

National Statistician's Data Ethics Advisory Committee

Correspondence Minute

Papers reviewed via correspondence, September 2021.

1. **Addendum (3) to previous paper: To determine the population-level relative risk of hospitalisation or death that COVID-19 presents to people with different socio-demographic characteristics and co-morbidities. NSDEC (20)12.**
- 1.1 Jonny Tinsley and Hannah McConnell provided a paper on the Coronavirus risk factor research previously reviewed in June 2020, December 2020 and February 2021. The addendum outlined updates to the project in the following areas.
 - i. The paper set out new additional data sources that correlate with new research questions for investigation. This includes:
 - a. The COVID-19 vaccination data supplied by NHS Digital (NHSD) for England. This data has the same underlying source as data currently used but has undergone more data cleaning and validation through NHSD systems. This data also includes information on adverse reactions to the vaccination.
 - b. The COVID-19 Ethnic Category Data Set, which is a derived dataset owned by NHS Digital, that records NHS patient ethnicity indicators. This will provide greater in depth understanding of ethnicity data within the linked pandemic dataset.
 - c. The Improving Access to Psychological Therapies (IAPT) data will be used to expand the mental health analysis;
 - d. The Annual Population Survey (APS) which provides a large representative sample of the England and Wales population, where the characteristic information for this group is more up to date compared with the 2011 Census data.
 - e. The genomics testing data will allow for the investigation of differences in outcome according to the variant of COVID-19 involved in an infection. This data refers to the genomics of the virus, not the infected individual.
 - f. An early extract of the Census 2021 data, for internal use, in late 2021. For purposes of this addendum, this early extract will only be used to link to the existing Census 2011 based cohort of subjects that has been the basis of the risk factors project to date. This will allow up to date characteristics information on occupation, ethnicity, disability status and religion.
 - ii. The feasibility and quality testing of linking hashed DWP Customer Information System (CIS) data, that ONS already holds, with the Census 2011 data. This is a separate and new project, which is closely related to the main COVID-19 risk factors research conducted by ONS, that has been commissioned by HM Treasury/Cabinet Office, and will be delivered in partnership with DWP and the Department for Health and Social Care (DHSC). This project aims to investigate the long-term employment and health outcomes due to COVID-19 particularly on disproportionately impacted groups, such as disabled people.

- iii. The paper also outlined plans to extend the research to include analysis for Wales. To date there has only been records for NHS patients in England but not in Wales, thus restricting analysis to England. This work will link medical records for patients in Wales to the 2011 Census and other datasets to fill this gap.
- 1.2 The Committee acknowledged the clear public interest and therefore supported the developments outlined in the addendum. The Committee raised the following points for further consideration by the researcher:
- i. Due to the very rich and wide-ranging nature of the dataset produced by this study, with a high degree of very sensitive content, the NSDEC raised the importance of ensuring that all trusted research environments that may potentially hold this data in the future are evaluated to have the equivalent level of security as the ONS, if this data is made available more widely. This could be achieved through use of the Digital Economy Act accreditation standard.
 - ii. The NSDEC welcomed the restriction of the data to research and statistical purposes only and stress the importance that this is adhered to. The NSDEC did note, however, that institutions may believe that data should be made available for operational purposes if it identifies those with a higher health risk as there is a moral obligation to approach them directly. It was advised that this would require wider public engagement before this level of 'targeting' is adopted.
 - iii. Given the focus of this project on public health, the Committee agreed that most of the public may not immediately understand or anticipate why income-related data would be used for this project. Therefore, it was suggested that a highly transparent approach be taken to communicate the use of this data in order to maintain public confidence.
 - iv. The NSDEC supported the investigation into the feasibility of using the DWP data as outlined above, but requested that this element of the project be brought back to NSDEC for further consideration if it expands beyond the feasibility stage.
 - v. Legal advice was requested regarding the Ethnic Category Data Set. Furthermore, the NSDEC requested further detail on the ethnicity information and questioned whether this had been created by algorithmic analysis, without the individual's input/consent.
 - a. The researchers confirmed that this data was being shared with the ONS via the appropriate legal gateways. Further assurance was also given regarding the production of this dataset, confirming that no algorithmic tool was used, rather the ethnicity data was derived from the most recent value recorded in one of the underlying datasets.

1.3 Action - Jonny Tinsley and Hannah McConnell to provide responses and assurances to the secretariat