13th Meeting of the
National Statistician's Data Ethics Advisory Committee

Minute, Agenda and Papers

Tuesday 24 July 2018
10:30 – 14:15
Board Room, UK Statistics Authority
London

UK STATISTICS AUTHORITY

NATIONAL STATISTICIAN'S DATA ETHICS ADVISORY COMMITTEE

Agenda

Tuesday 24 July 2018
Board Room, Drummond Gate London
10:30am – 2:10pm

Chair: Moira Gibb

Apologies: Vanessa Cuthill, Colin Godbold, Annie Hitchman, Emma Uprichard

(10:30am to 12:40pm)

1 10:30 am	Minute and matters arising from the previous meeting	NSDEC Chair
2 10:40 am	Testing the use of dependent interviewing with administrative data	NSDEC(18)16 Peter Jones
3 11:05 am	Crime Survey for England and Wales – 10-15 year old questionnaire – Cyber Crime Module	NSDEC(18)17 Mark Bangs
4 11:30 am	Sexual orientation, gender identity and armed forces question implementation: 2021 Census, England and Wales	Oral report Garnet Compton

Lunch (12:30pm to 1:00pm)
(1:00pm to 2:10pm)

5 1:00 pm	Collecting and using social media data for the production of statistics and research	NSDEC(18)18 Leone Wardman
6 1:20 pm	Social media analysis of public perceptions to inform question development	NSDEC(18)19 Karina Williams
7 1:40 pm	Investigating Mental Health Analysis Programme – Phase 1	NSDEC(18)20 Neil Bannister
8 2:00 pm	Any other business	

Next meeting: Tuesday 30 October 2018
Location: One Drummond Gate, London

National Statistician’s Data Ethics Advisory Committee

Minute

Tuesday, 24 July 2018
Board Room, Drummond Gate, London

Present
Members
Dame Moira Gibb (Chair)
Mr Stephen Balchin
Mr Robert Bumpstead
Ms Vanessa Cuthill
Mr Keith Dugmore
Ms Isabel Nisbet

UK Statistics Authority
Dr Simon Whitworth
Mr Petros Saravakos

Office for National Statistics
Mr Peter Jones and Ms Natalie Shorten (for item 2)
Mr Mark Bangs and Mr Joe Traynor (for item 3)
Mr Garnett Compton, Ms Amie Kamanda, Ms Michelle Monkman, Ms Maria King, Ms Rose Elliot (for item 4)
Ms Leone Wardman (for item 5)
Ms Karina Williams and Ms Natalie Gillson (for item 6)
Mr Neil Bannister and Ms Dannielle Cornish (for item 7)

Apologies
Mr Colin Godbold
Ms Annie Hitchman
Ms Marion Oswald
Ms Emma Uprichard

1. Minutes and matters arising from the previous meeting

1.1 Members were informed that the minute of the twelfth meeting had been agreed by correspondence.

1.2 The Chair invited the NSDEC Secretariat to provide an update on projects previously considered by NSDEC. The meeting heard that the project proposal from the Scottish Government Justice Analytical services, which looked to assess the risk of reoffending posed by offenders to inform policy planning for support services, was still undergoing revisions consistent with NSDEC’s recommendations. All other projects had been revised and signed off by the Secretariat.

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

2. Testing the use of dependent interviewing with administrative data [NSDEC(18)16]

2.1 Mr Peter Jones and Ms Natalie Shorten, from the ONS Population and Public Policy Integrated Data Division (IDD), presented a proposal to conduct a Population Coverage Survey to estimate and remove over-coverage in the Statistical Population Dataset which IDD have developed through linking numerous different administrative data sources. This proposal was referred to the Committee by the ONS Data Governance Committee.

2.2 Interviewers would contact a small sample of households in areas with high-churn, where over-coverage is more common. Responses from the survey would be used to develop a methodology to address the issue of over-coverage within the SPDs.

2.3 Members were informed that interviewers would be potentially disclosing names and addresses of previous residents to respondents. It was felt that this was not consistent with ONS's established policies and public commitments to protecting personal data. The UK Statistics Authority Board would need to approve a change to this policy if this work was to progress.

2.4 The meeting identified that harm could potentially be caused to some people through the releasing of personal data in the proposed way. The application did not offer sufficient mitigations to reduce the risk of any harm being caused to people as a result of the release of personal data through the proposed work.

2.5 Part of the proposed work involved using data from the publicly available Electoral Roll to test the method in the field. Given the under coverage on the publicly available Electoral Roll, it was felt that the statistical benefits of doing this work would be extremely limited and that instead other methods should be explored.

2.6 This project was rejected.

3 Crime Survey for England and Wales – 10-15 year old questionnaire – Cyber Crime Module[NSDEC(18)17]

3.1 Mr. Mark Bangs presented a proposal from the ONS Crime Statistics team to introduce a new self-completion module to the Crime Survey for England and Wales to capture online crime experiences of children aged 10-15 years old. A small sample of randomly selected households would be included in the survey and only one child would be selected at random in each household. Parents and guardians as well as children would be asked to provide consent before participating in the survey.

3.2 The meeting heard that advice on question development and the administration of the survey had been received from the National Society for the Prevention of Cruelty to Children (NSPCC). The proposal had also been considered by the NSPCC's research ethics committee. Members heard that advance materials would be provided to children and parents before the interview and that contact details to support children affected by online crime experience would be provided during the survey. It was also reported that the interviewers, from the research agency Kantar Public, would be specially trained and had experience of conducting similar interviews and would operate in accordance with the relevant safe guarding policy.

3.3 Members suggested that further work was required to make sure that the questions are clearly understood by all children taking part in the survey given the different levels of understanding a child at the age of ten is likely to have compared with a child at the age of fifteen.

3.4 It was agreed that some of the questions in the survey could mean that ONS may be in a position where they received information that could indicate potential harm. In instances such as this it needed to be clear in the application whether ONS's responsibility to the child's welfare would override their commitment to upholding the confidentiality of the responses. It was suggested that some of the questions could be amended to minimise the chances of such sensitive situations arising.

3.5 The committee heard that the ONS Crime Statistics team would consult with ONS Legal Services to clarify the Welsh language requirements.

3.6 The committee heard that interviewers from Kantar Public would comply with the Kantar Public Child Protection policy and that this was aligned with the ONS Safeguarding Policy.

3.7 The Committee agreed that the project required major revisions and will be considered further via correspondence

Action: Mr Mark Bangs to:

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.

4. Question implementation 2021 Census, England and Wales

4.1 Mr Garnett Compton, Ms Maria King, Ms Amie Kamanda, Ms Michelle Monkman, Ms Rose Elliot and Ms Jenny Knight from the ONS Census team, provided an oral report on progress with question implementation in the 2021 Census of England and Wales.

4.2 NSDEC welcomed this thoughtful and comprehensive presentation and were pleased that this work was consistent with NSDEC's principles.

5. Collecting and using social media data for the production of statistics and research[NSDEC(18)18]

5.1 Ms Leone Wardman, from the ONS Big Data team, provided an updated version of the ONS policy to collect and use social media data for the production of statistics and research. This policy aims to ensure that any statistical research across the ONS and the Government Statistical Service involving social media data is consistent, ethical and lawful.

5.2 A first draft of this policy was presented at the January 2018 NSDEC meeting. Members agreed that the proposed policy was a significant improvement and that all NSDEC's points raised in the previous meeting had been addressed. However, the meeting recommended that further assurances should be provided by ONS Legal Services to ensure the proportionality of collecting and using social media data for research and statistical purposes.

5.3 Members agreed that there should be an annual review of the policy, alongside case studies where it had been implemented.

5.4 The meeting heard that all uses of social media data would need to go through the ethics self-assessment process. Any projects presenting high ethical risk would be considered by NSDEC.

Action: Ms Leone Wardman and the NSDEC secretariat to

i. collaborate with ONS Legal services to conduct a proportionality test; and
ii. ensure that the policy is regularly reviewed and any use of social media data receives appropriate ethical consideration.

6. Social media analysis of public perceptions to inform question development[NSDEC(18)19]

6.1 Ms Karina Williams and Ms Natalie Gillson, from Data Collection Methodology, presented a project proposal to use social media data to conduct cognitive and qualitative research in order to develop a methodology to capture public views as expressed on social media networks. The proposal will initially focus on public views on gender identity.

6.2 The committee was informed that the proposed method would be supplementary to traditional methods of capturing the views of the public such as focus groups and stakeholder engagement events. The use of social media analysis imposes less of a burden on individuals than some forms of traditional public engagement and includes population groups which might be reluctant to participate in public events.

6.3 Members heard that ONS Data Collection Methodology would develop a robust methodology to address biases in the social media data to ensure that data is used following best practice and analysis can lead to valid conclusions. These methods would be published.

6.4 This project was approved.

7. Investigating Mental Health Analysis Programme – Phase 1

7.1 Mr Neil Bannister and Ms Danielle Cornish, from ONS Life Events and Analysis team introduced a project proposal to link census data, Improving Access to Psychological Therapies (IAPT) data and death registrations to assess the mortality risk of people with CMDs, comorbidities between mental and physical health problems, and investigate inequalities around mental health.

7.2 Members agreed that this project was potentially an excellent use of ONS's data and linkage expertise that could provide valuable new insights on mental health and therefore presented a potential significant public benefit.

7.3 This project was approved.

8. Any other business

8.1 There was no other business.

UK Statistics Authority

National Statistician’s Data Ethics Advisory Committee

NSDEC(18)16

Testing the use of dependent interviewing with administrative data

This project proposal has not been approved by the committee

NSDEC(18)17

Crime Survey for England and Wales – 10-15 year old questionnaire – Cyber Crime Module

This project is undergoing major revisions and will be published in due course

Sexual orientation, gender identity and armed forces question implementation:
2021 Census, England and Wales

Oral report

Garnett Compton

Amie Kamanda
Michelle Monkman
Maria King

UK Statistics Authority

National Statistician's Data Ethics Advisory Committee

NSDEC(18)18

Collecting and using social media data for the production of statistics and research

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 media analysis of public perceptions to inform question development

Start Date:	25 July 2018	**End Date: 2021 for Census 2021 work.**	

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

Office for National Statistics (ONS)

Project Summary

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

The Data Collection Methodology (DCM) team in ONS provides guidance and advice, and tests question designs for Census, Social and Business surveys. DCM have tested proposed 2021 Census question designs on gender identity, marital status and sexual orientation among other topics. Traditionally, DCM have conducted focus groups and interviews as preliminary research methods to explore attitudes on sensitive or new topics before incorporating the findings into question development for cognitive testing. There are advantages in adopting these methods such as gaining rich insight into topics. However, these methods can be costly, time consuming to collect, open to social desirability bias and not always representative of views of the broader population.

The publics’ use of social media is becoming a prominent feature of modern society. An ever-increasing part of our social interaction is taking place online. Analysing public comments and metadata on social media platforms is becoming common place within social research, academic institutes and government departments. Social media analysis is focused on understanding insight, meaning and values of human behaviour through social media platforms.

DCM considers social media analysis a beneficial approach to gaining background knowledge on the acceptability of a question topic to drive question development. Using this technique, we can capture data from diverse members of the public and capture opinions that are more likely to reflect accurate views. This is because the data is collected in a naturalistic and unobtrusive way and is less likely to be subjected to social desirability bias.

DCM are seeking ethical approval to conduct social media analysis as a consultation process for DCM use to inform question development for cognitive testing. We would like to carry out this social media consultation process for all topics that DCM are commissioned to test going forward. This method does not intend on replacing traditional stakeholder engagement methods that, for example, our clients carry out. DCM’s social media analysis would strengthen the evidence base of providing insight into public attitudes of the question topic in conjunction with traditional stakeholder engagement methods.

Section A
Project Details

A1 Legal gateways

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

A2 Ethical approval

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

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

---

A3 Proposed site of research select all that apply

☑ ONS

☐ SRS (formerly VML)

☐ HMRC Data Lab

☐ Other

please specify

☐ ADRC - England

☐ ADRC - Scotland

☐ ADRC - Northern Ireland

☐ ADRC - Wales

A4 Data subjects to be studied

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

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

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

Although we will not focus on specific population groups for this research, we will consider any biases based on the demographic characteristics using social media platforms.

Justification for focusing on these subsections or groups:

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

Aims

DCM are seeking ethical approval to analyse social media comments and metadata. This would help us gain better understanding of public views on question topics and therefore contribute to question development. In this application DCM are seeking to gain ethical approval to conduct social media analysis as a consultation process we adopt for all question topics we are commissioned to carry out question testing on.

Methods

DCM will develop their social media analysis process using the Gender Identity topic as our prototype to explore the feasibility of collecting evidence through social media analysis. But we intend on using this process as an initial consultation exercise for all topics we are commissioned to test questions on. Our initial topic will be Gender Identity because we have sufficient background knowledge on this topic and we have tested Gender Identity question designs for the 2021 Census question development.

Data collection

We aim to extract data within the parameters of the social media platforms' terms and conditions by using their publicly available Application Program Interfaces (APIs). Social media platforms we aim to extract data from are: YouTube (via Google API), Facebook, Reddit, and Twitter. Where there requires a log in to use the public API and a reason for analysis we will outline that the data is for ONS and not use personal accounts. These platforms were selected given they have public APIs and are platforms where a diverse range of members of the public post. The platforms themselves are largely considered neutral or without their own political agenda which potentially reduces bias in the data extracted.

Extracted data will include: metadata (e.g. time of post, tags), posts, comments, number of posts, 'likes' and 'dislikes'. The extracted data will remain within a secure, limited access research environment (on SharePoint) within the ONS DCM team. Extracted data will not contain identifiers (e.g. name, email address, geolocation, user names, photos identifying respondents) and it will not be possible to directly identify individuals from these data. These data will not be linked to other datasets for any purpose. DCM will own the dataset and the use of data will be recorded on the ONS Information Asset Register and within transparency reports, including what data we hold, and the purpose for which it has been used. Only trained members of DCM who have been through security clearance will be able to access the extracted data (raw data) and no data will leave the data that is stored within the limited access SharePoint folder.

Analysis

We aim to carry out qualitative analysis, sentiment analysis, explore most popular words used using word clouds, the number of ‘likes’ and ‘dislikes’, frequency of comments and opinions over time. The latter will be identified using specific word searches as defined in a variety of academic publications (e.g. https://onlinelibrary.wiley.com/doi/abs/10.1111/josl.12080) , or key terms from discussion titles.

Dissemination

Findings from the analysis will be shared with our internal clients in high level reports. Data will be presented at aggregate level, subject to appropriate disclosure controls. This will be shared within ONS to inform business areas of the evidence behind our question development recommendations and to inform internal business areas of our research. We will share paraphrased quotes internally within these reports. Our methodology (but not findings) will be presented more widely and at conferences as a knowledge sharing exercise.

When required to present the methodology of creating word clouds we will not share real quotes but construct our own quotes to avoid disclosing the identity of data subjects.

A6

Data use

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

Data Level

Please specify the name of the data set

Type of data	**Aggregate Data**	**Identifiable Data**	**De-identified personal data**	**Anonymised/ pseudo anonymised**
Administrative data *(please specify, e.g. Patient Register 2011, School Census 2012 etc, in the relevant options adjacent)*				
Big Data *(please specify e.g. Twitter data, smart meters and mobile phones, in the relevant options adjacent)*			Twitter, Facebook, Reddit, YouTube comments placed on these sites using the sites APIs to extract comments and metadata.	
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)

An advantage of conducting interviews and focus groups are that they provide rich data. However, there can be limitations with these methods. It is sometimes difficult to recruit diverse members of the public and those who do volunteer are sometimes within a certain demographic or hold an opinion that is not always representative of the diverse views of the population. These traditional methods are at risk of social desirability bias (due to researcher or participant presence). [17/07/2018 15:45] There are also difficulties in accessing views and gaining sentiments from the public who are in stealth mode. In data collection terminology, stealth mode refers to members of the population who do not want to be recruited in survey testing because they identify in a certain way, or hold a certain view, that they do not want others in the population knowing about. In capturing the views and sentiments of people in stealth mode, we will not disclose the identity of data subjects or expose sensitive information in a way that may cause them potential harm.

In ONS we need to design questions that capture accurate responses from members of the public including those who are in stealth mode. This enables ONS to produce accurate statistics that drives representative service provision. Recent developments within social research have enabled online opinions to be collected through social media platforms. Social media analysis is naturalistic and unobtrusive. Previous research into online behaviour has demonstrated that online opinions can reflect accurate attitudes rather than socially desirable views. This increases the chances of capturing public opinions from diverse members of society including those who might not have volunteered in research and therefore would have not have had their voices heard through traditional methods. Social media analysis has previously been conducted by UK government departments and academic institutes. For example, HMRC analysed a range of social media sites (Facebook, Twitter, Youtube, Reddit) to explore public opinion of HMRC services. They are considered a 'listening organisation' through their ability to use social media to gain public feedback that drives service improvements (MacNamara, 2016). Likewise ONS would capture public voices to develop survey questions that collects better statistics to drive policy and understanding of our societal needs.

Thus, it would be within the public interest to capture online opinions on potentially new or sensitive question topics. This will strengthen the evidence base on which to develop question designs and therefore capture data using robust methodology that will inform public policy on critical social matters.

It is also within the public interest for ONS to design a survey question that maximises the respondent's positive experience of completing a survey. If we do not understand the diverse public perceptions of a topic there is a risk that ONS design a question where respondents do not understand how to answer. Worst still, respondents might be excluded from being able to answer. This places burden on the respondent which has the potential to result in item-non-response and increase survey attrition rates. The result would be low response rates, less representative data to drive service provision, more resource required by ONS to impute and estimate the data gathered from the survey, dissatisfied members of the public and reputational risk to ONS for not asking questions the public can answer (or answer easily). Improving our surveys based on understanding the question topic, allows data to be captured in a more efficient way that minimises respondent burden, encourages survey completion and enables the production of better statistics. Better statistics can be used to inform the public on social trends, raise awareness on the topic, contribute towards the development of government policies but also inform practices which may directly or indirectly affect specific members of the UK population.

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

Through this research we will develop a social media analysis methodology which DCM will use as a tool and methodology to extract and analyse social media data to establish public attitudes on any question topic DCM are commissioned to test. Our methodology based on this consultation tool will meet the highest ethical standards and will be seen by NSDEC.

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

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

Based on the safeguards in place we anticipate low risk of harm or confidentiality breaches to users of social media sites. Data used in this research will not be published and is used as a consultation tool to facilitate ONS survey question development for testing. Data extracted via the social medias' public APIs will not contain any personal identifiers and no personal identifiers will not be extracted from the social media sites' metadata. Only the contents of the comments (and responses), number of likes/dislikes, number of comments, date of comment, in addition to other non-identifiable metadata will be extracted. While we are not extracting any identifiers using the public APIs, we cannot control for identifiers that users' mention within their comments. If we extract comments that have identifiers we will remove these identifiers from the high-level analysis and quotes. Comments which contain identifiers will not be shared or linked with other datasets.

We will share our high-level findings within ONS to provide insight into the question design recommendations DCM are proposing. We will only share paraphrased quotes with our internal clients. Verbatim quotes will not be shared and any quotes we need to demonstrate our methodological approach will be constructed by our team to ensure that they cannot be traced back to the original producer.
Data will be stored within limited access SharePoint folders within DCM. No raw data will leave the limited access SharePoint folder

A limited number of approved people within DCM will be working with the raw data. The researchers within DCM are learning to use big data and so could require technical and methodological support from the Big Data team and Data as a Service (DaaS). All members of staff will be trained and security cleared.
If there is a way to extract comments from adults only (16 years and over) using public APIs then this method will be adopted (we would need to find out whether this is feasible post ethical approval). While there is no intention of collecting social media comments from minors, in some cases this might not be possible to screen for (e.g. where age can be inferred from the contents of the comments). If identified as a minor from their comments we will remove the data from analysis. The approach we will follow when dealing with data produced by minors was decided following consultation by ONS Legal Services.
In terms of consent, ONS Legal Services have been consulted with regarding this project proposal and they agreed that we do not need to collect consent where data is collected for processing and statistical output production. The terms and conditions of these social media platforms state that their aggregate data could be used for research and it would be disproportionate to seek consent from each social media user. Furthermore, by asking for consent, the value of capturing snapshots of the publics' opinion might be lost as users could seek to edit their comments before they agree on us extracting their data.

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

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

While social media analysis is new to the DCM team there has been sufficient previous research conducted in academic institutes across the UK and UK government departments to inform the development of social media analysis. We have considered many social media projects as illustrated in academic publications and government dissemination and have developed a sound understanding of processes used in social media data analysis and have implemented any lessons identified in our proposal to mitigate risks.

We will adhere to the ONS social media policy principles which is currently in draft format and we will seek methodological and technical support from The Big Data team and Data as a Service (DaaS) which are teams within the same directorate as DCM and have ample experience in data analysis using data science methods.

All data and analysis will be stored within a limited access SharePoint folder. We will use Python or any coding that is compatible with the public APIs available for each social media platform. We will use the public API code provided by the social media site for legal adherence and consistency with the sites terms and conditions. Intermediate and final outcomes will be audited and the quality of outputs will be regularly checked ensuring that there is sufficient human oversight. Given the sensitive nature of this project we will ensure that any processes followed are fully compliant with the terms and conditions of service for the use of the APIs and reviewed by the relevant ONS Methods, Data and Research teams and ONS Legal Services.

The researchers have ample experience in analysing and coding large quantitative datasets and has extensive experience carrying out qualitative analysis on sensitive data. Further training will be sought to ensure that researchers fully understand all aspects of social research and data science using social media data. DCM approaches research on sensitive topics with discretion and are dedicated to ensuring the participants of our research are not subjected to any harm through the research we conduct.

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

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

The ONS Legal Services team have advised that this research proposal is in compliance with current data protection legislation and sufficient controls are in place to ensure the ongoing legal compliance of this research with the terms and conditions of service. If the latter change we will need to consult with ONS Legal Services again to make sure that our proposed used of data continues to be lawful.

B6 Collaboration and Sponsors

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

List of Collaborators/Sponsors	Details and relevant documentation relating to collaboration (you may attach copies of relevant documentation)
Office for National Statistics	Use of APIs provided by Google, Facebook, Twitter and Reddit in full compliance with their T&C.

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

Please list any public engagement activities (max 250 words)

Previous research has collected evidence on public perceptions of the use of digital, online and social media data. This research has demonstrated the positive impact of social media analysis on the public.

Research has explored the acceptability of collecting data through SMS from pregnant women on their attitudes towards infant feeding in Tayside, Scotland. This could be considered a potentially sensitive topic for some expectant mothers. These expectant mothers were positive and thought it was acceptable to collect opinions of this nature through SMS (Whitford et al., 2012).

Stellefson et al. (2014) explored patient engagement with YouTube videos providing advice on COPD management and treatment. They explored likes, dislikes and comments. The lack of engagement enabled the researchers to recommend an improvement in the quality of YouTube videos to improve healthcare engagement amongst patients.

Research published in the peer-reviewed Journal of Child Neurology has explored sensitive topics though social media analysis on YouTube. This research considered the public perception of Tourette syndrome on YouTube.

One peer-reviewed journal article explored cohorts' online interaction with 'Blackboard' to assess learning potential. Regression analysis predicted that the more meaningful comments posted on blackboard the better the quality of learning exhibited by the student cohort.

In the same light DCM wish to obtain social media data to explore perceptions and attitudes of sensitive topics to improve processes in question design development.

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

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

We would not be publishing the results of our social media analysis. We will present our methodology at conferences and papers which may be published on the ONS website. We will not link the data between extracted social media datasets or other data sources. The high-level findings from the research will be shared within ONS and paraphrased quotes may be shared but they will not be verbatim to avoid disclosure.

Raw data with identifiers removed will be retained within a limited access SharePoint for auditing purposes. Only selected members of the DCM team will have access to this area and no data will be shared outside the research team. The extracted raw data will be kept until 2021. Data without identifiers and aggregate data will be kept for longer, according to our retention policies, in the event of follow-ups/queries during quality assurance and auditing processes.

Section C
Responsible owner and applicant details

C1 Responsible Owner

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

Declaration to be signed by the responsible owner

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

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

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

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

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

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

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

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

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

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

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

Please consult the guidance document before filling in the application form

Project Title
Please provide a title indicative of the project

Investigating Mental Health Analysis Programme – phase 1: assessing the risk of having a mild or moderate diagnosed common mental health disorder (CMD) and the risk of death associated with CMDs.

Start Date: 1/8/2018 **End Date:** 1/10/2019

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

Office for National Statistics (ONS)

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

Improving the mental health of the nation is a current priority for the UK government. According to the Five Year Forward View (FYFV), an independent report from the Mental Health Taskforce, one in four adults experiences at least one diagnosable mental health problem in any given year. The FYFV sets out the start of a 10-year journey to transform mental health care. The Government's response to this report accepted the Taskforce's recommendations in full. ONS can inform this transformation, and provide better quality information on the mental health of the population, via a programme of analyses based on linking mental health treatment datasets with selected population-based data.

The project will focus on common mental health disorders (CMDs) such as anxiety and depression. Using a phased approach, we intend to look first at the mortality risk of people with CMDs, comorbidities between mental and physical health problems, and investigate inequalities around mental health. In the second phase, we will investigate income and employment transitions for patients who have been through mental health treatment. This application is for NS-DEC to consider the ethics regarding the first phase; a separate, application will be made for the second phase of the project.

The first phase will address existing evidence gaps on comorbidities between mental and physical health, improve understanding on the demographics of people with CMDs, and investigate whether some or all people with CMDs are more at risk of death than the general population. The Improving Access to Psychological Therapies (IAPT) programme in England treats some 900,000 patients a year with mild and moderate mental health conditions, and will be used as the source of data on mental health. The IAPT data for 2012 to 2017 will be linked to the 2011 Census to provide detailed demographic background, and to death registrations from 2012 to 2018.

The mortality analysis will focus on specific causes of death which may be connected to mental health (suicide, alcohol and drug abuse) as well as overall risk. In addition, the causes of death will be compared to the distribution of causes in the general population to identify any common comorbidity with life-threatening illnesses. This goes some way to provide insight into important issues raised by the FYFV:

"An important barrier to good care is the lack of appropriate data sharing to enable organisations to identify co-morbidities…People with poor mental health may require primary care, secondary physical care and social care, as well as mental health services, but the lack of linked datasets hinders effective provision."

The King's Fund and Centre for Mental Health report found that people with long-term physical health problems and comorbid mental illness disproportionately live in deprived areas. The linked census, IAPT and mortality data will be used to identify patterns in household type, area deprivation and geography, including transitions over time.

Findings from the Adult Psychiatric Morbidity Survey (APMS) also support the bringing together of physical and mental health. The survey found an association between CMDs and physical health, but only considered five key chronic conditions. Linkage to all deaths in England will provide a larger sample to analyse comorbidities between CMDs and less common causes of death not considered in the APMS.

The National Confidential Inquiry into Suicide and Homicide by people with Mental Illness (NCISH) used mortality data to examine suicide deaths in those who had recently accessed mental health services. This provided policy implications for suicide prevention; the proposed linkage would build on this work by considering all causes of death (not just suicide), and using patient records from IAPT (rather than clinician questionnaires), with an aim to achieve a more complete picture.

IAPT data is estimated to cover over 15% of people with CMDs in England. Because of the service's large, national scale and focus on people with mild and moderate mental health conditions, it provides a reasonable proxy for patterns and trends in the population of people with diagnosable CMDs. The IAPT data will be compared with the findings of the APMS (2007 and 2014) to assess likely issues of representativeness, such as the under-representation of specific population groups in the treatment cohort. People with severe mental health conditions not typically treated in the IAPT programme.

The analysis will aim to answer the following research questions:

Risk of having a mild or moderate CMD:
- How is the IAPT treatment population distributed compared to the whole Census population in terms of key demographics such as: age, sex, ethnicity, occupation and employment status, marital status, and household composition?
- Which population groups are most at risk of having a CMD as represented by starting IAPT treatment?
- Secondary: what does comparison of the demographics of IAPT patients with the APMS suggest about the representativeness of the cohort?

Risk of death connected with having a CMD:

- Are people who have been treated in IAPT at a higher risk of death by suicide, drug poisoning or alcohol than the wider population?
- Are people who have been treated in IAPT at a higher risk of death by any other causes?
- Does this vary by demographics, including: age, sex, ethnicity, occupation/NS-SEC, employment status, marital status, area deprivation?
- Are there particular groups identified as at higher risk?

The three-way linkage will provide an independent and more detailed demographic baseline than the IAPT data could do alone, and allow us to investigate if there have been changes in peoples' circumstances between the Census and treatment in IAPT (e.g. becoming disabled or living alone), which could only be undertaken by comparing the two sets of demographic data on the two datasets. In addition, the quality of the demographic data in the IAPT is unknown, and the Census is likely to be more reliable. Having the mortality data linked as well allows us to see the overall tends in mortality, plus to see if there is any relationship between changes in demographics and the cause of death outcomes.

The research is not aiming to look at individual level outcomes or to evaluate the IAPT treatment, but to look for trends in the aggregate data after linkage, to provide population level analysis to inform policy. Entry into IAPT treatment will be used as the main indicator if having a diagnosable CMD. The clinical data will not be analysed except to:

- Group the cohort into broad types of CMD
- Potentially, link successful/unsuccessful treatment outcome to risk of subsequent death.

Answers to the above questions will provide bodies such as Public Health England and NHS England with better intelligence to support the provision of physical and mental healthcare, ultimately saving lives. All outputs from the research for all phases will be published by ONS as aggregated data tables and statistics with the relevant statistical disclosure controls applied.

Section A
Project Details

A1 Legal gateways

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

NHS-Digital controlled IAPT data

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

ONS Mortality data

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

Census Data

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

A2 Ethical approval

Has the project being reviewed or is it expected to be reviewed by another ethics committee?	● **Yes** ○ **No**
If Yes please provide the name of the committee, the outcome and the date approved	

A3 Proposed site of research select all that apply

☑ ONS

☐ SRS (formerly VML)

☐ HMRC Data Lab

☐ Other

please specify

☐ ADRC - England

☐ ADRC - Scotland

☐ ADRC - Northern Ireland

☐ ADRC - Wales

A4 Data subjects to be studied

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

○ **Yes** ○ **No**

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

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

This work will focus only on adults (aged 18+) in England & Wales who accessed IAPT services in England between 2012 and 2017. Any patients aged under 18 years of age will be stripped from the dataset prior to any linkage. The patients aged 18 and over will be linked to the 2011 census and to mortality to flag patients who have died. But both living and deceased IAPT patients will be included in the analysis.

Justification for focusing on these subsections or groups:

The IAPT data contains information on patients who accessed IAPT services. There are therefore limitations to using this data, for example the mental health of the rest of the population, including those accessing secondary or tertiary mental health services, cannot be considered. Comparisons will be made with the APMS to assess how representative the cohort of IAPT patients is likely to be of the wider population with CMDs.

NHS D do hold data on patients accessing secondary mental health services (the mental health services dataset or ‘MHSDS’), including more severe conditions, and both community and hospital treatment settings, but these data are complex. We are aware of the limitations of restricting this first project to IAPT only, but will consider expanding the work to include MHSDS at a later stage to achieve a more representative dataset, and to consider conditions other than anxiety and depression, following the outcome of this initial project.

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

This project will link individual level death records for the years 2012 to 2018, and 2011 Census data to individual level IAPT records for the calendar years 2012 to 2017. The variables that the linked dataset will contain from both the Mortality dataset, Census and IAPT are listed in detail in Annex 1. Broadly, the linked dataset will contain information taken from the death certificate such as underlying cause of death, contributory cause(s) of death, and place of usual residence, alongside 2011 Census demographics and IAPT data around provisional diagnosis, and patient demographics.

The linkage will be based on the 900,000 or so referrals to IAPT each year (965,379 in 2016-17)

The linkage will work as follows

1. IAPT data will be linked by NHS Number to the Patient Register ONS holds to link patient name, date of birth, address, and other demographics needed for linkage to the IAPT data
2. IAPT will then linked by name and address to the 2011 Census data
3. Finally, this data will be linked to the ONS Mortality dataset

This will result in a linked person level file between IAPT, Census and Mortality. The dataset will contain all IAPT cases linked to census with mortality data being linked to the relevant cases. Cases of living patients in the IAPT data will not be removed from the dataset.

Once linked, identifiers will be removed from the linked data and this subset will be stored on a ONS secure data platform (most likely the new corporate solution - the Data Access Platform (DAP), see Annex 2), where two to three researchers will have access to undertake the analysis. All staff who have access to the data will have SC security access and training to ensure they handle data correctly (see section B3).

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

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

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

Type of data	**Data Level** *Please specify the name of the data set* **Aggregate Data**	**Identifiable Data**	**De-identified personal data**	**Anonymised/ pseudo anonymised**
Administrative data *(please specify, e.g. Patient Register 2011, School Census 2012 etc, in the relevant options adjacent)*		IAPT 2012-2017 Mortality 2012-2018 Patient Register 2012-2017 Only available to those conducting the linkage	The analytical dataset will contain de-identified, linked person level data. The analytical dataset, detailed below, will only be available to researchers working on this project.	
Big Data *(please specify e.g. Twitter data, smart meters and mobile phones, in the relevant options adjacent)*				
Survey Data *(please specify e.g.LFS, BRES, etc in the relevant options adjacent)*				
Census Data *(please specify year, e.g. Census 2011 in the relevant options adjacent)*		2011 Census Only available to those conducting the linkage		

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

Section B
Assessment against NSDEC ethical principles

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

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

Mental health is a high priority in government policy, and comorbidities between mental and physical health, as well as inequalities in mental health, are of increasing interest within health policy.

In their response to the FYFV, NHS England set an objective for the majority of new CMD services to be integrated with physical healthcare by 2020/21. This is in line with a King's Fund report and the 2014 APMS, which both provided evidence for the strong links between mental and physical health. Similar research to this proposal, though focused solely on suicide deaths rather than all causes, has been conducted in the NCISH.

This project will add to the evidence base by:

a) providing information on many physical conditions (rather than a focus on only a few key health problems, as in the APMS)
b) providing a detailed demographic context, including information such as ethnicity, sexual orientation, occupation, marital status
c) Investigating inequalities.

Investigating the links between mental health, mortality, and co-morbidity has clear benefits for the public. By determining physical and mental health conditions that commonly co-occur, the government can target its health services to better meet the needs of patients resulting in a better patient experience, and ultimately could saves lives. For example, it may be that a particular cause of death has an increased prevalence in patients with CMD compared to the general population; by ensuring policy makers and clinical staff are aware of this, prevention and intervention could be more targeted.

The second benefit of the project is analysis of inequalities in mental health, in line with the FYFV "*focus on tackling inequalities. Mental health problems disproportionately affect people living in poverty, those who are unemployed and who already face discrimination".* The King's Fund found that people with long term physical health problems and comorbid mental illness disproportionately live in deprived areas. This analysis would allow detailed geographical mapping of those with a CMD who died from particular causes, and analysis by deprivation deciles. Other demographic variables could also be used for inequalities analysis to investigate any difference in premature mortality in certain demographic groups of IAPT users (age, sex, or occupation) versus the general population. Obtaining this information will benefit the public by allowing healthcare providers to target groups who may be disproportionately affected by physical and mental health problems, and subsequently reduce premature mortality due to co-morbidities.

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

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

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

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

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

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

The linked data will be stored on an ONS Secure Data Platform (most likely the DAP). Annex 2 provides detail on the security of the DAP solution. The linkage will be undertaken by a staff from a dedicated team in ONS, specialised in linking sensitive identifiable datasets. They have been trained in data protection. The linkage will be undertaken using NHS Number as a linkage key between the Patient Register spine and IAPT, to join the name and address to the IAPT data. Then NHS Number, name, address and date of birth will be used as linkage keys to mortality and Census data. The linked data will be stored in the secure DAP environment.

The linked data will be stripped of identifiable variables before being accessed by up to three researchers to undertake the analysis. The research will take place in DAP. Record level data will not leave this secure environment.

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

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

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

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

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

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

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

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

No new technologies are to be used.

The linkage will be undertaken using probabilistic match keys. We are confident that the matching should have a high success rate, but until we start the linkage work it is not possible to be absolutely certain.

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

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

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

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

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

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

Main condition:

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

Special category condition:

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

To further comply with data protection legislation-

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

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

B6 Collaboration and Sponsors

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

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

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

Please list any public engagement activities (max 250 words)

While we have not sought the views of the public for this specific project, the research is of clear public benefit as the findings will be used to inform mental health service provision (see section B1).

Regarding public support for mental health research, the public engagement findings from the FYFV explained that "*many people discussed the importance of addressing the broader determinants of good mental health and mental health problems, such as good quality housing, debt, poverty, employment, education, access to green space and tough life experiences such as abuse, bullying and bereavement*". Linkage of the 2011 Census to later IAPT referrals will answer part of this question raised by the public, by assessing whether changes in life circumstances (such as employment or divorce) are commonly found in people with CMDs.

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

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

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

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

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

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

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

As part of this, having case studies that show the positive impact of data sharing is important for securing public buy-in. To have these, we must start somewhere; the wide-ranging stakeholder interest, evidence gaps that could be filled, and potential practical value of this project's outcomes to service delivery mean it will contribute to helping make this case.

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

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

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

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

Following feedback from experts in this group we will then decide how best to disseminate the findings moving forward. No microdata will be shared outside of the small select group of researchers in the Health Analysis and Life Events division in ONS who will access to the de-identified linked data.

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

Section C

Responsible owner and applicant details

C1 Responsible Owner

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

Declaration to be signed by the responsible owner

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

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

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

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

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

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

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

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

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

Full Name:	**Position:**
Address:	**Email:**

Telephone:

Organisation:

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

List of mortality variables

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

Wpla10	Workplace code
REGDETS	ONS unique identifier for death records.
CODT	Cause of death text
CORINQT	Coroner’s text

List of IAPT variables

Variable	**Description**
LOCAL PATIENT IDENTIFIER (EXTENDED)	A number used to identify a PATIENT uniquely within a Health Care Provider. It may be different from the PATIENT's casenote number and may be assigned automatically by the computer system. Where care for NHS patients is sub-commissioned in the independent sector or overseas, the NHS commissioner PAS Number should be used. If no NHS PAS Number has been assigned the independent sector or overseas PAS Number should be used.
PERSON GENDER CODE CURRENT	The classification is phenotypical rather than genotypical, i.e. it does not provide codes for medical or scientific purposes. Notes: National Code 'Not Known' means that the sex of a PERSON has not been recorded National Code 'Not Specified' means indeterminate, i.e. unable to be classified as either male or female. National Codes: 0 Not Known 1 Male 2 Female 9 Not Specified
ETHNIC CATEGORY	The ethnicity of a PERSON, as specified by the PERSON. Note: ETHNIC CATEGORY is the classification used for the 2001 census. The Office for National Statistics has developed a further breakdown of the group from that given, which may be used locally. National Codes: White A British B Irish C Any other White background Mixed D White and Black Caribbean E White and Black African F White and Asian G Any other mixed background Asian or Asian British H Indian J Pakistani K Bangladeshi L Any other Asian background

	Black or Black British M Caribbean N African P Any other Black background Other Ethnic Groups R Chinese S Any other ethnic group Z Not stated National code Z - Not Stated should be used where the PERSON has been given the opportunity to state their ETHNIC CATEGORY but chose not to.1
RELIGIOUS OR OTHER BELIEF SYSTEM AFFILIATION GROUP CODE	A Baha'i B Buddhist C Christian D Hindu E Jain F Jewish G Muslim H Pagan I Sikh J Zoroastrian K Other L None M Declines to Disclose N Patient Religion Unknown
SEXUAL ORIENTATION (CURRENT)	The SEXUAL ORIENTATION of a PATIENT. National Codes: 1 Heterosexual 2 Gay/Lesbian 3 Bi-sexual 4 PERSON asked and does not know or is not sure * Z Not Stated (PERSON asked but declined to provide a response) *1
EX-BRITISH ARMED FORCES INDICATOR	An indication of whether the PERSON is an ex-member of the British Armed Forces, i.e. army, navy or air force, or is a dependant of a PERSON who is an ex-services member. National Codes: 01 Yes - Currently Serving (including reservists) (Retired 1 April 2012) 02 Ex-services member 03 Not an ex-services member or their dependant 04 Dependent of a Current Serving Member (Retired 1 April 2012) 05 Dependant of an ex-services member UU Unknown (PERSON asked and does not know or is not sure) ZZ Not Stated (PERSON asked but declined to provide a response)2

LONG TERM PHYSICAL HEALTH CONDITION INDICATOR (IMPROVING ACCESS TO PSYCHOLOGICAL THERAPIES)	An indication of whether the PATIENT has a Long Term Physical Health Condition, as stated by the PATIENT. National Codes: Y Yes N No U Unknown (PERSON asked and does not know or is not sure) Z Not Stated (PERSON asked but declined to provide a response)1
DISABILITY CODE	The DISABILITY of a PERSON. This could be where: the PERSON has been diagnosed as disabled or the PERSON considers themself to be disabled. A PERSON can have more than one DISABILITY CODE. National Codes: 01 Behaviour and Emotional 02 Hearing 03 Manual Dexterity 04 Memory or ability to concentrate, learn or understand (Learning Disability) 05 Mobility and Gross Motor 06 Perception of Physical Danger 07 Personal, Self Care and Continence 08 Progressive Conditions and Physical Health (such as HIV, cancer, multiple sclerosis, fits etc) 09 Sight 10 Speech XX Other NN No DISABILITY ZZ Not Stated (PERSON asked but declined to provide a response)1
REFERRAL REQUEST RECEIVED DATE	The date the REFERRAL REQUEST was received by the Health Care Provider.
IMPROVING ACCESS TO PSYCHOLOGICAL THERAPIES OPT IN DATE	The date a PATIENT choses to be considered for treatment by an Improving Access to Psychological Therapies Service, where the Improving Access to Psychological Therapies Opt-In Model is used. The IMPROVING ACCESS TO PSYCHOLOGICAL THERAPIES OPT IN DATE will start the Improving Access to Psychological Therapies Referral To Treatment Measurement if the Improving Access to Psychological Therapies Opt-In Model is used.

PROVISIONAL DIAGNOSIS (ICD)	This is the provisional PATIENT DIAGNOSIS for the main condition treated or investigated during the relevant episode of healthcare. Note: ICD-10 diagnostic codes are at least four characters in length. The first character is always alphabetic. Where an undivided three character code is used, the fourth character must be filled with 'X'. Fifth characters should be used in accordance with the National Clinical Coding Standards for (ICD-10). Where they are not used the character must be filled with a '-'. The sixth character of the code is used to designate an asterisk or dagger indicator in ICD-10; it may be an 'A' or 'D'.
YEAR AND MONTH OF SYMPTOMS ONSET (IMPROVING ACCESS TO PSYCHOLOGICAL THERAPIES)	The year and month the PATIENT first experienced the mental health symptoms, as stated by the PATIENT.
PREVIOUS SYMPTOM INDICATOR	An indication of whether this is a recurrence of a previously diagnosed condition, as stated by a PERSON. National Codes: Y Yes N No U Unknown (PERSON asked and does not know or is not sure) Z Not Stated (PERSON asked but declined to provide a response)

IMPROVING ACCESS TO PSYCHOLOGICAL THERAPIES CARE SPELL END CODE	The reason for the termination of an Improving Access to Psychological Therapies Care Spell as determined by the CARE PROFESSIONAL. National Codes: ASSESSED ONLY 10 Not suitable for Improving Access to Psychological Therapies Service - no action taken or directed back to referrer 11 Not suitable for the Improving Access to Psychological Therapies Service - signposted elsewhere with mutual agreement of PATIENT 12 Discharged by mutual agreement following advice and support 13 Referred to another therapy service by mutual agreement 14 Suitable for Improving Access to Psychological Therapies Service, but PATIENT declined treatment that was offered 15 Deceased (Assessed Only) 97 Not Known (Assessed Only) ASSESSED AND TREATED 40 Stepped up from low intensity Improving Access to Psychological Therapies Service 41 Stepped down from high intensity Improving Access to Psychological Therapies Service 42 Completed scheduled treatment 43 Dropped out of treatment (unscheduled discontinuation) 44 Referred to non Improving Access to Psychological Therapies Service 45 Deceased (Assessed and Treated) 98 Not Known (Assessed and Treated)
END DATE (IMPROVING ACCESS TO PSYCHOLOGICAL THERAPIES)	Notes: END DATE (IMPROVING ACCESS TO PSYCHOLOGICAL THERAPIES) is the same as attribute ACTIVITY DATE where the ACTIVITY DATE TYPE is National Code 'End Date'. END DATE (IMPROVING ACCESS TO PSYCHOLOGICAL THERAPIES) is the date the PATIENT is deemed by the CARE PROFESSIONAL to have completed treatment and discharged from the Improving Access to Psychological Therapies Service.1

ADULT MENTAL HEALTH CARE CLUSTER CODE (FINAL)	The Adult Mental Health Care Cluster assigned to a PATIENT. National Codes: 00 Care Cluster 0 - Variance (unable to assign ADULT MENTAL HEALTH CARE CLUSTER CODE) 01 Care Cluster 1 - Common Mental Health Problems (Low Severity) 02 Care Cluster 2 - Common Mental Health Problems (Low Severity with Greater Need) 03 Care Cluster 3 - Non-Psychotic (Moderate Severity) 04 Care Cluster 4 - Non-Psychotic (Severe) 05 Care Cluster 5 - Non-Psychotic Disorders (Very Severe) 06 Care Cluster 6 - Non-Psychotic Disorder of Over-Valued Ideas 07 Care Cluster 7 - Enduring Non-Psychotic Disorders (High Disability) 08 Care Cluster 8 - Non-Psychotic Chaotic and Challenging Disorders 09 Care Cluster 9 - Cluster Under Review - Note: This CARE CLUSTER is under review and should not be used. 10 Care Cluster 10 - First Episode Psychosis 11 Care Cluster 11 - Ongoing Recurrent Psychosis (Low Symptoms) 12 Care Cluster 12 - Ongoing or Recurrent Psychosis (High Disability) 13 Care Cluster 13 - Ongoing or Recurrent Psychosis (High Symptoms and Disability) 14 Care Cluster 14 - Psychotic Crisis 15 Care Cluster 15 - Severe Psychotic Depression 16 Care Cluster 16 - Dual Diagnosis 17 Care Cluster 17 - Psychosis and Affective Disorder (Difficult to Engage) 18 Care Cluster 18 - Cognitive Impairment (Low Need) 19 Care Cluster 19 - Cognitive Impairment or Dementia Complicated (Moderate Need) 20 Care Cluster 20 - Cognitive Impairment or Dementia Complicated (High Need) 21 Care Cluster 21 - Cognitive Impairment or Dementia Complicated (High Physical or Engagement)
EMPLOYMENT STATUS	The current EMPLOYMENT status of a PERSON. National Codes: 01 Employed 02 Unemployed and actively seeking work 03 Undertaking full (at least 16 hours per week) or part-time (less than 16 hours per week) education or training as a student and not working or actively seeking work 04 Long-term sick or disabled, those receiving government sickness and disability benefits 05 Looking after the family or home as a homemaker and not working or actively seeking work 06 Not receiving government sickness and disability benefits and not working or actively seeking work 07 Unpaid voluntary work and not working or actively seeking work 08 Retired ZZ Not Stated (PERSON asked but declined to provide a response)

PSYCHOTROPIC MEDICATION USAGE	An indication of whether the PATIENT is taking Psychotropic Medication, as stated by the PATIENT. National Codes: 01 Prescribed but not taking 02 Prescribed and taking 03 Not Prescribed UU Unknown (PERSON asked and does not know or is not sure) ZZ Not Stated (PERSON asked but declined to provide a response)
STATUTORY SICK PAY INDICATOR	An indication of whether a PERSON is currently receiving Statutory Sick Pay, as stated by the PERSON. National Codes: Y Yes N No U Unknown (PERSON asked and does not know or is not sure) Z Not Stated (PERSON asked but declined to provide a response)
PHQ-9 TOTAL SCORE	The Patient Health Questionnaire-9 (PHQ-9) is a type of ASSESSMENT TOOL. The Patient Health Questionnaire-9 is designed to facilitate the recognition and diagnosis of depression in PATIENTS. For PATIENTS with a depressive disorder, a Patient Health Questionnaire-9 can be calculated and repeated over time to monitor change. The responses for each of the 9 items are as follows: 0 - Not at all 1 - Several days 2 - More than half the days 3 - Nearly every day The score ranges are: 0 - 4 None 5 - 9 Mild 10 - 14 Moderate 15 - 19 Moderately Severe 20 - 27 Severe If one or two values are missing from the score, then they can be substituted with the average score of the non-missing items. Questionnaires with more than two missing values should be disregarded. The recommended cut off score for the Patient Health Questionnaire-9 is 9, so a PERSON scoring 10 or above will be considered to be suffering from clinically significant symptoms of depression.

List of Census 2011 variables

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

Annex 2: Overview of Data Access Platform (DAP)

Statistical data processing

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

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

Data security

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

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

Any other business