

Background

To access secure research data in a safe setting provided by a [DEA Accredited Processor](#), researchers must submit a research project or exploratory analysis accreditation application. This can be done in parallel with the [Accredited Researcher application process](#).

This example application is designed to be used alongside the [Research Project Accreditation Application Guidance](#), also available on the UK Statistics Authority website. This guidance provides researchers with examples of how best to complete a Research Project Application Form, to be assessed by the [Research Accreditation Panel](#). The examples below are taken from multiple projects that were accredited and allowed to proceed.

This guide provides specific detail on what is expected from research project applications that are submitted for accreditation. The guide has pinpointed a number of sections of the research project application form where particular detail is required. Researchers should use the examples and guidance provided in this document when completing research project applications. Researchers that require further support with research project applications should contact [their associated DEA accredited data processing environment](#).

7 Abstract of the research project: *

Include a short description of the project and its benefits, **in no more than 100 words.**

This project would analyse how the tax system affects the decisions made by business owners, particularly the decisions over whether to incorporate, how much to invest and in what assets, and the extent to which they retain profits in their companies. Business owners are an increasingly important part of the workforce, and often receive preferential tax treatment – this research will cast light on how the tax system affects a variety of business outcomes, including growth, investment, as well as the extent of tax avoidance.

Sticking to this **word limit is key**. The research project abstract may be published on the public register of research projects.

Write all terms out in **full on first use**, then use **acronyms after**.

8 Purpose of Research Project: *

Provide a detailed description of the purpose for which the data are requested, describing the aims of the study/research in no more than 500 words. Where research is part of a larger programme please include details below.

We request data from the Small Business Surveys and Longitudinal Small Business Surveys to complement our ongoing research using data from administrative tax records and firm accounts that are being accessed in **HM Revenue and Customs (HMRC) Datalab**.

This broader research project aims to understand how policy affects business start-up and subsequent outcomes. Business owners are an important group – they account for 40% of the growth in the workforce since 2008. The government has a range of policies, including tax breaks, targeted at people running their own businesses in the hope this will encourage entrepreneurial activity. But there is a great deal of variation in the types of people who operate businesses, and poorly targeted policies can lead to inefficiencies and tax avoidance. Analysis by the Office for Budget Responsibility and HM Revenue and Customs suggests that increased growth in the owner-manager population (compared with employment growth) will reduce tax receipts by £3.5 billion in the next five years alone, with growth in self-employment costing a further £1 billion.

There is thus widespread interest in understanding more about business owners, including how their behaviour affects tax receipts. **Through this study we seek to address the following questions:** how do different aspects of the tax system affect incentives to incorporate? How does the behaviour of incorporated business owners differ from unincorporated business owners? What do these changes in legal forms mean for aggregate incomes and tax receipts? What would be the effects of various potential reforms e.g. removing Entrepreneur's Relief?

Ensure **all research questions** in a single project application are **directly related**. Unrelated research themes undertaken by the same researcher must be presented in a separately submitted research project application. These project applications will be considered separately.

Specify your research aims, including a **clearly defined hypothesis** (or hypotheses), or **research questions** to be answered.

9.2 Please explain how the methods selected will lead to valid conclusions and the extent to which this method will realise the public benefit of the project.

The level of detail of the explanation should be proportionate to the complexity of the methods chosen. This should not exceed 500 words.

How the methodological approach answers the research aims

We will use national ONS controlled census and HES data that has been linked to COVID-19 mortality, and the COVID Infection Survey. Within these datasets, we will **use statistical modelling to examine whether the increased risk in minority ethnic groups is explained by differences in underlying factors** linked to social inequality, deprivation or chronic disease prevalence. We will also seek to understand **whether the factors that help explain differences in risk between ethnic groups have changed between the first and second waves of the pandemic.**

How the methods used will lead to valid conclusions

To **quantify age adjusted risk of COVID-19 mortality across sex stratified groups we will use survival modelling, reporting absolute and relative risk** (Cox proportional hazards, checked for assumptions). Ethnicity will be defined based on census data and relative risk will be calculated against the reference of White British.

To **investigate the relative importance of selected sociodemographic and clinical risk factors we will use Cox proportional hazard models.** Considered sociodemographic factors will include general health, disability, multiple deprivation (IMD), household deprivation, household tenure, social grade, level of highest qualification, household size, multigenerational household, household with children, key worker type, key worker in the household (yes, no), exposure to disease.

Considered clinical factors will include cardiovascular disease (i.e. myocardial infarction, heart failure), cancer, influenza pneumonia, kidney disease, serious mental illness, respiratory failure. Multicollinearity will be assessed, and an appropriate index variable selected as appropriate.

Factors will be assessed for their overall association with the outcome (COVID-19 mortality), their association with the exposure of interest (ethnicity) as well as whether the strength of association with the outcome (COVID-19 mortality) is modified by ethnicity (using interaction terms). **Significant interactions will be explored by ethnic stratification.** **We will also undertake logistic regression and cluster analysis to investigate whether symptoms of COVID-19 differ by ethnic group.** This work will be run in collaboration with experts in machine learning and cluster analysis (applicants Evgeny and Gorban).

How the methods enable the public good of the research

This work will unpick **why minority ethnic groups may be at increased risk and whether this increased risk is spread equally across the population.** For example, is the increased risk in minority ethnic groups explained by living in more deprived areas or by living with more people? **Addressing these questions will help inform public health priorities and actions in the short-term** (e.g. understanding the sociodemographic and clinical factors that predict greatest risk could help effectively target populations with vaccination and prevention policies) **and longer-term** (e.g. quantifying how policies aimed at targeting markers of deprivation generally will affect ethnic inequalities specifically).

The UK Statistics Authority is seeking assurance that the choice of the methodology is **appropriate to achieve the research aims and proposed public good** of the research project.

Please include the following in your explanation:

- o Include an explanation of **how your methodological approach will answer the research aims/questions** set out in the project application.
- o Explain how the methods used will **lead to valid conclusions** and **enable the public good** of your research to be achieved.

10 Data Required⁴:

10.1 List the title(s) and years/releases of the dataset(s)⁵ you will use. Make sure you include enough detail to uniquely identify the data you want.

Wealth and Assets Survey, Waves 1-5, 2006-2016: Secure Access (with additional waves as these become available)

Annual Survey of Hours and Earnings, 1997-2018: Secure Access (with additional waves as these become available)

10.2 If you intend to bring in any data for your project, give details of the data including who the owner is and provide evidence that the owner has given permission for their data to be used by you for this research.

We would like to bring in external publicly available data on:

1. average housing prices at the local authority, year/month level (publicly available data from HM Land Registry/ONS - <http://landregistry.data.gov.uk/app/ppd/> ; <https://www.gov.uk/government/statistical-data-sets/uk-house-price-index-data-downloads-july-2019>)
2. inflation, year/month level (publicly available data from ONS - www.ons.gov.uk/economy/inflationandpriceindices/datasets/consumerpriceinflation)
3. rate of planning refusal on major projects, at the local authority, year level (statistics produced by MHCLG and available on an Open Government licence)
4. Share of land 'undevelopable' because it lies within an area of greenbelt, national park, area of special scientific interest or area of outstanding national beauty (calculated using GIS software from Ordnance Survey data available on an Open Government Licence) inflation (from ONS)
5. Share of developable land that is exposed to geological risk (such as risk of landslide, flooding, shrink swell etc), calculated from British Geological Survey Geo Sure 5km hex grid (available on Open Government Licence).
6. Share of land already developed from the Land Cover Map of Great Britain (1990)

We will derive summary variables from these datasets at the local authority level and merge these to ASHE/WAS based on the Local Authority in which an individual resides or works (either in the year/month of interview or in a previous year).

Provide enough details about the various data you wish to access to allow them to be **unambiguously identified, including the years** of the data you wish to access. Remember not to use acronyms.

Ensure you have included **all datasets** you are requesting access to. Subsequent requests for additional data may require further consideration by the Research Accreditation Panel. **Ensure requested time period of the datasets are proportionate** to your research questions and outcomes.

Ensure the project application specifies which regions of the UK are within scope of this research, which is supported by geographical scope of the datasets requested

If you intend to bring in other data into a secure environment, be that publicly available or secure data, you should **describe the data, its source and who the owner is**. It is **important to detail all data** you wish to use, as **subsequent requests for additional data may require further consideration** by the Research Accreditation Panel.

Explain why you are applying for access to legally protected data. **Provide detail on the additional insight this data will provide** that publicly available data cannot give you. .

10.3 Explain why access to legally protected (unpublished) data is needed. Please state what other data sources have been considered and why they are not sufficient for your purposes.

The ASHE micro-data is the best data available for examining dynamics of pension contributions. It has a long time series (allowing age/time effects to be distinguished and examined, and comparisons before and after recent policy reforms such as automatic enrolment to be made), a large sample size (allowing robust analysis, and analysis of subgroups such as relatively low level geography) and it is high quality (being completed by employers there is less risk of measurement error). The panel aspect also allows pension behaviour of individuals to be examined with respect to changes in their own earnings or other employment characteristics.

The Wealth and Assets Survey data is best placed for examining questions that require information on the wider household wealth portfolio, since it collects comprehensive and detailed measures of wealth, or broader household characteristics. The Secure Access version of WAS is required as for some aspects of analysis (e.g. examining the substitution between pension and housing saving) as we require knowing location of residence at a more disaggregate level than region. The Secure Access version also far more easily facilitates longitudinal analysis, which is important for our analysis that examines how saving behaviour changes as individuals' circumstances change.

Explain why you have applied for data at **the specific level you have requested**. Provide detail on why aggregate data would not be sufficient.

Ensure you demonstrate how the research project serves the public interest as set out in the **Research Code of Practice and Accreditation Criteria**. Public interest should be articulated in a way that meets the required standard, the UK Statistics Authority's Centre for Applied Data Ethics has produced guidance for researchers to support with the articulation of the public good of a research project. Please visit the website for further guidance.

Provide as much detail as possible about how your analysis will provide a public good, including **references to specific policies** or decisions that you anticipate the research will provide an evidence base for.

Only complete the public good statements that **apply to your research project**.

13 Public Good:

13.1 Please describe how your research project will provide a public good. Complete all the sections that apply.

Public Good	Describe how this research project will provide this public good
Provide an evidence base for public policy decision-making	Small business owners are an increasingly important part of the workforce, and often attract preferential tax treatment and are targeted by a range of other policies. These are often justified by the desire to promote economic growth and boost productivity, but can lead to costly tax avoidance and a misallocation of people and capital. This research project will provide vital evidence on how policy affects a range of business outcomes, and the potential impact of various policy reforms.
Provide an evidence base for public service delivery	
Provide an evidence base for decisions which are likely to significantly benefit the UK economy, society or quality of life of people in the UK	We will be able to quantify the costs and benefits of various existing policies and potential reforms. These have important implications for the government finances and the wider economy. Favourable tax treatment of small business can be costly – for example, Entrepreneurs' Relief (a reduced rate of capital gains tax applied to business assets) is estimated to cost the government in excess of £2.4 billion a year.
To replicate, validate or challenge Official Statistics	
To replicate, validate or challenge existing research	
To significantly extend understanding of social or economic trends or events by improving knowledge or challenging widely accepted analyses	Small business owners have accounted for 40% of growth in the workforce since 2008 , so understanding how policy affects both this trend, and the economic impact of their growth is vitally important.
To improve the quality, coverage or presentation of existing statistical information	

Key aspects

Population Coverage: Who will likely benefit (e.g., are the benefits focused on a particular group within society or will they impact the population as a whole?).

Type of Impact: The overall way in which you are contributing to public value (e.g., are you primarily influencing decision-making, service delivery or policy development, or focused more on extending understanding, replicating, validating, challenging or improving statistics or research?). Remember, public good can be achieved by doing just one of these things or all of them.

Impact Strategy: What practical steps might you take to ensure that the research is seen, and that the public good of the research is realised?

Scale of Impact: The likely size of the public benefit that you expect to achieve (e.g., is this work potentially influencing a policy that will have substantial social or economic impact?).