15th Meeting of the
National Statistician's Data Ethics Advisory Committee

Minute, Agenda and Papers

Friday 8 February 2019

10:30 – 13:50

Board Room, UK Statistics Authority
London

UK STATISTICS AUTHORITY

NATIONAL STATISTICIAN'S DATA ETHICS ADVISORY COMMITTEE

Agenda

Friday 8 February 2019
Drummond Gate London
10:30am – 1:50pm

Chair: Dame Moira Gibb

(10:30am to 12:00pm)

1 10:30am	Minute and matters arising from the previous meeting	NSDEC Chair
2 10:40am	Qualitative research to inform a feasibility study of whether a child abuse prevalence survey could be effective	NSDEC(19)01 Meghan Elkin
3 11:00am	Update on 'Extending the Crime Survey for England and Wales (CSEW) to include a new module of questions on the online behaviour of children aged 10-15 years [NSDEC(18)17]'	Presentation Meghan Elkin
4 11:20am	Migration Statistics Transformation Programme	Presentation Rebecca Briggs
5 11:40am	Using linked administrative datasets to understand international migrants' interactions with the health sector	NSDEC(19)02 Nicola Rogers

Lunch (12:00pm to 12:30pm)

(12:30pm to 1:50)

6 12:30pm	Linking Welsh Cancer Data to the ONS Longitudinal Study	NSDEC(19)03 Jim Newman
7 12:50pm	Data for children: assessing the impact of family background on the outcomes of children in England	NSDEC(19)04 Bonnie Lewis
8 1:10pm	ONS approach to communicating and engaging on the public benefits of integrated data	Presentation Andy Teague
9 1:30pm	Any other business	

National Statistician’s Data Ethics Advisory Committee
Minute

Friday, 8 February 2019
Boardroom, Drummond Gate, London

Present
Members
Dame Moira Gibb (Chair)
Mr Rob Bumpstead
Ms Vanessa Cuthill
Mr Keith Dugmore
Mr Colin Godbold
Ms Annie Hitchman
Ms Isabel Nisbet
Ms Marion Oswald
Dr Emma Uprichard

UK Statistics Authority
Dr Simon Whitworth

Office for National Statistics
Ms Meghan Elkin (for items 2 to 3)
Ms Rebecca Briggs (for items 4 to 5)
Ms Nicola Rogers (for items 4 to 5)
Mr Jim Newman (for item 6)
Ms Bonnie Lewis (for item 7)
Mr Andy Teague (for item 8)
Ms Lara Phelan (for item 8)

Natcen
Ms Sarah Sharrock (for item 2)

Apologies:

1. Minutes and matters arising from the previous meeting

1.1. The Chair welcomed members to the fifteenth meeting of the National Statistician’s Data Ethics Advisory Committee (NSDEC).

1.2. Members approved the minutes from the previous meeting.

1.3. Dr Whitworth updated the meeting with progress on actions from previous meetings. Most actions were complete or in progress and would soon be complete.

2. Qualitative research to inform a feasibility study of whether a child abuse prevalence survey could be effective [NSDEC(19)01]

2.1 Ms Meghan Elkin, Head of ONS Crime Survey team, and Ms Sarah Sharrock, Senior Researcher at NatCen, briefed NSDEC that ONS had commissioned NatCen to conduct feasibility research to inform an ONS decision about the viability of a pilot study on the prevalence of child abuse. This would involve conducting a small number of interviews with the following;

i. children aged 11-17;
ii. young people aged 18-25 who were victims of abuse as a child;
iii. parents of children who are known victims;
iv. parents of children who are not known victims; and
v. people working with children to determine whether a prevalence survey could be effective.

2.2 The Committee heard that numerous stakeholders, such as the Centre of Expertise on Child Sexual Abuse and the NSPCC, had suggested that a survey on child abuse could provide valuable information for policy makers, service providers and practitioners to help improve the experience of victims and hopefully over time, reduce the prevalence of child abuse. It was made clear to the committee that this research would not ask respondents to answer questions on any experiences of child abuse.

2.3 The following comments were raised in the discussion that followed:

- NSDEC were assured that NatCen had appropriate experience in collecting sensitive data from children;
- more information on how the interviewees would be selected and how these interviews would be conducted should be included in the application;
- ONS should consult internationally to learn from any other countries who had done similar work;
- ONS should learn from the earlier work done by the NSPCC which was cited in the application;
- it was recognised that the lessons learned from this small-scale feasibility research could be used to inform the child cyber module on the Crime Survey of England and Wales; and
- ONS should assure themselves that the security and safeguarding measures that NatCen have in place are sufficient for research of this type.

2.4 This project was approved subject to Ms Elkin providing the following:

- sharing the cover letters that will be sent to the parents and the children with the committee;
- making it clear in the application that there will always be two people present throughout this work;
- providing clarity in the application on how the gatekeepers will identify the people to interview; and
- sharing any changes to this proposal that are requested by the NatCen ethics committee with NSDEC.

3. Update on ‘Extending the Crime Survey for England and Wales (CSEW) to include a new module of questions on the online behaviour of children aged 10-15 years [NSDEC(18)17]’

3.1. Ms Elkin updated NSDEC on the work that has been completed in response to NSDEC’s comments on including a new module on cyber crime in the 10-15 year olds Crime Survey of England and Wales. The meeting heard that there was a demand for data on this topic from across government departments and the voluntary sector.

3.2 The Committee heard that ONS had continued to develop the risk rating approach with NSPCC and Kantar, the organisation ONS had commissioned to administer the survey, conducted testing of the approach and carried out additional reviews of survey and safeguarding materials.

3.3 This project was approved subject to Ms Elkin doing the following:

- clarifying Kantar's data protection and privacy policies;
- clearly communicating the following to the child:
  i. that a letter will be sent (to the child and parent/guardian) with a risk score for their online activity,
  ii. that if they are under 16 the parent/guardian(s) have a right to request access to the child's answers
  iii. that if the parents make a legal request, ONS may have to share the answers the child gave with them; and
  iv. providing NSDEC with progress reports on the roll out of the module at future meetings.

4. Migration Statistics Transformation Programme

4.1. Ms Rebecca Briggs, from the ONS Migration Statistics Transformation Programme, presented an overview of ONS' plans to use administrative data to transform the way it produces migration statistics. The Committee were informed that all uses of data would be consistent with the DEA Statistics Statement of Principles which was approved by Parliament.

5. Using linked administrative datasets to understand international migrants' interactions with the health sector [NSDEC(19)02]

5.1 Ms Nicola Rogers from the ONS Migration Statistics Transformation Programme presented a project to link non-clinical Hospital Episodes Statistics data to HMRC/DWP Migrant Worker data, HMRC PAYE and Self-Assessment data, and Higher Education Statistics Agency (HESA) data. This would be done to better understand, at an aggregate level, migrants' interactions with the health sector to improve understanding of the interactions and contribution of migrants while they are in the UK.

5.2 This project was approved subject to Ms Rogers doing the following:

i. being clear in the application about how a migrant is defined;
ii. providing a better explanation of why each of the variables from the Hospital Episodes Statistics data was required; and
iii. clearly explaining to external audiences about the public good of this work.

6. Linking Welsh Cancer data to the ONS Longitudinal Study [NSDEC(19)03]

6.1 Mr Jim Newman, from the ONS Longitudinal Study (LS) team, presented a proposal to continue linking Welsh Cancer data to the LS. NSDEC heard that the LS is an established longitudinal study which contains linked census and life events data for a 1% sample of the population of England and Wales from

1971. It was suggested that there was a long tradition of ONS making deidentified LS data available safely and securely to approved researchers for approved research for the public good.

6.2 NSDEC recognised that the continuation of the linkage of Welsh cancer data to the LS provided a very powerful tool to understand how social factors relating to the individual, their household structure and their physical environment can influence cancer outcomes.

6.3 This project was approved.

7. Data for children: assessing the impact of family background on the outcomes of children in England [NSDEC(19)04]

7.1 Ms Bonnie Lewis, from the ONS Centre for Equalities and Inclusion, presented a proposal to use the All Education Dataset for England (a longitudinally linked cohort dataset designed by Department for Education in collaboration with ONS Admin Data Census) to better understand the impact of family background, school type and geography on educational outcomes and social mobility.

7.2 The committee was informed that this work would also explore the utility of this data for research purposes. If utility was proved through this research, the data would be made available to approved researchers as part of the Administrative Data Research Partnership. To access this data under the Research strand of the Digital Economy Act (DEA), researchers would have to provide evidence to the Research Accreditation Panel, which will govern the access to data under the Research strand of the DEA, that the ethics of their research had been appropriately considered.

7.3 It was suggested that it may be useful to include information on Further Education students in the research and that this analysis should be not used as a predictor for children’s educational outcomes.

7.4 This project was approved subject to Ms Lewis doing the following:

i. rewording the research question on which local authorities add the most value;
ii. publishing the results from the analysis; and
iii. rewording the title of the project to more accurately reflect the scope of the work.

8. ONS approach to communicating and engaging on the public benefits of integrated data

8.1 Mr Andy Teague from ONS’s Statistical Design and Research team, and Ms Lara Phelan, from ONS Communications Division, presented the ONS approach to public engagement on the public benefits of using integrated data for research and statistics. The approach including the following:

i. creating clear positioning and messages which reduce any concerns and issues about data collection;

ii. providing clear and easy to consume information about ONS and the way it collects and uses data, via public facing channels such as ONS website and public-facing teams; and
iii. demonstrating the value of integrated data by using case studies to show the benefits of integrated data.

8.2 NSDEC offered support for this work and stressed the importance of producing robust case studies which clearly show how the use of data for research and statistics can have real impacts on people’s lives. These case studies need to be clear about the benefits and be understandable to the public. Specific tailored communication approaches may be required when using certain types of data, such as social media data, for research and statistical purposes.

9. Any other business

9.1 There was no other business.

NSDEC(19)01

National Statistician’s Data Ethics Advisory Committee

Application for Ethical Review

Please consult the guidance document before filling in the application form

---

Project Title

Please provide a title indicative of the project

Qualitative research to inform a feasibility study of whether a child abuse prevalence survey could be effective

Start Date:	Feb 2019	**End Date:**	December 2019

Project Sponsor(s)

Please list the project sponsor(s)

Office for National Statistics (ONS)

Project Summary

Please provide a brief high level summary of the research giving necessary background (max 250 words)

In September 2015, the National Statistician’s Crime Statistics Advisory Committee established a Task and Finish Group to make recommendations for improvements to the official statistics on child abuse. One of the group’s key recommendations was that Government should commission a UK-wide prevalence study of all forms of abuse and neglect of children to establish a reliable time series of data.

It would be very costly to run such a survey nationally with a sample large enough to produce robust results, and there would always be some bias in the estimates obtained given the sensitivity of the topic. However, the case made for running such a survey is strong (as detailed in Section B1). Therefore, it was recommended to the National Statistician that ONS should conduct a feasibility study, to determine whether a prevalence survey could be effective. This was agreed, and in 2018 the ONS Centre for Crime and Justice started work on the feasibility study.

To date, the feasibility study has involved desk research and stakeholder engagement. However, this type of research can only tell us so much. As the next step, we plan to carry out qualitative research to explore some of the issues in more detail. This would enable us to better assess whether it’s possible to successfully overcome these issues.

The results of the qualitative research will be used to inform a decision on whether the feasibility study should continue. If the research and evidence gathered to that point suggest that it could be possible to make a prevalence survey work, we would continue the study and look to set up a pilot survey. The qualitative research will be an important step in helping us to determine whether there is any value in running a pilot. If the research shows that there are insurmountable issues, we will be able to conclude that a prevalence survey couldn’t be successful prior to embarking on a costly pilot.

The issues we will explore through the qualitative research are:

- Would parents/guardians give their consent for their child to participate in such a survey?
- What concerns would parents/guardians have about their child participating in such a survey?
- How would parents/guardians feel about the survey being conducted in school?
- What support/guidance/information would parents expect to be provided for themselves and their child if their child participated in such a survey?
- Would parents expect to be able to find out what responses their child gave in the survey?
- What would parents expect to happen if their child did disclose having experienced abuse in the survey?
- Would children be willing to participate in such a survey?
- What benefits do children think there are to such a survey?
- If they do participate, would children complete the survey truthfully?
- What concerns would children have about participating in such a survey?
- What concerns would children have about disclosing abuse in such a survey?
- How would children feel about the survey being conducted in school?
- What concerns do practitioners have about such a survey?
- What impact do practitioners think such a survey would have on victims?
- What support do practitioners think would need to be available to participants following the survey?

The project will be carried out in 2 phases, which will allow the research tools, recruitment approach etc to be reviewed after phase 1, with any necessary changes then implemented for phase 2. Phase 1 will cover research with practitioners and young people. Information from children and parents will be collected in phase 2. Advice will be sought from practitioners and young people on how best to approach the research with children. This will then inform how phase 2 should be conducted.

This application is to gain approval for the qualitative research part of the feasibility study only. ONS are commissioning NatCen Social Research to conduct this research. NatCen are very experienced in conducting research with children and other stakeholders about challenging and sensitive subjects, including:

- a project on safeguarding against child sexual abuse in residential schools for the Independent Inquiry into Child Sexual Abuse
- a study with victims of sexual violence for the Sentencing Council
- exploring suicide prevention among men for the Department of Health
- a project on extremism for the Foreign and Commonwealth Office
- a project on hate crime for Her Majesty’s Inspectorate of Constabulary and Fire & Rescue Services

Section A
Project Details

A1 Legal gateways

Please provide the assessment of the legal gateways of the project as provided by Legal Services

No legal gateways are required for this project as data will be collected directly from the data subjects with informed consent.

A2 Ethical approval

Has the project being reviewed or is it expected to be reviewed by another ethics committee?	○ **Yes** ○ **No**
If Yes please provide the name of the committee, the outcome and the date approved	NatCen will also seek approval for this project from their internal Research Ethics Committee, which is made up of senior NatCen staff and external experts. The project will go to their ethics committee after the contract has been awarded. We will update NSDEC after the project has been to NatCen's Research Ethics Committee if any changes are made to the project proposal as a result.

A3 Proposed site of research select all that apply

☐ ONS

☐ SRS (formerly VML)

☐ HMRC Data Lab

☐ ADRC - England

☐ ADRC - Scotland

☐ ADRC - Northern Ireland

☑ Other

☐ ADRC - Wales

The research will be carried out in a quiet location agreed on with the participants. This will be driven by discussions with gatekeepers to ensure it is a safe setting and children will be given options so they can choose a setting they are most comfortable with

A4 Data subjects to be studied

Does the study include all subsections of the population (i.e. all ages, sex, ethnic groups etc?)

○ **Yes** ○ **No**

If no please detail which subsections with justification(s) below:

Subsections of the population (including vulnerable groups) the project focuses on:

Children and young people aged 11 to 25
Parents/guardians of children aged 11 to 17
Practitioners who work with children

Justification for focusing on these subsections or groups:

It has not yet been determined what the lower age limit for participation in the prevalence survey should be, but it has been agreed that it will not be lower than 11 because of difficulties in collecting information from children of such a young age. In a previous study conducted by the NSPCC information was collected about abuse amongst children aged under 11 from parents/guardians. However, NSPCC have recommended that this approach should not be repeated as the costs far outweigh the benefits.

Therefore, the children taking part in this research project will cover a variety of ages between 11 and 18 to gather a range of views from across the age range that would be likely to be included in a prevalence survey.

In addition, it is likely that the upper age limit for participation in the prevalence survey would be 25, because research shows that disclosure rates go up over the age of 18. Therefore, a sample of adults aged between 18 and 25 will be included in the research. Not only is it important to explore issues around the participation of this group in a prevalence survey, but adult survivors of child abuse could also give their views on how they would have felt about participating in a survey of child abuse as a child.

Parents will be included to explore issues around them giving consent for their child to participate in a survey of child abuse, what concerns they would have, etc. Practitioners who work with children will also be included in the research to provide additional insight into the issues around a prevalence survey and a different perspective to parents and children.

A5 Please provide details of the research protocol or methodology (e.g. data linkage, web scraping etc) (max 500 words)

In-depth qualitative methods will be used to gather evidence for the feasibility study. A combination of interviews and focus groups will be carried out to capture the views of the different participant groups on issues related to conducting a survey of child abuse.

Children and young people
Face-to-face encounters will be used with children and young people to ensure the research team can develop rapport and enable participants to speak openly, confidentially and safely about the subject.

Children will be offered the opportunity to participate either by themselves or in small groups of between 2-4 participants, including friendship groups. This approach has worked well in encouraging participation in other research projects of a sensitive nature, and using friendship groups has encouraged children to be more open. Children will also have the option of having a key worker or parent/guardian present if that would make them feel more comfortable about taking part.

NatCen will conduct 12-16 interviews/mini groups with children aged 11-17. Children will be split into two age groups (11-14 and 15-17) because of the nature of the topic under discussion and level of engagement of each. This will enable distress or discomfort to be better monitored during discussion, and for children to feel more confident around peers closer to their age. Some of the children will be known victims of abuse, others will not. Known victims will be identified to take part in the research through gatekeeper organisations.

A focus group will be conducted with young people aged 18-25 who were victims of abuse as a child.

Children and young people who were victims of abuse as a child will be identified to take part in the research through gatekeepers, i.e. through organisations who work with victims, such as the NSPCC, Barnardo’s, and the Office of the Children’s Commissioner. The exact gatekeeper organisations to be used will be discussed and agreed between ONS and NatCen as part of the recruitment strategy. Participation in the study will not be dependent on the victims having been through the criminal justice system in relation to their abuse, given that a key aim of a prevalence survey would be to identify those victims who are not currently known.

Any children or young people who are known to be currently experiencing, or have recently experienced, any form of abuse will be excluded. NatCen will be clear to gatekeepers about the requirements and look to incorporate any guidance about how best to involve their service users in the research. This will include ensuring that participation in the research will not put any participants at risk. For example, considering who the perpetrator of abuse was and whether this has any implications for whether it is safe for the victim to participate in the study.

Parents/guardians and practitioners
Face-to-face encounters will also be used with parents/guardians, with 2 focus groups conducted with parents. One will involve parents of children who are known victims and one will be with parents of children who are not known victims. Parents of known victims will be recruited via gatekeeper organisations.

NatCen will conduct 2 focus groups with practitioners, or if it is not possible to arrange a focus group due to the practitioners' schedules, interviews will be carried out instead. This will include practitioners from a range of organisations that work with/support children and young people. For example, social workers who work with children on a regular basis as well as those who work with children in a more specialised capacity in relation to abuse. These participants will be recruited through established relationships with organisations such as the NSPCC and Barnardo's. These focus groups/interviews could be carried out online or over the phone, which may be preferable if participants have busy schedules or if they have concerns about sharing information with competing organisations.

Gatekeeper organisations
The gatekeeper organisations will be given a briefing which provides an overview of the research and what taking part involves. Those that are happy to support the study will be contacted by a researcher from NatCen to discuss the objectives of the research. This will include a discussion around the different groups of people that we hope to include in the study.

The gatekeepers will pass on an information leaflet and privacy notice to people they have identified that could potentially participate in the study, explaining the purpose of the study and emphasising the voluntary nature of the research.

Gatekeepers will then share the contact details (first name, phone number and email address where relevant) of any individuals who consent to this with the NatCen research team, so that NatCen can then make contact with the potential participants.

Facilitation
The mini groups with children and focus groups with young people and parents of victims will be facilitated by two researchers. Having a second researcher in the room will enable a simultaneous focus on the research tasks (facilitating the discussion, focussing on the topic guide coverage and managing the dynamic in the room) and on supporting participants (including debriefing outside the group with any individual who becomes distressed due to the sensitivity of the topic under discussion).

Research tools
A range of research tools will be developed to enable full exploration of the research questions with each group. For example, visual aids such as traffic light cards will be used so participants can indicate if they want to pause or stop, without having to verbalise their wishes. Enabling techniques such as vignettes will be used with children to enable them to focus on abstract ideas and hypothetical scenarios.

Data collection encounters will last between around 45 minutes for younger children and 90 minutes for focus groups with adults. No data collection will happen on a Friday afternoon or weekend when there may not be support services available to refer to if required.

The project will be carried out in 2 phases, which will allow the research tools, recruitment approach etc to be reviewed after phase 1, with any necessary changes then implemented for phase 2. Phase 1 will cover focus groups and interviews with practitioners and young people. Information from children and parents will be collected in phase 2. Advice will be sought from practitioners and young people on how best to approach the research with children. This will then inform how phase 2 should be conducted. If the results from phase 1 suggest that it would not be appropriate to carry out the research with children, or that the proposed age range of children included in the research needs amending, this will be taken into consideration when planning phase 2.

Incentives

All participants will be given an incentive for taking part. This will be in the form of a Love2shop voucher. The level of incentive given will be in line with the standard within the industry and will help to support the recruitment of participants within the project timetable. It will be made clear to participants that they will still receive their voucher even if they withdraw before completion of the interview/focus group.

A6

Data use

Please specify the data used **by the research team** including any timeframes *e.g. LFS data 2014-15*

	Data Level *Please specify the name of the data set*			
Type of data	**Aggregate Data**	**Identifiable Data**	**De-identified personal data**	**Anonymised/ pseudo anonymised**
Administrative data *(please specify, e.g. Patient Register 2011, School Census 2012 etc, in the relevant options adjacent)*				
Big Data *(please specify e.g. Twitter data, smart meters and mobile phones, in the relevant options adjacent)*				
Survey Data *(please specify e.g.LFS, BRES, etc in the relevant options adjacent)*				
Census Data *(please specify year, e.g. Census 2011 in the relevant options adjacent)*				
Other *(please specify e.g. Ordnance Survey Address register in the relevant options adjacent)*				Anonymised data collected through interviews/focus groups

Section B
Assessment against NSDEC ethical principles

B1 Principle 1: The use of data has clear benefits for users and serves the public good.

Please outline the proposed benefits of the project (max 500 words)

There is no data source that provides a current measure of the prevalence of child abuse. The data sources that do provide information on child abuse are dependent on victims being known to the authorities or having sought help, but many victims remain hidden.

A national child abuse prevalence survey would provide valuable information which would be widely used by policy makers, service providers, and practitioners to help improve the service provided to victims and hopefully over time, reduce the prevalence of child abuse. Organisations working to help support victims of child abuse and reduce victimisation regularly require prevalence statistics in their work.

The need for a national prevalence survey has been recognised by a number of external organisations, who have publicly called upon ONS to implement such a survey. The Home Office funded Barnardo's to establish a Centre of Expertise on child sexual abuse. This centre published a scoping report in July 2017, which concluded that the understanding of child sexual abuse has improved but there is still a lack of a comprehensive picture and an inability to monitor changes over times. The report stated that the first step for future work was *"To advocate for the UK Government to commit to commissioning a regular CSA prevalence study, under the auspices of the Office for National Statistics".*

This echoed a call that the NSPCC have repeatedly made. They commented in their 2017 'How safe are our children?' report that it isn't enough to simply know the numbers of reported cases of children being abused and neglected; we need to understand the full scale of child maltreatment to know how many more children need to be reached, and *"…that's why there's an urgent need for the UK Government to commission a new UK-wide study that looks at the prevalence of all forms of abuse and neglect".*

The report went on to say: *"We need a clear picture to inform how we all can do our part to keep children safe. We are calling on the UK Government to now step up and make that happen. We need to count every child who has been abused – they are all counting on us".*

Other key stakeholders such as the Office of the Children's Commissioner, the Independent Inquiry into Child Sexual Abuse and the Department for Education are also supportive of the work we are doing, recognising the need for better data on child abuse to fill gaps in knowledge, with the aim of identifying what can be done differently in future.

The most comprehensive study of child abuse that has been carried out in the UK was by the NSPCC in 2009. The 2009 study consisted of interviews with parents/guardians of children aged under 11, interviews with children aged 11 to 17 with additional information provided by their parents/guardians, and interviews with young people aged 18 to 24. The reports based on the 2009 study have been cited many times, and continue to be almost 10 years after the field work was completed. The acceptability of these figures is now being called into question given their age, however, there are no other robust prevalence figures available. It is important to understand current prevalence levels to be able to provide effective responses for maximum impact for children.

A robust estimate of the number of victims and survivors of child abuse would provide a basis for policy making and inform the design, delivery and funding of services aimed at preventing and responding to child abuse. The information gathered from a survey would enable an evidence-based approach to tackling child abuse.

As well as providing an overall prevalence estimate, a large-scale national survey could also help to understand the problem within sub-groups of the population, could be used to establish a baseline against which changes over time can be tracked, and provide detailed information about the nature and impact of child abuse. All of this information would enable effective interventions and prevention strategies to be designed and targeted.

A number of countries have successfully carried out surveys of child abuse including:

- The National Survey of Children's Exposure to Violence in the US - covered conventional crime, child maltreatment and neglect, peer and sibling offending, sexual assault, witnessing and indirect exposure to violence, and online offending. Participants to the survey included children aged 11-17. The survey was administered mainly by telephone, using random digit dialling.
- The Child Sexual Abuse Questionnaire in Switzerland - measured child sexual abuse and child sexual exploitation. Follow up questions included the frequency, age at first occurrence, characteristics of perpetrators, location of each occurrence and disclosure and treatment. Participants were children aged 13-20, who completed a self-completion CASI interview in schools.
- The Violence Against Children Survey - led by the Centers for Disease Control and Prevention as part of the Together for Girls partnership. Covers experiences of sexual, physical and emotional violence prior to 18 years old. The survey has been administered in a number of countries, such as Tanzania, Kenya and Indonesia. Participants to the survey are children aged 13-24, data are collected through an in-person interview with trained interviewers.

Data from a prevalence survey would have clear benefits for users and serve the public good. The feasibility study is needed to determine whether such a survey could be successful in the UK, and qualitative research is an important part of this study as it will provide valuable evidence to help determine whether conducting a pilot survey would be worthwhile.

B2 Please outline any intended future use for products (such as linked data sets or tools) produced as a result of the research and how they will be accessed. (max 250 words)

The findings of the research will be used to feed into the feasibility study decision on whether a pilot prevalence survey should be conducted. In addition, some of the information we find out about conducting surveys on sensitive topics could potentially be relevant to the Crime Survey for England and Wales (CSEW). We will therefore consider whether any of the findings/lessons learned from the research could be applied to the CSEW.

There are currently no other intended uses, however, it is possible that the findings of the research could be used to inform any future work on child abuse that charities or other organisations may wish to pursue. Any future uses would only be based on the published research findings rather than datasets.

B3

Principle 2: The data subject's identity (whether person or organisation) is protected, information is kept confidential and secure, and the issue of consent is considered appropriately.

Please outline how data security, confidentiality and informed consent is safeguarded in this project(max 500 words)

Informed consent
A key ethical consideration for the project is collecting informed consent. Each participant group will be provided with tailored, age-appropriate information sheets, setting out the research aims and ensuring that participants fully understand what taking part involves. If children and young people have particular communication needs, the information will be provided in an appropriate mode and format following discussion with the relevant gatekeeper and/or parent/guardian. The ongoing nature of consent will be explained, and that withdrawal is possible up until the point of data analysis. It will be ensured that potential participants understand that there will be no negative consequences from opting out and that they can withdraw at any time. Parental consent will also be obtained for the children taking part. Draft versions of the consent forms, information sheets, etc will be shared with NSDEC for comment before being used.

Confidentiality
Participants will not be identifiable in the research outputs and all participants will know that their participation will be anonymised and kept confidential. No identifying details will be included in presentations and reports to ONS. This will be communicated transparently as part of the recruitment process, as well as any limitations to confidentiality, i.e. confidentiality may need to be broken through the disclosure and safeguarding procedures in case of issues arising during interviews or focus groups.

It is recognised that some of the younger participants may need additional help to understand the concepts of confidentiality and anonymity. Care will be taken to ensure that this is communicated to them in age-appropriate language.

Confidentiality and anonymity will also be clearly articulated immediately prior to data collection. As some participants will be part of a group, interviewers will tell them not to discuss other participants' responses outside of the group setting. Immediately after data collection the interviewers will check that participants are comfortable with their contribution to the discussion.

Subject to a participant's consent, all interviews and discussions will be audio recorded on encrypted devices and transcribed verbatim for detailed analysis. The recordings will be deleted as soon as they have been transcribed. Transcripts, and any other documents which include identifiable information, will be deleted three months after the project ends.

Safeguarding
The data collection for this project will not focus on capturing information about experiences of abuse, participants will only be asked to discuss issues relating to conducting a survey of abuse. However, we are aware that it is possible some discussion of experiences of abuse will arise during the focus groups and interviews and this may be upsetting and traumatising for the children, their parents, and the researchers. The NatCen staff who will work on this project have had extensive training on how to conduct research on sensitive topics. They will be fully briefed before conducting the fieldwork on the content of the study and how to manage the interviews and focus groups.

The research will be carried out in a way that ensures participants feel supported throughout the study. The format, structure and content of the interviews will be carefully planned to ensure that participants are eased into the sensitive subject and that discussion is ended on a neutral or more positive note as far as possible. The recruitment and fieldwork materials will be carefully designed with particular attention paid to how to describe and frame difficult and sensitive language. Children will be offered the option of participating in individual interviews or mini-groups depending on their preference and what makes them feel more at ease. All interviews/mini-groups will be carried out in comfortable, neutral environments. Children will have the option of having a key worker/parent or guardian present during encounters if that would make them feel more comfortable.

Consent will be monitored throughout the interview process and the researchers will pick up on hesitation and non-verbal signs to identify when a participant may be uncomfortable or distressed. Protocols for handling such situations in both interview and mini-group settings will be developed to ensure participant welfare is always prioritised. For example, acknowledging discomfort and asking participants whether they wish to move onto a different topic, take a break, or bring the discussion to a close.

All participants, and their parents/guardians where relevant, will be provided with a list of relevant organisations that can provide advice and support. Examples of these organisations are NSPCC, Samaritans, Childline, NAPAC, MOSAC, The Children’s Society, Coram Children’s Legal Centre, and Family Lives. This information will be given to all participants to avoid any perception of stigmatisation and because the researchers are not best placed to identify those who are most likely to benefit from the information.

NatCen will also work with the gatekeeper organisations to ensure appropriate signposting and support mechanisms are in place for all participants and options for follow-up support will be explored with the NSPCC. This is likely to include a free-phone line that participants can use to access further support and advice, and this provision will be clearly communicated to all participants.

The research staff will also be offered appropriate support given the sensitive nature of the topic. NatCen have a number of ways of making sure interviewers are safeguarded against upsetting experiences, including holding regular team de-briefs, offering counselling and scheduling interviews at times when support services will be available if required.

Disclosure

If a participant reveals something that needs to be disclosed, they will be told how this information will be shared. Any information that indicates a participant or other person may be at risk of harm will be escalated. If children give their consent, the researchers will be made aware of the nature of the historic abuse experienced by victims so they are able to determine whether experiences discussed during data collection need to be escalated. Previous experience has shown that informing researchers in this way also helps to create an environment where the child feels comfortable and is happy to share their thoughts and feelings.

If somebody is believed to be at risk of harm, the situation will be discussed with NatCen’s Disclosure Board and a decision will be taken about what actions are necessary within 24 hours. If somebody is in immediate danger, an emergency response will be triggered to ensure the safety of those concerned. This will include contacting the police/emergency services.

As this work will be carried out by NatCen, it is NatCen’s safeguarding policy that will be applied rather than ONS’s. This approach is the same as that taken for the Crime Survey for England and Wales, which is carried out by Kantar Public on behalf of ONS. The ONS safeguarding policy states that when ONS commissions other organisations to collect data, ONS must be satisfied that the safeguarding policy of that research organisation is consistent with ONS’s policy. We are satisfied that NatCen’s policy meets ONS standards as the principles of the two policies are the same:

- Information will be passed on to the appropriate authority if it is believed there is a risk to someone's safety
- In an emergency situation, such as immediate danger to life, staff will contact the emergency services as soon as possible
- In a non-emergency situation, staff must report any concern for somebody's welfare to those specified in the policy so that a decision can be made within 24 hours about the appropriate action to be taken
- When a decision is made to pass information on to the appropriate authority, it will also be decided who else should be informed and what level of detail needs to be shared
- Support is provided to staff who encounter potentially distressing situations or information

NatCen will keep us informed of any cases where the policy needs to be applied.

Data security

NatCen has robust physical, electronic and procedural safeguards in place to store client and participant information, and all confidential data are handled in accordance with NatCen's rigorous data security and protection policies. NatCen is fully accredited to ISO 27001 (International Standard for Information Security) and is subject to annual external audits in order to maintain this accreditation, ensuring ongoing compliance.

Team members will work only on dedicated password protected laptops, and should a laptop be lost or stolen, the data contained on the hard drive is inaccessible. The digital audio recorders that will be used to record interviews and discussions are password protected to prevent unauthorised access. Documents or files containing data about a participant are identified by serial number and stored securely and separately from documents containing identifiable information.

All digital data will be stored on a secure network with access restricted to named team members. NatCen's firewall and virus software is constantly updated and their system is compliant with Cyber Essentials PLUS and ISO 27001. Laptops, digital audio recorders and hard copy data will be stored in locked cabinets when not in use.

NatCen have provided ONS with copies of their ISO and Cyber Essentials certificates, as well as information on their vulnerability scanning and pen testing. ONS Security are satisfied with NatCen's security set up.

B4 Principle 3: The risks and limits of new technologies are considered and there is sufficient human oversight so that methods employed are consistent with recognised standards of integrity and quality.

Please describe how the any risks from new technologies are been mitigated as well as any quality assurance activities in the project (max 500 words)

No new technologies will be used for this research.

B5 Principle 4: Data used and methods employed are consistent with legal requirements such as the Data Protection Act, the Human Rights Act, the Statistics and Registration Service Act and the common law duty of confidence

Please describe the legal frameworks pertinent to this project (max 500 words)

NatCen have the necessary procedures in place to ensure all aspects of this research project comply with current UK data protection legislation. All data collected by NatCen will be managed in accordance with institutional data security policies and in line with GDPR. All participants will be provided with information sheets that will set out how their data will be used and their rights under data protection legislation. The privacy statement required by GDPR will also be included in these information sheets.

NatCen are completing a Data Protection Impact Assessment (DPIA) for this work which will be reviewed and agreed by ONS, including the ONS Legal team, before data collection starts. As well as explaining the processing that will be carried out, the DPIA will also identify and assess the risks of the project and identify measures to reduce these risks.

All NatCen staff who will work on the project are BPSS (Baseline Personnel Security Standard) and DBS (Disclosure and Barring Service) cleared, and some have enhanced DBS clearance.

The NatCen team all have bachelor's degrees, and some also have master's degrees. These cover a range of subjects including sociology, social research methods, social and cultural psychology, criminology, and public policy. The team are highly experienced in qualitative research, and in working on projects of a sensitive nature and with vulnerable groups, such as the projects listed in the Project Summary section of this application.

B6 Collaboration and Sponsors

Please describe the project sponsors and the **legal gateways** to acquire, process use and share their data

List of Collaborators/Sponsors	**Details and relevant documentation relating to collaboration (you may attach copies of relevant documentation)**
Office for National Statistics	
NatCen Social Research	ONS issued an invitation to tender for the project, and NatCen were the successful bidder for this tender

B7 Principle 5: The views of the public are considered in light of the data used and the perceived benefits of the research

Please list any public engagement activities (max 250 words)

Users have been informed that we are carrying out a feasibility study via our ‘Improving Crime Statistics progress update’ published on the ONS website. We have been carrying out stakeholder engagement as part of the feasibility study, which has allowed us to gather the views of key expert stakeholders such as the Home Office, the Office of the Children’s Commissioner, the Department for Education, the Department of Health and Social Care, the Independent Inquiry into Child Sexual Abuse, the Centre of Expertise on Child Sexual Abuse, and the NSPCC.

However, this sort of research can only tell us so much and the qualitative research would allow the views of the public to be considered as part of the feasibility study, by gathering views from children/young people, parents and practitioners who work with children.

It is necessary to involve children in the research, because otherwise we can only make assumptions based on existing literature and the opinions of adults. Whilst this provides valuable information to inform the conclusions of the feasibility study, the voice of children also needs to be listened to. Recommending the costly implementation of a national prevalence survey without having spoken to children about their participation in such a survey would be a big risk. Speaking to children is not only important to identify what barriers there may be to a prevalence survey, but also to identify potential solutions to issues identified.

B8 Principle 6: The access, use and sharing of data is transparent, and is communicated clearly and accessibly to the public

How will the findings of the research be disseminated? (max 500 words)

At the end of the feasibility study we will publish a report providing information on the findings and conclusions of the study. This report will include a summary of the data collected via the qualitative research. No individuals will be identifiable in the report. The report will be published on the ONS website and will be available to the public free of charge.

Section C
Responsible owner and applicant details

C1 Responsible Owner

Full Name: Meghan Elkin	**Position:** Head of ONS Centre for Crime and Justice
Address:	**Email:**
	Telephone:
	Organisation: Office for National Statistics

Declaration to be signed by the responsible owner

I have met with and advised the applicant on the ethical aspects of this project design (applicable only if the responsible owner is not the Applicant).

I understand that it is a requirement for all researchers accessing the data to have undergone relevant training and to have either relevant security clearances or approved researcher status in order to access the data.

I am satisfied that the research complies with current professional, departmental and other relevant guidelines.

I will ensure that changes in approved research protocols are reported promptly and are not initiated without approval by the National Statistician's Data Ethics Advisory Committee.

I will provide notification when the study is complete if it or fails to start or is abandoned.

I will ensure that all adverse or unforeseen problems arising from the research are reported in a timely fashion to the National Statistician's Data Ethics Advisory Committee.

I will consider all advice received from the National Statistician's Data Ethics Advisory Committee and should I be unable to implement any of the recommendations made, I will provide reasoning in writing to the Committee.

Signature: .... **Date:** ....

C2 Applicant Details (if applicant is not the responsible owner)

Full Name: Alexa Bradley	**Position:** Head of Crimes Against Children, ONS Centre for Crime and Justice
Address:	**Email:**
	Telephone:
	Organisation: Office for National Statistics

NSDEC(18)17

National Statistician's Data Ethics Advisory Committee

Application for Ethical Review

Please consult the guidance document before filling in the application form

Project Title

Please provide a title indicative of the project

Extending the Crime Survey for England and Wales (CSEW) to include a new module of questions on the online behaviour of children aged 10-15 years

Start Date:	April 2019	**End Date:**	N/A (Continuous survey questions)

Project Sponsor(s)

Please list the project sponsor(s)

Office for National Statistics (ONS)
Home Office

Project Summary

Please provide a brief high level summary of the research giving necessary background (max 250 words)

A new module of survey questions has been developed for inclusion in the existing children's Crime Survey for England and Wales (CSEW) to address an evidence gap concerning young people's internet behaviour and their vulnerability to online crimes such as 'sexting', online bullying and online grooming.

The adult CSEW is a well-established voluntary face-to-face victimisation survey in which people resident in households in England and Wales and aged 16 and over are predominantly asked, via structured interviews, about their experiences of crime in the previous 12 months. These interviews are carried out using computer-assisted personal interviewing (CAPI) where interviewers record responses to the questionnaire on tablets.

The child survey runs as part of the CSEW, covering 10 to 15-year olds. The new survey questions will form part of an existing self-completion section of the survey, which follows at the end of the main interviewer administered section on victimisation. Self-completion by respondents using a tablet computer is used to collect information on topic areas that they could feel uncomfortable talking about to an interviewer.

The large variety of online platforms now accessible to young children has led to increased public awareness of the potential for their misuse. There is a need for greater information on children’s behaviour online and their risk of becoming victims of online crimes. The introduction of new questions into the CSEW child survey is supported by a range of government departments and organisations including the Home Office and National Society for the Prevention of Cruelty to Children (NSPCC).

These new questions will provide important information to build a much fuller picture of children’s online experiences, behaviours and perceptions, which will be useful in identifying possible risk factors associated with online victimisation. The findings will help policy makers, across government and the voluntary sector, to focus their efforts on preventing victimisation of children online. It should also enable the development of support services and information channels to help parents keep their children safe online.

The new questions will be introduced into the child survey from April 2019. Procedures have been put in place to address ethical considerations around the implementation of these new questions. These procedures have been reviewed by the NSPCC Research Ethics Committee and we have taken on board advice from this committee (for example, in adapting the advance materials provided to parents and children). We have also worked with the NSPCC to design appropriate safeguarding procedures for children whose survey responses indicate that they may be at risk of harm.

This work and application have also been informed by earlier conversations with NSDEC. We will continue to keep NSDEC informed on progress following the implementation of this module.

Section A
Project Details

A1 Legal gateways

Please provide the assessment of the legal gateways of the project as provided by Legal Services

The CSEW is a voluntary survey conducted under Sections 20 and 26 of the Statistics and Registration Services Act 2007. Once collected, Section 38 of the Act allows the Statistics Board to use the information for any of their functions, and Section 39 of the Act ensures the information is protected in law.

A2 Ethical approval

Has the project being reviewed or is it expected to be reviewed by another ethics committee?	[✓ ○ **Yes**] ○ **No**
If Yes please provide the name of the committee, the outcome and the date approved	NSPCC Research Ethics Committee (NSPCC REC).

The committee has provided us with valuable advice based on their substantial experience in carrying out research with children, and we have made some amendments to our proposal based on this advice. This includes developing a procedure for action to be taken if potential safeguarding concerns are flagged as a result of responses given in the survey.

A3 Proposed site of research select all that apply

☑ ONS

☐ SRS (formerly VML)

☐ HMRC Data Lab

☑ Other

Survey contractors – Kantar Public

☐ ADRC - England

☐ ADRC - Scotland

☐ ADRC - Northern Ireland

☐ ADRC - Wales

A4 Data subjects to be studied

Does the study include all subsections of the population (i.e. all ages, sex, ethnic groups etc?) ○ **Yes** ○ **No**

If no please detail which subsections with justification(s) below:

Subsections of the population (including vulnerable groups) the project focuses on:

Children aged 10-15, resident in households in England and Wales.

Justification for focusing on these subsections or groups:

The aim of the project is to provide data on children’s online behaviour and the extent to which they are victims of specific online crimes such as cyber bullying, sexting and online grooming. The survey module will be added to the existing child survey, covering 10 to 15 year olds, already running as part of the CSEW.

The CSEW is necessarily restricted to people living in residential households by the nature of its sampling frame, the Postcode Address File (PAF), which does not include population living in group residences (for example, student halls of residence) or other institutions.

The CSEW is carried out on behalf of ONS in England and Wales only. Crime statistics are a devolved matter and therefore separate surveys are run in Scotland (the Scottish Crime and Justice Survey) and Northern Ireland (the Northern Ireland Crime Survey).

A5 Please provide details of the research protocol or methodology (e.g. data linkage, web scraping etc) (max 500 words)

The proposed new module of questions would be added into the existing child survey which forms part of the CSEW. The survey is designed to be representative of the population of households in England and Wales and people living in those households. It uses the small users' Postcode Address File (PAF), which is widely accepted as the best general population sampling frame in England and Wales. The survey employs a partial cluster sample design in less heavily populated areas, but the design becomes un-clustered over a three-year period. The survey is weighted to adjust for possible non-response bias to ensure the sample reflects the profile of the general population.

The survey is carried out by an external contractor, currently Kantar Public, who employ a panel of interviewers who are specially trained to administer the CSEW. The interviewer visits each sampled address, establishes the eligibility of the household and selects one adult from the household as the respondent. At these sampled addresses the interviewer will also establish whether there are any eligible 10-15-year-olds in the household to interview.

Like the adult CSEW questionnaire, the interviewer administers the main section of the child survey, which consists of a core personal victimisation module that asked respondents about their experiences of a range of crimes over the past twelve months. There is a self-completion module which follows at the end of the face-to-face interview. This module allows respondents to provide answers directly via a tablet computer, on topic areas that they might feel uncomfortable talking about to an interviewer. The existing self-completion module on the 10-15 survey covers subjects such as internet usage, bullying, truanting from school, alcohol and drugs, carrying knives and street gangs.

The new questions will be added to this self-completion module, and will ask children about their experiences, behaviours and perceptions relating to interactions with others in an online environment, specifically in relation to online bullying, speaking with and meeting strangers online, and sending and receiving sexual images/messages.

Given the nature of the topics covered in the new module, the safeguarding of children is of utmost importance. Data collected in the new module relates to past experiences, and therefore cannot identify a current safeguarding issue. However, some survey responses (or combinations of responses) could indicate patterns of behaviour to suggest that a child may be putting themselves at ongoing risk through their online activity. We have designed a new safeguarding approach in collaboration with NSPCC and Kantar Public, considering a range of factors including:

- The extent to which we would be able to infer a current safeguarding concern through the proposed data collection;
- The confidentiality of children's responses, and;
- The possible risks associated with notifying parents of their child's behaviour.

Our safeguarding process will inform both the parent/guardian and the child of an assessment of the child's level of risk based on their responses to the new questions. This assessment will be provided for every child who completes the self-completion module. This process offers a balance between maintaining a degree of confidentiality (the risk rating does not disclose specific responses) whilst still highlighting potential concerns to the parent/guardian and child. The risk rating will be calculated using scores pre-assigned to responses that suggest a level of risk, with thresholds applied to band the scores in a low, medium or high risk category:

- Low risk - the answers given suggest that the child is taking reasonable precautions when engaging in the types of online activity covered in the survey.
- Medium risk - the answers given suggest that the child is taking some precautions when engaging in the types of online activity covered in the survey, but some of their activities might open them up to risk.
- High risk - the answers given suggest that the child has engaged in potentially risky online behaviour in the past and may have had a negative experience as a result of this.

The interviewer will have no access to the responses given by the child, and the risk rating will be calculated post-interview as part of the data processing by Kantar Public. The information will be communicated separately to both parent/guardian and the child in the form of a letter issued by Kantar Public.

The letter will explain that the child's responses have been assessed and given a risk rating along with a tailored explanation of what that particular risk rating means and where further information can be accessed, including details of the NSPCC helpline and website. Each letter will encourage discussion between parent/guardian and child about the child's online activity and how they can stay safe online. The letter will also be tailored for cases where a) it has not been possible to calculate a risk rating because the child has not answered any of the questions, and b) where a child has received a low risk rating but has not answered one or more of the questions (because in both cases non-response could potentially be masking some risky online activity). The number for the ONS Survey Enquiry Line will also be provided for people to use if they have further questions. A separate risk rating leaflet will also be sent out alongside the letter, which will explain more about how the risk rating has been calculated.

Both the parent/guardian and the child will be informed of this procedure as part of the consent process prior to the interview being conducted.

NSPCC support the inclusion of this module in the crime survey and believe the safeguarding and support procedures we have in place are sufficient. NSPCC have been consulted in the development of the risk rating system and believe that it is appropriate and fit for purpose.

A6

Data use

Please specify the data used **by the research team** including any timeframes *e.g. LFS data 2014-15*

Data Level

Please specify the name of the data set

Type of data	**Aggregate Data**	**Identifiable Data**	**De-identified personal data**	**Anonymised/ pseudo anonymised**
Administrative data *(please specify, e.g. Patient Register 2011, School Census 2012 etc, in the relevant options adjacent)*				
Big Data *(please specify e.g. Twitter data, smart meters and mobile phones, in the relevant options adjacent)*				
Survey Data *(please specify e.g.LFS, BRES, etc in the relevant options adjacent)*			Crime Survey for England and Wales (CSEW)	
Census Data *(please specify year, e.g. Census 2011 in the relevant options adjacent)*				
Other *(please specify e.g. Ordinance Survey Address register in the relevant options adjacent)*				

Section B
Assessment against NSDEC ethical principles

B1 Principle 1: The use of data has clear benefits for users and serves the public good.

Please outline the proposed benefits of the project (max 500 words)

Crime statistics produced from CSEW data are used by the Home Office, police forces, other government agencies, academics, and voluntary organisations for a variety of purposes. For example, the development and monitoring of crime and justice policy; public safety campaigns; raising awareness of particular forms of crime; and academic research. They also help to ensure that information on trends in different crime types in England and Wales are available to help inform the choices and decisions of the general public. Further breakdowns, such as offence type and victim characteristics, allow a greater depth of understanding about crime.

While the 10-15 CSEW is well established in terms of its existing topics, it does not include many questions which address the increasing interest and concern around children's internet behaviour. The large variety of online platforms now accessible to children from a young age has led to increased public awareness of children's vulnerability to online predators and to internet crimes such as 'sexting', online bullying and online grooming.

The new questions added to the child survey will be used to explore children's interactions with others in an online environment and the extent to which they are victims of online crimes. Currently the scale of such crimes against young children remains unclear and statistics based on cases referred to the police or support services can only provide a partial picture. Including new questions on the child survey will help address some of this evidence gap, as one of its key benefits is its ability to provide a picture including those incidents that have not been reported to the authorities.

The topics covered are central to ongoing public dialogue about children's use of the internet and social media platforms. Having new data on these issues would help inform this important debate, and further raise awareness of the risks young people can face online.

These new data would also help fill an important evidence gap around the prevalence of certain online behaviours amongst children, and the extent to which children may have been victims of online crimes. There are clear benefits to improving our understanding in these areas, including:

- providing more information to help parents keep their children safe online
- building the evidence base for those in local government and the voluntary sector managing support services for children
- aiding decision makers (both nationally and locally) and informing policy aimed at preventing victimisation of children online.

There is a demand for data in this area across a range of government departments including the Home Office, Department for Education (DfE) and the Department of Culture, Media and Sport (DCMS). There is also substantial interest among the voluntary sector.

B2 Please outline any intended future use for products (such as linked data sets or tools) produced as a result of the research and how they will be accessed. (max 250 words)

CSEW datasets are produced on a rolling quarterly basis. CSEW datasets held by ONS – and shared with Home Office and other government departments – do not contain any identifiable data. This information is only held by the survey contractor, Kantar Public, for the purpose of administering the survey.

The main mechanism for the dissemination of findings is through statistical bulletins, articles and data tables published on the ONS website. It is anticipated that, once we have a full year of survey data from the new module, we would publish an article and accompanying aggregate data tables, to release the new findings into the public domain.

Concerning the release of microdata, the annual CSEW dataset (the quarterly dataset that covers the full survey year of April-March) is shared with the UK Data Archive and the ONS Secure Research Service (SRS). Through these services, researchers can apply for access to the data. These datasets are subject to further statistical disclosure control since individual level information is included.

For the 10-15's data, the file supplied to the UK Data Archive is made available under the archive's End User License. All data from the sensitive questions asked during the self-completion modules are redacted from this dataset, and other potentially disclosive variables are also removed. Therefore, data from the new questions on children's online behaviour and experiences would not be included.

The file supplied to the SRS does include the sensitive variables from the self-completion module, but statistical disclosure control is applied to redact any potentially disclosive variables.

The statistical disclosure control procedures applied have been formally agreed with the ONS statistical disclosure control team.

B3 Principle 2: The data subject's identity (whether person or organisation) is protected, information is kept confidential and secure, and the issue of consent is considered appropriately.

Please outline how data security, confidentiality and informed consent is safeguarded in this project (max 500 words)

Informed consent

In line with other voluntary surveys, the information about the content of the survey contained in the CSEW's advance material is necessarily short and summarised, and they are designed to give a high-level overview of the background and purpose of the survey.

For all children taking part in the 10-15's survey, verbal consent from a parent or guardian and the child is required before conducting the interview, and the full name of the parent/guardian giving consent and their relationship to the respondent is noted. A parental information card containing details of the content of the child interview is given to parents/guardians before obtaining consent for the interview. Based on advice from the NSPCC Research Ethics Committee this has been updated to inform parents of the sensitive topics covered by the survey. This also explains that a risk rating will be provided after the interview, but that no specific responses given by the child will be shared.

Following consent from the parent/guardian, the interviewer explains the survey to the child and gives them a child survey leaflet. This leaflet, which has been designed to be easily understood by children, describes the purpose of the survey, what the information is used for, the confidentiality of the information they give, and details of children's support services. At the start of all interviews, the interviewer is required to confirm that the child has read and understood the information and is happy to continue with the interview. They also emphasise the option to choose not to answer any question in the survey.

At the start of the self-completion module, the child is presented with a screen which again explains that, following completion of the interview, a letter will be sent to them and their parent/guardian(s) about whether their online activity is low, medium or high risk. They are required to select a box on the screen to confirm that they have understood this information before they can move on and complete the self-completion module.

As with the existing survey, proposed questions have been written so that they can be understood by a child. The questions include a 'don't wish to answer' option if the child wants to skip the question. If they don't know the answer to something, they can say 'don't know'. For some of the self-completion questions, the respondent can ask for assistance from the interviewer, if the child gives permission for them to view the screen. However, because of the sensitive nature of the new questions on online behaviour and experiences, these questions will only be asked if the respondent completes the self-completion module on their own. If they ask for help from the interviewer, they will not be asked the new questions.

If a respondent becomes visibly distressed during the survey, the interviewer will suggest that the interview is terminated, with the option to continue if the respondent wants to.

The information leaflets and risk rating letters have been reviewed by ONS's Materials Strategy Advisory Group and ONS Communications. Advice given during these reviews has been taken on board in finalising the materials.

Safeguarding

As with any research involving children, particularly in the home, safeguarding of the child is of paramount importance. Should a CSEW interviewer observe something in the course of their work which leads them to believe that a child is at risk, Kantar Public have a clear Child Protection Policy in place. This policy is consistent with the ONS safeguarding policy in terms of its approach. Where an interviewer is concerned about the safety of a child they are required to report this to a named person in Kantar Public, who will follow a specified procedure to determine an appropriate course of action, in consultation with other designated individuals in Kantar Public. The policy also highlights the need for interviewers to contact the police and social services in cases where there is a clear and immediate danger to a child.

This Child Protection Policy would be applied to situations observed by the interviewer while they are in a respondent's home, or disclosures made by a child, outside of the formal survey. If, for example, at the end of the interview a child discloses to an interviewer that they are being abused, the Child Protection Policy sets out a process for interviewers to report this, and for an appropriate response to be determined and set in motion.

In terms of the new questions on children's online behaviour and experiences, these questions will gather information from respondents about past experiences and therefore will not provide information that would enable anyone to determine that a child is currently at risk of harm. The proposed module contains no free-text answers where the respondent could potentially disclose a current danger. The new safeguarding process of informing the parent/guardian and child of the child's level of risk based on their responses to the survey reflects that we will not be able to identify current danger through the survey responses. It will allow us to highlight potential ongoing safeguarding concerns whilst maintaining a degree of confidentiality for the respondent.

All children who complete the new self-completion module will be presented with a screen at the end providing advice on keeping safe online and details of appropriate support services. The screen will include the top ten tips for staying safe online, taken from www.safetynetkids.org.uk, and will provide a link to where more information on keeping safe online can be found on the NSPCC website. Contact details for Childline will be provided in case a child wishes to speak to anyone about any of the issues raised in the survey.

Confidentiality

The confidentiality of respondents and the information they provide as part of the CSEW is treated with the utmost importance. Respondent's personal information (including name, address and email address) are held securely by Kantar Public, and kept separately from the information that is collected as part of the interview. Kantar's privacy policy states that this information will not be made available to anyone without their agreement, unless it is for research purposes only or if required by law. This policy is in line with GDPR and published on Kantar's website. A link to the policy is included in the survey materials.

Name and address details are replaced by a unique ID number on the final dataset delivered to ONS. Kantar Public will supply an SPSS dataset containing this de-identified CSEW data to ONS on a quarterly basis. ONS are then responsible for further processing of the dataset. All published outputs will be based on aggregate CSEW data and no individual level data will be published.

The confidentiality of the survey is emphasised in the survey's advance materials that is provided to households and respondents. In addition, the use of self-completion for the new questions allows respondents to feel more at ease when answering these sensitive questions as they do not have to disclose their responses to the interviewer.

However, the use of the risk rating approach means that amendments have had to be made to the confidentiality of the survey. If a parent makes a request to see their child's responses after receiving the risk rating, we may have to share this information with them. All guidance around the confidentiality of the module has been updated to reflect this. The children's information leaflet explains that: a letter will be sent (to the child and parent/guardian) with a risk score for their online activity; that if they are under 16, parent/guardian(s) have a right to request access to the child's answers and if they made a legal request, we may have to share the answers the child gave with them.

The informed consent screen at the beginning of the self-completion module, where children are required to indicate that they understand the confidentiality statement before completing the module, also states that their parents have a right to request access to the information they provide and that if they made a legal request, we might have to share their answers with them.

If we are required to share any responses with parents/guardians, we will inform the child that we are doing so.

Data security

All data collected are held on secure systems. Kantar Public adhere to the ISO 27001 standard for the data security, which is the international standard for the management of information security. An anonymised data file is transferred to ONS using a secure data transfer system and held on the secure ONS IT system. This does not include all personal information collected, but does include Output Area (not address) and date of birth. Access to the data file is restricted to named members of ONS staff involved in the analysis and production of crime statistics.

Personal information is retained by Kantar for no longer than 12 months, unless:

- required to do so by law

- where respondents have agreed they can be re-contacted by Kantar for validation or follow-up research
- where respondents have agreed their details can be passed to another research organisation for follow up research.

However, children are not currently asked if they are willing to be re-contacted or if they agree for their details to be passed to another research organisation.

All sharing of CSEW data between ONS and other government departments is governed by Microdata Release Panel (MRP) data sharing agreements. These require that the data are held on secure systems and are only accessible to named individuals. This is in compliance with the Statistics and Registration Service Act 2007. The methods used are compliant with the principles in the Data Protection Act.

In sharing the data, statistical disclosure control procedures are formally agreed with the ONS disclosure control team and applied to the dataset, and all potentially disclosive variables are redacted. Figures produced from an unweighted base of less than 50 respondents are generally suppressed for data quality purposes as the estimates are unlikely to be robust. Estimates based on an unweighted base of less than 3 respondents are always suppressed for statistical disclosure control purposes.

All our policies and procedures for data protection has been reviewed to meet the standards and requirements of the General Data Protection Regulation (GDPR). Accountability and governance measures are in place to ensure that we understand and adequately disseminate and evidence our obligations and responsibilities; with a dedicated focus on privacy by design and the rights of individuals.

B4 Principle 3: The risks and limits of new technologies are considered and there is sufficient human oversight so that methods employed are consistent with recognised standards of integrity and quality.

Please describe how the any risks from new technologies are being mitigated as well as any quality assurance activities in the project (max 500 words)

The CSEW is conducted using a combination of well-developed Computer Assisted Personal Interviewing (CAPI) and Computer-Assisted Self-Interviewing (CASI) programming on a tablet computer.

The survey has been designed specifically with the needs of child respondents in mind. Audio-CASI, where the respondent can listen to the questions and the possible set of answers using headphones, is offered to all children filling in the self-completion module. The main rationale for using Audio-CASI is to help people with literacy problems and reading difficulties complete the section in private. Audio-CASI is offered to everyone as children may feel embarrassed disclosing any reading or learning difficulties.

Data from the survey is held and analysed in a commonly used statistical software package (SPSS).

B5 Principle 4: Data used and methods employed are consistent with legal requirements such as the Data Protection Act, the Human Rights Act, the Statistics and Registration Service Act and the common law duty of confidence

Please describe the legal frameworks pertinent to this project (max 500 words)

ONS require Kantar Public to comply with all legal requirements, including the Data Protection Act and the Statistics and Registration Service Act, for both data collection and storage purposes.

B6 Collaboration and Sponsors

Please describe the project sponsors and the **legal gateways** to acquire, process use and share their data

List of Collaborators/Sponsors	**Details and relevant documentation relating to collaboration (you may attach copies of relevant documentation)**
Home Office	The sharing of data from the CSEW between ONS and Home Office is governed by a Memorandum of Understanding
Office for National Statistics	

B7 Principle 5: The views of the public are considered in light of the data used and the perceived benefits of the research

Please list any public engagement activities (max 250 words)

As part of the scoping and development work, interviews were conducted by Kantar Public with experts in the field to gain a better understanding of online crime affecting children. These interviews included a wide range of individuals from academia, government departments (Home Office, DCMS, DfE) and charities (NSPCC) and the information gathered was used to inform the content of the proposed survey module.

Following this, qualitative work was undertaken through focus groups with parents and depth interviews with 10-15 year olds, with two main purposes:

1. To understand parents' attitudes towards their children's internet activities, including:
    - The types of discussion they have with their children about online safety
    - How their children find out about online security
    - Concerns about their children's activity online
    - Steps taken to protect their children online
    - Views on children participating in survey research of this kind

2. To explore young people's online behaviour and attitudes to security, and any exposure to negative online experiences, including:

- What children use the internet for and how they use it
- The importance of safety concerns to children
- The extent children are concerned about or affected by upsetting experiences online
- Perceptions of risky behaviour
- Parental awareness

Information gathered through this process was used extensively in the development of the questions, to ensure that they used language that children would understand and relate to, and that they would be interpreted as intended. As the testing covered all relevant ages and both genders, we know that the language used in the questions is appropriate for 10-15 year olds. Therefore there is no requirement to use different language for the younger children in this age group compared with the older children.

In focus groups with parents no concerns were raised about the sensitive topics covered in the proposed questions and the response from participants was positive.

Following the introduction of the risk rating approach, a small pilot was carried out by Kantar Public with the aim of:

1. Understanding how children and parents interpret the information about the survey and risk rating
2. Understanding whether the confidentiality statements affect children's responses to the module
3. Understanding the reaction of children and parents to the risk rating and survey materials

The pilot involved an initial interview with the parent, followed by an interview with the child, and then a final interview with the parent.

For parents/guardians the pilot showed that:

- All parents were happy for their child to participate in the survey. Most had no concerns or further questions about the survey, but some were concerned about the level of details the questions would cover and whether they were appropriate to be asked to a child. Once it was explained that the questions had been specifically tailored for children and all questions included an option to not answer, they were happy for their child to participate
- Some parents commented that the survey would be useful to help them and their child to understand online risk
- Most parents thought the risk rating letter was easy to read and understand and appreciated that contact details for ONS and NSPCC were provided. Some said they would like more information such as examples of the online behaviour covered by the survey. As a result, a link to where the questions could be found online was added into the risk rating letter following the pilot
- Most parents had a good understanding of the fact that the risk rating only covered behaviours/experiences included in the survey and not all online behaviour. For some parents there was some concern that their child might be at risk in areas that weren't covered
- All parents said they would discuss the letter and rating with their child. Following the pilot, the parent and child risk rating leaflets were altered to ensure they look sufficiently different from each other, helping to distinguish that the information on each leaflet is intended for a different audience
- Parents were more concerned about high risk ratings and understanding what their child was doing that put them at risk. Many suggested that they would look at the questions and discuss this further with their child.

For children the pilot showed that:

- Some children didn't pay attention to the confidentiality statement and the statement being quite long and wordy made it hard for some children to understand. The confidentiality statement was therefore simplified following the pilot and more clearly signposted on the advance information leaflet
- Most children didn't think about the confidentiality statement when completing the survey, saying that they would still answer honestly. Only one participant admitted that they weren't completely honest when answering some of the questions
- Most children didn't have any concerns about the risk rating system but understanding of what the risk rating meant varied. As a result, more explanations about the risk rating, and the differences between the low, medium and high ratings, were added to the materials provided to parents and children following the pilot
- In general, the children understood the survey questions and didn't have any concerns about them. Children who had experienced bullying found those questions more sensitive but were still willing and able to answer them
- All children thought everyone should receive a risk rating, regardless of their rating. Some commented that it would be important to them to know what their risk rating was, to help them know if they were doing the right things or not or whether they needed helped to stay safe online

We will continue to consider the views of the public in relation to this by carefully monitoring the implementation of the new module, including any impacts on CSEW response rates and data quality. If a significant negative impact is found, we will then decide on appropriate action.

B8 Principle 6: The access, use and sharing of data is transparent, and is communicated clearly and accessibly to the public

How will the findings of the research be disseminated? (max 500 words)

Once sufficient data has been collected from the new questions, it is our intention to publish outputs, most likely in the form of an ONS article and accompanying data tables. Initially these are likely to be badged as Experimental Statistics while we build understanding of how the questions are working. These outputs will be based on de-identified CSEW data which will be used to create aggregate data tables for publication. These will help identify the nature of children's online activities and the scale of online crimes committed against them.

All crime statistics and analysis produced by ONS from both the CSEW and other sources are published on the ONS website under the terms of the Open Government Licence. These are based on aggregate data only and are available to the public free of charge.

Publication dates are planned in advance and pre-announced on the statistics release calendar on both the gov.uk and ONS websites at least 6-8 weeks before the agreed date.

Section C
Responsible owner and applicant details

C1 Responsible Owner

Full Name: Meghan Elkin	**Position:** Head of ONS Centre for Crime and Justice
Address:	**Email:**
	Telephone
	Organisation: Office for National Statistics

Declaration to be signed by the responsible owner

I have met with and advised the applicant on the ethical aspects of this project design (applicable only if the responsible owner is not the Applicant).

I understand that it is a requirement for all researchers accessing the data to have undergone relevant training and to have either relevant security clearances or approved researcher status in order to access the data.

I am satisfied that the research complies with current professional, departmental and other relevant guidelines.

I will ensure that changes in approved research protocols are reported promptly and are not initiated without approval by the National Statistician's Data Ethics Advisory Committee.

I will provide notification when the study is complete if it or fails to start or is abandoned.

I will ensure that all adverse or unforeseen problems arising from the research are reported in a timely fashion to the National Statistician's Data Ethics Advisory Committee.

I will consider all advice received from the National Statistician's Data Ethics Advisory Committee and should I be unable to implement any of the recommendations made, I will provide reasoning in writing to the Committee.

Signature: .... **Date:** ....

C2 Applicant Details (if applicant is not the responsible owner)

Full Name: Alexa Bradley	**Position:** Head of Crimes Against Children, ONS Centre for Crime and Justice
Address:	**Email:**
	Telephone:
	Organisation: Office for National Statistics

UK Statistics Authority

National Statistician’s Data Ethics Advisory Committee

Migration Statistics Transformation Programme

Presentation

Rebecca Briggs

NSDEC(19)02

National Statistician’s Data Ethics Advisory Committee Application for Ethical Review

Please consult the guidance document before filling in the application form

Project Title
Please provide a title indicative of the project

Using linked administrative datasets to understand international migrants’ interactions with the health sector.

Start Date:	January 2019	**End Date:**	March 2020

Project Sponsor(s)
Please list the project sponsor(s)

Office for National Statistics (ONS)

Project Summary
Please provide a brief high level summary of the research giving necessary background (max 250 words)

This research aims to help build a robust evidence base for international migration statistics as set out in the ONS and Government Statistical Service (GSS) transformation programme. The research aims to meet our commitment to provide information users have told us they need to know on how international migrants contribute and interact with the health sector (both as users and as employees), to better inform policy and planning at the UK and sub regional level.

This research will assess the feasibility of using linked datasets to better understand international migrants’ interactions with the health sector as users. We aim to provide a balanced view on these interactions by making comparisons with UK nationals or UK born as appropriate. No diagnosis or treatment codes contained in Hospital Episodes Statistics (HES) will be required for this project. The research is to produce aggregate statistics only.

Current research and evidence available on the contribution international migrants make to the UK health sector is limited. There are no official statistics and no regular releases of data. Most evidence concludes that migrants make a more significant contribution to the NHS workforce than they consume through use of services but acknowledge that there are limited data to quantify this specifically (e.g. Kings Fund, 2015; Migration Advisory Committee, 2018).

There are more data available on migrant NHS employees than on migrant service users with analysis of use of services sometimes focused on proxy measures related to the age and sex structure of different population groups, or analysis of associations rather than on actual known counts.

We will build on work presented by the Administrative Data Census (ADC) Project to NS-DEC in October 2018 on 'Improving administrative based population and migration estimates using HES data'. We aim to link HES to other public sector datasets such as the HMRC/DWP Migrant Worker data and Higher Education Statistics Agency (HESA) data to allow for analysis of specific subgroups (e.g. migrant workers or students). The linkage to other sources is crucial to enable analysis of migrants and improvements in the evidence base.

This research project on service users aims to fill identified gaps in the evidence base by making better use of existing data and present these data in a way that allows for greatest benefit to decision makers and data users to serve the public good. We will:

- consider different definitions of migrants and investigate the definitions possible within the data
- assess data quality, coverage and suitability
- categorise groups of migrants within health datasets using linked administrative data and disaggregate by migrant type (e.g. long-term, short-term or regular migrant)
- assess migrants aggregate use of health services by quantifying their interactions within health datasets, how this varies at a subnational level and against appropriate comparison groups (e.g. UK nationals)
- propose new measures of interaction and recommend how we monitor change over time
- provide regional and where appropriate, non-disclosive local and small area analysis.
- investigate the possibility of understanding multiple service use by linking multiple health service use datasets (e.g. HES and Improving Access to Psychological Therapies (IAPT))

We will be restricted in how we can define migrants based on the data sources we use and this will vary between sources. In our initial analysis of the health data using the PDS we will be able to identify recent migrants i.e. those who have recently arrived after living abroad for 3 months or more. By linking to further public sector data sources we may be able to use nationality, citizenship or travel patterns to further categorise migrant groups i.e. those who are here short-, long-term or make regular journeys for work or other reasons. We will provide clarity on the definitions used throughput this work and work with our expert group to align with practice and policy where possible.

The research aims to provide aggregate data on the international migrant population at different geographical levels in England. We will only use underlying data for the purposes of linking datasets and producing these statistics – ONS does not track, monitor or identify individuals within the data. As HES is a dataset for England this research will be restricted to analysis for England only. Separately we are discussing the possibility of similar work for the devolved nations.

Analysis of existing and new health administrative data sources will contribute fresh insights to the international migration evidence base and validate assumptions made from the data and analysis currently available. Administrative data sources can provide new and act as a regular source of data on how international migrants interact with different sectors and the contributions they make. Robust evidence on the use of health services by different types of migrants will inform the planning of these services at a national and local level. Without this policy and decision makers will need to continue to rely on limited evidence at a critical time when the policy context is changing, largely because of the UK exiting the EU.

Section A

Project Details

A1 Legal gateways

Please provide the assessment of the legal gateways of the project as provided by Legal Services

Section 45A of the Statistics and Registration Service Act provides a permissive power for most Crown bodies and public authorities to share the data they hold in connection with their functions, with ONS. This is the legal gateway used for ONS to receive most of the data required for this project including:

- Patient Register (PR) and Personal Demographic Service (PDS)
- English and Welsh School Census
- Higher Education Statistics Agency (HESA)

In accordance with NHSD request ONS has served a Notice under section 45C SRSA to require the disclosure of the following datasets:

- Hospital Episode Statistics (HES)

In addition, Section 45 provides a similar power specifically for HMRC allowing ONS to receive HMRC/DWP Migrant Worker Scan (MWS) and the HMRC PAYE and Self-Assessment data.

All of the above gateways permit ONS to use the received data for any of our statistical functions.

ONS is still in discussions with NHSD regarding the acquisition of the Improving Access to Psychological Therapies (IAPT) data, but it is envisioned that this will also be obtained by serving a Notice under section 45C SRSA.

A2 Ethical approval

Has the project being reviewed or is it expected to be reviewed by another ethics committee? ○ **Yes** ○ **No**

If Yes please provide the name of the committee, the outcome and the date approved

A3 Proposed site of research select all that apply

☑ ONS

☐ SRS (formerly VML)

☐ HMRC Data Lab

☐ Other

please specify

☐ ADRC - England

☐ ADRC - Scotland

☐ ADRC - Northern Ireland

☐ ADRC - Wales

A4 Data subjects to be studied

Does the study include all subsections of the population (i.e. all ages, sex, ethnic groups etc?) ○ **Yes** ○ **No**

If no please detail which subsections with justification(s) below:

Subsections of the population (including vulnerable groups) the project focuses on:

International migrants

Justification for focusing on these subsections or groups:

The overall aim of the research is to understand the contribution to and interactions international migrants have with the health sector. To ensure data are presented in context we will include appropriate comparator populations such as UK nationals or UK born.

A5 Please provide details of the research protocol or methodology (e.g. data linkage, web scraping etc) (max 500 words)

We are using the detail provided by linking primary and secondary healthcare datasets at a record level to assess international migrants' interactions with the health sector.

Data linked in the Secure Research Environment (SRE) will use a pseudonymisation approach. Data linked in the Data Access Platform (DAP) environment will be linked using personal identifiers such as name, date of birth and sex. The variables required for the feasibility research and to carry out analysis are listed in detail in Annex 1. Feasibility research requires that we make use of a wide range of variables to ensure we are then able to recommend the most appropriate variables for our new measures, once the feasibility research is complete we will look to reduce the list of variables required for future use.

Where possible, data sources will be linked using a common identifier. Hospital Episode Statistics (HES) will be linked to the Patient Register (PR) and Personal Demographic Service (PDS) and when available the Improving Access to Psychological Therapies (IAPT) datasets using NHS number.

Linking HES and PR/PDS will enable the identification of new international immigrants within HES, using the National Health Authority Information System (NHAIS) flag 4 from the PR/PDS, and for the attachment of personal identified to facilitate linkage to further datasets.

In the absence of common identifiers across the other public-sector data sources, records can be linked using a combination of their demographic information (taken from the PR/PDS), such as name, date of birth, sex and postcode. Triangulation of the HES-PR/PDS data to the Migrant Worker Scan (MWS) and other public sector data sources will provide a Nationality variable.

As the project progresses we will explore the feasibility of linkage to other datasets such as Home Office Exit Checks, Higher Education Statistics Agency (HESA), English School Census (ESC) and HMRC PAYE and Self-Assessment. This will allow for the comparison of the usage of the health sector by different parts of the migrant population such as students and workers. Linkage to these wider sources will be dependent on plans for ONS to acquire appropriate access to the data.

We are interested in ‘activity’ based datasets as they inform us of an international migrant’s time in the UK (long-term, short-term or regular) – providing vital information about the services they interact with and use. We will provide context to this analysis of international migrants by presenting a balanced view. We will present appropriate comparison populations, such as UK nationals, and standardise the data as appropriate. We will test current assumptions about the use of services by international migrants when compared with relevant groups.

Outputs from this research will explore the use of these healthcare services, including multiple service use at regional and where appropriate, non-disclosive local and small areas.

This research will complement the work of the Administrative Data Census (ADC) Project (presented to NS-DEC in October 2018), inform the transformation of the migration and population statistics systems, and build on our recent analysis of administrative datasets.

Critically, the research will allow ONS to fill a key evidence gap identified by the Migration Advisory Committee (MAC, 2018) – to allow policy-makers to monitor and evaluate the impact of future policy changes.

Research outputs and dissemination are described in section B8.

A6

Data use

Please specify the data used **by the research team** including any timeframes *e.g. LFS data 2014-15*

Type of data	Data Level *Please specify the name of the data set* Aggregate Data	Identifiable Data	De-identified personal data	Anonymised/ pseudo anonymised
Administrative data *(please specify, e.g. Patient Register 2011, School Census 2012 etc, in the relevant options adjacent)*		Hospital Episodes Statistics 2009-10 - 2017-18 Patient Register 2012-2017 Patient Demographic Service 2016 – 2017 Improving Access to Psychological Therapies 2012-2018 Higher Education Statistics Agency 2011-2017 English School Census 2009-2018 Migrant Worker Scan to 2017 HMRC PAYE and Self-Assessment to 2017/18 Exit checks 2015-2018		Hospital Episodes Statistics 2009-10 - 2017-18 Patient Register 2012-2017 Patient Demographic Service 2016 - 2017 Improving Access to Psychological Therapies 2012-2018 Higher Education Statistics Agency 2011-2017 English School Census 2009-2018 Migrant Worker Scan to 2017 HMRC PAYE and Self-Assessment to 2017/18 Exit checks 2015-2018

Big Data *(please specify e.g. Twitter data, smart meters and mobile phones, in the relevant options adjacent)*				
Survey Data *(please specify e.g.LFS, BRES, etc in the relevant options adjacent)*				
Census Data *(please specify year, e.g. Census 2011 in the relevant options adjacent)*				
Other *(please specify e.g. Ordinance Survey Address register in the relevant options adjacent)*				

We anticipate using the most up to date data available for each dataset as they become available and where appropriate their historical time series, as such some of the dataset dates may change as new data become available over time. Use of and linkage to all these data sources will be dependent on plans for ONS to acquire appropriate access to the data.

Section B
Assessment against NSDEC ethical principles

B1

Principle 1: The use of data has clear benefits for users and serves the public good.

Please outline the proposed benefits of the project (max 500 words)

Following user engagement and consultation (see section B7), in May 2018 our migration statistics transformation update set out plans to look at the interactions migrants have while they are in the UK with the economy and society (including labour market, education and health sectors). These analyses aim to provide further clarity on what available data can tell us on how international migrants interact with different sectors, what contribution they make and what further work we need to do to provide better evidence to better inform policy-makers and planners.

This research project aims to benefit users by filling gaps in the evidence base to improve our understanding of the interactions and contribution of migrants while they are in the UK, particularly at the local level. Specifically, this research project focuses on international migrants use of health services but forms part of a wider project to understand how migrants interact with health and social care services (as both users and employees) and how this information can help build robust migration statistics when combined with other public sector data and provide a balanced view on the contribution international migrants make.

This in turn would benefit the public through the ability of policy-makers and planners to monitor and evaluate the impact of future policy changes and to ultimately ensure improved and more timely allocation of resources (both for services and workforce planning) nationally and locally, now and in the future.

There has been increasing demand for evidence on these interactions as the policy context has changed, largely because of the UK exiting the EU. We therefore need to be able to inform decision-makers in the near term, but also develop data sources and evidence to inform future policy and implementation.

This research project builds on work set out in the ‘Improving administrative based population and migration estimates using Hospital Episode Statistics data’ (presented to NS-DEC in October 2018) and further looks at their interactions within health data sets.

The proposed research is in line with the ONS strategy of ‘Better Statistics, Better Decisions’ and presents a clear benefit in assisting in the development and evaluation of public policy.

Using administrative data and filling evidence gaps benefits the public by ensuring that ONS provides value with statistics that support society’s need for information. This is in line with the Code of Practice for Official Statistics – “Value means that the statistics and data are useful, easy to access, remain relevant, and support understanding of important issues.”

B2 **Please outline any intended future use for products (such as linked data sets or tools) produced as a result of the research and how they will be accessed.** (max 250 words)

The analytical data set, which will include de-identified personal information from the linked datasets, will be held for as long as it is needed for research and statistical purposes such as: reproducibility of the research; time series analysis; further analysis/research to support government policy.

Access to the de-identified data will be limited to those working on this project (the research team). After the project has been completed, we will hold the data in Data Access Platform (DAP) and the Secure Research Environment (SRE). Researchers and data managers with access to the data will have had comprehensive training on access and outputting these types of sensitive data, and will have Security Clearance.

The results of the initial research (and any further developments) will be published on the ONS website in a paper with aggregated tables as appropriate that meet the ONS disclosure control policy. Outputs will be developed to best meet user needs.

The access to all datasets and Research Outputs will be limited to those developing these outputs. The Data as a Service team in ONS monitors the holding of such datasets in the secure environments. This monitoring ensures datasets are reviewed and deleted when they are no longer required.

B3 Principle 2: The data subject's identity (whether person or organisation) is protected, information is kept confidential and secure, and the issue of consent is considered appropriately.

Please outline how data security, confidentiality and informed consent is safeguarded in this project(max 500 words)

The linked data will be stored on an ONS Secure Data Platforms (both in the SRE and DAP). Annex 2 provides detail on the security of both solutions.

The linkage will only be undertaken by staff in ONS, specialised in linking sensitive datasets. In all cases those with access to the data are subject to appropriate security clearance, are required to sign a confidentiality declaration, and undergo rigorous data handling training which emphasises their personal and legal responsibilities. The linked data will be stored in the secure environments.

Consent is not being sought for this study. Consent is not required by law (see processing conditions set out in section B5), nor would it be practical or proportionate to obtain it for ethical reasons. The data being used has already been collected, so seeking consent would require trying to retrospectively contact all individuals.

Results from the research will only be released by ONS staff once they have confirmed that they contain no risk of identifying an individual. All research outcomes will be subject to statistical disclosure controls to mitigate the potential for disclosures of small populations.

The research will provide aggregate statistics on the international migrant population. ONS does not track, monitor or identify individuals within the data.

B4 Principle 3: The risks and limits of new technologies are considered and there is sufficient human oversight so that methods employed are consistent with recognised standards of integrity and quality.

Please describe how the any risks from new technologies are been mitigated as well as any quality assurance activities in the project (max 500 words)

Where the linking of data cannot be made using unique identifiers (NHS number) demographic information will be used to link data using probabilistic match keys. Based on our experience with similar datasets we are confident that the matching should have a high success rate, but until we start the linkage work it is not possible to be certain.

The linkage will use the experience of the Administrative Data Census (ADC) Project who have identified and developed expertise in the appropriate methods. All processes will have adequate human control to ensure appropriate quality assurance arrangements are in place.

B5 Principle 4: Data used and methods employed are consistent with legal requirements such as the Data Protection Act, the Human Rights Act, the Statistics and Registration Service Act and the common law duty of confidence

Please describe the legal frameworks pertinent to this project (max 500 words)

The legal basis for acquiring the HES data will be Section 45A of the Statistics and Registration Service Act. This removes any other duty of confidence on the supplier, including common law duty of confidence.

The project falls within the use of data for the production of statistics and is compliant to the Data Protection Act 2018 and the Statistics and Registration Service Act (SRSA) 2007.

All data have been acquired using established legal gateways cleared by ONS Legal Services (section A1 and B6).

However, the data share must be needed for ONS functions, and not contravene data protection legislation. To assess the former, there is a code of practice underpinning this power that the data share is assessed against. We believe that acquiring this data for this purpose is in line with the code of practice. Acquiring HES and IAPT data is in line with this code of practice, and internal scrutiny was provided by the ONS Data Governance Committee (DGC). The DGC approved acquisition of HES data following ethical approval from NS-DEC for the 2 HES projects in October 2018, and for IAPT following ethical approval on a project in July 2018. The DGC approval was for the whole ONS acquisition of HES and IAPT data and included approval for use for a wide range of purposes, including this one, subject to ethical approval.

In terms of the data protection legislation then the necessary conditions of fair processing have been met as follows:

Main condition:

(e) Public task: the processing is necessary for you to perform a task in the public interest or for your official functions, and the task or function has a clear basis in law.

Special category condition:

(j) processing is necessary for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) based on Union or Member State law which shall be proportionate to the aim pursued, respect the essence of the right to data protection and provide for suitable and specific measures to safeguard the fundamental rights and the interests of the data subject.

To further comply with data protection legislation-

1) ONS will only obtain and use the minimum data needed for the success of project outcomes – i.e. answering the research questions as stated in the project summary
2) Following data linkage researchers will only handle anonymised data
3) Strict, published disclosure control protocols will be adhered to with any outputs produced
4) There is a need for data controllers to inform individuals how their data are being used. For ONS we consider it would be disproportionate effort to notify each person individually, since there will be no direct impact on them. Instead details of the data shares will be published on ONS website along with other privacy information on how we use data and how we keep it secure. In addition, NHS D will publish the Data Access Request Service application form that ONS will complete as part of the process to access the HES data.

ONS have conducted a Data Protection Impact Assessment (DPIA) for health data under the Statistics and Registration Services Act 2007, as amended by the Digital Economy Act, 2017 which is consistent with ICO guidelines and best practice.

B6 Collaboration and Sponsors

Please describe the project sponsors and the **legal gateways** to acquire, process use and share their data

List of Collaborators/Sponsors	Details and relevant documentation relating to collaboration (you may attach copies of relevant documentation)
ONS	
The project falls within the use of data for the production of statistics and is compliant to the Data Protection Act 2018 and the Statistics and Registration Service Act (SRSA) 2007. All data have been acquired using established legal gateways cleared by ONS Legal Services (section A1 and B6).	Section 45A of the Statistics and Registration Services Act 2007 (as inserted by section 79 of the Digital Economy Act 2017) permits any public authority to disclose to the Statistics Board any information they hold in connection with their functions. Information so disclosed may be used by the Statistics Board for any of their functions as set out in the SRSA 2007 and the Census Act 1920. The legal basis for processing information under this agreement is provided by the GDPR under Article 6(1)(e) where the "processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller". and For processing information related to special categories of personal data is provided under Article 9(2)(j) where "processing is necessary for archiving purposes in the public interest, scientific or historic research purposes or statistical purposes in accordance with Article 89(1) based on Union or Member State law which shall be proportionate to the aim pursued, respect the essence of the right to data protection and provide for suitable and specific measures to safeguard the fundamental rights and the interests of the data subject". Personal data are provided under Article 9(2)(j) where "processing is necessary for archiving purposes in the public interest, scientific or historic research purposes or statistical purposes in accordance with Article 89(1) based on Union or Member State law which shall be proportionate to the aim pursued, respect the essence of the right to data protection and provide for suitable and specific measures to safeguard the

	fundamental rights and the interests of the data subject".
Cross-Government steering group for migration statistics and the GSS Migration Statistics Working Group (collaborators)	This projects forms part of the GSS wide migration statistics transformation programme. The Cross-Government steering group for migration statistics will set the strategic direction for the full range of migration statistics and analysis produced across government to provide a co-ordinated and coherent approach; and oversee the effective implementation of the migration statistics transformation programme. The GSS Migration working group reports to the steering group. Membership of the groups include data owners (i.e. NHS Digital) and policy departments (e.g. Home Office and Department of Health and Social Care). These groups will not have direct access to any of the data outlined in this application, access to the data will be restricted to ONS project staff. The groups will provide appropriate expert (data and policy) input and will be consulted to ensure the work aligns with the overall transformation programme. Non-disclosive aggregated tabulations of the results of this project will be initially shared with the GSS Migration Statistics Working Group for quality assurance purposes and treated as official sensitive until sign of and publication. See also section B8.

B7 Principle 5: The views of the public are considered in light of the data used and the perceived benefits of the research

Please list any public engagement activities (max 250 words)

This research forms part of plans to transform the information that the Government Statistical Service (GSS) produces on migration to meet the changing user needs, and which will put administrative data at the core of our evidence on international migration (UK) and on population (England and Wales) by 2020.

The importance of improving the statistics and the value of the GSS programme of work has been recognised by the Home Affairs Select Committee's report and the government's response.

This research and the migration statistics transformation programme are supported by several published papers, updates and engagement activities taking place over the last few years.

- Consultation on International Migration Statistics Outputs (2016/17) – this specifically highlighted the need for local data and using administrative data

- International migration data and analysis: Improving the evidence, February 2017 – outlined the initial ambition for developing international migration statistics across the GSS
- Population and Public Policy forum in September 2017 – a broad community of stakeholders contributed to the migration programme of work
- Office for Statistical Regulation round table events in 2017/18 on 'Improving the public value of the UK's international migration statistics'
- Migration statistics transformation update: May 2018
- Report on international migration data sources: July 2018
- Building our understanding of the migration evidence (section 9, August 2018)

In January 2019 ONS published a research engagement report on their findings from the feasibility research on using linked administrative data to provide international migration flows and will be seeking feedback on user needs across the international migration evidence base. Several engagement activities are planned in early 2019 to ensure this research and the wider programme will meet user needs to serve the public good.

Regarding public acceptability, the 2017 National Data Guardian Review of Data Security, consent and opt outs, and the subsequent DHSC public consultation, included public acceptability research on the sharing of health data. The NDG review stated:

> 'people hold mixed views about their information being used for purposes beyond direct care. Some are concerned primarily with privacy and are suspicious that information might be used by commercial companies for marketing or insurance.

The NDG review also recommended that data flowing to ONS for the purposes of official statistics be exempt from any opt out, and DHSC have since adopted this position in policy – meaning the data will still flow even if someone has opted out of their data being shared beyond their direct care.

This, alongside the passing of the statistics section of the Digital Economy Act bill itself, shows the will of parliament is for the national statistics institute to be able to acquire and link data for better statistics to support better decisions in the public interest (and that there is trust in ONS to keep the data secure and be independent).

B8 **Principle 6: The access, use and sharing of data is transparent, and is communicated clearly and accessibly to the public**

How will the findings of the research be disseminated? (max 500 words)

The ONS Centre for International Migration have an agreed internal communications strategy and will work closely with colleagues across the ONS and the GSS to ensure that our use of these data and the research findings are disseminated fully via the ONS website and other appropriate channels (e.g. conference presentation).

This project forms part of a wider workstream on the interactions of international migrants with the economy and society. Outputs will be coordinated across the workstream and overall programme to ensure the greatest reach, ease of interpretation for users and written in a way as to limit misinterpretation of findings.

We continue to develop our understanding of health data through the Statistical Quality and Health Data Working Groups where we discuss the data and the detail of the variables with data suppliers such as NHS Digital. This also allows us to understand any changes to the systems and the effect these could have on the data supplied.

The HES data will be acquired under section 45A of the SRSA 2007, as amended by the Digital Economy Act 2017. All datasets acquired under section 45 of the SRSA will be explained and justified on the ONS website. In addition, NHS Digital will publish the Data Access Request Service application form that ONS will complete as part of the process to access the HES data. This is in line with the transparency principle within the code of practice that underpins these powers.

To ensure that messages are clear, simple and to limit misinterpretation of the research we will work closely with data owners, statistical and analytical colleagues across the GSS such as NHS Digital via our GSS Migration Statistics Working Group. The non-disclosive aggregated tabulations of the results of this project will be initially shared with this Working Group for quality assurance purposes. As the research forms part of a GSS wide migration statistics transformation programme research findings will be discussed within this group and research written up considering the current policy context to best inform decision making.

This will help us to ensure that the work is completed in line with the Code of Practice by presenting statistics, data and explanatory material impartially and objectively. Further to this we have recently published a series of revised principles and policies on the use, management and security of data. These set out how we use, manage and secure data, while providing the best standard of statistical information for the public.

Part of the feasibility research will be to understand the quality and coverage of the health data as well as the quality of linkage to other data sources. This will enable us to make recommendations about the appropriate use in the context of understanding the strengths and limitations of the data and any linkages we make.

Our colleagues across the GSS and the Department of Health and Social Care are fully supportive of this work. We will use their subject knowledge expertise in the source datasets to help explain patterns or trends, and help identify any areas for further investigation.

We regularly consult users for feedback on the type and format of our outputs. We will work closely with our publishing support and content design team to ensure that outputs are tailored to key audiences and to inform public debate. We are likely to need to produce a range of outputs to meet different user needs (e.g. technical reports describing the data to high level summary findings) and will be asking users for their input as part of our engagement work over the next few months.

Section C
Responsible owner and applicant details

C1 Responsible Owner

Full Name: Rebecca Briggs	**Position:** Head of Migration Analysis (Grade 6)
Address:	**Email:**
	Telephone:
	Organisation: ONS

Declaration to be signed by the responsible owner

I have met with and advised the applicant on the ethical aspects of this project design (applicable only if the responsible owner is not the Applicant).

I understand that it is a requirement for all researchers accessing the data to have undergone relevant training and to have either relevant security clearances or approved researcher status in order to access the data.

I am satisfied that the research complies with current professional, departmental and other relevant guidelines.

I will ensure that changes in approved research protocols are reported promptly and are not initiated without approval by the National Statistician's Data Ethics Advisory Committee.

I will provide notification when the study is complete if it or fails to start or is abandoned.

I will ensure that all adverse or unforeseen problems arising from the research are reported in a timely fashion to the National Statistician's Data Ethics Advisory Committee.

I will consider all advice received from the National Statistician's Data Ethics Advisory Committee and should I be unable to implement any of the recommendations made, I will provide reasoning in writing to the Committee.

Signature:

Date:

C2 Applicant Details (if applicant is not the responsible owner)

Full Name: Nicola Rogers	**Position:** Principal Social Researcher (G7)
Address	**Email:**
	Telephone:
	Organisation: ONS

Annex 1: Variables to be held on the final analysis file

List of Hospital Episodes Statistics (HES) variables

The variables form part of three data collections which form the whole HES dataset namely data on admitted patient care (APC), accident and emergency (A&E) and outpatient appointment. Usually variables about 'admissions' refer to admitted patient care where patients may stay overnight or be seen as a day case, 'attendances' refer to attendance at A&E and appointments to outpatient appointments. People may not appear on any collection, or could appear on one, two or all three.

Admission (ADM_CFL, ADMIDATE, ADMIFLAG, ADMISORC, ELECDATE, ELECDUR, FAE)

Information on admission is needed as it will allow us to identify interactions with the health service. The date will be vital for understanding, in aggregate, the number of interactions over a given period. The admission flag variables will also provide confirmation that an interaction was valid. The ELECDATE variable is the date on which a consultant decides to admit a patient and will help understand waiting times which will be important to explore.

Administrative Category (ADMINCAT, ADMINCATST)

These variables identify whether the records are associated with an NHS or a private patient. This is particularly important as it will allow us to compare, in aggregate, interactions between these records as well as identify records that might not appear on other datasets due to accessing private healthcare.

A&E Attendance (AEARRIVALMODE, AEATTENDCAT, AEATTENDDISP)

The three variables listed will allow us to understand in aggregate admittances to A&E, whether there had been previous interactions or transfer onto another service. This is essential for understanding interactions with different health services.

A&E Linkage Key (AEKEY, AEKEY_FLAG, MATCH_RANK)

These variables are vital for linking different episodes and interactions by the same patient together. It will allow us to link between the A&E and APC datasets which will help us to analyse, in aggregate, use of multiple health services.

Age (ACTIVAGE, ADMIAGE, ARRIVALAGE, ENDAGE, MATAGE, STARTAGE)

There are six variables on age that will be in the final analysis dataset. Each of these variables will give us an indication of age at different points during interactions with the health services e.g. on arrival, on admission, on discharge etc. We will also use these variables to quality assure dates of birth recorded on the system.

Antenatal (ANASDAT, ANTEDUR)

This information about antenatal appointment and days of stay would be key indicators of measuring interactions with a specific part of the health service.

Appointment & Attendance (APPTDATE, ATENTYPE, ATTENDED, ATTENDKEY, ATTENDKEY_FLAG, DNADATE)

The six variables listed above will be essential for quantifying interactions, in aggregate, with the health service for outpatients. They will provide vital information on dates of interactions, number of interactions, and whether appointments were attended (sign of activity).

Date of Birth (DOB, DOB_CFL, DOBBABY_N, MOTDOB, MOTDOB_CFL, MYDOB)
There are several variables associated with dates of birth. The six listed above either relate to the patient, baby or mother so provide different information that is essential for us to use when linking records and quality assuring age. Age is a key characteristic that will inform our aggregate analysis of how different population groups interact with the health service.

Discharge (DIS_CFL, DISDATE, DISDEST, DISFLAG)
The four variables listed above will be vital for understanding length of interactions with the health service. By connecting episodes, we can explore how many interactions different population groups have with the health sector and for how long.

Episode Dates & Duration (BEDYEAR, EPIDUR, EPIE_CFL, EPIEND, EPIS_CFL, EPISTART)
The six variables listed above will be essential for quantifying interactions with the health service for patients within the APC data. They will provide vital information for the date of interactions, number of interactions and crucially how long episodes last.

Episode Type (EPIKEY, EPIORDER, EPISTAT, EPITYPE)
Episode type is essential for us to understand the type of service being used whether it's general, birth or maternity. This along with the other variables such as episode order will be key to us analysing interactions over time by espisode type.

Ethnicity (ETHNOS, ETHRAW, ETHRAWL)
Ethnicity is important for analysis which we will use to control for differences between population groups. There are three variables here that each contain a different level of detail and we need to assess completeness and recommend how they should be used.

Geography (AT_RESIDENCE, GPPRSTHA, HOMEADD, PCFOUND, POSTDIST, RESCTY, RESHA, RESLADST)
Geographies will form a vital part of our analysis. Information on postcode is vital for linking to other datasets. The other variables listed above also correspond to different levels of geography e.g. local authority district, health authority and country. These will be important for us when looking at differences in characteristics between populations i.e. where they live and service use.

HES ID (ENCRYPTED_HESID, HESID_ORIG, MATCH_ID)
This ID number will be useful to link patients' records across a number of years. This will be vital for analysing a series of interactions over time or between collections to better understand health service interactions.

Maternity (BIRORDER_N, MATERNITY_EPISODE_TYPE, NUMBABY, NUMPREG, NUMTAILB)
The variables above relating to births and maternity will be vital for understanding the use of these health services. Information on number of births, previous pregnancies should help us to connect episodes and quantify how interactions might differ between different populations.

NHS Number (NEWNHSNO, NEWNHSNO_CHECK, NHSNOIND)
This is an essential variable that is needed for linkage to other datasets. The NHS number status and check variables will also allow us to quality assure where matches might be missing.

Other Dates (ARRIVALDATE, FYEAR, INYRFLAG, OPDATE_NN, PARTYEAR, PEREND, SUBDATE)

Understanding dates of interactions is vital for our analysis. It will allow us to explore, in aggregate, length of interaction with the health service. There are a number of different date variables within HES and we have included them here as each brings something different to the analysis; for example date on arrival, referral, discharge etc. We can therefore also test the quality of dates within the data and understand which dates are relevant and of high quality for our analysis.

Referral to Treatment (RTTPEREND, RTTPERSTART, RTTPERSTAT)
Information on referral to treatments would be a really useful sign of interaction with the health service. Understanding the types of interactions between referrals and treatments will be important to see how different population groups use the health service.

Sex (SEX, SEXBABY_N)
This is an essential analysis variable which will allow us to make comparisons between different population groups.

Spell (SPELBGIN, SPELDUR, SPELEND)
A spell is made up of a number of episodes, marking the entire journey of a patient through the health service from admission to discharge. This is vital for us in quantifying interactions, in aggregate, and provide information on number of interactions as well as how long these last.

List of Improving Access to Psychological Therapies (IAPT) variables

Activity Suspension Start & End Dates, Identifier and Reason
These variables will allow us to identify, in aggregate, the end of interactions with this health service. The reason may also be important for us as it could indicate how many people may have left the country for a period of time and be used to quality assure similar indicators in other data sources.

Appointment Date, Type and Attended Code
The three variables listed above will be essential for quantifying interactions with the health service. They will provide vital information for the date of interactions and crucially, if appointments were attended, a sign of activity.

Characteristics (Disability, Ethnicity, Date of Birth, Gender, Postcode)
The characteristics listed above will be vital for making comparisons between different population groups and how they interact with health services. Date of birth will allow us to derive age and postcode different geographies in order to better understand interactions with the health sector. These characteristics will also be vital for linkage to other data sources.

Clinical Contact Duration of Appointment
This information will be useful as a potential measure of the extent/intensity of health use or service provided.

Consultation Medium Used
This variable could potentially be useful to see, in aggregate, use of the service. It could also give an indication of the extent/intensity of the use of the service.

Employment Status, Support and Statutory Sick Pay Indicator

Information about employment could be useful when linking to other data sources as a form of quality assurance within the data. These variables will also be important for exploring patterns of health service use related to employment activity. This will form part of the characteristics analysis which will be vital for making comparisons between different population groups and how they interact with health services.

Other Dates (End date, Opt-in Date, Referral Date)

Understanding dates of interactions is vital for our analysis. It will allow us to explore length of interactions with the health service as well as how they may have changed over time. There are a few different date variables within IAPT and we have included them here as each brings something different to the analysis; for example date on arrival, referral, end of treatment etc. We can therefore also test the quality of dates within the data and understand which dates are relevant and of high quality for our analysis.

Source of Referral for Mental Health

This variable will be explored to see how it could potentially be used as a sign of activity indicator.

Therapy Type

This information will be useful as a potential measure of the extent/intensity of health use or service provided.

Unique ID (NHS Number & Local Patient ID)

These variables will be key for linkage. The NHS Number can be used to link to other health datasets and the local patient ID used to link episodes within IAPT. This will enable us to better understand interactions, in aggregate, across the health service.

Annex 2: Overview of Data Access Platform (DAP) the Secure Research Environment (SRE)

Statistical data processing

Advances in technology and statistical methods create enormous opportunities to use new sources of data for the public good. With the Digital Economy Act, there is now a legal framework for ONS to increase its data sources from external organisations, both within Government and in the commercial sector.

To support this data transformation, ONS has developed an integrated, single environment – the Data Access Platform (DAP) – to host data and analytical applications. DAP facilitates the processing and analysis of more data in richer and more complex forms, integrating administrative and commercial data sources supported by appropriate methods and standards.

The Statistical Research Environment (SRE) has been designed to store and process all the information being used for our research. The SRE has been designed specifically to address the privacy and security concerns that may arise when statistics are produced using data from multiple administrative sources. The SRE is located on an ONS protected site. It may be accessed only by authorised and security cleared researchers, data processing and security staff. All access is recorded, monitored and regularly audited by ONS Security Managers by reviewing technical, procedural and CCTV records. Measures have been put in place to ensure that all the data we access and use is protected appropriately. More specifically, arrangements comply with the terms and conditions set out in our service level agreements with individual data suppliers, with mandatory government standards, and our statutory responsibilities.

The SRE is used for all research and statistical work being carried out by those working for the Admin Data Census project. At present this includes the following tasks or activities:

- pre-processing (i.e. initial quality checks, standardisation and georeferencing of key variables, derivation of new variables, creation and addition of match keys, transformation
- production of summary statistics that cannot be created after data transformation (e.g. distribution of month of birth)
- data linkage and matching
- production of statistical population datasets
- coverage assessment and adjustment
- derivation of trial outputs (currently population estimates at local author
- level but in the longer term will include details for smaller areas as well as estimates for a range of socio-demographic characteristics)
- validation and quality assurance of statistical outputs
- disclosure checks
-

Data security

Security of data is a key priority for ONS. DAP takes a robust approach to security that is risk-based and holistic, covering people, process and technology. DAP security is based on two key security governance and management layers. The first layer is a set of security principles to inform design and operation; the second layer distils these principles into specific security controls within the platform.

All security controls have been developed following recognised security standards and guidance from within Government, including from the Cabinet Office, the National Cyber Security Centre and the Centre for the Protection of National Infrastructure. They also adhere to international standards and best practice, as set by ISO 27001 and the Information Security Forum.

The SRE is fully isolated from all other systems and networks whether external or internal to ONS. Within the environment, technical safeguards exist to ensure only authorised research can take place. Any ‘unusual’ activity is detected, assessed and acted on. Unauthorised devices, software or connections are not permitted in the SRE under any circumstances and protective measures are in place to enforce this policy. The SRE has been assessed, tested and fully accredited in accordance with requirements specified by the UK government’s National Technical Authority for Information Assurance (CESG) to store, process and protect the data currently being used for research purposes by the Admin Data Census project. Arrangements are in place to ensure that the accreditation is reviewed as necessary to take account of any future data acquisitions.

NSDEC(19)03

National Statistician's Data Ethics Advisory Committee

Application for Ethical Review

Please consult the guidance document before filling in the application form

Project Title

Please provide a title indicative of the project

Linking Welsh cancer data to the ONS Longitudinal Study

Start Date:	01.01.2016	**End Date:**	None

Project Sponsor(s)
Please list the project sponsor(s)

Office for National Statistics

Project Summary

Please provide a brief high level summary of the research giving necessary background (max 250 words)

The ONS Longitudinal Study (LS) is a study containing linked census and life event data for a one per cent sample of the population of England and Wales. It was started with a sample from the 1971 Census. Data from subsequent censuses, civil registration events relating to births and deaths, NHS registration events and cancer registrations have been added since Census Day in 1971.

This application seeks ethical approval to include Welsh cancer data in the LS from 2016 onwards.

The cancer data that is already held in the LS originated from cancer registries in England and Wales. ONS played a key role in the annual processing of these data and held the data on the Model 204 (M204) system. ONS withdrew from this role when M204 was decommissioned following the processing of 2015 cancer data.

To be able to include Welsh cancer data in the LS from 2016 onwards, ONS needs to establish the supply of data from the Welsh Cancer Intelligence & Surveillance Unit (WCISU). Gaining ethical approval for the proposed use of Welsh cancer data is an essential part of the data acquisition process.

Section A
Project Details

A1 Legal gateways

Please provide the assessment of the legal gateways of the project as provided by Legal Services

Sections 44 and 45A of the Statistics and Registration Service Act 2007

A2 Ethical approval

Has the project being reviewed or is it expected to be reviewed by another ethics committee? ○ **Yes** ○ **No**

If Yes please provide the name of the committee, the outcome and the date approved

A3 Proposed site of research select all that apply

☑ ONS

☑ SRS (formerly VML)

☐ HMRC Data Lab

☐ Other

please specify

☐ ADRC - England

☐ ADRC - Scotland

☐ ADRC - Northern Ireland

☐ ADRC - Wales

A4 Data subjects to be studied

Does the study include all subsections of the population (i.e. all ages, sex, ethnic groups etc?) ● Yes ○ No

If no please detail which subsections with justification(s) below:

Subsections of the population (including vulnerable groups) the project focuses on:

Justification for focusing on these subsections or groups:

A5 Please provide details of the research protocol or methodology (e.g. data linkage, web scraping etc) (max 500 words)

ONS works in collaboration with NHS Digital (NHSD) to link ‘new’ data with the data already held in the LS. This process involves ONS sending identifiable personal data (e.g. name, post code, date of birth) to NHSD to be traced on the MIDAS system. The results of this tracing activity enable the linkage of data within the LS.

ONS does not retain the names of data subjects once the linkage has been completed. The anonymised, fully-linked dataset, including cancer data, is held on a secure server at ONS and can only be accessed by a small number of people.

A bespoke data extract is created for each new research project. These data extracts only include the subjects and variables required for the research in question. Analysis takes place within ONS’s Secure Research Service (SRS) by Approved Researchers and outputs are subjected to clearance checks to ensure they are not disclosive before being released out of the SRS. All LS projects need to be approved by the ONS Microdata Release Panel.

A6

Data use
Please specify the data used **by the research team** including any timeframes *e.g. LFS data 2014-15*

Type of data	**Data Level** *Please specify the name of the data set* **Aggregate Data**	**Identifiable Data**	**De-identified personal data**	**Anonymised/ pseudo anonymised**
Administrative data *(please specify, e.g. Patient Register 2011, School Census 2012 etc, in the relevant options adjacent)*			i) Civil registration data relating to births and deaths, 1971–2016 ii) NHS registration data, 1971–2016 iii) Cancer data, 1971-2015	
Big Data *(please specify e.g. Twitter data, smart meters and mobile phones, in the relevant options adjacent)*				
Survey Data *(please specify e.g.LFS, BRES, etc in the relevant options adjacent)*				
Census Data *(please specify year, e.g. Census 2011 in the relevant options adjacent)*			1971, 1981, 1991, 2001 and 2011 – the full census record of each LS member and anyone living in the same household as an LS member	
Other *(please specify e.g. Ordinance Survey Address register in the relevant options adjacent)*				

Section B

Assessment against NSDEC ethical principles

B1 Principle 1: The use of data has clear benefits for users and serves the public good.

Please outline the proposed benefits of the project (max 500 words)

The LS is a highly valued resource with a long history of research carried out in the public interest, most notably in the field of mortality and health inequalities. It has been used many times to investigate aspects of cancer incidence and survival. The ability to link cancer data to census data provides a very powerful tool to understand how social factors relating to the individual, their household structure and their physical environment can influence outcomes.

Current projects include:
i) *Adult cancers near overhead power lines* - Elliott, Toledano and Shaddick;
ii) *Causal effect of education on cancer risk and survival* - Potente and Monden
iii) *Variations in bowel cancer survival by individual characteristics and area type* - Sturley, Downing, Norman and Morris

Previously published work based on the LS includes:
i) *A prospective study of the independent effects of parity and age at first birth on breast cancer incidence in England and Wales* – Leon - International Journal of Cancer – 1989
ii) *Socio-economic differentials in cancer among men* - Davey Smith, Leon, Shipley and Rose - International Journal of Epidemiology – 1991
iii) *A descriptive study of occupation and bladder cancer in England and Wales* – Dolin - British Journal of Cancer - 1992
iv) *The incidence of cancers among second generation Irish living in England and Wales* – Harding - British Journal of Cancer – 1998
v) *Living arrangements and place of death of older people with cancer in England and Wales: a record linkage study* - Grundy, Mayer, Young and Sloggett - British Journal of Cancer - 2004

In response to an outputs consultation ONS carried out in 2010, the LS received widespread praise and support, including:

- *ONS Longitudinal Study is world class resource for all areas of epidemiology* – Prof. David Blane
- *The Longitudinal Study has provided insights into longer terms trends in relation to data recorded in the census, which cannot be derived in other ways* – Prof. Alison Macfarlane
- *The Longitudinal Study is one of the largest long-running cohorts worldwide, with follow-up health data for nearly 40 years, this makes it extremely valuable to the work of SAHSU [Small Area Health Statistics Unit]* – Dr. David Stokes

B2 Please outline any intended future use for products (such as linked data sets or tools) produced as a result of the research and how they will be accessed. (max 250 words)

Welsh cancer data will be fully integrated into the LS and made available to Approved Researchers for research projects that are considered to serve the public good. The ethics of these projects will be considered through the ethical self-assessment tool. Those projects considered ethically risky will be considered by NSDEC.

The methods in place to control and manage access to the data are outlined in B3.

B3 Principle 2: The data subject's identity (whether person or organisation) is protected, information is kept confidential and secure, and the issue of consent is considered appropriately.

Please outline how data security, confidentiality and informed consent is safeguarded in this project(max 500 words)

From its inception in the 1970s, it has always been recognised that the LS is a very sensitive dataset and that access to it needs to be strictly controlled. The LS now holds data on approximately 1.2 million members, each having their own unique combination of data.

Access to and use of LS data is governed by a range of controls:

- The master version of the LS database can only be accessed by two people from specific desks.
- Name and address information from the various data sources is only retained while processing of that data takes place.
- After each annual processing cycle, a subset of the data is created to support research - the LS Research Database (LSRD) – which can only be accessed by LS User Support Officers (USOs).
- The LSRD does not include any direct identifiers and access to certain variables (date of birth, day and month of death, codes identifying geography at a low level) is restricted to a small number of ONS staff working at specific desks.
- USOs create bespoke extracts of data for researchers that only include the data subjects and variables required for their research.
- Researchers wishing to make use of the LS must be accredited as an Approved Researcher before they are given access to their data extract.
- LS projects need to be approved by both the Microdata Release Panel, which oversees the provision of identifiable data to organisations outside ONS, and the LS Research Board, which reviews whether the LS is the most appropriate data source for the research being proposed.
- Approved Researchers can only access LS data via ONS's Secure Research Service (SRS) facility.

- Approved Researchers are unable to copy, print or extract data from the SRS and must submit all required outputs from their research for clearance by ONS staff – this ensures there is no risk of disclosure of an individual's information.

B4 **Principle 3: The risks and limits of new technologies are considered and there is sufficient human oversight so that methods employed are consistent with recognised standards of integrity and quality.**

Please describe how the any risks from new technologies are been mitigated as well as any quality assurance activities in the project (max 500 words)

No new technologies will be used.

B5 **Principle 4: Data used and methods employed are consistent with legal requirements such as the Data Protection Act, the Human Rights Act, the Statistics and Registration Service Act and the common law duty of confidence**

Please describe the legal frameworks pertinent to this project (max 500 words)

Sections 44 and 45A of the Statistics and Registration Service Act (SRSA) 2007 form the legal basis for the Welsh Cancer Intelligence & Surveillance Unit to share these data with ONS for inclusion in the LS.

The common law duty of confidence is not breached through disclosure of the information to ONS because there is a strong public interest in the disclosure and because there will be no direct effect, detrimental or otherwise, on any individual as a result of the disclosure.

Further, confidentiality is maintained since in accordance with section 39 of the SRSA, personal information held by ONS in relation to the exercise of any of its functions cannot be disclosed and must remain confidential. Unlawful disclosure of the information by ONS would constitute a criminal offence.

In any event, any disclosure of information under section 45A of the SRSA does not breach any obligation of confidence owed by the public authority making the disclosure.

B6 Collaboration and Sponsors

Please describe the project sponsors and the **legal gateways** to acquire, process use and share their data

List of Collaborators/Sponsors	**Details and relevant documentation relating to collaboration (you may attach copies of relevant documentation)**
Welsh Cancer Intelligence & Surveillance Unit (WCISU)	ONS is drafting an application for Welsh cancer data via WCISU's Data Request & Analysis Service. This process requires evidence of 'Ethical Committee Approval', hence this submission to NSDEC.

B7 Principle 5: The views of the public are considered in light of the data used and the perceived benefits of the research

Please list any public engagement activities (max 250 words)

While we have not sought the views of the public for this specific work, there has been a long history of public good research using the LS.

The statistics section of the Digital Economy Act (DEA) itself, shows the will of parliament is for the national statistics institute to be able to acquire and link data for better statistics to support better decisions in the public interest (and that there is trust in ONS to keep the data secure and be independent). Health data are in scope for the statistics strand of the DEA powers, and there was extensive and transparent public consultation during the parliamentary passage of these powers and the codes of practice which underpin them.

For ONS's part, research was conducted into the public acceptability of ONS acquiring and linking admin data for the Admin Data Census programme. It found that:

- Around three quarters of people do not object to data held by other government departments being shared with ONS
- The public are supportive of data sharing when personal or public benefits can be demonstrated and are communicated effectively
- Data linking and storage are more acceptable if personal data are anonymised (i.e. name, address and other personal identifiers are removed)
- Public confidence in ONS is high, with 78 per cent stating that they think the organisation adequately protects the confidentiality of the personal information it collects
- When provided with reassurance about security and privacy, the public broadly support ONS re-using administrative data to produce statistics

B8 **Principle 6: The access, use and sharing of data is transparent, and is communicated clearly and accessibly to the public**

How will the findings of the research be disseminated? (max 500 words)

In order to be accredited with Approved Researcher status, a researcher must agree to their inclusion in a list of all ONS Accredited Researchers and projects, published on the ONS website. They must also agree to publish the results of all research projects completed through the Approved Researcher scheme. ONS maintain a public record on the ONS website of Accredited Researchers, their institutions and projects under the Approved Researcher scheme, see: www.ons.gov.uk/file?uri=/aboutus/whatwedo/statistics/requestingstatistics/approvedresearcherscheme/arlistjanuary2019forpublication.xls

ONS openly promotes the existence of the LS, the data that it includes and the types of research and analysis that it is used for. There is a dedicated section on the ONS website about the LS here: www.ons.gov.uk/aboutus/whatwedo/paidservices/longitudinalstudyls.

Support for users of the LS is also provided by a team funded by the Economic & Social Research Council (ESRC). The Centre for Longitudinal Study Information & User Support (CeLSIUS) is based at University College London (UCL) and is headed by Professor Nicola Shelton. The staff that provide face-to-face support to LS users are UCL employees, but work at ONS's London office at Drummond Gate. CeLSIUS also has a website describing the LS here: www.ucl.ac.uk/iehc/research/epidemiology-public-health/research/celsius.

Members of the LS team at ONS and the CeLSIUS team regularly promote use of the LS at conferences and seminars. The principal aim of researchers using the LS for their projects is to publish their work in print or online journals, or to present their work at conferences.

Section C

Responsible owner and applicant details

C1 Responsible Owner

Full Name: Rich Pereira	**Position:** Deputy Director, Population Statistics Division, Office for National Statistics (also Information Asset Owner of the Longitudinal Study)
Address:	**Email:**
	Telephone:
	Organisation: Office for National Statistics

Declaration to be signed by the responsible owner

I have met with and advised the applicant on the ethical aspects of this project design (applicable only if the responsible owner is not the Applicant).

I understand that it is a requirement for all researchers accessing the data to have undergone relevant training and to have either relevant security clearances or approved researcher status in order to access the data.

I am satisfied that the research complies with current professional, departmental and other relevant guidelines.

I will ensure that changes in approved research protocols are reported promptly and are not initiated without approval by the National Statistician's Data Ethics Advisory Committee.

I will provide notification when the study is complete if it or fails to start or is abandoned.

I will ensure that all adverse or unforeseen problems arising from the research are reported in a timely fashion to the National Statistician's Data Ethics Advisory Committee.

I will consider all advice received from the National Statistician's Data Ethics Advisory Committee and should I be unable to implement any of the recommendations made, I will provide reasoning in writing to the Committee.

Signature: .... **Date:** ....

C2 Applicant Details (if applicant is not the responsible owner)

Full Name: Jim Newman	**Position:** Head of Longitudinal Study Development, Population Statistics Division, Office for National Statistics
Address:	**Email:**
	Telephone:
	Organisation: Office for National Statistics

NSDEC(19)04

National Statistician's Data Ethics Advisory Committee

Application for Ethical Review

Please consult the guidance document before filling in the application form

Project Title

Please provide a title indicative of the project

Data for children: assessing the impact of family background on the outcomes of children in England

Start Date: 28/01/2019 **End Date:** 31/08/2019

Project Sponsor(s)
Please list the project sponsor(s)

Office for National Statistics Centre for Equalities and Inclusion

Project Summary

Please provide a brief high level summary of the research giving necessary background (max 250 words)

The Office for National Statistics (ONS) has linked components of the All Education Dataset for England (AEDE) to 2011 Census records for a proof of concept (PoC) internal study.

The All Education Dataset for England (AEDE) is a longitudinally linked cohort dataset designed by Department for Education in collaboration with ONS Admin Data Census.

This dataset is child-centric but with links to others within the address/household (e.g. parents, siblings).

No other dataset provides such comprehensive coverage of a single cohort of children. The data covers around 2 million children in 2011 and their adults/family/household members which is sufficient for providing accurate and reliable statistics for several different disaggregation's.

The analysis of the PoC dataset will be completed by ONS analysts from the Centre for Equalities and Inclusion. The Centre for Equalities and Inclusion aims to improve the evidence base for understanding equity and fairness in the UK today, enabling new insights into key policy questions. They are a multi-disciplinary convening centre based at the Office for National Statistics, bringing together people interested in equalities data and analysis from across central and local government, academia, business and the third sector.

A child's socioeconomic background is an important determinant in their chances of future life success. The performance gap between advantaged and disadvantaged children develops from an early age and widens throughout students' lives.

The proposed research aims to better understand the impact of family background, school type, and geography on educational outcomes and social mobility; with a focus on the barriers faced by vulnerable children.

The initial analysis will aim to describe the family background of the children in the dataset, as a way of understanding the scope and potential of the dataset for future analysis. It will also focus on the following research questions:

1) Do parent's/familial qualifications, employment status, occupation and industry of work effect educational outcomes;
2) How does educational outcome differ between children with siblings and those without;
3) How is the population of children (aged 15 to 18 years) distributed across school types in terms of key demographics;
4) How do educational outcomes differ when comparing disadvantaged children to non-disadvantaged children?

The results will inform the work of the Children's Commissioner who is prioritising understanding the impact of vulnerability and identifying hidden issues which are affecting the most vulnerable children. Identification of these issues informs the Commissioner's requests for the annual budget where she lobbies for more equitable distribution of resources and specific funding to support the most vulnerable children. Better data on the scale and burden of vulnerabilities will ensure that funding is equitably distributed amongst those who require support.

The research will also inform the work of the Government Equalities Office (GEO) who are prioritising ensuring equitable career progression for women and vulnerable groups. Disparities which arise during a child's education go on to have a lasting impact on their socio-economic status throughout later life on average. Programs which support women fulfilling their potential in the workplace and increase their participation in the labour market must consider the population of future employees, and how their educational experience and outcomes influences their later life decisions.

This research will further the work being done across government to prove the usage of linked data for understanding nuanced interactions between factors which lead to disadvantage throughout the life course. For example, Department for Educations Longitudinal Educational Outcomes Dataset and ONS's Race Disparity Audit Linked Dataset are just two examples of existing linked data projects which are proving the value of linked data for population research. This project however, is the first to take a child centric view and include information necessary to understand complex vulnerabilities and the relationships between the child and their family.

The research is designed to respond to evidence gaps which have been identified by stakeholders working with the Centre for Equalities and Inclusion. The outcomes of the research will determine whether this methodology and type of data linkage can address evidence gaps (proving the concept). Results concerning the utility of the data will be communicated to Department for Education who supplied the education records. If the evidence is sufficient the intention is that this will form the business case for future supplies of data for use in ADRP funded projects.

Projects such as this are essential for achieving the goal of the Admin Data Research Partnership (ADRP) to maximise the potential of administrative data as a resource for high-quality research in the UK. The success of the partnership will pave the way for the UK to meet its ambition to become a global leader in big data infrastructure and create something which is world leading. Every research project which makes full use of these linked data resources will contribute to this aim by addressing major societal challenges and providing evidence for improving public service provision. The ADRP has taken over the work of the Admin Data Research Network (ADRN) and uses some of the practices and procedures developed for the protection of data, which are referenced in this document.

The research proposal has been put forward to the Data for Children analysis working group. The group consists of stakeholders from the Children's Commissioner, the Government Equalities Office, the Department for Education, the Race Disparity Unit, the Social Mobility Commission, the Sustainable Development Goals (ONS) and the Education Efficiency Measurement Unit (ONS).

Section A
Project Details

A1 Legal gateways

Please provide the assessment of the legal gateways of the project as provided by Legal Services

This project is being undertaken within ONS, using data ONS already holds for its statistical function. Therefore no legal gateway is required.

A2 Ethical approval

Has the project being reviewed or is it expected to be reviewed by another ethics committee? ○ **Yes** ○ **No**

If Yes please provide the name of the committee, the outcome and the date approved

Data for Children was presented through an oral report at the January 2018 NS-DEC meeting (Annex A). At this time, there was not a specific project to approve, however members were supportive of the approach for research in the public good but recommended: additional exploration on the public acceptability of this initiative; and additional information on the future role of NSDEC in providing ethical oversight of the resulting projects which used the linked data.

Progress towards the recommendation on public acceptability is given in section B7.

A3 Proposed site of research select all that apply

☐ ONS	☐ ADRC - England
☑ SRS (formerly VML)	☐ ADRC - Scotland
☐ HMRC Data Lab	☐ ADRC - Northern Ireland
☐ Other	☐ ADRC - Wales

please specify

A4 Data subjects to be studied

Does the study include all subsections of the population (i.e. all ages, sex, ethnic groups etc?) ○ **Yes** ○ **No**

If no please detail which subsections with justification(s) below:

Subsections of the population (including vulnerable groups) the project focuses on:

The Data for Children dataset takes children and young people who were aged between 15 and 29 in 2015 and their historical education records back to 2011. The children present in 2011 are 15-18 who successfully linked to Census 2011.

Justification for focusing on these subsections or groups:

This selection was necessary as the aim of the dataset is to provide a child-centric view of the factors influencing educational outcomes. Data on children is typically hard to collect outside of administrative sources as through Census and surveys we rely on an adult to complete the information on behalf of the child. This, coupled with the limitations of large surveys to ask ‘difficult’ questions which would highlight vulnerability has contributed to vulnerability remaining hidden and the scale of the issues difficult or impossible to identify.

The All Education Dataset for England (AEDE) is a longitudinally linked cohort dataset designed by Department for Education in collaboration with ONS Admin Data Census. The age groups included in the data are thus those required by Admin Data Census to assess the utility of the data in replacing the qualifications question on Census. As we are re-purposing this data for a broader use we are restricted to including the cohort present in AEDE. The AEDE cohort design in 2011 includes around 100% of children and young people aged between 15 and 18 enrolled in a DfE funded school at the time of the Spring School Census. As the selection is made on age and no other characteristics, the data should be representative of those children and young people aged 15-29 in 2015.

The cohort is large enough to protect the identify of children at low levels of disaggregation and provides crucial information about their lived experiences during education.

A5 Please provide details of the research protocol or methodology (e.g. data linkage, web scraping etc) (max 500 words)

Necessity of datasets and consideration of alternatives:
To fully inform the research questions, and enable evidence based policy making, evidence needs to have wide geographic coverage, include a large proportion of the population and cover a wide range of relevant variables.

Survey data does not contain enough records for disaggregation to either small geographies or other sub-groups relevant for the research such as ethnic minorities. No other dataset provides such comprehensive coverage of a single cohort of children. The data covers around 2 million children in 2011 and their adults/family/household members which is sufficient for providing accurate and reliable statistics for several different disaggregation's.

Other sources which contain similar data do exist such as the Avon Longitudinal Study and the Millennium Cohort Study however these cover much smaller populations and do not contain annual updates of all variables so the data is less timely.

Key Variables:
The Data for Children data will contain the variables relevant to the assessment of child outcomes such as attainment records and factors which may influence outcomes such as family structure and employment status, geography, carer status (in 2011) and other socio-demographic characteristics.

Methods, tools and techniques:
Within the ONS Statistical Research Environment (SRE) the Census 2011 data has been linked to the English School Census (ESC) 2011 records contained within the All Education Dataset for England (AEDE) to produce the child cohort. The cohort was then linked to the attainment records from AEDE for 2011-2016.

The AEDE dataset contains the following data:

- National Pupil Database (NPD data), containing personal identifiers of individuals who have been in the educational system in the academic years 2001/02 to 2014/15, along with agreed attributes collected from the English Schools Census, Alternative

Provision Census, Pupil Referral Unit Census, Awarding Body information (KS4 and KS5 examination data); and
- Higher Education (HE) Data, containing personal identifiers of individuals who have been in Higher Education in the academic years 2002/03 – 2014/15, along with agreed attributes collected from the HE Student Record; and
- Further Education (FE) Data, containing personal identifiers of individuals who have been in Further Education in the academic years 2002/03 – 2014/15 along with agreed attributes collected from the Individual Learner Record (ILR).

The NPD data (School Census and attainment components) for students between the ages of 15 and 18 enrolled in the educational system in 2011 to 2015 was the only data from the AEDE taken across for this study.

The linkage between Census 2011 and AEDE was completed using name, DOB, address and sex matchkeys. Individuals were then linked longitudinally to the child cohort using the ‘pupil matching ref’, DfE’s unique student identifier. All matching was carried out using hashed data in accordance with ONS TTP safeguarding policies and principles, including clear separation between people who access identifying data, and those who access attributes relating to those individuals.

All linkage has been completed by Data as a Service (DaaS).

Following linkage, the identifying match keys and any unique identifiers will be removed from the data and the subsequent matched data set will then be transferred to the ONS Secure Research Service (SRS) where it will be accessible only by ONS researchers who have undergone the necessary training and this research will be carried out.

Descriptive analysis will focus on the effect of family background, school type, geography, and vulnerable groups on educational attainment and progression. The full list of research questions are provided in Annex B.

Research outputs, sharing arrangements and future access to data.
Following completion of the research a report summarising findings and an assessment of the feasibility of using the data for this purpose will be produced and shared with key stakeholders including a working group with representatives from DfE, the Children’s Commission and the Admin Data Research Partnership.

A6

Data use

Please specify the data used **by the research team** including any timeframes *e.g. LFS data 2014-15*

Data Level

Please specify the name of the data set

Type of data	**Aggregate Data**	**Identifiable Data**	**De-identified personal data**	**Anonymised/ pseudo anonymised**
Administrative data *(please specify, e.g. Patient Register 2011, School Census 2012 etc, in the relevant options adjacent)*			All Education Dataset for England 2007-2015	
Big Data *(please specify e.g. Twitter data, smart meters and mobile phones, in the relevant options adjacent)*				
Survey Data *(please specify e.g.LFS, BRES, etc in the relevant options adjacent)*				
Census Data *(please specify year, e.g. Census 2011 in the relevant options adjacent)*			Census 2011	
Other *(please specify e.g. Ordinance Survey Address register in the relevant options adjacent)*				

Section B
Assessment against NSDEC ethical principles

B1 Principle 1: The use of data has clear benefits for users and serves the public good.

Please outline the proposed benefits of the project (max 500 words)

A child's socioeconomic background is considered an important determining factor in their chances of future life success. The performance gap between advantaged and disadvantaged children develops from as early as 10 years old and widens throughout students' lives (OECD, 2018). Research is therefore needed to better understand the barriers and gateways to social mobility.

Those who attend grammar schools do, on average, better than similar children in the comprehensive system; furthermore, children from deprived backgrounds are less likely to attend grammar schools (IFS, 2016). Research has also shown that children from more affluent families attend better performing schools (Teach First, 2017).

These educational inequalities also contribute to geographic inequalities, with house prices near the best schools higher than in the surrounding areas for both primary and secondary schools (DfE, 2017). A Social Mobility index, has been developed, recognising that a young person's chance of getting on in life is affected by where they live (Social Mobility and Child Poverty Commission, 2016).

The Department for Education and Skills have used linked data to track students through school, college, university and into the labour market (DfE & IFS, 2018). Through this analysis they have attempted to account for differences in background and in prior educational outcome. However, information on background has been limited to using eligibility for free school meals as an indicator for socio-economic status, and tax credit information to identify children in lone parent families. Socio-economic class has been found to be a bigger predictor of educational outcome than gender and ethnicity (DfE, 2007), it would therefore be interesting to use data from Census to explore this further.

Policy has also focussed on disadvantaged children, widening participation and funding for children in deprived areas, and improving educational outcome for looked after children. This proposed analysis is aligned with the Children's Commissioner's work on vulnerable children (2017), these include:
- Children from single-parent households
- Young carers
- Children from minority ethnic backgrounds

This background literature highlights the policy relevance of this topic.

As stated previously the goal of the Admin Data Research Partnership is to maximise the potential of administrative data as a resource for high-quality research in the UK. This research project is one of the first to make use of the new model for creating linked data resources (as previously discussed at NSDEC in January 2018). For this data resource to provide sorely needed evidence for public policy supporting children and vulnerable young people, foundational research such as this must happen to provide the initial evidence of the value of the data. Policy decisions should not be made based on evidence from unproven sources as there is a risk that the data could be misinterpreted or used incorrectly.

This study will provide the foundational analysis required to fully understand the feasibility of using this data for the intended purposes and will support future analysis which will contribute directly to the priorities set by the Children's Commissioner, the Government Equalities Office and other government departments. I.e., this analysis is essential to ensure that the public benefit of the data can be realised.

B2 Please outline any intended future use for products (such as linked data sets or tools) produced as a result of the research and how they will be accessed. (max 250 words)

The Data for Children datasets are currently only available for feasibility research until the utility of the dataset can be proven to the supplier (Department for Education). As stated in section B1, this foundational work is essential to ensure that the data is robust and appropriate to use for the desired research.

The data which will be used for this project is the first wave of the Data for Children dataset. Unless a subsequent research project is identified by researchers and approved by the Research Accreditation Panel (RAP) to utilise this deidentified data it will not be retained however the methodology and lessons learned from the creation of the dataset will be shared with stakeholders to inform any request they might make for a future version of the dataset to be built. There is a statistical need, for example, to provide analysis against more categories of vulnerability, such as children in receipt of free school meals and children with special educational needs.

B3 Principle 2: The data subject's identity (whether person or organisation) is protected, information is kept confidential and secure, and the issue of consent is considered appropriately.

Please outline how data security, confidentiality and informed consent is safeguarded in this project (max 500 words)

ONS TTP safeguarding document is published here:
https://www.ons.gov.uk/aboutus/whatwedo/programmesandprojects/theadministrativedataresearchnetworkcollaboration

ONS will provide secure data linkage and a secure environment.

Full account has been taken of the risks associated with:
- securing access to the data to be linked;
- the management of both the identifying data and attributes contained within the
- administrative data sources;
- the linkage process;
- the identification of population subgroups within the linked data;
- transfer of the de-identified linked file to the SRS.

Rather than operating a traditional 'rules based' approach to output checking, the SRS operates a 'principles based approach'. This forms part of the “5 safes approach” to data access adopted by the SRS.[1]

B4

Principle 3: The risks and limits of new technologies are considered and there is sufficient human oversight so that methods employed are consistent with recognised standards of integrity and quality.

Please describe how the any risks from new technologies are been mitigated as well as any quality assurance activities in the project (max 500 words)

The data will be linked by members of the Data Integration Team at the ONS who have many years' experience in linking and handling sensitive data, including from the Census. The data linkage algorithm proposed is based on deriving hash-keys from the identifying variables and then linking these, rather than using the identifying variables directly. This algorithm has been developed by ONS specifically to ensure that individual identities are protected. This data linkage process has been used successfully for several ONS data linkage projects, including linkage between Census 2011 data and other datasets.

Additionally, there will be significant human oversight so that the methods employed are consistent with recognised standards of integrity and quality. The SRS team will ensure that any output leaving the SRS is aggregate and compliant with disclosure control rules.

B5

Principle 4: Data used and methods employed are consistent with legal requirements such as the Data Protection Act, the Human Rights Act, the Statistics and Registration Service Act and the common law duty of confidence

Please describe the legal frameworks pertinent to this project (max 500 words)

Data will only be used for purposes set out for which approvals have been obtained by NSDEC, ADRP, DfE and MRP. Data will only be accessed by ONS researchers on behalf of the Children's Commissioner, Government Equalities Office, Race Disparity Unit and the Social Mobility Commission or ONS under the Digital Economy Act. Research conducted under Digital Economy Act researchers will be considered on a case-by-case basis following the completion of an ONS ethical self-assessment toolkit and application to NSDEC if required.

No identifying variables will be made available in the SRS. It will not be possible to re-identify individuals in the matched dataset, given it will not be possible to extract record level data from the SRS (and all outputs undergo disclosure control) or import data with the relevant

---

[1] See https://blog.ons.gov.uk/2017/01/27/the-five-safes-data-privacy-at-ons/

identifiers (nor are researchers permitted to bring this information in to the research in any form).

All staff working on the data will be trained in using large, complex administrative databases and handling sensitive data. This training will be documented. All researchers working on the data in the SRS will also need to have the same training as external approved researchers, which ensures they are equipped to work with individual level, sensitive population datasets.

Data will be kept in the SRS and only outputs (tables and figures) which have been disclosure controlled by an expert team in ONS will be allowed to be taken out of the SRS for publication purposes.

The data processing is in line with GDPR rules as detailed in section A1.

B6 Collaboration and Sponsors

Please describe the project sponsors and the **legal gateways** to acquire, process use and share their data

List of Collaborators/Sponsors	**Details and relevant documentation relating to collaboration (you may attach copies of relevant documentation)**
Department for Education	Data supplier, will be kept informed of research planned and progress.
ONS	The Admin Data Research curation team is facilitating the supply of data.
Members of analysis working group: Children's Commissioner, GEO, DFE, RDU & the Social Mobility Commission, Sustainable Development Goals, Efficiency Measurement Unit (Education branch)	Initial meeting held on 25th July 2018 Follow-up meeting to agree research proposal held on 3rd December 2018

B7 Principle 5: The views of the public are considered in light of the data used and the perceived benefits of the research

Please list any public engagement activities (max 250 words)

The design of this research was discussed at a meeting of theme partners who either represent the needs of children and young people and/or conduct research to provide evidence for policy;

- the Department for Education who are responsible for children's services and education, including early years, schools, higher and further education policy, apprenticeships and wider skills in England,
- Department for Work and Pensions,
- The National Foundation for Educational Research whose remit is to generate evidence and insights that can be used to improve outcomes for future generations everywhere, and to support positive change across education systems[2],
- Institute for Fiscal Studies who aim to promote effective economic and social policies by better understanding how policies affect individuals, families, businesses and the

---

[2] https://www.nfer.ac.uk/about-nfer/what-we-do/

government's finances[3]. The IFS has published a number of reports concerning the outcomes of children and young people[45],

- The Children's Commissioner who speaks up for children and young people so that policymakers and the people who have an impact on their lives take their views and interests into account when making decisions about them. The office of the Children's Commissioner is independent of Government and Parliament, and has unique powers to help bring about long-term change and improvements for all children, particularly the most vulnerable[6]. The Children's Commissioner has compiled a vulnerability framework which this research aims to contribute to[7],
- Next Steps is helping researchers, government, teachers and others understand what life is like for generation. Next Steps was set up to study young people's (born between 1980 to 2000) experiences of secondary school and their transitions from compulsory schooling to further education, training or the labour market[8],
- Early Interventions Foundation who champion and support the use of effective early intervention to improve the lives of children and young people at risk of experiencing poor outcomes[9],
- Alliance for Useful Evidence who have been commissioned by DfE to lead a consortium with support from the Social Care Institute for Excellence, Futuregov and the consultancy Traversum, to develop and incubate a new What Works Centre for Children's Social Care. They have partnered with a number of organisations to offer tailored support on using evidence to better inform their policies and programmes[10],
- Avon Longitudinal Study of Parents and Children (ALSPAC). Based at the University of Bristol, ALSPAC, also known as Children of the 90s, is a world-leading birth cohort study. Between April 1991 and December 1992 we recruited more than 14,000 pregnant women into the study and these women (some of whom had two pregnancies or multiple births during the recruitment period), the children arising from the pregnancy, and their partners have been followed up intensively over two decades. ALSPAC has conducted a series of assessments of the study and continually engages with the data subjects concerning how the data is used and what the benefits of the study are[11].

Previously, government research has tended to focus on the general population whereas the aim of the Data for Children project is to enable identification of small sub-groups of vulnerable young people. The ability to protect the identify of individuals in the dataset and appropriate training for researchers sufficiently protects the privacy rights of the children/young people in the dataset. An advantage of ADRP is that the data needs only be provided by a supplier once to ONS (rather than multiple times to multiple research agencies) which addresses significant concerns: the risk of data loss during transfer and the risk of misinterpreting analysis due to incorrect assumptions made about the data. By managing linkage and research in a more co-ordinated way and sharing knowledge these risks are addressed. The stakeholders agreed that there are significant public benefits to be seen from this project and that the risks to privacy are being appropriately managed by ONS and the agreements between ONS and suppliers.

The research proposal has been agreed by the working group convened by the Centre for Equalities and Inclusion with representatives from the following organisations:

---

[3] https://www.ifs.org.uk/

[4] https://www.ifs.org.uk/publications/7256

[5] https://www.ifs.org.uk/publications/10029

[6] https://www.childrenscommissioner.gov.uk/about-us/the-childrens-commissioner-for-england/

[7] https://www.childrenscommissioner.gov.uk/our-work/vulnerable-children/

[8] https://nextstepsstudy.org.uk/home/about/history/

[9] https://www.eif.org.uk/

[10] https://www.alliance4usefulevidence.org/advice-and-advocacy/

[11] http://www.bristol.ac.uk/alspac/about/

- the Children's Commissioner,
- the Government Equalities Office (GEO) who leads work on policy relating to women, sexual orientation and transgender equality. They are responsible for a range of equalities legislation,
- the Department for Education (DFE),
- the Race Disparity Unit (RDU) convened by PM Theresa May in 2016 to deliver publicly accessible data 'shining a light' on ethnic disparities in the UK. The RDU publishes data on ethnic disparities across all outcomes, including education[12],
- the Social Mobility Commission (SMC) monitors progress towards improving social mobility in the UK, and promotes social mobility in England. SMC is an advisory non-departmental public body, sponsored by the Department for Education[13]. The SMC focuses on education as it is viewed as fundamental to social mobility and support the government's education reform programs[14],
- Sustainable Development Goals (SDGs) specifically focus on the rights and interests of children across several Goals and targets, and
- the Efficiency Measurement Unit (Education branch) is an internal ONS team responsible for assessing the productivity of publicly funded education[15].

This engagement with stakeholders from government departments and other relevant organisations, is aimed at ensuring that our analysis is relevant to the public debate and policy needs. Additionally, these groups represent the needs of the children and young people who are featured in the data. The scope of the project means that it is not feasible to interview or consult with all data subjects and individual consent is not required as the data is accessed lawfully under the SRSA. However, each of the stakeholders has responsibility for representing the views and needs of young people across the UK so their input is essential to ensuring that the research is in line with feedback collected by these groups.

B8 Principle 6: The access, use and sharing of data is transparent, and is communicated clearly and accessibly to the public

How will the findings of the research be disseminated? (max 500 words)

The research findings (in particular, assessments of the utility of the data) will be shared with the working group for this topic which includes representatives from DfE, Children's Commission and ADRP. The findings will also be shared with DfE data guardians where they form the basis for business cases for further uses of the data or new data supplies.

The methodology used and findings concerning the appropriateness of the model will be shared at research conferences such as Data for Policy 2019.

High level details of the project are published on the ADRP website.

---

[12] https://www.ethnicity-facts-figures.service.gov.uk/

[13] https://www.gov.uk/government/organisations/social-mobility-commission

[14] https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/710294/DfE-Social_Mobility_Commission_report.pdf

[15] https://www.ons.gov.uk/economy/economicoutputandproductivity/publicservicesproductivity/methodologies/sourcesandmethodspublicserviceproductivityestimateseducation

If the results of the analysis are robust enough to contribute meaningfully to the discussion and body of research on the issues then ONS will publish a research report summarising the findings. Given that this is a feasibility project and the utility of the data is not known before research begins, no firm commitment to publication of results can be made. Any proposed publication will be reviewed by DfE prior to publication.

Section C
Responsible owner and applicant details

C1 Responsible Owner

Full Name: Pam Everett	**Position:** G5 Data as a Service
Address:	**Email:**
	Telephone:
	Organisation: Office for National Statistics

Declaration to be signed by the responsible owner

I have met with and advised the applicant on the ethical aspects of this project design (applicable only if the responsible owner is not the Applicant).

I understand that it is a requirement for all researchers accessing the data to have undergone relevant training and to have either relevant security clearances or approved researcher status in order to access the data.

I am satisfied that the research complies with current professional, departmental and other relevant guidelines.

I will ensure that changes in approved research protocols are reported promptly and are not initiated without approval by the National Statistician's Data Ethics Advisory Committee.

I will provide notification when the study is complete if it or fails to start or is abandoned.

I will ensure that all adverse or unforeseen problems arising from the research are reported in a timely fashion to the National Statistician's Data Ethics Advisory Committee.

I will consider all advice received from the National Statistician’s Data Ethics Advisory Committee and should I be unable to implement any of the recommendations made, I will provide reasoning in writing to the Committee.

Signature: ………………………………………… **Date:** ……………………

C2 Applicant Details (if applicant is not the responsible owner)

Full Name: Bonang Lewis	**Position:** Senior Research Officer, Centre for Equalities and Inclusion
Address:	**Email:**
	Telephone:
	Organisation: Office for National Statistics

Annex A – NSDEC minutes

January 2018 NSDEC minutes section 6:

6. ADRN: Developing persistent dataset for projects within the Data for Children research theme
    6.1 Mr Paul Jackson, from the Administrative Data Research Network (ADRN), and Mr Leon Feinstein, Director of Evidence for the Children's Commissioner for England, presented on the thematic ADRN proposal, Data For Children, to create a persistent dataset by linking data from the National Pupil Database and other sources.
    6.2 The linked dataset, stored in a secure research environment, would potentially enable research into the household, family, social, educational, and economic determinants of different life outcomes for children.
    6.3 Members were supportive of the suggested approach for research in the public good but recommended additional exploration on the public acceptability of this initiative and additional information on the future role of NSDEC in providing ethical oversight of the resulting projects which used the linked data.

Annex B – Analysis plan

1. Research questions for Proof of Concept dataset:

This proposed analysis focuses on the effect of family background, geography, and vulnerable groups on educational attainment. The justification for this focus is explained in the section 3.

The specific attainment measures included in the analysis will depend on the variables included in the proof of concept dataset. We expect that this first iteration of the dataset will only include information on the type of qualification gained. Detail on the subjects and grade outcomes will hopefully be included in future iterations of the dataset. This will allow us to measure educational progression. When this data becomes available we will seek advice from our working group on how best to measure educational progression from key stage 4 to 5.

We will seek guidance from the Statistical Design and Research (SDR), who will be setting up an education working group to define an agreed concept of educational attainment and progression.

1.1 Analysis of childrens' characteristics and their household context:

This section provides descriptive statistics on children's personal characteristics and their household context. This will be useful in determining the quality of the data and the potential uses of the data.

Tables – counts and rates:

- Number of children by age
- Number of children by ethnicity
- Number of children by religion
- Number of children by migrant status
- Number of children by disability status
- Number of children by local authority
- Number of children by carer status

- Number of children by household type
- Number of children by parents' highest qualification
- Number of children by parents' employment status
- Number of children by parents' occupation educational attainment
- Number of children by parents' National Statistics Socio-economic Classification (NS-SEC)

1.2 Questions related to family background:

This section aims to provide new insights on the effects of family background on educational attainment, using information on parent's and family characteristics linked from Census 2011.

a) Do parent's/familial qualifications, employment status, occupation, industry of work and socio-economic classification affect educational attainment?

Tables – counts and rates:

- Number of children by household type and educational attainment
- Number of children by parents' highest qualification and educational attainment
- Number of children by parents' employment status and educational attainment
- Number of children by parents' occupation and educational attainment
- Number of children by parents' National Statistics Socio-economic Classification (NS-SEC) and educational attainment

1.3 Question related to geography:

This section aims to identify which geographic areas are desirable for children and their families, i.e. which areas attract children and families to complete their education.

This analysis will use National Statistics Census geographies of pupils' home address from the English Spring School Census 2011-2016, which is collected in January. This will highlight some home address movements since the 2011 Census.

Mover status captures non-movers, movers and school leavers. Movers are defined as children, aged 15 to 18 years in 2011, who had a different home address recorded when they completed their education (key stage 5), over the period 2011 to 2015, compared to their home address at the time of Census 2011. Those whose home address doesn't change over the period are defined as non-movers. Those who leave school before 18 (and have no geography linked) are classified as school leavers.

This section will capture only children who have moved both spatial unit and school, the key demographic characteristics of children who move.

a) How does geographic mobility differ by various personal and household characteristics?

Tables – counts and rates:

- Number of children by mover status
- Number of children by mover status and age
- Number of children by mover status and ethnicity
- Number of children by mover status and religion
- Number of children by mover status and household type
- Number of children by mover status and parents' highest qualification
- Number of children by mover status and parents' employment status
- Number of children by mover status and parents' occupation
- Number of children by mover status and parents' National Statistics Socio-economic Classification (NS-SEC)
- Number of children by mover status and urban/rural household address

2. Potential future research:

Findings from the analysis of the Proof of Concept dataset will be used to support the development of a second wave of data with permissions to produce publishable research.

Future iterations of the dataset will hopefully also include information on:

- The subjects and grades obtained
- Vulnerability groups:
  - ✓ Special educational needs and/or disability (SEND)
  - ✓ Pupil Referral units
  - ✓ Excluded pupils, and those at risk of exclusion
  - ✓ Free School Meals
  - ✓ Looked after children, and children in need
- School type
- Ofsted rating
- The ability to identify school leavers

This additional data will allow us to answer the following research questions.

2.1 Questions related to subject choices and grade attainment:

Educational outcomes relates to educational grade attainment and progression from key stage 4 to key stage 5.

a) Does the sex of the parent and child impact educational outcome?

The analysis below will look at how the occupation and sector of the parents' work correlates with how their children do at school in the subjects that correspond to that sector, e.g. if the parents are scientists do both the sons and daughters do as well at science equally and does it matter which of parents is the scientist.

This analysis will also include information on household type, i.e. to identify whether single parent families are impacted differently from other family types.

Tables – counts and rates:

- Number of children by parents' sex and occupation and child's sex and educational outcome, 2007/08 to 2015/16
- Number of children by parents' sex and National Statistics Socio-economic Classification (NS-SEC) and child's sex and educational outcome, 2007/08 to 2015/16
- Number of children by parents' employment status and educational outcome, 2007/08 to 2015/16

We will also analyse educational attainment between siblings. This will be defined as the difference between siblings' results at the same key stage as well as the comparison of their progression between key stages.

b) How does educational outcome differ between children with siblings and those without?

c) How does educational outcome of children compare to their siblings?

d) How does the child's sex compare to their older siblings affect their educational outcome?

These three questions can be explored and broken down by different characteristics (for example; sibling with disability, different races and religions, vulnerable groups etc).

The following analysis relates to families' geographic mobility and the impact this has on educational outcomes. It is important to note that families can move for many different reasons, which we will not be able to capture from the data, i.e. whether they are moving for work, to be closer to family or for access to better schools. Furthermore, the geography data only covers the period 2011 to 2015, so we won't be able to capture those moving to better schools before their children enter the education system.

e) How does moving home and school affect educational outcomes?

Tables – counts and rates:

- Number of children by mover status, household type and educational attainment
- Number of children by mover status, parents' highest qualification and educational attainment
- Number of children by mover status, parents' employment status and educational attainment
- Number of children by mover status, parents' occupation and educational attainment
- Number of children by mover status, parents' National Statistics Socio-economic Classification (NSSEC) and educational attainment
- Number of children by mover status, urban/rural household address and educational attainment
- Number of children by mover status, parents' sex and occupation and child's sex and educational attainment
- Number of children by mover status, parents' sex and National Statistics Socio-economic Classification (NS-SEC) and child's sex and educational attainment

f) Scatter plots of local authorities (LA) will be used to explore the relationships between moving, the characteristics of the local area and educational outcomes. The following variables will be compared:

- median attainment key stage 4 by LA
- median attainment key stage 5 by LA
- median educational progress by LA
- proportions of people that stay by LA

- proportions of school leavers (those who leave before 18) by LA
- house prices by LA
- index for multiple deprivation by LA

g) Which local authorities add the most value in terms of educational outcome?

A multivariate model will be used to determine which local authorities have the most educational progression, holding constant the effects of other factors.

2.2 Questions related to school type:

This section aims to identify any key demographic characteristics of children (and their families) attending different school types. Given the educational reforms in recent years, we will ensure that the school type captured relates to same year as educational attainment. Similarly, Ofsted ratings will relate to the year that the qualification was obtained.

a) How is the population of children (aged 15 to 18 years) distributed across school types in terms of key demographics?

Tables – counts and rates:

- Number of children in different school types by age
- Number of children in different school types by ethnicity
- Number of children in different school types by religion
- Number of children in different school types by household type
- Number of children in different school types by parents' highest qualification
- Number of children in different school types by parents' employment status
- Number of children in different school types by parents' occupation
- Number of children in different school types by parents' National Statistics Socio-economic Classification (NS-SEC)

The population for the proposed analysis is restricted to children who change school over the period 2011 to 2016 but do not change home address. This is so that we can isolate the effect of moving school separate to moving geography on educational attainment. The latter is covered in section 4.3.

School movers is a derived variable defined as children, aged 15 to 18 years in 2011, who had a different school address recorded when they completed their education (key stage 5), over the period 2011 to 2016, compared to their school address at the time of Census 2011. School mover status captures whether children remain in the same school or move school over the period (school movers).

This section aims to identify any key demographic characteristics of children who move school.

b) How is the population of non-movers distributed compared to the rest of the population in terms of school types (i.e. grammar, comprehensive, academy schools)?

Tables – counts and rates:

- Number of children in different school types by mover status

c) Does whether a child moves school or not differ by various personal and household characteristics?

Tables – counts and rates:

- Number of children by school mover status and age
- Number of children by school mover status and ethnicity
- Number of children by school mover status and religion
- Number of children by school mover status and household type
- Number of children by school mover status and parents' highest qualification
- Number of children by school mover status and parents' employment status
- Number of children by school mover status and parents' occupation

- Number of children by school mover status and parents' National Statistics Socio-economic Classification (NS-SEC)

The assumption is that children move to better schools to improve their educational outcomes.

Comparative school rating is a derived variable used to compare the school rating of the initial school to the school the child moves to.

d) Do children move to a better school over the period?

Tables – counts and percentage relative to total child population in each spatial unit

- Number of children who move school and average comparative school rating
- Number of children who move school and average comparative school rating by local authority
- Number of children who move school by school type

e) How does moving school affect educational outcomes?

Tables – counts and rates:

- Number of children by school mover status and educational outcome
- Number of children by school mover status and educational outcome by school type
- Number of children by school mover status and educational outcome by school rating

2.3 Questions related to additional vulnerability groups:

The definition of vulnerable children is consistent with that developed by the Children's Commissioner[16]. Vulnerable children are defined as those with one or more of the following characteristics:

1. Children from minority ethnic backgrounds
2. Young carers
3. Children who have special educational needs and/or disability (SEND)
4. Children in Pupil Referral units
5. Excluded pupils, and those at risk of exclusion
6. Children in receipt of Free School Meals
7. Looked after children
8. Children in Need

Analysis will provide breakdowns by vulnerability type and whether or not children are disadvantaged in multiple ways.

a) How does educational attainment differ when comparing disadvantaged children to non-disadvantaged children?

Tables – counts and rates:

- Number of disadvantaged children by educational attainment and age
- Number of disadvantaged children by educational attainment and ethnicity
- Number of disadvantaged children by educational attainment and religion
- Number of disadvantaged children by educational attainment and household type
- Number of disadvantaged children by educational attainment and parents' highest qualification
- Number of disadvantaged children by educational attainment and parents' employment status
- Number of disadvantaged children by educational attainment and parents' occupation
- Number of disadvantaged children by educational attainment and parents' National Statistics Socio-economic Classification (NS-SEC)

---

[16] Children's Commissioner, Defining vulnerability: Definitions, frameworks and groups: Technical paper 2 in Children's Commissioner project on vulnerable children, July 2017 https://www.childrenscommissioner.gov.uk/wp-content/uploads/2017/07/CCO-TP2-Defining-Vulnerability-Cordis-Bright-2.pdf

2.4 Statistical modelling questions:

The analysis described above uses descriptive statistics to explore the effects of various personal and household characteristics on educational outcome. Following this analysis, we will also look to analyse the data further using multivariate modelling to determine which factors have the most significant impact.

Potential ideas for this analysis include:

a) Which of the characteristics explored have most impact on educational attainment?
b) Which of the characteristics explored have the most significant effect on educational progress?

UK Statistics Authority

National Statistician’s Data Ethics Advisory Committee

ONS approach to communicating and engaging on the public benefits of integrated data

Presentation

Andy Teague

Any other business