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

Minute, Agenda and Papers

Wednesday 27 January 2016

10:30 – 15:00

Board Room, UK Statistics Authority
Drummond Gate, London

National Statistician’s Data Ethics Advisory Committee

Minute

Wednesday, 27 January 2016
Boardroom, Drummond Gate, London

Present
Board Members
Mr Robert Bumpstead (Deputy Chair)
Mr Colin Godbold
Ms Annie Hitchman
Dr Dean Machin
Mr Neil McIvor
Ms Isabel Nisbet
Ms Marion Oswald
Mr Osama Rahman
Mr Hetan Shah

UK Statistics Authority
Mr Adil Deedat
Dr Simon Whitworth
Ms Natalie Shorten
Mr Ross Young (for item 3)

Office for National Statistics
Mr Owen Abbot (for item 4)
Mr Alistair Calder (for item 5)
Mr Nick Stripe (for item 6)
Mr Mark Gautrey (for item 6)
Ms Helen Colvin (for item 6)
Ms Jane Naylor (for item 7)
Mr Thomas Smith (for Item 7)

Apologies:
Mr Ian Cope (Chair)
Professor Martin Severs

1. Minutes and matters arising from the previous meeting

1.1 In the absence of the Chair, Mr Ian Cope, the meeting was chaired by Deputy Chair, Mr Robert Bumpstead.

1.2 The Deputy Chair informed the committee that the minute of the second meeting had been agreed and signed off by correspondence. The minute, agenda and papers from the last meeting are now published on the UK Statistics Authority website.

1.3 The Deputy Chair welcomed Mr Hetan Shah to his first meeting of the committee.

1.4 Progress with actions from the previous meeting held on 14 October were reviewed. Most actions were complete or on the agenda for further discussions. The secretariat informed the meeting that the action relating to appeals against NSDEC decisions for ADRN projects would be taken forward with the Chair of the ADRN Board.

Action: Secretariat to include the minute from the previous meeting in the Board pack in addition to the link currently provided

2. Chair’s report

2.1 The Deputy Chair informed members that the ethical principles for NSDEC had been provided to the Authority Board, as agreed in the last meeting on 14 October. The principles, which are owned by the National Statistician, are now published on the UK Statistics Authority’s web pages.

2.2 It was agreed that the meeting on 19 April will take place at the ONS Titchfield office. The Secretariat will contact members in due course to confirm start and end times for the meeting. The site visit will also include a tour of the Census Transformation Programme and a presentation on data shares.

2.3 The meeting was of informed of the interim findings from Professor Sir Charles Bean’s review of economic statistics. These findings included ensuring appropriate ethical scrutiny and safeguards around data shares. The final report will be published on March 16.

2.4 Members considered recent media coverage of big data, ethics and statistics. The committee agreed that it would be useful to see a coordinated programme of communication activity at the next meeting.

2.5 Ms Oswald informed the meeting that she will be organising a conference on trust, risk, information and the law at the University of Winchester on 27 April. Speakers include Renate Samson from Big Brother Watch and Sir David Omand former Director of Government Communications Headquarters (GCHQ). Ms Oswald also invited the Chair of NSDEC to present at this conference, as an opportunity to highlight the role of the committee.

2.6 The Deputy Chair concluded his report by updating the committee on current and newly proposed data shares.

2.7 It was agreed that more information relating to the data shares be provided at future meetings, using less jargon.

Action: Secretariat to circulate interim report from Professor Sir Charles Bean

Action: Secretariat to confirm with members timings for the meeting on 19 April

Action: Secretariat to discuss with the Chair on his return presenting at the trust, risk, information and the law conference

Action: Secretariat to present co-ordinated programme of communication activity at the next meeting

3. Developments in data sharing

3.1 The Deputy Chair introduced Mr Ross Young, who leads on Data Access Policy and Legislation at the UK Statistics Authority. Mr Young provided members with an overview of the Authority’s objectives in improving and expediting access to data for research and statistical purposes and contrasted these with obstacles and challenges when using the current legislation.

3.2 Mr Young provided NSDEC with an overview of progress with proposals for the new legislation for access to data for statistical and research purposes.

3.3 It was suggested that given current powers within existing legislation care should taken in framing the legislation as new. Members considered different aspects of the legislation and issues including accessing data from government and business.

3.4 The proposal for ethical oversight of data shares was welcomed by the committee. It was suggested that ethical considerations should also be given in developing clauses as well as when implementing them.

4. Revised: Estimating ethnicity from names [NSDEC(16)01]

4.1 The Deputy Chair summarised the discussions held on the "Estimating ethnicity from names" at the last meeting on 14 October and reminded members of the clarifications sought:

- i. the intellectual property rights for the tool and who, following further development, would be able to make use of it;
- ii. the measures University College London (UCL) and users of the tools have taken to ensure they have satisfied the requirements of the Data Protection Act around processing of sensitive personal data and other legal requirements; and
- iii. ongoing use of the tool.

4.2 Mr Abbott informed the meeting that further to discussions with UCL, ONS would now own the intellectual property rights for the tool. ONS would not be providing any identifiable data and would therefore be compliant with requirements of the Data Protection Act.

4.3 The meeting heard that access to the tool would require registering for use and providing information about the proposed use. UCL would also be provided with a perpetual license.

4.4 NSDEC was also informed of relevant disclaimers highlighting to users that once downloaded the tool should not be passed to other users.

4.5 The committee welcomed changes provided in light of previous requests for clarifications. Members, however, remained unsatisfied with the proposed mode of access and suggested that, under these proposals, the tool might be transferred to a third party who hadn't registered to use the tool.

4.6 In addition, members discussed the legal and ethical concerns about information generated by the tool and the appropriateness of use of ONS data for the development of a tool which generates potentially sensitive data about individuals.

4.7 It was agreed that the National Statistician be advised that, whilst there are good uses to be made of the tool, with the proposed access arrangements, there is a high risk to the organisation.

4.8 Members suggested that the mode by which a tool is provided is reconsidered with an alternative provided to NSDEC.

Action: Secretariat to work with Mr Abbott to propose alternative arrangements, subject to the National Statistician's views

5. Address register

5.1 The Deputy Chair introduced Mr Alistair Calder, from Population Methodology and Statistical Infrastructure Division in ONS. Members heard that this item was for information but that Mr Calder would return with specific applications in due course.

5.2 Mr Calder provided NSDEC with an overview of ONS's plans for an address register, which is integral to the success of the 2021 Census. The presentation highlighted the complexities in forming an accurate address register.

5.3 The meeting was informed that address information could be used to target more resources at addresses likely to yield lower response rates (such as student addresses). The targeting of addresses in this way raises some issues which will benefit from ethical consideration. Benefits to the public could include reduced costs, improved accuracy and more effective targeting of resources.

5.4 Members suggested that, should ONS wish to target addresses based on the types of individuals residing in them (for example students), then this would need to be well circulated.

6. Application: Use of civil registration data as a sampling frame for third party surveys [NSDEC(16)02]

6.1 Mr Nick Stripe, ONS Life Events and Population Sources division presented an overview of the work. Mr Stripe explained that over the past five years ONS have received increasing numbers of requests to run surveys on behalf of third parties, such as the Department for Health. ONS's involvement in these surveys can range from survey design through to output or using ONS data (such as births and deaths registration) as a sampling frame.

6.2 The meeting was informed that legal advice had been sought and that the use of the data for such purposes was legal with fair processing requirements met.

6.3 Mr Mark Gautrey presented Annex B which provided detail on the National Survey of Mother and Infant Health, which is conducted by the National Perinatal Epidemiology Unit at the University of Oxford. Mr Gautrey explained that ONS's involvement in the survey was to provide a sampling frame for mothers who had recently given birth.

6.4 Ms Helen Colvin presented Annex A providing an overview of the National Survey of Bereaved People (VOICES). Ms Colvin informed the committee that there is uncertainty as to whether the survey will be put on hold next year by NHS England.

6.5 Members saw the public benefits of both surveys and made the following comments and suggestions:

i. for both studies there needs to be further development of systems to help ensure that those who decline to participate are flagged and not sampled in similar surveys in the future;
ii. more information as to how registrants data is used should be provided at the time of registering a vital event; and
iii. for the VOICES survey the information leaflet for prospective participants should:
    - be re-structured so that prospective participants can see immediately who is organising and funding the research; and
    - include more information about who ONS are.

6.6 The meeting agreed that both projects could proceed and could be used as precedent for similar surveys. The committee recommended minor revisions, with the secretariat to follow up to ensure that comments and suggestions made in 6.5 are implemented.

6.7 Any future proposals relating to sub-national surveys should be made as separate applications to NSDEC.

6.8 The Deputy Chair thanked the presenters on behalf of the committee for an excellent and professional paper and presentation.

Action: Secretariat to follow up with Life Events and Population Sources to check on progress with recommendations made

7. Application: Twitter pilot [NSDEC(16)03]

7.1 Ms Jane Naylor, from ONS Population Methodology and statistical Infrastructure division, provided an overview of ongoing research using geo-located Twitter data in understanding patterns of mobility. The research also explores the use of Twitter data in sentiment analysis following ONS releases, to understand whether overall Twitter views expressed by a release were; positive, negative or neutral.

7.2 It was suggested that the Twitter pilot application for ethical review may not be relevant as some of the research had been undertaken with findings already published. However, others suggested that as the research was ongoing and at this stage was for feasibility purposes only, the project could still be considered.

7.3 Members agreed that whilst informative, the sentiment analysis was not relevant to the application and should be removed.

7.4 This particular data set raised a number of issues, including the extent to which Twitter users were aware that their geo-location settings were turned on or off. Members discussed the extent to which Twitter users were aware that their Twitter data could be accessed publicly or purchased by a third party.

7.5 The quality of data was also discussed. At present only one to two per cent of all tweets have geo-location turned on, but there is no information available about this group of tweeters and whether they made an informed choice to turn this function on.

7.6 All members agreed that there should be greater clarity around how Twitter data is used in statistical feasibility research and this should be communicated clearly to the public. There is also a need to better define and describe the user need that the feasibility research was intended to meet.

7.7 The committee did not reach a consensus on the Twitter research. Members agreed with the Deputy Chair's suggestion that the research should be allowed to continue for the time being, so long as there were no changes to the use of data from that already outlined in the application and while consultation with the National Statistician was undertaken to consider the next steps.

Action: Secretariat to update NSDEC before the next meeting

8. Any other business

8.1 Members reported no other business.

UK STATISTICS AUTHORITY

NATIONAL STATISTICIAN'S DATA ETHICS ADVISORY COMMITTEE

Agenda

Wednesday, 27 January 2016
Board Room, One Drummond Gate, London
10:30am – 3:00pm (coffee from 10:00am)

Chair: **Mr Robert Bumpstead (for Mr Ian Cope)**

Apologies: **Mr Ian Cope**
Professor Martin Severs

(10:30am to 12:00pm)

1 10:30am	Minute and matters arising from the previous meeting	Mr Robert Bumpstead
2 10:40am	Chair's report	Oral Report Mr Robert Bumpstead
3 11:00am	Developments in data sharing	Oral Report Mr Ross Young
4 11:45am	Revised: Estimating ethnicity from names	NSDEC(16)01 Mr Owen Abbott

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

(12:30pm to 3:00pm)

5 12:30pm	Address register	Oral Presentation Mr Alistair Calder
6 1:15pm	Use of civil registration data as a sampling frame for third party surveys	NSDEC(16)02 Mr Nick Stripe
7 2:15pm	Twitter pilot	NSDEC(16)03 Ms Jane Naylor
8 2:45pm	Any other business	

Next meeting: Tuesday 19 April 2016

Location: Titchfield

National Statistician's Data Ethics Advisory Committee

Minute

Wednesday 14 October, 2015
Boardroom, Drummond Gate, London

Present
Board Members
Mr Ian Cope (Chair)
Mr Robert Bumpstead (Deputy Chair)
Mr Colin Godbold
Ms Annie Hitchman
Ms Isabel Nisbet
Ms Marion Oswald
Professor Martin Severs
Dr Dean Machin

UK Statistics Authority
Mr Adil Deedat
Dr Simon Whitworth

Office for National Statistics
Mr Owen Abbot (for item 8)

Apologies:
Mr Neil McIvor
Mr Osama Rahman
Mr Hetan Shah

1. Minute and matters arising from the previous meeting

1.1 The minute of the meeting on 14 October 2015 was agreed by correspondence and signed off by the Chair.

Chair’s report

Mr Robert Bumpstead

List of Annexes

Annex A **Acquisition of new data sources, Adil Deedat, NSDEC Secretariat, Central Policy Secretariat, 19 January 2016**

Annex A- Acquisition of new data sources

Data Set	Information
PAYE and Benefits information	**Update:** These data have now been received (late 2015) and are currently being used in feasibility research.
All England Education Dataset	**Update:** Expected Spring 2016 – with ONS providing analytical services to BIS under section 22 on the Statistics and Registration Service Act.
Health Demographic	Data expected late 2016 – May be delayed due to Care.data patient opt-out implementation.
Driver and Vehicle Licensing Agency data	• Initial discussions taking place with the data owner so that ONS can understand the potential of the data that could be available. • DVLA data which could provide information on both car ownership and drivers.
TV Licensing	• Work starting to understand the legal position with regards to the data. • Could provide information for an address register.
Armed forces	**Update:** Record level data no longer being sought however ONS is in the process of gaining access to new aggregate figures.
Valuation Office Agency Data	• Legal gateway to support acquisition of VOA data was created through 'The Statistics and Registration Service Act 2007 (Disclosure of Revenue Information) Regulations 2015 in March 2015. • Initially acquired on behalf of the Housing Market Indices branch within Prices Division with the purpose of producing a 'single definitive house price index and accompanying statistics'. • Feasibility data set comprising all Billing Authorities delivered in June 2015 with full data set delivered in November 2015 with an enduring, monthly data supply to update and maintain the full data set thereafter. • Current acquisition work will enable Census Transformation Programme (CTP) and Social Survey Division (SSD) access to VOA data; this is currently at business case approval stage.
National Border Target Centre	• Data is to be used to support feasibility work in improving the International Passenger Survey (IPS) sampling and weighting frame. A series of unrounded aggregate data extract reports on inbound and outbound data between the dates of 1 October to 7 October 2015 was delivered on 19 November 2015. • Further data extracts may be requested depending on the results of the initial feasibility work.

Developments in data sharing

Mr Ross Young

UK Statistics Authority

National Statistician’s Data Ethics Advisory Committee

NSDEC(16)01

Revised: Estimating ethnicity from names

Purpose

1. This paper outlines a re-submitted project proposal from the Office for National Statistics, (ONS) which looks to measure the quality of, and further develop, a tool, which estimates ethnicity from names. The project will be run in collaboration with University College London (UCL). This proposal was presented at the last meeting and is now being revisited, following NSDEC’s advice for major revisions.

Recommendations

2. Members of NSDEC are invited to:

   - consider the changes made, in light of NSDEC recommendations; and
   - advise the National Statistician to approve the project presented at **Annex A**.

Background

3. ONS and UCL have been developing a tool, which estimates the ethnicity distribution of a population, based on names.

4. At the last meeting on 14 October 2015, NSDEC members saw the benefits of such a tool but agreed that the project should be classed as requiring major revisions and should be further considered by the Committee. The Committee sought clarification on:
   i. the intellectual property rights for the tool and who, following further development, would be able to make use of it;
   ii. the measures UCL and users of the tools have taken to ensure they have satisfied the requirements of the Data Protection Act around processing of sensitive personal data and other legal requirements; and
   iii. ongoing use of the tool.

5. These changes have been highlighted in relevant sections presented at **Annex A**.

Natalie Shorten, Central Policy Secretariat, UK Statistics Authority, 14 January 2016

List of Annexes

Annex A	**Revised application: Estimating ethnicity from names, Mr Owen Abbott, Methodology, ONS, 10 January 2016**

UK Statistics Authority

National Statistician’s Data Ethics Advisory Committee

Application for Ethical Review

The Application Process

This is an application form for applying for ethical review from the National Statistician’s Data Ethics Advisory Committee (NSDEC). You should use the additional guidance when completing this form.

The application form should be completed in ***plain English*** which is understandable to lay members and all abbreviations should be explained the first time they are used. The form should contain sufficient information to ensure a thorough ethical review can take place.

Please word process the form using Arial or Times New Roman font, size 11. Where necessary expand text boxes on the form to accommodate answers, but ensure word counts are adhered to where specified.

Where sections are not relevant to your study please mark as N/A.

On completion the responsible owner should sign the application form and send to:
nsdec@statistics.gsi.gov.uk

UK Statistics Authority

Section A
Application Details

A1	**Responsible Owner**

Full Name: [redacted]	**Position:** Branch Head, Sample Design and Estimation Branch, Population Methodology and Statistical Infrastructure Division.
Address: Room 2400, ONS, Segensworth Road, Titchfield, PO155RR	**Email:** [redacted]
	Telephone: [redacted]
	Organisation: Office for National Statistics

Declaration to be signed by the responsible owner

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

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

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

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

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

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

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

Print Name: [redacted]
Signature: [redacted]
Date: 18th January 2016

UK Statistics Authority

A2	**Applicant Details** (if applicant is not the responsible owner)

Full Name	**Position**
Address:	**Email:**
	Telephone:
	Organisation:

A3	**Project Information**

Project Title: Estimating ethnicity from names	
Start Date: March 2015	**End Date: Estimated March 2016**
Project Sponsor (select all that apply)	
☐ ONS ☑ Collaboration ☐ ADRN ☐ Other (please specify)	

A4	**Collaboration and Sponsors**

List of Collaborators/Sponsors	**Details and relevant documentation relating to collaboration** (you may attach copies of relevant documentation)
Office for National Statistics **University College London, Department of Geography**	This is a Joint project funded by ESRC under their secondary data analysis initiative phase 2 – 2013. The project was initiated by UCL, who invited ONS to contribute. The funding provided by ESRC is for UCL staff only. ONS and UCL have agreed and signed a MOU for this project. The MOU covers: data access, terms and conditions for data access, stakeholder engagement, publications clearance, breach and dispute procedures and project termination.

	ONS and UCL have agreed that the Intellectual Property Rights for the tool would be held by ONS (as the data used to develop the tool is ONS data). UCL will be granted a perpetual licence to use and distribute the software (but not sell it). **Previous ethical reviews** The use of the existing tool for NHS applications was reviewed by the Health Research Authority (HRA) who gave a favourable ethical review for a project titled "Small area, geodemographic profiling of health needs"[1]. The review highlighted specific patient benefits through application of this approach, for instance the identification of better medication for specific groups and increased participation in screening programmes.

A5	**Proposed Site of Research** (select all that apply)

Where will the research take place?

☑ ONS ☑ VML ☐ HMRC Data Lab

☐ ADRC-E ☐ ADRC-NI ☐ ADRC-W ☐ ADRC-S

☐ Other (please Specify)

Is this a secure site?

☑ Yes ☐ No

---

[1] See ethical approval record at http://www.hra.nhs.uk/news/research-summaries/small-area-profiles-of-health-needs/

Section B
Project Details

B1	**Please provide a brief high level summary of the research giving necessary background** *(max 500 words)*

ONS is exploring ways to produce new or update existing outputs from data such as those held for administrative purposes, particularly where there is a strong user need. One method is to apply a statistical model or algorithm to predict a new variable from those already included on such sources.

ONS is considering the potential for using name data in these models. There are existing commercial tools which estimate ethnicity from data including forenames and surnames, however their quality and origin is unknown. Often, administrative sources contain names but do not contain important characteristics such as ethnic group. There is a strong user demand for ethnicity statistics both in their own right and also combined with other variables.

This proposal is for a joint project with academics to measure the quality of an existing tool which estimates ethnicity from name data, with a view to improving the tool based on the findings. Users will have to register to be able to download and use the resulting software tool, which will include accompanying metadata and user guidance so that users of the tool can obtain their own estimates together with quality measures.

The existing tool has been successful, in that it has been used by a range of health care and other organisations. However, it was built using data that either did not include the whole population (e.g. the public version of the Electoral Roll) or were unlikely to fully represent the UK population (e.g. consumer data). Crucially, such sources do not include individuals' self-assignments of their ethnic groups, making the classification more remote from the population that is being classified than is desirable. It has not used ONS data previously, other than in comparison with published area level census outputs. This has meant that the predictions can be poor for some population groups, for instance specific groups whose names have become anglicised (e.g. those of Caribbean ancestry) or those from groups that consider themselves assimilated into British society (e.g. bearers or Irish names). The extent of the uncertainty of the predictions using the current tool is unknown. In addition, the tool provides estimates which are a mixture of ethnicity, nationality and religion – having discussed with UCL the intention is to produce estimates only of self-assigned ethnicity and language spoken, avoiding the vaguer terms that are necessary in the current classification in the absence of Census data.

2011 Census data would be used to measure the quality of the estimates. The census provides self-classified ethnicity which is more aligned with user requirements for outputs, whereas the existing data for creating the tool is based entirely upon pairings of given and family names.

The estimation methodology is based on the use of clustering algorithms which group forenames and surnames into groups based upon observed pairings of given names and surnames. The resulting classification will be probabilistic in nature – for each name there will be an associated probability that the individual belongs to any of a number of ethnic and linguistic groups.

B2 **Data Use**

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)*	UCL project members will only have access to aggregate level 2011 Census data in the VML.	ONS project members will use identifiable 2011 Census data to prepare aggregate level data.		
Other *(please specify e.g. Ordinance Survey Address register in the relevant options adjacent)*				

B3 **How will information be kept confidential and data kept secure?**
(max 500 words)

Only security checked ONS staff have access to the individual level census data on a secure server. UCL will only have access to aggregate level diagnostics as detailed in the study protocol.

ONS staff within methodology group will prepare and clean the 2011 Census unit level data within the CDME (Census Data Management Environment) environment. UCL will provide an algorithm which ONS will import into the CDME, run against the cleaned census data, produce agreed aggregate diagnostic tables and apply agreed thresholds and then submit these for export from the CDME. The census data custodian will check and approve the export from the CDME, whereupon the diagnostic tables will be transferred into the standard VML. The UCL researcher will then access the diagnostic tables through controlled access to the standard VML. UCL will then update and refine their algorithm, and pass the revised algorithm back to ONS to rerun and generate revised diagnostics and thus iterate around the process.

Only the UCL researchers named in the ESRC research proposal (Paul Longley, James Cheshire, Alex Singleton and Muhammad Adnan) and Kira Kowalska (a UCL Phd student) will work on this project. UCL researchers needing to access the diagnostics will have to be ONS approved researchers in order to access the VML (as per standard VML access protocols). The ONS methodology staff working on this project are Owen Abbott, Helen Ross and Adriana Castaldo.

The diagnostic tables will mostly consist of proportions (not counts) to minimise the perception of disclosure. In addition, thresholds will be applied to minimise disclosure risks (e.g. to ensure that very rare names do not appear in diagnostic outputs). For diagnostics that do not include forename or surname fields, a threshold of 3 will be applied. For diagnostics that include forename or surname fields, a threshold of 10 will be applied. These diagnostics are not identifiable. The data will be accessed in a controlled manner through the standard VML, which is being used as the mechanism to access this sensitive data for this project only.

The resulting tool will be disseminated through a website that will require users to register and provide information about their intended use of the tool, including that they will comply with relevant legislation (e.g. data protection act). This will be assessed by ONS and/or UCL representatives as to whether the applicant is genuine and the intended use is appropriate, ethical and for the public good. Permission will not normally be refused, on the assumption that the application is for the public good. Following clearance, users will then be able to download and use the tool alongside their own data. We will ask users to register for each separate use of the tool so that we can monitor usage and benefits.

B4	**Please provide details of the research protocol or methodology** ***(e.g. data linkage, web scraping etc) (max 500 words)***

Names are a valuable link to family history and to cultural heritage, and can thus have important applications to understanding migration and population structures. Since 2003, research led by Professor Paul Longley at University College London (UCL) has investigated the geographical concentrations of names and ethnicity in the British Isles, and devised methods to analyse this data for a range of applications.

This has resulted in a tool which allows users to estimate the ethnicity distribution of a population for which names (forename/surname combinations) are available. The tool has proved useful for organisations, particularly those in the health sector, to be able to add a predicted ethnicity class onto their patient database records where previously it was not available, or as a way of updating local profiles for service needs (e.g. for translators) as census data becomes out of date.

The project aims to explore how 2011 Census data, which included both names and self classified ethnicity and language spoken, can be used to understand and improve the accuracy of the estimates. This will enable a robust assessment of the existing tool's performance using the Census top level 18 ethnic groups, and suggest where improvements could be made. The project will tune and improve the existing tool. A new version will be disseminated as described in section B3, along with user guidance on how to interpret the estimates. This will enable users to apply the tool and understand its strengths and weaknesses, helping them to know where they can and cannot use it.

Study protocol

ONS staff will apply the existing tool to a subset of 2011 Census data. These will include records with ethnicity, language spoken, religion, country of citizenship, area of the country, age (which bears a correspondence to the popularity of many forenames) and names. They will be used to produce aggregate diagnostics for UCL showing the success of their algorithm. Thresholds will be applied to these diagnostic tables, designed so that the frequency of a particular forename or surname is above a minimum number. This will be to ensure that no person can be uniquely identified within the tables, and therefore unique forenames or surnames will be excluded and not used in this project. Similar rules will be applied to cross tabulations. These procedures have been assessed by ONS disclosure control experts to ensure we maintain our commitment to the public on the confidentiality of their personal information. These aggregate, non-identifiable data will be made available to UCL through the VML, where the UCL staff will analyse the diagnostics and make improvements to their algorithm. The revised algorithm can then be passed back to ONS so that it can be re-applied to the secure microdata to derive new diagnostics. This process will be iterative. All UCL staff who have access to the non-identifiable data will have signed the Census Confidentiality Undertaking. No results or analysis will be allowed to leave the VML if they are disclosive or include any personal data that will identify an individual.

B5	**Please outline the proposed benefits of the project** ***(max 500 words)***

The main evidence for the benefit to the public comes from the application of the existing software.

The UCL software emerged from collaborations with two primary care trusts (PCTs) in ethnically diverse London boroughs to improve ethnicity designations in medical records, and targeted public health initiatives. For instance, a London borough used the software in a pilot project seeking to increase extremely low rates of breast cancer screening amongst women of African Caribbean descent. It used the names classification to identify the ethnic groups of women who missed screening, and then targeted resources and information accordingly, leading to an increase in the uptake of screening among African Caribbean women. This led to specific patient benefits through earlier identification of breast cancer for that specific group.

Other PCTs have used the tool to analyse GP referral patterns and admissions to accident and emergency facilities to measure equality of service usage.

Between 2008 and 2013, over 15 PCTs, local authorities and other government organisations have licensed the tool; for example, in 2011–2012 the Health Protection Agency (now Public Health England) used it in a survey of hepatitis and other blood borne viruses to explore whether transmission was related to ethnicity, while NHS Lothian licensed it in 2010 to assess access to public health services such as smoking cessation. Most of these public sector applications are for seeking improvements to services, ensuring equality of access to public services or assessing whether the quality of that service differs across ethnic groups.

However, the quality of the existing software is unknown. Therefore, the decisions made for these and future applications were in the absence of information about the accuracy of the outputs. Were this to be provided, then the decisions made would be better informed. In addition, if the underlying quality were to be improved by using additional information on language spoken, age, religion and citizenship then this would also provide additional benefits to service providers and the public.

Thus, for example: health researchers would be able to augment imprecise descriptors (e.g. 'African') that are appended to many health records in order to measure and monitor the effectiveness of (preventive and remedial) interventions across sub-populations; public attitude surveys could be used to identify the degree of cross community support for reassurance policing; universities would be able to establish the effects of their ‘widening participation’ initiatives in selection and recruitment policies in different subjects of study; and employers would be able to investigate the representativeness of their recruitment procedures with respect to their local labour markets. ONS plan to explore this sort of approach for providing improved ethnicity estimates either between censuses or in the absence of a census.

B6	**Please outline any ethical issues that might arise from the proposed study and how they will be addressed** ***(all research projects have some ethical considerations, so this section must not be left blank)***

a) There could be a perception that the tool uses census data as a ‘lookup’ for a name to provide an ethnicity. We would mitigate this by:

- ensuring that the tool always indicates the level of uncertainty for its estimates, e.g. through the provision of probabilities of a forename/surname combination belonging to a particular ethnic group (see section B6).
- Providing clear documentation on how the tool works (i.e. that it does not contain any census data)

b) There could be a risk that this tool is misused. For instance, someone with access to a list of names could use the tool to estimate ethnicity to discriminate against that group. This could happen anyway using the existing tool or existing commercial alternatives (or through a person guessing a persons ethnicity from names). This project will provide information about the accuracy of estimates which may reduce the likelihood of this happening, although it could not be prevented. As described in section B3, we will safeguard against this by asking users to register and provide information about them and their intended usage before they can download the software. Whilst this may not completely prevent determined misuse, it will provide some protection.

c) There is a risk that the tool is labelled as ‘ONS approved’, and therefore it might be perceived to have a error free output despite the intended provision of transparent methodology and information about quality.

d) There is a risk that some of the categories in the classification may be ambiguous or inappropriate. This arises in the existing tool because it does not use any data on the groups that name bearers would assign themselves to. Use of census data will greatly reduce this risk because predictions will be modelled using information on the groups that individuals assign *themselves* to.

e) There is a risk that users will be insufficiently aware of the uncertainty that is associated with the estimates. This was apparent in a previous version of the Onomap website that allowed users to enter a single name and receive a single predicted ethnic group without any associated level of uncertainty or possible alternatives. We have since discussed with UCL who have taken down the site, and we do not envisage using this approach in the future. In addition, all estimates will always be accompanied by an indication of uncertainty.

f) There are a number of existing tools marketed by commercial organisations. They do not necessarily provide information on the quality of the outputs and are not transparent in their methodology. There is a risk that users of such tools are currently making poor decisions for want of better alternatives and because the quality of estimates are not understood. This project will enable registered access to a tool with known providence and enable them to make better decisions through providing information on the quality of outputs and on the methods used to create the estimates.

g) The use of name data for anything other than data matching (or to produce primary outputs like baby names which are of public interest) is not something that ONS has undertaken before – although linking names to age is deemed acceptable elsewhere (see http://fivethirtyeight.com/features/how-to-tell-someones-age-when-all-you-know-is-her-name/), is modelled by commercial organisations in the UK and can be estimated for young people in the UK (excluding immigrants) using Birth Registration data.

h) There might be a perception that we are giving individual level census data to UCL. Preserving the confidentiality of personal information provided by the public on their census questionnaire remains ONS's utmost priority. The tool will not have any census data within it, and this will be made clear in the documentation. Only security checked ONS staff have access to the individual level census data on a secure server. UCL will only have access to aggregate level diagnostics as detailed in the study protocol.

i) There is a risk that organisations use the tool to derive ethnicity instead of collecting it directly. For instance, if they have a statutory requirement to measure service provision under the Equalities Act. This could happen anyway using the existing tool. This project will provide information about quality which may reduce the likelihood of this happening, as there may be restrictions around the quality of such measurement, although it could not be prevented. In some cases, if the quality is good then this might provide a benefit in reducing burden on the public and costs.

B7	**How will the findings of the research be disseminated?**

Through research papers outlining the research methodology and findings – these will be authored by UCL but cleared by ONS.

Through the software being made available for download (on a website requiring registration as described in section B3), together with metadata, user guidance and the above papers. The software will allow the user to import a list of forenames and surnames, and will return an output dataset which includes for each name the most likely ethnic group together with an estimated probability of that being correct, as shown below (using made-up data):

a) Input table

Forename	**Surname**
Wayne	Rooney
Didier	Drogba
Asmir	Begovic
Gareth	Bale
David	Silva
Sulzeer	Campbell
Shinji	Okazaki

b) Output table

Forename	**Surname**	**Estimated ethnic group**	**Estimated Probability that estimate is correct**
Wayne	Rooney	White: English/ Welsh/ Scottish/ Northern Irish/ British	0.92
Didier	Drogba	Black/ African/ Caribbean/ Black British: African	0.83
Asmir	Begovic	White: Any other White background	0.89
Gareth	Bale	White: English/ Welsh/ Scottish/ Northern Irish/ British	0.85
David	Silva	White: Any other White background	0.85
Sulzeer	Campbell	Black/ African/ Caribbean/ Black British: Caribbean	0.75
Shinji	Okazaki	Asian/ Asian British: Any other Asian background	0.91

If the research methodology allows, we would like the output to include the likelihood of that name belonging to all ethnic groups (with the corresponding probabilities that will obviously be small), so that the output might look like (assuming there are only really 5 ethnic groups for illustrative purposes (although the tool will actually use the standard 18 group ethnic categories used in the census outputs) and a ‘-‘ implies that a probability is less than 0.01):

c) Output table including all ethnicities

Forename	Surname	Estimated probability of being				
		White: English/ Welsh/ Scottish/ Northern Irish/ British	White: Any other White background	Black/ African/ Caribbean/ Black British: Caribbean	Black/ African/ Caribbean/ Black British: African	Asian/ Asian British: Any other Asian background
Wayne	Rooney	0.92	0.05	0.02	0.01	-
Didier	Drogba	0.01	0.02	0.13	0.83	-
Asmir	Begovic	0.10	0.89	-	-	-
Gareth	Bale	0.14	0.85	-	-	-
David	Silva	0.07	0.85	0.03	0.02	0.03
Sulzeer	Campbell	0.02	0.05	0.75	0.17	-
Shinji	Okazaki	0.02	0.05	-	-	0.91

The standard 18 group ethnic categories used in the census outputs are below:

White
English/ Welsh/ Scottish/ Northern Irish/ British XXX
Irish XXX
Gypsy, Traveller or Irish Traveller * XXX
Any other White background XXX
Mixed/ Multiple ethnic groups
White and Black Caribbean XXX
White and Black African XXX
White and Asian XXX
Any other Mixed/ Multiple ethnic background XXX
Asian/ Asian British
Indian XXX
Pakistani XXX
Bangladeshi XXX
Chinese XXX
Any other Asian background XXX
Black/ African/ Caribbean/ Black British
African XXX
Caribbean XXX
Any other Black/ African/ Caribbean background XXX
Other ethnic group
Arab XXX
Any other ethnic group XXX

The inclusion of additional, more detailed, categories will be considered during the research period, based on consultation and advice from advisory groups alongside disclosure issues.

Section C
Details of Data Subjects

C1	**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:

N/A

Justification for focusing on these subsections or groups:

N/A

C2	**Please detail consent given to use data specified in section B2**

The 2011 Census Information Asset Owner has provided consent to use the data through the VML access mechanisms, given that the data made available to UCL are aggregates and thresholds are applied to ensure so individual can be identified.

C3	**If you are using data held by a third party please detail how you will obtain this**

N/A

Address register

Mr Alistair Calder

UK Statistics Authority

National Statistician's Data Ethics Advisory Committee

NSDEC(16)02

Use of civil registration data as a sampling frame for third party surveys

Purpose

1. The purpose of this paper is to provide NSDEC with information on the work of ONS's Life Events and Population Sources division, in relation to life events surveys. These surveys are based on samples taken from administrative registers of births and deaths.
2. This paper is intended as an encompassing note to complement specific papers on the National Survey of Bereaved People **(Annex A)** and the National Survey of Mother and Infant Health **(Annex B)**, both presented as example items for the Committee's consideration.

Recommendations

3. Members of NSDEC are invited to consider advising the National Statistician approving ONS activities in using civil registration data to run a suite of surveys. Specifically we seek:
   - guidance on whether new requests for running surveys of a similar nature to those presented at **Annex A** and **Annex B** can proceed on precedent or following the commissioning body receiving ethical approval from an external ethics committee, such as health or academic research ethics committees; and
   - the committees views on running small sub-national surveys based on registration data.

Background

4. Researchers and policy makers are becoming increasingly interested in utilising birth and death registers as a sampling frame to measure citizens' perceptions on a range of care pathways in the National Health System, from mother and infant health to end of life care. ONS is in a unique position, being the sole provider of the sample frame for these types of surveys. The sample frames are our statistically coded and validated registers of births and deaths.

5. It is a legal requirement to register births, marriages and deaths, which occur in England and Wales, with the Local Registration Service. Under section 42 of the Statistics and Registration Service Act 2007 (SRSA) the Registrar General for England and Wales (effectively the General Register Office (GRO)) shares this data with ONS. We code and validate this data in order to produce a suite of aggregate National Statistics. We also share microdata via lawful gateways to enable further statistical research into health and demography. Section 22 of the SRSA allows ONS to provide statistical services to others. Compiling sampling frames for social surveys is considered a statistical use of the registers.

6. Since 1995 ONS has used birth and death registration data as sampling frames for maternity and mortality based surveys. The surveys have been commissioned by third parties, such as NHS England and external researchers.

7. The number of requests to use birth and death registration data as sampling frames has increased over time, from a single survey request every couple of years to two to

three requests a year. Not only have the amount of requests increased, but so has the complex nature of issues encountered.

Discussion

Legal Position

Lawful gateway to use data in this way

8. Section 42 (S42) of the Statistics and Registration Service Act (SRSA) provides ONS with cover to receive the data and to use it for any statistical function. Section 22 (S22) of the SRSA provides cover for us to use the data for statistical services to any person inside or outside the United Kingdom. See **Annex C**, ONS Legal Services have confirmed that this is a lawful use of the data.

Fair Processing

9. The issue of 'Fair Processing' has also been investigated. ONS Legal Services have confirmed ONS is currently meeting Fair Processing principles by informing potential survey participants at the outset how ONS has received their information and by taking an 'opt in' approach to respondent participation, rather than an 'opt out' approach. In other words, once contacted about the survey, those sampled must opt in to actually take part. They do this by returning their agreement to participate either to ONS or to the third party researcher, depending on survey design. See **Annex D**.

10. ONS is further strengthening its commitment to Fair Processing by:
    i. working with GRO to have privacy notices in local registration offices, explaining how informants' data may be used; see **Annex E**. Although not compulsory to display this notice, GRO expect registration offices will comply with this best practice;
    ii. seeking approval from the NSDEC before administering a survey; and
    iii. providing potential respondents with further details on how ONS has contacted them, by having a statement on the ONS website explaining how we receive and use registration data.

Why do researchers contact ONS?

11. Civil registration data is supplied to ONS by GRO. ONS has a history of involvement in life event surveys, as prior to independence in 2008, GRO was a division of the office. Since 2008 the GRO have been clear they consider surveys to be a statistical service, which should be provided by ONS.

12. ONS is in a unique position. The SRSA provides the legal basis for ONS to receive civil registration data and for it to be used for statistical purposes; this can be supplemented by researchers applying for Approved Researcher status. ONS is also considered an expert in designing, administering and producing outputs from surveys.

Current criteria for considering a survey request

13. In addition to assessing whether the request is feasible and that data and resource are available, further considerations are also made:

    i. Is the applicant an Approved Researcher or is the work to be undertaken for the purpose of assisting the Secretary of State or Welsh Ministers?
    ii. Has the research obtained external ethics approval?
    iii. Will the survey outcomes be used for the public good?

iv. Does the content of the questions in the questionnaire and other survey materials (leaflets, letters etc) meet Government Social Research and ONS guidelines?
v. The researcher must agree to an ‘opt in’ approach to respondent participation.

Survey services provided by ONS

14. Services vary depending on customer requirements but may include a combination of the following:

	Services Provided	**Amount of surveys service provided for since 2010**
1	Designing, drawing and quality assuring the sample	14
2	Devising or advising on the questionnaires and accompanying documentation	5
3	Printing questionnaires and documentation	6
4	Quality assuring addresses by adding missing house numbers and postcodes	14
5	Screening data i.e. remove any baby deaths from maternity surveys	12
6	Adding name and address details to letters and questionnaires	14
7	Monitoring responses and sending reminder letters	14
8	Analysis of survey results and their publication	5
9	Provision of respondents/non respondents data (e.g. age of mother by 5 year age band and country of birth of mother)	8
10	Providing the respondent with the ONS Survey Enquiry Line to help field their queries and issues.	5
11	Provision of organisations who can provide support if the questionnaire is sensitive in nature [e.g. Cruse Bereavement Care charity)	13
12	Responding to additional queries or complaints from respondents	8

Provision of services to run sub-national surveys

15. Currently, all surveys ONS have run using registration data as a sample frame have been national (England, or England & Wales) surveys. ONS has had a number of enquires recently asking if we can run surveys at the sub-national level. This interest has come from local councils, academic students and Clinical Commissioning Groups who wish to run local surveys with small sample sizes, i.e. 250 – 1000 respondents. Issues that need to be considered in undertaking sub-national surveys are:

i. **Respondent burden** – need to avoid the same respondents being selected for different studies by selecting deaths occurring at different times for each different study.
ii. **Respondent selection in future surveys** - when respondents opt out of participating in the survey, there is currently no robust method available for identifying them for removal from sampling in future surveys. Consequently, they are removed from the current survey, but can be re-contacted for future surveys if they register another death. It is possible to avoid re-contacting about the deceased, but not the person who registers a death.

iii. **Value of the research** – smaller geographical studies, based on small samples and/or very specific causes of death, may be argued to have less value for the public good, compared to large scale national surveys.

Mark Gautrey, Life Events and Population Sources Division, ONS, 19 January 2016

List of Annexes

Annex A	**Application: Use of deaths data in the National Survey of Bereaved People (VOICES Survey), Neil Bannister, Life Events and Population Sources, ONS, 19 January 2016**	
	Annex A1	**Example survey form: The National Survey of Bereaved People**
	Annex A2	**Information leaflet for respondents to the VOICES Survey**
Annex B	**Application: use of births registration data in the National Survey of Mother and Infant Health, Mark Gaurtey, Life events and Population Sources Division, ONS and the National Perinatal Epidemiology Unit, 19 January 2016**	
Annex C	**Relevant extracts from the Statistics and Registration Service Act**	
Annex D	**Relevant extracts from the Data Protection Act**	
Annex E	**Privacy notice at the General Registrars Office**	

UK Statistics
Authority

National Statistician’s Data Ethics Advisory Committee

Application for Ethical Review

The Application Process

This is an application form for applying for ethical review from the National Statistician’s Data Ethics Advisory Committee (NSDEC). You should use the additional guidance when completing this form.

The application form should be completed in ***plain English*** which is understandable to lay members and all abbreviations should be explained the first time they are used. The form should contain sufficient information to ensure a thorough ethical review can take place.

Please word process the form using Arial or Times New Roman font, size 11. Where necessary expand text boxes on the form to accommodate answers, but ensure word counts are adhered to where specified.

Where sections are not relevant to your study please mark as N/A.

On completion the responsible owner should sign the application form and send to: nsdec@statistics.gsi.gov.uk

UK Statistics Authority

Section A
Application Details

A1	Responsible Owner

Full Name [redacted]	**Position:** Head of End of Life Care Analysis
Address: Rm 1.059, ONS Government Buildings, Cardiff Road Newport NP10 8XG	**Email:** [redacted] **Telephone:** [redacted] **Organisation:** ONS

Declaration to be signed by the responsible owner

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

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

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

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

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

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

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

Print Name: [redacted]
Signature: [redacted]
Date: 15 January 2016

UK Statistics Authority

A2 Applicant Details (if applicant is not the responsible owner)

Full Name	**Position**
Address:	**Email:**
	Telephone:
	Organisation:

A3 Project Information

Project Title: National Survey of Bereaved People (VOICES)

Start Date: May 2016	**End Date:** April 2017

Project Sponsor (select all that apply)

☐ ONS

☐ ADRN

☐ GSS

☐ Collaboration

☑ Other (Please specify) **NHS England**

A4 Collaboration and Sponsors

List of Collaborators/Sponsors	**Details and relevant documentation relating to collaboration** (you may attach copies of relevant documentation)
NHS England	

UK Statistics Authority

A5	**Proposed Site of Research** (select all that apply)

Where will the research take place?

☑ ONS

☐ VML

☐ HMRC Data Lab

☐ ADRC-England

☐ ADRC-Northern Ireland

☐ ADRC-Scotland

☐ ADRC-Wales

☐ Other (please specify).....

Is this a secure site?

☑ Yes ☐ No

Section B
Project Details

B1	**Please provide a brief high level summary of the research giving necessary background** ***(max 500 words)***

The National Survey of Bereaved People (VOICES, Views of Informal Carers – Evaluation of Services) collects information on bereaved peoples' views on the quality of care provided to a friend or relative in the last three months of life, for England. The survey has now been run for five years and was commissioned by the Department of Health in 2011 and 2012, and NHS England from 2013. It is administered by the Office for National Statistics (ONS).

VOICES data provides information to inform policy requirements, including the End of Life Care Strategy, published by the Department of Health in July 2008. This set out a commitment to promote high quality care for all adults at the end of life and stated that outcomes of end of life care would be monitored through surveys of bereaved relatives

(Department of Health, 2008). Recently, the Liverpool Care Pathway, which provided a protocol for end of life care, has received criticism (Department of Health, 2013). As a result, new questions were added to the VOICES survey in 2014 to measure changes in delivery of care, while the Liverpool Care Pathway is replaced.

The VOICES sample is selected from the deaths registrations database. Eligible deaths occurring between 1st January and 30th April each year form a sampling frame for a stratified random sample of approximately 49,000 deaths.

A postal survey is sent to people who registered the death of their friend or relative. The survey contains 60 questions on the quality of end of life care provided to the deceased in the last three months of life and includes a free text section for any other information that the respondent wishes to provide on the back page (See **Annex A1**). The field period runs from 1st September until mid December each year, which allows for a period of between four and nine months prior to first contact.

Results from the survey are published annually at national level and can also be combined into more than one year of data for analysis at sub regional level. For instance, combining data from the 2011 and 2012 field periods provides a larger dataset, which enables more reliable estimates at sub regional level. Sub regional analysis has previously been conducted for Primary Care Trusts using the 2011-12 data and NHS Area Teams; using 2012-13 (these are health geographies for England, which were in use during the different time periods).

Datasets are also being prepared for the UK Data Service at special licence (SL) and end user licence (EUL) level to enable secondary analysis of the data by third party researchers. This is standard for data collected within ONS and will maximise the use of the data by making it accessible to third parties, who can request access to the data for their own purposes.

B2	**Data Use**

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)*		Deaths Registration Database. Deaths occurring between 1st Jan and 30th April form the		

		sampling frame		
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)*				

B3	**How will information be kept confidential and data kept secure?** ***(max 500 words)***

- Data is held on a secure server, accessible only by authorised staff.

- The survey is posted to a respondent, enclosing only the respondent's name and address and the name of the deceased.

- ONS make a confidentiality pledge to respondents in the information leaflet (see **Annex A2**) provided with the questionnaire:

  "ONS will not give personal information that identifies you to anyone else. Your survey responses will only be identified by an anonymous Study ID number. All the information collected will be kept strictly confidential and secured against unauthorised access. We would also like to make absolutely clear that no information that could identify you will be used in any reports or articles we write.

  If you add comments at the end of the questionnaire this may be used by approved researchers, to inform studies on people's personal experiences. Please try not to write the names of any people or places that may identify you. Any names written will be removed to protect your identity. You will not be asked to include your name (or the name of your relative, partner or friend) on the questionnaire."

- Respondents are given a unique ID number and name and address information is separated from personal details of the deceased (e.g. cause of death, place of death information, geography, age etc).

- Surveys are returned to the ONS where they are scanned with name and address information redacted.

- Datasets combining sampling information and survey responses exclude personal name and address information.

- Datasets planned for release to NHS England and the UK Data Service will have Microdata Release Panel (MRP) approval and Data Access Agreement (DAA) controls in place. Datasets for release under EUL are currently being discussed with the Survey Data Collection (SDC) team to ensure they protect respondents against identification.

- A method to redact free text information has been developed. Any free text released by ONS will be redacted and protected by MRP and DAA procedures.

B4	**Please provide details of the research protocol or methodology** ***(e.g. data linkage, web scraping etc)*** ***(max 500 words)***

- The VOICES survey was piloted by Southampton University and the ONS. This work aimed to ensure that VOICES was a suitable tool for collecting information on quality of end of life care nationally. As the survey was conducted outside NHS premises and did not involve NHS patients or staff, it was not necessary to seek NHS ethical approval. Ethical approval was, therefore, sought from the University of Southampton, School of Health Sciences' Ethics Committee. In addition, the project protocol was reviewed by the Head of Health Analysis at ONS.

- In 2014 changes were made to the VOICES survey to update it in line with end of life care policy developments. The survey underwent cognitive testing with bereaved volunteers from the ONS and recruited through the National Council of Palliative Care. Changes were made to ensure questions had good reliability and were worded as sensitively as possible.

- In the annual VOICES survey, deaths occurring between 1st January and 30th April are selected from the deaths registrations database.

- Ineligible deaths are excluded. Eligible deaths are those where:
  Deceased aged 18+
  Death was not sudden
  Deceased's usual place of residence was England
  Death was registered by a friend or family member
  Place of death was home, hospital, care home or hospice
  Address is available for person who registered the death

- A random sample of approximately 49,000 is selected from eligible deaths, stratified by place of death (home, hospital, care home and hospice), cause of death (cancer, cardiovascular disease and all other non-external (suicide or accidental) causes) and geography (NHS Area Team, 25 groups).

- A paper survey is mailed to the person who registered the death at the start of the data collection period in September. The initial survey pack includes an invitation letter, the questionnaire, an information leaflet and a reply paid envelope.

- The survey is phrased according to the gender of the respondent (e.g. ‘how long had *she* been ill?’) and contains 60 questions about the quality of care by setting and service provided. Respondents are invited to provide further details of the care in their own words at the back of the survey.

- Two further reminders are sent to non-responders at four week intervals.

- Returned questionnaires are scanned electronically with identifiable details redacted.

- Data is cleaned to remove erroneous results and weighted to adjust for biases in the sampling and response processes.

B5	**Please outline the proposed benefits of the project** *(max 500 words)*

The VOICES survey provides the only source of national information on the quality of end of life care. The data provides information to inform policy requirements, such as the NHS Outcomes Framework indicator 4.6 ‘improving the experiences of care at the end of life’.

The data is used by health care providers and policy makers for informing service provision and is also widely used by charities for assessing the needs of service users and by academics for investigating broader issues in relation to end of life care.

End of life care is high on the political agenda. Recent criticisms of the Liverpool Care Pathway have highlighted concerns of withdrawal of fluid and nutrition in the final stages of life and the survey has been amended to measure changes in satisfaction with these aspects of care.

B6	**Please outline the ethical issues that might arise from the proposed study and how they will be addressed** ***(all research projects have some ethical considerations, so this section must not be left blank)***

1. **The use of the deaths registrations database sending unsolicited requests for participation in a sensitive survey.**
   There is no mention, at the time of registering a death, that the information provided

may later be used for sending an unsolicited invitation to participate in a sensitive survey. Despite this, the information registered is used as the basis for selecting a sample. The use of the data for this purpose is fully authorised by Act of Parliament and no prior consent is needed. See C2 for further information.

2. **Receiving the survey can cause distress for the respondent**
   VOICES can be a distressing survey for some respondents and we provide contact details for the charity Cruse Bereavement Care on the survey and accompanying information leaflet. The survey is provided in a paper based form, with a deadline of over three months to enable the respondent to take their time to complete it or pass it on to someone more suitable. Respondents are contacted between 4 and 11 months post bereavement, allowing time for the most acute period of grief to pass and avoiding significant dates, such as Christmas and the anniversary of a death.

   VOICES has undergone cognitive testing, both when it was adapted for use as a national survey and when the survey was adapted in line with policy changes in 2014. In 2014, volunteers for cognitive testing included people who had been bereaved in the preceding months, recruited from both the ONS and the National Council of Palliative Care. Attention was given to ensuring that questions were worded as sensitively as possible.

3. **Respondents may disclose serious abuse**
   The VOICES team has recently received allegations of serious abuse within a care home. Standard procedure is to write to the respondent, explaining that we cannot act on individual experiences and directing them to the Care Quality Commission. ONS prioritises the protection of the individual’s confidentiality and will not release identifiable information to a third party for non statistical purposes.

4. **Use of free text data**
   The back page of the survey contains a blank page for respondents to tell us anything they would like to say about the care provided. Many of the respondents provide significant amounts of further details, attaching additional pages. The VOICES team have developed a method for redacting this information and potentially making it available for qualitative research analysis. We are hoping to be able to release the data to other researchers and have begun putting a process in place for this. The data would not be released in an identifiable format and would be provided as redacted PDF documents of the written information (transcription has proven too labour intensive). There is nothing published from this yet.

   However, in 2015 alone, there were 12,000 free text comments. This is a significant amount of information to redact and too much information for use in qualitative research. A potential method is to randomly select cases to redact from the sample, meaning that all data has a chance of being used, but realistically not all of this information will be used in further research.

   Despite this, respondents often welcome the chance to share the details of their experiences, for reasons such as cathartic release and the opportunity to share their experiences or raise complaint, rather than just ‘tick boxes’. The removal of the free text could potentially reduce the response to the survey.

5. **Reminder letters**
   Up to two reminder letters are sent to people who have not responded, four and eight weeks after the initial mail out. These all have options for declining with a reply paid envelope enclosed to enable respondents to opt out easily. Every effort is made to ensure that people who have responded are not re-contacted. This is done by logging refusals which come in from the survey enquiry line and scanning returns from the mail. A list of reminder letters is created after all post for the day has been receipted to ensure that the list is as up to date as possible.

6. **Respondent selection in future surveys**
   When respondents opt out of participating in the survey, there is no robust method available for identifying them for removal from sampling in future surveys. Consequently, they are removed from the current survey, but can be re-contacted for future surveys if they register another death. It is possible to avoid re-contacting about the deceased, but not the person registering a death.

7. **Method of contact for respondents**
   Respondents are provided with the name and address of the survey manager, the telephone number of the survey enquiry line and a Stamped Addressed Envelope (SAE). This enables respondents to contact us to ask questions, complain or decline to participate.

   Respondents who are distressed or make complaints about the survey are sent a personalised letter of apology and explanation by the research team, and removed from further reminder lists.

8. **Informed consent**
   The survey contains an information leaflet and invitation letter with the questionnaire. The aim of these is to ensure that the respondent has enough information to make an informed assessment about whether they would like to participate. The letter clearly states that the survey is voluntary and clear instructions are provided for how respondents can opt out of participating.

B7	**How will the findings of the research be disseminated?**

Findings are disseminated through an annual national bulletin published on the ONS website. This includes a bulletin and downloadable tables of results. The information published is tabular and does not contain identifiable results. Previously, NHS England has also requested a combined bulletin for two years of data at NHS Area Team level.

The published results are reproduced as an indicator on the Health and Social Care Information Centre and also further disseminated by the National End of Life Care Intelligence Network and charities, such as Marie Curie and Macmillan. Individual NHS Area Teams have also used the data to produce information and monitor the quality of care within their own areas.

Datasets at individual level are also being prepared for the UK Data Service at special licence (SL) and end user licence (EUL) level to enable secondary analysis of the data by third party researchers. This is standard for data collected within ONS and will maximise the use of the data by making it accessible to third parties, who can request access to the data for their own purposes.

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.**

It is intended to make datasets for each year of the survey available on the UK Data Service archive under special licence and end user licence terms. Datasets will also be provided to NHS England. The purpose of releasing these datasets is to enable further analysis by academics and charities and by the NHS, who commission the survey.

Section C
Details of Data Subjects

C1	**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 following criteria is used to determine eligible cases for the survey:

Deceased aged 18+
Death was not sudden
Deceased's usual place of residence was England
Death was registered by a friend or family member
Place of death was home, hospital, care home or hospice
Address is available for person who registered the death
Death occurred between the 1st January and 30th April in the survey year

Justification for focusing on these subsections or groups:

Deaths of under 18s are excluded due to the greater sensitivities around the death of a child and different epidemiology of health care needs for this group.

Sudden deaths are excluded, as the person would have been unlikely to have received end of life care.

The survey is commissioned by NHS England and applicable to those residents living within the location.

Deaths registered by people other than friends or family members are deemed unlikely to hold adequate personal experiences on the care provided to the deceased.

Asking for opinions from professionals may bias the results of the survey.

Place of death as home, hospital, care home and hospice includes only those areas of interest to provision of end of life care.

Adequate address details must be available for a respondent to be contacted. Those residing outside of the UK are also contacted.

The sample selection period remains the same each year to reduce seasonal variation in the results and enhance comparability between years.

C2	**Please detail consent given to use data specified in section B2**

The rights by which ONS holds deaths registrations information is as follows: the General Register Office (GRO) send it to the ONS so that deaths information can be used for statistical purposes. This is carried out under the Births and Deaths Registration Act 1953 and the Statistics and Registration Service Act 2007, and because it is required by law, there is no need for any specific consent for use of the information.

The ONS is empowered to receive this information under s42 of the Statistics and Registration Service Act. Under s22 of the Act, ONS may provide statistical services for another organisation, such as running a survey when requested by another government department or a university. Under s38, ONS is authorised to use information received under s42 with prior consent from the person we received it from.
use much of the information which it lawfully holds, for example from death registrations, for surveys or other statistical purposes.

C3	**If you are using data held by a third party please detail how you will obtain this**

N/A

Office for
National Statistics

NHS
England

000000

TEST PRINT
ADDRESSLINE1
ADDRESSLINE2
ADDRESSLINE3
ADDRESSLINE4
ADDRESSLINE5
ADDRESSLINE6
ADDRESSLINE7
ADDRESSLINE8

VOICES

VIEWS OF INFORMAL CARERS -
EVALUATION OF SERVICES

Dear TEST PRINT

Invitation to help with the VOICES survey of experiences of care in the last months of life.

If you would like to receive this information in large print please call our Survey Enquiry Line on 0800 298 5313

We are writing to you because you registered the death of .
We appreciate that this may be a very difficult time for you, but we would like to invite you to take part in the VOICES survey. VOICES gathers information on your experiences of health care services in the last months of life and is used by NHS England and other health care charities to monitor and improve the services they provide. For more information please read the information leaflet enclosed.

If you **do** wish to take part please complete the questionnaire and return it to ONS in the pre-paid envelope by **Friday 11th December 2015**.

If you **do not** wish to participate please tick the box on page 15 of the questionnaire and return it to ONS in the pre-paid envelope provided. This will ensure that you do not receive reminder letters.

Your views are important and will help improve health care for patients in England. We apologise if this enquiry has caused you any distress and hope that you feel able to take part in this study.

Yours sincerely

Neil Bannister
Head of End of Life Care Analysis
Office for National Statistics

000000

☑ **Answer the questions by putting a tick in the most appropriate box or boxes.**
If you wish to change your answer, cross through the answer you do not want.

- ☒ Excellent
- ☐ Good
- ☑ Fair

If she died suddenly, please go to question 38.
Otherwise, please continue with the questions below.

1 **How long had she been ill before she died?**

Tick one only

- ☐ She was not ill - she died suddenly - **go to question 38**
- ☐ Less than 24 hours
- ☐ One day or more but less than one week
- ☐ One week or more but less than one month
- ☐ One month or more but less than six months
- ☐ Six months or more but less than one year
- ☐ One year or more

2 **Did she spend any time at home during the last three months of life?**

Tick one only

- ☐ Yes - **go to question 3**
- ☐ No - she was in a care home for the whole 3 months - **go to question 12**
- ☐ No she was in hospital - **go to question 24**

Care at Home

These questions are about care at home - not in a care home

3 **When she was at home in the last three months of life, did she get any help at home from any of the services listed below?**

These may be provided by different organisations, such as voluntary organisations, a private agency or social services.

Tick all that apply

- ☐ A district or community nurse (a nurse in uniform who comes to the house)
- ☐ A Macmillan nurse, hospice home care nurse or specialist (a care nurse who visits or telephones to talk and advise on medications and other aspects of care. Not in uniform)
- ☐ A Marie Curie nurse (someone who comes to the house for a few hours or overnight to care for the patient)
- ☐ Any other nurse at home
- ☐ Home care worker, home care aide or home help
- ☐ Social worker / support worker
- ☐ Counsellor
- ☐ Religious leader
- ☐ Meals-on-wheels or other home delivered meals
- ☐ Hospice at home
- ☐ Occupational therapist (OT)
- ☐ Rapid response team (team of nurses and home care workers who provide care over the short term to allow someone to remain at home and prevent hospital admission)
- ☐ She did not receive any care
- ☐ Don't know
- ☐ Something else

000000

4 **When she was at home in the last three months of life, did all these services work well together?**

Tick one only

☐ Yes, definitely
☐ Yes, to some extent
☐ No, they did not work well together
☐ She did not receive any care
☐ Don't know

5 **Overall, in the last three months of care, do you feel that you and your family got as much help and support from health and social services as you needed when caring for her?**

Tick one only

☐ Yes, we got as much support as we needed
☐ Yes, we got some support but not as much as we needed
☐ No, although we tried to get more help
☐ No, but we did not ask for more help
☐ We did not need help

6 **During the last three months of her life, while she was at home, how well was her pain relieved?**

Tick one only

☐ Does not apply - she did not have any pain
☐ Completely, all of the time
☐ Completely, some of the time
☐ Partially
☐ Not at all
☐ Don't know

Urgent Care Provided Out of Hours

7 **In the last three months of life, while she was at home, did she ever need to contact a health professional for something urgent in the evening or at the weekend?**

Tick one only

☐ Not at all in the last three months - **go to question 12**
☐ Once or twice - **go to question 8**
☐ Three or four times - **go to question 8**
☐ Five times or more - **go to question 8**
☐ Don't know - **go to question 12**

8 **The last time this happened, who did she contact, or who was contacted on her behalf?**

Tick all that apply

☐ Her GP or the out-of-hours number
☐ NHS 111 (formerly NHS Direct)
☐ District nurses
☐ Macmillan nurses
☐ She used her 'lifeline' pendant
☐ A hospice
☐ 999
☐ Someone else

9 **What happened as a result? Was she ...**

Tick one only

☐ Visited by her GP at home

☐ Visited by another GP at home

☐ Visited by a nurse at home

☐ Visited by a hospice doctor at home

☐ Given medical advice over the phone

☐ Given another number to ring to get medical advice

☐ Advised to go to an out-or-hours GP surgery

☐ Advised to go the GP surgery when it opened

☐ Advised to go to an Accident and Emergency Department at a hospital

☐ Advised to call 999

☐ Something else

10 **Overall, on this last occasion, do you think that the health services responded in the right way?**

Tick one only

☐ Yes

☐ No

☐ Not sure

11 **Overall, do you feel that the care she got when she needed care urgently in the evenings or weekends in the last three months of her life was:**

Tick one only

☐ Excellent

☐ Good

☐ Fair

☐ Poor

☐ Don't know

District & Community Nurses

If she had care in the last three months from district and community nurses - **go to question 12.**
If she did not **- go to question 15.**

12 **How often, in the last three months of her life, did the district or community nurse visit (at the most frequent time)?**

Tick one only

☐ More than once a day

☐ Every day

☐ 2-6 times a week

☐ Once a week

☐ 2-3 times a month

☐ Less often

☐ Don't know

13 **How much of the time was she treated with respect and dignity by the district and community nurses in the last three months of her life?**

Tick one only

☐ Always

☐ Most of the time

☐ Some of the time

☐ Never

☐ Don't know

14 **Overall, do you feel that the care she got from the district and community nurses in the last three months of her life was:**

Tick one only

☐ Excellent

☐ Good

☐ Fair

☐ Poor

☐ Don't know

Care from the GP

15 **In the last three months, how often did she see the GP she preferred to see?**

Tick one only

☐ Always or almost always

☐ A lot of the time

☐ Some of the time

☐ Never or almost never

☐ She didn't try to see a particular GP

☐ She did not need to see a particular GP - **go to question 20**

16 **How much of the time was she treated with respect and dignity by the GPs in the last three months of her life?**

Tick one only

☐ Always

☐ Most of the time

☐ Some of the time

☐ Never

☐ Don't know

17 **Were you able to discuss any worries and fears you may have had about her condition, treatment or tests with the GPs in the last three months of her life?**

Tick one only

☐ I had no worries or fears to discuss

☐ Yes, I discussed them as much as I wanted

☐ Yes, I discussed them, but not as much as I wanted

☐ No, although I tried to discuss them

☐ No, but I did not try to discuss them

18 **Overall, if the GP visited her at home in the last three months, how easy or difficult was it to get him/her to visit?**

Tick one only

☐ Very easy

☐ Fairly easy

☐ Fairly difficult

☐ Very difficult

☐ She wanted the GPs to visit but they would not visit

☐ Does not apply - the GP did not need to visit

☐ Don't know

19 **Overall, do you feel that the care she got from the GP in the last three months of life was:**

Tick one only

☐ Excellent

☐ Good

☐ Fair

☐ Poor

☐ Don't know

000000

Last Care Home

20 **Did she live or stay in a care home at any time during her last three months of life?**

Tick one only

☐ Yes she was in a care home

☐ No - **go to question 24**

☐ Don't know - **go to question 24**

21 **How much of the time was she treated with respect and dignity by the staff at the last care home she stayed in?**

Tick one only

☐ Always

☐ Most of the time

☐ Some of the time

☐ Never

☐ Don't know

22 **During the last three months of her life, while she was in the care home, how well was her pain relieved?**

Tick one only

☐ Does not apply - she did not have any pain

☐ Completely, all of the time

☐ Completely, some of the time

☐ Partially

☐ Not at all

☐ Don't know

23 **Overall, do you feel that the care she got from the care home in the last three months of her life was:**

Tick one only

☐ Excellent

☐ Good

☐ Fair

☐ Poor

☐ Don't know

Last Hospital Stay

24 **Did she live or stay in hospital at any time during her last three months of life?**

Tick one only

☐ Yes

☐ No - **go to question 29**

☐ Don't know - **go to question 29**

25 **During her last hospital admission, how much of her time was she treated with respect and dignity by the hospital doctors and nurses?**

Please answer for both doctors and nurses

Doctors	Nurses	
☐	☐	Always
☐	☐	Most of the time
☐	☐	Some of the time
☐	☐	Never
☐	☐	Don't know

26 **During her last hospital admission, how well was her pain relieved?**
Tick one only

☐ Does not apply - she did not have any pain
☐ Completely, all of the time
☐ Completely, some of the time
☐ Partially
☐ Not at all
☐ Don't know

27 **Did the hospital services work well together with her GP and other services outside of the hospital?**
Tick one only

☐ Yes, definitely
☐ Yes, to some extent
☐ No, they did not work well together
☐ Don't know

28 **Overall, do you feel that the care she got from the staff in the hospital on that admission was:**

Please answer for both doctors and nurses

Doctors	Nurses	
☐	☐	Excellent
☐	☐	Good
☐	☐	Fair
☐	☐	Poor
☐	☐	Don't know

Last Hospice Stay

29 **Did she live or stay in a hospice at any time during her last three months of life?**

Tick one only

☐ Yes
☐ No - **go to question 33**
☐ Don't know - **go to question 33**

30 **How much of the time was she treated with respect and dignity by the hospice doctors and nurses?**
Please answer for both doctors and nurses

Doctors	Nurses	
☐	☐	Always
☐	☐	Most of the time
☐	☐	Some of the time
☐	☐	Never
☐	☐	Don't know

31 **During the last three months of her life, while she was in the hospice, how well was her pain relieved?**
Tick one only

☐ Does not apply - she did not have any pain
☐ Completely, all of the time
☐ Completely, some of the time
☐ Partially
☐ Not at all
☐ Don't know

32 **Overall, do you feel that the care she got from the staff in the hospice in the last three months of her life was:**

Tick one only

☐ Excellent

☐ Good

☐ Fair

☐ Poor

☐ Don't know

Experiences in the last two days of life

33 **How much of the time was she treated with respect and dignity in the last two days of her life?**

Please answer for both doctors and nurses

Doctors	Nurses	
☐	☐	Always
☐	☐	Most of the time
☐	☐	Some of the time
☐	☐	Never
☐	☐	Don't know

34 **Please look at the following statements and tick the answer box that corresponds most with your opinion about the help she received in the last two days of life**

Tick one box for each question (a-c)

	Strongly Agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree	Does not apply	Don't know
(a) There was enough help available to meet her personal care needs (such as toileting needs)	☐	☐	☐	☐	☐	☐	☐
(b) There was enough help with nursing care, such as giving medicine and helping her find a comfortable position in bed	☐	☐	☐	☐	☐	☐	☐
(c) The bed area and surrounding environment had adequate privacy for her	☐	☐	☐	☐	☐	☐	☐

000000

35 **As far as you are able to say, how much do you agree with the following statements about the overall level of care given by health and social care professionals to her in the last two days of life?**

Tick only one response per statement

a) In the last two days of life she had sufficient pain relief

☐ Strongly agree
☐ Agree
☐ Neither agree nor disagree
☐ Disagree
☐ Strongly disagree
☐ Pain relief was not needed
☐ Not sure

b) In the last two days of life she had support to eat or receive nutrition if she wished

☐ Strongly agree
☐ Agree
☐ Neither agree nor disagree
☐ Disagree
☐ Strongly disagree
☐ Food/nutrition was not needed
☐ Not sure

c) In the last two days of life she had support to drink or receive fluid if she wished

☐ Strongly agree
☐ Agree
☐ Neither agree nor disagree
☐ Disagree
☐ Strongly disagree
☐ Drink/fluid was not needed
☐ Not sure

d) In the last two days of life care and attention were given to problems apart from pain, thirst and hunger

☐ Strongly agree
☐ Agree
☐ Neither agree nor disagree
☐ Disagree
☐ Strongly disagree
☐ Does not apply
☐ Not sure

e) In the last two days of life her emotional needs were considered and supported

☐ Strongly agree
☐ Agree
☐ Neither agree nor disagree
☐ Disagree
☐ Strongly disagree
☐ Does not apply
☐ Not sure

f) In the last two days of life her spiritual and/or religious needs were considered and supported

☐ Strongly agree
☐ Agree
☐ Neither agree nor disagree
☐ Disagree
☐ Strongly disagree
☐ Does not apply
☐ Not sure

This question continues overleaf

continued from question 35 overleaf

Tick only one response per statement

g) In the last two days of life efforts were made to make sure she was in the place she most wanted to be cared for

☐ Strongly agree
☐ Agree
☐ Neither agree nor disagree
☐ Disagree
☐ Strongly disagree
☐ Does not apply
☐ Not sure

36 **Overall, how much do you agree with the following statements about communication between you and health care professionals in the last two days of her life?**

Tick one box for each question (a-c)

	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree	Don't know	Not applicable
a) I/we were kept informed on her condition and care	☐	☐	☐	☐	☐	☐	☐
b) I/we had enough time with staff to ask questions and discuss her condition and care	☐	☐	☐	☐	☐	☐	☐
c) I/we understood information provided to us	☐	☐	☐	☐	☐	☐	☐

37 **How much do you agree with the following statement?**

In the last two days of her life you had a supportive relationship with the health care professionals.

Tick one only

☐ Strongly agree
☐ Agree
☐ Neither agree nor disagree
☐ Disagree
☐ Strongly disagree
☐ Does not apply
☐ Not sure

Circumstances Surrounding Her Death

38 **Did she know she was likely to die?**

Tick one only

- ☐ Yes, certainly
- ☐ Yes, probably
- ☐ No, probably not
- ☐ No, definitely not
- ☐ Not sure

39 **In your opinion, did the person who told her she was likely to die break the news to her in a sensitive and caring way?**

Tick one only

- ☐ Yes, definitely
- ☐ Yes, to some extent
- ☐ No, not at all
- ☐ Not sure
- ☐ Does not apply - they did not know she was dying
- ☐ Does not apply - they did not tell her she was dying

40 **Were you contacted soon enough to give you time to be with her before she died?**

Tick one only

- ☐ Yes
- ☐ No
- ☐ I was already there
- ☐ It was not clear that she was going to die soon
- ☐ I couldn't have got there anyway

41 **Where did she die?**

Tick one only

- ☐ In her own home
- ☐ In the home of another family member or friend
- ☐ In a hospital ward
- ☐ In a hospital Accident and Emergency Department
- ☐ In a hospital Intensive Care Unit
- ☐ In a hospice
- ☐ In a care home
- ☐ In an ambulance on the way to hospital or hospice
- ☐ Somewhere else

42 **Did she ever say where she would like to die?**

Tick one only

- ☐ Yes - **go to question 43**
- ☐ No - **go** to **question 45**
- ☐ Not sure - **go to question 45**

43 **Where did she say that she would like to die?**

Tick one only

- ☐ At home
- ☐ In a hospice
- ☐ In a hospital
- ☐ In a care home
- ☐ She said she did not mind where she died
- ☐ She changed her mind about where she wanted to die
- ☐ Somewhere else

000000

44 **Did the health care staff have a record of this?**

Tick one only

☐ Yes

☐ No

☐ Not sure

45 **Do you think she had enough choice about where she died?**

Tick one only

☐ Yes

☐ No

☐ Not sure

☐ She died suddenly

46 **On balance, do you think that she died in the right place?**

Tick one only

☐ Yes

☐ No

☐ Not sure

47 **Were you or her family given enough help and support by the healthcare team at the actual time of her death?**

Tick one only

☐ Yes, definitely

☐ Yes, to some extent

☐ No, not at all

☐ Not sure

48 **After she died, did staff deal with you or her family in a sensitive manner?**

Tick one only

☐ Yes

☐ No

☐ Not sure

☐ Does not apply, I didn't have any contact with the staff

49 **Looking back over the last three months of her life, was she involved in decisions about her care as much as she would have wanted?**

Tick one only

☐ She was involved as much as she wanted to be

☐ She would have liked to be more involved

☐ She would have liked to be less involved

☐ She was not able to be involved

☐ Not sure

50 **Looking back over the last three months of her life, were you involved in decisions about her care as much as you would have wanted?**

Tick one only

☐ I was involved as much as I wanted to be

☐ I would have liked to be more involved

☐ I would have liked to be less involved

☐ Not sure

51 **Looking back over the last three months of her life, were any decisions made about her care that she would not have wanted?**

Tick one only

☐ Yes

☐ No

☐ Not sure

52 **Overall, and taking all services into account, how would you rate her care in the last three months of life?**

Tick one only

☐ Outstanding

☐ Excellent

☐ Good

☐ Fair

☐ Poor

☐ Not sure

53 **Since she died, have you talked to anyone from health and social services, or from a bereavement service, about your feelings about her illness and death?**

Tick one only

☐ Yes

☐ No, but I would have liked to

☐ No, but I did not want to anyway

☐ Not sure

Information About You Both

54 **What was your relationship to her? Were you her:**

Tick one only

☐ Husband / Partner

☐ Son / Daughter

☐ Brother / Sister

☐ Son-in-law / Daughter-in-law

☐ Parent

☐ Other relative

☐ Friend

☐ Neighbour

☐ Staff in care home

☐ Warden (sheltered accommodation)

☐ Other official

☐ Someone else

55 **What is your age?**

☐ 18 - 19

☐ 20 - 29

☐ 30 - 39

☐ 40 - 49

☐ 50 - 59

☐ 60 - 69

☐ 70 - 79

☐ 80 - 89

☐ 90 +

56 **Are you:**

☐ Male

☐ Female

57 **Please could you indicate which ethnic group <u>you</u> belong to:**

Tick one only

White

☐ English / Welsh / Scottish / Northern Irish / British

☐ Irish

☐ Gypsy or Irish Traveller

☐ Any other white background

Mixed / Multiple ethnic group

☐ White and Black Caribbean

☐ White and Black African

☐ White and Asian

☐ Any other mixed background

Asian / Asian British

☐ Indian

☐ Pakistani

☐ Bangladeshi

☐ Chinese

☐ Any other Asian background

Black African / Caribbean / Black British

☐ African

☐ Caribbean

☐ Any other Black / African / Caribbean background

Other ethnic group

☐ Arab

☐ Any other ethnic group

58 **Please could you indicate which ethnic group in your opinion she belonged to:**

Tick one only

White

☐ English / Welsh / Scottish / Northern Irish / British

☐ Irish

☐ Gypsy or Irish Traveller

☐ Any other white background

Mixed / Multiple ethnic group

☐ White and Black Caribbean

☐ White and Black African

☐ White and Asian

☐ Any other mixed background

Asian / Asian British

☐ Indian

☐ Pakistani

☐ Bangladeshi

☐ Chinese

☐ Any other Asian background

Black African / Caribbean / Black British

☐ African

☐ Caribbean

☐ Any other Black / African / Caribbean background

Other ethnic group

☐ Arab

☐ Any other ethnic group

59 **What was her age when she died?**

☐ 18 - 19
☐ 20 - 29
☐ 30 - 39
☐ 40 - 49
☐ 50 - 59
☐ 60 - 69
☐ 70 - 79
☐ 80 - 89
☐ 90 +

60 **What was her religion?**

☐ No religion
☐ Christian (all denominations)
☐ Buddhist
☐ Hindu
☐ Jewish
☐ Muslim
☐ Sikh
☐ Any other religion

We would be grateful if you could return your questionnaire to us in the pre-paid envelope provided
If you require a replacement envelope
or if you have any other questions
please phone the Survey Enquiry Line
on 0800 298 5313

If you feel that you would like to talk about your feelings
or discuss painful memories brought back by
completing this questionnaire, please call:

Cruse Bereavement Care 0844 477 9400
or e-mail: helpline@cruse.org.uk

☐ DECLINE - Please tick this box if you decide not to complete this survey.

If you would like to tell us why, please write any comments below.

____

____

____

____

____

000000

Please use the space below if there is anything you would like to say about the care provided. Information written here may be combined with information on other people's experiences, and provided to NHS and ONS approved researchers, to inform studies on improving end of life care. All information provided to researchers will be anonymised. To help us ensure confidentiality please do not give names of people or places.

Thank you for completing the VOICES survey, we will not ask you for any further information.

000000

Office for National Statistics

NHS England

VOICES Survey – Experiences of care in the last months of life

INFORMATION LEAFLET

If you would like to receive this information in large print, or if English is not your first language and you would like interpreter services, please call our Survey Enquiry Line on 0800 298 5313

The Survey Enquiry Line is open Monday to Thursday 9am to 9pm, Friday 9am to 8pm, and Saturday 9am to 1pm.

You are being invited to take part in a questionnaire-based research study of bereaved people called the VOICES Survey (Views Of Informal Carers – Evaluation of Services). Before you decide whether or not you would be willing to participate, it is important that you understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish.

We understand that coping with the loss of a loved one is not easy and we really appreciate you taking the time to read this information.

What is the purpose of the VOICES Survey?

VOICES is a survey of bereaved carers who provided support and care to a relative, partner or friend. It covers experiences in the last months of life and will be used nationally to monitor and improve services provided. Although participation in VOICES will not help you directly, we hope that the information you give us will enable us to improve people’s experiences of care at the end of their lives and improve services provided to bereaved relatives and friends.

Why have I been chosen?

You have been chosen to take part because you registered a death in the past year. The Office for National Statistics (ONS) will not share any personal information about you with anyone. Your survey responses will be shared with NHS England and their approved researchers using only an anonymous Study ID number. This ensures that the information you provide is totally confidential, in accordance with the Data Protection Act 1998.

What will taking part involve?

We would like you to fill in a questionnaire which will take around 30 minutes. It asks about the care and support both you and your relative/friend received in the last months of their life and whether their needs were fully met. Your experiences are very important, so please feel free to be completely open and honest.

If you do not think you are the best person to complete the questionnaire, please pass it on to whoever you feel would be the best person to complete it. If you would like to complete the questionnaire with the help of your family or a group of friends, you are welcome to do so.

Most of the questions can be answered by ticking the most appropriate box. If you make a mistake or wish to change your answer, simply cross through the answer you do not want. If you would prefer not to answer a question, please go on to the next one. We would be very grateful for any additional comments that you would like to make in the spaces provided.

To return the completed questionnaire, please use the enclosed pre-paid envelope. You can request a replacement envelope by phoning the Survey Enquiry Line on 0800 298 5313. The closing date for returning completed questionnaires is **Friday 11th December 2015**.

Do I have to take part?

Taking part is completely voluntary. If you do decide to take part you may change your mind and choose not to continue in the research at any time, without having to give a reason for doing so. If you decide not to participate, you can let us know by ticking the box on the back of the questionnaire and returning it to ONS in the pre-paid envelope. This will ensure that you do not receive reminder letters.

What are the possible disadvantages of taking part?

Some people find it distressing to think about the care that their loved ones or close friends received during the last months of their lives. Answering questions about care at the end of life can bring back painful memories. If you find it distressing, you can stop completing the questionnaire at any time and choose not to continue. Please let us know if you do not wish to participate, to avoid us sending you any reminder letters.

We are working with Cruse Bereavement Care services, a charitable organisation that provides help and support to those who have lost loved ones. If you feel that you would like to talk about your feelings or discuss painful memories brought back by completing the questionnaire, please contact Cruse Bereavement Care by telephone on 0844 477 9400 or by e-mail at helpline@cruse.org.uk

How will the information I give be kept confidential?

ONS will not give personal information that identifies you to anyone else. Your survey responses will only be identified by an anonymous Study ID number. All the information collected will be kept strictly confidential and secured against unauthorised access. We would also like to make absolutely clear that no information that could identify you will be used in any reports or articles we write.

If you add comments at the end of the questionnaire this may be used by approved researchers, to inform studies on people's personal experiences. Please try not to write the names of any people or places that may identify you. Any names written will be removed to protect your identity. You will not be asked to include your name (or the name of your relative, partner or friend) on the questionnaire.

The information collected will be retained electronically and securely stored for 10 years. Paper forms are disposed of securely.

How is it that ONS hold death registration information?

The General Register Office, part of Her Majesty's Passport Office (HMPO), shares death registration information with ONS on the basis that it is not disclosed to anyone in an identifiable format without the individual's specific permission. HMPO's data sharing principles can be found at https://www.gov.uk/government/publications/ips-privacy-policy. The legal basis by which ONS collects and holds data is set out in the Statistical Registration and Service Act 2007.

Who is funding and organising the study?

The VOICES study is funded by NHS England and run by the Office for National Statistics.

What will happen to the results of the study?

Information obtained from the questionnaire will be entered into a database and analysed by ONS, NHS England and our approved researchers. At the end of the project, the findings will be presented in reports and tables. Results from the last survey can be found at the following link:

http://www.ons.gov.uk/ons/rel/subnational-health1/national-survey-of-bereaved-people--voices-/2014/index.html

The information we get from this project will help to improve the quality of end of life care provided in England.

If you would like more information about the study, please call our Survey Enquiry Line on 0800 298 5313.

Thank you again for taking the time to read this information. We are confident that this study will make a difference to improving the way that care is delivered to people at the end of their lives.

National Bereavement Survey (VOICES), Office for National Statistics, Room 1.059, Government Buildings, Cardiff Road,
NEWPORT NP10 8XG

UK Statistics Authority

National Statistician's Data Ethics Advisory Committee

Application for Ethical Review

The Application Process

This is an application form for applying for ethical review from the National Statistician's Data Ethics Advisory Committee (NSDEC). You should use the additional guidance when completing this form.

The application form should be completed in ***plain English*** which is understandable to lay members and all abbreviations should be explained the first time they are used. The form should contain sufficient information to ensure a thorough ethical review can take place.

Please word process the form using Arial or Times New Roman font, size 11. Where necessary expand text boxes on the form to accommodate answers, but ensure word counts are adhered to where specified.

Where sections are not relevant to your study please mark as N/A.

On completion the responsible owner should sign the application form and send to:
nsdec@statistics.gsi.gov.uk

Section A
Application Details

A1	**Responsible Owner**

Full Name: [redacted]	**Position:** Professor of Statistical Epidemiology
Address: [redacted]	**Email:** [redacted]
	Telephone: [redacted]
	Organisation: University of Oxford

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.

Print Name: [redacted]
Signature:
Date 20 January 2016

UK Statistics Authority

A2	**Applicant Details** (if applicant is not the responsible owner)

Full Name	**Position**
Address:	**Email:**
	Telephone:
	Organisation:

A3	**Project Information**

Project Title: A national survey of mother and infant health	
Start Date: 1st April 2016	**End Date:** 31st December 2017
Project Sponsor (select all that apply)	
☐ ONS ☐ ADRN ☐ GSS ☐ Collaboration ☑ Other (Please specify).................................Department of Health (Policy Research Programme)	

A4	**Collaboration and Sponsors**

List of Collaborators/Sponsors	**Details and relevant documentation relating to collaboration** (you may attach copies of relevant documentation)
The study will be sponsored by the University of Oxford. The study will be led by Maria Quigley and a small research team from NPEU (National Perinatal Epidemiology Unit). As stated in section B4, an Advisory Group will be formed to give input on the design, conduct and analysis of the study, and will help with dissemination. There will not be any other collaborators.	

UK Statistics Authority

A5	**Proposed Site of Research** (select all that apply)

Where will the research take place?

☐ ONS

☐ VML

☐ HMRC Data Lab

☐ ADRC-England

☐ ADRC-Northern Ireland

☐ ADRC-Scotland

☐ ADRC-Wales

☑ Other (please specify)……………………………………National Perinatal Epidemiology Unit, University of Oxford.

Is this a secure site?

☑ Yes ☐ No

Section B

Project Details

B1	**Please provide a brief high level summary of the research giving necessary background** ***(max 500 words)***

The postnatal period is an important time of transition for mothers, babies and their families. The first six months in particular, is a critical period of transition, during which many mothers are adapting to parenthood, need family and social support, and are planning their return to work. Infant feeding patterns change dramatically during this period, most notably due to the introduction of solids. Mothers are also particularly at risk of experiencing mental health problems during this period.

In England, recently delivered women have only a few routine postnatal checks and most women will have no routine checks after the first 6-8 weeks. Infants have several routine health checks in the first 6-8 weeks, after which point they are not routinely reviewed again until an assessment at 9-12 months, except for their routine immunisations (at 2, 3, 4, 12-13 months and beyond). Hence, there is very little in the way of routine monitoring and collection of information about the health and wellbeing in mothers and their babies in the postnatal period between 6-8 weeks and 9-12 months.

As part of the Department of Health Policy Research Unit for Maternal Health and Care programme of work, NPEU propose to conduct a national survey of mother and infant health in England in 2016-17.

A random sample of up to 20,000 women giving birth in England will be drawn by ONS from birth registration records. The sample selection will be over two specified time periods, with a postal questionnaire sent to the women at around 6 months following the birth. The proposed time periods are September - October 2016 (when infants born in March - April 2016 will be 6 months old) and January - February 2017 (when infants born in July - August 2016 will be 6 months old).

The questionnaire will be different to the one used in the NPEU National Maternity Surveys, in that it will focus more on postnatal women’s health and wellbeing and that of their baby and their access to health care. Thus, the women will be asked about their own health, their baby’s health and use of services in the six months following birth. The women will also be asked about smoking, infant feeding and other health-related behaviours and about their plans to (or experiences of) return to work. As these data items are not currently collected at this time point, either routinely or as part of regular surveys, these data will provide unique information, which can be used to help improve postnatal monitoring or support and can help guide policies in the postnatal period, such as routine postnatal checks, weaning and returning to work.

B2	**Data Use**

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)*		Birth Registration Data (only to be used by ONS) Death Registration Data (only to be used by ONS when screening for infant deaths)		Birth registration data* in responders and non-responders (to be used by NPEU)

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)*				

* During 2 specified short periods in 2016. The following data fields will be required:

- Baby's sex
- Age group of mother and father (16-19, 20-24, 25-29, 30-34, 35-39, 40+)
- Registration status (married, sole, joint with same address, joint with different address)
- Index Multiple Deprivation quintile (1, 2, 3, 4, 5)
- Mother's National Statistics Socio Economic Classification (as already coded, available in 10% only)
- Government Office Region (as already coded, North East, North West etc)
- Country of birth

B3	**How will information be kept confidential and data kept secure?** ***(max 500 words)***

ONS will select the sample and send out the questionnaires. The sample will be held on a secure server, accessible only by authorised staff. The survey will be posted to respondents, enclosing only their name and address. The literature will reassure the respondent that their personal details will not be shared with NPEU or others.

Limited, anonymised data on respondents and non-respondents will be made available to NPEU for comparison with respondents and for calculation of survey weights. This will be transferred electronically from ONS to the NPEU in password protected form.

Completed questionnaires will be returned by the women to the NPEU and in the short-term, will be stored in locked filing cabinets. The questionnaires will then be given to a professional company, who will scan the questionnaires and enter the data. These electronic data and scanned forms will be transferred to the NPEU in password protected form and stored within NPEU. The questionnaires (hard copies) will also be sent back to the NPEU (by courier), where they will be stored in locked filing cabinets within the NPEU, before being destroyed.

The professional company will sign confidentiality agreements before doing any of this work.

For those completing questionnaires online, the secure NPEU/maternity survey website will only allow access by individual URN and individual password (contained in the individual invitation letters), for both online and phone access, and will be securely managed by the external contractor responsible for electronic data capture from the questionnaires.

The following physical arrangements are in place: the NPEU is located in the Richard Doll Building at the University of Oxford Old Road Campus, Headington, Oxford. The building requires electronic card token access by authorised and authenticated users only. The NPEU offices also require electronic card token access which is given to only authorised members of NPEU staff. The building is situated at the University of Oxford Old Road Campus in Headington, which has 24/7 onsite security by University of Oxford security staff.

The NPEU shared data is held on 2 logical Novell volumes accessible by members of the unit only if they are granted specific trustee rights to that data. The physical servers and their storage arrays are secured at all times. When the data is no longer required, it is destroyed using a Netware process called shredding, and can no longer be retrieved or salvaged. Any backups which have been taken off the data are then overwritten in due course, according to the retention period associated with them.

B4	**Please provide details of the research protocol or methodology** *(e.g. data linkage, web scraping etc)* *(max 500 words)*

An advisory group will be formed to provide input to content, design, conduct and analysis of the study, and will help in disseminating the findings. This group will include members of user groups and their representatives, enabling views of the public to be considered. Members of professional bodies and experts in maternal and child health will also be included.

A random sample of approximately 20,000 women giving birth in England will be drawn by ONS from birth registration records. The sample selection will be over two specified time periods with a postal questionnaire sent to women around 6 months postnatally. The proposed time periods are September - October 2016 (for infants born in March - April 2016) and January - February 2017 (for infants born in July - August 2016).

The women will be asked about their own health care and use of services in the six months following birth (using validated questions from the 2014 National Maternity Survey); their baby's health, infant feeding, smoking, and plans to (or experiences of) return to work (using validated questions from the 2010 UK Infant Feeding Survey); and their mental health and wellbeing, using a standard measure. The women will be offered the option of completing the questionnaire online.

ONS will use birth registrations data to identify a sample of women and send a postal questionnaire, together with an invitation letter and information leaflet.

Completion and return of questionnaires will be taken as implicit consent. Once completed, questionnaires, identifiable only by a unique reference number, will be returned to NPEU and then to the contractor for data entry. ONS will be informed about returned questionnaires and refusals to prevent reminders being sent.

Reminder letters will be sent by ONS to non-responders two weeks after the first mailing. Two weeks later, a further questionnaire will be mailed, and two weeks after that, another letter may be sent to non-respondents. If no response is obtained, no further contact will be made. In the final reminder, women will be offered the choice of completing the previously sent questionnaire, or a very short version. NPEU are exploring the possibility of being able to complete the short version using a mobile phone.

Prior to each mail out, ONS will use death registration data to screen for infant deaths; therefore, if a child has been sampled and subsequently dies they will be removed from the sample.

ONS will be asked to provide anonymised individual-level data on responders and non-responders, so NPEU can compare responders and non-responders, calculate survey weights to adjust for non-response and apply weights in the analysis of the responders.

Data analysis will include standard summary statistics and statistical regression models (using survey weights) for each topic. Where appropriate, results will be compared with those from similar surveys (e.g. Infant Feeding Surveys) and trends over time explored. Finally, response rates will be compared according to different methodologies (e.g. short versus long questionnaire; option of text message response to key questions).

B5	**Please outline the proposed benefits of the project** *(max 500 words)*

The project is part of a programme of research, which is funded by the Department of Health Policy Research Programme in Maternal Health and Care. Our advisors at the Department of Health have prioritised this survey as one of the important projects in our current programme of research on maternal health and care. From a policy and care perspective, it is important to collect data on maternal and infant health in the first year after birth. Infant weaning and development, mothers returning to work and women’s mental health during this time are key topics and data on these are not currently collected, either routinely or as part of regular surveys. Hence, these data will provide unique information, which can be used to help improve monitoring and support for recent mothers and can help guide policies in the months following childbirth and the transition to parenthood.

There are unlikely to be major direct benefits of the project to the women in the study, although some women welcome the opportunity to write about their experiences and may find some benefit from doing so. However, as stated above, this research is highly policy relevant and the impact of this research will have benefits to those women having subsequent children and to those giving birth in the future.

B6	**Please outline the ethical issues that might arise from the proposed study and how they will be addressed** ***(all research projects have some ethical considerations, so this section must not be left blank)***

Receiving the survey could be deemed to be intrusive

Participation in the survey will be entirely voluntary. For most women, the burden of participation will be limited to completion of a questionnaire. However, a small number of women may find the questionnaire intrusive. The Information leaflet and letter will explain that participation is entirely voluntary and that women can return a blank questionnaire, indicating that no further contact should be made if this is what they prefer. The Information leaflet and questionnaire will also recommend that if they are distressed they should consult their GP or health visitor. Prior to each mailing, the birth cohort will be screened against the death civil registrations. If a child is identified as now being deceased, then they are removed from the sample, in order to alleviate the possibility of causing distress to non-respondents.

The use of data has clear benefits for users and serves the public good
Our results have the potential to lead to improved services or policies related to the health and well-being of mothers, their infants and their families during the postnatal period. For example, data from the Infant Feeding Surveys between 1975 and 2010 showed that since 2000, there has been a trend towards introducing solids at older ages. The most recent national data to be collected was the 2010 Infant Feeding Survey, in which 75% of babies had started solids by 5 months, including 30% who had started solids by 4 months. This is in contrast to guidance from World Health Organisation, National Institute for Health and Care Excellence and the Department of Health, who recommend that solids should be introduced at 6 months. Our data will be the first national data to show whether the trend towards delaying the introduction of solids has continued for babies born in 2016. This will inform national policy on infant feeding. Similarly, the data we collect on mental health will identify the burden of disease at 6 months postnatally, what factors are associated with mental health problems at this time and whether the support given is adequate and timely.

New technologies
Our research does not use new technologies. The only aspect of our study that could be regarded as new technology is that we are considering whether it is feasible to develop a short version of the questionnaire, which can be entered and submitted using a mobile phone.

Legal considerations
The NPEU is covered by the University of Oxford Data Protection Act registration (registration number: Z575783X). However, the study data will be anonymised and non-identifiable within the terms of the Data Protection Act 1998.

Views of the public
An advisory group will be formed to provide input to content, design, conduct and analysis of the study, and will help in disseminating the findings. This group will include members of user groups and their representatives, enabling views of the public to be considered. Members of professional bodies and experts in maternal and child health will also be included.

The access, use and sharing of data is transparent
The research data that will be held in the NPEU will be anonymised and held securely, as described (B3), and this will be communicated in the study information leaflet and on the NPEU/maternity survey website.

B7	**How will the findings of the research be disseminated?**

The main outputs of the project will be journal articles, conference presentations, and a report. These will be produced by NPEU and disseminated to different audiences using a variety of methods:

1. Research communities will be reached by publishing findings in appropriate journals and presenting at conferences. The conferences will be chosen depending on what the key findings are and who should be targeted. Possible conferences include the Society for Infant and Reproductive Psychology annual conference (for maternal mental health), Nutrition and Nurture in Infancy and Childhood Conference (for infant feeding), the Unite/Community Practitioners and Health Visitors Association Annual Professional Conference 2015 (targeting health visitors) and Society for Social Medicine (targeting researchers in public health).

2. The journal articles will be 'open access'.

3. A report of the key findings will be available for stakeholders, such as the participants, pregnant or recently delivered women and their families, and the professionals and user groups who support them, or provide commission services. The report will be available to download on the NPEU website. The Information Leaflet will explain this and give the link to the website at the time of recruitment.

4. The researchers will also use other channels for dissemination, as appropriate, including, the Press Offices of Oxford, via briefings, the NPEU website and Twitter account.

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.**

The electronic dataset and scanned forms will be held by the NPEU research team on secure servers. The questionnaires will be stored in the NPEU in locked filing cabinets in the short term and then they will be destroyed. It has been the policy of the NPEU, as a Department of Health funded Policy Research Unit since 1978, to hold research data in perpetuity where this is permitted. On completion of the study and following publication of the study results, datasets will be encrypted and archived in a secure electronic archive. Future access will be controlled by the data custodian (Director of the NPEU; currently Prof Jenny Kurinczuk) and would be subject to further regulatory approvals, should access be required for any purpose other than that outlined in this protocol. The NPEU is covered by the University of Oxford Data Protection Act registration (registration number: Z575783X). However, the study data (electronic and questionnaires) will be anonymised and non-identifiable within the terms of the Data Protection Act 1998.

UK Statistics Authority

Section C
Details of Data Subjects

C1	**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:

A random sample of mothers of babies born in two time periods in 2016 (see B1).
Mothers aged under 16 years at birth registration will be excluded.
Mothers of babies who have died before 6 months of age will be excluded.

Justification for focusing on these subsections or groups:

We are interested in mother and infant health around 6 months postnatally and, therefore, we are focusing on a random sample of eligible women.

As with previous maternity surveys, there would be ethical concerns in contacting women whose baby had died, or very young mothers (under 16 years). Hence, these groups of women will be excluded.

C2	**Please detail consent given to use data specified in section B2**

Lawful gateway to use data in this way
Section 42 (S42) of the Statistics and Registrations Service Act (SRSA) provides ONS with cover to receive the data and to use it for any statistical function. Section 22 of the SRSA provides cover for us to use the data for statistical services to any person inside or outside the United Kingdom. ONS Legal Services have confirmed that this is a lawful use of the data.

Fair Processing
The issue of 'Fair Processing' has also been investigated. ONS Legal Services have again confirmed ONS is currently meeting Fair Processing principles by informing potential survey participants at the outset how ONS has received their information and by taking an 'opt in' approach to respondent participation, rather than an 'opt out' approach. In other words, once contacted about the survey, those sampled must opt in to actually take part. They do this by returning their agreement to participate either to ONS or to the third party researcher, depending on survey design.

Questionnaire return will be taken as implicit consent. The Invitation letter and Information leaflet will make it clear that participation is entirely voluntary and that the information they disclose in the questionnaire will be anonymous. These documents will also make it clear that the data specified in B2 may also be made available to the researchers, but that these data will be completely anonymised and it will not be possible to identify individual women.

C3	**If you are using data held by a third party please detail how you will obtain this**

The only data used will be from birth registrations (via ONS) and from the survey questionnaires (which will be mailed out by ONS).

Annex C – Relevant Sections of the Statistics Registration and Services Act

Section 42 Information relating to births and deaths etc

(1)The Registrar General for England and Wales may, for the purpose of the exercise by the Board of any function, disclose to the Board any information to which this section applies.

(2)This section applies to—

(a)any information entered in any register kept under the Births and Deaths Registration Act 1953 (c. 20);

(b)any other information received by the Registrar General in relation to any birth or death;

(c)any information entered in the Adopted Children Register maintained by the Registrar General under section 77 of the Adoption and Children Act 2002 (c. 38);

(d)any information entered in any marriage register book kept under Part 4 of the Marriage Act 1949 (c. 76);

(e)any information relating to a civil partnership which is recorded under the Civil Partnership Act 2004 (c. 33) at the time of the formation of the civil partnership.

[F1(ea)any information recorded under section 9 of the Marriage (Same Sex Couples) Act 2013 and regulations made under that section on the conversion of a civil partnership into a marriage;**]**

[F2(f)any other information received by the Registrar General in relation to any marriage or civil partnership.**]**

(3)In subsection (2)(b) “birth” has the same meaning as in the Births and Deaths Registration Act 1953.

[F3(4)The Board may disclose to a person mentioned in subsection (4A) any information referred to in subsection (2)(a) to (c) which is received by the Board under this section, or any information which is produced by the Board by analysing any such information, if—

(a)the information consists of statistics and is disclosed for the purpose of assisting the person in the performance of functions exercisable by it in relation to the health service, or

(b)the information is disclosed for the purpose of assisting the person to produce or to analyse statistics for the purpose of assisting the person, or any other person mentioned in subsection (4A), in the performance of functions exercisable by it in relation to the health service.

(4A)Those persons are—

(a)the Secretary of State,

(b)the Welsh Ministers,

(c)the National Health Service Commissioning Board,

(d)a clinical commissioning group,

(e)a local authority,

(f)a Local Health Board,

(g) an NHS trust established under section 18 of the National Health Service (Wales) Act 2006,

(h)the National Institute for Health and Care Excellence,

(i)the Health and Social Care Information Centre,

(j)a Special Health Authority,

(k)the Care Quality Commission, and

(l)such other persons as the appropriate authority may specify in a direction given for the purposes of this section.

(4B)For the purposes of subsection (4A)(l), the appropriate authority is—

(a)in relation to a direction to be given for purposes relating only to Wales, the Welsh Ministers, and

(b)in any other case, the Secretary of State.**]**

(5)Any information disclosed under subsection (4) may be disclosed in such form as the Board considers appropriate for the purpose specified in that subsection.

[F4(5A)A direction under subsection (4A)(l) must be given by an instrument in writing.

(5B)Sections 272(7) and 273(1) of the National Health Service Act 2006 apply in relation to the power of the Secretary of State to give a direction under subsection (4A)(l) as they apply in relation to powers to give a direction under that Act.

(5C)Sections 203(9) and 204(1) of the National Health Service (Wales) Act 2006 apply in relation to the power of the Welsh Ministers to give a direction under subsection (4A)(l) as they apply in relation to powers to give a direction under that Act.**]**

(6)In subsection (4) “health service” has the same meaning as in the National Health Service Act 2006 (c. 41).

[F5(7)In subsection (4A)—

- “ clinical commissioning group ” and “ Special Health Authority ” have the same meaning as in the National Health Service Act 2006;
  “ local authority ” has the same meaning as in section 2B of that Act of 2006. **]**

Section 22 Statistical services

(1)The Board may provide statistical services to any person in any place within or outside the United Kingdom.

(2)The services which may be provided under this section include in particular—

(a)providing information, advice and technical assistance in relation to statistics;

(b)providing quality assessment in relation to statistics;

(c)conducting statistical surveys and analysis;

(d)collecting, adapting and developing data.

Annex D – Relevant extracts from the Data Protection Act

The Data Protection Act

The Data Protection Act requires you to process personal data fairly and lawfully. The requirement to process personal data fairly and lawfully is set out in the first data protection principle and is one of eight such principles at the heart of data protection. The main purpose of these principles is to protect the interests of the individuals whose personal data is being processed. They apply to everything you do with personal data, except where you are entitled to an exemption.

First data protection principle

In practice, it means that you must:

i. have legitimate grounds for collecting and using the personal data
ii. not use the data in ways that have unjustified adverse effects on the individuals concerned;
iii. be transparent about how you intend to use the data, and give individuals appropriate privacy notices when collecting their personal data;
iv. handle people’s personal data only in ways they would reasonably expect;
v. make sure you do not do anything unlawful with the data.

What does fair processing mean?

i. Processing personal data must above all else be fair, as well as satisfying the relevant conditions for processing. “Processing” broadly means collecting, using, disclosing, retaining or disposing of personal data, and if any aspect of processing is unfair, there will be a breach of the first data protection principle – even if you can show that you have met one or more of the conditions for processing.
ii. Fairness generally requires you to be transparent – clear and open with individuals about how their information will be used. Transparency is always important, but especially so in situations where individuals have a choice about whether they wish to enter into a relationship with you. If individuals know at the outset what their information will be used for, they will be able to make an informed decision about whether to enter into a relationship, or perhaps to try to renegotiate the terms of that relationship. Assessing whether information is being processed fairly depends partly on how it is obtained. In particular, if anyone is deceived or misled when the information is obtained, then this is unlikely to be fair.
iii. The Data Protection Act says that information should be treated as being obtained fairly if it is provided by a person who is legally authorised, or required, to provide it.

Annex E

Privacy Notice

Personal information which you are required by law to provide for a registration will be kept by the relevant local registration officer. The local registration officer to whom you supply information will also send a copy of this information to the General Register Office (GRO) for England and Wales so that a central record of all registrations can be maintained.

A copy of any register entry will be provided to any applicant, provided that they supply enough information to identify the entry concerned and pay the appropriate fee. The copy may only be issued in the form of a paper certified copy (a “certificate”). An application for a certificate may be made to either the Local Register Office or to the GRO.

The GRO makes indexes, for the central record of registrations, publicly available in order to help members of the public identify the registration they might need. The Register Office also makes a local index available for this purpose.

The Local Register Office may also choose to make the information contained within local indexes available on line. This will be done in order to help members of the public identify the registration they might need. Any information placed on line must be done in manner which is compliant with the Data Protection and Human Rights Act.

Additionally, confidential information for statistical purposes which you are required by law to give to the local registration officer, and other information provided voluntarily, will be passed to the UK Statistics Authority for the preparation and supply of statistics.

As well as providing certificates, local registration officers and the GRO may make registration information available to other organisations, for the following purposes:

1. Statistical or research purposes
2. Administrative purposes by official bodies e.g. ensuring their records are up to date
3. Fraud prevention or detection, immigration and passport purposes

For further information on data held by the Registrar General visit

https://www.gov.uk/government/publications/ips-privacy-policy

Staff at this local registration office will be able to provide further information on data held by the registration service.

UK Statistics Authority

National Statistician's Data Ethics Advisory Committee

NSDEC(16)03

Twitter pilot

Purpose

1. This paper presents a proposal from the ONS Big Data team, for ongoing research using geo-located Twitter data. The aim of the research is to investigate and establish patterns of mobility.

Recommendations

2. Members of NSDEC are invited to consider the project application at **Annex A** and advise the National Statistician to:
   - i. approve the research and allow it to proceed;
   - ii. approve the research subject to minor revisions;
   - iii. recommend major revisions to the research and request the proposal be resubmitted to a future meeting once implemented; and
   - iv. reject the research advising that the research be stopped from proceeding.

Background

3. ONS is exploring alternative data sources, which could improve the accuracy, frequency and timeliness of population estimates.
4. Population estimates are widely used in the planning and provision of services, such as building schools, railways and hospitals. It is therefore, key that policy makers have as accurate information possible if services are to be fit for purpose.
5. Currently, internal migration, a key component of population estimates, is measured by looking at changes in General Practitioner (GP) registrations between periods and aggregating these to measure total flows in to and out of areas, such as local authorities.
6. Whilst this method works well when estimating population changes for families and the elderly, younger adult groups, in particular students and young adult males are more difficult to measure. This is because these groups are less likely to re-register with a GP when changing address.
7. As a quarter of Twitter users are aged 18 to 24, and most users are likely to tweet more regularly than interacting with a GP, Twitter could be a credible source for improving internal migration estimates, and thus population estimates. Twitter data could also be used to quality assure estimates of moves in to and out of areas.
8. Whilst Twitter data is publicly available and ONS may legally access it (so long as Twitter's 'rules of the road' are adhered to), it is unclear as to what extent Twitter account holders understand about their data are being shared.

Adil Deedat, NSDEC Secretariat, UK Statistics Authority, 14 January 2016

List of Annexes

Annex A Application: Twitter Pilot, ONS Big Data, Ms Jane Naylor, 19 January 2016

National Statistician's Data Ethics Advisory Committee

Application for Ethical Review

The Application Process

This is an application form for applying for ethical review from the National Statistician's Data Ethics Advisory Committee (NSDEC). You should use the additional guidance when completing this form.

The application form should be completed in ***plain English*** which is understandable to lay members and all abbreviations should be explained the first time they are used. The form should contain sufficient information to ensure a thorough ethical review can take place.

Please word process the form using Arial or Times New Roman font, size 11. Where necessary expand text boxes on the form to accommodate answers, but ensure word counts are adhered to where specified.

Where sections are not relevant to your study please mark as N/A.

On completion the responsible owner should sign the application form and send to:
nsdec@statistics.gsi.gov.uk

UK Statistics Authority

Section A
Application Details

A1	Responsible Owner

Full Name: [redacted]	**Position: ONS Big Data Project lead**
Address: **ONS** **Segensworth Road** **Fareham** **Hants** **PO15 5RR**	**Email:** [redacted] **Telephone:** [redacted] **Organisation: ONS**

Declaration to be signed by the responsible owner

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

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

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

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

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

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

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

Print Name: [redacted]
Signature: [redacted]
Date: 19 January 2016

UK Statistics Authority

A2	**Applicant Details** (if applicant is not the responsible owner)

Full Name	**Position**
Address:	**Email:** **Telephone:** **Organisation:**

A3	**Project Information**

Project Title: **ONS Big Data team – Twitter pilot**

Start Date: 01/01/14	**End Date: 31/03/16 – although could continue after this date – dependent on funding**

Project Sponsor (select all that apply)

☑ ONS

☐ ADRN

☐ GSS

☐ Collaboration

☐ Other (Please specify)..................................

A4	**Collaboration and Sponsors**

List of Collaborators/Sponsors	**Details and relevant documentation relating to collaboration** (you may attach copies of relevant documentation)
Cardiff University (COSMOS – Collaborative Online Social Media Observatory) Southampton University (Social Science and Web Science)	Although no formal collaborative arrangements are in place we have engaged with a number of different experts in this field, as listed. The aim of this engagement is to share expertise and to provide technical support and quality assurance for our work.

UK Statistics Authority

A5	**Proposed Site of Research** (select all that apply)

Where will the research take place?

☑ ONS

☐ VML

☐ HMRC Data Lab

☐ ADRC-England

☐ ADRC-Northern Ireland

☐ ADRC-Scotland

☐ ADRC-Wales

☐ Other (please specify)..

Is this a secure site?

☑ Yes ☐ No

Section B

Project Details

B1	**Please provide a brief high level summary of the research giving necessary background** ***(max 500 words)***

Twitter is a micro-blogging platform where users post short messages, or “tweets”, with a limit of 140 characters. Users tweeting from a Smartphone, or other devices providing location services, may choose to provide a precise GPS location. These are referred to as geo-located tweets. Although less than two percent of tweets are geo-located, the volumes of data are still considerable, with hundreds of thousands of such tweets being sent every day within Great Britain.

Methods have been developed to infer a user's place of residence from geo-located tweets. Analysis over time indicates changes in residence and aggregating these changes over areas provides estimates of the number of users moving in and out of local authorities. This provides intelligence about population mobility in these areas.

The focus for the analysis of geo-located tweets has been on internal migration (moves between local authorities), as this is a key component of change for sub-national population estimates in the United Kingdom. At present, the main source of internal migration is the GP patient register. However, a well documented issue with the GP patient register is that students and young men in particular, are less likely to re-register with a GP when they change address, compared with the general population. In 2014, 9% of the population were between the ages of 18 and 24. This group makes up almost a quarter of all Twitter users (eMarketer, 2015). Twitter is not representative of the population. However, higher proportions of young people use Twitter in comparison to older people and Twitter is more popular with certain socio-economic groups. Therefore, the premise is that geo-located Twitter data could be particularly useful for gaining insight into student age migration.

Work has also been undertaken to better understand what type of people are represented in the sample of geo-located Tweets and to develop methods to produce more representative estimates to understand and analyse mobility patterns.

B2	**Data Use**

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)*		Twitter data (all geo-located Tweets from 1 April to 31 October 2014)		
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)*	2011 Census data used for comparative analysis			
Other *(please specify e.g. Ordinance Survey Address register in the relevant options adjacent)*				

B3	**How will information be kept confidential and data kept secure?** ***(max 500 words)***

Data was collected on all geo-located tweets sent within Great Britain over a seven month period (1 April to 31 October 2014). This involved collecting data through a combination of real-time collection through the Twitter API [1]and procurement of a bulk point in-time extract (due to the large volumes of tweets required). Although at an individual level and potentially identifiable (through username and user bio) this data is public and can be accessed through the public Twitter API. The decision was made and agreement sought from GNIP (a reseller of data from whom ONS purchased the bulk extract) that the data could be held and analysed within the ONS Innovation Lab[2]. The lab has been designed to only hold public or open data and hence does not have the same level of security as the standard ONS network. However it is a private network with a standard broadband connection and firewall. The data used for this project was password protected and only accessible by the team.

The Census aggregates used within this analysis were also public data and therefore required no additional security and hence analysis could be undertaken within the labs.

---

[1] Application Programming Interface – the Twitter API facilitates public access to Twitter data

[2] An environment that has been set up to help facilitate research into new technologies and open source tools, new sources of public or open data and to develop associated skills. Labs have been set up on both the Newport and Titchfield ONS sites and are completely separate from the main ONS network and therefore provide a route for easily accessing open source tools without compromising ONS security.

B4	**Please provide details of the research protocol or methodology *(e.g. data linkage, web scraping etc)* *(max 500 words)***

Following collection of geo-located tweets for the seven month period a number of processes were run to create a final clean version of the data, e.g. removal of Twitter accounts not related to individuals but say businesses, removal of non-GB tweets etc. Preliminary analysis was then undertaken e.g. distribution of volumes of tweets by user, investigation of user persistence.

Mobility analysis

Having obtained a clean data set of tweets, the next step is to organise them into a framework that will support analysis of mobility patterns. The broad approach is to cluster tweets by location[3], identify those clusters that are in residential areas[4] and define the residential cluster for each user with the highest number of tweets as being the most likely location of usual residence. This location is referred to as the *dominant residential cluster*. It is proposed that the dominant residential cluster can also be calculated for different time periods, for example, by month. Any changes in the dominant residential cluster across time would signal a de facto change in residence.

A number of analyses were then made of the resulting clusters, including aggregate comparisons with 2011 Census data and a specific analysis of student mobility.

The next stage of the work involves developing methods to produce representative estimates from the geo-located tweets. This has involved 3 steps:

- Where possible demographics (age, gender, occupation, location) are inferred from the Twitter user's metadata. Text string analysis is used to pull out relevant numbers and words which can then be matched to key demographics of interest with a reasonable degree of confidence.
  For example:

Twitter metadata				Inferred demographics			
Username	Personal description	Name	Location	Sex	Age	Location	Occupation
DarkKnightDan	21.Proud supporter of Exeter City and Manchester United. Ferrari in F1 are my team. Uni film student and avid gamer.	Daniel Way	Devon	M	21	Devon	Student
DanthemanHilder	Director, buisness owner and gold medalist in life.	Dan Hilder	Hastings	M	-	Hastings	Business Director

---

[3] Using DBSCAN, a spatial clustering algorithm

[4] Using AddressBase the definitive source of address information for Great Britain

fever_x	21. London. Psychology & Criminology graduate. Blair Waldorf wannabe.	Ellie Fever	Ealing, London	F	21	Ealing	-
EveKirwan	Its never wot it seems!! Hometown Dublin Nursing in London NHS....	EvvvE	London	F	-	London	Nurse

So the inferred demographics in the table above have been derived manually. A larger training data set that has been classified in this way is then used to develop the algorithms or models that are applied to the full data set.

- Questions have been run for 3 months on the ONS ad-hoc Opinions Survey – this has provided a pooled responding sample of size of just under 3000 individuals. Respondents were asked about their frequency and use of Twitter. Key demographics are also collected on all respondents.

- The information collected/derived from the 2 steps above are combined within an estimation framework to produce more representative estimates from Twitter.

Although individual level data are required to derive the clusters and demographics, this analysis is solely concerned with aggregate patterns.

More detail on the methodology, analysis and results are provided in the report linked here http://www.ons.gov.uk/ons/about-ons/who-ons-are/programmes-and-projects/the-ons-big-data-project/index.html

B5	**Please outline the proposed benefits of the project** ***(max 500 words)***

As mentioned in section B1, the current main source for analysing migration patterns in the United Kingdom is the GP patient register. It is known that the GP patient register is not necessarily a reliable source for this, especially amongst student ages and young men. This is because these groups are less likely to re-register with a GP when they change address, compared with the general population. Almost a quarter of Twitter users are between the ages of 18-24. Therefore, one of the benefits of this project is that the results can be used to improve population estimates in the United Kingdom.

Another benefit is that it may even be possible to produce estimates for different population bases, which cannot be produced from existing data sources, such as day time or seasonal populations. Although we recognise that there are numbers of quality issues associated with the use of Twitter, in particular bias and the stability of the data, Twitter data is available at near real time and much more frequently than official sources. The results to date demonstrate that there is useful intelligence in the data. Analysis of monthly flows in and out of local authorities, known to have high student populations, have shown strong patterns of mobility that follow the cycle of the academic year, i.e. movements out of the local authority in May/June followed by movements into the local authority in September (after the summer break). Again, this will help ONS to understand the population and how different groups move around the country.

B6	**Please outline the ethical issues that might arise from the proposed study and how they will be addressed** ***(all research projects have some ethical considerations, so this section must not be left blank)***

A number of ethical issues have arisen within this work to date and have been anticipated for future work in this area.

Informed consent

The data used in this project is in the public domain, anyone can access the data through the public Twitter API and so we have the legal right to access this data. However, do Twitter users realise that their bio, the content of the tweet and also their location (if they have geo-location enabled) is all in the public domain?

We have evidence that some Twitter users do not have this understanding. Following the release of the ONS report on this work (as linked above) comments were made on Twitter by individuals questioning whether ONS had the authority to access and use this data and whether it was right that we were doing so.

In addition through our exploratory analysis of the geo-located tweets used in this study we identified a 25 per cent drop in daily volumes during the second half of September 2014. Investigations into the reason for this decline in volumes during September identified a link with the release of the iPhone iOS8 operating system. This included changes to how privacy and location are managed[5]. An analysis of tweets by device type shows that this decline is indeed almost entirely explained by a decline in volumes from iPhone devices. This suggests that many iPhone users took the opportunity to exert greater control over their location settings which subsequently impacted the overall volume of geolocated tweets which raises the question as to whether these users were fully aware of what was happening to their data prior to the release. These users would have provided consent for their location data to be shared through the operating system and the supporting applications. However, this does not mean that these users were fully aware of what was happening with their location data.

Twitter Rules of the Road

We initially collected data for this study through the public stream of the Twitter API. A decision to stop collecting data through the Twitter API was due to advice from Twitter that the application was in breach of the Twitter Developer Rules[6]. In June 2014, the Beyond 2011 Privacy Advisory Group[7] reviewed all four ONS big data pilots and questioned whether the Twitter pilot was operating within the relevant developer rules. The pilot team could not initially establish with certainty whether the application was operating within the rules.

---

[5] https://support.apple.com/en-us/HT203033

[6] https://dev.twitter.com/overview/terms/agreement-and-policy

[7] See www.ons.gov.uk/ons/about-ons/who-ons-are/programmes-and-projects/beyond-2011/privacy-impact-assessment.pdf (p.18)

However, correspondence with Twitter UK established that it was not.

Although no definitive reason was given, the main issue appeared to be around the scale of the data collection operation. The Twitter Developer Rules states:

"If your application will need more than 1 million user tokens, you must contact us about your Twitter API access, as you may be subject to additional terms." (Section 1c)

To ensure compliance the application developed by the API was halted. The pilot was advised to contact GNIP (a reseller of data, now owned by Twitter) to discuss requirements with a view to purchasing the required data.

GNIP advised that as the data supply agreement covers the use of Twitter data, this could also be applied to cover data already collected by ONS through the API. Thus, this data was combined with purchased data to minimise project costs. Additional data was procured covering the period 15 August to 31 October 2014 and from 1 April to 10 April 2014 to give seven full months of data.

The collection of data through the Twitter API is a sometimes thorny subject. In conversations with other organisations collecting data through the Twitter API, there is generally an awareness of the Twitter Developer Rules, but these are not always followed. For example, in one organisation, a risk-based assessment was made that the worst case would be that Twitter would simply block access to their API key. In any case, there is certainly little evidence of Twitter enforcing their developer rules. These rules emphasise principles of courtesy and "being a good partner" rather than enforcement and sanctions. Thus, although the pilot could have probably continued to collect data through the Twitter API without material consequences, this would be inconsistent with the ONS aim of being a good partner organisation.

In conclusion, any large scale use of Twitter data, including any future extension of this work, would require commercial arrangements to acquire data. Based on the experience of this pilot, this would be a small fraction of the cost of running a similarly sized survey. Although there are clearly major issues around representativeness of data, there may be a business case for procuring Twitter data, providing it offers sufficient benefit. We collaborate and share expertise with a number of different partners around Twitter analysis, e.g. academics, international National Statistics Institutes and other Government Departments. We will only share results and expertise with these organisations and not data.

Deriving demographics

As described above our analysis involves the derivation of demographics from a Twitter user's bio. We recognise the methods are fairly crude but this may raise ethical concerns around the assignment of specific characteristics at the individual level, e.g. age, gender, occupation.

B7	**How will the findings of the research be disseminated?**

A report has already been published see : http://www.ons.gov.uk/ons/about-ons/who-ons-are/programmes-and-projects/the-ons-big-data-project/index.html
And presentations given at a number of different conferences.
As the work progresses we will continue to publish working papers on the ONS website and provide updates at conferences etc.

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.**

At present the main outputs from this research are reports and working papers. In the future we could potentially produce experimental estimates derived from the twitter data using the established estimation framework, these would not be at an individual level but would be aggregating for certain geographies or population groups.

Section C
Details of Data Subjects

C1	**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 considers all Twitter users, in particular those with geo-location enabled. Deriving the demographics of Twitter users does allow us to focus in on specific groups if required – ie students for mobility or certain subsections of the population for sentiment analysis

Justification for focusing on these subsections or groups:

The initial focus has been on students since there are known quality issues for students and young men in particular in the GP patient register (the current method for identifying internal migration).

When undertaking sentiment analysis it may be useful to understand different opinions or different sections of the population.

C2	**Please detail consent given to use data specified in section B2**

The Twitter data used in this project is in the public domain, anyone can access the data through the public Twitter API and so we have the legal right to access this data. More detailed is provided under B6, ‘Informed Consent’ and ‘Twitter Rules of the Road’ section.

C3	**If you are using data held by a third party please detail how you will obtain this**

Purchased through GNIP as outlined in B6, data cost $8,700 for 89 days of data.

Any other business