coverage, as well as commercial data. A list of considered datasets is shown below, linked to themes which they may be most relevant in capturing:

Dataset	**Themes appropriate**
National Pupil Database (equivalents for devolved administrations)	Education, learning and training
Longitudinal Education Outcomes dataset	Education, learning and training, Work
Cohort Studies (e.g. Millennium Cohort Study)	Family and home, personality traits
Understanding Society (and previous panel surveys)	Family and home, personality traits
Hospital Episode Statistics	Health
Time Use Surveys	Learning and training, family and home, health

Annual Population Survey and Labour Force Survey (including longitudinal versions)	Work, Health
Survey of Adult Skills, Programme for International Student Assessment	Work, education
Census	Family and home
National Child Measurement Programme	Health
Workplace Employment Study	Work
Commercial Vacancy data	Work
Employer Skills Survey	Work
Reoffending Statistics	Crime

As we plan to consider a wide range of indicators, we expect to have data gaps in relation to sufficiently capturing skills, knowledge, competencies and attributes of individuals directly throughout their lives. This is key in capturing primary outcomes. We also think there will be data gaps in accurately measuring personality traits and certain health conditions and both of their impacts on labour market and education outcomes, as well as gaps in timely indicators relating to time spent doing specific skill-accumulating activities. Finally, we think there will be more data gaps with factors particularly in the life stages before children attend school, though these are known to have a very large impact on children's outcomes throughout their whole lives.

Methods for measuring indicators

When we are trying to capture the outcome from having more knowledge, skills, competencies and attributes across the different life stages of an individual's life, as well as secondary outcomes, we propose to use more sophisticated statistical inference methods and techniques. This is so that benefits from inputs can more accurately be assigned to certain factors.

Causal relationships

Casual relationships are when one variable has a direct influence on another. To produce these casual relationships, it is important we use the correct model and techniques, while controlling for factors that have a direct impact on our outcome variable. For example, having a learning disability could lead you to accumulate less skills and knowledge (outcome), but the exact relationship of these factors can only come about if we control your other health factors and learning environment. If we do not control for these in our model, the relationship between learning disabilities and accumulating skills and knowledge may be over-estimated.

When modelling our outcomes, we have considered potential issues that may arise. The use of the ordinary least squares (OLS) approach may be useful to obtain a correlation between one variable and another. However, for our work, this type of modelling may not allow us to obtain causal estimates. Much of the human capital literature emphasises accurately estimating the benefit of a certain driver where the data cannot measure all drivers. Additionally, the non-random nature of the data, where outcomes earlier in life determine who is selected into data, may affect the validity of the models. To overcome these issues, we have researched several potential models and techniques that can help us to obtain causal estimates that may alleviate some of these concerns.

Models across life stages

Our approach to producing these outcome indicators involves capturing the impact of factors across the different life-stages of the individual and controlling for them in the next life stage. The life-stages of an individual from childhood (pre-school, primary, secondary), young adult (college, university), career (early, late) and retirement will have different factors as well as, in some cases, similar factors impacting on their human capital and outcomes.

Using one example of our approach, childhood factors, such as adverse childhood events and family structure, can have an impact on a child's skill development. Using this relationship between childhood factors and skills, we can take the predicted values of our childhood models and use these values as a control variable in the working adult regression model. As we want to see the impact on an outcome such as earnings, from adult factors, by controlling for childhood factors, we can try to obtain a causal relationship between our interested adult factor on our outcome. It is important to emphasise that while much of the human capital literature looks at the impact from an investment or change in an individual's human capital on a labour market outcome, we are looking go beyond these outcomes and consider the wider impact on skills and well-being.

Model selection criteria

When selecting the appropriate model, it is important to recognise their limitations when trying to establish a causal relationship between our variables of interest. Obtaining this causal relationship is often difficult in statistics. Where possible, we will look to use methods and frame our statistics in a way to try and control for these limitations.

An issue that often arises from the human capital literature is one of **endogeneity**, which is when the relationship between the independent variable (X) and the error term (E) are correlated. This leads to the relationship between the independent (X) and the dependent variable (Y) to be over- or under-estimated. There are three reasons why in our work this can occur. (1). When unobserved characteristics, or omitted variables, such as ability and motivation, are not controlled for in our model, this can lead to relationships to be over-stated. For example, when estimating the returns from in-work training on wages, by not controlling for the ability and motivation of the individual, this could over-estimate the relationship between the course and earnings in the future.

Another form of endogeneity can arise from simultaneity bias, or reverse causality, where the independent variable (X) is jointly determined with the dependent variable (Y). An example of this can be seen with education and health. Having a higher education may lead you to have better health, as you may be more aware of your behaviour (smoking, alcohol consumption) and the benefits of regular exercise and diet. However, having better health may also lead you to have a higher education, as you may be healthier to undertake higher education, be more motivated, or similar. Because of this, it can be difficult to see which direction the relationship flows.

Measurement error, which is the difference between measured value and its true value, is also a form of endogeneity. Due to the difficulty in collecting data on a variable that truly affects economic behaviour, a variable might be computed which may not be a precise measure of the economic variable, for example, capturing reporting wages as a measure of actual wages.

To alleviate the issues around endogeneity in the model, we propose several methods. One approach is to use a two-stage least squares, or instrumental variable approach. This involves finding an instrument which has a relationship with the endogenous independent variable but does not have a relationship with the dependent variable. One example of this is estimating the returns to schooling. The number of years of schooling influences earnings, but educational levels are not randomly assigned (individuals choose their level of education). To deal with this issue, Card (1993) used geographic differences in the proximity to a college as an exogenous variable and found that return to education is substantially higher. Another approach we have considered is using a fixed effects approach, which would allow us to deal with any unobservable factors, as it will take out any component that is constant over time.

Selection bias is the bias introduced by the selection of individuals or a group for analysis, where the sample obtained is not representative of the population intending to be analysed. One form of selection bias in our analysis, as explained previously, tends to be around motivation and ability bias, where people with higher motivation and ability are likely to do better in school, participate in in-work training, and tend to have higher earnings and health outcomes. Another form of selection bias is called “sorting gain”, which assumes that people who benefit the most from university will attend university. One other problem we could face in our work is around selection bias on those employed and unemployed. Those in employment will tend to have higher earnings than those not in the labour force, hence our results will be biased. This has been dealt with in the literature, using methods such as the Heckman correction.

While it is important that we consider the relationship between human capital and labour market outcomes, and the methods that would allow us to achieve more reliable estimates of this relationship, it also important to consider ways in which we can measure skills more directly. By measuring skills more directly as an outcome, we can get a better understanding of what factors have an impact on the attainment of different skills, such as reading, writing, and mathematics. In addition, selection bias can be somewhat alleviated when considering certain skills developed by individuals through different stages of their life. As we explain later, we want to consider the human capital of those at school age, working age, and retired. By considering the factors that impact skill development as a child, for example, mental impairments or motivation, this would allow us to control for these factors when the individual reaches working age. We can then see how these childhood and adulthood factors influence labour market outcomes, well-being, and further development of skills. By controlling for more of these drivers in our regression model, we can alleviate some concerns around selection bias in our model.

It is also important to consider that some variables in our model may have a **non-linear effect**. For example, there is a non-linear relationship between personal well-being and age (curvature is a feature of the relationship), as personal well-being tends to be higher for younger people, which deteriorates as they reach middle age and improves as they get older towards retirement. While age will be used as a control variable in many of our models along with other non-linear variables, the difficulty of non-linear relationships as an interest variable is when interpreting the impact on the outcome variable. For example, we can see the growth of skills being a non-linear relationship, as a child with a strong initial understanding in maths may have a higher compound growth in learning than a child with little initial understanding in maths. We can partially control for this by running separate models for different life stages, using models that allow for non-linear terms, and also consider running models for specific parts of the population, allowing us to work out each independent impact.

We will also look to consider **interaction effects**. This can relate to when the causal variable on an outcome depends on the state of a second causal variable. For example, when trying to capture the impact on earnings from doing in-work training, we might want to consider interacting the effect from in-work training with that of having a supportive manager. This would allow us to see the difference in earnings between those that train and those that train and have a supportive manager. In order to present these effects as indicators, we will try and extract the total effect size, using techniques such as pathway analysis. Additionally, we may present indicators that have interaction effects jointly, or even present jointly-interacting indicators explicitly. In the example above, this may mean showing two outcome indicators on earnings of in-work training - one for those that do not have supportive managers and showing a separate indicator for the extra premium that may occur from also having supportive managers.

As we are trying to measure the impact of human capital across an individual's lifetime, it is important to recognise what is sometimes termed intertemporal complementarity, where certain factors may have **compound effects across life** on an individual if the impact is when they are a child, compared to when they are an adult. An example of this is family structure. A child being bought up in a single parent home may have lower skills and knowledge when compared to a child being bought up in a two-parent home, as the single parent may have less resources and time to invest in a child. This has a knock-on impact to future skills development. If the impact of being brought up in a single-parent home is only during a child's development, then its effect should be captured through outcomes at that life stage. What we want to capture later in life is specific factors that lead adults to accumulate further skills and knowledge, for example, participating in in-work training or education, work experience etc.

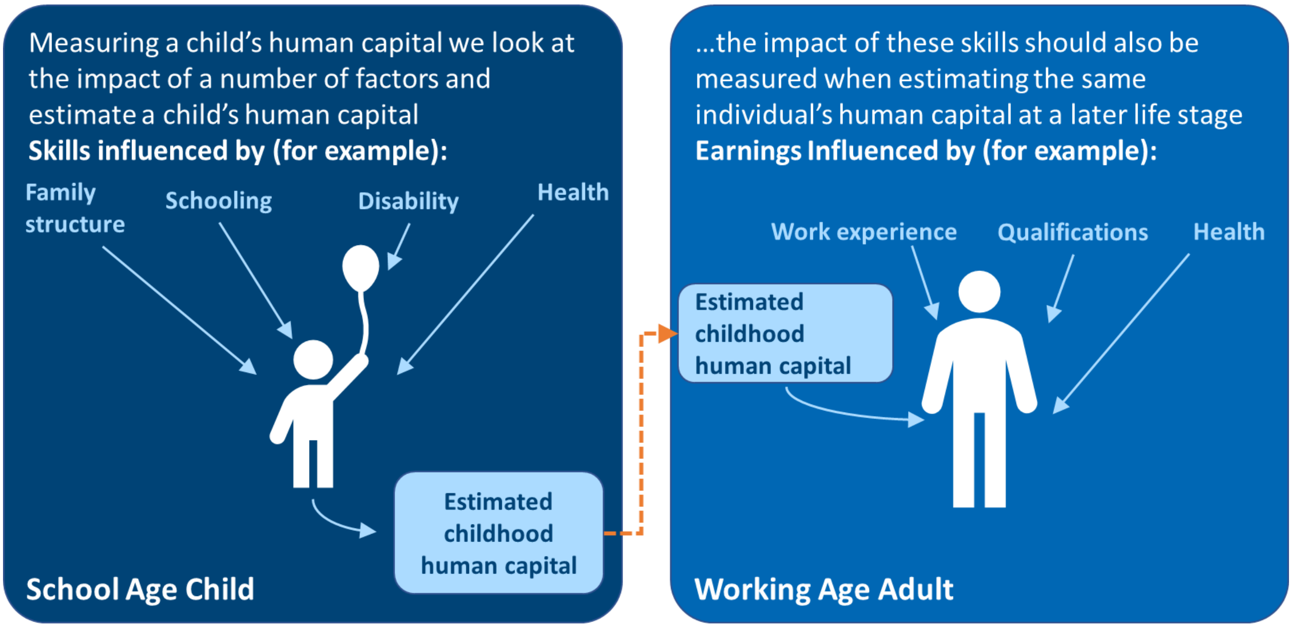
To achieve this, we can use longitudinal data sources that would allow us to track the same individual over time. While much of the human capital literature focuses on the returns to earnings, following an investment in human capital, it is difficult to use earnings as an outcome for those at school age. Therefore, we need to look broader and capture the skills, knowledge, competencies and attributes of young individuals as outcomes, and how factors such as family structure, health, and among others, have an impact on how young people accumulate human capital.

There are many reasons why we are taking this approach.

1. By controlling for factors that impact human capital accumulation as a child and as a working adult in one model, it may lead to an over-specification bias of our model, leading to insignificant relationships and biased results. Note, there are multiple life-stages so here we just use the example across two of the life stages
2. As longitudinal data sources have smaller sample sizes, controlling for many factors will lead to a small sample to produce our estimates.
3. As some data sources we have considered do not have an extensive longitudinal element, it is important we capture the impact of child hood factors on human capital accumulation as a working adult. Therefore, this approach would allow us to control for childhood factors and use data linkage techniques such as propensity score matching to transfer these predicted values across to other datasets, which do not have a longitudinal element.

Data linkage

As these indicators cross many different themes, we recognise the data limitations of our approach and potential solutions that can help us to alleviate some of these limitations. Lots of drivers interact with each other so we would need one dataset to combine a large amount of information, though such a dataset doesn’t currently exist. Additionally, some data sources do not measure individuals over time, so it is important we capture the impact of childhood factors on, for example, skills and how this has an impact on a working age adult. This is demonstrated in the diagram below:



One solution to this may be through linking survey and administrative data sources, using data science and econometric techniques This would allow us to produce a complete dataset to produce the desired indicators.

Given the breadth of themes and inputs we want to control for in our regression model, it is understandable that many of these elements may not be contained in just one data source. Therefore, to control for these personal demographic and economic factors, such as skills, health conditions, labour market outcomes, and personality traits, it is necessary for us to link several survey or administrative data sources together to ensure we can produce more reliable relationships between our variable of interest and outcome. For example, while the annual population survey contains information on the self-reported health of an individual, we may also want to expand on these health indicators using other sources of data that contain more detailed health conditions of the population. This could be done through linking an administrative health source to an existing survey.

It is also important to consider that not all data sources used to produce these indicators are longitudinal, therefore, capturing child and adult factors that impact on human capital in one data source may not always be possible when producing all our indicators. This is where data linkage can help to bring together information of individuals through different stages of their life. As explained in the previous section, by capturing the factors that impact a child and using these predicted values, we can use data linkage techniques such as propensity score matching to transfer these values to a dataset which contains information on a working age adult. Therefore, we can control for childhood factors in data sources which do not have a longitudinal element.

Administrative data is great for granularity, which can be hard to capture in a survey because of recall and social desirability biases, but it doesn't give you opinions or rationales. For example, administrative data can tell us which child received what vaccinations at what age, and which children weren't vaccinated, but the survey data could tell us why the unvaccinated kids didn't receive vaccinations.

It is also important to consider how this would influence our models. A limitation we may face is in terms of variables (derived or raw) from linked datasets is "missingness". These aren't necessarily missing in the sense of the data not being recorded but missing in the sense that if an event didn't occur, therefore, there won't be a record of that person being in a dataset. For example, not everyone has been to hospital. This can lead to populations in each group being too small to do statistical analysis. On a related issue, when matching individuals on an administrative and survey data source, issues may arise where the number of people matched is a fewer than the number of people in the sample. This can arise through individuals in the survey or administrative data source not being found in either one, which can be due to multiple reasons, which may lead to a loss of power within a statistical model.

What Regression Models are we planning to use?

Regression Model	Limitations	Criteria
Two-Stage Least Squares (2SLS) Two-Stage least squares (2SLS) regression analysis is a statistical technique that is used in the analysis of structural equations. This technique is the extension of the OLS method. It is used when the dependent variable's error terms are correlated with the independent variables.	Difficulty in finding exogenous instruments which are correlated with the independent variable and not with the dependent variable or any unmeasured confounding variables Even if a good instrument is found, it may underestimate the impact of the relationship between the independent variables and the dependent variable. Instrumental variable performance in small samples may be poor, leading to inconsistent and biased estimates.	☒ Fixed effects ☒ Interaction terms ☐ Selection bias ☒ Non-Linear ☒ Intertemporal ☒ Endogeneity Inc. omitted variable bias, reverse causality, measurement error.

Pathway Analysis Pathway analysis aims to provide estimates of magnitude and significance of hypothesised causal connections between sets of variables.	The casual relationship between one variable and another can only go in one direction. This is an issue when it comes to reverse causality. The variables must be time-ordered. Can only be used when there is a steady causal progression across or down a path diagram, there cannot be a feedback loop. Relationships in the path diagram must be tested by multiple regression. The intervening variables all must serve as dependent variables in multiple regression analyses.	☐ Fixed effects ☒ Interaction terms ☐ Selection bias ☐ Non-Linear ☐ Intertemporal ☐ Endogeneity Inc. omitted variable bias, reverse causality, measurement error
Difference-in-difference A difference-in-difference model is designed to estimate the causal effects from certain policy interventions and policy changes that do not affect everybody at the same time in the same way. This model is useful when instrumental variables is deemed unsuitable.	The model requires the treatment and control group to have a similar baseline. If intervention is determined by baseline outcome, or the comparison groups have different outcome trend, the model cannot be used. The technique cannot be used if the composition of groups pre/post change are not stable.	☒ Fixed effects. ☒ Interaction terms ☒ Selection bias ☒ Non-Linear ☒ Intertemporal ☒ Unobservable ☐ Endogeneity Inc. omitted variable bias, reverse causality, measurement error
Regression Discontinuity Designs Regression Discontinuity Design (RDD) is a quasi-experimental evaluation option that estimates the impact of an intervention through an observed "running" variable exceeding a known cut off point.	If another treatment occurs at the same cut-off value as the treatment you're interested in, it may be difficult to separate out the treatment you're interested in. The estimated effects are unbiased if the functional form of the relationship between the treatment and outcome is correctly modelled. An example of this can be a non-linear relationship being mistaken as a discontinuity. Regression discontinuity designs can be invalid if individuals can precisely manipulate the running variable. For example,	☐ Fixed effects. ☒ Interaction terms ☐ Selection bias ☒ Non-Linear ☒ Intertemporal ☐ Endogeneity Inc. omitted variable bias, reverse causality, measurement error.
Simultaneous Equation Regression Model Simultaneous Equation Model (SEM) is a model in the form of a set of linear simultaneous equations. This model is jointly determined by equations in the system, where the system exhibits some type of similarity or causation between the X and Y variables.	Often requires a theoretical model to base the empirical inference on, which can be more subjective.	☐ Fixed effects ☐ Interaction terms ☐ Selection bias ☒ Linear or Non-Linear ☒ Intertemporal ☐ Unobservable ☐ Endogeneity (measurement error, omitted variable bias, reverse causality).