# 14th Meeting of the
# National Statistician's Data Ethics Advisory Committee

_____

## Minute, Agenda and Papers

_____

Tuesday 30 October 2018
10:30 – 14:15
Drummond Gate, UK Statistics Authority
London

*UK STATISTICS AUTHORITY*

*NATIONAL STATISTICIAN'S DATA ETHICS ADVISORY COMMITTEE*

*Agenda*

*Tuesday 30 October 2018*
*Drummond Gate London*
*10:30am – 2:15pm*

**Chair: Dame Moira Gibb**

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

| | | |
|---|---|---|
| 1<br>10:30am | Minute and matters arising from the previous meeting | NSDEC Chair |
| 2<br>10:40am | ONS strategy towards the use of health data for statistical purposes | Oral Report<br>Ben Humberstone and Jonny Tinsley |
| 3<br>10:55am | Improving administrative based population and migration estimates using Hospital Episodes Statistics data | NSDEC(18)21<br>Vicky Field |
| 4<br>11:15am | Feasibility study to model health state prevalence estimates for use in the derivation of health state life expectancies in England | NSDEC(18)22<br>Chris White |
| 5<br>11:35am | Census–Refugee Matching: Feasibility, 2009 cohort (CRM:2009) | NSDEC(18)23<br>Jon Wroth-Smith and Rebecca Gillespie |
| 6<br>11:55am | ONS uses of DWP data for Statistical Purposes | Oral Report<br>Jon Wroth-Smith and Matt Brown |

**Lunch (12:15pm to 12:45pm)**

**(12:45pm to 2:15pm)**

| | | |
|---|---|---|
| 7<br>12:45pm | Informing GLA's local development planning responsibilities under the New London Plan | NSDEC(18)24<br>Nick O'Donnell |
| 8<br>1:05 pm | Estimation of price elasticities of demand for alcohol and tobacco products in the United Kingdom | NSDEC(18)25<br>Nick O'Donnell |
| 9<br>1:25pm | ADRP – Developing persistent dataset for projects within the Data for Children research theme | Oral Report<br>Peter Stokes, Emma Gordon and Paul Jackson |
| 10<br>1:55pm | Child abuse feasibility study | Oral Report<br>Alexa Bradley |
| 11<br>2:10pm | Any other business | |

# National Statistician's Data Ethics Advisory Committee
## Minute

Tuesday, 30 October 2018
Meeting Room 10, Drummond Gate, London

**Present**
**Members**
Dame Moira Gibb (Chair)
Mr Rob Bumpstead
Mr Stephen Balchin
Ms Vanessa Cuthill
Mr Keith Dugmore
Mr Colin Godbold
Ms Annie Hitchman
Dr Brent Mittelstadt
Ms Isabel Nisbet
Dr Emma Uprichard

**UK Statistics Authority**
Dr Simon Whitworth

**Office for National Statistics**
Mr Ben Humberstone (for items 2 to 4)
Ms Vicky Field (for items 2 to 4)
Mr Jonny Tinsley (for items 2 to 4)
Mr Jon Wroth-Smith (for items 5 to 6)
Mr Matt Brown (for item 6)
Mr Peter Stokes (for items 7 to 9)
Ms Alexa Bradley (for item 10)

**Administrative Data Research Partnership**
Ms Emma Gordon (for item 9)
Mr Paul Jackson (for item 9)

**Home Office**
Dr Becky Gillespie (for item 5)

**Apologies:**
Ms Marion Oswald


**1.      Minutes and matters arising from the previous meeting**
1.1.    The Chair welcomed members to the fourteenth meeting of the National
         Statistician's Data Ethics Advisory Committee (NSDEC).

1.2.    Members approved the minutes from the previous meeting.

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

**2.    ONS strategy towards the use of health data for statistical purposes**

2.1    Mr Ben Humberstone, ONS Deputy Director of Health and Life Events, presented an overview of the strategy towards the use of health data for statistical purposes. The meeting heard that linked government health data could provide new insight into the drivers of key patterns, such as inequality, which had the potential to have a real impact on lives. It was reported that ONS would never seek to identify individuals and no data shared with ONS would be onwardly shared, nor used for operational purposes. All staff handling the data would be trained in their responsibilities and security cleared to work with the data.

2.2    NSDEC welcomed this opportunity to hear about the strategy and suggested the production of high quality case studies showing the public good of longitudinal analysis was important to show the public the value of the linking of data across the life course.

**3.    Improving administrative based population and migration estimates using Hospital Episodes Statistics data [NSDEC(18)21]**

3.1.    Ms Vicky Field from ONS Integrated Data Division (IDD) presented a proposal to use Hospital Episodes Data (HES) to establish whether people are present in the population to improve the accuracy of the population estimates that could be produced from linked administrative data.

3.2.    NSDEC was assured that this work would be consistent with the Digital Economy Act Statistics Statement of Principles and would meet the organisation's obligations under the GDPR.

3.3.    It was suggested that it was not clear why all of the variables listed in annex 1 were required for the purposes of improving the population estimates. NSDEC advised that Ms Field should take another look at these variables and fully justify why each of these variables were required for this work.

3.4.    Members were informed that work was being conducted to look at the possibility of using surveys to collect information on people who don't interact with the administrative sources. NSDEC felt that this should be made clear in the application. The reasons why ethnicity was required for this research should also be made clear given it is classified as sensitive data under GDPR.

3.5.    This project was approved subject to Ms Field making the suggested amendments to the application.

Action: Ms Field to;
      i.    fully justify why all variables were required for the purposes of improving population/migration estimates;
     ii.    make clear that surveys may be used to capture people who don't interact with the administrative data; and

iii.     explain why data on ethnicity was required for this project.

**4.     Feasibility study to model health state prevalence estimates for use in the derivation of health state life expectancies in England [NSDEC(18)22]**

4.1.     Mr Chris White from ONS Health and Life Events presented a project proposal to investigate how the strengths of the HES data could be used to improve the accuracy of healthy life expectancy statistics using correlational analysis and data linkage.

4.2.     NSDEC approved the first Phase of this study and advised that the second phase of the project should be considered at a future NSDEC meeting. NSDEC were informed that all the data used for this project would be deleted should the first Phase of the project not be successful.

**5.     Census–Refugee Matching: Feasibility, 2009 cohort [NSDEC(18)23]**
5.1.     Dr Rebecca Gillespie from Home Office (HO) presented a proof of concept exercise assessing the feasibility of matching HO asylum grant (HOAG) data to ONS Census data, in order to identify refugees in Census data. The matched data has the potential to inform further exercises to aid the development of integration and dispersal policies and increase understanding of refugee outcomes.

5.2.     The meeting heard that stakeholder engagement had taken place based on the advice NSDEC had provided when they first discussed this project in April. This engagement had shown that stakeholders were supportive of the research.

5.3.     NSDEC were assured that HO would only be given access to deidentified linked data during this project and that the results of the matching would be published in the form of a technical report.

5.4.     Members stated that any future use of this data should come back to NSDEC for ethical consideration.

5.5.     NSDEC felt that this proposal had been considerably strengthened in line with the advice they had previously provided and approved the project.

**6.     ONS uses of DWP data for Statistical Purposes**
6.1.     Mr Matt Brown from ONS Data as a Service presented a high-level overview of the department's approach to using DWP data for statistical purposes and how this was consistent with NSDEC's ethical principles.

6.2.     The following comments were made in the discussion that followed:
i.     To maintain public trust in their use of data ONS should be clear that they would not make data they get for statistical purposes under the Digital Economy Act (DEA) available to researchers without the permission of the department who supply the data.

ii. ONS should be transparent about the data it uses and the variables used for research and statistical purposes.

iii. Given the volume of data that ONS could potentially access under the DEA, the importance of robust governance to ensure appropriate consideration of the use of data was stressed.

iv. All uses of health data should continue to be considered by NSDEC.

## 7. GLA's local development planning responsibilities under the New London Plan [NSDEC(18)24]

7.1. Mr Peter Stokes from ONS Research Support and Data Access presented a project by Arup, a private research company, commissioned by the Greater London Authority, to use the Interdepartmental Business Register to study economic development and the potential for economic change in Oxford Street and the West End.

7.2. NSDEC were assured that the focus of this research was economic development and change rather than housing.

7.3. NSDEC approved this project subject to all outputs being shared with ONS prior to publication

## 8. Estimation of price elasticities of demand for alcohol and tobacco products in the United Kingdom [NSDEC(18)25]

8.1 Mr Peter Stokes presented a project from Deloitte to use the Living Costs and Food Survey to explore the scope to increase taxes on cigarettes and alcohol.

8,2 NSDEC were assured that all outputs would be shared with ONS prior to publication. It was suggested that ONS should insist that Deloitte inform ONS if Deloitte badge this work as having ONS support at any point.

8.3 This project was approved.

## 9. Administrative Data Research Programme (ADRP)

9.1 Dr Emma Gordon from ADRP's Strategic Hub and Mr Peter Stokes reported that the ESRC are working in partnership with ONS to strategically acquire and curate admin data and make these available to researchers in a secure way, to drive high-quality research and impact around prioritised policy themes within the UK. NSDEC were informed that ADRP has 4 Primary Goals:

i. Trust and Sustainability;
ii. Research for the Public Good;
iii. Researcher Access and Support; and
iv. Available Data for Research

## 10. Child abuse feasibility study

10.1. Ms Alexa Bradley from ONS Crime Statistics reported that ONS plan to explore the feasibility of conducting a national survey on the prevalence of child abuse. NSDEC were informed that this work would be put out to tender

and the exact scope and detail of this work would be agreed with the successful bidder.

10.2. NSDEC stated that they would need further details to provide ethical assurance of this work. Whilst the detail is not available the committee expressed considerable concern about the difficulties of this approach and wished to consider details of the method as early as possible.

Action: Ms Alexa Bradley to:
    i.      ensure that the invitation to tender make clear that this work should be considered by NSDEC; and
    ii.     to bring the project to NSDEC for ethical scrutiny as soon as possible when the details of the project are known.

## 11.    Any other business
11.1. There was no other business.

# ONS strategy towards the use of health data for statistical purposes

## Oral report

Ben Humberstone
Jonny Tinsley

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

---

**Please consult the guidance document before filling in the application form**

---

**Project Title**
*Please provide a title indicative of the project*

Improving administrative based population and migration estimates using Hospital Episodes Statistics data.

---

**Start Date:** October 2018      **End Date:** September 2019

---

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

Office for National Statistics (ONS)

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

Greater use of administrative data is central to the Office for National Statistics (ONS) plans to modernise the population and migration statistics system in England and Wales. It is the aspiration of UK government that the 2021 Census will be the last of its kind, and that sources of data already collected within government will be used to produce future census statistics. The Administrative Data Census (ADC) project is leading on this within ONS and a recommendation to government is due in 2023.

One of the key milestones for Population and Public Policy is to put administrative data at the core of migration statistics in 2019 and administrative data at the core of population statistics in 2020. Since 2015, ADC have been publishing annual Research Outputs to demonstrate progress in estimating the size of the population using administrative data. While the results are encouraging, users have expressed the need for ONS to prove that a robust framework can be put in place to adjust for coverage errors[1] in the administrative data.

---

[1] Coverage errors include UK residents that have not yet registered for government services (under-coverage), and individuals with outdated registrations that have moved address or left the country (over-coverage).

Over-coverage (records with outdated registrations who have moved address or left the country) on administrative sources is a challenging problem that needs to be understood and adequately adjusted. Having taken advice from academics and colleagues at other National Statistical Institutes (NSIs), additional data needs to be collected to measure and adjust for over-coverage on the administrative sources. Using Hospital Episode Statistics (HES) will enable us to explore its potential for improving our methodology through interactions with secondary care services. These interactions can be used as additional evidence of "activity"[2] and thereby confirm residence, or non-residence of individuals, which may not be available in the other data sources. **No diagnosis or medical data contained in HES will be required for this project.**

Ethnicity is also a high-priority population characteristic as it is an important demographic component of populations. Ethnicity data are used in resource allocation and policy development, and by organisations to monitor and meet their statutory obligations under the Equality Act 2010. This year, we published new ethnicity estimates investigating a new approach combining administrative and survey data.

Using combined administrative and survey data showed promise for the ability to produce estimates of ethnicity by local authority. However, the survey data are less accurate for smaller ethnicities due to small sample sizes, which means the combined administrative and survey data estimates were also less accurate. We believe that we will be able to use the more comprehensive HES ethnicity data to further improve our models by providing more detailed information of the population by LA and ethnic group.

Currently, the Census holds detailed ethnicity information at small levels of geography and for smaller sub-groups of the population. The use of administrative data provides the opportunity to provide equivalent information on a more timely basis. While the English School Census (ESC) data provides ethnicity for school-age children it doesn't include children at independent schools, children having home schooling or those being educated somewhere other than at school. HES data could provide much wider proportional coverage of ethnicity for the population generally in England and Wales rather than being limited to a subset of the population currently provided by the ESC.

---

[2] "Activity" can be defined as an individual interacting with an administrative system, for example for National Insurance or tax purposes, when claiming a benefit, attending hospital appointments or updating information on government systems in some other way. Only demographic information (such as name, date of birth and address) and dates of interaction are needed from such data sources to improve the coverage of our population estimates.

Section A
**Project Details**

**A1**     **Legal gateways**

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

---

**NHS-Digital controlled HES data**

Section 45A of the Statistics and Registration Services Act 2007 (as inserted by the Digital Economy Act 2017) provides a legal gateway for all public authorities, including NHS Digital (NHS D), to share the information they hold with ONS. This legal gateway over-rides any restriction on disclosure, including patient confidentiality, though data protection legislation must still be complied with. ONS can use the data received under this gateway in pursuit of any of its statutory functions.

**Statistical Population Dataset[3]**

Section 45A of the Statistics and Registration Services Act 2007 (as inserted by the Digital Economy Act 2017) permits the sharing of personal identifiers from the Customer Information System for use in population statistics.

Section 43 of the Statistics and Registration Services Act 2007 permits the sharing of personal identifiers from the Patient Register and Personal Demographic Service for use in population statistics.

English and Welsh School Census S45A of the Statistics and Registration Services Act 2007 (as inserted by the Digital Economy Act 2017) permits the sharing of personal identifiers from the English and Welsh School Census for use in population statistics.

Section 45A of the Statistics and Registration Services Act 2007 (as inserted by the Digital Economy Act 2017) permits the sharing of personal identifiers from the Higher Education Statistics Agency.

**Census Data**

This is collected by ONS under their powers in the Census Act 1920 (and associated subordinate legislation), and can be used by ONS for any of its functions.

Any additional legislative requirements of using specific administrative data is agreed with data providers and will be scrutinised by both the ONS Legal Services and the legal services department of the data providers.

---

[3] The Statistical Population Dataset (SPD) includes several pseudo-anonymised data sets including (NHS Patient Register, HMRC Customer Information System, Higher Education Statistics Agency, School Census, Patient Demographic Service). More information can be found at
https://www.ons.gov.uk/census/censustransformationprogramme/administrativedatacensuspr oject/methodology/methodologyofstatisticalpopulationdatasetv20

## A2    Ethical approval

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

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

## A3    Proposed site of research select all that apply

☑ ONS                                    ☐ ADRC - England

☐ SRS (formerly VML)                     ☐ ADRC - Scotland

☐ HMRC Data Lab                          ☐ ADRC - Northern Ireland

☐ Other                                  ☐ ADRC - Wales

   please specify

## A4    Data subjects to be studied

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

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

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

Justification for focusing on these subsections or groups:

<br>

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

The Administrative Data Census (ADC) project team are using the detail provided by linking a range of datasets at record level to develop a methodology to produce population estimates by sex and age to levels of geography currently published by the Population Estimates Unit. The annual population estimates for England and Wales, cover the regions and counties of England, and local authorities and their equivalents, plus Lower and Middle Layer Super Output Areas and more recently Output Area. At present, we have focussed on producing admin based population estimates down to Output Area level. Westminster parliamentary constituencies, electoral wards and National Parks in England and Wales and clinical commissioning groups in England are also produced as part of the official population estimates, but as yet these have not been covered as part of the ADC project.

Statistical Population Dataset (SPD) population estimates are produced by linking administrative records. In the absence of common identifiers across the different data sources, our linkage methodology has been based on matching records using combinations of their name, date of birth, sex and postcode. We currently link administrative and survey data using a pseudonymisation approach to preserve the privacy of information about individuals and households. The variables required for linkage and to carry out the activity and ethnicity research are listed in detail in Annex 1.

Due to inconsistencies that exist in the capture of this information, two types of error can occur in the matching process. These are: (1) false positives – incorrect matches that have been made between records, and (2) false negatives – missed matches that have not been correctly identified between records.

We have an interest in accessing all "activity"-based information, indicative of any interactions with primary and secondary healthcare, for all patients recorded on the Personal Demographic Service (PDS). This includes HES data where interactions with a hospital service are recorded as separate episodes that may not be evident on the GP Patient Register (PR) or PDS which solely provide coverage of people registered with an NHS General Practitioner in England and Wales.

The availability of some "activity" data from the Personal Demographic Service (PDS) and some Department for Work and Pensions (DWP) benefits data has enabled us to develop the methodology for assigning records to the most likely address, but we require more access to additional data sources that confirm most recent address for a record on the SPD. This will help put the right records in the 'right' place, leading to more robust population estimates at small areas.

We will also begin targeting the use of "activity" data to help remove records from the SPD if there is no evidence they are still present in the population, or to include records that are present but do not appear on multiple administrative sources.

These developments will also feed into work on the transforming the population statistics system including improving estimates of international migration. ONS are reviewing how this can be achieved; for example, how alternative data sources can be used to identify movement both to and from the UK and within the UK. This work builds and expands upon our research on an Administrative Data Census to move into a better understanding of population and migration using all available sources.

It is also known that not every individual interacts with services captured by the administrative data that ONS receive, this can result in coverage errors. This is a challenge that ONS also have with the current system, that not everyone completes a Census form. To adjust for this a capture re-capture method is used via the Census coverage survey. The ADC project is researching into the potential of using surveys in a similar way for administrative data.

Our latest Research Output on ethnicity estimates highlighted the limitations current admin data sources to cover certain population groups. Information on ethnic group is not widely collected on administrative data sources, HES is one of the few, and while HES will not give complete coverage of the population it is one of the few sources where ethnicity is captured at all. Although ONS recognises these coverage limitations, it realises that are methodological approaches that can be used to mitigate these.

| A6 | **Data use** Please specify the data used **by the research team** including any timeframes *e.g. LFS data 2014-15* |
|----|----|

**Data Level**
*Please specify the name of the data set*

| Type of data | Aggregate Data | Identifiable Data | De-identified personal data | Anonymised/ pseudo anonymised |
|---|---|---|---|---|
| **Administrative data** *(please specify, e.g. Patient Register 2011, School Census 2012 etc, in the relevant options adjacent)* | | Hospital Episodes Statistics 2009-10 - 2017-18<br><br>Patient Register 2012-2017<br><br>Patient Demographic Service 2016 - 2017 | The analytical dataset will contain de-identified, linked person level data. | Hospital Episodes Statistics<br><br>Statistical Population Dataset (see section A1)<br><br>Only available to those conducting the linkage |

| | | | | |
|---|---|---|---|---|
| | | Only available to those conducting the linkage | | |
| **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)* | | 2011 Census  Only available to those conducting the linkage | | |
| **Other** *(please specify e.g. Ordinance Survey Address register in the relevant options adjacent)* | | | | |

## Section B
## Assessment against NSDEC ethical principles

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

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

This project aims to benefit users by producing more accurate and timely population estimates, which can be used in a number of applications, including pensions forecasting, provision of local services and migration statistics. This project is also working closely with the wider statistical community and over time should help improve coherence in the use of administrative data across the GSS. Users here include:

- local or sub-national government

- Voluntary / Community / Non-profit organisations
- Government Departments
- Education or university
- Private/retail sector

Although the currently used administrative sources are valuable, they have significant issues of under- and over-coverage of the population. This requires development of more technical methods and new surveys to produce population statistics to the required quality. The Admin Data Census team is conducting research to see if it is feasible to fill gaps in current population statistics by using additional administrative data sources. This would benefit users (see above) by better meeting their needs as captured in several public engagement initiatives (see section B7).

To improve the quality of the population estimates, we are exploring a rules-based approach that makes best use of "activity" and HES can provide this through confirmed interactions with an administrative data system. It is expected that HES will provide better evidence of residence or non-residence and more up-to-date intelligence of the correct address where there is conflict with the other data sources currently being used. It is also expected to overcome issues of lag that are common on non-activity based data sources.

Using administrative data benefits the public by ensuring that the Integrated Data Division within ONS provides value for money by comparing alternative approaches. This is in line with the Code of Practice for Official Statistics – "*administrative sources should be fully exploited for statistical purposes, subject to adherence to appropriate safeguards.*"

This research is an important element of our plans to transform to a new administrative data-led system for population and migration statistics for England and Wales by spring 2020. The current system is rebased every 10 years using the Census estimates resulting in declining quality over time, administrative data based population estimates have the potential to retain consistent levels of quality over time. Improved population estimates between census periods and an improved basis for weighting social surveys would allow policy makers to be more proactive and more accurate in planning and making provisions for public services. This would benefit the public through improved and more timely allocation of resources.

Information on ethnic groups is a high-user need, as demonstrated by the 2021 Census Topic Consultation. Quality ethnicity estimates are important in decision making, resource allocation and planning, including understanding impact of policies and programmes on sub-groups of the population. The census is currently the key method to capture information on ethnic groups at the lowest levels of geography, however these statistics become outdated. Surveys offer an alternative, but due to sample sizes these do not always meet user needs. Producing ethnicity population estimates using administrative data provides the opportunity to generate more timely statistics at lower levels of geography that are of high quality and that are robust. This will in turn enable effective decisions at the local level.

Improved population estimates and realising the potential of administrative data can aid us in evaluating and improving the quality of existing population statistics and enhance the coherence of the wider statistical system. The proposed research is in line with the ONS strategy of 'Better Statistics, Better Decisions' and presents a clear benefit in assisting in the development and evaluation of public policy.

In addition, a better understanding of the impact of international migrants at a local level will be essential in providing local authorities with the statistics they need to plan the provision of services to meet increasing demand due to immigration.

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

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

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

The results of the initial research (and any further developments) will be published on the ONS website as an article with aggregated tables that meet the ONS disclosure control policy.

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

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

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

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

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

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

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

**B4**

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

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

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

Probabilistic matching is a mathematically formulated approach to calculate agreement scores between candidate pairs of linked records. For the Statistical Population Dataset V2.0, we have used this approach to identify further links between PR and CIS records that have not already been matched using deterministic matching in SPD V1.0. Below is a summary of the process:

- all residual (non-linked) PR and CIS records that remain unmatched following the deterministic matching are compared with each other; for each single PR record, comparisons are made with all CIS records on forename, surname, date of birth, sex and postcode

- PR and CIS records that partially agree on at least two of these five matching fields are retained as potential candidates in the match pool

- an Expectation-Maximisation algorithm is used within the probabilistic matching framework to calculate agreement scores between all of the candidate pairs in the match pool

- an analysis of the score distribution of candidate pairs is used to determine a threshold for accepting or rejecting links in the match pool; those candidate pairs with agreement scores higher than the threshold are accepted as matches to be included in SPD V2.0; those below are excluded

This approach will also be possible with the HES data when it is linked to the other named data sources. All processes will have adequate human control to ensure appropriate quality assurance arrangements are in place.

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

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

The legal basis for acquiring the HES data will be Section 45A of the SRSA 2007 as amended by the Digital Economy Act 2017. This removes any other duty of confidence on the supplier, including common law duty of confidence.

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

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

However, the data share must be needed for ONS functions, and not contravene data protection legislation. To assess the former, there is a code of practice underpinning this power that the data share is assessed against. We believe that acquiring this data for this purpose is in line with the code of practice. Final sign-off that this is the case will be made by the ONS Data Governance Committee and this will be sought after ethical approval for the project.

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

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

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

To further comply with data protection legislation-

1) ONS will only obtain and use the minimum data needed for the success of project outcomes – i.e. answering the research questions as stated in the project summary

2) Researchers will only handle anonymised data

3) Strict, published disclosures control protocols will be adhered to with any outputs produced

4) There is a need for data controllers to inform individuals how their data are being used. For ONS we consider it would be disproportionate effort to notify each person individually, since there will be no impact on them. Instead details of the data shares will be published on ONS website along with other privacy information on how we use data and how we keep it secure. In addition, NHS D will publish the Data Access Request Service application form that ONS will complete as part of the process to access the HES data.

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

**B6**     **Collaboration and Sponsors**

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

| List of Collaborators/Sponsors | Details and relevant documentation relating to collaboration (you may attach copies of relevant documentation) |
|---|---|
| ONS | |
| The project falls within the use of data for the production of statistics and is compliant to the Data Protection Act 2018 and the Statistics and Registration Service Act (SRSA) 2007.<br><br>All data have been acquired using established legal gateways cleared by ONS Legal Services (section A1 and B6). | Section 45A of the Statistics and Registration Services Act 2007 (as inserted by section 79 of the Digital Economy Act 2017) permits any public authority to disclose to the Statistics Board any information they hold in connection with their functions. Information so disclosed may be used by the Statistics Board for any of their functions as set out in the SRSA 2007 and the Census Act 1920.<br><br>The legal basis for processing information under this agreement is provided by the GDPR under Article 6(1)(e) where the "processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller".<br><br>and |

| | For processing information related to special categories of personal data is provided under Article 9(2)(j) where "processing is necessary for archiving purposes in the public interest, scientific or historic research purposes or statistical purposes in accordance with Article |
|---|---|
| | 89(1) based on Union or Member State law which shall be proportionate to the aim pursued, respect the essence of the right to data protection and provide for suitable and specific measures to safeguard the fundamental rights and the interests of the data subject". Personal data are provided under Article 9(2)(j) where "processing is necessary for archiving purposes in the public interest, scientific or historic research purposes or statistical purposes in accordance with Article 89(1) based on Union or Member State law which shall be proportionate to the aim pursued, respect the essence of the right to data protection and provide for suitable and specific measures to safeguard the fundamental rights and the interests of the data subject". |

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

*Please list any public engagement activities* (max 250 words)

Public responses to the Beyond 2011 Programme Consultations were overwhelmingly supportive that ONS should continue developing methods for producing population estimates using administrative and survey data.

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

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

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

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

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

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

With the introduction of the Digital Economy Act 2017 amendments to the SRSA 2007, which broadens ONS's powers to acquire data, further public acceptability / public license work is planned.

We have continued to consult with our users on our Research Outputs via a series of local authority engagement events and have encouraged feedback more generally on our Research Outputs using the Administrative Data Census contact email. The feedback from local authorities on our population estimates has been encouraging but has highlighted the need for more 'activity' data, such as HES, to strengthen the exclusion rules and reduce the over-coverage on the SPD.

A 2021 Census Topic Consultation was held in June 2015, where it was identified that ethnic group had a high user need. Additionally, we have gathered feedback on our research for producing ethnicity estimates from combining administrative and survey data. This feedback acknowledged that we still had some way to go in our research and pointed to the reduced quality of the estimates for the smaller ethnic groups. This is likely to be a result of the relatively small sample size of the survey data used in the model and therefore the number of people sampled within detailed geographical areas and population subgroups is limited. It was acknowledged that for our next steps we should continue to encourage the improved collection of and access to ethnicity data, such as that contained in Hospital Episode Statistics.

Feedback on the Digital Economy Act, Part 5: Information Sharing Codes and Regulations Consultation and the consultation outcome on the Statistics Statement of Principles and Code of Practice on changes to data systems found broad agreement and strong support for the use of a range of data sources for research and statistical purposes that serve the public good.

The Statement of Principles and Procedures (see above) also recognised – "that data useful for the production and publication of official statistics and for statistical research is held by an increasing number of public and private bodies in ever larger quantities, and that the proliferation of useful data sources will continue in the future." This includes health and social care data which was noted by several respondents.

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

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

We've continued to develop our understanding of the data through Statistical Quality Working Groups (SQWGs) where we discuss the data and the detail of the variables with data suppliers such as NHS Digital. They also allow us to understand any changes to the systems and the effect these could have on the data supplied. For example, during the last year, SQWGs helped ONS prepare for the replacement of a previously supplied dataset with an alternative dataset.

More activity data are required to make this possible and we expect data from Hospital Episode Statistics to provide this need.

In the latest Annual Assessment documenting ONS's evaluation of its ability to move to an Administrative Data Census in the next decade, the importance of gaining more sources of administrative data, including "activity" data, to improve the quality of these estimates is noted.

The publication also references the migration statistics transformation update published in May 2018 highlights our plans to produce estimates of migration with administrative data at the core. Importantly, the update states as part of the Government Statistical Service migration development plan we need to explore "the possibility of producing migration flows over time from integrated data sources and identifying signs of activity within the system. This will bring together the work on EU and non-EU nationals to explore the feasibility of different approaches to putting administrative data at the core of migration statistics".

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

The non-disclosive aggregated tabulations of the results of this project will be initially shared with analytical colleagues in departments such as NHS D and NHS-England for quality assurance purposes. These organisations are directly involved in the delivery and/or analysis of HES data and have all agreed that this project will fill evidence gaps. We will use their subject knowledge expertise in the source datasets to help explain patterns or trends, and help identify any areas for further investigation.

Outcomes of this work will be published on the ONS website, as Research Outputs. We will work with the other organisations who publish analyses of HES to ensure an overall coherent picture is presented – for example, providing links to and a short explanation of the ONS research alongside their analysis.

**Section C**
**Responsible owner and applicant details**

| C1 | **Responsible Owner** |
|----|----|

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

### 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.

**C2**

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

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

**List of HES variables**

| Field name | Field description | Justification |
|---|---|---|
| ACTIVAGE | Age at activity date | Confirmation/identification of patient's age. Quality assurance. |
| ADM_CFL | Admission date check flag | Quality assurance of the data in the source - understand whether the data is valid/not valid. |
| ADMIAGE | Age on admission | Confirmation/identification of patient's age. Quality assurance. |
| ADMIDATE | Date of admission | Identify date of activity. |
| ADMIFLAG | Admission episode flag | Validation of admission episode - confirm interaction. |
| ADMINCAT | Administrative category | Identifies type of patient. Private patients not recorded on current datasets and may not be included in current admin data population estimates. Therefore, this could lead to identification of missed records on other data sources. |
| ADMINCATST | Admin category at start of episode | Identifies type of patient. Private patients not recorded on current datasets and may not be included in current admin data population estimates. Therefore, this could lead to identification of missed records on other data sources. |
| ADMISORC | Source of admission | Identification of communal establishments such as care homes and potential prison populations. Both populations are difficult to identify and therefore may be missed in the current admin data population estimates. |
| DIS_CFL | Discharge date check flag | Quality assurance of the data in the source - understand whether the data is valid/not valid. |
| DISDATE | Date of discharge | Sign of life/activity. Understand dates of activity. |
| DISDEST | Destination on discharge | Identification of communal establishments such as care homes and potential prison populations. Both populations are difficult to identify and therefore may be missed in the current admin data population estimates. |
| DISFLAG | Discharge episode flag | Sign of life/activity. Date of activity. |
| DOB | Date of birth - patient | Confirmation/identification of a patient's date of birth. Essential linkage variable. |
| DOB_CFL | Date of birth check flag - patient | Quality assurance of the data in the source - understand whether the data is valid/not valid. |
| DOBBABY_N | Birth date (baby) | Confirmation/identification of the baby's date of birth. Essential linkage variable. |
| ENCRYPTED_HESID | Encrypted HES ID | This identifier will assist in linking HES data, particularly to transform from financial years to calendar years to complete analysis. |
| EPIDUR | Episode duration | Sign of life/activity. Understand dates of activity. |

| EPIE_CFL | Episode end date check flag | Quality assurance of the data in the source - understand whether the data is valid/not valid. |
|---|---|---|
| EPIEND | Date episode ended | Sign of life/activity. Understand dates of activity. |
| EPIKEY | Record identifier | Link patient data across different years. |
| EPIORDER | Episode order | Understand sequence of episodes and therefore identify the more recent information available. |
| ETHNOS | Ethnic category | Confirmation /identification of a patient's ethnicity. |
| ETHRAW | Ethnic character (audit version) | Confirmation /identification of a patient's ethnicity. |
| FYEAR | Financial Year | Reference year |
| HESID_ORIG | Patient ID - HES generated (original) | This identifier will assist in linking HES data, particularly to transform from financial years to calendar years to complete analysis. |
| HOMEADD | Postcode of patient | Confirmation/identification of a patient's postcode district.    Essential linkage variable. |
| INYRFLAG | In Year flag | Understand whether activity occurred in HES year, help to form rules around inclusion of records. |
| MYDOB | Date of Birth - month and year | Confirmation/identification of the mother's date of birth. Quality assurance. |
| NEWNHSNO | NHS number | Essential linkage variable |
| NEWNHSNO_CHECK | NHS Number valid flag | This code potentially explains why a patient's NHS number is absent. Quality assurance. |
| NHSNOIND | NHS number status indicator | This code potentially explains why a patient's NHS number is absent. Quality assurance. |
| PARTYEAR | Year and month of data | Reference year |
| PCFOUND | Postcode Found | Confirmation/identification of a patient's postcode. Quality assurance. |
| SEX | Sex of patient | Confirmation/identification of a patient's sex. Essential linkage variable. |
| SPELBGIN | Beginning of spell | Understand order of episodes/spells for more recent information. |
| SPELDUR | Duration of spell | Sign of life/activity. Understand dates of activity. |
| SPELEND | End of spell | Sign of life/activity. Understand dates of activity and order. |

IDD Data Sources Summary Table – Contains source descriptions and variable list for other admin data sources used in the project.

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

### Statistical data processing

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

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

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

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

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

### Data security

Security of data is a key priority for ONS. DAP takes a robust approach to security that is risk-based and holistic, covering people, process and technology. DAP security is based on two key security governance and management layers. The first layer is a set of security

principles to inform design and operation; the second layer distils these principles into specific security controls within the platform.

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

---

**Please consult the [guidance document](#) before filling in the application form**

---

**Project Title**

*Please provide a title indicative of the project*

Feasibility study to model health state prevalence estimates for use in the derivation of health state life expectancies in England using the relationships found in linked Hospital Episode Statistics (HES) and 2011 Census records.

---

**Start Date:**    1/12/2018        **End Date:**        31/12/2019

---

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

Office for National Statistics (ONS)

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

**Background**

Health state life expectancy measures are estimates of years lived in favourable states of health, such as years spent in good general health and years lived free from disabling health conditions. They provide an overview of the health of a population and enable sub-groups of the population to be compared spatially and socioeconomically.

The coverage of current estimates includes UK and constituent countries, [England's regions, combined authorities, upper tier local authorities](#), Scotland's Council areas,

Wales's Unitary Authorities and Health Boards and Northern Ireland's local government districts.

Health state life expectancies are important for public health monitoring and decision making:

- They are included in Public Health England's (PHE) [overarching indicators](#) in the public health outcomes framework, where they are used to monitor health improvement locally and health inequality nationally.
- Healthy life expectancy and disability-free life expectancy provided evidence to the [state pension age review](#) as key indicators of fitness to work to higher state pension ages and its fairness in different locations and among those exposed to different levels of deprivation.

The current method of producing health state life expectancy estimates has limitations. These estimates currently rely on self-reported measures of general health and disability collected in the Annual Population Survey. At local level, they rely on small samples, which often inadequately represent the oldest old, while data is not collected for those aged under 16 years. The small samples lead to wide confidence intervals for sub-national areas which restrict the ability to detect statistically meaningful changes, set health improvement targets for healthy life expectancy, and measure the effectiveness of local decision-making.

Administrative sources such as Hospital Episodes Statistics (HES), and the information it includes on common health conditions, can help with these limitations. In particular, HES provides good coverage of the prevalence of activity limiting conditions at a local level. These conditions include mental health conditions, cardiovascular diseases, respiratory diseases, dementia and Alzheimer's Disease, digestive diseases, endocrinal diseases, cancer and musculoskeletal conditions.

The project will investigate whether these strengths of the HES data can be used to improve the accuracy and granularity of health state life expectancy statistics, and in turn improve local public health decision making. The project will take a phased approach with phase one being the subject of this application and phase 2 being subject to a later submission to the committee.

**Research Phase 1 (the basis of this application)**

Phase 1 will be split into two sub-phases, phase $1_a$ and $1_b$.

In phase $1_a$ we will undertake correlational analyses using unlinked HES, annual population survey and 2011 Census data. We will estimate age and sex specific prevalence of morbidity treated or managed by hospital services, using diagnostic information available in hospital activity for sub-national upper tier local authorities in England in the financial periods 2010/11 and 2011/12. We will then test the correlation between hospital activity based prevalence estimates and those reported in Annual Population Survey data and self-reports of general health and activity restriction reported at 2011 Census. This will be done for upper tier local authorities in England and by Index of Multiple Deprivation 2015 deciles.

In phase $1_b$, we will individually link individuals found in hospital episode statistics to the 2011 Census for the periods 2010/11 and 2011/12, and to death registrations in the period 28th March 2011 to December 2017. Linkage will enable us to investigate the relationship between hospital activity diagnoses found in HES and self-reports of general health and activity restriction in the 2011 Census. This relationship will shed insight into which conditions (sets of conditions) predict future self-reports of health states, thereby clarifying

how much HES data is able to discriminate subjective self-reports of health status using a statistical modelling approach. These modelling coefficients can then be applied to Annual Population Survey weighted data to provide population level counts of those in 'good', 'fair' and 'bad' general health and those 'not limited', 'limited a little' and 'limited a lot'.

The analysis will aim to answer the following research questions:

Phase $1_a$:

- How is the HES population distributed compared to the whole Census population in terms of key demographics such as: age, sex, ethnicity, socioeconomic class, area deprivation as compared with 2011 Census in the periods 2010\11 and 2011\12?

- Do upper tier local authority and deprivation decile age-group and sex specific derived prevalence of hospital treated and managed health conditions in the periods 2010\11 and 2011\12 have a statistically significant relationship with self-reported general health and activity restriction found in 2011 Census?

Phase $1_b$:

- Is there a statistically robust relationship between an individual's age and sex specific prevalence for different conditions treated or managed in hospital and their self-reported age and sex specific prevalence of general health and activity restriction recorded at Census 2011.

- Does this relationship hold when data is aggregated to area and deprivation decile level?

- Does the relationship with self-reported health status strengthen when an individual's length of stay, average number of episodes per admission, number of admissions per year and number of health conditions recorded across their admission and episode history is taken into account?

- Do socio-demographic variables available at Census and additionally collected in Annual Population Survey records, such as socioeconomic class, educational attainment, economic position improve the model fit over and above HES factors?

- How well does the adjustment of Annual Population Survey data based on the prevalence of conditions found in hospital activity fit census based estimates of general health and activity restriction in reweighted Annual Population Survey 2010-12?

The three-way linkage of HES, 2011 Census and death registrations will provide an independent and more detailed demographic baseline than the HES data could do alone, and allow us to model health state prevalence based on stronger evidence of the prevalence of conditions being treated. The use of HES data will reduce the reliance on imputation of data to those age groups which APS data does not cover or where APS data is scant, such as the very old and the smallest local areas such as Rutland. It will also improve precision by deriving weighted counts of health states.

Phase 2

In phase 2, we will investigate to what extent self-reports of general health and activity restriction predict future cause specific hospital activity and mortality. We will also apply the

predictive coefficients for factors found in modelling between hospital activity and self-reports of general health and activity restriction found at 2011 Census to future years of HES data to enable future hospital activity to be used as an explanatory factor. Census characteristics associated with self-reported health states will also be applied to those characteristics available in APS data as additional explanatory factors to estimate future prevalence of health states post census. These estimates will then be used to calculate health state life expectancy at local level and for deprivation deciles. The research questions to be addressed in phase 2 will be:

- Does the inclusion of HES data as an explanatory factor in health state prevalence modelling improve the plausibility of health state life expectancy trajectories at sub-national level compared to reliance on survey data used in past releases?

- Do the modelled prevalence estimates with HES as an explanatory factor have a similar relationship to future mortality risk as that found in linked Census and death registrations data?

**Section A**
**Project Details**

## A1    Legal gateways

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

---

**NHS-Digital controlled HES data**

Section 45A of the Statistics and Registration Services Act 2007 (as inserted by the Digital Economy Act 2017) provides a legal gateway for all public authorities, including NHS Digital (NHS D), to share the information they hold with ONS. This legal gateway over-rides any restriction on disclosure, including patient confidentiality, though data protection legislation must still be complied with. ONS can use the data received under this gateway in pursuit of any of its statutory functions.

**ONS Mortality data**

Section 42 of the Statistics and Registration Services Act (SRSA) 2007 enables ONS to receive mortality data from the Registrar General and to use that data to support any of its statutory functions.

**Census Data**

This is collected by ONS under their powers in the Census Act 1920 (and associated subordinate legislation), and can be used by ONS for any of its functions.

---

## A2    Ethical approval

---

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

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

---

## A3 Proposed site of research select all that apply

- ☑ ONS
- ☐ SRS (formerly VML)
- ☐ HMRC Data Lab
- ☐ Other

  please specify

- ☐ ADRC - England
- ☐ ADRC - Scotland
- ☐ ADRC - Northern Ireland
- ☐ ADRC - Wales

## A4 Data subjects to be studied

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

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

Justification for focusing on these subsections or groups:

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

HES data includes NHS Number, date of birth, sex and postcode. It does not include full name.

The Census 2011 data does include all common demographics used for data linkage but it does not include NHS number.

The patient registration information already held by ONS includes all the demographic variables commonly used for linkage and NHS number. It will therefore be used as a bridge between the Census and HES data.

Our data integration team who are separate from our analysis team, trained in data protection, and are security cleared, will complete this work. This small team (only 2-3 people will access the identifiable data) will link 2011 patient registration information to Census 2011 data using demographic matchkeys. They will then add NHS number to as many Census records as possible. Similar linkage work between these two sources has already been done and refined as part the ONS Admin Data census work.

In addition, the data integration team will add name to the HES data by linking HES data to the appropriate year of patient registration information. Once the above are complete, the data integration team will link the census and HES data to one another.

ONS are already in the process of linking census 2011 to ONS Mortality data. The mortality data if of good quality (accuracy, coverage, completeness), and includes name, address, date of birth, sex and NHS number.

All linkage work will be completed in the ONS secure Data Access Platform (see annex 2). The end result will be a linked person level file between HES, Census and Mortality.

Once linked, identifiers will be removed from the linked data and this subset will be stored on the ONS secure Data Access Platform. Here, two or three ONS researchers will have access to undertake the analysis. All staff who have access to the data will be substantive ONS employees, SC cleared, and trained to handle data correctly (see section B3).

The research questions to be answered by the research team are outlined in the Project Summary section.

Products of the analysis will be published on the ONS website and the aggregated tabulations and statistics will be assessed for disclosure control against the current Mortality Disclosure Control protocol.

Correlational analysis and statistical modelling will be applied to estimate age, sex, local area and decile specific health state prevalence based on a range of variables available from HES, 2011 Census and Death Registrations.

**A6** **Data use Please specify the data used** by the research team **including any timeframes** *e.g. LFS data 2014-15*

**Data Level**
*Please specify the name of the data set*

| Type of data | Aggregate Data | Identifiable Data | De-identified personal data | Anonymised/ pseudo anonymised |
|---|---|---|---|---|
| **Administrative data** *(please specify, e.g. Patient Register 2011, School Census 2012 etc, in the relevant options adjacent)* | | HES 2010/11, 2011/12<br><br>Mortality 2011-2018<br><br>Patient Registration information 2011-2012<br><br>Only available to those conducting the linkage | The analytical dataset will contain de-identified, linked person level data.<br><br>The analytical dataset, detailed below, will only be available to researchers working on this project. | |
| **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)* | | 2011 Census<br><br>Only available to those conducting the linkage. | | |

| **Other** *(please specify e.g. Ordinance Survey Address register in the relevant options adjacent)* | | | | |
|---|---|---|---|---|

**Section B**
**Assessment against NSDEC ethical principles**

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

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

This project will add to the evidence base by enabling a better understanding of which health conditions requiring hospital treatment are associated with self-reports of general health and activity restriction in the population.

Improvements in health state prevalence measurement are essential to authentically measure years lived in different health states, and provide scope to measure years lived in a wider range of health states such as 'Fair' general health and separating out the extent of activity restriction.

Department of Health and Social Care (DHSC) will benefit from improvement in the rigor of estimates of health state life expectancy in the future, to evaluate policies designed to improve the health of the poorest fastest and narrow health inequalities. DHSC pay ONS for the delivery of two Healthy Life Expectancy overarching indicators for measuring area specific health improvement and monitoring the socioeconomic inequality. They have also submitted written support for this project to progress.

Healthy life expectancy measurement has been recently extended as a high level measure of health status in the combined authorities of Greater Manchester and the West Midlands. The Greater Manchester Population Health Plan has HLE as a measure of health need and a guide to target health improvement actions. West Midlands Combined Authority has commissioned ONS to provide HLE by quintiles of deprivation for measurement of inequalities within the combined authority area, and is used as an indicator in their inclusive growth strategy. Local public health teams operating in local authorities generally will be better able to benchmark themselves against national and regional estimates of health state life expectancy and with other authorities with similar levels of area deprivation.

By getting health state life expectancy estimates on a stronger data foundation, the accuracy and precision of these measures will facilitate meaningful evaluation of health improvement programmes implemented in local public health business plans and enable them to set health improvement targets as indicated by health state life expectancy. It will also give users a steer to identify outliers where health state life expectancy is better or worse than would be expected based on census characteristics or measures of deprivation.

Public Health England and Department for Health and Social Care will be able to track the inequality in health state life expectancy across area deprivation deciles and detect meaningful change in the inequality with improvements in precision. At present, there is insufficient precision in estimates to detect significance in non-overlapping periods, making setting targets for change difficult.

Improving lives: the future of work, health and disability, a command paper published by the Work and Health Unit, part of the Department for Work and Pensions and the Department for Health and Social Care, setting out their joint strategy, used disability-free life expectancy to illustrate the gaps in years lived with and without health conditions that cause activity restriction. The presence of such a wide inequality was cited as a primary obstacle to

entering and maintaining a place in the labour market. A benefit of this project will be to improve the objectivity and precision of disability-free life expectancy at local area level and give local areas the means to evaluate the impact of the improving lives initiative.

The Pensions Act 2014 introduced a requirement for the Secretary of State of the Department for Work and Pensions to periodically review the rules on State Pension age and report to Parliament. Health state life expectancy at sub-national level and by measures of socioeconomic status was used in the most recent [review](#) in 2017. With the state pension age set to increase in future years, an assumption will be to extend working lives to higher state pension ages. A key judgement to support such an assumption is fitness to work at older ages, and whether improvements in health state life expectancy are keeping track with life expectancy.

In the State Pension age review health state life expectancies formed part of the evidence base to support decision-making on the fairness of increases and its plausibility. However, the review team pointed to the subjective nature of the measure and stronger links with objective health care impact would strengthen the measures for use in pensions policy making as it would provide a stronger link to the tracking with life expectancy. Therefore, another key benefit of this project will be improved measures of health state life expectancy on which to base pensions policy.

DWP have also provided written support for this research to progress.

This project has the potential to transform the measurement of health state life expectancy and increase its utility in informing policy decisions in the areas of work, pensions, health and social care. The improved data foundation, data linkage and prospective analytical approach will enhance the objectivity of these measures by containing a health care usage component. The use of administrative data will also improve the precision of estimates which currently suffer from wide confidence bounds which restrict the ability to detect timely changes.

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

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

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

The results of the initial research (and any further developments) will be published on the ONS website as an article with aggregated tables that meet the ONS disclosure control policy.

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

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

---

The phase 1 research is not aiming to look at individual level outcomes or to evaluate treatment, but to look for trends in the aggregate data after linkage, to provide population level analysis to improve the measurement of health state life expectancy by placing it on a stronger foundation of diagnostic data requiring treatment.

Both the individual and linked datasets will be stored on the ONS Secure Data Platform. Annex 2 provides detail on the security of the DAP solution. Record level data will not leave this secure environment.

The linkage will be undertaken by staff from a dedicated team in ONS, specialised in linking sensitive identifiable datasets. They have been trained in data protection.

The linked data will be stripped of identifiable variables before being accessed by up to 3 researchers to undertake the analysis.

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

Aggregated data will be published at the National Level (England) and then assessed to see if relevant sub-national tables can be produced following the published Mortality Disclosure Control policy. This applies threshold rules to table cells to ensure that individuals are not able to be identified in the aggregated tabulations. The research team will also consult the HES analysis guide published by NHS-D which includes guidelines on disclosure control:

https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics/users-uses-and-access-to-hospital-episode-statistics

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

Consent is not being sought for this study. Consent is not required by law (see processing conditions set out in section B5), nor would it be practical or proportionate to obtain it for ethical reasons. The data being used has already been collected, so seeking consent would require trying to retrospectively contact all individuals. Further, the study is not concerned with individuals; identifying information is being used for linking only and then being removed before analysis takes place, so there will be no impact on individuals.

The data for analysis will require use of data from the entire population including individuals aged under the age of 16 years old. This is required because health state life expectancy

calculation, which uses a partitioned life table methodology, needs health prevalence estimates across the age range in discrete age-groups; for those aged under 16, no survey data of adequate depth is available at sub-national level.

| B4 | **Principle 3: The risks and limits of new technologies are considered and there is sufficient human oversight so that methods employed are consistent with recognised standards of integrity and quality.** |
|---|---|
| | *Please describe how the any risks from new technologies are been mitigated as well as any quality assurance activities in the project* (max 500 words) |

No new technologies are to be used.

We are confident that the matching should have a high success rate – previous ONS research has shown that a good match rate can be achieved between the Census and the 2011 patient register, and NHS number quality (completeness, accuracy) on HES and death registrations since 2011 is good. However, until we start the linkage work it is not possible to be absolutely certain.

All processes will have adequate human control to ensure appropriate quality assurance arrangements are in place.

| B5 | **Principle 4: Data used and methods employed are consistent with legal requirements such as the Data Protection Act, the Human Rights Act, the Statistics and Registration Service Act and the common law duty of confidence** |
|---|---|
| | *Please describe the legal frameworks pertinent to this project* (max 500 words) |

The legal basis for acquiring the HES data will be Section 45A of the SRSA 2007 as amended by the Digital Economy Act 2017. This removes any other duty of confidence on the supplier, including common law duty of confidence.

However, the data share must be needed for ONS functions, and not contravene data protection legislation. To assess the former, there is a code of practice underpinning this power that the data share is assessed against. Acquiring HES data is in line with this code of practice, and internal scrutiny was provided by the ONS Data Governance Committee which approved ONS acquiring HES data for a wide range of purposes including this one.

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

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

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

To further comply with data protection legislation:

5) ONS will only obtain and use the minimum data needed for the success of project outcomes. One specific example of this is that only the minimum diagnosis information required to enable robust feasibility testing of the methods being developed by this (and other planned ONS health projects) will be acquired from NHS Digital at this stage. Only once it is clear that robust statistics that will enable better public health decision making for the public good can be produced successfully, will additional diagnosis information be sought.

6) Researchers will only handle anonymised data

7) Strict, published disclosures control protocols will be adhered to with any outputs produced

8) There is a need for data controllers to inform individuals how their data are being used. For ONS we consider it would be disproportionate effort to notify each person individually, since there will be no impact on them. Instead details of the data shares will be published on the ONS website along with other privacy information on how we use data and how we keep it secure. In addition, NHS D will publish the Data Access Request Service application form that ONS will complete as part of the process to access the HES data.

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

## B6    Collaboration and Sponsors

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

| List of Collaborators/Sponsors | Details and relevant documentation relating to collaboration (you may attach copies of relevant documentation) |
|---|---|
| ONS | |
| | |
| | |

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

*Please list any public engagement activities* (max 250 words)

While we have not sought the views of the public for this specific project, the research is of clear public benefit as per section B1.

ONS owns and regularly analyses Census and mortality data. Therefore, this section will focus in on the public acceptability of sharing and statistical research involving health data such as HES:

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

> 'people hold mixed views about their information being used for purposes beyond direct care. Some are concerned primarily with privacy and are suspicious that information might be used by commercial companies for marketing or insurance.
>
> Others prioritise the sharing of information to improve health and social care, and for research into new treatments. There is broad support for data being used in running the health and social care system when the benefits of doing so are clearly explained, but people think that anonymised information should be used wherever possible.
>
> The Review also heard very strong views from providers, commissioners, researchers and public bodies that high quality person-level data is needed to run the health and social care system, and to support research…
>
> Because of the importance of earning public trust, the Review concluded that people should be able to opt out of their personal confidential data being used for purposes beyond their direct care unless there is a mandatory legal requirement or an overriding public interest.'

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

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

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

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

With the introduction of the Digital Economy Act 2017 amendments to the SRSA 2007, which broadens ONS's powers to acquire data, further public acceptability work is planned.

In summary, all the above is of course a fine balancing act between privacy and research for the public good. The inclusion of health data in the statistics strand of the new DEA powers was important to enable such research, and in at least some cases we should consider that it may be unethical to *not* complete work – for example where it could save lives. On the other hand, we must be transparent about such uses, ensure the data are kept secure, and welcome external challenge. This ethics committee has an important role to play here.

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

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

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

The non-disclosive aggregated tabulations of the results of this project will be initially shared with analytical colleagues in NHS D for quality assurance and we will consult them on disseminating the findings moving forward. No microdata will be shared outside of the small group of researchers in the Health Analysis and Life Events division in ONS who will access the de-identified linked data.

Outcomes of this work will be published on the ONS website, badged as experimental statistics in an article. We will work with the other organisations who publish analyses of HES to ensure an overall coherent picture is presented – for example, providing links to and a short explanation of the ONS research alongside their analysis.

All outputs from the research for all phases will be published by ONS as aggregated data tables and statistics with the relevant statistical disclosure controls applied. Methodological articles explaining how HES data has been used, will be published first, and any adoption of a method that encompasses HES data in routine reporting of health state life expectancy

releases will require a formal consultation with stakeholders before implementation in routine statistics.

## Section C
## Responsible owner and applicant details

| C1 | Responsible Owner |
|----|-------------------|

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

### Declaration to be signed by the responsible owner

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

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

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

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

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

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

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

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

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

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

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

**List of mortality variables**

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

**List of Census 2011 variables**

| Variable | Description |
| --- | --- |
| TYPACCOM | Accommodation type |
| ACTLW | Activity last week |
| COB | County of Birth |
| HEALTH | General Health |
| HOURS | Hours worked |
| DISABILITY | Long-term health problem or disability |
| MAINLANG | Main Language |
| MARSTAT | Marital and Civil Partnership status |
| POSITION | Position in a communal establishment |
| CARER | Provision of unpaid care |
| RELIGION | Religion |
| STUDENT | Schoolchild or full-time student indicator |
| SEX | Sex |
| SLEEPRUGH | Sleeping rough identifier |
| ADULTLSPUK11 | Adult lifestage |
| AGE | Age |
| AGEARRPUK11 | Age of arrival in the UK |
| SCGPUK11 | Approximated social grade |
| AFIND11 | Armed forces member and dependents indicator |
| DEPRIVED | Classification of household deprivation |
| CECTMCEWS11 | Communal establishment management and type |
| ECOPUK11 | Economic activity |
| ETHNICID | Ethnic group |
| ETHPUK11 | Ethnicity |
| HLQPUK11 | Highest level of qualification |
| HHCHUK11 | Household composition |
| AHCHUK11 | Household composition (Alternative classification) |
| HHLSHUK11 | Household lifestage |
| INDGPUK11 | Industry |
| LRESPUK11 | Length of residence in the UK |
| AFROPUK11 | Member of armed forces |
| NSSEC | National Statistics Socio-economic Classification |
| OCCPUK112 | Occupation (Sub-major group) |
| MAINLANGPRF11 | Proficiency in English |
| RELPUK11 | Religion (Grouped) |
| TENHUK11 | Tenure of household |
| UNEMPHIST | Unemployment history |
| YRARRPUK11 | Year of arrival in the UK |

**List of Hospital Episode Statistics variables (split by the three different HES datasets)**

**Admitted patient care dataset**

| Field name | Field description |
| --- | --- |
| ACSCFLAG | Ambulatory Care Sensitive Condition Flag |
| ACTIVAGE | Age at activity date |
| ADM_CFL | Admission date check flag |
| ADMIAGE | Age on admission |
| ADMIDATE | Date of admission |
| ADMIFLAG | Admission episode flag |
| ADMIMETH | Method of admission |
| ADMINCAT | Administrative category |
| ADMINCATST | Admin category at start of episode |
| ADMISORC | Source of admission |
| AEKEY | Record identifier |
| AT_RESIDENCE | Area Team of Residence |
| BEDYEAR | Bed days within the year |
| CAUSE | Cause code |
| CURRWARD_ONS | Current electoral ward (ONS) |
| DIAG_COUNT | Count of diagnoses |
| DIAG_NN | All Diagnosis codes |
| DIS_CFL | Discharge date check flag |
| DISDATE | Date of discharge |
| DISDEST | Destination on discharge |
| DISFLAG | Discharge episode flag |
| DISMETH | Method of discharge |
| DISREADYDATE | Discharge ready date |
| DOB | Date of birth – patient |
| DOB_CFL | Date of birth check flag - patient |
| ELEC_CFL | Date of decision to admit check flag |
| ELECDATE | Date of decision to admit |
| ELECDUR | Waiting time |
| ENDAGE | Age at end of episode |
| EPIDUR | Episode duration |
| EPIE_CFL | Episode end date check flag |
| EPIEND | Date episode ended |
| EPIKEY | Record identifier |
| EPIORDER | Episode order |
| EPIS_CFL | Episode start date check flag |
| EPISTART | Date episode started |
| EPISTAT | Episode status |
| EPITYPE | Episode type |
| ETHNOS | Ethnic category |
| ETHRAW | Ethnic character (audit version) |
| ETHRAWL | Ethnic category (audit version) |
| FAE | Finished Admission Episode |
| FYEAR | Financial Year |

| | |
|---|---|
| GPPRSTHA | Strategic Health Authority area where patient's GP was registered |
| GRIDLINK | Ordnance Survey grid reference |
| HESID_ORIG | Patient ID - HES generated (original) |
| HOMEADD | Postcode of patient |
| INYRFLAG | In Year flag |
| LSOA11 | Lower Super Output Area (LSOA11) |
| MATCH_RANK | MATCH_RANK |
| MATCHID | Patient identifier (HES generated) - basis of match |
| MSOA11 | Middle Super Output Area, 2011 |
| MYDOB | Date of Birth - month and year |
| NEWNHSNO | NHS number |
| NEWNHSNO_CHECK | NHS Number valid flag |
| NHSNOIND | NHS number status indicator |
| OPDATE_NN | Date of operation |
| PARTYEAR | Year and month of data |
| PCFOUND | Postcode Found |
| POSTDIST | Postcode district of patient's residence |
| RESCTY_ONS | County of residence (ONS) |
| RESGOR_ONS | Government office region of residence (ONS) |
| RESLADST_ONS | Local authority district (ONS) |
| RTTPEREND | RTT period end date |
| RTTPERSTART | RTT period start date |
| RTTPERSTAT | RTT period status |
| SEX | Sex of patient |
| SPELBGIN | Beginning of spell |
| SPELDUR | Duration of spell |
| SPELEND | End of spell |
| STARTAGE | Age at start of episode |
| STARTAGE_CALC | Age of patients at start of episode, babies restated |
| SUBDATE | Submission date |

## Outpatient dataset

| Field name | Field description |
|---|---|
| ACTIVAGE | Age at activity date |
| ADMINCAT | Administrative category |
| APPTDATE | Appointment date |
| AT_RESIDENCE | Area Team of Residence |
| ATENTYPE | Attendance type |
| ATTENDED | Attended or did not attend |
| ATTENDID | Attendance identifier |
| ATTENDKEY | Record identifier |
| ATTENDKEY_FLAG | Attendence Key Flag |
| CARERSI | Carer support indicator |
| CURRWARD_ONS | Current electoral ward (ONS) |

| | |
|---|---|
| DIAG_COUNT | Count of diagnoses |
| DIAG_NN | Diagnosis |
| DNADATE | Last DNA or patient cancelled date |
| DOB | Date of birth – patient |
| DOB_CFL | Date of birth check flag – patient |
| ETHNOS | Ethnic category |
| ETHRAWL | Ethnic category (audit version) |
| FYEAR | Financial Year |
| GRIDLINK | Ordnance survey grid reference |
| HESID_ORIG | Patient ID – HES generated (original) |
| HOMEADD | Postcode of patient |
| LSOA11 | Lower Super Output Area (LSOA11) |
| MATCH_RANK | MATCH_RANK |
| MSOA11 | Middle Super Output Area, 2011 |
| MYDOB | Date of Birth - month and year |
| NEWNHSNO | NHS Number |
| NEWNHSNO_CHECK | NHS Number valid flag |
| NHSNOIND | NHS number status indicator |
| OACODE11 | Census Output Area, 2011 |
| PARTYEAR | Year and month of data |
| PCFOUND | Postcode Found |
| POSTDIST | Postcode district of patient's residence |
| PRIORITY | Priority type |
| REQDATE | Referral request received date |
| RESCTY_ONS | County of residence (ONS) |
| RESGOR_ONS | Government office region of residence (ONS) |
| RESLADST_ONS | Local authority district (ONS) |
| RTTPEREND | RTT period end date |
| RTTPERSTART | RTT period start date |
| RTTPERSTAT | RTT period status |
| SEX | Sex of patient |
| SUBDATE | Submission date |

## Accident and emergency dataset

| Field name | Field description |
|---|---|
| ACTIVAGE | Age at activity date |
| AEARRIVALMODE | Arrival mode |
| AEATTENDCAT | Attendance category |
| AEATTENDDISP | Attendance disposal |
| AEKEY | Record identifier |
| AEKEY_FLAG | AEKEY Flag |
| AEPATGROUP | Patient group |
| AEREFSOURCE | Source of referral for A&E |
| ARRIVALAGE | Age on arrival |
| ARRIVALDATE | Arrival date |

| | |
|---|---|
| AT_RESIDENCE | Area Team of Residence |
| CARERSI | Carer support indicator |
| CURRWARD_ONS | Current electoral ward (ONS) |
| DIAG_NN | A&E diagnosis |
| DOB | Date of birth |
| EPIKEY | Record identifier |
| ETHNOS | Ethnic category |
| FYEAR | Financial Year |
| GRIDLINK | Ordnance Survey grid reference |
| HOMEADD | Postcode |
| LSOA11 | Lower Super Output Area |
| MATCH_RANK | MATCH_RANK |
| MSOA11 | Middle Super Output Area, 2011 |
| NEWNHSNO | NHS Number |
| NEWNHSNO_CHECK | NHS Number valid flag |
| NHSNOIND | NHS number status indicator |
| OACODE11 | Census Output Area, 2011 |
| PARTYEAR | Year and month of data |
| PCFOUND | Postcode Found |
| PCON_ONS | Westminster parliamentary constituency (ONS) |
| PEREND | Reporting period end date |
| RESCTY_ONS | County of residence (ONS) |
| RESGOR_ONS | Government office region of residence (ONS) |
| RESLADST_ONS | Local authority district (ONS) |
| RTTPEREND | RTTP period end |
| RTTPERSTART | RTT period start |
| RTTPERSTAT | RTT period status |
| SEX | Sex of patient |
| SUBDATE | Submission date |

## Annex 2: Overview of Data Access Platform (DAP)

### Statistical data processing

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

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

### Data security

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

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

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

---

**Please consult the guidance document before filling in the application form**

---

## Project Title
*Please provide a title indicative of the project*

> Census–Refugee Matching: 2009 (CRM:2009)

---

**Start Date:** December 2018    **End Date:** December 2019

---

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

> Home Office Analysis & Insight Directorate

---

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

> The **Census-Refugee Matching: 2009** exercise aims to assess the feasibility of matching ONS Census data to Home Office Asylum grant (HOAG) data to understand refugee outcomes.
>
> As a signatory to the 1951 Refugee Convention (and subsequent legislation) the UK is obliged to offer protection to those awarded refugee status and *ensure access to the courts, to primary education, to work, and the provision for documentation*. Refugee and asylum policy is the departmental responsibility of the Home Office, and is multi faceted, covering the initial decision process and dispersal of asylum seekers, as well as integration initiatives for refugees, relating to employment and English language. In order to fulfil this departmental responsibility, and develop and assess relevant policies, the Home Office requires an understanding of the socio-economic outcomes of refugees. It is however widely accepted that quantitative evidence in the area of refugee outcomes is lacking, and this is largely attributed to a lack of datasets which permit refugees to be identified (Ruiz & Vargas-Silva, 2018)[4].
>
> CRM:2009 is potentially the first step in informing this important evidence gap by exploring the feasibility of matching HOAG data to Census data. It is necessary to conduct an initial proof of concept study using one year's worth of data – 2009 - to understand linkage feasibility. Directly identifiable data on HOAG is limited to name, date-of-birth, sex and country-of-birth and this could produce issues for linkage as it is not known how unique combinations of this data are. 2009 is the earliest year that HO holds good quality digital data on refugees, although the short time span between

---

[4] *Journal of Economic Geography*, Volume 18, Issue 4, 1 July 2018, Pages 855–885, https://doi.org/10.1093/jeg/lby027

grant and Census will limit the usability of any findings from this initial project. If this feasibility stage (stage 1) is successful, NSDEC approval will be sought for further linkage exercises (stage 2).

If matching successful, the results from stage 1 will provide experimental findings on location of residence, employment outcomes and English proficiency. These are important dimensions of community cohesion (Cohesion and faiths unit, 2005) and integration (Home Office, 2004) and as such are key refugee policy interests for the Home Office. Stage 2 will explore matching HOAG data to Census 2021. As such CRM:2009 is potentially the first stage in the creation of a rich, local level data set permitting valuable analysis of the socio-economic outcomes of refugees.

## Section A
## Project Details

### A1　Legal gateways

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

**In accordance with 1st Data Protection principle, the data sharing is fair** as it is consistent with the HO Personal Information Charter (HO departmental Privacy Notice), which states that any personal data held by the HO on individuals may be shared with other government departments and agencies and that data that we hold on applicants may be used to undertake statistical and analytical analysis. A Fair Processing Notice (FPN) is included as part of the declaration signed by the individual when applying for asylum.

**Access to ONS Census data by individual Home Office researchers is permissible** under the Approved Researcher (AR) gateway in S.39 of the Statistics and Registration Service Act 2007, allowing ONS personal information to be accessed for research purposes by researchers holding the necessary AR accreditation. HO researchers accessing the data will be accredited accordingly.

**The data processing is in accordance with the EU General Data Protection Regulation (**GDPR) as it meets conditions in both article 6 (required for processing personal data) and article 9 (required for processing special category data), specifically
-   Article 6 condition**:** the processing is necessary in order to perform a task in the public interest or for official functions, and the task or function has a clear basis in law.
-   Article 9 condition: processing is necessary for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) based on Union or Member State law which shall be proportionate to the aim pursued, respect the essence of the right to data protection and provide for suitable and specific measures to safeguard the fundamental rights and the interests of the data subject.

## A2    Ethical approval

**Has the project being reviewed or is it expected to be reviewed by another ethics committee?**

○ **Yes**    ○ **No**

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

CRM was considered at the April 2018 NS-DEC meeting. The proposal was rejected as it did not meet requirements under principle 1 and principle 5.

In order to address **principle 1 (clear public benefit)** workshops have been held with policy leads to refine the policy usages of the matched dataset. These uses are outlined in section B1.

An engagement event was held with members of the National Asylum Stakeholder Forum (NASF) in early October to address **principle 5 (public engagement)**. Views were sought on data usage, research questions and potential benefits of the work. Further information on this engagement is provided in section B7.

## A3    Proposed site of research *select all that apply*

☐ ONS

☑ VML

☐ HMRC Data Lab

☐ Other

    please specify

☑ ADRC - England

☐ ADRC - Scotland

☐ ADRC - Northern Ireland

☑ ADRC - Wales

## A4    Data subjects to be studied

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

○ **Yes**    ⦿ **No**

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

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

CRM: 2009 will focus on all asylum grants for a single cohort year (2009), and will test a process which would only be applied to those refugees for which there is accessible digital data available.

Justification for focusing on these subsections or groups:

**Focus on asylum grants**: Grants of asylum are the only data that HO holds that will identify refugees. If matching is successful, subsequent descriptive analysis will inform what we know about the socio-economic outcomes of refugees. Causality will not be implied and so control comparisons are not required. It will however be informative to compare outcomes to appropriate nationality groups and the wider population, which are identifiable from published, aggregate level census data.

**Focus on 2009**: It is considered necessary to conduct an initial proof of concept study using one year's worth of data to understand linkage feasibility. 2009 has been selected as the cohort year as digital data prior to this point is considered to be less complete. Additionally, 2009 is the year prior to 2011 with the largest volume of *resettled* refugees (947, compared to an average of 350 for the 5 preceding years). The selection of this year also permits an assessment of dispersal policy at a point desired by policy (grants of asylum will be between 15-26 months prior to Census day). However, it should be noted that the short time span between grant and Census will limit the usability of any findings from this initial project, hence why this is regarded as an initial proof of concept permitting further work as future data sources become available.

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

**Necessity of datasets and consideration of alternatives:**
In order to fully inform the research questions, and enable evidence based policy making, evidence needs to have wide geographic coverage, include a large proportion of the population and cover a wide range of relevant variables. As the HO have no way of contacting refugees (refugees are not required to keep in contact with the HO after they have been granted asylum), the only way to gather this evidence is via linking the administrative data that HO holds on those granted asylum (HOAG data) to other sources. Matching HOAG data with Census will permit a wide range of relevant variables (such as indicators relating to employment, housing and English language proficiency) to be explored in one dataset, and produced via one data matching exercise. The only alternative would be a series of data matching exercises with a range of stakeholders/OGDs, however this is also untested and would increase the privacy risks given the need to share data more widely and carry out multiple linkages. It has not been possible to identify a less intrusive means to create the rich, local level data provided by the Census.

**Key Variables:**
A full list of variables is provided in Annex B. Essentially the HOAG data will contain the match keys (i.e. name and date of birth) and other variables relevant to the assessment of HO policy (e.g. whether the individual received asylum support). These policy variables will also be included in the matched data set, alongside Census variables

relating to 3 key evidence areas (location of residence, employment outcomes and English language proficiency) or within group comparisons (ie. nationality, age, sex).

Published statistics show that in 2009 there were 8,387 total grants and 947 resettled refugees. It is therefore estimated that there will be between 9-10k records in the 2009 HOAG cohort.

**Methods, tools and techniques:**
A 2009 cut of HOAG data will be transferred to the ONS secure research environment (SRE), where it will be linked to Census 2011 data using name, DOB and country of birth as match keys. All matching will be carried out using hashed data in accordance with ONS TTP safeguarding policies and principles, including clear separation between people who access identifying data, and those who access attributes relating to those individuals.

Following linkage, the identifying match keys will be removed from the data and the subsequent matched data set will then be transferred to the ONS Secure Research System (SRS) where it will be accessible only by accredited HO researchers who have undergone the necessary training. Descriptive analysis in the SRS will explore the potential of this matched data set to evidence refugee outcomes relating to employment, English language proficiency and area of residence/housing. The matched data set will also include any HO records that were not linked to Census data, to enable an assessment of linkage bias. As per the matched records, no direct identifiers – name or date of birth - will be included.

ONS will retain the 2009 HOAG data set in pseudonymised form. This will facilitate future linkage exercises (such as matching to 2021 Census) for which additional NSDEC applications will be submitted.

**Research outputs, sharing arrangements and future access to data.**
Following completion of the linkage, a report outlining the aims of the exercise, details of the methodology (including quality criteria and the extent to which these were met), and high level descriptive counts will be published. This will inform the work of other analysts looking to do similar data linkage exercises, and will benefit the wider research community, as well as any potential future work within Government.

| **A6** | **Data use** Please specify the data used **by the research team** including any timeframes *e.g. LFS data 2014-15* |
| --- | --- |

**Data Level**
*Please specify the name of the data set*

| Type of data | Aggregate Data | Identifiable Data | De-identified personal data | Anonymised/ pseudo anonymised |
| --- | --- | --- | --- | --- |
| **Administrative data** *(please specify, e.g. Patient Register 2011, School Census 2012 etc, in the relevant options adjacent)* | | | HOAG 2009 | |
| **Big Data** *(please specify e.g. Twitter data, smart meters and mobile phones, in the relevant options adjacent)* | | | | |
| **Survey Data** *(please specify e.g.LFS, BRES, etc in the relevant options adjacent)* | | | | |
| **Census Data** *(please specify year, e.g. Census 2011 in the relevant options adjacent)* | | | Census 2011 | |
| **Other** *(please specify e.g. Ordinance Survey Address register in the relevant options adjacent)* | | | | |

# Section B
# Assessment against NSDEC ethical principles

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

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

The Home Office has a responsibility to 'protect refugees' (3.2) and in its single department plan to achieve this it aims to:  *"Deliver a progressive, trusted and efficient asylum system which provides consistent, timely and good quality decisions [and] supports those in need of protection"*

The CRM:2009 project is the first step in developing a rich data source to support the Home Office in achieving this goal. Whilst the short time span between grant year and Census in this initial proof of concept exercise will produce findings of an experimental nature, if the project is successful this exercise will mark the start of a programme of work matching HOAG data to 2021 Census data, providing valuable insights into the socio-economic outcomes of refugees. The success of the project will be measured by the match rate between 2011 Census and HOAG data. If the match rate is sufficient analysis will explore three themes: location of residence, employment and English language proficiency (the full research questions are presented in Annex B). Insights in to these three areas have the potential to inform asylum policy, both in its high-level design and its focus on an operational level. It is anticipated that:

...analysis of **location of residence** will help:
- assess whether **dispersal policy** is successful in establishing a *local connection* (ie whether refugees remain in dispersal area). An effective dispersal policy is key in balancing pressures on local authorities and services across the UK and avoiding the concentration of refugees in particular areas.
- develop **integration strategies**. Understanding refugee outcomes compared to the outcomes of the local population will assist in identifying areas for further research to unpick relevant issues to improve integration. (Research questions 4-15.)

...analysis of **employment outcomes** (research questions 9-12) will help identify:
- occupations that need **greater support** to help refugees enter the UK labour market, enabling development of specific interventions.
- industries where refugees have been successful in gaining employment, informing  **employer engagement** and potentially leading to more work placements
- opportunities for **entrepreneurial support**

...analysis of **English language proficiency** (research questions 13-15) will help identify:
- which languages are needed for **cultural orientation** products
- where resource for ESOL (English for Speakers of Other Languages) provision should be targeted.

This project therefore has clear possible benefit for the refugee population in terms of providing experimental findings in the policy areas outlined above, as well as testing the methodology required to create a rich data source on the socio-economic outcomes of refugees. A strong evidence base in the area of refugee integration has the potential to benefit all. Well integrated communities are not just more harmonious places to live, they also facilitate economic and social benefits – an extension of opportunity and

| | |
|---|---|
| | prosperity, and lower levels of prejudice and hate crime (Integrated Communities Strategy , 2018). |

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

The intention is to use the method of linkage developed during the CRM:2009 exercise to inform further matching exercises, and increase the utility of findings. For example, matching the same 2009 cohort to Census 2021, will provide a better assessment of outcomes in the 3 key areas, allowing progress over time to be measured. It will also provide more current data, permitting a more relevant consideration of policy options. Additionally matching multiple cohorts to Census 2021 will allow between group comparisons and increase relevance of findings to policy development.

Separate applications will be made for exercises subsequent to CRM:2009.

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

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

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

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

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

In particular, full account has been taken of the risks associated with:
- transfer of administrative data from data owner to ONS;
- securing access to the data to be linked;
- the management of both the identifying data and attributes contained within the
- administrative data sources;
- the linkage process;
- the identification of population subgroups within the linked data;
- transfer of the de-identified linked file to a secure environment; and
- secure access by researchers at the secure environment.

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

ONS will retain the pseudonymised data for 7 years after delivering the dataset in case there is a requirement to carry out follow-up linkage work. The de-identified dataset will be retained according to the ADRN data retention policies to ensure that the benefits of this research are realised. Within this time the dataset may be used for additional linkage subject to approval of data owners and ethical review of the subsequent projects.

It is not possible to gain informed consent for the matching exercise as the HO do not have contact details for the individuals concerned (refugees are not required to keep in contact with the HO after they have been granted asylum). Consent is not considered to be an appropriate legal basis for this data processing given the imbalance of power between the two parties (it is likely that refugees would feel obliged to consent) and so this has not been cited as a relevant GDPR condition. This in accordance with advice from the ICO who state that:

*Public authorities, employers and other organisations in a position of power over individuals should avoid relying on consent unless they are confident they can demonstrate it is freely given[6].*

Additionally, a key element of consent is enabling participants to withdraw. Given data is anonymised in the final stages it isn't possible to enable removal of data if consent is withdrawn.

Previous HO research with a small group of refugees suggested that this group of refugees were content for their data to be used for research purposes if a) their data was anonymised and b) such research went on to inform the development of asylum policy and improve the experience of refugees. As CRM meets both these conditions, and given the safeguarding in place to govern the use and handling of this data, both the Home Office and Members of the National Asylum Stakeholder forum feel that this sentiment would apply to CRM (see section B5 for further details of stakeholder consultation).

---

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

[6] See https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/lawful-basis-for-processing/consent/

**B4**

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

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

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

**B5**

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

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

Data will only be used for purposes set out for which approvals have been obtained by NSDEC, ADRN, HO and MRP. Data will only be accessed by Home Office accredited researchers, via the SRS. Access by Digital Economy Act accredited researchers will be considered on a case-by-case basis.

No identifying variables will be made available in the SRS. It will not be possible to re-identify individuals in the matched dataset, given it will not be possible to extract record level data from the SRS (and all outputs undergo disclosure control) or import data with the relevant identifiers (nor are researchers permitted to bring this information in to the research in any form).

All staff working on the data will be trained in using large, complex administrative databases and handling sensitive data. This training will be documented. All researchers working on the data in the SRS will also need to have ONS approved researcher status, which ensures they have experience in working with individual level, sensitive population datasets.

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

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

| B6 | **Collaboration and Sponsors** |

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

| List of Collaborators/Sponsors | Details and relevant documentation relating to collaboration (you may attach copies of relevant documentation) |
|---|---|
| Office for National Statistics | **Access to ONS Census data by individual Home Office researchers is permissible** under the Approved Researcher (AR) gateway in S.39 of the Statistics and Registration Service Act 2007, allowing ONS personal information to be accessed for research purposes by researchers holding the necessary AR accreditation. HO researchers accessing the data will be accredited accordingly. |
| Home Office | **Data is already lawfully held by the Home Office** |

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

*Please list any public engagement activities* (max 250 words)

The Home Office's principal means of engagement with its external asylum Non-Government Organisation (NGO) stakeholders is the National Asylum Stakeholder Forum (NASF). NASF recognises that stakeholders' roles are independent of the HO, and aim to scrutinise and hold to account the HO in the running of the UK asylum and resettlement processes. The Asylum Strategic Engagement Group (SEG) is part of the NASF and is comprised of approximately 15 NGOs with a national remit, ideally represented by their Chief Executives. Between them, they have a service/evidence base and/or advocacy/policy profile that gives them a collective overview of the UK asylum and resettlement systems.

A full list of asylum SEG members is presented in Annex C and include UK based charities such as the Refugee Council (whose mission is to be a *champion of refugee rights, transforming refugees' lives, holding governments to account and changing public attitudes for the better*) and the United Nations High Commissioner for Refugees (UNHCR), a global organisation *dedicated to saving lives, protecting rights and building a better future for refugees*.

In September, HOAI presented the CRM:2009 proposals at the quarterly asylum SEG meeting. The concept, data processes, research questions and data usage were clearly outlined and discussed further after the presentation. The project was well received and stakeholders expressed interest in expanding project proposals to answer additional research questions. Where appropriate we have amended project proposals (for example, by including Census data on general health in the returned matched data) but this wasn't possible for all requested additions due to limitations in data availability or a need to restrict the volume and content of data shared.

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

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

> Following completion of the linkage, a report outlining the aims of the exercise, details of the methodology (including success criteria and the extent to which these were met), and high level descriptive counts will be published. This will inform the work of other analysts looking to do similar data linkage exercises, and will benefit the wider research community. The report will be available to all but specifically disseminated to relevant research groups across Whitehall, academia and the private sector.
>
> Additionally we are looking to set up stakeholder group (through our previous engagement with the NASF SEG), and will feedback findings from the linkage to participating members.

## Section C
## Responsible owner and applicant details

| **C1** | **Responsible Owner** |

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

## Declaration to be signed by the responsible owner

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

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

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

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

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

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

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

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

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

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

**Annex A      CRM:2009 NSDEC application,**

**Research questions for CRM:2009**

Overarching:
Is it possible to link HOAG data to Census 2011 data?

Attribute:
1.  Which LAs are refugees living in?
2.  How many remain in dispersal area? How far have they moved?
3.  How many refugees have moved within the last year? What proportion stays within LA/region?
4.  Do refugees move to areas of existing diaspora? Does this effect employment/language outcomes?
5.  How many refugees are owner occupiers or in private or social rented housing? What proportion live in overcrowded accommodation?
6.  What proportion of refugees are living in housing areas targeted for renewal?
7.  What proportion of refugees are living in most deprived 10% LA wards?
8.  What is the employment rate of refugees?
9.  What industries and occupations do refugees work in?
10. What are the characteristics of unemployed refugees? (inc. previous occupation)
11. How many refugees are self employed?
12. How many degree qualified refugees are in low skilled jobs?
13. How many refugees speak English as a main language?
14. What proportion can speak English proficiently? Not at all?
15. What other languages are spoken by refugees?

**Table 1. Variables in CRM:2009 *matched* data set**

| HOAG variables | | RQ |
|---|---|---|
| 1. | Dispersal/resettlement area | RQ2 |
| 2. | Asylum support | All RQs (to explore within group differences) |
| 3. | Grant type | |
| 4. | Accompanied by dependants | |
| Census 2011 variables | | RQ |
| 5. | Passports held/Nationality | All RQs (to explore within group differences) |
| 6. | Sex | |
| 7. | Age | |
| 8. | Country of birth | |
| 9. | Year of arrival | |
| 10. | Age of arrival in UK | |
| 11. | Ethnicity | |
| 12. | General Health | |
| 13. | Long term health problem/disability | |
| 14. | Current residence area | RQs 1-7 |
| 15. | Address one year ago | RQ3 |

| 16. | Tenure | RQ5 |
| --- | --- | --- |
| 17. | Overcrowded household | |
| 18. | Economic activity | RQs 8-11 |
| 19. | Industry | RQs9-10 |
| 20. | Occupation | |
| 21. | Main Language | RQs 13, 15 |
| 22. | Proficiency in English | RQ 14 |

**Annex B CRM:2009 NSDEC application**

NASF Asylum SEG: Members and consultation outcomes

**On the 12th September 2018 Home Office Analysis and Insight presented the CRM:2009 proposals to members of the Asylum Strategic engagement group (SEG), part of the National Asylum Stakeholder Forum (NASF).**

**Members** of the SEG are listed in Table 1. All except Amnesty International and the Welsh Refugee Council attended the meeting and heard the presentation. The presentation and key discussion points were circulated to all SEG members after the meeting.

Table 1: SEG members

| |
|---|
| Amnesty International |
| Asylum Aid |
| British Red Cross |
| Freedom from Torture |
| International Organisation for Migration (IOM) |
| Migrant Help |
| Refugee Action |
| Refugee Children's Consortium |
| Refugee Council |
| Scottish Refugee Council |
| UNHCR |
| UK Lesbian and Gay Immigration Group |
| Welsh Refugee Council |
| Immigration Law Practitioners Association |

# ONS uses of DWP data for Statistical Purposes

## Oral report

Jon Wroth-Smith
Matt Brown

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

---

**Please consult the guidance document before filling in the application form**

---

## Project Title
*Please provide a title indicative of the project*

Informing GLA's local development planning responsibilities under the New London Plan

---

**Start Date:** 16.11.18          **End Date:** 31.01.19

---

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

**Greater London Authority**

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

The Mayor of London has published the Draft New London Plan[7] which sets out expectations and development targets for employment and housing (among other areas) in the city. The Greater London Authority (GLA) would like to determine whether these targets are achievable in the current circumstances and what additional measures may be necessary to ensure they are met. In order to answer these questions, they have commissioned the professional services firm Arup[8] to study economic development in Oxford Street and the West End, studying how future changes might affect Oxford Street and the West End's importance as an employment centre and international retail and entertainment destination.

---

7      https://www.london.gov.uk/what-we-do/planning/london-plan/new-london-plan/draft-new-london-plan/
8 https://www.arup.com/

The study will develop a holistic understanding of the West End economy and its potential to change and grow based on trends in the retail industry, sectoral restructuring and accessibility changes to the West End. Changes in accessibility are particularly topical given the advent of the Elizabeth Line (the Crossrail project) and Westminster Council's recent withdrawal from the plans to partially pedestrianise Oxford Street. They generate an evidence base for future developments of the Oxford Street district and wider West End including a thorough review of risks and opportunities. The results of this project will also contribute to Greater London Authority's ability to meet its obligations as a part of London's Central Activity Zone.

GLA generate numerous reports on social and economic developments in London and publishes them freely on their website[9]. This project, like those before it, while make transparent the reasons for GLA's policy decision-making, as well as sharing the outcomes of their research for public and academic interest.

## Section A
## Project Details

| A1 | Legal gateways |
|---|---|

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

The ONS Approved Researcher scheme is the legal gateway being used to access the ONS-owned data. This complies with the Statistics and Registration Service Act 2007.

The ONS Microdata Release Panel (MRP) have approved the proposal, as there is a legal gateway to access the data, it is deemed appropriate use of ONS data and a public benefit has been demonstrated.

| A2 | Ethical approval |
|---|---|

**Has the project been reviewed or is it expected to be reviewed by another ethics committee?**

◯ Yes    ◉ No

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

---

[9] https://www.london.gov.uk/what-we-do/research-and-analysis/gla-economics-publications

**A3**  **Proposed site of research** select all that apply

---

☑ ONS

☑ SRS (formerly VML)

☐ HMRC Data Lab

☐ Other

please specify

☐ ADRC - England

☐ ADRC - Scotland

☐ ADRC - Northern Ireland

☐ ADRC - Wales

**A4**  **Data subjects to be studied**

---

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

◉ **Yes**    ○ **No**

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

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

n/a

Justification for focusing on these subsections or groups:

n/a

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

The researchers' specific area of interest is The West End – traditionally conpirising Oxford Street, Regent Street and Bond street. This area does not align with a standard ONS administrative division, which limits the number of datasets suitable for this purpose. However the Inter-Departmental Business Register (IDBR) provides the necessary level of geographical data. These data will be accessed in the ONS Secure Research Service (SRS) in de-identified form.

The outputs will be aggregated statistics measuring levels of business growth and employment, and changes in the same in Oxford Street and the West End, by sector, by year. This will be a baseline of statistical evidence for GLA's decisions about the area going forward. The sectors being assessed are food and drink, retail, entertainment and cultural institutions. The researchers are primarily interested in data from the 2008 financial crisis to present, but will extend the work back to 2000 for historical context.

Although Oxford Street and The West End is a small area geographically, the density of commercial activity in this area is such that there is no danger of the researchers identifying individual businesses. For comparison, the West End contributes Gross Value Added of £51 billion per year, while the whole of Wales contributes a GVA of £52 billion[10]. This means there is plentiful scope for business data can be aggregated to protect business identity, without losing the degree of geographical and sectoral detail necessary for GLA's purposes.

---

[10]https://newwestend.com/wp-content/uploads/2016/08/NEW_0015_The_West_End_Powerhouse_DIGITAL.pdf (page 3)

| A6 | **Data use** Please specify the data used **by the research team** including any timeframes *e.g. LFS data 2014-15* |
|---|---|

**Data Level**

*Please specify the name of the data set*

| Type of data | Aggregate Data | Identifiable Data | De-identified personal data | Anonymised/ pseudo anonymised |
|---|---|---|---|---|
| **Administrative data** *(please specify, e.g. Patient Register 2011, School Census 2012 etc, in the relevant options adjacent)* | | | | |
| **Big Data** *(please specify e.g. Twitter data, smart meters and mobile phones, in the relevant options adjacent)* | | | | |
| **Survey Data** *(please specify e.g.LFS, BRES, etc in the relevant options adjacent)* | | | Interdepartmental Business Register, 2000-2018 | |
| **Census Data** *(please specify year, e.g. Census 2011 in the relevant options adjacent)* | | | | |
| **Other** *(please specify e.g. Ordinance Survey Address register in the relevant options adjacent)* | | | | |

# Section B
# Assessment against NSDEC ethical principles

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

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

GLA's primary purpose for this project is to improve their ability to meet the growth and housing targets mandated by the draft New London Plan. The evidence base provided by this project will allow them to make better informed decisions about how those targets may be met, as well informing future policy decisions.

GLA is a strategic regional authority, with powers encompassing a range of policy areas, such as transport, economic development, and fire and emergency planning. This research will provide clear evidence to enable GLA to deliver its statutory objectives, assess how growth and employment targets are met, and enable better informed decisions. These decisions will lead to the evaluation and development of policy in London with clear benefits to businesses in a wide array of sectors as well as to residents in London.

Given the economic importance of this area, policy decisions based on this survey are expected to have national impact. And beyond the New London Plan, the information from this project will help inform future economic development decisions by the GLA. These purposes have direct benefits for the social and economic health of the London region and consequently the UK as a whole.

GLA's numerous reports on social and economic developments in London are published freely on their website[11]. This project's publication will make transparent the reasons for GLA's policy decision-making, as well as sharing the outcomes of this research for public and academic interest.

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

No such products are being produced as a part of this project.

---

[11] https://www.london.gov.uk/what-we-do/research-and-analysis/gla-economics-publications

**B3**

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

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

Access to, analysis and use of the data will only take place within the ONS Secure Research Service (SRS) environment. All outputs will be checked by the Statistical Support team to ensure that they do not disclose confidential or personal identifiable data and to confirm that they are proportional and necessary to the project. This process ensures that disclosure control and the confidentiality of data subjects is protected. Outputs that identify small groups or individuals will not be allowed out of the SRS environment.

The researchers are accredited as ONS Approved Researchers. ONS will have sight of the final report before it is published.

**B4**

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

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

No new technologies are being employed.

**B5**

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

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

Access to the potentially disclosive data will be in the Secure Research Service environment and via an approved legal gateway (Approved Researcher scheme). This complies with the Statistics and Registration Service Act 2007 and the methods used comply with the principles in the Data Protection Act.

## B6   Collaboration and Sponsors

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

| List of Collaborators/Sponsors | Details and relevant documentation relating to collaboration (you may attach copies of relevant documentation) |
| --- | --- |
| Arup | Contracted researchers |
| Greater London Authority | Sponsor |
| | |

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

*Please list any public engagement activities* (max 250 words)

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

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

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

ONS has a commitment to transparency and publishes a register of all Approved Researchers on their website. This includes their organisations, current projects, sponsors and the datasets being used.

The outcomes of this research will be published by GLA in support of the policies they have informed. GLA have a strong track record of publishing the studies that inform their planning and policy decisions[12].

---

[12] https://www.london.gov.uk/what-we-do/research-and-analysis/gla-economics-publications

## Section C

## Responsible owner and applicant details

**C1**     **Responsible Owner**

| Full Name: | Position: |
|---|---|
| Address: | Email: |
| | Telephone: |
| | Organisation: |

## Declaration to be signed by the responsible owner

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

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

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

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

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

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

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


**Signature:** _____ **Date:**

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

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

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

---

**Please consult the guidance document before filling in the application form**

---

## Project Title

*Please provide a title indicative of the project*

| |
|---|
| Estimation of Price Elasticities of Demand for Alcohol and Tobacco Products in the United Kingdom |

---

**Start Date:**   11.06.18        **End Date:**   30.06.18

---

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

| |
|---|
| Deloitte |

## Project Summary

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

---

This is an unsponsored project by Deloitte, which is replicating, updating and expanding existing statistical work by HMRC[13] which suggested that elasticities for alcohol and tobacco products appeared to be increasing over time. Deloitte will use the Living Costs and Food Survey to re-examine the price elasticities of alcohol and tobacco products in the UK.

The project's outputs will be, firstly, a set of elasticity estimates for these products. Secondly it will identify changes and trends in these elasticities over time. Thirdly it will go beyond the scope of the original HMRC research by including new products such as e-cigarettes.

This evidence will be used to form conclusions about which "policy levers" are available to the UK government. For example, whether increased taxes on alcohol and tobacco products

---

[13]   https://www.gov.uk/government/publications/estimation-of-price-elasticities-of-demand-for-alcohol-in-the-uk
https://www.gov.uk/government/publications/econometric-analysis-of-alcohol-consumption-in-the-uk
https://www.gov.uk/government/publications/econometric-analysis-of-cigarette-consumption-in-the-uk

would result in increased tax income and/or decreased public consumption (the two not being mutually exclusive).

The project is intended to be a showcase of Deloitte's work and expertise, as well as being of academic value, improving the body of public evidence on this economic matter. It will confirm the outcomes of prior research, as well as updating and expanding it. The researchers hope that the outcomes will also be of interest to HMRC by providing updated statistical evidence relevant to HMRC policy decision-making.

## Section A
## Project Details

| A1 | Legal gateways |

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

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

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

| A2 | Ethical approval |

**Has the project been reviewed or is it expected to be reviewed by another ethics committee?**

◯ Yes　　◉ No

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

| A3 | Proposed site of research | select all that apply |

☑ ONS

☐ ADRC - England

| ☑ SRS (formerly VML) | ☐ ADRC - Scotland |
|---|---|
| ☐ HMRC Data Lab | ☐ ADRC - Northern Ireland |
| ☐ Other | ☐ ADRC - Wales |
| please specify | |

**A4**  **Data subjects to be studied**

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

◯ **Yes**    ⦿ **No**

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

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

n/a

Justification for focusing on these subsections or groups:

n/a

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

Deloitte will use de-identified Living Costs and Food Survey (LCFS) data to construct a large cross-sectional dataset in the Secure Research Service (SRS), covering household consumption for alcohol and tobacco products each year from 2008 to 2016.

The LCFS provides data on volumes of goods and services consumed by households, by category, and the prices of these goods and services. Non-response bias and misreporting is to be expected for a sizeable proportion of households. Therefore, the researchers will use a statistical model that treats censored variables as an omitted variables problem[14].

Prices will need to be calculated as the ratio of expenditure to volumes consumed. For households which consume zero volumes of a particular alcohol or tobacco product, the price will be calculated based on averages by household type or region. Control factors likely to influence household consumption trends – such as income, socioeconomic group, household size, and region – will also be compiled in the dataset. Prices will be adjusted for inflation using the publicly available Consumer Price Index statistics for alcohol and tobacco.

Once completed, this dataset will be used as a source of descriptive statistics for those households that consume alcohol and tobacco products. The researchers will use this to identify trends in consumption, which will be compared to trends in tax revenues raised, as identified in the publicly available HMRC data[15]. This will provide sufficient information for price elasticities to be calculated with economic modelling techniques. Control variables will be used to account for other influences on household consumption.

---

[14] Heckman Correction Model or Tobit Model
[15] https://www.gov.uk/government/statistics/hmrc-tax-and-nics-receipts-for-the-uk

**A6**

**Data use**

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

| Type of data | Data Level *Please specify the name of the data set* | | | |
|---|---|---|---|---|
| | **Aggregate Data** | **Identifiable Data** | **De-identified personal data** | **Anonymised/ pseudo anonymised** |
| **Administrative data** *(please specify, e.g. Patient Register 2011, School Census 2012 etc, in the relevant options adjacent)* | Consumer Price Index data | | | |
| **Big Data** *(please specify e.g. Twitter data, smart meters and mobile phones, in the relevant options adjacent)* | | | | |
| **Survey Data** *(please specify e.g.LFS, BRES, etc in the relevant options adjacent)* | | | Living Costs and Food Survey, 2006-2016: Secure Access (SN 7047) | |
| **Census Data** *(please specify year, e.g. Census 2011 in the relevant options adjacent)* | | | | |
| **Other** *(please specify e.g. Ordinance Survey Address register in the relevant options adjacent)* | | | | |

# Section B
# Assessment against NSDEC ethical principles

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

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

As previously described, this project will replicate, expand and update existing work. This is increasing the depth of public knowledge regarding these social and economic matters. Deloitte's work is also expected to be of academic interest for projects concerned with alcohol and tobacco consumption habits. It may form the basis for future research projects, just as Sousa (2014) and Czubeck and Johal (2010) gave rise to this project.

The paper's conclusions are expected to identify "policy levers" that the UK government could use to affect consumer spending and consumption habits. Changing goods prices have second-order effects: households purchase less or more, or shift to purchasing alternatives. This project will describe how changes in (traditionally volatile) alcohol and tobacco duties would likely affect household consumption levels in the UK.

Publishing estimated price elasticities will allow organisations that work with tobacco and alcohol users to anticipate changes in throughput based on any future price changes. This would include addiction charities and support services, as well as the National Health Service.

As this project is based on original HRMC research, Deloitte hope to present their findings to HMRC for consideration. While HMRC would be in no way obligated by Deloitte's conclusions, they may be of interest. Its findings, based on a mix of publicly available data and ONS data, could be replicated if deemed relevant to future policy decisions. For example, they may be relevant to decisions about minimum unit pricing.

The price elasticities generated for use in HMRC's forecasting analysis use the same approach as Sousa (2014) and Czubeck and Johal (2010). If this project can improve upon hat methodology then it would be relevant to HMRC's work, improving the quality of their forecasting analysis if adopted.

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

No future use is intended for any such products of this project. Only non-disclosive statistical outputs will be removed from the SRS.

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

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

Access, analysis and use of the data will only take place within the secure ONS Secure Research Service (SRS) environment. All outputs will be checked by the ONS Statistical Support team to ensure that they do not disclose confidential or personally identifiable data and to confirm that they are proportional and necessary to the project. This process ensures that disclosure control and the confidentiality of data subjects is protected. Outputs that identify small groups or individuals will not be allowed out of the Secure Research Service environment.

The researchers are accredited as ONS Approved Researchers. ONS will have sight of the final report before it is published.

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

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

No new technologies are being employed.

| B5 | **Principle 4: Data used and methods employed are consistent with legal requirements such as the Data Protection Act, the Human Rights** |
|---|---|

**Act, the Statistics and Registration Service Act and the common law duty of confidence**

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

Access to the potentially disclosive data will be in the Secure Research Service environment and via an approved legal gateway (Approved Researcher scheme). This is in compliance with the Statistics and Registration Service Act 2007. The methods used are compliant with the principles in the Data Protection Act

## B6 Collaboration and Sponsors

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

| List of Collaborators/Sponsors | Details and relevant documentation relating to collaboration (you may attach copies of relevant documentation) |
|---|---|
| Deloitte | Research organisation |
| | |
| | |

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

*Please list any public engagement activities* (max 250 words)

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

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

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

ONS has a commitment to transparency and publishes a register of all Approved Researchers on their website. This includes their organisations, current projects, sponsors and the datasets being used.

The analysis and results will be published as part of a final report on the Deloitte website. This is expected be approximately a year after the research begins.

## Section C
## Responsible owner and applicant details

**C1**   **Responsible Owner**

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

# Declaration to be signed by the responsible owner

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

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

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

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

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

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

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


**Signature:** _____ **Date:**

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

**Full Name:**

**Address:**

**Position:**

**Email:**

**Telephone:**

**Organisation:**

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

## Oral report
Peter Stokes
Emma Gordon
Paul Jackson

**Child abuse feasibility study**

**Oral report**
Alexa Bradley

# Any other business