Eleventh Meeting of the National Statistician's Data Ethics Advisory Committee

---

Minute, Agenda and Papers

Tuesday 16 January 2018

10:15 – 14:40

Board Room, UK Statistics Authority
London

UK STATISTICS AUTHORITY

NATIONAL STATISTICIAN'S DATA ETHICS ADVISORY COMMITTEE

Agenda

Tuesday 16 January 2018
Board Room, Drummond Gate London
10:15am – 2:40pm

Chair: **Mr Ian Cope**

Apologies: **Keith Dugmore, Rob Bumpstead, Martin Severs**

(10:15am to 12:00pm)

1 10:15am	Minute and matters arising from the previous meeting	Ian Cope
2 10:20am	Chair's report	Oral report Ian Cope
3 10:35am	Ethical review of the Crime Survey for England and Wales	NSDEC(18)01 Fiona Aitchison Joe Traynor
4 10:55am	Linking mortality and prescription data	NSDEC(18)02 Neil Bannister
5 11:10am	Students' suicide statistics	NSDEC(18)03 Neil Bannister
6 11:25am	ADRN – Developing persistent dataset for projects within the Data for Children research theme	Oral report Peter Smith
7 11:45am	DCMS – Update on the Ethics Commission on Data Use	Oral report Nicholas Dodd

Lunch (12:10am to 12:30pm)

(12:30pm to 2:40pm)

8 12:30pm	Self-assessment process for ethical consideration	Oral report Peter Fullerton Petros Saravakos
9 12:45pm	Ethical guidelines for social media research	NSDEC(18)04 Owen Abbott Leone Wardman
10 1:00pm	Ethnicity from names	NSDEC(18)05 Owen Abbot
11 1:15pm	An Exploratory study for estimating the outflow of Welsh Speakers from Wales to England from the 2011 Census and the Patient Register	NSDEC(18)06 Martin Parry
12 1:30pm	ADRN: Identifying household and family risk factors for hospital admissions in children	NSDEC(18)07 Steve Bond
13 1:45pm	ADRN: Social and Economic Predictors of the severe mental disorders	NSDEC(18)08 Steve Bond

14 2:00pm	MRP: Transport Model Development for West Berkshire Council	NSDEC(18)09 Nick O'Donnell
15 02:15pm	Ethical considerations for ONS acquisition of NHS Digital Data	Oral report Jonny Tinsley
16 2:35pm	Any other business	

Next meeting: Tuesday 24 April 2018
Location: One Drummond Gate, London

National Statistician's Data Ethics Advisory Committee

Minute

Tuesday, 16 January 2018
Board Room, Drummond Gate, London

Present
Members
Mr Ian Cope (Chair)
Ms Vanessa Cuthill
Mr Keith Dugmore
Mr Colin Godbold
Ms Annie Hitchman
Dr Brent Mittelstadt
Ms Isabel Nisbet
Ms Marion Oswald

UK Statistics Authority
Dr Simon Whitworth
Mr Petros Saravakos

Office for National Statistics
Ms Fiona Aitchison (for item 3)
Mr Neil Bannister (for items 4 and 5)
Mr Peter Fullerton (for item 8)
Mr Owen Abbott (for items 9 and 10)
Mr Steve Bond (for items 12 and 13)
Mr Nick O'Donnell (for item 14)
Mr Jonathan Tinsley (for item 15)

Other
Mr Paul Jackson and Mr Leon Feinstein (for item 6)
Mr Nicholas Dodd (for item 7)
Mr Glyn Jones (for item 11)

Apologies
Professor Martin Severs
Mr Robert Bumpstead
Dr Emma Uprichard

1. Minutes and matters arising from the previous meeting

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

1.2. Members were informed that the minute of the eleventh meeting had been agreed by correspondence. The minute, agenda and papers from the last meeting are now published on the UK Statistics Authority website.

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

2. Chair's report

2.1. The Chair provided members with an update on projects previously considered by NSDEC. The meeting heard that all projects which received major revisions in the previous meeting had been revised and these projects would be presented at this meeting.

2.2. Members received an update on the ongoing engagement between the NSDEC secretariat and staff from the Department for Culture Media and Sport (DCMS) about the development of the Centre for Data Ethics and Innovation to enable and ensure safe, ethical and ground-breaking innovation in Artificial Intelligence and data-driven technologies. The secretariat had also participated in a series of workshops organised by the Government Digital Service (GDS) to provide feedback on the next iteration of the GDS Data Ethics Framework, expected to be formally published within the next month.

2.3. The meeting also heard that the NSDEC Secretariat had provided advice on setting up an ethics committee to the West Midlands Police (WMP), who are establishing an independent ethics committee which will advise the Commissioner and the Chief Constable on whether proposed data analysis projects meet certain ethical standards.

2.4. The Chair concluded his report by informing members that Osama Rahman will be stepping down from NSDEC after two and a half years. The meeting was also informed that the Chair will be shortly starting a secondment to Stats New Zealand and thus he will be stepping down as NSDEC Chair after two and a half years. A new chair will be appointed by the National Statistician in due course.

3. Ethical review of the Crime Survey of England and Wales [NSDEC(18)01]

3.1. Ms Fiona Aitchinson, from the ONS Crime Team, presented a re-submitted application to review the Crime Survey for England and Wales (CSEW), a well-established ONS survey conducted by a third party (Kantar Public) on behalf of ONS.

3.2. Members agreed that significant improvements had been made in this application. However, they suggested that the following further improvements should be made:

i. ONS should collaborate with the National Society for the Prevention of Cruelty to Children (NSPCC) to strengthen the language in the supporting material and advance letter to make clear the voluntary nature of the survey;
ii. members requested more clarity on how parental consent was obtained for survey participants between 16-17 years old;
iii. the committee asked for further assurance that survey participants were only re-contacted when they had provided their permission to do so; and
iv. members recommended that it be made clear to respondents that information would be passed to other research organisations only if respondents opted to be re-contacted for further research.

3.3. This project was approved subject to minor revisions.

Action: Ms Aitchison to update the application to:

i. ensure that informed and voluntary consent is obtained, especially when parental consent is required, and this is clearly communicated to participants in the advance letter;
ii. clarify in the application how parental consent is acquired for participants aged between 16-17 years old;

iii. **provide assurance that the confidentiality of data subjects is protected and that participants cannot be re-contacted without their permission after completing the survey; and**

iv. **consult with NSPCC to make clear the voluntary nature of the survey in the advance letter.**

4. Linking mortality and prescription data [NSDEC(18)02]

4.1. Mr Neil Bannister, from the ONS Life Events team, and Ms Margaret Dockey, from the NHS Business Services Authority (NHSBSA), presented an updated proposal to conduct a feasibility study to develop an evidence base to investigate the relationship between suicide statistics and prescription drugs.

4.2. The meeting heard that no information about specific drugs would be disclosed in any research outcomes, which would be subject to strict ONS mortality statistical disclosure controls.

4.3. Members were satisfied with the proposed confidentiality safeguards put in place by ONS but requested more assurances regarding the training and accreditation of the research team and the confidentiality safeguards in place at NHSBSA.

4.4. The meeting requested that Mr Bannister also liaise with the NHS Health Research Authority to determine if the project would require additional ethical approval by a health research ethics committee.

4.5. This project was approved subject to minor revisions.

Action: Mr Neil Bannister to work with NHSBSA to:

i. **clarify whether the Health Research Authority (HRA) would need to consider the ethical aspects of the research via a research ethics committee; and**

ii. **clearly articulate in the application the level of accreditation and training of researchers from NHS Business Services Authority (NHSBSA) as well as the confidentiality safeguards in place in NHSBSA.**

5. Linking suicide deaths data to Higher Education Student Registry Data [NSDEC(18)03]

5.1. Mr Neil Bannister presented a proposal to link individual higher education record data to suicide mortality data to improve the quality of information around student suicides. This information would provide better intelligence to relevant public health and higher education bodies to help prevent student suicides and feed into national policy around suicide prevention measures (e.g. National Suicide Prevention Report).

5.2. Members required more information on how students are defined within the scope of this research project. This should also provide assurances that minors are not included given that some higher education institutions follow more flexible admission age criteria.

5.3. The meeting heard that any research outcomes will be presented at a regional level subject to ONS statistical disclosure controls to prevent the re-identification of data subjects. These statistical disclosure controls should be made clear in the application.

5.4. This project was approved subject to minor revisions.

Action: Mr Neil Bannister to make clearer in the application:

i. **how students are defined and ensure that minors are not included in the age groups examined; and**
ii. **how the risk of re-identification is mitigated by applying appropriate disclosure controls.**

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.

Action: Dr Simon Whitworth to discuss with Mr Paul Jackson the potential role of NSDEC in providing ethical oversight for potential projects using this dataset.

7. Update on the Ethics Commission on Data Use

7.1. Mr Nick Dodd from the Data Use and Ethics team in DCMS provided an update to the NSDEC members on the recent developments in establishing the Centre for Data Use and Innovation.

7.2. Members heard of the potential to develop collaborative relationship between NSDEC and the new body and offered advice on establishing the new body given their extensive experience in a variety of research ethics subjects.

8. Self-assessment process for ethical consideration

8.1. Mr Peter Fullerton, from the ONS Data Science Campus, and Mr Petros Saravakos presented an update on the plans to develop a process of ethical self-assessment based on the NSDEC principles and precedents established over the past three years. The self-assessment, would be used across ONS to identify ethical issues in research projects and determine which projects require further ethical consideration by NSDEC. Self-assessments would be reviewed by senior managers and the NSDEC secretariat, to ensure that robust, consist decisions are reached. Researchers would receive training to enable them to complete the self-assessment.

8.2. It was reported that the NSDEC secretariat had tested the process in several ONS research areas and received positive feedback. The process was presented to the Data Governance Committee who were supportive of the suggested approach. The meeting heard that the ONS Chief Security Officer had endorsed the proposed self-assessment framework as a necessary and sensible way forward.

Action: The NSDEC Secretariat to:

i. **present the outcomes of the self-assessment six-month pilot to future NSDEC meetings; and**
ii. **arrange for an audit of the piloted projects by one or two committee members to ensure its consistency.**

9. Ethical guidelines for social media research[NSDEC(18)04]

9.1. Mr Owen Abbot, from the ONS Big Data team, presented early work on the development of a policy to set out the practices and procedures that ONS staff would follow if using social media data for research and statistical purposes. Any policy would be developed in line with the NSDEC ethical principles.

9.2. Members recommended that the policy should provide clear guidance to researchers on:

i. the legal frameworks which allow access to the data;
ii. the retention of social media data;
iii. mitigating re-identification risk via data linkage;
iv. assessing the privacy impacts; and
v. handling data produced by the use of social media networks by minors.

Action: Mr Abbott to work with the NSDEC Secretariat to establish a task and finish group to address the NSDEC comments and present a policy at a future meeting for further consideration.

10. Ethnicity from names [NSDEC(18)05]

10.1. Mr Abbott provided an update on the project proposal which looks to measure the quality of, and further develop a tool to estimate ethnicity from names. The project will be run in collaboration with University College London (UCL).

10.2. Members were satisfied with the progress that had been made since they last discussed the project. Members suggested that the application should provide more information on the predictive accuracy of the estimates produced by the tool. Mr Abbott reassured the committee that the tool would provide information about the accuracy of the estimates produced and was more accurate than commercially available software.

10.3. Members agreed that the project could proceed subject to:

i. further engagement with other population groups to provide them with the opportunity to feedback on the tool; and
ii. completion of the additional security testing of the tool and the hosting website.

10.4. This project was approved subject to minor revisions.

Action: Mr Abbott to:

i. liaise with other population groups to provide them with the opportunity to feedback on the tool; and
ii. ensure that additional security testing of the tool and the hosting website is completed before the project can proceed.

11. ADRN - An Exploratory study for estimating the outflow of Welsh Speakers from Wales to England from the 2011 Census and the Patient Register [NSDEC(18)06]

11.1. Mr Glyn Jones presented a re-submitted proposal to link 2011 Census to Patient Register data in England to estimate the outflow of Welsh speakers from Wales to England. The meeting heard that the project was of considerable importance to public authorities and the general public in Wales.

11.2. Mr Jones assured the committee that statistical disclosure controls were in place and results would be presented at a high enough geographical level to fully mitigate against any risk of re-identification.

11.3. Members also recommended that the project outcomes should not be used to inform national and local policy in a way that would disadvantage other population

groups. Mr Jones informed the committee that a wide array of sources would be used to avoid any direct or indirect dis-benefits to population groups outside the scope of this research.

11.4. The committee approved this project.

12. ADRN – Identifying household and family risk factors for hospital admissions in children [NSDEC(18)07]

12.1. Mr Steve Bond, from ONS Data as a Service, presented a proposal by University College London, via the Administrative Data Research Network, to examine environmental, household and risk factors for hospital admissions in children. It was reported that the research would be conducted within the ONS Secure Research Service and would use a linked dataset comprising of Hospital Episode Statistics, ONS birth and death registration data and NHS birth notification data.

12.2. The committee approved this project.

13. Social and Economic Predictors of the severe mental disorders, the SEP-MD study [NSDEC(18)08]

13.1. Mr Steve Bond presented a proposal by Kings College and South London and Maudsley Trust, via the Administrative Data Research Network, to link census data to electronic Patient Records to assess mental health inequalities within the Trust's area and to inform the development of social and public health interventions.

13.2. Members suggested that any consent arrangements currently in place should also comply with the updated Caldicott Principles. Furthermore, members recommended that the application should provide further clarity on whether minors would be included in the study and improve the language in the application so as not to confuse correlation and causality.

13.3. The meeting heard that the dataset would be retained in the Secure Research Environment for 12 months after producing the linked dataset. Members noted that the proposed retention period was not proportional to the three year length of the study and suggested that the application considered a longer retention to ensure the availability of data throughout the project life cycle and the reproducibility of the outcomes of the study.

13.4. This was approved subject to minor revisions

Action: Mr Bond to:

- **i. provide more information on whether the proposed arrangements regarding consent meet the Caldicott principles;**
- **ii. tighten the language in the application and the title;**
- **iii. clarify whether minors are included in the research; and**
- **iv. provide more clarity on the retention of the data to ensure it is proportional to the duration of the project.**

14. MRP: Transport Model development for West Berkshire Council [NSDEC(18)09]

14.1. Mr Nick O'Donnell presented a proposal referred to NSDEC by the Microdata Release Panel which was made by WSP, a global professional services firm, to use de-identified 2011 Census travel to work data in the SRS together with aggregate GPS data from mobile phone networks and traffic survey data. This project aimed to inform the development of local transport policy and evaluate current plans in line with West Berkshire Council's local transport model plan.

14.2. The committee approved this project.

15. Ethical considerations of ONS acquisition of NHS Digital Data

15.1. Mr Jonny Tinsley invited the committee to advise on the ethical aspects of using activity and clinical data from the Hospital Episodes Statistics dataset for research and statistical purposes. ONS could access this data using the legal gateway provided by the Statistics and Registration Service Act 2007 as amended by the Digital Economy Act 2017.

15.2. The meeting heard that the acquisition of health data to produce new and improved existing statistical outputs presented a wide array of potential significant public benefits. Mr Tinsley reassured the committee that all projects using these datasets will be subject to ethical consideration by NSDEC.

15.3. Members recommended further engagement with the public to capture their views on the acquisition and use of datasets containing clinical data. It was suggested that it was important that ONS was transparent in communicating the potential benefits of research projects using this data.

16. Any other business

16.1. The Committee thanked Mr Ian Cope for his work as chair of NSDEC.

National Statistician's Data Ethics Advisory Committee

Minute

Tuesday, 10 October 2017
Board Room, Drummond Gate, London

Present
Members
Mr Ian Cope (Chair)
Mr Keith Dugmore
Mr Colin Godbold
Ms Isabel Nisbet
Ms Marion Oswald
Mr Osama Rahman
Dr Emma Uprichard

UK Statistics Authority
Dr Simon Whitworth
Mr Petros Saravakos

Office for National Statistics
Mr David Johnson (for item 3)
Mr John Flatley (for item 4)
Ms Fiona Aitchinson (for item 4)
Mr Jon Wroth-Smith (for item 5)
Ms Becky Tinsley (for item 5)
Mr Neil Bannister (for item 7)
Mr Peter Stokes (for items 8 and 9)
Mr Owen Abbot (for item 10)

Ministry of Justice
Mr Steve Ellerd-Elliot (for item 6)

Apologies:
Professor Martin Severs
Mr Robert Bumpstead
Ms Vanessa Cuthill
Ms Annie Hitchman
Dr Brent Mittelstadt

1. **Minutes and matters arising from the previous meeting**

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

1.2. Members were informed that the minute of the ninth meeting had been agreed by correspondence. The minute, agenda and papers from the last meeting are now published on the UK Statistics Authority website.

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

2. **Chair's report**

2.1. The Chair welcomed three new members to the committee.

i. Dr Emma Uprichard: Reader, Deputy Director, Centre for Interdisciplinary Methodologies, University of Warwick;

ii. Dr Brent Mittelstadt: Postdoctoral Research Fellow in data ethics at the Oxford Internet Institute; and

iii. Mr Stephen Balchin: Interim Head of Profession for Statistics at Department for Work and Pensions.

2.2. The Chair provided members with an update on projects previously considered by NSDEC. The meeting heard that:

i. All the revisions to the proposal to estimate income from administrative data have been implemented. Estimates will be published as banded distributions at lower geographical levels.

ii. Progress has been made in the project to measure the quality of, and further develop, a tool which estimates ethnicity from names. An update was offered to the committee and a revised application is expected at the next meeting.

iii. Researchers involved in the proposal to use ONS data to examine the link between design skills and economic outcomes have clarified how design occupations and design industries will be defined.

iv. All revisions to the proposal to enhance the evidence base on the impact of the heritage sector to the UK economy have been implemented. Researchers from the private research consultancy Ortus Economic Research have clarified the research outputs and how conservation officers are defined.

v. Researchers from the private economic research consultancy Frontier Economics, who are conducting research to evaluate the impact of the apprentice pay policy in England have clarified how the Individual Learners Records will be used.

2.3. The meeting heard that the NSDEC secretariat had received positive feedback from members about the training day organised on 1 September 2017.

2.4. Members received an update on the research and statistics strands of the Digital Economy Act (DEA). It was reported that a consultation on the codes of practice was launched on 21 September and would run until 2 November 2017.

2.5. The meeting heard that the secretariat had participated in a series of workshops organised by the Government Digital Service (GDS) to provide feedback on the GDS Data Science Ethical Framework. This offered the opportunity to present the work of NSDEC. The secretariat also took part in a workshop with the Department of Culture Media and Sport (DCMS) to discuss the establishment of an expert Data Use and Ethics Commission to provide ethical advice to regulators and Parliament on the use of data and emerging data technologies.

2.6. NSDEC members were informed that the Nuffield foundation is working in partnership with the British Academy, the Royal Society, the Alan Turing Institute and the Royal Statistical Society in undertaking the necessary groundwork to launch an independent Convention on Data Ethics in 2018. The Convention would connect the work of data scientists and academics from other disciplines to public policy and to practitioners and decision-makers in the public and private sectors. The aim of this independent convening space, is to identify key issues, deliberate on how these might be addressed, and propose areas for research that would progress the debate.

3. **Forward planning for managing future work**

3.1. Dr Simon Whitworth and Mr David Johnson, from the Data Science Campus, presented early plans to develop a process of ethical self assessment based on the NSDEC principles which could be rolled out across the ONS research community. This self assessment will be used to decide what projects need to scrutinised by the committee. It is proposed that the self assessments will be scrutinised by an ONS senior manger and the NSDEC secretariat to quality assure the self assessment and ensure that due process had been correctly followed Researchers will receive training to enable to complete the self assessment.

3.2. Members were informed that this project is necessary to cope with the Committee's likely greater workload due to the expected increase in the number of projects following the implementation of the DEA. Initially, this work is taking place in collaboration with the ONS Data Science Campus who are supportive of the move towards the suggested process of self-assessment.

3.3. Members were supportive of the initiative but suggested that further work is required to accurately define the scoring system, the weightings and outline high risk areas. NSDEC asked for the secretariat to work with DSC to produce a detailed proposal for self assessment which will be considered at the next meeting.

Action: The Secretariat to work with the Data Science Campus to update the proposal based on the committee's feedback and present a detailed proposal for ethical consideration via self-assessment at the next meeting.

4. **Ethical review of the Crime Survey for England and Wales[NSDEC(17)20]**

4.1. Mr John Flatley presented a proposal from the ONS Crime Team to review the Crime Survey for England and Wales (CSEW), a well-established ONS survey conducted by a third party (Kandar Public) on behalf of ONS. This review was brought to the committee in light of a complaint raised by a member of the public about the lack of ethical consideration of some of the voluntary questions in the survey that asked about experience of sexual abuse during childhood.

4.2. Members agreed that asking these questions presented a clear public benefit in providing a realistic and accurate view of sexual abuse during childhood in England and Wales. However members noted that the survey does not capture crimes for age groups (16-17 years old) and does not capture cyber crime.

4.3. Members requested that Mr Flatley reviewed the guidance and training offered to interviewers to ensure that:

i. they comply with the ONS safeguarding policy;
ii. prepare interviewers to handle distressing situations;
iii. offer the respondents early warning of the sensitivity of the questions asked; and
iv. do not pressurise potential respondents into answering the question.

4.4. The meeting requested that Mr Flatley provided more detail on the data security arrangements and safeguards to limit the risk of re-identification appropriate to the sensitivity of the survey questions.

4.5. Members suggested further work to be done together with the National Society for the Prevention of Cruelty to Children (NSPCC) to refine the language used in the advance letter, supporting material and questionnaire.

4.6. This project received major revisions

Action: Mr John Flatley to work with NSPCC to:

i. **update the supporting material and advance letter;**
ii. **review the questions and scope of the survey with NSPCC;**
iii. **provide assurances on the security and confidentiality of the data; and**
iv. **resubmit a proposal at the next NSDEC meeting.**

5. **Integrated Data Enabling Analysis and Statistics (IDEAS)**

5.1. Mr John Wroth Smith, from ONS Data as a Service Division, and Ms Becky Tinsley, from the ONS Administrative Data Census Division, presented on a data infrastructure project being developed by ONS to store separate de-identified data sets in such a way that they can be consistently and coherently joined together in a timely fashion.

5.2. Members heard that this project will improve the quality of the data linkage and will facilitate the production of more timely and responsive statistics without compromising on data security. Similar projects have been successfully implemented in Wales (SAIL databank) and New Zealand (Integrated Data Infrastructure).

5.3. Members were supportive of the project and thought that this project would enable a consistent approach to linkage to be applied across government which would facilitate wider collaboration across government for research and statistics. Given the potential sensitivity of the having a unique id number assigned to individual records further engagement is required to inform the public about this work.

6. **Opening access to the Ministry of Justice data**

6.1. Mr Steve Ellerd-Elliot, Chief Statistician and Deputy Director of Justice Statistics Analytical Services at the Ministry of Justice (MoJ), presented on the MoJ's approach to open data and transparency. The presentation focused on the new developments to safely and responsibly open up the data produced and held by MoJ to the public domain. This included the data security and ethical safeguards in place.

6.2. Mr Steve Ellerd-Elliot agreed to a closer collaboration between NSDEC and the MoJ. NSDEC will provide ethical consideration for projects using MoJ data for the production of statistics.

7. **Linking prescription data to mortality data[NSDEC(17)21]**

7.1. Mr Neil Bannister from the ONS Life Events team presented a proposal to conduct a feasibility study to provide an evidence base to investigate the relationship between suicide statistics and prescription drugs. This work will investigate the:

i. extent of suicides related to prescription drugs in England; and
ii. relationship between specific prescription drugs and suicides in England;

7.2. Members were satisfied that the application complied with the Data Protection Act but requested further legal advice about how proposal complied with the Statistics and Registration Services Act and the Human Rights Act. Members also requested the following:

i. more clarity on the aims and expected outcomes of the research;
ii. assurance about how the risk of re-identification of individuals known to researchers (e.g. public figures) will be mitigated; and
iii. more information about the age groups and geographies covered in the application.

7.3. This project received major revisions

Action: Mr Neil Bannister to:

i. **clarify in the application the research aims and outcomes;**
ii. **clearly specify the age groups and geographies covered;**
iii. **provide assurances on limiting the risk of re-identification by the researchers;**
iv. **consult ONS Legal Services regarding compliance with the Human Rights Act and SRSA; and**
v. **resubmit a proposal at the next NSDEC meeting.**

8. **MRP Projects considered via correspondence [NSDEC(17)22]**

8.1. Mr Pete Stokes presented the research proposal "A study of European citizens in the UK", which was an Approved Researcher project and had been approved with minor revisions via correspondence. Mr. Stokes thanked the committee for considering this application by correspondence.

9. **MRP Projects considered via precedent [NSDEC(17)23]**

9.1 Mr Pete Stokes presented the evidence that had been used to pass the eight projects by precedent since the NSDEC meeting in July. Members were satisfied with the decisions made and the evidence upon which the decisions were made. Members suggested that they should be provided with more detail in the future about the strength of the precedent used.

Action: The Secretariat to work with the ONS Researcher Support and Data Access team to decide on a classification of precedent decisions and present this to the next meeting.

10. **Ethnicity from names**

10.1. Mr Owen Abbott from ONS's Big Data Team provided an updated on the project proposal which looks to measure the quality of, and further develop, a tool, to estimate ethnicity from names. The project will be run in collaboration with University College London (UCL).

10.2. Members heard that the proposed tool will now produce aggregate outputs and include a disclosure control mechanism (differential privacy) based on advice by the ONS Statistical Disclosure Control team. Mr Abbott reassured the committee that the tool will provide information about the accuracy of the estimates produced and is currently more accurate than commercially available software.

10.3. Members were satisfied with the progress that has been made since they last discussed the project. Members suggested the following further work:

i. a pilot using census data should be run within ONS to update the thresholds, taking into consideration age and ethnic groups; and
ii. further engagement with other population groups to provide them with the opportunity to feedback on the tool.

Action: Mr Owen Abbott to:

i. **update threshold levels to mitigate the risk of re-identification;**
ii. **liaise with other population groups to provide them with the opportunity to feedback on the tool; and**
iii. **resubmit a proposal when this work has been completed.**

11. **Any other business**

11.1. There was no other business.

1.	2.	3.	4.	5.	6.	7.	8.	9.	10	11.	12	13	14	15	16

Chair’s report

Mr Ian Cope

1.	2.	3.	4.	5.	6.	7.	8.	9.	10	11.	12	13	14	15	16

UK Statistics Authority

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

Ethical review of the Crime Survey for England and Wales

Start Date: 1982 **End Date:** N/A

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)

The Crime Survey for England and Wales (CSEW) was first carried out in 1982 and has run on a continuous basis since 2001. Responsibility for the survey transferred from the Home Office to ONS in 2012. It is a voluntary face-to-face victimisation survey in which people resident in households in England and Wales are predominantly asked about their experiences of a range of crimes in the 12 months prior to the interview. Respondents to the survey are also asked about their perceptions of crime and anti-social behaviour as well as their attitudes towards police and the criminal justice system.

The interviewer will administer the main section of the survey, however, self-completion modules have been used to collect information on topic areas that respondents could feel uncomfortable talking about to an interviewer since 1992. The use of self-completion on tablet computers allows respondents to feel more at ease when answering questions on sensitive issues, due to increased confidence in the privacy and confidentiality of survey responses.

A number of the topics included in the self-completion section of the survey rotate each year. Currently there are three rotating topics that will each appear once every three years:

- Nature of serious sexual assault since age 16 (first introduced in April 2005)
- Nature of domestic partner abuse (first introduced in April 2006)
- Experience of abuse during childhood (first introduced in April 2015)

This ethical review is focused on the abuse during childhood module, which is due to be included in the CSEW for the April 2018 to March 2019 survey year.

In recent years, high profile reports of child sexual abuse in a number of institutions including the BBC, schools, hospitals, and care homes, as well as cases involving prominent people, have led to increased public awareness. However, the scale of child abuse remains unclear as there is a lack of comprehensive statistics covering all offences – particularly those not reported to the authorities. The CSEW is an important tool for addressing some of this evidence gap, as one of its key benefits is its ability to provide a more comprehensive picture of crime in terms of offences that have not been reported to the police. The first results from the CSEW abuse during childhood module were published in August 2016 in the report 'Abuse during childhood: Findings from the Crime Survey for England and Wales, year ending March 2016[1]'.

The design of the questions on abuse during childhood was built on work carried out by the NSPCC and incorporates similar wording to that used in other CSEW questions on sexual abuse. The questions around sexual assault use explicit language to mirror the legal definition of rape and assault by penetration as set out in the Criminal Justice and Public Order Act, 1994 and the Sexual Offences Act, 2003. When the Home Office originally developed the questions, testing found it was necessary to be explicit as public understanding of, for example what constitutes a rape, does not align well with the definition in law. NSPCC have also been actively involved in a review of both the existing questions and new ones planned for inclusion in 2018-19 as part of recent survey development work.

The CSEW has an annual sample size of ~35,000 households and all respondents aged between 16 and 74 are asked to complete the self-completion section. A small number of complaints have been received regarding the explicit language used in some of the survey's self-completion modules, which cover particularly sensitive topics such as sexual abuse and abuse during childhood.

Previous discussions around possible softening of the wording of these questions have resulted in key users of the survey stating that they would be against such a change as this would lead to a reduction in the estimated prevalence of sexual offences and introduce a discontinuity to an important time series. Key users include Home Office Violence against Women and Girls policy team, the Independent Inquiry Child Sexual Abuse (IICSA), academics, and third sector organisations providing support services to victims and survivors. In March 2017 we raised this issue with an external Steering Group including representatives from a number of these groups, who strongly advocated that the wording should not be changed.

Feedback has also suggested that the survey materials (e.g. advance letter, leaflet) do not always adequately prepare respondents for the content of these sections. As a result of both this and previous feedback from the National Statistician's Data Ethics Advisory

[1] https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/articles/abuseduringchildhood/findingsfromtheyearendingmarch2016crimesurveyforenglandandwales

Committee, work has been conducted with the survey contractor to review these materials ahead of the 2018/19 survey. This has included amending the letter to clearly advise respondents that the survey is voluntary; added wording on the fact that the survey asks about people's personal experiences of crime; and encourages respondents to visit the survey website for more detail on the content of the survey, confidentiality and help or support information.

The CSEW has evolved over the past few years and the ethical implications of the inclusion of sensitive topics in the survey have not been reviewed since 2012 when ONS took over responsibility for the survey from the Home Office.

As part of the regular rotation of CSEW self-completion modules, this project aims to use the CSEW module on abuse during childhood for the 2018-19 iteration of the survey to:

- Provide data to address some of the evidence gap on the prevalence of child abuse, making use of the benefit of the CSEW's coverage of crimes not reported to the police
- Support the work of the IICSA
- Producing and publishing independent statistics on child abuse, a topic that is of key public interest
- Inform an important area of public policy.

ONS produce and publish the majority of crime statistics based on data from the CSEW, including abuse during childhood. However, CSEW data are also shared with a number of other government departments (Home Office, Ministry of Justice, Department for Transport and Department of Education) who use the information to publish a range of statistics (such as the annual Misuse of Drugs publication) and to inform policy development.

The disclosure-controlled data is also deposited with the UK Data Archive on annual basis, which allows approved researchers and academics to access the data for planned research projects. The same datasets can also be accessed via the ONS run Virtual Microdata Laboratory (VML).

Section A
Project Details

A1 Legal gateways

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

The Crime Survey for England and Wales 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

A3 Proposed site of research select all that apply

- ☑ ONS
- ☐ VML
- ☐ HMRC Data Lab
- ☐ Other
- ☐ 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:

The CSEW only includes adults (aged 16 and over) living in residential households in England and Wales. A related children's survey is conducted for children aged 10 to 15 years of age living in households where an adult has already taken part in the main survey. However, the proposed module of questions on abuse during childhood would, like all CSEW self-completion modules, be asked only of adults aged 16-74.

Justification for focusing on these subsections or groups:

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, care homes or 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). Neither of these surveys currently measure abuse during childhood.

Previously the self-completion modules were only asked of those aged 16-59. The age limit was raised from April 2017 following research that assessed uptake from older participants, alongside consideration of ease of use and other factors, to determine whether the upper age limit could be removed. Testing showed that willingness to answer the self-completion declined as age increased, with the proportion of people completing the modules decreasing significantly for those aged 75 and over.

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

Data Collection

The CSEW 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. This is a comprehensive list of addresses compiled by Royal Mail for the purpose of mail delivery and contains no names or personal details. 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 current sample size is 35,000 households, with a minimum of 650 in each police force area and a response rate of between 70-75% is achieved annually. Calibration weighting is applied to adjust for differential non-response.

The survey is carried out by an external contractor, currently Kantar Public, who employ a panel of interviewers who are trained to administer the CSEW. Interviews are carried out on a face-to-face basis; the interviewer will visit each sampled address, establish the eligibility of the household (e.g. excluding second homes) and select one adult from the household as the respondent. Although identifiable information such as the respondent's name and address is collected for the purposes of administering the interview, this information is held separately by Kantar Public and is replaced by a unique ID number on the final dataset delivered to ONS. This is standard practice for social research surveys and any risk is

mitigated by adherence to approved data security arrangements and protocols.

The interviewer will administer the main section of the survey, however, self-completion modules are used to collect information on topic areas that respondents could feel uncomfortable talking about to an interviewer. The use of self-completion on tablet computers allows respondents to feel more at ease when answering questions on sensitive issues, due to increased confidence in the privacy and confidentiality of the survey.

The abuse during childhood self-completion module has only been run once previously, in the survey year 2015/16. It asks adult respondents about their experiences of abuse during their childhood before the age of 16 and covers subjects including psychological abuse, physical abuse and sexual assault. A full list of the questions included is contained in the attached Annex A.

Data Processing and Publication

Kantar Public supply an SPSS dataset containing de-identified CSEW data to ONS on a quarterly basis.

ONS are then responsible for further processing of the dataset to apply weighting processes and create a range of derived variables required for regular crime statistics publications:

- Quarterly crime statistics bulletins (main focus on headline levels of crime)
- Focus On series (more in depth analysis of topics such as violence or property crime)
- Specific articles and publications (one off or less regular pieces focused on specific types of crime e.g. abuse during childhood, domestic abuse or public perceptions of crime)

All of these publications use statistics based on aggregate CSEW data. No individual level data are published. The three key measures used are: the incidence rate (the number of crimes experienced per household or adult), the prevalence rate (the proportion of the population who were victims of an offence once or more), and the estimated number of victims.

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.

The quarterly dataset is shared with the Home Office crime statistics team via secure systems for the purpose of producing statistical publications (similar to the ONS use of this data) and for internal analysis for policy development.

The annual dataset (the quarterly dataset that covers the full survey year of April-March) is shared with the UK Data Archive and the VML. These datasets are subject to further statistical disclosure control due to the fact that individual level information is included, but the specific controls applied differ between the two due to the protections in place on the outputs from these systems.

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 (including abuse during childhood) is redacted from this dataset. Other potentially disclosive variables such as open text fields are also removed. Other demographic variables such as age, marital status and geography are recoded to provide less detailed information.

The file supplied to the VML does include the sensitive variables from the self-completion,

but statistical disclosure control has been applied to redact any potentially disclosive variables such as open text fields.

The statistical disclosures control procedures applied in both cases have been formally agreed with the ONS statistical disclosure control team.

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)*				
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)*			CSEW (all years)	
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, including 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 by offence, region, and characteristics of victims, allow a greater depth of understanding about crime.

The CSEW has included questions on abuse in adulthood for many years, but user demand from both government and the third sector for data on abuse during childhood led to the introduction of a new module in 2015/16.

The 2015/16 abuse during childhood CSEW module provided the first official estimates of adults' experiences of being abused as children. The questions built on work previously done by the National Society for the Prevention of Cruelty to Children (NSPCC). In addition to the benefit of providing more recent and up-to-date data, re-running this module in 2018/19 would give the opportunity to extend these estimates to a wider range of the population (following the increase in the age limit of the survey's self-completion section), include additional questions on pertinent topics (see below) and provide evidence to confirm the robustness of the estimates from the previous version. The NSPCC have been actively involved in a review of both the existing questions and new ones planned for inclusion in 2018-19 as part of recent survey development work.

In recent years, reports of child sexual abuse in a number of institutions including the BBC, schools, hospitals, and care homes have increased public awareness of child sexual abuse. Running in parallel have been a series of high profile cases involving people in prominent positions in public life. However, while there is a growing evidence base on child abuse, including large rises in the number of offences reported to the police since 2013, the scale of child abuse remains unclear. Statistics based on cases referred to the police or social services can only ever provide a partial picture since much child abuse is hidden. The CSEW is able to address some of this evidence gap, as one of its key benefits is its ability to provide a more comprehensive picture of crime in terms of offences that have not been reported to the police.

A recognised public demand for better statistics on child abuse led the National Statistician's Crime Statistics Advisory Committee (NSCSAC) to set up a Task and Finish group to map the available data, identify gaps and make recommendations for what is a challenging area of crime to measure. The proposal to include the abuse during childhood on the CSEW in 2018/19 forms part of a package of work on child abuse statistics that ONS has established in response to this decision by NSCSAC.

The need for child abuse questions to be retained on the CSEW was also highlighted by the Centre of Expertise on Child Sexual Abuse in a briefing paper associated with their report

published in July 2017[2]. The data have been used in the ONS publication 'Abuse during childhood: Findings from the Crime Survey for England and Wales, year ending March 2016[3]' and provide key information for policy development within government (e.g. by the Home Office) as well as within the third sector (NSPCC).

ONS are also working with the Independent Inquiry into Child Sexual Abuse (IICSA to ensure that the intended 2018/19 CSEW module on abuse during childhood will take account of their research requirements, particularly in relation to children in custody. This ongoing independent statutory inquiry was initiated in March 2015 and previously requested data from the CSEW on sexual abuse of children in custody. However, this specific topic was not included in the 2015/16 module. ONS also plan to seek consent to re-contact respondents to this module for the purpose of conducting further research for the IICSA (details of IICSA plans are contained in Annex L).

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. Researchers can apply to the VML or the UK Data Archive for access to de-identified individual level survey data for each survey year (April to March) from the CSEW. In the case of the VML this will also including data from the abuse during childhood module.
This is the only version that would contain a full year's sample for modules such as abuse during childhood, which are not included on the survey every year.

It is also anticipated that the ONS publication 'Abuse during childhood: Findings from the Crime Survey for England and Wales, year ending March 2016', which was published in August 2016, will be updated and expanded using information collected from the proposed 2018/19 version of the abuse during childhood CSEW module.

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)

Interview Safety

All interviewers employed by the survey contractor, Kantar Public, carry an identity card and will show this to respondents when they visit their home. Respondents are advised of this fact in the advance letter they receive before the interviewer calls. The letter also provides respondents with both phone and email contact details for Kantar Public and a link to the survey website.

Interviewers also have a police calling card that they can show to respondents. This highlights police support for the survey (in the form of the National Police Chief's Council) as well as ONS and Home Office involvement.

[2] https://www.csacentre.org.uk/research-publications/scale-and-nature-of-child-sexual-abuse-and-exploitation-report/

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. They are designed to give a high level overview of the background and purpose of the survey and we rely on interviewers providing more detail in response to individual questions. However, work has been conducted with the survey contractor to review these materials ahead of the 2018/19 survey. This has included amending the letter to clearly advise respondents that the survey is voluntary, added wording on the fact that the survey asks about people's personal experiences of crime; and encouraging respondents to visit the survey website for more detail on the content of the survey, confidentiality and help or support information. In addition, the advance leaflet that accompanies the letter provides further details on the types of questions that respondents may be asked. ONS have consulted the NSPCC regarding the wording of all advance materials (letter, leaflet and website), including the sections related to the voluntary nature of the survey.

If a 16 or 17 year old is selected as the main respondent for their household then consent must be obtained from a parent or guardian before the interview can take place. Permission is verbal, but the fact that this permission has been obtained must be formally recorded on the survey instrument by the interviewer. If permission is refused then no interview will take place.

As is standard practice, the more sensitive questions, including those on abuse during childhood are placed at the end of the face-to-face survey once interviewers have managed to establish rapport and trust. Prior to handing the tablet computer over to the respondent for the most sensitive questions on experience of domestic violence, sexual assault or child abuse, interviewers are trained to advise the respondents of the confidentiality of the survey and seek their consent to continue with the interview (current response rates for the self-completion section are ~94%). Consent is then recorded by the interviewer.

Kantar Public's guidance to interviewers on the administration of the self-completion section, safeguarding the well-being of respondents who take part and obtaining parental consent for 16 to 17-year-old respondents is detailed in Annex E.

It is possible that the inclusion of sensitive questions in the self-completion modules (including abuse during childhood) may cause possible distress to respondents. However, there are various measures taken to mitigate this issue as far as possible:

- The self-completion section includes a number of warning screens ahead of the most sensitive questions where respondents are made aware of the potentially upsetting nature of the topic and the explicit language used in some questions. They are also reminded that they can ask the interviewer for help at any point, or simply record that they 'do not wish to answer' the question. (Annex F). Respondents are asked to provide confirmation that they have read this information.

- All interviewers attend a full survey briefing before beginning work on the survey, one section of which focuses on how to interview vulnerable respondents. They are also provided with documents specifically designed with vulnerable respondents in mind, and a short video offering guidance on this subject. While, the emphasis is primarily on interviewing elderly respondents, the principles covered also apply to vulnerable respondents in general.

- Interviewers receive training on what to do if a respondent becomes distressed. (see Annex E). This includes terminating the interview if a respondent is too upset or

doesn't want to continue.

- Interviewers are also trained to offer respondents a help card (Annex G), which contains contact details for a number of relevant support services. This information is also available on the survey[4] website and is in the process of being directly incorporated into the self-completion section.

In contrast, it should also be noted that some respondents who have experienced the types of crime covered in the self-completion module welcome the opportunity to be able to record it.

Kantar Public's policy on data quality, data security and safeguarding vulnerable respondents are set out in Annexes H & I.

Confidentiality

The confidentiality of both respondents and the information they provide as part of the CSEW is treated with the utmost importance. Name and address details are held securely by Kantar Public, and kept separately from the any information that is collected as part of the interview. These details are not passed on to any other organisation – including the Office for National Statistics – without the respondent's permission. Any information collected is used solely for research purposes and no individual can be identified from the results.

At the end of the survey, respondents are asked for their permission for their details to be shared for the purpose of taking part in further research being conducted on behalf of either the Office for National Statistics (ONS) or another government department. If they agree then the only information that would be shared is their contact details such as their name and address. If permission is withheld then their information will only be used for the production of aggregate statistics and their name and address will not be shared.

The abuse during childhood module of the survey also contains a separate request for the respondent's permission for Kantar Public to pass their name, contact details and specific relevant survey responses to the ONS for the purpose of carrying out further research on behalf of the Independent Inquiry into Child Sexual Abuse (IICSA). This request has been developed in collaboration with stakeholders, including the IICSA and ethical experts at the NSPCC. Again, if permission is withheld then their name and address will not be shared and will not be matched to any of their survey responses.

The confidentiality of the survey is emphasised in the survey's advance materials (Annex B & C) and explained fully on the survey website. In particular the survey leaflet clearly states that the respondent's name and address details will not be passed onto any other organisation or matched to any of their survey responses without their specific consent. When responders are aged 16 or 17 years' old then consent must be obtained by their parent or guardian. A leaflet that explains the survey's data security arrangements in more detail to respondents is available on request (Annex D).

In addition, sensitive questions on topics such as drug use or sexual abuse are collected via self-completion using tablet computers. This allows respondents to feel more at ease when answering questions on sensitive issues, due to increased confidence in the privacy and confidentiality of the survey as they do not have to disclose their responses to the interviewer.

[4] http://www.crimesurvey.co.uk/FurtherHelp.html

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 (Annex D). The data is transferred to ONS using a secure data transfer system and held on the secure ONS IT system. Access to it is restricted to named members of ONS staff involved in the analysis and production of crime statistics.

A full description of the data security arrangements in place for the transfer of data within Kantar Public (including tablet computers) are set out in Annex K. These processes meet industry standards and have been approved by ONS IT data security.

All sharing of CSEW data between ONS and other government departments, including the Home Office, 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.

De-identified data and statistical disclosure control

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 annual dataset (the quarterly dataset that covers the full survey year of April-March) is shared with the UK Data Archive and the VML. These datasets are subject to further statistical disclosure control due to the fact that individual level information is included, but the specific controls applied differ between the two due to the protections in place on the outputs from these systems.

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 (including abuse during childhood) is redacted from this dataset. Other potentially disclosive variables such as open text fields are also removed. Other demographic variables such as age, marital status and geography are recoded to provide less detailed information.

The file supplied to the VML does include the sensitive variables from the self-completion, but statistical disclosure control has been applied to redact any potentially disclosive variables such as open text fields.

The statistical disclosures control procedures applied in both cases have been formally agreed with the ONS statistical disclosure control team. All published CSEW statistics that include estimates with an unweighted base of less than 3 respondents are always suppressed for statistical disclosure control purposes.

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 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. Data from the survey is held and analysed in a commonly used statistical software package (SPSS). No particular use of new technology is made as part of these processes.

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 (Annex D) and the Statistics and Registration Service Act, for both data collection and storage purposes.

Kantar Public's policy on data quality, data security and safeguarding vulnerable respondents are set out in Annexes H & I.

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	The sharing of data from the CSEW between ONS and Home Office is governed by a Memorandum of Understanding (Annex J).
Home Office	

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 the CSEW underpins a range of crime statistics holding National Statistics status, under the Code of Practice for Official Statistics, ONS are obliged to consult with users on significant changes that may affect the quality or availability of the statistics.

Recent public consultations include:

June – July 2017: Changes to the Crime Survey for England and Wales

This consultation was carried out in relation to required cost savings from the survey. It put forward the potential impacts of removing questions, reducing sample size and reducing response rates and invited user views to ensure that the best approach for the majority of

users could be undertaken.

July – August 2016: Review of treatment of high frequency repeat victimisation in Crime Survey for England and Wales estimates

This consultation related to technical changes to the methodology used to balance the accuracy of estimates of the numbers of crime against sample variation caused by respondents with high levels of repeat victimisation.

It is known that issues relating to child abuse are of high relevance to the public and widely covered by media organisations. Public engagement related to the abuse during childhood module included gathering the views of groups such as the NSPCC who represent the views of wide range of users and the general public.

The questions included in the abuse during childhood module were developed based on a similar survey conducted by the NSPCC. Further user acceptability testing of questions planned for inclusion in 2018/19 is currently being planned.

The report[5] on the results from the abuse during childhood module was subject to pre-publication peer review by key users such as the Home Office and the NSPCC.

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)

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.

[5] https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/articles/abuseduringchildhood/findingsfromtheyearendingmarch2016crimesurveyforenglandandwales

Section C
Responsible owner and applicant details

C1 Responsible Owner

Full Name: [redacted]	**Position:** [redacted]
Address: [redacted] Road,	**Email:** [redacted]
	Telephone: [redacted]
	Organisation: [redacted]

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:	**Position:**
Address:	**Email:**
	Telephone:
	Organisation:

UK Statistics Authority

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 prescription data to suicide and drug misuse deaths data

Start Date:	**End Date:**
January 2018	December 2018

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

Office for National Statistics
NHS Business Services Authority (NHSBSA)

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

This feasibility project aims to link ONS data on deaths from suicide and drug-related causes with prescriptions data held by NHS Business Services Authority (NHSBSA) for the period April 2015 to December 2016. The research will provide new insight and evidence for drug misuse and suicide prevention bodies to ultimately help reduce the number of deaths from suicide and drug misuse. The project will help us to establish initial estimates around the following questions:

- Are there prescription drugs or classes of drugs (antidepressants, other psychotropic drugs, and analgesics) that individuals were taking prior to accidental or intentional overdose, or to suicide by another method?
- Where specific drugs such as methadone being prescribed to individuals at the correct dosage (according to clinical guidelines), and did these people then die by accidental drug poisoning? This follows on from findings in an ONS project analysing detailed coroner records that showed that some heroin users are being prescribed methadone in doses lower than the clinical guidelines

- What is the proportion of deaths where a prescription drug is mentioned in coroners text as being related to the drug-misuse death or suicide, and has been prescribed 6 months prior to death
- Investigate if it is possible to tell whether the person was in receipt of the prescription at the time of the death (as opposed to receiving it historically)
- Investigate how the above questions vary by drug (or class of drug), age, gender, cause of death (accident/suicide) and geographic locality.
- To investigate if the data can provide insight into whether the deceased had obtained multiple prescriptions to assist in their death

The results will inform policy on prevention of suicides and drug related deaths, improve understanding of how the use of prescription drugs contributes to these deaths, and therefore provide opportunities to reduce future deaths from these causes. The analysis will provide NSHBSA with increased understanding of a patient's clinical pathway and contribute towards patient safety with a view to preventing further similar deaths.

National outputs will be published by ONS in collaboration with NHSBSA as aggregated data tables, and a written report with the relevant statistical disclosure controls applied. Both organisations will have joint authorship and logos of both organisations included in the publication.

The relationship between prescribing patterns and suicides or accidental deaths of drug addicts is of considerable public health interest. A report by the Home Affairs Committee in 2014 identified a growing problem of 'doctor shopping' whereby addicts obtain multiple prescriptions. As many as 1.5 million people are addicted to prescription drugs in the UK. The National Suicide Prevention Strategy Group has called for better statistics on suicide, while the House of Commons Health Committee in 2016-17 highlighted in their report the need to examine various opportunities for prevention.

The independent Mental Health Taskforce[1] recommendations is to reduce suicide rate in England by 10% by 2020/2021 and for every local area to have in place a multi-agency suicide prevention plan. The research outcomes of the study will contribute directly to provide evidence to meet government initiatives such as the House of Commons Health Committee Suicide prevention interim report 2016-17 which highlighted the need for the Government to re-examine its own policy areas to identify missed opportunities to reduce the risk of suicide[2].

If useful results are found we will make the case to make this a more regular piece of work or extended the work to include a control group from the population. Any extension would be undertaken in agreement with NHSBSA and via a application to the National Statistician's Data Ethics Advisory Committee.

Version .0.2

[1] https://www.england.nhs.uk/wp-content/uploads/2016/02/Mental-Health-Taskforce-FYFV-final.pdf
[2] https://publications.parliament.uk/pa/cm201617/cmselect/cmhealth/300/300.pdf

UK Statistics Authority

Section A
Project Details

A1 Legal gateways

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

Section 22 of the Statistics and Registrations Service Act (2007) applies to this work. It gives ONS with the legal gateway to provide a data processing and analytical service to NHSBA. It also permits the linkage of our mortality data to the prescription data. The SRSA also protects the confidentiality of the information provided (section 39 of the act).

ONS and NHSBSA have a signed Data Sharing Agreement which specifically lays out how the data is to be shared and linked. The document states the specific time period for holding the linked data, specific conditions around the sharing of the data including security and the legality of using the data. The agreement will be signed by the ONS and NHSBSA Head of Business Areas, the Caldicott Guardians and Senior Information Risk Owners (SIRO).

Throughout the project NHSBSA will remain the Data Controller of the prescriptions data at all times and ONS will be the Data Processors acting on behalf of, and only in accordance with the instructions of NHSBSA.

The project has been assessed using the NHS Health Research Authority to ascertain if the project requires approval from the NHS Research Ethics Committee. The self assessment (attached as an annex) states no the project does not need approval from the Committee.

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
- ☐ 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:

The study will contain only persons who have died from suicide or drug misuse in England.

Suicides will be indentified using the National Statistics definition of suicide which includes all deaths from intentional self-harm deaths and where the intent was undetermined. Drug related deaths will be based on deaths identified as drug poisonings (involving both legal and illegal drugs) or drug misuse (involving controlled drugs).

No other cause of death will be included in the data.

Justification for focusing on these subsections or groups:

All deaths from the causes specified above will be included. No specific demographic group (sex, age etc) will be targeted in this work.

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

This project will link individual level death registrations for suicides and drug related deaths with a corresponding subset of patient prescriptions data held by NHSBSA. The extract will contain identifiable information regarding deaths where the deceased's usual place of residence was England and identified as being a death from suicide or drug misuse for all ages including individuals under the age of 18, for the period April 2015 to December 2016.

The linkage will work as follows. ONS will create an extract of data of suicide and drug related deaths that occurred for the specified period. This will contain the relevant mortality data and NHS numbers. This will be stored on a secure server with access provided to two data managers. No researchers will have access to this data. ONS will then provide the list of NHS numbers to NSHBSA with a unique non-disclosive identifier attached (a sequential

number from 1 to 12,500). No other data from the mortality file will be provided to NHSBSA.

The data will be transferred to NHSBSA via the secure MoveIT tool, using the current ONS guidance. NHSBSA will then link this data to their prescriptions data of each individual for the six month prior to death, and provide a linkage flag. They will then transfer back the prescription data with the unique identifier, not the NHS Number, via the MoveIt tool. ONS data managers will then place the extract in a restricted SQL database hosted on a secure ONS server. In the SQL database they will then link this data to the subset of deaths data held by ONS using the l unique identifier. Data managers will ensure that the relevant geographies are contained on the file (down to local authority level) and then will load the a de-identified (NHS Number removed) individual record level linked dataset to another restricted SQL database for a restricted number of researchers (three)to access. This linked dataset will not contain any directly identifiable data, such as NHS number, name or address or postcode to reduce the chance of a researcher identifying the deceased (i.e. known public figures or neighbours or other acquaintances).

This method reduces the risk of disclosure of individuals by researchers by reducing the amount of identifiable variables they have access too. The transfer method mitigates the risk of malicious or accidental disclosure of identifiable data while being passed between organisations. Staff having access to identifiable data will not be involved in the ensuring research.

The research questions to be answered would include:

- Are there prescription drugs or classes of drugs (antidepressants, other psychotropic drugs, and analgesics) that individuals were taking prior to accidental or intentional overdose, or to suicide by another method?
- Where specific drugs such as methadone being prescribed to individuals at the correct dosage (according to clinical guidelines), and did these people then die by accidental drug poisoning? This follows on from findings in an ONS project analysing detailed coroner records that showed that some heroin users are being prescribed methadone in doses lower than the clinical guidelines
- Where specific drugs such as methadone being prescribed to individuals at the correct dosage (according to clinical guidelines), and did these people then die by accidental drug poisoning? This follows on from findings in an ONS project analysing detailed coroner records that showed that some heroin users are being prescribed methadone in doses lower than the clinical guidelines
- What is the proportion of deaths where a prescription drug is mentioned in coroners text as being related to the drug-misuse death or suicide, and has been prescribed 6 months prior to death
- Investigate if it is possible to tell whether the person was in receipt of the prescription at the time of the death (as opposed to receiving it historically)
- Investigate how the above questions vary by drug (or class of drug), age, gender, cause of death (accident/suicide) and geographic locality.
- To investigate if the data can provide insight into whether the deceased had obtained multiple prescriptions to assist in their death

All aggregated tables will be produced adhering to the ONS mortality disclosure control policy ensuring the data is non-disclosive. This could mean that common drugs are grouped together, or separate tables created for sex and age breakdowns.

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)*			Linked prescription and mortality data based on the identifiable data. Provided to 3 research officers.	
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)*				

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 project is designed to provide insight into the possible circumstances around suicide and drug-related deaths. The project is expected to deliver increased insight through data and analytics to identify prescribing patterns for patients who later died from suicide or suffered drug related deaths. This initiative will help provide the NHSBSA with an increased understanding of a patient's clinical pathway and contribute towards patient safety with a view to preventing further similar deaths.

The work has a clear benefit to helping enhance public health around drug misuse and Public Health England Alcohol, Drugs and Tobacco Division is the supportive of this work.

The project will help provide evidence to help answer the questions asked by the Home Affairs Committee report in December 2013 which identified a growing problem of "doctor shopping", whereby prescription drug addicts sign up to multiple GPs to get multiple prescriptions.

This project is in line with other government initiatives such as the House of Commons Health Committee Suicide prevention interim report 2016-17, which highlighted the need for the Government to re-examine its own policy areas to identify missed opportunities to reduce the risk of suicide[3]. This is aligned with the independent Mental Health Taskforce[4] recommendations to reduce suicide rate in England by 10% by 2020/2021 and for every local area to have in place a multi-agency suicide prevention plan.

Each output would be assessed to make sure the data is non-disclosive following the ONS guidelines. ONS is used to dealing with sensitive drug related death and suicide data and are fully aware of the sensitivity of the outputs. To avoid the potential for 'copy cat' deaths we are careful not to publicly release information about new suicide methods. However we do report these in confidence to the National Suicide Prevention Advisory Group and PHE to inform them of new areas of prevention. We would apply the same rules to this analysis.

We would like to produce results at a local level to enable policy makers and health authorities to launch public health initiatives at a local level to reduce drug related deaths and suicides. This is aligned with the National Strategy goal to drive delivery of its aims at a local level to prevent further families, friends, colleagues and communities from experiencing the tragedy of suicide and drug misuse deaths. This will however depend on the results of the analysis and whether the data at the local level complies with our disclosure control rules.

This exploratory study is part of a suite of enhancements being made to the ONS suicide data. We have another project investigating suicides in higher education students. We have recently published new analysis looking at suicide risk in occupations.

[3] https://publications.parliament.uk/pa/cm201617/cmselect/cmhealth/300/300.pdf
[4] https://www.england.nhs.uk/wp-content/uploads/2016/02/Mental-Health-Taskforce-FYFV-final.pdf

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 linked data will be held by ONS for up to 3 years specified in the Data Sharing Agreement with NHSBSA. Any extension to this would be in agreement with NHSBSA.

The data available to researchers in ONS will be limited as specified in section A5. The researchers and data managers that will have access to the data will have had comprehensive training on access and outputting these types of sensitive data (see section B3).

Products of the analysis will be published on the ONS website and the aggregated tablutation will have been assessed for disclosure control against the current Mortality Disclosure Control protocol which is published on the ONS website,

If useful results are found we will make the case to make this a more regular piece of work or extended the work to include a control group from the population. Any extension would be undertaken in agreement with NHSBSA and via a application to the National Statistician's Data Ethics Advisory Committee.

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)

To protect peoples identities the methodology we propose clearly separates duties around the linkage to ensure only the minimum people necessary have contact with the disclosive data both within ONS but also with the separate involvement of the NHSBSA in the process.

The process is as follows. Two ONS data managers will create an extract of suicide and drug related deaths that occurred for the specified research period. This data will be stored on a SQL database that they only have access to. The data mangers will then send a list of NHS numbers and unique non-disclosive identifier to NHSBSA via the secure transfer tool MoveIT. No mortality data will be sent to NHSBSA. NHSBSA will then link this data to their prescriptions data of each individual for the six month prior to death, and provide a linkage flag. They will then transfer back the prescription data with the unique identifier, not the NHS Number, via the MoveIt tool. The ONS data managers will then load this data into the restricted SQL database and link this data to the subset of mortality data using the sequential unique identifier. The data managers will ensure that the relevant geographies are contained on the file, strip out all identifiers (NHS Number) and then will provide the linked data to another restricted SQL database for 3 researchers to access. The linked data will not have any directly identifiable data contained, such as NHS number, name or address or postcode.

The NHSBSA has many safeguards in place for staff dealing with patient level data. Staff on joining the organisation as part of signing their contract must sign a confidentiality statement. NHSBSA hold continuous on the job training for all staff including annual online training

modules for Information Governance related matters. Staff are constantly reminded of the IG policies in place which must be adhered to. There are clear channels to Information Governance colleagues if staff have any concerns or are faced with issues when dealing with sensitive data.

NHSBSA have clear procedures in place which staff follows when disclosing patient identifiable information. These have been signed off by the NHSBSA Caldicott Guardian. These procedures specify who NHSBSA can provide this level of data to, whether a data sharing agreement needs to be in place and how to determine if the legal powers are in place to share the data. These procedures are regularly monitored by the Caldicott Guardian, to ensure compliance to the Caldicott principles, and advice is sought when staff are unsure of a request for patient level data. They are also audited by Health Group Internal Audit who provide an objective and independent assurance, analysis and consulting service to the Department of Health and its arms length bodies, bringing a disciplined approach to evaluating and improving the effectiveness of risk management, control and governance processes.

Access to the data bases on the server is restricted to authorised users in the ONS HALE team by username/password logins. Access lists are reviewed quarterly. Record level data will not leave this secure environment.

Aggregated data will be published at the National Level (England) and then assessed to see if relevant sub-national tables can be produced following the published Mortality Disclosure Control policy. This applies threshold rules to table cells to ensure that individuals are not able to be identified in the aggregated tabulations

Results from the research will only be released by ONS staff once they have confirmed that they contain no risk of identifying an individual or a household. This is assured by the statistical disclosure controls the ONS HALE team currently have in place.

No consent is required by the data subjects or their families as the Data Protection Act does not apply to deceased.

The data will contain deaths of individuals under the age of 18 years old. These deaths will be treated in same sensitive manner as the deaths over the age of 18 with regards to the disclosure control.

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 are involved in the process and safeguards are in place (see section B3) to make sure that any risks introduced by the proposed linkage are mitigated. All processes 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 Statistics and Registrations Service Act (2007) applies to this work, as it provides ONS with the legal gateway to provide the data processing service to NHSBSA (section 22 of the act), and for ONS to link our mortality data to the prescription data (section 42 of the act). The SRSA also protects the confidentiality of the information provided (section 39 of the act).

It is important to note that the Data Protection Act does not apply to this project due to the analysis being undertaken on deceased individuals, The DPA only applies to living individuals, however the disclosure control methods we will use will ensure that no decreased individual or family member related to the individual will be identifiable.

With regards to the Human Rights Act. There is no interference with the right to family life and privacy (HRA Article 8) of the deceased or their surviving family members. No information will be collected about family members and they will not be contacted or identified in any circumstances. The data will be fully anonymised so there is minimal risk of causing any harm or distress by a breach of privacy. The research methods used are sanctioned by law (see above) and proportionate to the public interest in the protection of health.

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)**
NHSBSA	Statement of support provided via email. In Annex 1 and the Data Sharing Agreement (can be provided on request). See section A1 for information regarding the legal gateway

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)

Whilst the views of the public have not been sought with regards to the specific research, there is a clear public benefit for the analysis (see section B1).This project will be breaking new ground. As far as we can tell linkage of prescription data with suicide and drug deaths has not been undertaken before. ONS presented the proposal of this work at the National Suicide Prevention Strategy Advisory Group. The membership has a wide representation including the public; it is made of a mix of government bodies, charities, health providers, academics and champions for families affected by suicide. There were no objections from the group and the Chair has provided a statement of support:

'This will be a very helpful piece of work. In relation to suicide, the most important questions are:

1. Numbers of the main individual and classes of prescription drugs used in fatal self-poisoning, especially individual antidepressants and other psychotropic drugs, and analgesics including opioid-containing analgesics; overall and by age and gender group.

2. The same information related to prescription volume; overall and by age and gender group.

3. In the case of opioids, the numbers of prescribed and non-prescribed drugs used in fatal self-poisoning; overall and by age and gender group.'

These research questions are included in section A5.

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 non-disclosive aggregated tabulations of the results of this project will be initially shared with the NHSBSA, Public Health England, and the National Suicide Prevention Strategy Advisory Group (NSPSAG). Following feedback from experts in this group we will then decided how best to disseminate the findings. No microdata will shared outside of the small select group of researchers and data managers in the HALE division in ONS who will access the linked data

The analysis will be published on the ONS website in collaboration with NHSBSA as experimental statistics.

Section C
Responsible owner and applicant details

C1 Responsible Owner

Full Name: [redacted]	**Position:** [redacted]
Address: [redacted]	**Email:** [redacted]
	Telephone: [redacted]
	Organisation: [redacted]

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:** ----------

UK Statistics Authority

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

Full Name: [redacted]	**Position:** [redacted]
Address: [redacted]	**Email:** [redacted]
	Telephone: [redacted]
	Organisation: [redacted]

Annex 1 –
NHS Health Research Authority Self Assessment for the project

Go straight to content.

MRC Medical Research Council

NHS Health Research Authority

Do I need NHS REC approval?

To print your result with title and IRAS Project ID please enter your details below:

Title of your research:

Linking Prescription Data to Drug Related Deaths and Suicides

IRAS Project ID (if available):

Your answers to the following questions indicate that **you do not need NHS REC approval for sites in England.** However, **you may need other approvals.**

You have answered ***'YES'*** to: Is your study research?

You answered ***'NO'*** to all of these questions:

Question Set 1

- Is your study a clinical trial of an investigational medicinal product?
- Is your study one or more of the following: A non-CE marked medical device, or a device which has been modified or is being used outside of its CE mark intended purpose, and the study is conducted by or with the support of the manufacturer or another commercial company (including university spin-out company) to provide data for CE marking purposes?
- Does your study involve exposure to any ionising radiation?
- Does your study involve the processing of disclosable protected information on the Register of the Human Fertilisation and Embryology Authority by researchers, without consent?
- Is your study a clinical trial involving the participation of practising midwives?

Question Set 2

- Will your study involve research participants identified from, or because of their past or present use of services (adult and children's healthcare within the NHS and adult social care), for which the UK health departments are responsible (including

cisiontools.org.uk/ethics/EngresultN1.html

Result - England

services provided under contract with the private or voluntary sectors), including participants recruited through these services as healthy controls?
- Will your research involve collection of tissue or information from any users of these services (adult and children's healthcare within the NHS and adult social care)? This may include users who have died within the last 100 years.
- Will your research involve the use of previously collected tissue or information from which the research team could identify individual past or present users of these services (adult and children's healthcare within the NHS and adult social care), either directly from that tissue or information, or from its combination with other tissue or information likely to come into their possession?
- Will your research involve research participants identified because of their status as relatives or carers of past or present users of these services (adult and children's healthcare within the NHS and adult social care)?

Question Set 3

- Will your research involve the storage of relevant material from the living or deceased on premises in the UK, but not Scotland, without an appropriate licence from the Human Tissue Authority (HTA)? This includes storage of imported material.
- Will your research involve storage or use of relevant material from the living, collected on or after 1st September 2006, and the research is not within the terms of consent from the donors, and the research does not come under another NHS REC approval?
- Will your research involve the analysis of DNA from bodily material, collected on or after 1st September 2006, and this analysis is not within the terms of consent for research from the donor?

Question Set 4

- Will your research involve at any stage intrusive procedures with adults who lack capacity to consent for themselves, including participants retained in study following the loss of capacity?
- Is your research health-related and involving prisoners?
- Does your research involve xenotransplantation?
- Is your research a social care project funded by the Department of Health?

1.	2.	3.	4.	5.	6.	7.	8.	9.	10	11.	12	13	14	15	16

UK Statistics Authority

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 suicide deaths data to Higher Education Student Registry data

Start Date:
January 2018

End Date:
January 2019

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

ONS

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

Prevention of student suicides is a high priority in Government health policy. The Secretary of State for Health alongside Public Health England have emphasised that research into these suicides is crucial to provide evidence to assist prevention measures. The priority of such research has been highlighted by the clusters of suicides in both Bristol and York Universities in 2017.

The Secretary of State for Health in September 2017 blogged on the Times Higher Education Website:

'….we know when it comes to suicide you can make big improvements in prevention. This can only be done when you have accurate and reliable data – and we know sometimes the figures for student suicides can be understated if, for example, "student" is not recorded properly as someone's occupation. So Universities UK, Public Health England and the ONS are now working with researchers from University of Worcester to encourage more accurate reporting of student suicides.'

This research proposal aims to improve the quality of information around student suicides by linking individual higher education record data to suicide mortality data. The linkage will be undertaken using keys such as name and address of the deceased for the period 2000-01 to 2012-13, for England and Wales. The proposed linkage will provide a much richer and accurate source of data than the current ONS estimates, allowing us to answer questions such as:

- How many student suicides are there annually in England and Wales and is this number increasing?
- Are there groups of students at particular risk of suicide both demographically(i.e. age, ethnicity) and geographically?
- Are there particular clusters of suicides in terms of methods of suicide, geographic location?

The answers to the above questions will provide evidence to help prevent student suicide by providing better intelligence to the relevant public health and higher education bodies. Ultimately saving lives

The results should also present evidence to feed into national policy around specific suicide prevention measures such as the National Suicide Prevention Report and the Mental Health Taskforce and the Step Change Mental Heath Higher Education initiative this area. It will also show if there are particular clusters in certain higher education institutions.

National outputs will be published by ONS as aggregated data tables with the relevant statistical disclosure controls applied. If clusters are discovered in the data, an assessment will be made about the sensitivity of releasing this information publicly, to mitigate risk of 'copy cat' deaths. The approach would be agreed on a case by case basis with PHE, the lead government agency on suicide prevention. PHE would be able to use any evidence of clusters to work with specific Higher Education Institutes and relevant Local Government to put specific prevention measures and policies in place.

If successful we will look at making this an annual output.

UK Statistics Authority

Section A
Project Details

A1 Legal gateways

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

The Statistics and Registration Service Act 2007 (Disclosure of Higher Education Student Information) Regulations 2009 made on 25 November 2009 (SI 3201/2009), made pursuant to section 47 of the Statistics and Registration Service Act 2007, give Higher Education Funding Council for England (HEFCE) and Higher Education Funding Council for Wales (HEFCW) power to disclose to ONS certain data relating to higher education students for use by ONS for the purposes of producing population statistics, making arrangements for a census and assessing census returns. They also permit ONS to disclose some of that data to NISRA for the purposes of assisting the ONS to produce population statistics. HEFCE and HEFCW have each decided to exercise their powers to disclose data to ONS and through asking Higher Education Statistics Agency (HESA) Services to provide the data to ONS on their behalf.

The Memorandum of Understanding (MoU) has been signed by representatives of ONS, HESA ltd, HEFCE, HEFCW, Department for the Economy Northern Ireland & Northern Ireland Statistics and Research Agency

For this specific project we have asked HESA specifically if the planned research fits within the remit provided in the MOU. Both the Higher Education Funding Council for England and HESA are content for their data to be used for the purpose of this project under the above current agreement with ONS (ONS's Permitted Purposes). HESA are also content for the Welsh HESA data to be used for the same purpose. The written communication has been seen by the NSDEC Secretariat.

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

- [x] ONS
- [ ] SRS
- [ ] 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:

The study will contain persons who have died from suicide in England and Wales and who have been identified as having been in higher education at the time of their death for the academic years 2000-01 to 2012-13.

Suicides will be indentified using the National Statistics definition of suicide[1]. The data will also be linked to accidental hanging's and poisonings. This is line with the National Suicide publication where hard-to-code accidents are also included as these may be suicides. We will also include all accidental hangings and poisonings as a coroner may be more likely to record a death as accidental for this age group. It has been recommended by an academic to include these

The HESA Student record is collected in respect of all students registered at a reporting higher education (HE) provider who follow courses that lead to the award of a qualification(s) or HE provider credit, excluding those registered as studying wholly overseas. **Higher education students** for the purpose of HESA's data collection are those students on courses for which the level of instruction is above that of level 3 of the Office of Qualifications and Examinations Regulation (Ofqual) Qualifications and Credit Framework (QCF), or the former Qualifications and Curriculum Authority (QCA) National Qualifications Framework (NQF) (e.g. courses at the level of Certificate of HE and above).

[1]https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/bulletins/suicidesintheunitedkingdom/2015registrations?lang=welsh

It may be possible for a small number of records to be for students under the age of 18 years of age. We would filter those individuals out of the dataset using the age of student variable or the date of birth of the individual.

Individuals will be included in the study if their date of death coincides with the date of study on the HESA dataset, or within 6 months of finishing studying.

No other cause of death will be included in the data. Data will only be for deceased 18 years of age or over.

Justification for focusing on these subsections or groups:

All deaths that fall within the National Statistics definition of suicide and that are' hard to code' and who have been identified as having been in higher education at the time of their death, will be included. No specific demographic group (sex, ethnicity etc) will be targeted in this work.

We will include all narrative conclusions in England and Wales resulting from an external cause of death (an injury or poisoning) rather than a disease as it is possible that coroners may not have specifically stated the death as a suicide. Some narrative conclusions clearly state the intent (for example, accidental) and mechanism (for example, hanging, poisoning) of death. However, in other cases, the coroner may not indicate unambiguously whether the fatal injury was accidental, intentional or otherwise. We define deaths where the intent has not been specified as "hard-to-code". The rules for coding cause of deaths mean that, if no indication of intent has been provided by the certifier, a death from injury or poisoning must be coded as accidental. We include these deaths in our National Statistics Suicide bulletin, even though they are outside the scope of the definition

---

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

This project will link individual level death registrations for suicides to individual level student records for the academic years 2000-01 to 2012-13. Suicides will be identified using the National Statistics Definition which includes all deaths from intentional self-harm deaths where the intent was undetermined, and extra tables which included cases where they have been coded 'hard to code'. For the purposes of this study only deaths of persons 18 years and over will be included.

The variables that the linked dataset will contain from both HESA and the Mortality file are listed in detail in Annex 2. Broadly the linked dataset will contain information taken from the death certificate such as cause of death, place of death and date of death, alongside HESA data around mode of study (part-time/fulltime), institute of study, start and end dates.

The linkage will work as follows. HESA has already provided ONS with the student data for other data linkage projects in the office as outlined in the legal gateway section. This is stored in the ONS Statistical Research Environment. The HESA data will be securely passed to the SQL restricted server via a secure encrypted USB stick (Iron Key) to the

Health Analysis and Life Events Division (HALE) server. This process has been signed-off by the ONS Information Assurance team. The HESA data which will contain names and addresses will be placed in a secure partition of the restricted SQL database on this server, along with the subset of suicides data with names and address of the deceased. Two Data Managers will have access to this partition – no one else. They will both be **security cleared and trained** in handling sensitive data.

The Data Managers will link the data using keys such as names, date of birth, place of usual residence, university address. Staff having access to identifiable data will not be involved in the ensuring research.

Once linked identifiers are removed from the linked data this subset will be stored on separate restricted SQL database on the same server, where two to three ONS researchers will have access to undertake the analysis. The researchers will not have access to the identifiable data as it is best practice to keep separate the staff working on the identifiable data and those on the de-identified data. Staff in HALE have SC security access and training to ensure they handle data correctly (see section B3)

The research questions to be answered will include:

- How many student suicides are there annually in England and Wales and is this number increasing?
- Are there groups of students at particular risk of suicide both demographically(i.e. age, ethnicity) and geographically?
- Are there particular clusters of suicides in terms of methods of suicide, geographic location?
- How does the new student suicide estimates match with the current ONS estimates?

In some instances the findings from the analysis may result in us wanting to refer to particular individual coroners text which ONS hold. This text provides more detail around the cause of death and circumstance around the death. The text would mainly be used if we identify any particular clusters of suicides to see if there are common elements to the deaths. This information would not be published but aggregate data will be passed securely to PHE to assist them in any prevention work.

The outputs of the analysis will be a statistical article and associated aggregated tables. These will be published as experimental statistics on the ONS website and the aggregated tabulations will have been assessed for disclosure control against the current Mortality Disclosure Control protocol (which is published on the ONS website).[2]

[2] https://www.ons.gov.uk/file?uri=/methodology/methodologytopicsandstatisticalconcepts/disclosurecontrol/guidanceforbirthanddeathsstatistics/disclosurecontrolguidanceforbirthanddeathstatisticsv2tcm77351338.pdf

UK Statistics Authority

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)*		Person level ONS mortality data – the academic years 2000-01 to 2012-13 Person level HESA data for the academic years 2000-01 to 2012-13 Data only available to 2 Data Managers	Linked HESA and mortality data based on the identifiable data. Provided to 3 ONS research officers.	
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)*				

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)

Student suicides are an emerging high priority in health policy. The Secretary of State for Health alongside Public Health England have emphasised that research into these suicides is crucial to provide evidence to assist prevention measures. The priority of the work has been highlighted by the clusters of suicides in both Bristol and York Universities reported in 2017. This project will go some way to providing a better national picture of student suicides in England and Wales.

The independent Mental Health Taskforce[3] recommendations is to reduce suicide rate in England by 10% by 2020/2021 and for every local area to have in place a multi-agency suicide prevention plan. The research outcomes of the study will contribute directly to provide evidence to meet government initiatives such as the House of Commons Health Committee Suicide prevention interim report 2016-17 which highlighted the need for the Government to re-examine its own policy areas to identify missed opportunities to reduce the risk of suicide[4].

The Secretary of State for Health recently blogged on the Times Higher Education Website:

'….we know when it comes to suicide you can make big improvements in prevention. This can only be done when you have accurate and reliable data – and we know sometimes the figures for student suicides can be understated if, for example, "student" is not recorded properly as someone's occupation. So Universities UK, Public Health England and the ONS are now working with researchers from University of Worcester to encourage more accurate reporting of student suicides.'

Currently ONS provide an estimate of 161 student suicides in 2016 using occupation data recorded on the death certificate. We identify students using the National Statistics Social Economic Classification (NSSEC) category 'Full time student'. This estimate is not robust as it is based on the information provided by the informant at the time of death registration and is likely to be an underestimate, missing groups like part-time students.

The exploratory linkage of individual death registration data for suicides to HESA records will provide more accurate estimates of student suicide in higher education (see annex 2 for list of variables to be contained in the linked dataset). The analysis will provide more insight into if there are patterns of suicide by broad study areas, year groups, full or part-time students, location. Also we may be able to identify if there are particular clusters occurring in particular geographic locations, or particular methods of suicides used. Identifying clusters will enable PHE to work with Local Authorities, charities and Universities to put in place specific prevention measures for the institution.

Each output would be assessed to make sure the data is non-disclosive following the ONS guidelines. ONS is used to dealing with suicide data and are fully aware of the sensitivity of the outputs. To avoid the potential for 'copy cat' deaths we are careful not to publicly release

[3] https://www.england.nhs.uk/wp-content/uploads/2016/02/Mental-Health-Taskforce-FYFV-final.pdf
[4] https://publications.parliament.uk/pa/cm201617/cmselect/cmhealth/300/300.pdf

information about new suicide methods or clusters. However we do report these in confidence to the National Suicide Prevention Advisory Group and PHE to inform them of new areas of prevention. We would apply the same statistical disclosure rules to this analysis.
It is important to note that ONS would not produce any type of tabulations that ranked Higher Education Institutes by numbers of suicides. That is not the purpose of this work.

All this information will help inform Public Health England and Public Health Wales in their prevention measures. The evidence will also help provide other partners such as NHS England, Local Authorities, Department of Education and the Universities themselves with information to help prevent suicide.

In summary, the research outcomes should enable the production of more accurate evidence base to inform public policy makers, government departments, academic institutions, and voluntary organisations. This could lead to more informed interventions to prevent suicides in higher education.

This study is part of a suite of enhancements being made to the ONS suicide data. We have another project investigating links between prescription drugs and suicides. We have recently published new analysis looking at suicide risk in occupations.

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 linked data will be held by ONS for up to 3 years specified in the MOU with HESA. Any extension to this would be in agreement with HESA.

The data available to researchers in ONS will be limited as specified in section A5. The researchers and data managers that will have access to the data will have had comprehensive training on access and outputting these types of sensitive data (see section B3).

Products of the analysis will be published on the ONS website and the aggregated tabulation will have been assessed for disclosure control against the current Mortality Disclosure Control protocol which is published on the ONS website.

If useful results are found we will make the case to make this a more regular piece of work or extend the work to include a control group from the population. Any extension would be undertaken in agreement with HESA and via an application to the National Statistician's Data Ethics Advisory Committee.

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 HESA data is currently stored in the ONS Statistical Research Environment. The specified HESA data containing research variables (see Annex 2) and linkage variables (including person's name, home and term address, postcode and data of birth) will be transferred to HALE to allow us to link this data to the mortality file. The data will be transferred internally via encrypted USB stick (Iron Key) into the HALE restricted environment in ONS. This environment is a secure server based on the ONS Titchfield site which hosts a number of SQL databases containing all the ONS record level mortality data from suicides through to abortions.

Data Managers will also extract mortality data for suicides for the same time period as the HESA data and store that data in the same SQL database. This database will only permit access to two named HALE Data Managers who will be responsible for the proposed linkage. Linkage will be undertaken using the keys name, address (place of usual residence from the death certificate & home and term), postcode (place of usual residence from the death certificate &home and term), and date of birth. The linked dataset will be stored in the secure SQL database.

The linked data will be stripped of identifiable variables before being released to another SQL database in the same HALE Environment for three ONS researchers to undertake the analysis. The research will take place in the ONS, on a restricted access server environment. LEO. Access to the data bases on the server is restricted to authorised users in the ONS HALE team by username/password logins. Access lists are reviewed quarterly. Record level data will not leave this secure environment.

ONS researchers in the HALE have a high level of training in handling person level death information, This includes:

- additional security vetting to the level of Security Cleared. This has been provided due to the sensitive nature of the data they have access to
- trained in data protection through Office wide courses and have a thorough understanding the mortality disclosure control policy
- the research and outputs from the analysis will be overseen by the Assistant Deputy Director of the division, an expert in Health Related Disclosure Control
- all staff are highly aware of the sensitive nature of the data we hold.

To mitigate the risk of identification of the deceased in any publication the data will be aggregated. Aggregated data will be published at the National Level (England and Wales) and then assessed to see if relevant sub-national tables can be produced following the published Mortality Disclosure Control policy. This applies threshold rules to table cells to ensure that individuals are not able to be identified in the aggregated tabulations. Cells with small counts (less than 3 records) would likely be suppressed in the aggregated tables to protect identification of the deceased. Also categories in tables would be combined if cells counts were too small. For example 5 year age bands may be applied rather than single year of age. The statistical disclosure control threshold of three cells is selected in line with best practice as identified in the ONS Mortality Disclosure Control policy

Results from the research will only be released by ONS staff once they have confirmed that they contain no risk of identifying an individual or a household. This is assured by the statistical disclosure controls the ONS HALE team currently have in place.

No consent is required by the data subjects or their families as the Data Protection Act does not apply to deceased.

The data will not contain deaths of individuals under the age of 18 years old.

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 are involved in the process and safeguards are in place (see section B3) to make sure that any risks introduced by the proposed linkage are mitigated. All processes have adequate human control to ensure appropriate quality assurance arrangements are in place. The linkage will be undertaken using probabilistic matching using keys such as name and address. We are confident that the matching should have a high success rate based on the linkage methodology used but until we start the linkage work it is not possible to be absolutely certain.

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 Statistics and Registrations Service Act (2007) applies to this work, as it provides ONS with the legal gateway to provide the data processing service to HESA (section 22 of the act), and for ONS to link our mortality data to the higher education data (section 42 of the act). The SRSA also protects the confidentiality of the information provided (section 39 of the act).

It is important to note that the Data Protection Act does not apply to this project due to the analysis being undertaken on deceased individuals, The DPA only applies to living individuals.

With regards to the Human Rights Act there is no interference with the right to family life and privacy (HRA Article 8) of the deceased or their surviving family members. No information will be collected about family members and they will not be contacted or identified in any circumstances. The data will be fully anonymised and subjected to strict Statistical Disclosure Controls so there is minimal risk of causing any harm or distress by a breach of confidentiality. The research methods used are sanctioned by law (see above) and proportionate to the public interest in the protection of health.

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)
Professor Louis Appleby – Chair of the National Suicide Prevention Advisory Board	Providing guidance in how to report the findings of the work to reduce to ensure we take account of the sensitivities around 'copy cat' deaths. No access to disclosive data will be provided (see letter of support in Annex 1)
Helen Garnham, National Programme Manager for Public Mental Health - PHE	Providing guidance in how to report the findings of the work to reduce to ensure we take account of the sensitivities around 'copy cat' deaths. No access to disclosive data will be provided (see letter of support in Annex 1)
David Gunnell, Prof. Epidemiology, Bristol University	Assistance in designing the analysis, and interpretation of results. No access to disclosive data will be provided (see letter of support in Annex 1)

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)

ONS has sought feedback from the National Suicide Prevention Strategy Advisory Group. The membership is a mix of government bodies, charities, health providers, academics and family champions bereaved by suicide. There was broad support for the work and no objections provided. The Chairman, Prof. Appleby explained the importance of this work for the group in a recent email correspondence (annex 1):

'The limited information we have at the moment suggests a recent rise in the number of suicides by students and further studies are urgently needed into causal factors. However, the starting point for better understanding of the problem is more accurate figures and it is for this reason that the HESA-ONS linkage would be so valuable to future prevention.'

Also in recent months ONS has been contacted to provide data on student suicides via Parliamentary Questions (House of Lords), media, charities and non-government organizations which shows that there is a growing public interest in this area of research and a gap in knowledge that ONS can fill.

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 non-disclosive aggregated tabulations of the results of this project will be initially shared with Public Health England, and the National Suicide Prevention Strategy Advisory Group (NSPSAG). These organisations/ groups have are involved as PHE has the lead remit in government for suicide prevention and the NSPSAG is responsible for the monitoring of the government suicide prevention strategy Following feedback from experts in this group we will then decided how best to disseminate the findings moving forward. No microdata will be shared outside of the small select group of researchers and data managers in the HALE division in ONS who will access the linked data.

If clusters are discovered in the data, an assessment will be made about the sensitivity of releasing this information into the public domain, to ensure we minimise any risk of 'copy cat' deaths. This is type of assessment is common when ONS are handling suicide data as we do not want to put in the public domain new novel methods of suicide, or highlight publicly current clusters of suicides. Discussions would be had with PHE, the lead government agency on suicide prevention to agree the best approach in using the data in these circumstances.

The outputs of the analysis will be a statistical article published on the ONS website and associated aggregated tables. These will be published as experimental statistics on the ONS website and the aggregated tabulations will have been assessed for disclosure control against the current Mortality Disclosure Control protocol (which is published on the ONS website).[5]

[5] https://www.ons.gov.uk/file?uri=/methodology/methodologytopicsandstatisticalconcepts/disclosurecontrol/guidanceforbirthanddeathsstatistics/disclosurecontrolguidanceforbirthanddeathstatisticsv2tcm77351338.pdf

Section C
Responsible owner and applicant details

C1 Responsible Owner

Full Name: [redacted]	**Position:** [redacted]
Address: [redacted]	**Email:** [redacted]
	Telephone: [redacted]
	Organisation: [redacted]

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)

UK Statistics Authority

Full Name: [redacted]	**Position:** [redacted]
Address: [redacted]	**Email:** [redacted]
	Telephone: [redacted]
	Organisation: [redacted]

Annex: Variables to be held on the final analysis file

List of mortality variables

Variable name	Description
Agegroup 2	Age
Age in years	Age
Cerdets	Type of medical certificate presented
Certific	Certified cause of death indicator
Certype	Certification type
Cestrss	Communal establishment code
Cestrssrs	Communal establishment code
Ceststay	Duration of stay in communal establishment.
Ctryir	Country indicator code for usual residence of deceased
Doddy	Day of death
Dodmt	Month of death
DoR	Date of registration
DoD	Date of death
Empsecdm	Employment status (ONS code) of deceased or mother of deceased juvenile for SOC2000
Esttyped	Communal Establishment type where death occurred
Fic10und	Final underlying cause of death (ICD10)
Fimdth10	Final manner of death (ICD10)
Ic10f001-15	Final ICD10 code
Marstat	Marital status of deceased
NHSIND	NHS Establishment indicator
PCDPOD	Postcode of place of death
PCDR	Postcode of usual residence of deceased
Ploacc10	Place of occurrence of accident (ICD10) resulting in death.
Postmort	Whether a postmortem took place
refcor	Whether referred to coroner
regyr	Year of registration of death
Seccatdm	NS SEC operational category for deceased or mother of deceased.
sex	Coded sex of deceased
Soc2kdm	Standard occupation classification (2000) for deceased or mother of deceased child
Wpla10	Workplace code
REGDETS	ONS unique identifier for death records.
CODT	Cause of death text

List of HESA variables

Variable	Description
Ethnicity	Students domiciled are required to report their ethnic origin.
Nationality	This field defines the country of legal nationality. If the student has a dual nationality including UK they are coded as United Kingdom. If the student has dual nationality including non-UK EU country, they are classified with the relevant EU country code. If the student has neither UK or no-UK EU country, then they are classified by either nationality.
Mode of Study	Full-time and Sandwich/Full-time students are those normally required to attend the institution for periods amounting to at least 24 weeks within the year of study. Part time students are those recorded as studying part time, or studying full-time on courses lasting less than 24 weeks e.g. on block release or studying during the evenings. Writing Up students are those who normally expected to submit a thesis to the institution for examination have completed the work of their course and are not making significant demand on institutional resources. Sabbatical students are those that are on sabbatical years eg Holding Student Union office. Students on years out for any reason are returned as Dormant.
Campus Identifier	The Campus identifier is used to identify different campuses with a physical location. Note: Campus information is sent at the discretion of the institution.
Location of study	This describes the student's location of study. For those students studying outside the UK they are not counted.
Term Time type of Accommodation - TTAccom_HESA	The term time accommodation field identified where the student is living during their period of study.
Year of Student on this instance - Yearstu_HESA	This field is used to indicate the year number that the student is in since enrolling for a course/programme leading to the student's aim. Note this could be different from the year of course if the student has changed course or retaken a year.
Reason for ending instance	This field identifies the reasons why students leave.
Course Aim	The level of study the student is taking ie postgraduate, first degree and other undergraduate.
Start date of Instance	This is the date of the student's initial commencement of studies and may relate to a date prior to the current academic/financial year. The above format is for the input format. This is converted to SAS DDMMYY10 format.
End date of Instance	This is the date of the student's end of studies. The above format is for the input format. This is converted to SAS DDMMYY10 format.
Campus Name	eg Walsall Campus
Local Authority District of Campus	This refers to the Local Authority which the campus is based.
The time students have completed from the beginning of their study.	The amount of time in units that students have completed from the beginning of their study year (this normally includes time for examinations) Does not include study time, learning time or contact time. Where there are both part-time and full-time students following a programme of study, the expected length of study should

Variable	Description
	be the normal length applicable for the mode of study of the student.
Unit of Length related to the SPLENGTH variable	The units of length associated with SPLENGTH. Code Label 1 Years 2 Months 3 Weeks 4 Days 5 Hours 9 Not applicable

ADRN – Developing persistent dataset for projects within the Data for Children research theme

Oral report

Leon Feinstein
Paul Jackson

1.	2.	3.	4.	5.	**6.**	7.	8.	9.	10	11.	12	13	14	15	16

DCMS – Update on the Ethics Commission on Data Use

Oral report

Nick Dodd

1.	2.	3.	4.	5.	6.	7.	8.	9.	10	11.	12	13	14	15	16

Self-assessment process for ethical consideration

Oral report

Peter Fullerton

Petros Saravakos

1.	2.	3.	4.	5.	6.	7.	**8.**	9.	10	11.	12	13	14	15	16

UK Statistics Authority

National Statistician's Data Ethics Advisory Committee

NSDEC(18)04

Ethical guidelines for social media research

This paper is in draft and will be published in due course

1.	2.	3.	4.	5.	6.	7.	8.	**9.**	10	11.	12	13	14	15	16

UK Statistics Authority

National Statistician's Data Ethics Advisory Committee

NSDEC(18)05

Estimating ethnicity from names

This paper is in draft and will be published in due course

1.	2.	3.	4.	5.	6.	7.	8.	9.	**10**	11.	12	13	14	15	16

UK Statistics Authority

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

An exploratory study for estimating the outflow of Welsh speakers from Wales to England from the 2011 Census and Patient Register (ADRN Project Number PROJ-108)

Start Date: **End Date:**

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

Welsh Government (Knowledge and Analytical Services)

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

The proportion and number of people speaking and using Welsh is of considerable importance to the Welsh Government, public authorities and the general public in Wales. One of the statutory well-being goals of the Well-being of Future Generations (Wales) Act is "A Wales of vibrant culture and thriving Welsh language". Public bodies in Wales must develop well-being assessments and well-being plans to demonstrate their contribution to those goals. Furthermore, the Welsh Government has set out an ambition to reach a million Welsh speakers by 2050. In order to measure the number of Welsh speakers and in particular to monitor the ambition of a million speakers by 2050, it is important to be clear about the factors that may have an impact on this, including migration. Results from the 2011 Census showed that there had been a decrease in the number of Welsh speakers between 2001 and 2011 owing to migration and other demographic changes in the population (in addition to changes in peoples' skills between Censuses).[1] The Welsh Government's Welsh Language Strategy Evidence Review also notes "the fragmentation and rapid erosion of predominantly Welsh-speaking areas in the North and West (Y Fro Gymraeg) through large-scale migration and selective out-migration, particularly of young people".[2]

The loss of Welsh speakers from a community can have a significant impact on the characteristics and demography (in terms of language) of a local area. This in turn has a significant impact on planning services.

Little is known about how many Welsh speakers live outside Wales, or when Welsh speakers leave Wales. Previous work by the former Welsh Language Board, and later by the Welsh Government, broadly estimated the annual change in the number of Welsh speakers by modelling migration flows and estimating how many fluent Welsh speakers leave Wales annually using aggregate information from various incomplete data sources.[2,3] As results from the Welsh Language Use Survey 2013-15 showed, there are clear links between fluency and the use made of the Welsh language therefore the out-flow of Welsh speakers from an area will have an impact on the vitality and use made of the language in that area.[4]

Since none of the primary sources of data for the whole population in England collect information on Welsh language ability, it is very difficult to model the out-migration of Welsh speakers other than by using the Census Longitudinal Study (only once a decade) or the Pupil Level Annual School Census (coverage only for children) and these estimates of outward migration have been subject to large error. This research project would provide the opportunity to develop more accurate and robust estimates of the outflow of Welsh speakers from Wales to England.

Using the NHS patient register in England will allow us to understand the flow of Welsh speakers from the Census to locations in England and thus provide an evidence base to support future policies and strategies around the Welsh language, service provision and local development plans which in turn should provide better well-being outcomes for Welsh speakers.

This project will also be a proof of concept for the use of Census data in conjunction with administrative sources to fill an evidence gap that has long been the source of frustration for Welsh language researchers and policy makers.

The aims of the project will be:
(i) to estimate the number of Welsh speakers in Wales (according to the 2011 Census) that have migrated to England in the subsequent years using the NHS patient register as a data source.
(ii) to analyse the geographical origin and destination of those migrants (with the main focus being on the individual's location in Wales as recorded on the Census), and the demographic characteristics of Welsh speakers that leave Wales
(iii) to produce a summary dataset at local authority level that can be used for further analysis by Welsh Government that does not identify any individuals or attributes about individuals
(iv) to provide useful evidence to the Census transformation project on the use of administrative data to support Census-type outputs and to understand more about the demand and need for information on Welsh speakers outside Wales.

1. Welsh Government. 2012. 2011 Census: First Results on the Welsh Language [Online]. Available at: http://llyw.cymru/statistics-and-research/census-population-welsh-language/?skip=1&lang=en

2. Welsh Government. 2012. Welsh Language Strategy Evidence Review [Online]. Available at: http://gov.wales/statistics-and-research/7430545/?skip=1&lang=en

3. Welsh Language Board. Derivation of rough estimate of annual change in fluent

Welsh speakers.

4. Welsh Government and Welsh Language Commissioner. Welsh language use in Wales, 2013-15 [Online]. Available at: http://llyw.cymru/statistics-and-research/Welsh-language-use-survey/?skip=1&lang=en

Section A
Project Details

A1 Legal gateways

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

Census of Population data will be obtained from the Office for National Statistics. It will be obtained under Section 43 of 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
- ☐ 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:

Project will be focused on persons who identified themselves as Welsh speakers in the 2011 Census

Justification for focusing on these subsections or groups:

The purpose of the project is to understand the movement of Welsh speakers following the Census, as no other data exists on Welsh speakers living in England. General migration trends for the whole population are analysed elsewhere.

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

Data from the 2011 Census in England and Wales will be linked to data from the NHS Patients Register at the individual-level, only for those individuals living in Wales in the 2011 Census. Matching will therefore be conducted on name, sex, age and address (although some of this matching has already been done by ONS). The Census will form the spine point, to which will be appended those appearing in the NHS Patients Register in England. The final analytical dataset will therefore contain only information relating to individuals on both the Census in Wales and NHS Patients Register, rather than the entire Census dataset. The linked dataset will be non-identifiable.

The Census and NHS Patients Register will be used for data linkage as no other such data exists that would enable us to estimate migration patterns of Welsh speakers from Wales to England. The Census dataset has the advantage that it covers the whole population and includes information on the ability of the population (aged three and over) in Wales to speak Welsh.

The linkage will be carried out in the Statistical Research Environment (SRE) in ONS Titchfield. The SRE has been designed specifically to address any privacy and security concerns that may arise when datasets are linked together for research purposes. Robust

security controls are in place to ensure the safety of that information and to ensure that confidentiality is protected. The security controls are covered in greater detail in the ADRC-E Safeguarding Paper.
https://www.ons.gov.uk/file?uri=/aboutus/whatwedo/programmesandprojects/theadministrativedataresearchnetworkcollaboration/adrcsafeguardingpaper_tcm77-404473.pdf

The datasets are received, pre-processed and hashed individually by one team and then linked together after hashing by a separate team – so no person sees multiple datasets together unhashed or 'in the clear'. Pre-processing involves cleaning of the variables to standardise date variables, remove punctuation from name fields and geo-code and validate address information.

These datasets have been selected due to the Census being the best quality, broad coverage reference for Welsh language speakers and the Patient Register is assumed to have a relatively up to date address location for the individual.

The linked, de-identified record-level dataset will be checked to ensure it meets the security level of the SRS and then moved to SRS for the researchers to access. Any exports from SRS are subject to strict statistical disclosure control checks.

The research dataset will be deleted once the researchers have finished analysing it as per the ADRC-E protocol.

The Parliamentary Under Secretary of State for Health, Lord O'Shaughnessy has given his consent (on behalf of the Secretary of State for Health) for using the NHS Patients Register in England for this specific purpose.

Data will be accessed from either the ADRC-England safe setting, or if it would be possible to access the datasets through the SRS we could access using the Welsh Government's remote SRS access. The data subject's identity will be protected and the information will be kept confidential and secure. The linked data set accessed by the researchers will only be de-identified data and the data linking will be done by a trusted third party so that staff and researchers cannot see information which would identify individuals, for a strict separation of activities.

Demographic data required for the project will be sourced from the Census. This includes details such as age, sex, area of residence, ability in Welsh, economic activity and qualifications. The project only requires individual's demographic registration details from the NHS Patients Register - no information is required on any individual's clinical history. The resulting dataset will be a linked, de-identified data set. As per the conditions of access for the Administrative Data Service, a short summary of the research will be produced alongside an aggregate dataset that will be subjected to statistical disclosure controls. The linked data will only be retained for a period of 12 months and be destroyed securely at that point.

As the proposed study is exploratory, the methods used will predominantly be descriptive. The methods can however be grouped into several aims:

1. The first phase of the project will be to produce counts of Welsh speakers who appear on the NHS register and to use this information to estimate the annual level of outward migration of Welsh speakers to England since the 2011 Census. In this phase we would analyse the geographical location of those Welsh speakers through simple cross-tabs and GIS maps (limited to local authority level). These estimates would then be compared with previous estimates of the out-flow of Welsh speakers from Wales (by the Welsh Language Board and Welsh Government) to assess their reliability and accuracy.

The remaining phases of the project will depend on the numbers of people in question.

2. Assuming that there are sufficient numbers to produce worthwhile analysis the second phase of the project will be to develop an origin and destination matrix of Welsh speakers by local authority (with the main focus being on the origin local authority of the individual). We would expect to produce these matrices for different age groups to understand migration by students and non-students. For example, is the out-flow of Welsh speakers higher in local authority areas around the Wales-England border? Or is the out-flow of Welsh speakers higher from those local authorities with the highest densities of Welsh speakers?

3. The third phase of the project will be to present descriptive analyses of the characteristics of the Welsh speakers who leave Wales including (according to age, gender, qualifications, economic activity, and Welsh-speaking background of adults in the households etc.) to understand the characteristics of those Welsh speakers that move from Wales to England. This, again, will be done by simple cross-tabs and significance tests.

4. Again if numbers suffice we would anticipate also undertaking logistical regression analysis to understand the relationship between these variables and the likelihood of moving from Wales to England. This analysis will give us an insight into the characteristics of those Welsh speakers who appeared in the Census in Wales but have subsequently moved to England.

5. As this study is exploratory, we would also be conducting some simple descriptive analysis of the demographics of individuals who didn't identify as Welsh speakers in the Census but have subsequently moved to England. This will enable us to compare the demographic characteristics of those Welsh speakers moving to England with the demographic characteristics of the non-Welsh speakers, i.e. are Welsh speakers less likely than non-Welsh speakers to move to England?

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)*			NHS register, 2012 onwards	
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 2011	
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)

This project has substantial benefits to the evidence base around migration and specifically the Welsh language, as there is currently little information available about the out-flow and prevalence of Welsh speakers outside Wales. This project will improve our understanding of migration in Wales by enabling us to analyse migration patterns by demographic characteristics, in this instance migration by ability to speak Welsh. This will provide a significant improvement to the evidence base to support future policies and strategies towards promoting and facilitating around the Welsh language, service provision and local development plans which in turn should provide better well-being outcomes for Welsh speakers in Wales and England.

Previous work by the former Welsh Language Board, and later by the Welsh Government, estimated the annual change in the number of Welsh speakers by modelling migration flows and estimating how many fluent Welsh speakers leave Wales annually. This study would provide the opportunity to replicate, validate or challenge this existing research and provide more reliable estimates of the flow of Welsh speakers between Wales and England.

This is a long standing deficiency in the evidence base which is continually raised with Welsh Government particularly during debates on the content of the Census of Population. Most recently it was raised in a workshop held by ONS and Welsh Government with stakeholder groups concerning Welsh Language topics for the 2021 Census.

The project would provide significant societal benefits in terms of understanding the changing nature of Welsh language demographics in Wales, which will help understand what type of Welsh speakers move away from Wales. It will help us explain the changes in overall numbers of Welsh speakers and measure the contribution to this that can be attributed to outward migration. The outcome of this project will develop the evidence base for public policy decision-making around the Welsh language. Policy makers will be able to use the outcomes of this study to identify areas that have experienced the greatest outflow of Welsh speakers, and to understand the factors behind that migration and the implications for local service provision, community cohesion and local development plans. Having a better understanding of migration patterns and the impact that this has on the number of Welsh speakers will also aid the evaluation of public policy in how successful it is at increasing the number of Welsh speakers to one million by 2050.

Increasing the numbers of Welsh speakers is an important priority for the Welsh Government and they and public authorities are required to work towards the goal of a thriving Welsh language under the Well-being of Future Generations (Wales) Act. As part of this they are required to undertake needs assessments and set well-being objectives which will be based on evidence. To understand fully the prevalence of the Welsh language, they need to consider the impact of demographic changes on Welsh speakers, including the impact of migration. This will help produce informed policy to support the promotion of the Welsh language.

The project will also provide an additional evidence base to support public authorities, (particularly those who are in scope of the Welsh Language Standards), some of whom provide services across England and Wales (or the UK as a whole). Under the Standards those authorities will be expected to provide services through the Welsh language to those who need it and at present there is very little information to help them understand the impact of the Standards on their services to residents of England. There is significant interest from broadcasters such as BBC, S4C and Radio Cymru in understanding the prevalence of Welsh speakers outside of Wales to ensure that they are targeting their digital provision effectively. Therefore Welsh speakers in England will directly benefit through public authorities and broadcasters understanding the scale of Welsh language provision that are required and in turn provide better services for those citizens.

A wider benefit will be a contribution to the Census Transformation Programme, which is aimed at developing recommendations for modernising the Census in future by using more administrative data. A successful Census Transformation Programme has potentially huge benefits for society, value for money and the statistical infrastructure of the nation. The project will improve the understanding of the use of administrative data to support Census-type outputs and also help establish the need and demand for information about the Welsh language outside Wales. Understanding the scale and nature of outward migration may also support the targeting of future research in this area.

The key impact of the study will be to feed into the evidence base for policy-making around the Welsh language. Improvements in the evidence base would be expected, in due course, to deliver societal benefits in terms of meeting the goal within the Well-being of Future Generations (Wales) Act 2015 of a vibrant culture and thriving Welsh language.

The findings of the research will be presented to the Welsh Language Policy Board in the Welsh Government and to other groups such as the Welsh Language Partnership Council, chaired by the Minister for Lifelong Learning and the Welsh Language, who are responsible for giving advice and making representations to Ministers in relation to the Welsh language strategy. Policy officials and key stakeholders will be able to fully consider the outcomes and implications of the research study in light of the new Welsh language strategy, published in July 2017.

We would also publicise this research throughout the community of Welsh language researchers and stakeholders, alongside our statistical communities such as the Welsh Statistical Liaison Committee. We would also disseminate the findings to statisticians involved population and migration statistics, and as described above, the Census Transformation Programme team.

We would also seek to publicise the results through presentations for example in academic conferences or the National Eisteddfod for Wales, given the importance of the findings to the Welsh Government's cultural commitments. The Welsh Language Commissioner's officials have already expressed a strong interest in the outcome of this research for their own use.
Any research on the Welsh language is likely to have significant national media coverage within Wales and we will ensure the media is aware of the work and provide briefing to them if required.

Welsh Government also chair a meeting of UK public bodies who have a presence in Wales and we would present the findings to that meeting to ensure that UK bodies have a better understanding of the scale of Welsh Language service provision needed in England.

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)

As part of accessing the linked data resource through the ADRN, this full linked dataset will not be available for use by other researchers at a micro level. However the meta-data and the coding used in generating the initial data set will be retained for future use in order to re-create the data set, and perform other linkages between administrative and Census data. The aggregated dataset at local authority level created through this project will be used by Welsh Government and potentially other researchers for future analysis. The linked data will only be retained for a period of 12 months and be destroyed securely at that point.

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 data subject's identity will be protected and the information will be kept confidential and secure. The linked data set accessed by the researchers will only be de-identified data and the data linking will be done by a trusted third party so that staff and researchers cannot see information which would identify individuals, for a strict separation of activities: https://adrn.ac.uk/protecting-privacy/de-identified-data/trusted-third-parties

Data will be accessed from either the ADRC-England safe setting, or if it would be possible to access the datasets through the SRS we could access using the Welsh Government's remote SRS access.

Persons accessing either safe setting will undergo training in the use of these settings, in addition to training on statistical disclosure. The training undertaken will be the ADRN accreditation training to become an 'accredited researcher': https://adrn.ac.uk/media/1162/adrn013-accreditedresearcher_v0100_pub.pdf

ONS have sought and received consent from the Secretary of State for Health to provide access to onwardly disclose the patient register in line as per requirements in the Statistics and Registration Service Act 2007.

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)

There will be sufficient human oversight so that the methods employed are consistent with recognised standards of integrity and quality. The ADRC will be involved in the output checking for the project to ensure information released is aggregate and non-identifiable

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 proposed data linkage is based on the legal framework set via the Statistics and Registration Services Act and the Data Protection Act.

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	Researchers from Welsh Government will be undertaking the analysis
Welsh Government (Knowledge and Analytical Services)	

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)

There is no need for direct public engagement however there is increased interest from the Welsh Government, public bodies and voluntary sector organisations in Wales to obtain more information on migration patterns, in particular the migration patterns of Welsh speakers in light of the target of a million Welsh speakers by 2050. The Welsh Government has sought the views of a wide range of external stakeholders, including the charitable sector. In communicating our analytical plans for 2017-18 and beyond with our Welsh language stakeholders (including at events such as the ONS workshop on the Welsh language in the 2021 Census) we have signalled our intent to provide more definitive information on the migration patterns of Welsh speakers. It's also worth noting that the Administrative Data Research Network has undertaken significant public engagement of its own regarding widening the use of administrative data for the public good, and this research project falls under the umbrella of the work of the Network.

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)

A report will be published as a Government Statistical Service (GSS) publication on the Welsh Government website. The report will be accompanied by an accessible 'short report' or executive summary that will present a summary of the research findings and conclusions arising from the research project in an accessible manner for those who have an interest in the findings of the project but not the statistical detail, including policymakers.

An important aspect of this work will be supporting the production of metadata and user guides in relation to the data used for this project. In line with the conditions of access, a short summary of the research will be produced for the Administrative Data Service.

Section C

Responsible owner and applicant details

C1 Responsible Owner

Full Name: [redacted] **Position:** [redacted]

Address: [redacted]

Email: [redacted]

Telephone: [redacted]

Organisation: [redacted]

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:	**Position:**
Address:	**Email:**
	Telephone:
	Organisation:

UK Statistics Authority

1

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

Household, parental and environmental risk factors for hospital admissions in children

Start Date: **01/06/2018**

End Date:31/05/2022

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

University College London
Office for National Statistics

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

Emergency hospital admissions in children aged less than 15 years in England have increased by over 30% since 2000.[1] These admissions are highly stressful for parents and children, as well as costly for the National Health Service (NHS). Identifying risk factors for hospital admissions in children and determining their relative contribution can help policy makers in local and central government and the NHS identify areas where improvements in health service access and public health interventions are likely to lead to the greatest improvements in child health.

In this project, we will focus on examining environmental and household risk factors for

[1] Gill PJ, Goldacre MJ, Mant D, et al. Increase in emergency admissions to hospital for children aged under 15 in England, 1999-2010: national database analysis. *Arch Dis Child* 2013; 98: 328-34.

hospital admissions in children, and whether children whose parents were born abroad face barriers to accessing preventive primary and community health services, which in turn leads to the need for hospital admission.

Respiratory infections are the most common reason for hospital admission among children less than two years old in England. Environmental risk factors, including overcrowding, quality of housing and chronic exposure to outdoor and indoor air pollution during pregnancy and infancy have been found in smaller European studies to adversely affect children's respiratory health,[2,3] however there are few studies examining their effect on hospital admissions, and none from the UK.

In 2015, 28% of children in England were born to mothers who are born abroad, however the most recent UK studies of the impact on parental migration on access to health care for children are now over 25 years old. All children should have access to appropriate health care, and children whose parents were born abroad should not be disadvantaged.[4]

Although national hospital admission databases provide rich clinical data there is very limited information about the parents, household and local environment in which children grow up. This project will examine a birth cohort of all babies born in England up to one year before and after the 2011 Census date (2010-2012). This birth cohort is available in the Secure Research Service (SRS) as a result of linkage of ONS birth registration data, NHS birth notification data and longitudinal hospital admission data for both mothers and babies by City, University of London. We will link this birth cohort with de-identified 2011 Census data and aggregate small area data on air pollution to examine the impact of:

1. environmental factors e.g. air pollution, type of housing, indoor air pollution (measured via type of heating) and occupancy rate to the risk of respiratory hospital admissions in the first two years of life.
2. household factors i.e. country of birth, year of migration and knowledge of English to the risk of hospital admissions in children. Here we will focus on hospital admissions which are considered preventable via timely access to preventive primary and community health services.

We will adjust for other important risk factors including prematurity, congenital anomalies, and socio-economic status. These results will be used to inform policy makers in public health and the NHS, environmental experts, housing associations and local authorities in the development and evaluation of policies to improve child health and optimise children's use of health services.

2

[2] MacIntyre EA, Karr CJ, Koehoorn M, et al. Residential air pollution and otitis media during the first two years of life. *Epidemiology* 2011; 22: 81-9.
[3] Brugha R, Grigg J. Urban air pollution and respiratory infections. *Paediatr Respir Rev* 2014; 15: 194-9.
[4] Hjern A, Østergaard L, Norredam M, et al. Health policies for migrant children in Europe and Australia. *The Lancet*; 389: 249.

UK Statistics Authority

Section A
Project Details

A1 Legal gateways

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

Census data can be accessed by the approved researcher gateway in s.39 SRSA. Health data can be accessed under Section 251 of the Health and Social Care Act 2006. If patient demographics are to be used that will also be accessed under section 251 of the Health and Social Care Act 2006.

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

The project will be reviewed by the Health ResearchAuthority Confidentiality Advisory Group (CAG), an NHS Research Ethics Committee, and the ONS Microdata Release Panel (MRP). These applications are under development. We are seeking to update the longitudinal hospital data, by requesting HES data for years from 2015/16 onwards, for which we will apply to the Independent Group Advising on the Release of Data (IGARD) at NHS Digital. An application for ADRN approval for this project is also under development.

A3 Proposed site of research select all that apply

- ☑ ONS
- ☑ SRS
- ☐ 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:
This project focuses on child health, and we will examine health outcomes (hospital admissions and mortality) in children aged less than five years.

Justification for focusing on these subsections or groups:

This project focuses on hospital admissions in children. The risk factors for hospital admissions in children are different from those for adults. Children's health is strongly associated with the health and socio-economic status of their parents and their own birth characteristics (such as preterm birth and low birth weight). We therefore need to examine risk factors for hospital admissions in children separately from adults.

Note however that the information recorded at birth registration will be linked to mothers' and their live-in partners' Census records in order to examine health outcomes among children (See section A5 below). We will also use the mothers' hospital admission records to define some covariates of interest, such as delivery type (vaginal/Caesarean) and maternal comorbidities such as pre-eclampsia, and gestational diabetes.

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

Justification for datasets used and linkage requested

This project seeks to determine environmental, household and family risk factors for hospital admissions in children, with careful adjustment for multiple confounders to minimise bias. There are no aggregate datasets available that would allow such detailed analyses. We are requesting the use of the following datasets: **Hospital Episode Statistics, ONS birth registration, ONS death registration, NHS birth notification data.**

These four datasets have already been linked to create a birth cohort database by a team at City, University of London, led by Professor Alison Macfarlane. This linked birth cohort database is held in the Secure Research Service (SRS, formerly known as VML)

We are requesting to link this birth cohort database to **2011 Census data** for mothers and her resident partner to obtain information on overcrowding, housing type & tenure, type of heating (for project 1), and mother's and her partner's (if any) length of stay in the UK and knowledge of English (project 2). The link to Census data will also allow the use of individual-level variables to adjust more finely for sociopeconomic status (SES) including mothers' and her partner's education level, car ownership, and employment status. We will also request Census data on mothers' self-reported health status, to allow for better adjustment for maternal comorbidities.

Currently held birth cohort data

The currently existing birth cohort held at ONS in the SRS contains data on all births that occurred from1st January 2005 to 31st December 2014, with follow-up through longitudinal HES hospital records and ONS mortality records until 31st December 2015. Births are identified via ONS birth registration records linked to NHS birth notifications. The mother's and baby's longitudinal hospital records are linked via the birth notification (which has both the mother's and the baby's NHS number).

Patient identifiers such as mother's and baby's NHS numbers and postcode of residence are kept in an identifier file separately from the clinical and vital events data. They can be linked with a study ID (a random number unique to each baby and each mother in the cohort). The patient identifier file will only be accessed to extract variables for linkage, and not used for analysis.

Proposed data linkage between birth cohort and 2011 Census data

We propose to link births between 27th March 2010 and 27th March 2012 (that is, births up to one year before and one year after the 2011 Census) to data from 2011 Census questionnaires completed by the mothers in the cohort and their partners (or by the household reference person in the household in which the mothers live if this is not the mother or her partner). The proposed linkage is summarised below. A flowchart depicting the linkage process is provided in Appendix A:

1) Identifiers will be extracted by a member of the study team from the national birth cohort and transfer the file to the ONS Data Processing Team.

2) Identifiers will be extracted from a dataset of birth registrations held by the ONS Vital Statistics Output Branch in a separate area of the SRS. A member of the ONS Data Integration Team will do the data extraction and send the file to the ONS Data Processing Team. This step is necessary since linkage to Census requires mother's first and last name, and this is not available from the birth cohort files.

3) The ONS Data processing team will clean and encrypt the identifiers from the birth cohort file and the birth registration file, using the encryption algorithm described in the document

'ONS policy for safeguarding data whilst managing Admin Data Research Network projects[3]'. The files will be processed separately. The original (unencrypted identifiers) will be deleted after encryption.

4) The two files with encrypted identifiers will be transferred to the ONS Data Integration Team. The birth cohort file will first be linked to the birth registration file (to add encrypted identifiers based on mother's name), then to the 2011 Census using the encrypted identifiers including mother's name. The Data Integration Team hold a copy of the 2011 Census on their secure server, which includes encrypted identifiers and attribute data only (no actual identifiers). The Census household matrix will be used to extract data about the mother's live-in partner at the time of the Census.

5) The linked file will be transferred to the ONS Security and Permissions Team, who check the data, and assess the risk of disclosure, including granularity of variables and the output level. Data will only be released to the approved research team once the risk of disclosure has been minimised. The Census attribute data will be linked back to the birth cohort data via a record ID generated as a random number by the data processing team in step 3.

Proposed linkage to small area air pollution data:

The small area air pollution (AP) data is held by Dr Ai Milojevic at the London School of Hygiene and Tropical Medicine. The dataset contains annual average levels of 8 major air pollutants at 1 X 1 km grid resolution for the UK from 2003 to 2016. The grids have been mapped to full postcode. A further dataset contains data on distance to A roads and motorways for each full postcode in England. The AP data and the distance to road data do not contain information about individuals.

These datasets will be uploaded to the birth cohort area of the SRS and linked to the birth cohort via the mother's postcode at delivery. The linkage will be carried out by a researcher who is a member of the study team. To do the linkage, we will extract only the study ID and postcode from the identifier file of the birth cohort, which is held in a separate area of the SRS to the clinical data in the birth cohort. We will link the AP data to the postcode, then remove the postcode, and link the AP data to the birth cohort clinical data via the study ID (see Appendix B). The researcher will therefore not be able to see the clinical data, the postcode and the AP data simultaneously. The AP data will be converted to deciles of annual exposure before analysis to ensure that no individuals will be identified.

Proposed analyses:

With the enhanced birth cohort data, we will derive and validate a number of exposure variables (see Appendix A). We will calculate hospital admission rates (overall and diagnosis-specific) per 1000 child years according to age and each of the exposure variables.

To answer research question 1 (see project summary), we will examine the association between overcrowding, indoor air pollution and type of housing, and the risk of admission to hospital for respiratory infection using survival models (such as Cox proportional hazards regression or accelerated failure time models). Children will be censored at their first hospital admission for respiratory tract infection, death, or their second birthday, whichever occurs

[3]see: https://www.ons.gov.uk/file?uri=/aboutus/whatwedo/programmesandprojects/theadministrativedataresearchnetworkcollaboration/adrcsafeguardingpaper_tcm77-404473.pdf

first. We will adjust for potentially confounding variables (listed in Appendix C). We will also include the AP variables as covariates in the regression model. Multiple pollutants will be included simultaneously. Separate models will be fitted including distance to major road.

For research question 2 (see project summary), we will derive a categorical variable describing mother's country of birth, coded into United Nations population regions and length of stay in the UK. We will examine the association between this variable and the risk of total and diagnosis-specific hospital admissions, using similar methods as for research question 1. We will focus on common (>10,000 admissions/year) reasons for hospital admission that indicate access to health care services across the early life course:

a) neonatal feeding problems and jaundice, indicating access to care and advice provided by community midwives and health visitors shortly after birth

b) respiratory, urinary and gastrointestinal infections (in children <5 years) indicating ambulatory care sensitive conditions, potentially amenable to prompt primary care

c) multiple tooth extraction for dental caries (in children <5 years)– indicating access to public health and health visitor advice regarding healthy eating and dental care

We will also adjust for father's country of birth (coded into UK/non-UK born initially) and mother's and her partner's knowledge of English.

The main analyses will be carried out and presented at national (England) level. We will also carry out sub-national analyses for both research questions 1 and 2, examining differences between government office regions of England. For all analyses, we will adhere to strict disclosure control (see section B3) and no small numbers for tables or graphs will be extracted from the VML to ensure no individual can be identified. All outputs from the VML are assessed for disclosure risk by a trained employee of ONS independent to the study team.

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 episode Statistics for mothers and babies 2005-2014 (we will separately request an update from NHS Digital for data up to 2017) -NHS Birth Notifications 2005-2014 -ONS Birth and still birth registrations (2005-2014) -ONS death registrations for babies born between 2005-2014 Note - only a two year cohort around the Census 2011 will be linked to Census via ONS birth registration records. All births will be linked to air pollution data	

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 data	
Other *(please specify e.g. Ordinance Survey Address register in the relevant options adjacent)*	-Small area level data on air pollution and distance to major road (mapped to full post code)			

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 aim of this project is to identify household, parental and environmental risk factors for hospital admissions in children. We will also calculate population attributable risks for each of these factors, meaning that the relative contribution of, for example, overcrowded housing to emergency hospital admissions can be estimated. The results will allow the design and targeting of interventions, either through the NHS or other organisations (such as local authorities) to reduce the need for children to be admitted to hospital. This project therefore aims to improve the evidence base in order to save money for the NHS, reduce stress for families, and improve children's access to health services and health outcomes both in the short and long term. The specific beneficiaries of the project will be:

1) Public health departments in local authorities. This project will provide information on the relative importance of risk factors including housing quality, overcrowding, air pollution and health access barriers to hospital admissions in children, providing a sound evidence base for prioritising spending in order to improve child health. We will also carry out subnational analyses to examine whether risk factors for hospital admissions vary according to region within England. The project will also highlight which Census variables are most strongly associated with children's health status and need to be collected through other data sources when the Census is phased out.

2) The National Health Service (Clinical Commissioning Groups; CCGs): Part 2 of the project will examine whether all children have equal access to healthcare, independent of their parents' country of birth. If inequalities are found, this will provide evidence for CCGs to improve access to particular preventive health care services (this could include extending interpreting services or setting up outreach clinics).

3) Parents and civil society: This project will provide parents with information about the relative importance of housing quality, air pollution and health access barriers as risk factors for hospital admission in children. This evidence base can empower parents and civil society to push for improvements in local environments, housing quality, or healthcare access.

4) Other researchers: The enhanced mother-baby cohort created in this project will be a unique resource for research into maternal and child health research. We propose that the linked datasets created for this project should be kept securely at ONS, and that other researchers should be able to access the data for research purposes (given appropriate approvals, as discussed below). We will create metadata to go with the data resource. We will also make available R/Stata code used to analyse the data. Further, the methods used to link the data and the

potential to use modern study designs and statistical methodology (such as sibling control studies and causal mediation analyses) will also be of interest to researchers.

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)

As mentioned above, we anticipate that the linked datasets will be a rich resource for maternal and child health research, including research into environmental exposures during pregnancy and health at birth and early childhood.

The linked birth cohort dataset on its own is being made available to other researchers, subject to agreement with ONS, and City, University of London. The cohort database is held in the SRS. Access to the database for research, for purposes other than for which it was originally set up, also requires approvals by an NHS ethics committee, CAG, NHS Digital and the MRP.

We propose to also make the linked Census data available to other researchers. Access to the linked birth cohort and Census data for research purposes not covered in this application would, in addition to NHS ethics, CAG, NHS Digital and MRP approval, also require approval by the NSDEC.

The linked Census data would be held in a different project area to the rest of the birth cohort data, to ensure that only researchers with NSDEC approval can access them.

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 will provide secure data linkage and a secure environment in which approved researchers can access only the linked data relevant to their approved projects.

In particular, full account has been taken of the risks associated with:

- transfer of administrative data from data owner to ONS;
- 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 a secure environment; and
- secure access by researchers at the secure environment.

This is being done via the following controls:

- **Physical** – all linkage takes place behind high-security doors in a secure physical environment.
- **Procedural** – all data acquisition, import and export processes are subject to strict procedural controls.
- **Personnel** – only authorised employees holding Security Check clearance are permitted to enter the environment, and all access is recorded, monitored and audited by ONS Security Managers on a regular basis, through regular review of technical, procedural and CCTV records.
- **Technical** – the linkage facility is fully isolated from all other systems and networks. Within the environment, technical safeguards exist to ensure only authorised work can take place, and "unusual" activity is detected, assessed and acted upon. Electronic devices, (including mobile phones), software or connections are not permitted in the environment under any circumstances, and protective measures are in place to enforce this policy.

The linked attribute dataset will be protected at all times whilst in transit from the linkage facility to the SRS. Access to the data via the SRS is only possible through controlled rooms, located in ONS and other Government buildings. Before access is made available at any given location, it has to be approved by ONS.

Researchers will take no data into the room, and will take out no data. Results from the research will only be released by ONS staff once they have confirmed that they contain no risk of identifying an individual, household or business.

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.[4]

To summarise, research outputs created within the SRS are checked against threshold rules. When an output fails the threshold rule (for example a cell count less than 10) a discussion with the researcher takes place, giving the researcher an opportunity to demonstrate the output is not disclosive and therefore safe. By applying a principles based approach, rather than a black and white rule, context can be taken into consideration, however ultimately, it is the SRS statistical support team that make the final decision.

To limit the risk of re-identification we have introduced the following safeguards:

- we are compliant with government guidelines and best practice on assessing the risk of re-identification. The Government Communications Headquarters are content with the approach and controls in place;
- any future re-use of data will only take place following a completed risk assessment as described above; and
- any outputs will be subject to statistical disclosure controls and no aggregate outputs below the minimum cell count of 10 will be published;

The study is in the public interest so we propose to link the birth cohort and Census 2011 data securely within the safe data linkage environment at ONS. ONS will retain data for 12 months after delivering the dataset in case there is a requirement to carry out follow-up linkage work. The de-identified dataset will be retained according to the ADRN data retention policies

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

(https://adrn.ac.uk/media/174422/adrn-034_datareuseforresearchpurposes_02_00_pub.pdf) to ensure that the benefits of this research are realised. Within this time the dataset may be used for additional linkage subject to approval of data owners and ethical review of the subsequent projects.

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 encrypted identifiers 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.

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 for which approvals have been obtained by NSDEC, ADRN, CAG, NHS REC, NHS Digital and MRP.

No identifying variables will be made available to researchers and they will be kept separate from clinical data and Census responses. Variables from the Census and the national birth cohort will be categorised into broader categories to maintain confidentiality and prevent small counts in tables and figures.

All staff working on the data will be trained in using large, complex administrative databases and handling sensitive data (including NHS information governance training and Secure User Research Environment training). This training will be documented. All researchers working on the data in the SRS will also need to have ONS approved researcher status, which ensures they have experience in working 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.

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)**
University College London	UCL and ONS are both ADRC-E partner organisations. We will also apply for ADRN project approval.
City, University of London	Formal agreement between ONS and City, University of London to hold the birth cohort in the SRS. City has also applied to register the database with the ADRN. PH holds an honorary contract at City, University of London.
Office for National Statistics	
NHS Digital	Agreement between City, University of London, ONS and NHS Digital to hold the birth cohort data at SRS.
London School of Hygiene and Tropical Medicine	LSHTM is also an ADRC-E partner organisation.

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)

To date, we have consulted two public groups to examine the interest in the research and the acceptability of linking these data: the Great Ormond Street Hospital Parent and Carer advisory group (consisting of parents and carers of children who are frequent users of secondary care and other health services), and the Farr Institute Public Engagement Panel (consisting of persons with an interest in medical research based on electronic health records, many of whom have young children and/or grandchildren).

The responses we have had from these groups indicate broad support for the research: the groups consulted can see benefits to the public. They are generally supportive of data linkage, including between family members (as we are proposing), but also aware of the challenges in using these types of administrative data.

The main concerns raised were around a) ensuring data are kept safe at all times (which is why we propose linkage and analysis within the ONS secure data integration environment and the SRS) and b) informing the public their data are used for research in this way. To address b) we propose to develop a website (hosted by UCL) where we present information about the project, and a link to this website will be posted on the relevant section of the ONS website. The project will also be presented on the ADRN public website.

Our public engagement activities are ongoing. We are planning further public engagement activities, including with the National Childbirth Trust and the National Children's Bureau.

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)

As explained in section B7, have consulted various public groups about the use of linked health and Census data for this research.

We will publish the results of this research in peer-reviewed, open access journals. We will work with the ONS and UCL (the partner university) press offices to communicate the findings effectively to the public.

Further, we will develop a publically accessible website (hosted at UCL) where the methodology and the findings of these projects will be presented in an comprehensive manner.

We will participate in Research Open Days held at the UCL Great Ormond Street Institute of Child Health, and in public engagement events organised by the Administrative Data Research Centre for England (such as the Bloomsbury Festival) where we will present our research methodology and findings to the general public.

Section C
Responsible owner and applicant details

C1 Responsible Owner

Full Name: [redacted]	**Position:** [redacted]
Address: [redacted]	**Email:** [redacted]
	Telephone: [redacted]
	Organisation: [redacted]

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.

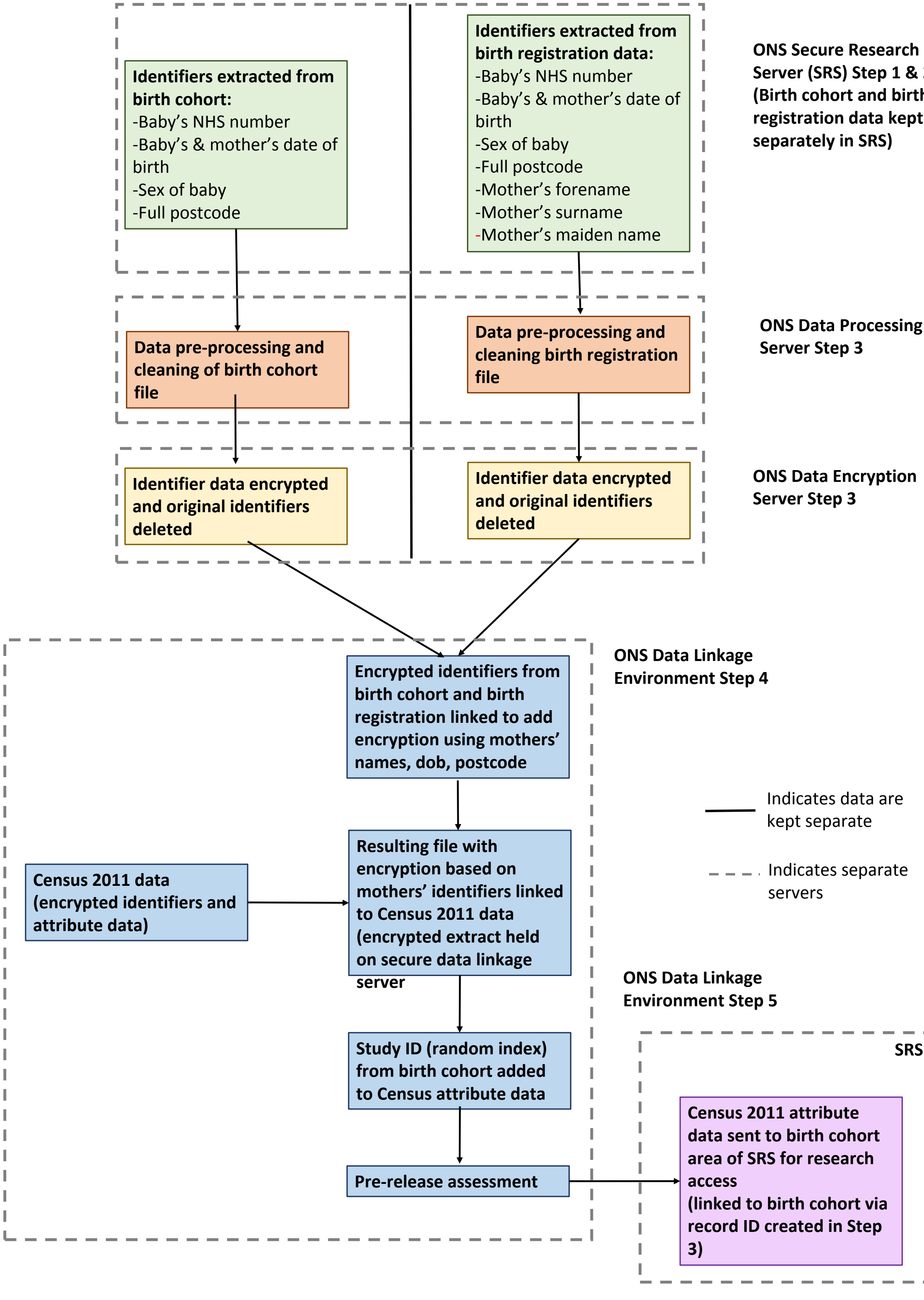
Signature: ---------- **Date:** ----------

UK Statistics Authority

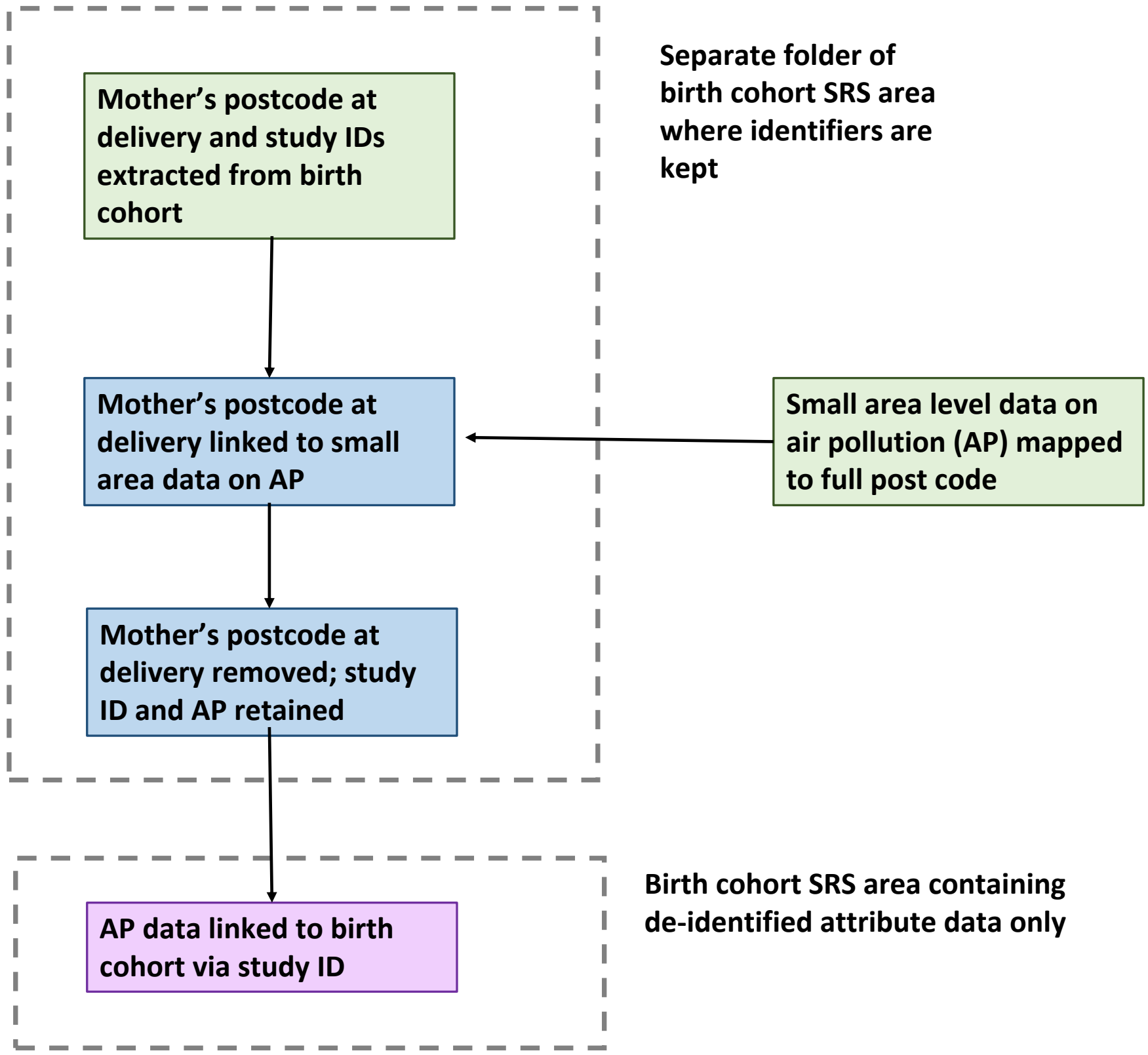
C2 Applicant Details (if applicant is not the responsible owner)

Full Name: [redacted]	**Position:** [redacted]
Address: [redacted]	**Email:** [redacted]
	Telephone: [redacted]
	Organisation: [redacted]

Appendix A: Data flows for linking Birth cohort to 2011 Census



Appendix B: Data flows for linking Birth cohort area-level air pollution (AP) data



Appendix C: Variables to be used in the proposed analyses

We will use the following variables in the linked data for the proposed analyses.

Outcome variables

Hospital admissions: total hospital admissions, total emergency hospital admissions, total planned admissions, hospital admissions for specific diagnoses: respiratory infections (research question 1) acute infections, neonatal feeding problems, multiple tooth extractions (research question 2)

Main exposure variables:

Housing: type of housing (flat/house/communal establishment), number of persons usually in household, number of bedrooms, number of rooms, central heating type, house ownership, landlord

AP exposure during pregnancy: 8 air pollutants measured at 1 X 1 km grids, or distance to major roads, mapped to mother's postcode at delivery

Parents' countries of birth: Mother's country of birth and father's country of birth (from birth registration, coded into WHO world regions), date of entry into the UK for mother and co-resident partner (from Census), mother's knowledge of English, co-resident partner's knowledge of English (from Census)

Confounding variables

Maternal socio economic status indicators: area-level deprivation (measured via the Index of Multiple Deprivation linked to the postcode at delivery in HES), mother's educational level, mother's National Statistics Socio-Economic Classification (NS-SEC), partner's educational level, partner's NS-SEC, access to car (all from Census)

Mother's partner socio economic status indicators: partner's educational level, partner's NS-SEC, (both from Census)

Birth/baby characteristics: baby's gender, month of birth, multiple birth status, birth weight (from ONS birth registration), presence of congenital anomalies or chronic conditions (from longitudinal HES data), mode of delivery (from mother's HES delivery record), gestational age (from NHS birth notifications), birth complications (from mother's and baby's hospital records), date of death (from ONS death records)

Mother's health: mother's chronic conditions (from mother's longitudinal hospital data), mother's self-reported health status, carer status, limited activities due to health problem (all from Census)

Mother's demographic characteristics: mother's marital status at delivery, joint or sole registration, father living at same address as mother indicator (if joint registration), mother's age at delivery (all from ONS birth records), mother's parity (from ONS birth records or derived from longitudinal HES data), mother's ethnic group (from Census).

Geographical variables: government office region (coded from maternal postcode at delivery)

1.	2.	3.	4.	5.	6.	7.	8.	9.	10	11.	**12**	13	14	15	16

UK Statistics Authority

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

Analysis of social and economic factors of severe mental disorders in Southeast London, UK

ADRN: Social and Economic Predictors of the severe Mental Disorders (SEP-MD study).

Start Date: June 2018 **End Date:** June 2021

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

King's College London/ South London & Maudsley Trust

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

Mental disorders can be debilitating and costly with a major adverse impact on an individual's social and occupational functioning. Attempts to characterise the association of socio-environmental risks with onset, course and outcomes of the mental disorders have been hampered by shorter follow-up, small samples, attrition and a lack of information on socio-environmental indicators. Having a clearer idea of socio-environmental risks could inform intervention development and contribute to a broader understanding of the social determinants of mental disorder course and outcomes. In addition, understanding the way in which broader social factors impact on service use (including in-patient admissions and bed days), involuntary admissions, prescribing and premature mortality, could help to inform service provision.

Previous research on mental health have relied on electronic health records, however a limitation is that data on socioeconomic status (e.g. education, occupation, tenure) and other important social indicators (migration status, household living circumstances, marital status and carers) and health states (e.g. self-rated assessment of disability and health) are not routinely collected. To address this, we propose a linkage of census data to Electronic Patient Records(EPRs) from a large mental health Trust (South London & Maudsley Trust(SLaM)). The SLaM EPR system, also known as 'CRIS' is a clinical database, covering a population of 1.3 million people in an ethnically diverse and geographically distinct region in south London. The linkage will be performed in the SRS (formerly known as the VML) and the data will only be accessible to approved researchers accessing the data in a secure ONS SRS location.

The linked dataset will not contain any identifiable information and there will be no data included on minors.

The findings of the research will allow an assessment of mental health inequalities and will help to inform the development of social and public health interventions and may lead to informing processes designed to eliminate mental health inequalities in service provision. The research on the linked dataset will be led by an ONS approved research team with expertise in epidemiology, public health and quantitative research methods, including advanced statistical methodologies and informatics. The team also comprise experienced senior clinicians with expertise in mental health.

UK Statistics Authority

Section A
Project Details

A1 Legal gateways

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

Census data can be accessed by the approved researcher gateway in s.39 SRSA. Health data can be accessed under Section 251 of the Health and Social Care Act 2006. If patient demographic data is to be used that will also be accessed under section 251 of the National Health Service Act 2006.

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

1. **The project has been submitted to the Confidentiality Advisory Group (CAG) full meeting on 7th December 2017.**
2. **The project has also been submitted to Oxford C Research Ethics Committee for ethics approval for a research database- meeting date is 26th Jan 2018**

A3 Proposed site of research select all that apply

- ☐ ONS
- ☑ SRS
- ☐ 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:

The study will focus on adults aged 15 or older and will not make any other restrictions.

Justification for focusing on these subsections or groups:

This is because the study focuses on severe mental disorders in adults, which has a different evidence base compared to children and young people

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

Background: There are longstanding controversies around incidence, course and outcome of severe mental illnesses(SMI) such as schizophrenia[8-10] and in particular the role of the social environment in predicting onset and course/ outcomes of SMI. Traditional electronic patient record systems lack detail on individual-level social disadvantage indicators for SMI, such as social isolation, poverty, household relationships and migration/citizenship, which has hampered previous enquiry.

This project will lead to the establishment of an anonymised electronic dataset of individuals with SMI, matched to non-SMI population-based controls. The proposed linkage will bring together information from census with clinical information from the mental health Trust to inform understanding of how the social environment impacts on onset, course and outcomes of SMI. The linked dataset will provide an evidence base to clinical commissioning groups, public health professionals and other practicing clinicans and to policy makers. In addition it will be possible to conduct research using statistical methods to address the following research aims:

Aims: To assess the role of a range of social factors (household poverty, social isolation, family support, neighbourhood social fragmentation) and ethnicity/ migration status in predicting:

1. Onset of severe mental illness (study 1).
   Public benefit: the analyses will help to identify modifiable risk factors in the

environment and the interplay of these in increasing the risk of severe mental illness. This will help to inform the development of interventions and inform service provision and commissioning within the local area
2. Hospital admissions/ involuntary admissions, length of stay and premature mortality in SMI, including FEP (First Episode Psychosis) (study 2).
3. Public benefit: The analyses will help to identify the interplay of social factors with severe mental illness in predicting poorer course and outcomes. This will inform the development of interventions and treatments to improve the course of illness. The research will also inform the provision of services and the targeting of treatments to the most vulnerable groups with severe mental illnesses.

Methods:
Setting and participants: South London & Maudsley(SLaM) Trust provides mental healthcare to a catchment of 1.2 million people[1] and operates a fully electronic health records system[1]. The Clinical Record Interactive Search(CRIS) system enables access to anonymised electronic patient records for secondary analysis from SLaM Trust, and has full ethical approvals[1]. CRIS will be linked to individual-level census records containing sociodemographic indicators on people resident in the UK, on census days in 2001 and 2011, leading to a longitudinal cohort of individuals with an SMI diagnosis followed to 2018 (study 2). The cohort will include individuals with severe mental illnesses from the age of 15 onwards. The age cut-off is based on the observation that the severe mental illness may occur before the age of 18. Census data from random population-based controls will also be added (for case-control study (study 1)). A further linkage to death certificates[3] will be included.

Study design: To address aim 1 we will use a case-control design. We will identify randomly sampled controls for identified 'cases'. Cases will be defined as individuals with an SMI diagnosis, controls will be individuals resident in the catchment area of the Mental Health Trust, not known to the Trust. The case-control design will permit assessment of association of exposures (e.g. household poverty, social isolation, family support, neighbourhood social fragmentation and ethnicity/ migration status) (derived from census records) with outcomes (onset of SMI). To address aim 2 we will create a cohort of individuals with SMI and will assess the association of social and clinical exposures with outcomes (in-patient admissions, use of the Mental Health Act, longer in-patient stays, mortality).

Linkage methodology:
The linkage will be conducted by the ONS. This will be through a matching methodology developed with the ONS using Patient Identifiable Information (PII) such as name, age, sex and address. Acceptable levels of matching will be agreed. The personal identifiers will then be stripped out of the dataset, leading to a final de-identified linked dataset, for analysis. The ONS approved research team will then have access to this anonymised dataset in order to conduct the research. See Section B4 below for further detail.

Selection of control groups and sample groups :
South London & Maudsley Trust provides near-monopoly mental healthcare to a well defined geographical catchment area of 1.34 million people in south London. The Trust operates fully electronic health records. Our sample group of people with severe mental illnesses will be selected from this clinical data resource using diagnoses according to standardised criteria (International Classification of Mental Disorders-10) for the severe mental disorders. The control group will be selected at random from the surrounding geographical area and will be individuals not known to the mental health provider.

Statistical methods: Comparison between the SMI cohort and population-controls will be made across social disadvantage indicators using statistical methods across two census

time-points (2001,2011).

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)*				
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 Census 2001		Census 2011 Census 2001	

Other *(please specify e.g. Ordinance Survey Address register in the relevant options adjacent)*	CRIS (clinical data from SLaM Mental Health Trust)		CRIS (clinical data from SLaM Mental Health Trust)	

UK Statistics Authority

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 are a number of concerns relating to onset and adverse outcomes in people with severe mental disorders (SMD) which the database linkage and the proposed research will aim to address. It is known that people with SMD have a life expectancy 15-20 years less than the general population and most causes of death are from preventable physical causes[1,2]. Certain individuals with SMD are more likely to be admitted to hospital and experience long admissions however the social factors driving these differences are poorly understood. In addition, urban living[3,4,6], being of an ethnic minority background[3,4] and adversity are associated with an increased risk of SMD. Employment can be a good indicator of recovery in people with SMD, however there is little data on the quality of employment which people with SMD achieve and how this correlates with disability. There are also a number concerns relating to ethnic inequalities and the SMDs. For example, certain Black Asian and Minority Ethnic (BAME) groups with SMDs are more likely to be admitted involuntarily to hospital than White British people. It is known that Black people with SMDs are more likely to experience social exclusion, relative to White British people with SMDs. To date, studies have been restricted by smaller sample sizes, difficulties in recruiting people from underserved populations and a paucity of information on important social indicators.

The linkage will lead to the following benefits:

1. Clarifying which factors in the urban environment are associated with an increased risk/ adverse outcomes in the SMDs. This could inform social and public health interventions in urban settings to reduce the risk of developing SMDs and improve outcomes.
2. Clarify whether migration status and factors such as household poverty and social isolation, play a role in increasing the risk of SMDs in ethnic minority groups. This will inform the development of interventions for these underserved populations.
3. Permit assessment of the association of SMDs with mortality risk, taking into account a range of individual-level confounders such as education, tenure and occupation. This will inform how far health-related behaviours (e.g. smoking, diet) and clinical risks (antipsychotic medication prescribing) with baseline disability account for mortality in these populations independent to socioeconomic status. This will inform the development of interventions to target premature mortality in people with SMD.
4. Clarify the role of social factors (e.g. social isolation, household poverty, migration status) alongside self-rated disability, in predicting admissions to psychiatric hospitals, especially in BAME groups relative to White British people with SMD. The proposed analyses will inform policy/legislation reviews, for example inform public health policy and anticipated updates to Mental Health Act legislation.
5. Clarify the role of social exclusion/ employment status in recovery in the SMDs. This will inform the development/targeting of social/employment-based interventions for people with SMD.
6. The proposed analyses will allow a detailed assessment of disparities in health service use and inform service provision/commissioning.

References
1. **Das-Munshi J,** Chang CK, Dutta R, Morgan C, Nazroo J, Stewart R, Prince MJ. Ethnicity and excess mortality in severe mental illness: a cohort study. *Lancet Psychiatry.* 2017

DOI: http://dx.doi.org/10.1016/S2215-0366(17)30097-4
2. Chang CK, Hayes RD, Perera G, Broadbent MTM, Fernandes AC, et al. (2011) Life Expectancy at Birth for People with Serious Mental Illness and Other Major Disorders from a Secondary Mental Health Care Case Register in London. PLOS ONE 6(5): e19590.
3. Schofield, P., Thygesen, M., **Das-Munshi, J.,** Becares, L., Cantor-Graae, E., Pedersen, C. & Agerbo, E. Ethnic density, urbanicity and psychosis risk for migrant groups – A population cohort study *Schizophrenia Research 2017. In press.* *http://www.schres-journal.com/article/S0920-9964(17)30170-6/abstract*
4. Vassos E Pedersen CB Murray RM Collier DA Lewis CM . Meta-analysis of the association of urbanicity with schizophrenia. Schizophr Bull . 2012; 38: 1118–1123.
5. Fearon, P., Kirkbride, J., Morgan, C., et al (2006). Incidence of schizophrenia and other psychoses in ethnic minority groups: Results from the MRC AESOP Study. Psychological Medicine, 36(11), 1541-1550.
6. James B. Kirkbride, Paul Fearon, Craig Morgan, Paola Dazzan, Kevin Morgan, Jane Tarrant, Tuhina Lloyd, John Holloway, Gerard Hutchinson, Julian P. Leff, Rosemarie M. Mallett, Glynn L. Harrison, Robin M. Murray, Peter B. Jones. Heterogeneity in Incidence Rates of Schizophrenia and Other Psychotic Syndromes. Findings From the 3-Center ÆSOP Study. *Arch Gen Psychiatry.* 2006;63(3):250–258. doi:10.1001/archpsyc.63.3.250

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)

There will be a linked dataset (clinical data linked to census) as a result of the research, which will only be accessible to approved researchers accessing the data in a secure ONS SRS location. The data owners will jointly comprise the ONS and South London & Maudsley Trust. In future if other researchers wish to access the data for analyses they will have to lodge separate applications to the data owners. All researchers will need full approvals to access the data.

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 will provide secure data linkage and a secure environment in which approved researchers can access only the linked data relevant to their approved projects.

In particular, full account has been taken of the risks associated with:

- transfer of administrative data from data owner to ONS;
- 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 a secure environment; and
- secure access by researchers at the secure environment.

This is being done via the following controls:

- **Physical** – all linkage takes place behind high-security doors in a secure physical environment.
- **Procedural** – all data acquisition, import and export processes are subject to strict procedural controls.
- **Personnel** – only authorised employees holding Security Check clearance are permitted to enter the environment, and all access is recorded, monitored and audited by ONS Security Managers on a regular basis, through regular review of technical, procedural and CCTV records.
- **Technical** – the linkage facility is fully isolated from all other systems and networks. Within the environment, technical safeguards exist to ensure only authorised work can take place, and "unusual" activity is detected, assessed and acted upon. Electronic devices, (including mobile phones), software or connections are not permitted in the environment under any circumstances, and protective measures are in place to enforce this policy.

The linked attribute dataset will be protected at all times whilst in transit from the linkage facility to the VML. Access to the data via the VML is only possible through controlled rooms, located in ONS and other Government buildings. Before access is made available at any given location, it has to be approved by ONS.

Researchers will take no data into the room, and will take out no data. Results from the research will only be released by ONS staff once they have confirmed that they contain no risk of identifying an individual, household or business.

Rather than operating a traditional 'rules based' approach to output checking, the VML operates a 'principles based approach'. This forms part of the "5 safes approach" to data access adopted by the VML.https://blog.ons.gov.uk/2017/01/27/the-five-safes-data-privacy-at-ons/

To summarise, research outputs created within the VML are checked against threshold rules. When an output fails the threshold rule (for example a cell count less than 10) a discussion with the researcher takes place, giving the researcher an opportunity to demonstrate the output is not disclosive and therefore safe. By applying a principles based approach, rather than a black and white rule, context can be taken into consideration, however ultimately, it is the VML statistical support team that make the final decision.

To limit the risk of re-identification we have introduced the following safeguards:

- we are compliant with government guidelines and best practice on assessing the risk of re-identification. The Government Communications Headquarters are content with the approach and controls in place;
- any future re-use of data will only take place following a completed risk assessment as described above; and
- any outputs will be subject to statistical disclosure controls and no aggregate outputs below the minimum cell count of 10 will be published;

ONS will retain data for 12 months after delivering the dataset in case there is a requirement to carry out follow-up linkage work. The linked dataset will be retained for the entire period of the research (3 years) with the prospect of a reasonable extension to account for unforeseen delays.

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)

There are a number of risks which we will aim to minimise:

1. **Quality of linkage**
   As this is a linkage of health data with non-health data there will not be a common identification number (eg. NHS ID), however a matching methodology will be developed with the ONS using Patient Identifiable Information (PII) and acceptable levels of matching will be agreed.

2. **Risks of identification**
   Identifiers will be used by the Trusted Third Party (the ONS) to undertake the linkage. These identifiers will be held separate to the attribute data and will not be accessible to the researcher. Data will be processed (i.e linked and assembled) in a separate (virtual) secure area to the secure area where the researchers will be analysing the data. All staff handling the data are required to have security clearance at a level appropriate for the data they are going to handle as per ADRN Secure Environment Policy (ADRN 032) . Adherence to data confidentiality is also part of the contract of any staff at ONS. In addition, staff listed in the data sharing agreement and with access to the data are required to sign a declaration that they understand these restrictions.

3. **Privacy Impact Assessment**
   The Privacy impact of this project has been assessed at two stages, firstly by the ADRN and subsequently by the ADRN Approvals Panel. The ADRN has considered the information on the project proposal (data sources linked, types of dataset and variables requested, level of geography, sensitivity of the variables requested and topic in study). It then carried out an evaluation of the risk of an individual being identified or information being disclosed about them from the data or from the project's output. A Privacy impact assessment report was then produced and presented to the ADRN approvals panel who considered the application and reviewed the ADRN proposal and the Privacy Impact Assessment. The Panel is responsible for deciding if they are confident that any potential privacy issues have been assessed, reported appropriately and addressed satisfactorily. I will consider whether the potential public benefit offered by the project outweighs the potential privacy impact.

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)

As a requirement of the NHS HRA for this project, We have lodged applications for ethical approvals for a new database The Research Ethics Committee (REC) meeting for this application has been booked to take place on 26th Jan 2018. We are requesting NSDEC ethical approval for the proposed use of the data.

Census data can accessed by the approved researcher gateway in s.39 Statistics and Registration Service Act 2007. Health data can be accessed under Section 251 of the National Health Service Act 2006. If patient demographic data is to be used that will also be accessed under section 251 of the National Health Service Act 2006.

The Statistics and Registration Services Act and the Approved Researchers Scheme will be the legal gateway to get the data into the Secure Research Service.

As the data linkage will involve access to confidential patient information without consent in England and Wales we have also applied to the Confidentiality Advisory Group (CAG) (Health Research Authority; HRA) for ethical approval of the clinical research.

We of course understand that NSDEC approvals will be contingent on favourable outcomes of the applications to the REC and the CAG.

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)**
South London & Maudlsey Trust (SLAM Trust)	ONS, SLAM Trust are data suppliers. ONS will provide the Trusted Third Party data linkage service on behalf of the ADRN. King's College London is the sponsor of this research and researchers from King's College London will conduct the research.
Kings College London	
Office for National Statistics	
Administrative Data Research Network	

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)

There has been much public engagement with respect to the Clinical Record Interactive Search (CRIS) system (the clinical records system from South London & Maudsley Trust (SLaM) which we are proposing to link to ONS census records). This includes:

- Dedicated service user and carer involvement in the development of the CRIS System, which informed the CRIS build in 2007/08 in parallel with the technical developments. The structures for patient-led oversight of CRIS were described for the original ethics approval in 2008 and have been maintained since then.

- We have developed the BRC Data Linkage Service User and Carer Advisory Group. The group is a regular meeting of people with lived experience of mental illness, all of whom have an interest in mental health research involving data linkage. The group meets regularly. The current project was presented to the group on 28th September 2017. There was group-wide acknowledgement of the importance of the project and this application has been developed to reflect the views expressed at the meeting.

- SLaM service users are also made aware of the CRIS resource and that health records data are being used for research. This has included the dissemination of the CRIS information leaflet and website content, 'road shows', communications with individual service-user and carer-representative groups, and online dissemination. These inform services users of the possibility that CRIS data may be linked with other data sources for secondary research purposes. These communication methods mention that service users may opt out at any time.

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 findings of the study will be widely disseminated through the following means:

1. Publication in academic journals which are open access
2. Presentation at departmental meetings, conferences and invited presentations
3. Dissemination to local commissioners, health care providers and clinicians through reports, presentations and newsletters
4. Through institutional websites (eg. through Kings Health Partners and the Administrative Data Research Network (ADRN))), blogs, social media
5. To service user groups through presentations and newsletters
6. For high impact findings through press releases developed in conjunction with funders and the university press relations team
7. To non-academic partners/ charities with an interest in the findings, eg. Public Health England

UK Statistics Authority

Section C
Responsible owner and applicant details

C1 Responsible Owner

Full Name: [redacted]

Position: [redacted]

Address: [redacted]

Email: [redacted]

Telephone: [redacted]

Organisation: [redacted]

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:** ----------

UK Statistics Authority

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

Full Name: [redacted]	**Position:** [redacted]
Address: [redacted]	**Email:** [redacted]
	Telephone: [redacted]
	Organisation: [redacted]

UK Statistics Authority

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

Project Title
Please provide a title indicative of the project

Transport Model Development for West Berkshire Council

Start Date: 1-Dec-17 **End Date:** 1-Feb-18

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

West Berkshire Council, Market Street, Newbury, RG14 5LD

Project Summary
Please provide a brief high level summary of the research giving necessary background

The Local Plan[1] is part of the overall Development Plan for West Berkshire, setting out the local planning policies for the Council. Part of that plan is the Local Transport Plan[2] (LTP) used to set the framework for the delivery of all aspects of transport and travel for West Berkshire.

The council has commissioned WSP, a global engineering consultancy group, to carry out a project to analyse trip patterns in the West Berkshire area using 2011 Census travel to work data in the Secure Research Service (SRS). The data will be used to construct trip matrices, which contain information about the travel demand between each zone of the West Berkshire Transport Model. Once completed the West Berkshire Transport Model will act as an evidence base to inform development and transport decision making in the local authority, and assess the Local Transport Plan policies. These will include:

- junction improvements
- parking strategies
- helping to make the best use of the roads

[1] http://info.westberks.gov.uk/localplan
[2] http://info.westberks.gov.uk/CHttpHandler.ashx?id=36911&p=0

- assessing the impact of proposed developments on the transport network

As a Local Highway Authority, the Council is required to produce an LTP that contains policies (that can be set out in a strategy or vision) and implementation plans. This is a statutory requirement introduced by the Transport Act 2000 and amended by the Transport Act 2008.

The updated LTP will provide an improved evidence base for public policy decision making and service delivery, with the aim of benefiting the local economy, society and quality of life for people in the West Berkshire area.

WSP Ltd will use de-identified 2011 Census travel to work data in the SRS as these data have detailed characteristics tables containing information on the origin and destination of trips, and the mode of transport used. WSP have a history of previously carried out this work using ONS census data for local authorities in 2015 for Wokingham Borough, Northamptonshire, West Sussex and Suffolk, and in 2016 for Dover.

Census data will be supplemented with data from other sources to understand the trip patterns in the area and produce a trip matrix. Specifically, WSP has commissioned a West Berkshire traffic survey and aggregated GPS data from mobile phone networks has been provided to WSP by Citi Logik. According to Regulation 14 of The Privacy and Electronic Communications Regulations 2003, phone user location data may be used without the consent of the user provided that the user cannot be identified from the data. None of this ancillary data will be imported into the SRS therefore there is no risk of it becoming identifiable through linking with census data.

The transport model's development will be described in documents published on the West Berkshire Council Website. In particular, under the Department for transport's WebTAG best practice guide for the conduct of transport studies (which is mandatory for projects or studies that require government approval) a Local Model Validation Report will be published. Among other things, this will detail the development of the trip matrix and describe the data used in its construction, ensuring that use of data will be transparent to the public.

Results extracted from the transport model (for example delay, vehicle flow etc.) will be included in these documents. However the de-identified Census data used to produce the trip matrix will not be published.

UK Statistics Authority

Section A
Project Details

A1 Legal gateways

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

The ONS Approved Researcher scheme is the legal gateway being used to access the data. This is in compliance with the Statistics and Registration Services Act 2007.

The ONS Microdata Release Panel (MRP) approved the proposal on the grounds that there was a legal gateway to access the data, it was deemed appropriate use of ONS data and a public benefit was demonstrated.

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
- ☐ 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:

Population aged 16 and over in employment (including students)

Justification for focusing on these subsections or groups:

This project is focussed on Travel to Work data, and the dataset is not applicable for other subsets of the population. .

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

The Census dataset requested is WM03EW at Output Area (OA) geography (Location of usual residence and place of work by method of travel to work by sex by student employment status (with 'outside UK' detailed)). This set contains information about the number of trips made via a number of transport methods, and at the most detailed level of geography. It covers all usual residents aged 16 and over in employment the week before the census. ;This will enable the most detailed understanding of travel patterns possible, so that traffic conditions can be represented accurately in the transport model.

The West Berkshire Transport Model distinguishes between internal and external travel zones (i.e.: zones within and zones outside of West Berkshire). This project is only concerned with internal zones. Accordingly, only data for the Census's Output Areas and Work Place Zones within these internal zones is required. Spatial analysis will be carried out using ArcMap to relate the Census-defined Work Place Travel Zones to the West Berkshire Transport Model zones. This work will be performed by the researchers in the SRS.

Output data will be aggregates of journeys between each transport model zone and will be formatted as origin/destination matrices. As aggregates, the outputs will contain no personally identifiable data and will be subject to statistical disclosure controls.

Outside of the SRS, these matrices can be supplemented with those generated from traffic surveys and mobile phone location data to obtain a more complete understanding of trip patterns in the West Berkshire area than was possible for the previous LTP. This is expected

to lead to more accurate results and better-informed decisions, leading to more efficient use of public resources.

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)*				
Big Data *(please specify e.g. Twitter data, smart meters and mobile phones, in the relevant options adjacent)*	Mobile phone location data (Not imported into SRS)			
Survey Data *(please specify e.g.LFS, BRES, etc in the relevant options adjacent)*	Traffic survey data (Not imported into SRS)			
Census Data *(please specify year, e.g. Census 2011 in the relevant options adjacent)*			Census 2011 'WM03EW at Output Area geography'	
Other *(please specify e.g. Ordinance Survey Address register in the relevant options adjacent)*				

Section B
Assessment against NSDEC ethical principles

B1 **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 will be a demonstrable public benefit from the research as the West Berkshire Transport Model will provide an evidence base for both policy and decision making in the West Berkshire area, and will be used to assess the Local Transport Plan policies.

The council's priorities are maximising people's access to services and the economic vitality, while promoting safety and sustainability. Increasing travel choice is a key goal; promotion and support of walking, cycling and public transport options contributes towards sustainability and minimises environmental harm and congestion. This transport model will help West Berkshire Council to build an evidence-based strategy to meet these goals.

The previous West Berkshire LTP was used to aid in the following projects/achievements:

- 100% of West Berkshire Schools developing their own transport plans. The implementation of these School Travel Plans has seen a significant reduction in the use of the car for the journey to school and an increase in active travel. In 2010 the achievements were demonstrated by 64% of journeys to school being made on foot, by bike, by bus or by rail.
- A consistent reduction of people killed and seriously injured on West Berkshire roads achieving the 2010 national reduction target
- The development of a Rights of Way Improvement Plan
- The development of a preferred freight route network for goods travelling through and within West Berkshire
- Extended highway maintenance schemes across the district have been delivered to ensure that the district's road network is in safe working order
- Extensive improvement works (considering all users) along the A4 corridor between Newbury and Theale
- New crossings delivered for pedestrians and cyclists to improve safety and accessibility
- Junction improvements - Wash Water/A343 junction which was redesigned to improve visibility and safety
- Bridge strengthening such as at Alfreys Bridge, Grazeley formed part of a wider bridge programme aimed at improving safety and access. The replacement of a number of rural footbridges across West Berkshire have also helped to improve vital rural links
- Traffic calming/management schemes - Hungerford Primary School, Church Gate/Green Lane, Thatcham and Stroud Green, Newbury
- Partnership working has increased and developed in order to deliver better results in terms of quality and value. Some examples of partnership working are the Cycle Forum, Berkshire Strategic Transport Forum, developing relationships with the health sector, the Highways Agency, Thames Valley Safer Roads Partnership, working with First Great Western on access projects and the Berkshire School Travel Excellence Programme (STEP)

B2 **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)

Access to data will only take place within the secure ONS Secure Research Service (SRS) environment and all outputs will be checked by ONS prior to release to ensure disclosure control and the confidentiality of data subjects is protected. All analysis and use of the data will be at a secure SRS setting at one of the ONS offices. The researcher is accredited as ONS Approved Researcher. ONS will have sight of the final report.

B3 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 are being used. The research methods employed will be openly available for further scrutiny or replication of results.

B4 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)

Access to the potentially disclosive data will be in the Secure Research Service environment and via an approved legal gateway (Approved Researcher scheme). 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.

B5 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)**
West Berkshire Council	Contract between West Berkshire Council and WSP Group.
Citi Logik	Contract. Privacy and Electronic Communications Regulations 2003

B6 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)

Whilst the views of the public have not been sought with regards to the research, there is a clear public benefit for the analysis (as described in B1 above). A public consultation on the Approved Researcher scheme in 2015/16 recommended that commercial organisations should be allowed to access ONS research data where there is a clear public benefit.

B7 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)

West Berkshire Council publish their LTP on their website at :
http://info.westberks.gov.uk/CHttpHandler.ashx?id=36911&p=0

ONS will include a link to the published report from the Approved Researcher web pages and we will work with WSP and West Berkshire Council to publish a detailed case study showcasing the impact of the research.

B8 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)

None.

UK Statistics Authority

Section C
Responsible owner and applicant details

C1 Responsible Owner

Full Name: [redacted]	**Position:** [redacted]
Address: [redacted]	**Email:** [redacted]
	Telephone: [redacted]
	Organisation: [redacted]

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: [redacted]	**Position:** [redacted]
Address: [redacted]	**Email:** [redacted]
	Telephone: [redacted]
	Organisation: [redacted]

Ethical considerations for ONS acquisition of NHS Digital Data

Oral report

Jonny Tinsley

1.	2.	3.	4.	5.	6.	7.	8.	9.	10	11.	12	13	14	**15**	16

Any other business

1.	2.	3.	4.	5.	6.	7.	8.	9.	10	11.	12	13	14	15	**16**