**Study protocol**: Risks of mortality and hospitalisation related to COVID-19 associated with learning disability using the OpenSAFELY platform

**25th January 2021**

The OpenSAFELY Collaborative, Elizabeth J Williamson\*, Helen I McDonald\*, Krishnan Bhaskaran, Kevin Wing, Alex J Walker, John Tazare, Laurie Tomlinson, Sebastian Bacon, Chris Bates, Helen J Curtis, Hannah Kuper, Harriet Forbes, Caroline Minassian, Caroline E Morton, Emily Nightingale, Amir Mehrkar, Dave Evans, Brian D Nicholson, Peter Inglesby, Brian MacKenna, Jonathan Cockburn, Will Hulme, Jess Morley, Ian Douglas, Christopher Rentsch, Rohini Mathur, Angel Wong, Anna Schultze, Richard Croker, John Parry, Frank Hester, Sam Harper, Rosalind M Eggo, Stephen Evans, Ben Goldacre†, Liam Smeeth†

† Joint principal investigators; \* joint first authors

Abstract

On March 11th 2020, the World Health Organization characterised COVID-19 as a pandemic. Various patient groups have been found to have higher risks of severe outcomes from COVID-19, such as death or hospitalisation, including those with chronic heart disease and respiratory disease. Studies have suggested large increases in risks of poor COVID-19 outcomes among people with learning disabilities. Whether this is restricted to those living in residential care or those with profound learning disability remains unknown.

The aim of this study is to use linked electronic health records within the OpenSAFELY platform to describe the risk of hospitalisation and mortality from COVID-19 in England among adults with learning disabilities compared to the general population.

Keywords

Learning disability; COVID-19; absolute risk; mortality; hospitalisation.

# Background and aims

On March 11th 2020, the World Health Organization characterised COVID-19 as a pandemic. Various patient groups have been found to have higher risks of severe outcomes from COVID-19, such as death or hospitalisation, including those with chronic heart disease and respiratory disease. Studies have suggested large increases in risks of poor COVID-19 outcomes among people with learning disabilities. Whether this is restricted to those living in residential care or those with profound learning disability remains unknown.

The aim of this study is to use linked electronic health records within the OpenSAFELY platform to describe the risk of hospitalisation and mortality from COVID-19 in England among adults with learning disabilities compared to the general population; to separate the risk associated with profound learning disability from milder learning disability; and to establish risk experience by people with learning disability among the population not yet included in the Phase 1 vaccination priority list.

# Methods

## Study Population, Outcome, Follow-up and Exposure

### Study design

Data will be accessed using OpenSAFELY, a data analytics platform created on behalf of NHS England to address urgent COVID-19 research questions (https://opensafely.org). OpenSAFELY includes the electronic health record (EHR) data of 24 million people currently registered with primary care practices using TPP SystmOne software, representing approximately 40% of the English population (see supplementary materials for more details).

Two population-based observational cohort studies of adult patients in England will be used. The first cohort will comprise adult patients (males and females, aged between 18 and 105 years) registered as of 1st March 2020 in a general practice which employs the TPP system. Patients with missing age or a recorded age over 105 years, missing gender, or missing postcode (from which much of the household and geographic information is calculated) will be excluded. Patients with a record of HIV will be excluded, due to the partial suppression of outcomes for this group, leading to poor outcome ascertainment. The second cohort will be similarly defined, but will include patients registered as of 1st September 2020 in a general practice which employs the TPP system. These two time periods roughly correspond to the two big “waves” experienced in England during 2020.

### Outcomes

The outcomes for this study are (i) COVID-19 related death and (ii) COVID-19 related hospitalisation.

Primary care data will be linked to mortality data from the Office for National Statistics (ONS) mortality data and to Secondary Uses Services Admitted Patient Care data. COVID-19 related deaths will be defined as deaths with an ICD-10 code of U071 or U072 anywhere on the death certificate. COVID-19 related hospital admissions will be defined as admissions with any ICD-10 admission diagnosis (not restricted to primary diagnosis) of U071 or U072.

### Follow-up

The first cohort study will begin on 1st March 2020 and will end on 31st August 2020 (inclusive). The second cohort study will begin on 1st September 2020 and will end on the last date at which outcome data is available.

For each cohort, follow-up for all eligible patients will begin at study start and will end on the first of the outcome of interest or study end date. Note that censoring for competing outcomes (e.g. death due to other causes) will not be undertaken because the focus is on the sub-distribution hazard. In sensitivity analyses, censoring will be undertaken to explore the cause-specific hazard.

### Exposure

Primary care records will be retrieved from the TPP SystmOne electronic health record system. These data include diagnoses (Read 3 CTV3), prescriptions (dm+d), basic sociodemographics and vital signs. Information on exposure and other covariates will be derived from primary care records retrieved from the TPP SystmOne electronic health record system, including diagnoses (Read 3 CTV3), prescriptions (dm+d) and basic sociodemographics.

Learning disability will be identified via codelists (see Appendix for details) and will be split into profound and milder disability.

## Covariates

Codelists used to define the covariates below can be found in the Appendix.

Demographic measures: age (continuous); sex (male or female); e[thnicity](https://codelists.opensafely.org/codelist/opensafely/ethnicity/) (5 category: White, South Asian, Black, Mixed, Other); deprivation (quintile of the index of multiple deprivation (IMD) derived from the patient’s postcode at lower super output area level). Local geographical region will be measured by the Sustainability and Transformation Partnership (STP). We do not have a comprehensive indicator of residential care so we will assume that households with 5 or more adults with codes indicating learning disability are residential care homes.

Obesity will be grouped using categories derived from the World Health Organisation classification of Body Mass Index (BMI; kg/m2): underweight <18.5 kg/m2; obese I 30-34.9; obese II 35-39.9; obese III 40+; or no evidence of obesity or being underweight, with BMI ascertained from weight measurements within the last 10 years, restricted to those taken when the patient was over 16 years old.

Comorbidities will be defined through combinations of clinical measurements, prescriptions, and recorded diagnoses. Other comorbidities to be included: diagnosed hypertension; [chronic cardiac disease](https://codelists.opensafely.org/codelist/opensafely/chronic-cardiac-disease/) including chronic heart failure, ischaemic heart disease, and severe valve or congenital heart disease likely to require lifelong follow up; atrial fibrillation; prior deep vein thrombosis or pulmonary embolism; [diabetes](https://codelists.opensafely.org/codelist/opensafely/diabetes/) (additionally using HbA1c within last 15 months to determine level of HbA1c control, grouped into <58 mmol/mol (good control), >=58 mmol/mol (poor control) and no recent measure); stroke; [dementia](https://codelists.opensafely.org/codelist/opensafely/dementia/); and o[ther neurological conditions](https://codelists.opensafely.org/codelist/opensafely/other-neurological-conditions/) (motor neurone disease, myasthenia gravis, multiple sclerosis, Parkinson's disease, cerebral palsy, quadriplegia or hemiplegia, malignant primary brain tumour, and progressive cerebellar disease).

Asthma will be grouped by use of oral corticosteroids as an indication of severity, with 2 or more prescriptions in the last year taken to indicate severe asthma); cystic fibrosis and associated diseases such as primary ciliary dyskinesia; and other r[espiratory disease](https://codelists.opensafely.org/codelist/opensafely/chronic-respiratory-disease/). Haematological malignancies and non-haematological malignancies will each be grouped according to time since diagnosis (<1 year, 2-<5 years, 5+years). Haematological malignancies will be considered separately from other cancers to reflect the immunosuppression associated with haematological malignancies and their treatment. Li[ver](https://codelists.opensafely.org/codelist/opensafely/chronic-liver-disease/) disease; solid organ transplant (any); and kidney function (ascertained from the most recent serum creatinine measurement taken in the last 5 years excluding the most recent fortnight, where available, converted into estimated glomerular filtration rate (eGFR) using the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation), with reduced kidney function grouped into no evidence of kidney impairment (no creatinine measurement or eGFR>=60 mL/min/1.73m2), stage 3 (eGFR in range 30-<60 mL/min/1.73m2) and stage 4-5 (<30 mL/min/1.73m2) will also be considered. Patients with a history of kidney dialysis or kidney transplant will be included in the category representing stage 4-5.

We will also include common autoimmune diseases including Rheumatoid Arthritis (RA), Systemic Lupus Erythematosus (SLE) or psoriasis; inflammatory bowel disease, asplenia (splenectomy or a spleen dysfunction, including sickle cell disease); and other immunosuppressive conditions including a condition inducing permanent immunodeficiency ever diagnosed, or aplastic anaemia or temporary immunodeficiency recorded within the last year.

Finally, we will include serious mental illness.

## Vaccine priority list

Patients already included in the Phase 1 vaccination priority list in the UK will be identified, as closely as possible, by applying the criteria using our existing codelists. These are: residents in a care home for older adults and their carers; 80 years and over and frontline health and social care workers; 75 and over; 70 and over and clinically extremely vulnerable individuals; 65 years and over; 16-64 years with underlying health conditions putting them at higher risk of serious disease and mortality; 60 years and over; 60+; 55+ and 50+.

The underlying health conditions putting them at higher risk of serious disease and mortality are: chronic respiratory disease, including chronic obstructive pulmonary disease (COPD), cystic fibrosis and severe asthma; chronic heart disease (and vascular disease); chronic kidney disease; chronic liver disease; chronic neurological disease including epilepsy; Down’s syndrome; severe and profound learning disability; diabetes; solid organ, bone marrow and stem cell transplant recipients; people with specific cancers; immunosuppression due to disease or treatment; asplenia and splenic dysfunction; morbid obesity; severe mental illness.

Clinically extremely vulnerable individuals include: solid organ transplant recipients; people with cancer who are undergoing active chemotherapy; people with lung cancer who are undergoing radical radiotherapy; people with cancers of the blood or bone marrow such as leukaemia, lymphoma or myeloma who are at any stage of treatment; people having immunotherapy or other continuing antibody treatments for cancer; people having other targeted cancer treatments that can affect the immune system, such as protein kinase inhibitors or PARP inhibitors; people who have had bone marrow or stem cell transplants in the last 6 months or who are still taking immunosuppression drugs; people with severe respiratory conditions including all cystic fibrosis, severe asthma and severe chronic obstructive pulmonary disease (COPD); people with rare diseases that significantly increase the risk of infections (such as severe combined immunodeficiency (SCID), homozygous sickle cell disease); people on immunosuppression therapies sufficient to significantly increase risk of infection; problems with your spleen, for example splenectomy (having your spleen removed); adults with Down’s syndrome; adults on dialysis or with chronic kidney disease (stage 5); women who are pregnant with significant heart disease, congenital or acquired; other people who have also been classed as clinically extremely vulnerable, based on clinical judgement.

It is not possible for us to identify carers in care homes or frontline health and social care workers. We have limited data on cancer treatments and certain comorbidities and conditions, e.g. cystic fibrosis is poorly recorded in primary care and pregnancy is hard to identify.

## Statistical Analysis

For each cohort study, we will describe demographic characteristics and comorbidities by learning disability status, compared with the general population. We will fit a Cox proportional hazards model, stratified by STP to account for differing patterns of infection over time in different regions, with days in study as the timescale. We will adjust for age using a cubic spline, or a linear term if a likelihood ratio test indicates no substantial improvement in fit using the splines. We will then fit fully adjusted models, including all covariates listed above. Hazard ratios for adults identified as having profound and milder learning disabilities will be estimated and presented with 95% confidence intervals and Wald test p-values. We will use robust standard errors to account for clustering by household.

We will then repeat the multivariable analysis, adding interaction terms between our residential care indicator and learning disability, to separately estimate hazard ratios for people with learning disabilities residing in the community.

We will then exclude all patients identified as being on the phase 1 vaccine priority list, as described above. In the remaining group, we will refit the multivariable models above to estimate hazard ratios for milder learning disability in the population not yet prioritised for vaccination (profound learning disability is in the priority list).

### Missing data

A large number of the candidate predictors will be fully observed, in the sense that an absence of a diagnosis is taken to indicate the absence of disease, as is typically assumed in electronic health record research. While this may lead to issues with misclassification and subsequent interpretation, this does not manifest itself in a missing data problem. Missing data will arise in some demographic variables and clinical measurements, with the later predominantly used to determine severity of certain conditions. The predictors that are expected to have missing data (with anticipated missing rates in brackets) are: ethnicity (~25%), BMI (~20%), hba1c (~20% of patients with diabetes), and kidney function (missingness likely in serum creatinine measurement). Our previous analyses in these data suggested that the missingness mechanism for ethnicity may be somewhat missing not at random in one region, but little evidence against missing at random in the other regions.

A complete case approach will be used for ethnicity, restricting the analysis to the sub-population in which ethnicity is measured. Patients with missing BMI will be assumed non-obese, on the assumption that obesity, if present, are likely to be recorded. Patients with no serum creatinine measurement will similarly be included in the “no evidence of poor kidney function”. Patients with diabetes but no Hba1c measurement will be included in a separate “diabetes, no Hba1c” category.

# Ethics and Information Governance

NHS England is the data controller; TPP is the data processor; and the key researchers on OpenSAFELY are acting on behalf of NHS England. This implementation of OpenSAFELY is hosted within the TPP environment which is accredited to the ISO 27001 information security standard and is NHS IG Toolkit compliant; patient data has been pseudonymised for analysis and linkage using industry standard cryptographic hashing techniques; all pseudonymised datasets transmitted for linkage onto OpenSAFELY are encrypted; access to the platform is via a virtual private network (VPN) connection, restricted to a small group of researchers, their specific machine and IP address; the researchers hold contracts with NHS England and only access the platform to initiate database queries and statistical models; all database activity is logged; only aggregate statistical outputs leave the platform environment following best practice for anonymisation of results such as statistical disclosure control for low cell counts. The OpenSAFELY research platform adheres to the data protection principles of the UK Data Protection Act 2018 and the EU General Data Protection Regulation (GDPR) 2016. In March 2020, the Secretary of State for Health and Social Care used powers under the UK Health Service (Control of Patient Information) Regulations 2002 (COPI) to require organisations to process confidential patient information for the purposes of protecting public health, providing healthcare services to the public and monitoring and managing the COVID-19 outbreak and incidents of exposure. Taken together, these provide the legal bases to link patient datasets on the OpenSAFELY platform. This study was approved by the Health Research Authority (REC reference 20/LO/0651) and by the LSHTM Ethics Board (reference 21863).

# Discussion

Results from these analyses will provide a clear description of the risks of COVID-19 hospitalisation and mortality experienced by adults with profound and milder learning disabilities compared with the general population; whether these risks appear to differ for people living in residential care homes compared with those living in the community, and will provide evidence about whether these risks have changed in the first and second large waves experienced in England during 2020.

### **Data and Software Availability**

All data were linked, stored and analysed securely within the OpenSAFELY platform (<https://opensafely.org/>). Detailed pseudonymized patient data are potentially reidentifiable and therefore not shared. We rapidly delivered the OpenSAFELY data analysis platform without

prior funding to deliver timely analyses on urgent research questions in the context of the global COVID-19 health emergency: now that the platform is established we are developing a formal process for external users to request access in collaboration with NHS England. Details of

this process will be published shortly on the OpenSAFELY website.

Data management was performed using Python 3.8 and SQL, with analysis carried out using Stata 16.1 and Python. All code is shared openly for review and reuse under an MIT open license. All code for data management and analysis is archived online at <https://github.com/opensafely/risk-prediction-research>. All clinical and medicines codelists are openly available for inspection and reuse at <https://codelists.opensafely.org/>.

### **Author Contributions**

B.G. conceived the platform and the approach; B.G. and L.S. led the project overall and are guarantors; S.B. led the software; E.J.W and K.B. led the statistical analysis with advice from HM; C.E.M. and A.J.W. led on codelists and implementation; and A.M. led on information governance. Contributions are as follows: data curation, C.B., J.P., J.C., S.H., S.B., D.E., P.I. and C.E.M.; analysis, E.J.W., K.B., A.J.W. and C.E.M.; funding acquisition, B.G. and L.S.; information governance, A.M., B.G., C.B. and J.P.; methodology, E.J.W., K.B., A.J.W., B.G., L.S., C.B., J.P., J.C., S.H., S.B., D.E., P.I., C.E.M., R.G., D.H., R.K. K. D-O, E.S. and R.P.; disease category conceptualization and codelists, C.E.M., A.J.W., P.I., S.B., D.E., C.B., J.C., J.P., S.H., H.J.C., K.B., S.B., A.M., B.M., L.T., I.J.D., H.I.M., R.M. and H.F.; ethics approval, H.J.C., E.J.W., L.S. and B.G.; project administration, C.E.M., H.J.C., C.B., S.B., A.M., L.S. and B.G.; resources, B.G., L.S. and F.H.; software, S.B., D.E., P.I., A.J.W., C.E.M., C.B., F.H., J.C. and S.H.; supervision, B.G., L.S. and S.B.; writing (original draft), E.J.W., J.T. All authors were involved in design and conceptual development and reviewed and approved the final manuscript.

### **Competing Interests**

All authors have completed the International Committee of Medical Journal Editors (ICMJE) uniform disclosure form at www.icmje.org/coi\_disclosure.pdf. C.B., J.P., F.H., J.C. and S.H. are employees of TPP. A.M. was interim Chief Medical Officer of NHS Digital April–Sept 2019 (left NHS Digital at the end of January 2020) and Digital Clinical Champion NHS England 2014–2015. All other authors have no competing interests.

### **Acknowledgements**

The OpenSAFELY Collaborative are grateful for all the support received from the TPP Technical Operations team throughout this work; for assistance from the information governance and database teams at NHS England and NHSX; and for additional discussions on disease characterization, codelists and methodology with H. Drysdale, B. Nicholson, N. DeVito, I. Lipska, J. Morley, J. Quint. TPP provided technical expertise and infrastructure within their data centre pro bono in the context of a national emergency.

### **Grant Information**

The OpenSAFELY collaborative has received funding from NIHR. The work of B.G. on better use of data in healthcare more broadly is currently funded in part by: the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre, NIHR Applied Research Collaboration Oxford and Thames Valley, the Mohn-Westlake Foundation, NHS England and the Health Foundation; all DataLab staff are supported by the grants of B.G. for this work. L.S. reports grants from Wellcome, MRC, NIHR, UKRI, British Council, GSK, British Heart Foundation and Diabetes UK outside this work; K.B. holds a Sir Henry Dale fellowship jointly funded by Wellcome and the Royal Society; H.I.M. is funded by the NIHR Health Protection Research Unit in Immunisation (a partnership between Public Health England and LSHTM); A.Y.S.W. holds a fellowship from BHF; R.M. holds a Sir Henry Wellcome fellowship funded by the Wellcome Trust; E.J.W. holds grants from MRC; R.G. holds grants from NIHR and MRC; I.J.D. holds grants from NIHR and GSK; H.F. holds a UKRI fellowship; B.D.N. holds an NIHR fellowship. The views expressed are those of the authors and not necessarily those of the NIHR, NHS England, Public Health England or the Department of Health and Social Care. The funders had no role in the study design; the collection, analysis and interpretation of data; the writing of the report; and the decision to submit the article for publication.

# References

[to be added]

# Appendix 1

Table A1. Codelists used to define candidate predictors and suspected COVID-19 (as proxy for infection burden)

|  |  |  |
| --- | --- | --- |
| Variable | Notes | Codelist |
| Suspected COVID-19 |  | <https://codelists.opensafely.org/codelist/opensafely/covid-identification-in-primary-care-suspected-covid-suspected-codes/2020-07-16/> |
| Ethnicity | 5 categories, obtained from 16 (White, South Asian, Black, Mixed, Other) | [https://codelists.opensafely.org/codelist/opensafely/ethnicity](https://codelists.opensafely.org/codelist/opensafely/ethnicity/) |
| Diagnosed hypertension |  | [https://codelists.opensafely.org/codelist/opensafely/hypertension](https://codelists.opensafely.org/codelist/opensafely/hypertension/) |
| C[hronic cardiac disease](https://codelists.opensafely.org/codelist/opensafely/chronic-cardiac-disease/) |  | <https://codelists.opensafely.org/codelist/opensafely/chronic-cardiac-disease/> |
| Atrial Fibrillation |  | <https://codelists.opensafely.org/codelist/opensafely/atrial-fibrillation-or-flutter/2020-07-30/> |
| Lower limb amputation | Combined with surgery for peripheral arterial disease to form a peripheral arterial disease variable | <https://codelists.opensafely.org/codelist/opensafely/amputation/2020-09-21/> |
| Prior deep vein thrombosis / pulmonary embolism |  | <https://codelists.opensafely.org/codelist/opensafely/venous-thromboembolic-disease/2020-09-14/> |
| D[iabetes](https://codelists.opensafely.org/codelist/opensafely/diabetes/) | Combined with Hba1c measure within 18 months to determine level of control | <https://codelists.opensafely.org/codelist/opensafely/diabetes/> |
| S[troke](https://codelists.opensafely.org/codelist/opensafely/stroke-updated/) |  | <https://codelists.opensafely.org/codelist/opensafely/stroke/> |
| Dementia |  | <https://codelists.opensafely.org/codelist/opensafely/dementia/> |
| O[ther neurological conditions](https://codelists.opensafely.org/codelist/opensafely/other-neurological-conditions/) |  | <https://codelists.opensafely.org/codelist/opensafely/other-neurological-conditions/> |
| Asthma | Combined with ICS prescriptions in past year to determine severity | <https://codelists.opensafely.org/codelist/opensafely/asthma-diagnosis/> |
| Cystic Fibrosis and associated conditions |  | <https://codelists.opensafely.org/codelist/opensafely/cystic-fibrosis/2020-07-20/> |
| R[espiratory disease other than asthma](https://codelists.opensafely.org/codelist/opensafely/chronic-respiratory-disease/) or cystic fibrosis |  | <https://codelists.opensafely.org/codelist/opensafely/other-chronic-respiratory-disease/2020-07-20/> |
| Non-haematological cancer | Grouped by time since diagnosis (<1 year, 2-<5 years, 5+years) | <https://codelists.opensafely.org/codelist/opensafely/cancer-excluding-lung-and-haematological/> |
| Haematological cancer | Grouped by time since diagnosis (<1 year, 2-<5 years, 5+years) | <https://codelists.opensafely.org/codelist/opensafely/haematological-cancer/> |
| Lung cancer | Combined with other non-haematological cancer | <https://codelists.opensafely.org/codelist/opensafely/lung-cancer/> |
| L[iver disease](https://codelists.opensafely.org/codelist/opensafely/chronic-liver-disease/) |  | <https://codelists.opensafely.org/codelist/opensafely/chronic-liver-disease/> |
| Kidney dialysis | Used if no kidney transplant since most recent dialysis | <https://codelists.opensafely.org/codelist/opensafely/dialysis/2020-07-16/> |
| Kidney transplant | Combined with non-kidney transplant for transplant indicator. Also used to determine which of dialysis/transplant is most recent. | <https://codelists.opensafely.org/codelist/opensafely/kidney-transplant/2020-07-15/> |
| Organ transplant (other than kidney) | Combined with kidney transplant for transplant indicator. | <https://codelists.opensafely.org/codelist/opensafely/other-organ-transplant/2020-07-15/> |
| A[splenia](https://codelists.opensafely.org/codelist/opensafely/asplenia/) |  | <https://codelists.opensafely.org/codelist/opensafely/asplenia/> <https://codelists.opensafely.org/codelist/opensafely/sickle-cell-disease/> |
| Autoimmune diseases ([rheumatoid arthritis, lupus, psoriasis](https://codelists.opensafely.org/codelist/opensafely/ra-sle-psoriasis/)) |  | <https://codelists.opensafely.org/codelist/opensafely/ra-sle-psoriasis/> |
| HIV | (used to exclude) | [https://codelists.opensafely.org/codelist/opensafely/hiv/2020-07-13/](https://codelists.opensafely.org/codelist/opensafely/hiv/2020-07-13/#full-list) |
| Other immunosuppressive condition | Temporary and aplastic anaemia within last year; permanent ever. | <https://codelists.opensafely.org/codelist/opensafely/permanent-immunosuppresion/>  <https://codelists.opensafely.org/codelist/opensafely/aplastic-anaemia/>  <https://codelists.opensafely.org/codelist/opensafely/temporary-immunosuppresion/> |
| Inflammatory bowel disease |  | <https://codelists.opensafely.org/codelist/opensafely/inflammatory-bowel-disease/2020-04-07/> |
| Learning disability, including Down’s syndrome | TO BE UPDATED | <https://codelists.opensafely.org/codelist/opensafely/intellectual-disability-including-downs-syndrome/2020-08-27/> |
| Serious mental illness |  | <https://codelists.opensafely.org/codelist/opensafely/psychosis-schizophrenia-bipolar-affective-disease/2020-07-09/> |