19th Meeting of the
National Statistician's Data Ethics Advisory Committee

_____

Minute, Agenda and Papers

_____

Tuesday, 4 February 2020

10:30 – 14:30

Board Room, UK Statistics Authority

London

UK STATISTICS AUTHORITY

NATIONAL STATISTICIAN'S DATA ETHICS ADVISORY COMMITTEE

Agenda

Tuesday 4 February 2020
Drummond Gate London
10:30am – 2:30pm

(10:30am to 12:20pm)

1 10:30am	Minute and matters arising from the previous meeting	Presentation Simon Whitworth
2 10:45am	Ethics in AI Training Session	Presentation Jasmine Grimsley
3 11:30 am	Pilot record linkage study: Vulnerable Persons and Vulnerable Children's Resettlement Scheme data	NSDEC(20)01 Nicky Rogers, Louisa Blackwell & Eveie Shaw
4 12:00am	Developing ONS' Household Financial Statistics through the linkage of administrative data to household financial survey data	NSDEC(20)02 Matt Greenaway

Lunch (12:20pm to 12:50pm)

(12:50pm to 2:30pm)

5 12:50pm	Investigating the risk factors and inequalities associated with infant mortality	NSDEC(20)03 Jonny Tinsley
6 1:20pm	Social Factors Impacting Offending and Resettlement in Northern Ireland	NSDEC(20)04 Shannon Browne & Rodney Redmond
7 1:40pm	Uses of Admin Data to support future population and social statistics	Presentation Peter Jones
8 2:00pm	Data Protection and Ethics Compliance	NSDEC(20)05 Rhys Nadin
9 2:20pm	Any other business	

National Statistician's Data Ethics Advisory Committee Minute

Tuesday, 4 February 2020
Boardroom, Drummond Gate, London

Present
Members
Dame Moira Gibb (Chair)
Mr Stephen Balchin
Mr Rob Bumpstead
Ms Vanessa Cuthill
Ms Isabel Nisbet
Dr Emma Uprichard

UK Statistics Authority
Dr Emily Mason-Apps
Mr Rhys Nadin (on phone for item 8)
Dr Simon Whitworth

Office for National Statistics
Ms Louisa Blackwell (for item 3)
Mr Matt Greenaway (for item 4)
Ms Jasmine Grimsley (for item 2)
Mr Peter Jones (for item 7)
Ms Nicky Rogers (for item 3)
Mr Jonny Tinsley (for item 5)
Ms Gemma Quayle (on phone for item 3)

Other
Ms Shannon Browne (on phone for item 6)
Ms Maria Calam (for item 3)
Mr Rodney Redmond (on phone for item 6)
Ms Eveie Shaw (for item 3)

Apologies
Mr Colin Godbold
Dr Brent Mittelstadt
Ms Marion Oswald

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

1.2 Members approved the minutes from the previous meeting.

1.3 Dr Whitworth updated the meeting with progress on actions from previous meetings. All actions were complete.

1.4   Dr Whitworth provided the Committee with an update on the ethics self-assessment process. The ethics self-assessment tool is being widely used by ONS researchers, as well as by external researchers across Government, academia and the commercial community.

2.   Ethics in AI Training Session

2.1   Ms Jasmine Grimsley from the ONS Data Science Campus presented this item. Ms Grimsley focussed on the potential sources of bias that can be introduced during different stages of an AI project, and how to validate and maintain AI systems.

2.2   The Committee welcomed this session and recognised the value this training adds in enabling all those involved with AI projects to be aware of the ethical risks that may present throughout every stage of the project lifecycle, right from consideration of data sources, through to interpretation and maintenance of outputs.

3.   Pilot record linkage study: Vulnerable Persons and Vulnerable Children's Resettlement Scheme data [NSDEC(20)01]

3.1   Ms Louisa Blackwell and Ms Nicky Rogers from ONS, along with Ms Maria Calam and Ms Eveie Shaw from the Home Office presented a proposal for a pilot project to link a sample of Vulnerable Persons and Vulnerable Children's Resettlement Scheme (VPRS/VCRS) data to administrative data held by ONS. The results of this pilot project will inform the feasibility of a longer-term project to design a longitudinal cohort study to monitor the integration outcomes of those included in the VPRS/VCRS for a 5-year period after their arrival in the UK. Insights gained from a longitudinal cohort study would be used to inform refugee and asylum policy in order to improve integration outcomes for these groups.

3.2   The Committee were informed that this project proposal had also been reviewed and approved by the Home Office Analysis and Insight Ethics Advisory Group, and that the researchers will continue to engage with this group throughout the timeframe of the project.

3.3   The following points were raised in the discussion that followed:
   i.   The researchers confirmed that this proposal is just for the feasibility stage of this work. A separate application will be presented to the Committee in the event that the longer-term longitudinal cohort study goes ahead.
   ii.   The linked dataset will be held and owned by ONS.
   iii. The dataset will be used to produce aggregate outputs to inform policy makers, and will not be used for operational purposes at any stage.
   iv.   The researchers are conscious of the potential challenges that may impact data linkage quality, and are engaging with colleagues internally, externally and internationally to enable knowledge sharing and awareness of best practice.
   v.   The Committee highlighted that this work may benefit from engaging academic and wider research communities.

vi. The Committee suggested that there should be more information included in the application about the potential benefits this work will provide for refugees.

3.4 The research team confirmed that if the longer-term longitudinal cohort study is deemed to be feasible then they would make an application to do this work to the Committee in the Autumn of 2020.

3.5 This project was approved, subject to more information about the benefits of this project for refugees being included in the application.

3.6 **Action: Ms Louisa Blackwell and Ms Nicky Rogers to provide an updated application to the secretariat.**

4. Developing ONS' Household Financial Statistics through the linkage of administrative data to household financial survey data [NSDEC(20)02]

4.1 Mr Matt Greenaway from ONS presented a proposal to conduct feasibility research to compare data from ONS' household financial surveys (Living Cost and Food Survey, Survey of Living Conditions, and the Wealth and Assets Survey) to DWP administrative data (Customer Information System, and Benefits and Income Dataset) on income and earnings. The purpose of this work is to understand how household financial statistics can be improved with the use of administrative data. Such improvements could include the measurement of high and low incomes, small-area income statistics, and accounting for survey non-response. This work feeds into a longer-term aim of putting administrative data at the core of household statistics, thereby limiting the need to ask additional questions via surveys which would result in cost saving and reduced respondent burden.

4.2 The Committee acknowledged that the information provided to respondents meets all legal requirements and approved the research. The Committee suggested that it would be worthwhile to conduct some research to explore respondents' understanding of the current phrasing and terminology used in respondent materials. This would also be beneficial for the wider transformation programme.

4.3 **Action: ONS to explore public understanding of the current phrasing and terminology used in survey respondent materials, and to report the findings of this work back to the Committee.**

5. Investigating the risk factors and inequalities associated with infant mortality [NSDEC(20)03]

5.1 Mr Jonny Tinsley presented a proposal to explore the risk factors and inequalities associated with infant mortality and stillbirth. The ONS already produce statistics on child mortality, which this work intends to build upon. This project will take a phased approached, with phase one linking ONS owned birth and mortality data to Census data. Phase two will expand on this work to explore other risk factors by linking this data to Hospital Episodes Data and, once acquired, the Maternity Services Dataset.

5.2 This work aims to fill evidence gaps by generating new insights in order to inform policy and guide better decisions in the design and provision of maternal healthcare, with the ultimate aim of supporting the Government's ambition to halve stillbirth and neonatal mortality rates by 2025.

5.3 The following points were raised in the discussions around this proposal:
   i. The researchers confirmed that this work aims to produce aggregate statistics that highlight potential risk factors to inform policy, and will not be used to create predictive models that indicate risk at an individual level.
   ii. The Committee highlighted the importance of ensuring that the findings of this work are communicated clearly and sensitively to enable appropriate and proportionate interpretation.
   iii. The Committee recommended that the researchers clearly acknowledge the limitations and biases that may exist in data sources when interpreting the findings of this work.
   iv. The Committee suggested the researchers consider what scope there would be to extend this work to include other known or potential risk factors that are not captured within the proposed datasets.
   v. The Committee suggested that the researchers tightly define what they mean by predictive models.

5.4 The Committee approved phase one of this work, subject to the researchers amending the application in light of the points discussed above. The Committee asked that the researchers present a revised application for phase two at a future meeting.

5.5 **Action: Mr Jonny Tinsley to update the application by addressing the points discussed in section 5.3 and provide this to the secretariat. Mr Tinsley to provide the Committee with a revised application for phase two of this research.**

6. Social factors impacting offending and resettlement in Northern Ireland [NSDEC(20)04]

6.1 Mr Rodney Redmond and Ms Shannon Browne from the Department of Justice (DoJ) in Northern Ireland, and Mr John Hughes from the Northern Ireland Statistics and Research Agency (NISRA) presented a project proposal aiming to improve the evidence base on the links between the offender population and mortality in Northern Ireland. The project will involve linking the Discharges and Reoffending Dataset (DoJ), the Custody Dataset (DoJ), and the General Registrar's Office Death Registration Dataset (NISRA). The researchers aim to specifically examine if there are high risk periods where risk of mortality is higher, for example in the first few weeks of imprisonment, or the immediate weeks post release from custody.

6.2 This project forms the first stage of a wide ranging DoJ research programme to better understand the issues impacting upon propensity to re-offend, and the risks that offenders may face when released from custody. The findings from this study will be used to produce a publicly available report to inform policy.

6.3 The following points were made in the discussion that followed:
  i. It was confirmed that the Committee were only being asked to consider the proposal for this specific project, and not the wider programme of work.
  ii. The linked dataset will be owned by NISRA and held within the NISRA Research Support Unit.
  iii. The Committee recommended that the researchers be vigilant of the possibility that some of the variables may be proxies for sensitive data.
  iv. The researchers explained that all outputs will go through strict statistical disclosure controls to safeguard against any possibility of direct or indirect identification of individuals.
  v. The Committee suggested that the researchers explore the utility of sequence analysis.

6.4 The Committee approved phase one of this work, subject to the researchers addressing the amendments discussed above in a revised application.

6.5 **Action: Mr Rodney Redmond and Ms Shannon Browne to provide an updated application to the secretariat.**

7. Uses of Administrative Data to support future population and social statistics

7.1 Mr Peter Jones from ONS presented an overview of ONS's transformation work which aims to move towards an admin led approach for the production of population, social and business statistics. This work aims to improve the breadth, quality, timeliness and geographic detail of ONS statistics.

7.2 The Committee acknowledged the scope and value of this programme of work, and emphasised the importance of seeking and incorporating the views of the public as this work progresses.

7.3 The Committee recommended the ethics self-assessment as a useful tool to consider the ethics of the different strands of this work, and welcome an update at a future meeting.

8. Data Protection and Ethics Compliance [NSDEC(20)05]

8.1 Mr Rhys Nadin from the UK Statistics Authority presented this item. This paper presented a data ethics compliance report following a review of the project "Extending the Crime Survey for England and Wales (CSEW) to include a new module of questions on the online behaviour of children aged 10-15 years [NSDEC(18)17]". This review confirmed that all the recommendations made by the Committee had been satisfactorily implemented. The paper also presented a Data Ethics Compliance Review Plan which proposes a list of projects for future reviews

8.2 The Committee welcomed this review and commented on the value of this process in providing feedback that the Committee can learn from.

8.3 The Committee made the following recommendations and comments:
  i. Mr Nadin and the secretariat should develop a list of clear objective criteria for selecting which projects are selected for review.

ii. To gauge how widely the recommendations and knowledge are disseminated throughout research teams, it would be beneficial to talk to a range of researchers who are working on the project, as well as the project lead.

iii. The Committee recommended conducting reviews for a sample of projects that have used the ethical self-assessment tool.

8.4 **Action: The UK Statistics Authority to include the points in section 8.3 in future compliance work.**

9. Any other business

9.1 There was no other business.

UK STATISTICS AUTHORITY

NATIONAL STATISTICIAN'S DATA ETHICS ADVISORY COMMITTEE

Ethics in AI Training Session

Presentation

Jasmine Grimsley

NSDEC(20)01

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

Please consult the guidance document before filling in the application form

Project Title
Please provide a title indicative of the project

Pilot record linkage study: Vulnerable Persons and Vulnerable Children's Resettlement Scheme data

Start Date: February 2020   **End Date:** October 2020

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

Office for National Statistics (ONS) Methodology Directorate

Home Office Migration and Border Analysis, Home Office Analysis and Insight (HOAI)

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

We are seeking NSDEC approval for a pilot project to link a sample of Vulnerable Persons and Vulnerable Children's Resettlement Scheme data (VPRS/VCRS) to administrative data within ONS.

If approved, the linkage will be undertaken by ONS data linkage experts within the ONS Data Access Platform (DAP) secure environment. Upon completion of the pilot project, results will be used to inform the feasibility of a longer-term project to design a longitudinal cohort study based on VPRS/VCRS. It is our intention to seek approval from the National Statistician's Data Ethics Advisory Committee (NSDEC) further down the line on the design and benefits of such a study prior to implementation.

The aim of this pilot project is to assess the feasibility of linking a sample of Vulnerable Persons and Vulnerable Children's Resettlement Schemes (VPRS/VCRS) data to administrative data held by the ONS to aid the development of linkage strategies for

different naming conventions for refugee nationals. In the first instance we plan to link VPRS/VCRS data to Home Office Exit Checks data, GP Patient Register (PR), and Patient Demographic Service (PDS) data. We have identified Exit Checks, GP Patient Register and Patient Demographic Service data as the best data sources to assess linkage rates and the viability of a future cohort study. We expect refugees to be present in the Exit Checks data, and as part of resettlement they will be registered with a GP. The sample of refugees will be for England and Wales only.

This will be the first stage of a proposed long-term project to design a longitudinal cohort study to support evaluation of the VPRS/VCRS. The longitudinal design will follow international best practice to support statistics on refugee integration.

The benefits from the pilot project are two-fold:

1) The pilot will provide unique new insights for linkage strategies for different naming conventions being developed by ONS for a wider programme of work to place administrative data at the core of our population and migration statistics system.

2) The pilot project will support decision-making on the outline feasibility and design of a UK cohort study to understand social integration outcomes (including employment, education, health and English Language) for these vulnerable groups.

Insights gained from a future longitudinal cohort study will be used to inform refugee policy to improve integration outcomes for this group, which ultimately will lead to refugees benefitting from schemes and policies that address their most pressing needs.
VPRS and VCRS were established by the UK Government to resettle vulnerable adults and children fleeing the current conflict in the MENA region[1]. The schemes are on track to resettle 20,000 people under the VPRS and intend to settle 3,000 children under the VCRS between 2015 and 2020. The Home Office currently collects monitoring data on refugees during their first 15 months of settlement in the UK. Data after this time period are not collected, in order to minimise the burden on refugees, local authorities and voluntary groups that provide these data. This is mainly due to the challenges in collecting these data directly from refugees in the longer-term as intensive caseworker support steps down. As integration is a long-term process, the Home Office would like to understand integration outcomes for a 5-year period after their arrival in the UK. These include English language classes, education and employment, secondary migration and health outcomes. Annex 1 lists key variables in the VPRS/VCRS monitoring dataset that will aid data linkage.

It is also widely accepted that quantitative evidence on outcomes for refugees is generally lacking and this is largely attributed to a lack of datasets which permit refugees to be identified (Ruiz & Vargas-Silva, 2018)[2]. The use of longitudinally linked administrative data for this population sub-group will not only fulfil this outcome, but will reduce data collection costs, reduce respondent burden and improve the quality and granularity of the statistics produced on these vulnerable groups.

ONS is currently progressing a programme of research which aims to place administrative data at the core of our population statistics system by 2020. We are working in partnership

---

[1] Middle East and North Africa Region consisting of 19 countries: Algeria, Bahrain, Egypt, Iran, Iraq, Israel, Jordan, Kuwait, Lebanon, Libya, Morocco, Oman, Palestine, Qatar, Saudi Arabia, Syria, Tunisia, United Arab Emirates, and Yemen

[2] *Journal of Economic Geography,* Volume 18, Issue 4, 1 July 2018, Pages 855-885

with colleagues across the Government Statistical Service, including the Home Office, to transform the statistics available, to improve the evidence base for migration, and to research and understand migrants' experiences in the UK. The use of administrative data to fill evidence gaps benefits the public by ensuring that both the Home Office and ONS add value through statistics that support society's need for information. This is in line with the Code of Practice for Official Statistics – "Value means that the statistics and data are useful, easy to access, remain relevant, and support understanding of important issues".

Section A
Project Details

A1   Legal gateways

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

Section 45A of the Statistics and Registration Service Act provides a permissive power for Crown bodies and public authorities to share the data they hold in connection with their functions, with ONS. This is the legal gateway the Home Office will use to share their data with ONS.

A legal gateway is not required for Patient Demographic Service (PDS), Patient Register (PR) and Exit Checks data as these are already held by ONS.

A2   Ethical approval

Has the project been reviewed or is it expected to be reviewed by another ethics committee? *(please delete as appropriate)*	**Yes**

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

Home Office Analysis and Insight (HOAI) Ethics Advisory Group reviewed the application on 20.01.2020. The group provides support and advise on ethical issues for analysts in the Home Office.

The group provided us with recommendations for appropriate data handling and data usage that have been applied to this application. They also advised on some long-term considerations that will be applied to the longitudinal project should the pilot be successful. This included a recommendation for a series of set 'Ethics Refreshers' to ensure that ethics does not become an initial tick-box exercise but is considered iteratively across the timeframe of the project.

The review concluded that the ethical considerations for this pilot data-linkage exercise have been appropriately considered. If the pilot results in a longitudinal cohort study, the HOAI Ethics Advisory Group can continue to support the development of this.

A3   Proposed site of research

Please provide information about where the research will take place, including detail of where any data linkage and/or data analysis will be conducted

Data will be held and linked within the ONS secure Data Access Platform (DAP). Only experienced ONS approved analysts will undertake data linkage in a restricted project area in DAP.

Following linkage, the identifying match keys will be removed to create a de-identified dataset. This dataset will then be moved to a separate project area in DAP for analysis by ONS experts. There will be a clear separation between ONS analysts who access identifiable data, and those who access the resultant de-identified linked/unlinked data.

A4   Data subjects to be studied

Does the study include all subsections of the population?
(i.e. all ages, sex, ethnic groups etc.) *(please delete as appropriate)*   **No**

*If **No,** please detail which subsections with justification(s) below:*

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

The pilot linkage project will focus on a sample of vulnerable persons and children who were resettled to the UK through the VPRS/VCRS because they were identified by the United Nations High Commission for Refugees (UNHCR) as being in the greatest need, including people requiring urgent medical treatment, survivors of violence and torture, and women and children at risk.

Justification for focusing on these subsections or groups:

The VPRS/VCRS aims to offer a safe and legal route to the UK for the most vulnerable refugees. The scheme purposely targets those in greatest need of assistance. The UK Government's approach is to take refugees directly from non-European countries where resettlement may be the only durable solution, often from the region bordering countries with conflicts. In the case of the VPRS/VCRS, this means those currently in Egypt, Iraq, Jordan, Lebanon and Turkey.

There are unique challenges in the use of linked administrative data, for example bias from linkage errors where records cannot be linked or are linked together incorrectly. This is particularly challenging where unique identifiers for linkage across data sources are not available and there is reliance on name, address, gender, date of birth and even nationality.
The treatment of different naming conventions in data linkage is equally challenging. Name can be a highly discriminative variable in data linkage, however the number of

different ways it can be structured can be problematic and we need to understand this and develop algorithms to optimise linkage. Missed linkages can result in bias if subgroups of records are more or less likely to link (Bohensky et al., 2010; Ford et al., 2006; Lariscy, 2011)[3].

Depending on the data source, studies have found that data quality varies according to several characteristics including age, sex, ethnicity and health status (Bohensky, 2015).[4] Potentially, this could lead to lower linkage rates in more vulnerable or deprived populations and therefore underrepresentation in our population and migration estimates and in reporting on outcomes for those resettled via the VPRS/VCRS.

Ultimately, if this pilot linkage project is successful it will inform future analysis of linked administrative data.

A5	**Research methodology and protocol**

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

As the Home Office does not hold up-to-date contact details for refugees (refugees are not required to keep in contact with the Home Office), the optimal way to gather this evidence is via linking the administrative data that the Home Office holds on those who have been resettled through the VPRS/VCRS to other administrative data sources.

The treatment of different naming conventions in data linkage is challenging and will affect data linkage quality. For example, if subgroups of records are less likely to link due to complexity around naming conventions. Equally, we need to assess linkage rates to evaluate the viability of a future cohort study. To do this we need to link to administrative data sources that have a high probability of including those on the VPRS/VCRS. We have therefore identified Exit Checks, GP Patient Register and Patient Demographic Service data as meeting this need since refugees will be present in the Exit Checks data and as part of resettlement they will be registered with a GP.

Linking to these sources will allow us to use a range of variables to improve linkage rates, for example name, date of birth, sex, nationality, and postcode. In addition, time/event variables such as application date/journey dates/registration dates with GPs will enable us to narrow down the search space and potentially reduce false positive linkages. Associative matching will allow the ability to make links based on family connections, for example using family names as some refugees may arrive in family groups.

Optimisation of record linkage is critical for this pilot study given the size of the refugee population group, the potential for false positive or negative matching to generate falsely linked or unlinked records, and known record linkage challenges for migrant groups.

The only alternative would be a series of data linkage exercises with a range of

---

[3] Bohensky, M, et al (2010) Data linkage: A powerful research tool with potential problems. *BMC Health Services Research* 12:480. Ford et al (2006) Characteristics of unmatched maternal and bay records in linked birth records and hospital discharge data. *Paediatric and Perinatal Epidemiology* 20: 329-337   Lariscy JT (2011) Differential record linkage by Hispanic ethnicity and age in linked mortality studies. *Journal of Aging Health* 23: 1263-1284

[4] Bohensky M (2015) Bias in data linkage studies. In: Harron K. Dibben C and Goldstein H (eds) *Methodological Developments in Data Linkage.* Chapter 4 London: Wiley.

stakeholders/Other Government Departments which would be a fragmented approach and the quality of data linkage would be difficult to control. Therefore, meaningful linkage diagnostics would be harder to collect. This approach would increase privacy risks given the need to share data more widely and carry out multiple linkages.

A full list of variables to be used in the pilot data linkage is provided in Annexes 1 and 2. We have also provided a justification for the inclusion of each variable. In the absence of common identifiers across the data sources, we propose linking records using a combination of name, date of birth, gender and geographic location as a starting point. These identifiers are available on the VPRS/VCRS monitoring datasets. Feasibility research requires that we make use of a wide range of variables to ensure we can recommend the most appropriate linkage strategy going forward, therefore we have included other variables such as nationality, time/event variables, and variables that will aid establishment of group/family membership to help the data linkage process via associative matching.

As we plan to use identifiable record level data, all data linkage will be undertaken in the ONS secure Data Access Platform (DAP) within a restricted project only accessed by ONS experienced named researchers. Following linkage, the identifying match keys will be removed to create a de-identified dataset which will be transferred to a separate project in DAP for analysis of characteristics (for example by age, sex and nationality). Analysis of the characteristics of linked and unlinked data will help us understand the representativeness of our linked dataset, for example we may be more successful in linking females than males or certain nationalities.

Analysis of the linked and unlinked data will also be undertaken to calculate match rates for each matching pass and calculation of false positive (where we have incorrectly linked records) or false negatives (where we have incorrectly not linked records). We will also use the pilot to model attrition rates that can be experienced by linking data longitudinally across time. This will help us assess the viability of the proposed cohort study. This will be undertaken by analysts with access to the identifiable data. Identifiable data will be deleted once the pilot study is completed. We will draw from both these analyses in the future development of matching algorithms for different naming conventions.

There will be a clear separation between ONS analysts who access identifying data, and those who access the resultant linked/unlinked de-identified data.

ONS will retain the de-identified linked dataset for the lifetime of the proposed cohort study should this go-ahead to inform design and linkage. Should the cohort study not go ahead all VPRS/VCRS data and linked datasets will be deleted. There will be no onward sharing of the linked data to the Home Office.

Following completion of the linkage, a report outlining the aims and outcomes of the exercise and methodology (including linkage strategy and quality criteria) will be shared with the Home Office and published via the ONS/Government Statistical Service Methodology websites. This work will be crucial to informing the work of analysts working on data linkage as part of the Population and Migration Statistics Transformation programme, but potentially could have wider use for analysts across the UK Government Statistical Service, the wider research community and inform international best practice on data linkage.

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 dataset

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)*		Home Office VPRS/VCRS data 2015-2019 for England and Wales Home Office Exit Checks Data 2015-2019 GP Patient Register Data 2015-2019 for England and Wales Patient Demographic Service Data 2015-2019 for England and Wales		
Big Data *(please specify e.g. Twitter data, smart meters and mobile phones, in the relevant options adjacent)*				
Survey Data *(please specify e.g.LFS, BRES, etc in the relevant options adjacent)*				

Census Data *(please specify year, e.g. Census 2011 in the relevant options adjacent)*				
Other *(please specify e.g. Ordinance Survey Address register in the relevant options adjacent)*				

Section B
Assessment against NSDEC ethical principles

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

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

The Home Office has a responsibility to 'protect vulnerable people and communities' (2.4) and in its single department plan to achieve this it aims to: "Identify and safeguard vulnerable people to protect victims of trafficking, modern slavery, detainees and children at risk of abduction".

This pilot project is the first step in developing a rich data source to support the Home Office in achieving this goal. If the pilot linkage project is successful, this will enable us to start designing a longitudinal cohort study with administrative data for these vulnerable groups which will provide valuable insights into their integration outcomes. Outcomes include English language, education and employment, secondary migration and health outcomes. These insights also have the potential to inform refugee and asylum policy and progress against integration outcomes.

Insights gained from a future longitudinal cohort study will be used to inform refugee policy in order to improve integration outcomes for this group. We plan to do this by sharing findings with local authorities and key stakeholders working with refugees. Ultimately refugees will benefit from schemes and policies that address their most pressing needs. For example, if the study reveals that certain demographic groups are less likely to find employment, this can inform targeted activity to support groups most in need/most likely to benefit.

The pilot project will act as a catalyst to filling an evidence gap and supporting evidence-based policy-making for these vulnerable groups. 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, for example, opportunity and prosperity and lower levels of prejudice and hate crime (Integrated Communities Strategy, 2018).

ONS is currently progressing a programme of research which aims to place administrative data at the core of our population statistics system by 2020. We are working in partnership with colleagues across the Government Statistical Service, including the Home Office, to transform the statistics available, improve the evidence base for migration, and to research and understand migrants' experiences in the UK. The pilot linkage study will help enrich ONS data linkage expertise, particularly for hard to reach populations. This valuable learning

will inform linkage work being undertaken as part of the Migration and Population Statistics Transformation Programme. Ultimately improvements to the evidence base for international migration will benefit the public through the ability of policy-makers, planners and service providers to monitor and evaluate the impact of future policy changes, and consequently leading to improved and timely allocation of resources where they are needed.

The use of administrative data to fill evidence gaps benefits the public by ensuring that both the Home Office and ONS provide value with statistics that support society's need for information. This is in line with the Code of Practice for Official Statistics – "Value means that the statistics and data are useful, easy to access, remain relevant, and support understanding of important issues".

B2	**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)

Data linkage using identifiers will be undertaken by ONS within a secure data environment (DAP). Only experienced ONS approved analysts will undertake data linkage in a restricted project area. All staff handling the data are required to have national security clearance for the data they are going to handle. Adherence to data confidentiality is also part of the contract of any staff at ONS. In addition, staff with access to the data are required to sign a declaration that they understand these restrictions.

Following linkage, the identifying match keys will be removed from the linked and unlinked datasets to create de-identified datasets. These de-identified datasets will then be transferred to a separate project in DAP. Analysts not involved in data linkage will carry out analysis on these datasets. There will be a clear separation between ONS analysts who access identifying data, and those who access de-identified linked/unlinked data.

A privacy notice is given to VPRS/VCRS refugees by Local Authorities when they are resettled in the UK which explains the use of their data. 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 is 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.[5]

Analysis on the identifiable data will include data linkage indicators (linkage rates, false positives etc) . Analysis by key characteristics will be undertaken on the de-identified linked and unlinked data to fully understand data representativeness. Analysis will only be released from the secure environment once disclosure control checks have confirmed that the analysis contains no risk of identifying an individual. We are seeking guidance from the ONS

---

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

Statistical Disclosure team on the best approach to statistical disclosure control.

Analysis will be presented as aggregate univariate and multivariate tables and shared with the Home Office. Such aggregate analysis will be used to inform the development of a longitudinal cohort study for this population sub-group and where appropriate, guide the development of linkage strategies for administrative data to support the Migration and Population Statistics Transformation Programme. The analysis will not be used for operational purposes by the Home Office.

Identifiable data will be deleted once the pilot study is completed. ONS will retain the de-identified linked dataset for the lifetime of the proposed longitudinal cohort study should this go-ahead to inform design and linkage. Should the longitudinal cohort study not go ahead all VPRS/VCRS data and linked datasets will be deleted. There will be no onward sharing of the linked data to the Home Office.

B3	**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)

Linkage will be undertaken by experienced ONS analysts using appropriate tried and tested data linkage methods. This pilot study provides an opportunity for ONS linkage experts to collaborate with the Home Office's Data Analytics Competence Centre (DACC) matching experts and this cross-fertilisation of skills and experience should deliver improvements in ONS matching capability, particularly for different naming conventions. The potential benefits for office-wide record linkage will be more representative record linkage studies, if we are able to reduce false positive matching. Methodological improvements flowing from this pilot study will be disseminated across ONS and the Home Office.

B4	**Principle 4: Data used and methods employed are consistent with legal requirements such as Data Protection Legislation[6], the Human Rights Act 1998, the Statistics and Registration Service Act 2007 and the common law duty of confidence**

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

The data used and methods we will employ are consistent with legal requirements under data protection legislation. We have also undertaken a Data Protection Impact Assessment for the proposed pilot project. This has confirmed that the way we plan to use these data are compliant with data protection legislation, and adequately protects the rights and freedoms of those whose data we are entrusted with.

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

---

[6] "Data Protection Legislation" means the full, applicable data protection framework as set out in the Data Protection Act 2018. This encompasses general processing (including the General Data Protection Regulation and the applied GDPR).

The data processing is in accordance with the EU General Data Protection Regulation (GDPR) as it meets conditions in article 6 (required for processing personal 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.

The proposed data sharing is consistent with the [Home Office Personal Information Charter](link) (HO departmental Privacy Notice). This charter notes that personal data held by the HO on individuals may be shared with other government departments and agencies and that data held on applicants may be used to undertake statistical and analytical analysis.

Our use of data is proportionate as the pilot linkage study can be undertaken on a sample of data rather than the entire VPRS/VCRS database. We have listed variables we plan to use for data linkage in Annexes 1 and 2.

A privacy notice is given to VPRS/VCRS refugees by Local Authorities when they are resettled in the UK which explains the use of their data.

B5	**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)

Migration and Borders Analysis at the Home Office have presented the VPRS/VCRS Monitoring findings at multiple regional away days. The potential to explore longer-term integration outcomes using administrative data was discussed as part of this work. Local Authorities and delivery partners have shown a keen interest in this work. The Resettlement Strategic Engagement Group (RSEG) are also in support of this work; the group focuses on the operational delivery of the UK's resettlement routes, providing expertise that will help inform the current and future work of the schemes as they merge into one into 2020. Membership of the Group is limited to around 28 stakeholders with direct experience of resettling refugees in the UK to provide targeted expertise that will help inform the delivery of the UK's resettlement schemes.

This pilot project will also inform plans to transform the information that the Government Statistical Service produces on migration to meet the demand for more evidence-based information on international migration and migrant sub-groups.

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

This research and the migration statistics transformation programme are supported by several published papers, updates and engagement activities taking place over the last few years. In January 2019 ONS published a [research engagement report](link) on their findings from the feasibility research on using linked administrative data to provide international migration flows and will be seeking feedback on user needs across the international migration evidence base. Several engagement activities took place during 2019 to ensure this research and the wider programme meets user needs to serve the public good.

B6	**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 pilot linkage project, a report outlining the aims and outcomes from the exercise, details of the methodology (including data linkage quality criteria and the extent to which these were met) and high-level descriptive statistics on the characteristics of the linked and unlinked data will be published. This work will be crucial to informing the development of a longitudinal cohort study for this population sub-group, as well as the work of analysts working on data linkage as part of the Population and Migration Statistics Transformation programme. It also will have wider use for analysts across the UK Government Statistical Service, the wider research community and inform international best practice on data linkage. The Home Office will also feedback to VPRS/VCRS stakeholder groups to seek their views on the pilot and of the potential longitudinal study that will follow.

B7	**Future use of products**

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

The de-identified linked and unlinked datasets will be held in the ONS secure environment for as long as they are needed for research and statistical purposes such as to inform the development of a longitudinal cohort study for this population sub-group. Analysts with access to the data will have had comprehensive training on access and analysing sensitive data and will have appropriate security clearance.

Access to the identifiable data will be limited to a small number of linkage experts for the purpose of record linkage only. Identifiable data will be deleted once the pilot study is complete.

Should the cohort study not go ahead all VPRS/VCRS data and linked datasets will be deleted. There will be no onward sharing of the linked data to the Home Office.

B8   Collaboration and Sponsors

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

List of Collaborators/Sponsors	Details and relevant documentation relating to collaboration (you may attach copies of relevant documentation)
Office for National Statistics	The 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 we plan to link to under this pilot project have been acquired using established legal gateways cleared by ONS Legal Services (Section A1 and B4)
Home Office	VPRS/VCRS data are already lawfully held by the Home Office (Section A1)

Section C
Responsible owner and applicant details

C1   Responsible Owner

Full Name: Dr Louisa Blackwell	**Position:** Principal Statistical Methodologist
Address:	**Email:**
	Telephone:
	Organisation: Office for National Statistics

Declaration to be signed by the responsible owner

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

I understand that it is a requirement for all researchers accessing the data to have undergone relevant training and to have either relevant security clearances or accredited/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: Louisa Blackwell   **Date:**   24/01/2020

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

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

Annex 1: List of linkage variables held in Home Office Vulnerable Person/Children dataset

Full name
Involves first, middle and last names, and any other names stated on official documents. The inclusion of full names in the pilot would especially be useful since not all datasets include national insurance numbers or person-specific identifiers other than names.

The linkage exercise would help to aid understanding of the following effects on data linkage:
- Variations in the spelling of the full name in a Latin alphabet due to errors in translation, a lack of knowledge of what likely variations of the same name may be, and when two separate names are misidentified as merely variations of one name.
- Some names will not have Arabic origins, and linkage on full names may be easier or more difficult for refugees from specific ethnic groups.

In addition, full names are helpful in ruling out false positive linkage of datasets based on other variables. This is especially applicable in the case of date of birth, considering the possibility that the 1st of January may be the most common date of birth occurring on official documents for some nationalities.

Date of birth
Involves date of birth from the refugee as stated in their official documents. As noted above, there is a possibility that the 1st of January is the most common date of birth among some nationalities, which is often assigned when no birth records exist, and no date of birth is known. This is especially common if someone is born in conflict.

The linkage exercise would help to aid understanding of the effect of certain common dates of births on linkage rates, and whether these effects vary by age.

In addition, information on date of birth can be used to provide clarity in cases of refugees with the same or similar names.

Gender
Involves the gender of the refugee. It is possible it only refers to self-reported gender, rather than someone's sex at birth. As the study covers vulnerable people, it may include those whose gender differ from their biological sex. However, the gender variable would not identify to which refugees this would apply.

The linkage exercise would help to aid understanding of the effect of gender on linkage rates, especially for linkage based on full names.

In addition, gender information can help prevent linkage errors in cases where male and female names do not differ much from each other, and where date of birth has not offered any additional insight.

Nationality
The nationality of a refugee in this case seems to have been determined by passport. This means that someone who self-identifies as being a Kurdish national with a Syrian passport is considered a Syrian national.

The linkage exercise would help to aid understanding of the effect that different nationalities may have on linkage rates. This could be brought about if as a group, certain nationalities interact with government institutions and other sources of administrative data differently than others. A nationality variable will also be able to help distinguish between separate identities especially when someone has both a common Arab name and a January 1st Birthday.

Postcode
The postcode consists of both the inward and outward segments and refers to the address of the refugee at the time of resettlement. This information will be outdated if an individual has moved since the initial resettlement.

The linkage exercise would help to improve understanding of the following effects on data linkage:
- Whether the quality of the linkage varies by region
- The number of individuals that move on from their first accommodation. Only the local authority that initially receives the refugees holds data on them, and any moves away from the local authority impact the availability of data.

Arrival date
Date of arrival. Information on arrival dates can be used to rule out linkages if one identity was present in an administrative data source before the arrival date.

Case ID
Unique reference number. This would aid establishment of group/family membership to help the data linkage process via associative matching.

Lead Case ID
Unique reference number. This would aid establishment of group/family membership to help the data linkage process via associative matching.

UK Visa Expiry Date
Date of visa expiry. To identify embark. Can be used to rule out linkages in administrative data sources if identity was not present.

Lead applicant
Contains full name. Involves first, middle and last names, and any other names stated on official documents. The inclusion of full names in the pilot would especially be useful since not all datasets include national insurance numbers or person-specific identifiers other than names. This would aid establishment of group/family membership to help the data linkage process via associative matching.

Group size
Number in group. This would aid establishment of group/family membership to help the data linkage process via associative matching.

Adults in case
Number of adults in case. This would aid establishment of group/family membership to help the data linkage process via associative matching.

Children in case
Number of children in case. This would aid establishment of group/family membership to help the data linkage process via associative matching.

Total number of people in household
Number of people in household. This would aid establishment of group/family membership to help the data linkage process via associative matching.

Change in family composition since arrival

If the family composition has changed at all since arrival. Data are coded to engagement, joined by other family member, marital/family separation (note that this can also include older children/relatives moving out), new-born, other, pregnancy, prison, family departure, marriage.

This would aid establishment of group/family membership to help the data linkage process via associative matching.

Annex 2: Linkage variables held in Exit Checks, Patient Register and Personal Demographic Service

<u>**Exit Checks**</u>
Nationality
In Exit Checks, nationality indicates which passport was used for an event. It is possible for one individual to have events attributed to them involving passports of several nationalities.

Date of birth
Date of birth according to passport. Not all records contain a DOB, and a larger proportion involves a DOB of 01/01.

Full name
As presented on a passport, in the order 'surname middle names first name'.

Sex (F/M/U)
As presented on a passport. Sex can be either male, female or unknown.

Arrival date
Exit Checks contains the date on which each event takes place, including arrivals into the country. Information on arrival dates can be used to rule out linkages if one identity was active in the system before the arrival date of another identity.

UK Visa Expiry Date
Date of visa expiry. To identify embark. Can be used to rule out linkages in administrative data sources if identity was not present.

CID Case type
Case Information Database case type to help identify VPRS in Exit Checks as a cross-validation.

CRS Endorsement
Central Reference System visa type to help identify VPRS in Exit Checks as a cross-validation.

<u>**Patient Demographic Service Data**</u>
NHS Number
Anyone who has ever interacted with the NHS in England and Wales has been assigned an NHS number. In this linkage exercise, it is particularly suitable for linking the PDS to the PR, which also includes NHS numbers.

Full name
Contains first given name, surname and other given names. Name is usually self-reported, so the name on the passport does not necessarily have to match the name on the PDS, or shortened versions may have been used.

Gender (F/M/Missing/Invalid)

Gender can be either male, female, missing or invalid. This variable is based on self-reported information, and in some cases sex and gender will not match.

DOB

As with the full name on the PDS, DOB is self-reported so there may be some linkage issues there.

Postcode

The postcode on the PDS has been provided by the patient at the time of treatment or when they registered with a GP. Moves that did not involve reregistering for a GP may not be reflected in the system. The data dictionary does not specify whether this variable contains the full postcode.

Previous postcode

If someone registers with a new GP, they are usually asked about their previous details to ensure that their medical records can be transferred. Previous postcode information enables linkage to Home Office data on an individual's initial residence in case they have since moved. This linkage exercise would shed light on whether

Date of patient UK entry

If patients are not UK citizens, they may be asked about the date they entered the UK. This information is likely to be flawed in some way, especially since it relies on recollecting events from the past. There is still value in including this variable in the linkage exercise. While the exact date of entry may not always match the records in Exit Checks, linkage may be possible based on months. Linkage would also expose to what extent we can rely on self-reported data.

Patient Register

NHS Number

Anyone who has ever interacted with the NHS in England and Wales has been assigned an NHS number. In this linkage exercise, it is particularly suitable for linking the PDS to the PR, which also includes NHS numbers.

Gender (F/M/Missing/Invalid)

Gender can be either male, female, missing or invalid. This variable is based on self-reported information, and in some cases sex and gender will not match.

Full name

Contains first given name, surname and other given names. Name is usually self-reported, so the name on the passport does not necessarily have to match the name on the PDS, or shortened versions may have been used.

Postcode

The PR does not provide a full postcode. Instead the postcode is presented without the last two letters, e.g. PO15 5.

DOB

As with the full name on the PDS, DOB is self-reported so there may be some linkage issues there.

NSDEC(20)02

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

Please consult the guidance document before filling in the application form

---

Project Title
Please provide a title indicative of the project

Developing ONS' Household Financial Statistics through the linkage of administrative data to household financial survey data

---

Start Date: 02-20   **End Date:**   02-21

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

Office for National Statistics

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

Improving ONS household financial statistics using administrative data has the potential to deliver significant public good – these statistics, which include statistics about low incomes and income distributions, are vitally important to both policy makers and to public discourse about the economy and society. We think administrative data are an important tool to improving these statistics, which are currently based predominantly on household surveys.

In order to work out how we might be able to deliver the benefits associated with using administrative data we need to conduct feasibility research – involving comparing, at both the aggregate and linked address and person level, data from ONS' household financial surveys to DWP administrative data on income and earnings. This would allow us to evaluate the quantitative and qualitative differences between survey and administrative realisations of key income-related concepts and therefore work out how to improve our statistics using these data.

This feasibility research is critical to helping us decide on how we might improve our household financial statistics, including potentially by -

- In the short term, investigating improvements to our current statistics to improve our measurement of high and low incomes, our small-area income statistics, and how we account for survey non-response.
- In the longer term, putting administrative data at the core of our household financial statistics, limiting the need to ask additional questions via surveys and resulting in cost savings and respondent burden improvements.

However, actually making these kinds of improvements is outside the scope of this project, which is solely for feasibility research.

This research will also have wider benefits across the Government Statistical Service (GSS), and we will collaborate across the ONS and with DWP, and share research results to help build a cross-GSS understanding of administrative income data.

Section A
Project Details

A1   Legal gateways

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

CIS and BIDS data is received from DWP under section 45A of the Statistics and Registration Service Act 2007. This gateway permits ONS to use the data received for any of their statistical functions. The Memorandum of Understanding (MoU) between ONS and DWP permits use of the data for the purposes set out in this application.

No gateway required for survey data as collected by and for ONS.

A2   Ethical approval

Has the project been reviewed or is it expected to be reviewed by another ethics committee? *(please delete as appropriate)*   **No**

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

A3   Proposed site of research

Please provide information about where the research will take place, including detail of where any data linkage and/or data analysis will be conducted

Research, including both linkage and analysis, will take place at ONS within our secure Data Access Platform

A4   Data subjects to be studied

Does the study include all subsections of the population?
(i.e. all ages, sex, ethnic groups etc.) *(please delete as appropriate)*   **Yes**

*If **No**, please detail which subsections with justification(s) below:*

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

Justification for focusing on these subsections or groups:

A5   Research methodology and protocol

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

Research would take place in five main stages –

1. Comparison of survey estimates and administrative data aggregates, using unlinked data. This will involve investigating in detail how estimates differ for different income components – such as income from benefits and PAYE income.

2. Data linkage at the person level between survey response files and administrative datasets. Since there is no unique identifier available across both the survey and administrative datasets, we would use standard linkage methods utilising 'hashed' (or pseudo-anonymised) name, address and date of birth. Individuals carrying out linkage will only have access to these hashed linkage variables, and neither individuals carrying out linkage or analysis will have access to unhashed (non-anonymised) personally identifiable variables.

3. Comparison of survey responses and administrative data at the individual level. Researchers carrying out this stage of the research would utilise a fully anonymised version of the dataset and would not have any access personal details such as name, address or date of birth, either hashed or unhashed.

4. Data linkage at the address level between survey sample files and administrative datasets (aggregated to the address level). Again, individuals carrying out linkage will only have access to linkage variables.

5. Analysis on response patterns using linked address level survey sample files (that is – is income related to non-response?), and what is the scope for using administrative data to improve survey methodology?

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 dataset

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)*				Customer Information system (CIS) and Benefits and Income Dataset (BIDs) data from DWP for 2012-2018. Data are pseudo anonymised and will be fully anonymised prior to analysis.
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)*				Household Financial Survey data (including Living Cost and Food Survey, Survey of Living Conditions, and Wealth and Assets Survey) for 2012-2018. Datasets will be pseudo-anonymised prior to linkage and fully anonymised prior to analysis.

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)

Household financial statistics are increasingly at the centre of policy making and public discourse about our economy and society. An accurate picture of household income, poverty and inequality is hugely important for the public good.

There are important limitations to ONS' current survey-based household financial statistics, as highlighted in the Bean Review of Economic Statistics and the Resolution Foundation's report 'unequal results'. The relevant surveys suffer from limited sample sizes and falling response (the Living Cost and Food Survey currently has a response rate of below 50%), and there is considerable evidence of measurement error in some survey estimates. A specific concern is that there is a lot of evidence demonstrating that that surveys do not perfectly capture the incomes of both high and low income individuals.

These are not just academic considerations – they have a real and important impact. One example of this is that widely evidenced under-reporting of benefits in surveys may limit the ability of policy makers to understand low incomes and how these have changed over time. In addition, small sample sizes mean that we cannot produce statistics at sufficiently low levels of geography, which limits the ability of policy makers to tackle poverty and disadvantage, which is often clustered at the local level.

Administrative data, and the research outlined in this proposal, is vital to us addressing many of these concerns. Very broadly speaking, we think administrative data can provide us with accurate (but partial) income information for a very large proportion private households, compared to the several thousand households captured annually in household financial surveys. The specific DWP administrative data we propose to use can provide a partial income measure for most of the UK population, and full details on state benefits and some other income components.

Administrative data are not perfect – they do not perfectly capture the individuals and concepts we're interested in, and research is needed to inform how administrative data can be used alongside survey data to improve our household income statistics. In the short term, the research outlined in this project will inform us how we can use administrative data to

make iterative improvements, in particular to help us capture high and low incomes better. In addition, this research will allow us to continue to improve our admin-based small area income statistics (as previously reviewed by NSDEC – item 4, July 2017) to help meet the policy demand for small area household income statistics, and also to investigate improving quality by using better non-response adjustments to survey-based estimates. By sharing the results of this research across the GSS we can ensure that these benefits are realised across government, not just by improving ONS statistics. DWP in particular are planning similar research in relation to the Family Resources Survey and we are keen to be able to learn from each other as our research progresses.

In the longer term, this research will let us evaluate whether we can use admin data at the core of our household financial statistics, potentially leading to a significant reduction in our use of face-to-face surveys or in the length of those surveys. In addition to continued improvements to the resulting statistics, this may also result in considerable cost savings and reduced respondent burden, as our face-to-face surveys are expensive and take on average over an hour and a half per respondent to complete.

B2	**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)

Data Security - All administrative and survey data will be stored within ONS' Data Access Platform, a secure, accredited environment with robust systems controlling access to data and the exporting of research results. Only aggregate data which has gone through disclosure control will be exported from this secure platform to ensure that no identifiable data is stored outside of a controlled environment.

Confidentiality – Staff carrying out linkage will be working with personally identifiable data which have been 'hashed' and are therefore anonymised. Only staff carrying out linkage will have access to these hashed data. Staff will only have access to income information which they require to carry out their research, and access will be removed once complete.

Consent
The research set out in this application would include -

> a) Person-level linkage of anonymised administrative data to pseudo-anonymised survey responses (in order to understand the difference between survey responses and administrative data).
> b) Address-level linkage of anonymised administrative data to pseudo-anonymised survey sampled addresses (in order to understand biases in survey non-response). This would include linkage to all sampled addresses, regardless of whether a response is obtained.

Respondents are informed that their data will be treated confidentially and used to produce non-identifiable statistics, but ONS do not make explicit reference to linkage in respondent materials.

This aspect of the research set out in this application is an area where we would like ethical

advice.

It is important to note that, to have enough data to carry out the research set out in this application, we need to link to a number of years of survey data. Ideally this would include quite a lot of historical data in order to understand the impact of changing economic circumstances and tax and benefits changes, but as a minimum we need multiple years of data to have sufficient sample size. This means that, were ONS to change the information we provide to respondents, it would take a number of years for us to be able to carry out the research set out in this application using data collected solely collected under this new basis. This would result in a very significant delay to the realisation of the benefits set out in B1.

B3	**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)

This research will utilise established statistical techniques in the fields of data linkage and survey estimation. Quality assurance will be carried out via consultation with ONS experts in these fields and, where appropriate, external experts including DWP, academic researchers in the domain of household income, and experts in the domain of survey methodology.

B4	**Principle 4: Data used and methods employed are consistent with legal requirements such as Data Protection Legislation[1], the Human Rights Act 1998, the Statistics and Registration Service Act 2007 and the common law duty of confidence**

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

DWP Administrative data have been shared with the ONS under the Statistics and Registration Service Act.

Use of data to safely and securely produce anonymous statistics that inform the public good is consistent with Data Protection and Human Rights legislation

---

[1] "Data Protection Legislation" means the full, applicable data protection framework as set out in the Data Protection Act 2018. This encompasses general processing (including the General Data Protection Regulation and the applied GDPR).

B5	**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)

We have received extensive feedback from a wide range of users on the value of this research. This includes feedback from numerous local authorities and voluntary organisations on the value that improved small-area estimates of income, which this research will help facilitate, would bring to them in terms of tackling deprivation and poverty, which is often clustered in small areas. Other public and private bodies – including other government departments, academia, and charities – have also emphasised, over a long period, the importance of improving our household financial statistics.

Expert users of our household financial surveys, such as the researchers represented at ONS' Income Statistics User round table, have told us that they are particularly concerned that household financial surveys are not adequately capturing the incomes of the poorest in society, which limits their usefulness in developing and targeting anti-poverty policies. There is a recognition among users that the use of linked administrative data has the potential to address these concerns.

We have not engaged with the public specifically on this research. ONS have commissioned broader research in collaboration with ESRC and Ipsos MORI to explore public understanding and views of administrative data and linkage. This research demonstrated people are particularly concerned about de-identification and disclosure risk, which we believe we will address effectively in this research proposal by working with anonymised data, restricting access, and utilising a highly secure environment. The research also demonstrated that people are more supportive of research using linked administrative data where it's possible to demonstrate a clear public good. We believe that is the case here – statistics related to poverty, disadvantage and inequality are widely considered to be important and highly relevant, and enabling effective policy to tackle poverty and disadvantage is widely considered to be in the public good.

B6	**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 will publish the results of this research on the ONS website. This paper will contain a full description of the research methodology, findings, and recommendations. We are proactively engaging with known users of household financial statistics to discuss this research and will continue to do so throughout the research and to discuss our findings.

B7	**Future use of products**

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

Anonymised linked datasets would be treated as a new information asset within ONS' secure Data Access Platform and added to ONS' information asset register. As a part of ONS' information management this asset would be reviewed periodically, and deleted when there is no longer a need for it.

These anonymised linked datasets may be utilised for further uses outside the scope the research in this proposal, subject to those uses having been through ethical review and data supplier approval. As a separate information asset, access to anonymised linked data would be controlled separately and conditional on this ethical/supplier approval.

As outlined above, no linked data containing personal identifiers will be created or retained.

B8	**Collaboration and Sponsors**

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

List of Collaborators/Sponsors	Details and relevant documentation relating to collaboration (you may attach copies of relevant documentation)
Department for Work and Pensions	MoU in place for data sharing As DWP is also exploring integrating admin data into DWP's Family Resources Survey, we expect to collaborate closely with DWP colleagues.
Office for National Statistics	

Section C
Responsible owner and applicant details

C1	**Responsible Owner**

Full Name: Matthew Greenaway	**Position:** Principal Statistician
Address:	**Email:**
	Telephone:
	Organisation: Office for National Statistics

Declaration to be signed by the responsible owner

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

I understand that it is a requirement for all researchers accessing the data to have undergone relevant training and to have either relevant security clearances or accredited/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: **Matthew Greenaway**   **Date:** 24/1/20

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

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

UK Statistics Authority
National Statistician's Data Ethics Advisory Committee

Investigating the risk factors and inequalities associated with infant mortality

Phase two of this project is undergoing major revisions and will be published in due course

NSDEC(20)04

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

Please consult the guidance document before filling in the application form

---

Project Title
Please provide a title indicative of the project

Social Factors Impacting Offending and Resettlement in Northern Ireland- Mortality within the Offending Population

---

Start Date: 10th February 2020   **End Date:**   30th June 2021

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

N/A

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

The Department of Justice (DoJ) in Northern Ireland is advocating the use of administrative data to progress a series of inter-related research projects to better understand issues impacting upon propensity to re-offend while also looking at risks individuals released from custody face that may lead to harm and even mortality.
Offenders face a challenging environment upon release from custody including substantial health risks, an inability to secure permanent accommodation, finding gainful employment and resettling in the community.
The purpose of this study is to improve the evidence base on the link between the offender population and mortality in Northern Ireland and to specifically examine if there are high risk periods where risk of mortality is higher, for example, the first few weeks of imprisonment or the immediate weeks post release from custody. The research will generate new insights into whether certain groups could be targeted during specific periods of time as part of strategies to reduce mortality levels. High mortality rates following release from prison have been well documented in a number of countries including GB. This study will also help to inform the work of the Post Release Deaths Project, a collaboration between the Northern Ireland Prison Service, Probation Board

Northern Ireland, the Northern Ireland Court Service and the South Eastern Health and Social Care Trust. The Prisoner Ombudsman in Northern Ireland recently called for more to be done to ensure those released from custody have access to services such as GPs, addiction support and housing, and this initiative aims to gain a better understanding of the volume and cause of death of former prisoners in order to better inform custodial interventions and service provision post release.

This study is the first stage of a wide ranging DoJ research programme examining an individual's propensity to offend and subsequently reoffend. Results of this study will be generalisable, will yield important timely outcomes and will inform subsequent research within the DoJ Programme. Further information on the three stages of this research programme is outlined below:

Stage 1 (*This project*) – As well as addressing the key aims outlined above, Stage 1 will be invaluable in assessing the quality and readiness of the DoJ data, with regard to matching to other administrative datasets.

Stage 2 – The second planned research study will expand on Stage 1 and examine the uptake of benefits by offenders post-release, as well as assessing the differences between those who have and have not reoffended. Access to benefits as quickly as possible post-release ensures individuals have financial means to assist their resettlement. Benefit uptake over the medium/longer term also provides a useful indicator of the ongoing resettlement journey for those released from custody.

Stage 3 –The third aspect of the research programme will specifically examine housing accommodation for offenders and will examine the barriers and difficulties encountered by these individuals that are detrimental to successful housing resettlement.

Programme findings will assist primarily the DoJ and also the Department for Communities (DfC) to continuously develop strategies to address offending and re-offending behaviour.

Section A
Project Details

A1   Legal gateways

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

- Digital Economy Act 2017; and
- General Data Protection Regulation (GDPR) 2018.

The Digital Economy Act (DEA) 2017 facilitates the linking and sharing of datasets held by public authorities for research purposes. Under the DEA, data can be processed and made available in a safe and secure way for research purposes.

The GDPR forms part of the data protection regime in the UK, together with the new Data Protection Act 2018 (DPA 2018). Sharing of personal data is legally allowable under Article 6 of the GDPR where there is an exemption for processing of personal data for statistics and research purposes. In this regard, the lawful basis for completing this piece of analysis is 'public interest'. As this project contains the use of special category data, the condition of processing has also been identified as 'research'.

We have consulted with the Data Protection Officer for the Department of Justice NI who was content with the expertise and security around the Northern Ireland Statistics and Research Agency's (NISRA) Administrative Data Research – Northern Ireland (ADR-NI) policies and processes for facilitating this research. A Data Protection Impact Assessment (DPIA) has been carried out.

A2   Ethical approval

Has the project been reviewed or is it expected to be reviewed by another ethics committee? *(please delete as appropriate)*   ~~Yes~~/No

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

A3   Proposed site of research

Please provide information about where the research will take place, including detail of where any data linkage and/or data analysis will be conducted

This research will be undertaken by the Northern Ireland Statistics and Research Agency (NISRA) via Administrative Data Research – Northern Ireland (ADR-NI). Data linkages and analyses will be undertaken in the secure setting of NISRA's Research Support Unit (RSU). All of the necessary data linkage work will be completed by NISRA's Trusted Third Party (TTP) in full accordance with agreed ADR-NI processes and security protocols. As with all projects, the finalised variable list and dataset will be agreed in conjunction with Data Custodians and researchers and will be documented in a project specific Data Sharing Agreement. In keeping with best practice, the finalised data set will be subject to a Disclosure Control Assessment prior to being made available to researchers in the secure environment. None of the demographic information used by the TTP for the necessary linkage work will be made available to the research team.

A4   Data subjects to be studied

Does the study include all subsections of the population?
(i.e. all ages, sex, ethnic groups etc.) *(please delete as appropriate)*   ~~Yes~~/No

*If **No**, please detail which subsections with justification(s) below:*

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

The project will focus on all offenders sentenced to a community-based sentence, a diversion, in custody or released from custody between April 2011 and March 2018.

Justification for focusing on these subsections or groups:

The purpose of this research is to improve the evidence base highlighting issues faced by offenders and the longer-term outcomes for those that have spent time in custody to enable statutory agencies to develop and resource the most appropriate services to assist resettlement of offenders pre-release and post-custody.

A5	**Research methodology and protocol**

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

The research will involve a retrospective cohort study of all people who received a non-custodial disposal at court, a diversionary disposal or who were in custody or released from custody in Northern Ireland between April 2011 and March 2018. The custodial population is estimated at 37,000 individuals over this time period. The population receiving a non-custodial or diversionary disposal over this period is estimated at 85,000 individuals.

Records of offenders (Department of Justice) will be linked to death registrations data (General Registrar's Office - NISRA).

Three main subgroups of offender will be identified across all analyses to distinguish between the main types of sentence. They are:

(i) *Sentenced prisoners* – i.e. offenders sentenced by the courts to immediate custody for criminal offences

(ii) *Remand prisoners* – includes those charged with an offence and whom the courts have ruled should be detained in custody pending trial; those whom the courts have permitted to be released on bail pending trial but have not as yet met the conditions (usually financial) of the bail; those who had been released on bail but have subsequently been re-admitted to prison because they breached a condition of bail; and those who have been found guilty by the court but have been ordered to be detained in custody pending sentence

(iii) *Non-custodial* – offenders receiving a non custodial sentence (including those convicted of an offence that received a suspended custodial sentence, a monetary penalty, a conditional discharge, community supervision or a diversionary disposal).

(I) Descriptive analyses

Socio-demographic (e.g. age, gender, geographical area, area deprivation) and offence specific (e.g. offence type, sentence type, offence history, sentence and custody length, cause of death if applicable) characteristics will be generated for
(I) Deaths occurring in prison;
(II) Deaths occurring post release from custody

The main cause of death will be assessed at a range of different time points (e.g. first few weeks in custody, first few weeks after release from custody)

Mortality rates among custody releases will be compared to those receiving a non-custodial sentence and the general population with the use of indirect standardization and adjustment for age and sex.

The utility of a sequence analysis will also be explored to assess the relationship between nature and order of contacts with the justice system and mortality.

Regression Models
Multivariate Cox proportional Hazards (CPH) models will be developed to model the following outcomes
(I) Mortality in prison;
(II) Mortality on release from prison
Univariate CPH models will be run to generate unadjusted mortality or re-incarceration hazards for each candidate predictor followed by a full multivariate model for each outcome. Key explanatory or predictors variables to be considered for inclusion include age, gender, marital status, geographical area, area deprivation, primary offence and offence history.

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 dataset

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)*			DOJ – Discharges and Reoffending Dataset (2011/12 to 2016/17) DOJ - Custody Dataset (2011/12 to 2017/18) NISRA – GRO Death Registrations Dataset (2011/12 to 2017/18)	

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

Section B
Assessment against NSDEC ethical principles

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

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

It is anticipated that this project will improve the empirical evidence base highlighting issues faced by offenders and the longer-term outcomes for those that have spent time in custody.

This evidence will assist statutory agencies in the development of custodial interventions and resettlement services for offenders, reducing the risk of reoffending while improving the welfare of serving or having served custodial sentences. The benefits of this being:
- A reduction in cost of crime to the economy;
- A reduction in the number of victims and the adverse impact crime has upon the quality of life of victims;
- The ability to better address factors contributing to offending behaviour, such as addictions and homelessness;
- More focussed and efficient Health and Social Care interventions; and
- The promotion of desistence, improving the quality of life of those previously engaged in offending behaviour

This study will also help to inform the work of the Post Release Deaths Project, a collaboration between the Northern Ireland Prison Service, Probation Board Northern Ireland, the Northern Ireland Court Service and the South Eastern Health and Social Care Trust. The Prisoner Ombudsman in Northern Ireland recently called for more to be done to ensure those released from custody have access to services such as GPs, addiction support

and housing and this initiative aims to gain a better understanding of the volume and cause of death of former prisoners in order to better inform custodial interventions and service provision post release.

B2	**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)

Data linkages and analyses will be undertaken in the secure setting of NISRA's Research Support Unit (RSU), part of ADR-NI, which is accredited by the UK Statistics Authority as an accredited processor under the research strand of the Digital Economy Act 2017.

Data which identifies individuals directly will travel by an accredited, secure route from the data provider to Trusted Third Party before being matched at the RSU facility where the researchers will work. Using a third party to link identifiers ensures that no organisation ever sees the entire linked dataset. Specific organisations are commissioned by the network and are therefore 'trusted' to undertake this function. Research teams will work in the secure facilities and will not transfer data outside of them. No-one is allowed to share data outside the proposed research team. Analytical outputs can only be shared if they have undergone statistical disclosure control and been cleared by the Network's expert staff.

With regards to consent, Personal Information Notices (PINs) include references to the use of administrative information for research and statistical purposes.

B3	**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 project does not suggest the use of new or novel technologies and/or methods and, as such, no risk is posed from this. There will be human oversight throughout each stage of the process, from data owners on providing data extracts, staff in the Trusted Third Party conducting linkage on demographic data, researchers accessing and analysing the data, and staff within the secure environment conducting statistical disclosure and quality control measures prior to release of outputs. This should ensure data quality is maintained throughout all stages of the process and that any issues with the quality of the data will be easily identified and addressed.

B4	**Principle 4: Data used and methods employed are consistent with legal requirements such as Data Protection Legislation[1], the Human Rights Act 1998, the Statistics and Registration Service Act 2007 and the common law duty of confidence**

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

Under General Data Protection Regulation (GDPR) and following on from ongoing public consultations, the lawful basis for this piece of analysis is 'public interest'. As this project contains the use of special category data, the condition of processing has also been identified as 'research'. It is deemed that undertaking this research would be fair and would best serve the interests of society as a whole.

Safeguarding measures, such as data minimisation and anonymization, have been put in place to ensure full compliance with Data Protection legislation. Within the current study, only the variables deemed necessary to conduct analysis will be accessed by the researchers. Prior to this, the data will be anonymised in line with the ICO Anonymization Code of Practice, with identifiers held by a Trusted Third Party with an agreement specifying no personal data will be accessed and no re-identification will take place.

B5	**Principle 5: The views of the public are considered in light of the data used and the perceived benefits of the research**

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

This project is the result of findings from consultations held in relation to the Northern Ireland Programme for Government (PfG). An eight week public consultation identified the need for the Government to do more to reduce offending and reoffending and make people feel safe in their communities. As a result of this, an empirical evidence base is now necessary in order for the Department of Justice to identify how we can tackle this issue raised within the consultation. No specific public engagements will take place following this study, however findings will be made publicly available and results will directly contribute towards the work of the Post Release Deaths Project in putting the appropriate interventions in place, with the aim of fulfilling public demands for a reduction in crime and offending across the country.

B6	**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)

Findings from the study will be used to form a report for policy colleagues within the Department of Justice's Reducing Offending Directorate. This report will also be made publically available via publication on the Statistics and Research section of the Department of Justice website. The release of the report will be announced on Gov.uk prior to publication, whilst announcements will also be made following publication on the corresponding intranet sites and Twitter and Facebook feeds for DOJ and NISRA. The paper will contain detailed methodology and links to the ADR-NI website to ensure the approach

---

[1] "Data Protection Legislation" means the full, applicable data protection framework as set out in the Data Protection Act 2018. This encompasses general processing (including the General Data Protection Regulation and the applied GDPR).

used is clearly visible. Findings will also contribute towards the Programme for Government (PfG), as part of the Executive's efforts to reduce crime and reoffending in Northern Ireland. It is anticipated that strategy documentation produced in relation to indicators for Outcome 7 will include reference to this study as a means of justification for any actions that have been taken forward.

B7   Future use of products

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

Following completion of each stage of research, the linked dataset(s) will be archived and disposed of by ADR NI in line with their retention policy of 5 years. It should also be noted that, whilst the data will be retained for a certain time period, this data cannot be linked to any other datasets due to the anonymised linkage process used by ADR NI and the Trusted Third Party. During the retention period, there is the possibility to re-access the linked dataset should any further questions arise based on the outcomes of the research. If it is necessary to re-access the data to answer any further questions in relation to the study, permission would be sought from the data owners and approval would need to be granted by NISRA's RSU. There will be no obligation to allow the reuse of this data – rather an encouragement, if necessary, to enable a more efficient and effective way of working.

B8   Collaboration and Sponsors

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

List of Collaborators/Sponsors	Details and relevant documentation relating to collaboration (you may attach copies of relevant documentation)
Department of Justice (DOJ)	Agreement in principle to provide criminal justice information for analysis
Northern Ireland Statistics & Research Agency (NISRA)	Agreement in principle to provide death registrations information for analysis

Section C
Responsible owner and applicant details

C1	**Responsible Owner**

Full Name: Shannon Browne   **Position:** Deputy Principal Statistician

Address: **Email:**

Telephone:

Organisation: Analytical Services Group, Financial Services Division, Department of Justice

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 accredited/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: **Shannon Browne**   **Date:**   28/02/2020

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

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

UK STATISTICS AUTHORITY

NATIONAL STATISTICIAN'S DATA ETHICS ADVISORY COMMITTEE

Peter Jones

Presentation

Uses of Admin Data to support future population and social statistics

UK STATISTICS AUTHORITY

NATIONAL STATISTICIAN'S DATA ETHICS COMMITTEE

Data Protection and Ethics Compliance

Purpose

1. This paper covers a Data Ethics Compliance report at **Annex A** on the ONS research project *"Extending the Crime Survey for England and Wales (CSEW) to include a new module of questions on the online behaviour of children aged 10-15 years"* and consider the role of the Data Protection Compliance Team in providing ongoing assurance to the National Statistician's Data Ethics Committee (the Committee) that research projects are continuing to adhere to ethical principles.

Recommendations

2. Members of the NSDEC are invited to:

   i.   Review the Data Ethics Compliance report at **Annex A**;
   ii. Consider the current approach to providing assurance to the Committee on implementation of actions, and;
   iii. Propose any projects which the Committee would like assurance on the level of compliance with the ethical principles in addition to, or in replace of, projects outlined in the high-level Data Ethics Compliance Review Plan for 2020/21 at **Annex B**.

Background

3. In order to provide the ONS with assurance that it is complying with data protection legislation, a function has been set up to support the Data Protection Officer to provide data protection auditing and compliance monitoring services. An audit/compliance reporting programme has been developed to cover all activities across the Office from a data protection perspective.

4. Good ethical practice and data protection legislation promote the appropriate use of data with appropriate safeguards. The organisation and the projects (involving the use of personal data) it sponsors, are required to adhere to data protection legislation where appropriate. However, good ethical practice ensures that the access, use and sharing of public data, for research and statistical purposes by the ONS, is ethical and for the public good.

5. Currently, the Data Protection Compliance team will follow up on actions given to projects which have been presented to the Committee for ethical review. This will provide an additional layer of assurance that research projects are in line with the Committee's ethical principles and offer the opportunity to give a view on the impact of the Committee.

6. The Data Protection Compliance Manager, along with the Data Protection Officer and the Committee's Secretariat, have developed a high-level data ethics compliance plan where we have identified several projects, based on risk, for review in 2020/21 (**Annex B**).

Discussion

7. The data ethics compliance review has confirmed that the ONS Centre for Crime and Justice has taken the advice of the Committee when extending the survey to include a new module of survey questions on the online behaviour of children aged 10-15 years.

We confirmed that all actions agreed at the NSDEC meeting in February have been satisfactorily implemented.

8. Through our Data Ethics Compliance reviews, the Committee will receive assurance that the actions assigned to a select number of research projects during ethical review have been implemented. However, we would like the Committee to consider this approach and whether it would like more in-depth ethics reviews on an ad-hoc basis, to ascertain whether all ethical principles are being adhered to in the development and application of project proposals. Alternatively, we can continue our approach of following up actions from Committee meetings on a regular basis.

9. We would also like the Committee to consider any projects which it feels it would like assurance on as a matter of urgency for reporting at the next meeting. We would also like the Committee to suggest any projects to add/replace on the Data Ethics Compliance review plan for 2020/21 **at Annex B**.

Conclusion

10. The Data Protection Team's ethical compliance reviews will provide NSDEC and the National Statistician with assurance that the project proposals which have been reviewed by the Committee have implemented the Committee's recommendations. However, the team can also offer the opportunity to provide more in-depth reviews of these projects and provide assurance that they are continuing to adhere to the all ethical principles and not just that the Committee's recommendations have been implemented. We would also welcome the Committee to suggest any specific reviews into ethical compliance of ONS projects.

Rhys Nadin, Data Protection Compliance Manager, February 2020

List of Annexes

Annex A   Data Ethics Compliance Review of Extending the Crime Survey for England and Wales (CSEW) to include a new module of questions on the online behaviour of children aged 10-15 years
Annex B   Data Ethics Compliance Reviews 2020/21

UK Statistics Authority

Data Ethics Compliance Review Report

Review of

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

January 2019

Table of Contents

Executive Summary

Introduction

A data ethics compliance review of the project 'Extending the Crime Survey for England and Wales (CSEW) to include a new module of questions on the online behaviour of children aged 10-15 years' was agreed as part of the 2019/20 Data Protection Compliance Programme.

The National Statistician's Data Ethics Advisory Committee (NSDEC) provides ethical advice that the access, use and sharing of public data for research and statistical purposes is ethical and for the public good. In reviewing project and policy proposals, NSDEC provides advice to the National Statistician on the basis of the following six ethical principles:

1. The use of data has clear benefits for users and serves the public good.
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.
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.
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.
5. The views of the public are considered in light of the data used and the perceived benefits of the research.
6. The access use and sharing of data is transparent and is communicated clearly and accessibly to the public.

In July 2018, the Head of the ONS Centre for Crime and Justice presented an application to extend the children's Crime Survey for England and Wales (CSEW), a ONS survey conducted by a third party (Kantar Public) on behalf of the ONS, to include a new module of questions on the online behaviour of children aged 10-15 years. The Committee agreed that the project required major revisions, and gave the ONS Centre for Crime and Justice the following four actions:

i. include in the application the support mechanisms available to children, parents and guardians;
ii. ensure that any safeguarding concerns are adequately addressed;
iii. consider amending the questions to ensure that they are understood by responders of all ages participating in the survey and that only necessary data is collected; and
iv. resubmit a proposal to be considered via correspondence.

A number of issues were resolved via correspondence between July 2018 and February 2019. At February's NSDEC meeting, the Head of the ONS Centre for Crime and Justice submitted a revised proposal along with an update of progress made against the actions given at July 2018's meeting. At the meeting, the project was approved subject to the ONS Centre for Crime and Justice doing the following:

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

Scope

The scope of this review was to provide assurance to NSDEC and the National Statistician that the ONS Centre for Crime and Justice has implemented the actions agreed at NSDEC February 2019's meeting in relation to extending the Crime Survey for England and Wales (CSEW) to include a new module of questions on the online behaviour of children aged 10-15 years.

Limitations

The review has not sought to provide assurance that the children's CSEW is continuing to apply all the controls to comply with the ethical principles in the application submitted to the committee in February 2019.

Conclusion

Through review of meeting minutes, relevant documentation and correspondence we confirmed that the Crime Survey for England and Wales has taken the advice of the committee when extending the survey to include a new module of survey questions on the online behaviour of children aged 10-15 years. We confirmed that all actions agreed at the NSDEC meeting in February have been satisfactorily implemented.

We confirmed that the ONS Centre for Crime and Justice clarified Kantar's safeguarding policy, and along with ONS' own policy was updated to ensure that they were aligned. We noted that this issue was also raised during an internal audit of the ONS' safeguarding policy and procedures in July 2018, and the action was reported as complete to the Audit Committee in March 2019. Through our testing we also established that communications to respondents and their parent/guardian(s) have been amended to include detail around the right to request access to the answers the child gave to ONS and the risk rating given for the child's online activity will be provided.

Detailed Findings

Ref	ACTION	FINDINGS	IMPLEMENTED
i	Clarifying Kantar's data protection and privacy policies;	Through review of email communications, we confirmed that the ONS Centre for Crime and Justice contacted Kantar to ensure that their safeguarding policies were aligned. Kantar and the ONS Centre for Crime and Justice reviewed each other's policies and it was felt that they were broadly similar. There was more focus in the Kantar policy where an issue is suspected, but the individual has not made a disclosure to the interviewer whereas the ONS policy gives more detail about the different signs of harm, which is not given in the Kantar policy. As the policies were similar, Kantar were happy to amend their policy to include more detail around the different signs of harm, as noted in the ONS policy, and ONS also amended their policy to include situations where an issue may be suspected, but not disclosed as outlined in Kantar's policy. We also noted that an action from an internal audit of ONS' safeguarding policies and procedures in July 2018, the ONS were required to review their 3rd party processes and policies around safeguarding to ensure that they were aligned. This action was reported to the Audit Committee in March 2019 as complete.	Yes
ii	Clearly communicating the following to the child: - that a letter will be sent (to the child and parent/guardian) with a risk score for their online activity, - that if they are under 16 the parent/guardian(s) have a right to request access to the child's answers	**Communications** We were provided with a leaflet which is given to all the respondents of the children's CSEW (online behaviour of children aged 10-15 years'), and through review identified that the leaflet: • details that after the interview the CSEW will send a letter to the respondent and their parent/guardian(s) about whether their online activity is low, medium or high risk, and; • states that the parents/guardian(s) have a right to request access to the information the respondent has provided, and that if a legal request is made, then the ONS might have to share the answers with them. Additionally, the leaflet also provides contact details (email and telephone number) for further questions, how the information gathered is used and online links to further guidance.	Yes

	- that if the parents make a legal request, ONS may have to share the answers the child gave with them; and - providing NSDEC with progress reports on the roll out of the module at future meetings.	We also noted that on the Crime Survey website, there is further guidance in relation to the survey module on 10-15 year-olds. Through review of this guidance we confirmed that it includes information on: • **Online Behaviour Risk Rating – T**his describes how the risk rating is calculated and that the rating will be shared with the parent/guardian(s) after the interview. There is also a link to further guidance on risk ratings on what each score means and contact details for further questions. • **Permission –** This states that permission will always be obtained from a parent or guardian before any interview is conducted with anyone aged 10-15. • **Questionnaire –** The Parent/Guardian(s) can also download the full questionnaire for the module and see the questions used in the online behaviour risk rating. **NSDEC Updates** Through review of minutes, agendas and papers from July 2019's NSDEC meeting, we confirmed that an update via a presentation from the Head of the ONS Centre for Crime and Justice on extending the CSEW. The presentation outlined that: • The module went live in April 2019; • the percentage of risk ratings from the first batch of surveys (90% Low, 6.2% medium and 3.6% high), and; • there have been no complaints or queries about the module. Through examination of the minutes, we confirmed that the Committee would review this update via correspondence. The Secretariat for NSDEC stated that no concerns or questions were raised by the committee on the update given.	

Annex B   High-level Data Ethics Compliance Review Plan - 2020/21

Proposed Review Area	Project Approved by NSDEC	2020/21
Students' suicide statistics	January 2018	Q1
TBC		Q2
TBC		Q3
TBC		Q4

UK STATISTICS AUTHORITY

NATIONAL STATISTICIAN'S DATA ETHICS ADVISORY COMMITTEE

Any other business