



Data Explained

Linking Census, Residential and Health Data to Investigate Disability and COVID-19 Health Outcomes in Scotland

Authors: Yusuff Adebayo Adebisi, Nick Bailey, Angela Henderson, Chris Dibben and Serena Pattaro

Date: August 2026

University of Glasgow | ADR Scotland

Summary

This Data Explained describes the linked administrative data resource used to investigate inequalities in COVID-19 infection, hospitalisation and mortality among adults with intellectual and physical disability in Scotland, with particular attention to household and residential context. The linkage enables disability, demographic and socioeconomic circumstances, living arrangements and national COVID-19 outcomes to be examined for the same individuals at population scale (Adebisi et al., 2026a; Adebisi et al., 2026b).

The resource combines: the Community Health Index (CHI) population register; the 2011 Scottish Census; CHI-to-Unique Property Reference Number (UPRN) Residential Linkage (CURL), with UPRN identifiers and property classification data sourced from Ordnance Survey AddressBase Premium; the Scottish Index of Multiple Deprivation (SIMD); Public Health Scotland COVID-19 testing, vaccination and shielding data; Scottish Morbidity Record 01 (SMR01) hospital admissions; and National Records of Scotland death registrations. The residential component therefore builds on the CURL infrastructure (Clark and Dibben, 2022). The final analytic cohort comprised 3,719,651 adults aged 16 years or older who were alive and resident in Scotland on 1 March 2020 and successfully linked to the 2011 Census, with follow-up to 30 April 2022 (Adebisi et al., 2026a).

This document explains how the datasets were used, how key variables were derived, where interpretation requires particular care, and what would make future research easier. The main cautions are: the nine-year gap between Census disability measurement; the pandemic, baseline-only measurement of residential context; changes in SARS-CoV-2 testing over time; heterogeneity within communal establishment categories; and the fact that the comparison group should not be interpreted as a non-disabled population (Adebisi et al., 2026a; Adebisi et al., 2026b).

Project details

The work forms part of a wider ADR Scotland programme investigating socioeconomic, household and environmental risk factors for COVID-19 in Scotland (ADR Scotland, 2021; Pattaro et al., 2025). The disability-focused study was designed to move beyond analyses that treat disability as a single clinical characteristic by linking disability status to the social and residential contexts in which people were living at the start of the pandemic. The study design and planned analyses were published in advance in BMJ Open (Adebisi et al., 2026b), and the linked cohort subsequently supported a national analysis of COVID-19 infection, hospitalisation and mortality among 3.7 million adults in Scotland (Adebisi et al., 2026a; Adebisi et al., 2026b).

The starting CHI population contained 4,892,658 people recorded as alive and resident in Scotland on 1 March 2020 (the date of the first COVID-19 case in Scotland). After excluding 759,698 people aged under 16 years, 4,132,960 adults were eligible for Census linkage. Of these, 3,719,651 (90%) were linked successfully to the 2011 Scottish Census and formed the analytic cohort. The cohort included 17,354 adults classified as having intellectual disability, 377,706 as having physical and/or sensory disability without intellectual disability, and 3,324,591 in the comparison group with no recorded intellectual or physical disability under the study definition (Adebisi et al., 2026a).

All linkage and analyses were undertaken within Scotland's national secure data environment. Data were pseudonymised before researcher access and analysed in the Scottish National Safe Haven, with linkage support from Public Health Scotland's electronic Data Research and Innovation Service (eDRIS). This model of secure, controlled access is part of Scotland's wider Safe Haven infrastructure (Gao et al., 2022). Ethical approval was granted by the University of Glasgow College of Social Sciences Research Ethics Committee (reference 400200099), and data linkage and research use were approved through the relevant Scottish Public Benefit and Privacy Panel for Health and Social Care and Scottish Government/National Records of Scotland governance routes (reference 2021-0119) (Adebisi et al., 2026b).

Initial research questions

The published protocol specified three linked questions. These required a data resource in which disability status, residential context and COVID-19 outcomes could be observed for the same people at population scale (Adebisi et al., 2026b).

1. How did the risks of first recorded SARS-CoV-2 infection, first COVID-19 hospitalisation and COVID-19-related mortality differ between adults with intellectual disability, adults with physical disability and the comparison group?
2. How were COVID-19 health outcomes associated with household living arrangements, including residential setting and household size?
3. Did residential setting modify the association between disability status and COVID-19 health outcomes, including departures from multiplicative and additive no-interaction models?

Research methodology

Study design and index date. We constructed a population-based retrospective cohort using linked, routinely collected administrative data. The index date was 1 March 2020. Eligible cohort members were adults aged 16 years or older who were alive and resident in Scotland at that date, registered within the CHI population spine, linked to a valid residential address/UPRN, and successfully linked to the 2011 Scottish Census. COVID-19 health outcomes were followed from 1 March 2020 to 30 April 2022 (Adebisi et al., 2026a; Adebisi et al., 2026b).

Linkage strategy. CHI provided the individual population spine and the residential address recorded around the index date. Residential addresses were linked through the CHI-UPRN Residential Linkage (CURL) infrastructure to Ordnance Survey AddressBase Premium, assigning a UPRN to each addressable location. The UPRN functions as a place identifier, so people linked to the same UPRN can be treated as co-resident for household derivation. The Basic Land and Property Unit (BLPU) classification attached to the UPRN was then used to distinguish private dwellings from communal establishments (Clark and Dibben, 2022). Census linkage supplied disability status and baseline sociodemographic information, while health and surveillance datasets supplied COVID-19 outcomes variables (Adebisi et al., 2026b).

Disability exposure. Disability status was derived from Question 18 of the 2011 Scottish Census. A sequential classification created three mutually exclusive groups. People reporting 'learning disability' were assigned to the intellectual-disability group regardless of co-occurring conditions. Among those not in that group, people reporting physical disability and/or sensory impairment (deafness/partial hearing loss and blindness/partial sight loss) were assigned to the physical-disability group. Everyone else formed the comparison group. This comparison group can contain people with mental health conditions, developmental disorders, learning difficulty, long-term illness or other conditions and must therefore be described as having no recorded intellectual or physical disability, not as 'healthy' or completely 'non-disabled' (Adebisi et al., 2026b).

Residential setting and household size. Residential setting was measured at baseline using BLPU property classifications linked through UPRN. Private dwellings included houses, flats/maisonettes, sheltered housing and houses in multiple occupation. Communal establishments included settings such as care homes, nursing homes, residential schools, halls of residence and hostels, and were divided into small (<50 linked residents) and large (≥ 50 linked residents) settings (Office for National Statistics, 2023; Adebisi et al., 2026b). Household size was derived by counting CHI-registered residents linked to the same UPRN at baseline, including children even though the analytic cohort itself was restricted to adults. It was grouped as 1, 2-4, 5-7, 8-10 and ≥ 11 residents. UPRN-based approaches have also been used to derive household composition from health-record address data in other UK settings (Johnson et al., 2021; Harper et al., 2024).

COVID-19 health outcomes. First recorded infection was the first laboratory-confirmed positive SARS-CoV-2 test during follow-up. First COVID-19 hospitalisation was identified from SMR01 as an inpatient admission with ICD-10 U07.1 or U07.2 in any diagnosis position, or an admission occurring within 14 days of a laboratory-confirmed positive test. COVID-19-related mortality was identified from National Records of Scotland death registrations when COVID-19 was recorded as an underlying or contributory cause, or when death occurred within 28 days of a laboratory-confirmed infection. These definitions follow the study protocol and Scottish COVID-19 reporting conventions and intentionally capture the first relevant event for infection and hospitalisation rather than recurrent episodes (Scottish Government, n.d.; Adebisi et al., 2026a).

Covariates. Demographic and socioeconomic variables included age, sex, ethnicity, educational qualification and SIMD 2020 deprivation quintile (Scottish Government, 2020). Household-context variables included multigenerational household status and whether another household member was recorded on the shielding list. Health-related variables included individual shielding status and other chronic conditions reported in the Census, while national vaccination records supplied dose information. Covariates were selected a priori and their temporal and causal roles were considered explicitly; later models that conditioned on shielding, chronic conditions and vaccination were interpreted as conditional associations rather than formal mediation analyses or total causal effects (Adebisi et al., 2026a; Adebisi et al., 2026b).

Statistical analysis. For the main disability analysis, Cox proportional hazards models estimated hazard ratios for infection, hospitalisation and mortality using sequential adjustment: unadjusted; age and sex; additional socioeconomic and ethnic factors; shielding; other chronic conditions; and vaccination. Crude incidence rates were expressed per 10,000 person-years. Secondary analyses stratified follow-up around the introduction of Scotland's Test and Protect programme and by age band (Adebisi et al., 2026a). The residential-setting analysis cross-classified disability and residential setting and tested multiplicative interaction using product terms and likelihood-ratio testing, and additive interaction using the relative excess risk due to interaction (RERI), as pre-specified in the protocol (Adebisi et al., 2026b).

Practical data-management point. The linked resource is not a single off-the-shelf dataset. Its value comes from the combination of separate administrative data sources with different reference dates, coding systems and data-generating processes. Researchers reusing these data should create a variable-provenance table before analysis, recording the source dataset, reference date, unit of observation, coding, derivation rule, missingness and whether the variable is fixed or time-varying.

Key variables

Dataset name	Variable(s)
Community Health Index (CHI) register	Population spine; CHI identifier for linkage; age/date of birth; sex; GP registration; alive/resident status; residential address at/around 1 March 2020.
CHI-UPRN Residential Linkage (CURL) + Ordnance Survey AddressBase Premium / Basic Land and Property Unit (BLPU)	Unique Property Reference Number (UPRN); property classification; residential setting (private dwelling, small communal establishment, large communal establishment); co-residence; household size (Clark and Dibben, 2022).
2011 Scottish Census	Question 18 long-term conditions used to derive intellectual disability and physical/sensory disability; ethnicity; highest educational qualification; other long-term/chronic conditions (Adebisi et al., 2026b).
Scottish Index of Multiple Deprivation (SIMD) 2020	Area-level deprivation quintile (1 = most deprived to 5 = least deprived), assigned from residential geography/postcode (Scottish Government, 2020).
Public Health Scotland COVID-19 testing data / COVID-19 Research Database	Date of laboratory-confirmed SARS-CoV-2 tests; first recorded positive test used to define the infection outcome (Scottish Government, n.d.; Adebisi et al., 2026a).
Scottish Morbidity Record 01 (SMR01)	Inpatient admission dates and diagnosis codes; first COVID-19 hospitalisation using U07.1/U07.2 and/or admission within 14 days of a positive test (Scottish Government, n.d.; Adebisi et al., 2026a).
National Records of Scotland death registrations	Date of death; underlying and contributory causes; COVID-19-related mortality based on certification and/or death within 28 days of a confirmed infection (Scottish Government, n.d.; Adebisi et al., 2026a).
Public Health Scotland national vaccination records	COVID-19 vaccine dose dates and dose count; analysis-specific vaccination status derived from these records.
Public Health Scotland National shielding / Highest Risk List	Individual shielding status; household-level shielding exposure after linkage to co-residents at the same UPRN.

Summary of comments on specific variables

Variable name(s)	Comments
Disability status	Derived sequentially from 2011 Census long-term-condition responses. 'Physical disability' in this project is shorthand for physical and/or sensory impairment among people without recorded intellectual disability. The measure is self-reported and predates the pandemic by about nine years (Adebisi et al., 2026b).
Comparison group	Not a healthy control group. It contains people not classified as having intellectual or physical disability and can include mental health conditions, developmental disorders, learning difficulty, long-term illness or other conditions (Adebisi et al., 2026b).
Residential setting	A baseline address-derived measure, not a direct measure of care intensity. BPLU classifications group heterogeneous environments. Small and large communal establishments differ in scale but can include several institution types with different populations and staffing arrangements (Clark and Dibben, 2022; Office for National Statistics, 2023).
Household size	Count of CHI-registered people linked to the same UPRN at baseline, including children. It is a co-residence measure, not a direct measure of crowding or bedroom density. Very large household categories overlap strongly with communal establishments, so household size and residential setting should be interpreted together (Johnson et al., 2021; Harper et al., 2024).
Multigenerational household	Defined algorithmically as a UPRN-linked household containing at least one person aged ≥ 65 years and at least one co-resident at least 20 years younger. It should not be interpreted as a complete typology of family relationships.
First recorded SARS-CoV-2 infection	Captures the first detected laboratory-confirmed infection, not all infections or reinfections. Testing access, policy and uptake changed markedly over the study period and could differ between residential settings and population groups (Scottish Government, n.d.; Adebisi et al., 2026a).
COVID-19 hospitalisation	Hybrid definition using diagnosis coding (ICD-10 U07.1/U07.2) and temporal linkage to a positive test. This improves capture but combines admissions identified through two routes and does not, by itself, establish that COVID-19 was the sole reason for admission (Scottish Government, n.d.; Adebisi et al., 2026a).
COVID-19-related mortality	Combined definition based on COVID-19 recorded as an underlying/contributory cause or death within 28 days of a confirmed infection. Researchers should state this explicitly because it is broader than using either death certification or the 28-day definition alone (Scottish Government, n.d.; Adebisi et al., 2026a).
Vaccination status	Dose information originates from national vaccination records. Coding should match the estimand and time scale of the analysis. In the published main association analysis, vaccination was categorised as 0, 1, 2 or ≥ 3 doses received during follow-up; dose dates allow alternative time-varying specifications where appropriate.
Highest qualification	Measured in the 2011 Census, so it represents educational status at that time rather than at the 2020 cohort index date. The Census 'full-time student' code also captured schoolchildren and people under 16 at enumeration, producing an important age/cohort interpretation issue nine years later.
SIMD 2020	An area-level measure. It should not be interpreted as individual income, wealth or personal socioeconomic position. People living in the same SIMD quintile can have very different individual circumstances.
Shielding variables	The Highest Risk List was a policy/clinical classification of people considered clinically extremely vulnerable, not a general measure of frailty, disability severity or overall health. Household shielding captures the presence of another listed person at the same UPRN, not the actual behaviour of household members.

BPLU, Basic Land and Property Unit; UPRN, Unique Property Reference Number; CHI, Community Health Index

How we dealt with data limitations

The most important exposure limitation was the age of the disability measure. Disability status came from the 2011 Census, approximately nine years before the pandemic. Intellectual disability is generally lifelong, whereas physical disability is more likely to be acquired or to change over time. We therefore avoided describing the Census measure as contemporaneous disability status, interpreted likely misclassification explicitly, and conducted age-stratified analyses. We also repeated the main models after excluding 907 people in the intellectual-disability group who reported upper-secondary or higher qualifications; the estimates were essentially unchanged (Adebisi et al., 2026a).

Recorded infection was especially vulnerable to changing ascertainment because community testing was restricted early in the pandemic. We therefore stratified follow-up at 28 May 2020, when Scotland's Test and Protect programme began. Elevated disability associations with recorded infection persisted in the later period, when community testing was much more widely available. Infection was nevertheless interpreted as recorded or detected infection rather than all SARS-CoV-2 infections (Adebisi et al., 2026a).

Linkage and cohort coverage. The final cohort represented 90% of age-eligible Community Health Index (CHI) records linked to the Census. Unmatched records were excluded rather than imputed because disability and several central covariates were only available from the Census. The possibility of selection arising from non-linkage remains and was reported as a limitation (Adebisi et al., 2026a).

Baseline-only residential context. Unique Property Reference Number (UPRN), household size and residential setting were defined at the index date. People can move, household composition can change and care arrangements can alter during a two-year follow-up. We therefore interpreted these variables as baseline residential context and avoided implying that every participant remained in the same arrangement throughout follow-up (Adebisi et al., 2026b).

Broad residential categories. Basic Land and Property Unit (BLPU) classifications provide objective property-based information, but they do not identify the exact care model, staffing level, support intensity or infection-control capacity experienced by an individual. Communal-establishment categories were therefore treated as broad contextual groupings rather than interchangeable care settings (Clark and Dibben, 2022; Office for National Statistics, 2023).

Outcome definitions and event structure. We used transparent national definitions and modelled first recorded infection and first qualifying hospitalisation. This avoided mixing recurrent events with first-event hazards, but it means the analyses do not describe reinfection burden, repeated admissions or conditional transitions from infection to admission and death (Scottish Government, n.d.; Adebisi et al., 2026a).

Confounding and causal interpretation. Administrative data do not capture every relevant potential determinant of COVID-19 health outcomes. Omissions include disability severity, functional limitation, care intensity, occupation, housing quality and informal care. Covariates were selected a priori, sequential adjustment was reported, and fully adjusted estimates were described as conditional associations rather than definitive causal effects. The interaction analysis likewise identifies statistical effect modification and does not by itself establish the mechanism generating it (Adebisi et al., 2026a; Adebisi et al., 2026b).

Disclosure control. All outputs left the National Safe Haven only after statistical disclosure-control checks. This is essential for protecting confidentiality in small disability-by-residential-setting cells, but it can limit the granularity of published tables and may require aggregation or suppression of very small counts (Gao et al., 2022; Adebisi et al., 2026b).

Suggested improvements recommended to data owners

- Make access more efficient and predictable: streamline application and amendment processes, provide clearer expected timelines, and enable reuse of approved linkage specifications and metadata where appropriate.
- Improve access to GP and primary-care data: research-ready linkage to diagnoses, consultations, prescribing, long-term-condition registers and relevant disability or support-need information would provide more current clinical context than Census data alone.

Additional data which would help to further develop the research

The current linkage can identify important population and residential inequalities, but it cannot directly observe several mechanisms that may explain them. The following additions would substantially strengthen future work, subject to public-benefit assessment, data availability and governance approval.

More contemporaneous disability and functioning data. Future linkage to Scotland's Census 2022 or other suitable data sources, if available for approved research, could update disability status and capture changes since 2011. Measures of functional limitation, severity, communication needs and support requirements would be particularly valuable.

Health and social care support data. Information on social-care packages, supported-living services, care-home residence, personal assistance, hours or intensity of support and unpaid/informal care would help separate residential setting from the care relationships that may generate close-contact exposure.

Longitudinal address and household histories. Repeated UPRN linkage would allow analysis of moves, entry to or exit from communal establishments, household formation/dissolution and changes in household size during the pandemic.

More detailed facility characteristics. For communal settings, data on facility type, resident capacity, staffing levels, staff movement between facilities, room sharing, infection-prevention infrastructure and outbreak history would greatly improve interpretation of setting-level effects.

Occupation and employment circumstances. Job type, workplace exposure, employment status and ability to work from home could help distinguish household exposure from occupational pathways, especially among working-age adults and co-residents.

Treatment and healthcare-pathway data. Linkage to antiviral use, intensive-care admission, treatment escalation, primary-care contacts and other markers of clinical management could help examine whether inequalities arise at exposure, diagnosis, admission, treatment or post-infection care.

Repeated infection and post-acute outcomes. Extending follow-up to reinfections and longer-term outcomes would allow the resource to address cumulative infection burden and post-COVID health consequences rather than only first acute events.

Works cited

- Adebisi, Y.A., Bailey, N., Henderson, A., Dibben, C. and Pattaro, S. (2026a) 'Intellectual and Physical Disability and Risk of COVID-19 Infection, Hospitalisation, and Mortality: A National Cohort of 3.7 Million Adults in Scotland', *Journal of Epidemiology and Global Health*, 16, 93. <https://doi.org/10.1007/s44197-026-00581-4>
- Adebisi, Y.A., Bailey, N., Henderson, A., Dibben, C. and Pattaro, S. (2026b) 'Intellectual and physical disability, household living arrangements and COVID-19 outcomes in Scotland: protocol for a retrospective cohort study', *BMJ Open*, 16(6), e117443. <https://doi.org/10.1136/bmjopen-2026-117443>
- ADR Scotland (2021) 'Investigating socioeconomic, household and environmental risk factors for COVID-19 in Scotland'. Available at: <https://www.adrscotland.org/our-work/our-projects/project/investigatingsocioeconomic-household-and-environmental-risk-factors-for-covid-19-in-scotland-1/>
- Clark, D. and Dibben, C. (2022) 'Adding a Residential Dimension to the Scottish Population Spine - CHI-UPRN Residential Linkage (CURL)', *International Journal of Population Data Science*, 7(3), 1946. <https://doi.org/10.23889/ijpds.v7i3.1946>
- Gao, C., McGilchrist, M., Mumtaz, S. et al. (2022) 'A National Network of Safe Havens: Scottish Perspective', *Journal of Medical Internet Research*, 24(3), e31684. <https://doi.org/10.2196/31684>
- Harper, G., Firman, N., Wilk, M. et al. (2024) 'Determining households from patient addresses and unique property reference numbers in general practitioner electronic health records', *International Journal of Population Data Science*, 9(1), 2379. <https://doi.org/10.23889/ijpds.v9i1.2379>
- Johnson, R.D., Griffiths, L.J., Hollinghurst, J.P. et al. (2021) 'Deriving household composition using population-scale electronic health record data - a reproducible methodology', *PLoS ONE*, 16(3), e0248195. <https://doi.org/10.1371/journal.pone.0248195>
- Office for National Statistics (2023) Communal establishment estimation and adjustment: Census 2021. Newport: Office for National Statistics. <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/methodologies/communalestablishmentceestimationandadjustmentcensus2021>
- Pattaro, S., Bailey, N., Dibben, C. (2025) 'Occupational differences in COVID-19 hospital admission and mortality risks between women and men in Scotland: a population-based study using linked administrative data', *Occupational and Environmental Medicine*, 82(3), pp. 128–137. <https://doi.org/10.1136/oemed-2024-109562>
- Scottish Government (2020) Scottish Index of Multiple Deprivation 2020. Edinburgh: Scottish Government. <https://www.gov.scot/collections/scottish-index-of-multiple-deprivation-2020/>
- Scottish Government (n.d.) Coronavirus (COVID-19) data: definitions and sources. Edinburgh: Scottish Government. <https://www.gov.scot/publications/coronavirus-covid-19-data-definitions-and-sources/>

Acknowledgements

This work was supported by UK Research and Innovation (UKRI) through the Economic and Social Research Council (ESRC) as part of the Administrative Data Research Scotland programme (grant numbers ES/R005729/1, ES/S007407/1, ES/W010321/1 and UKRI3324) and by an ESRC PhD Studentship at the University of Glasgow through the Scottish Graduate School of Social Science (SGSSS) (ES/P000681/1). This research used data assets made available by the National Safe Haven as part of the Data and Connectivity National Core Study, led by Health Data Research UK in partnership with the Office for National Statistics and funded by UK Research and Innovation (grant numbers MC_PC_20029 and MC_PC_20058). We thank Public Health Scotland's electronic Data Research and Innovation Service (eDRIS) for data-linkage and access support, and National Records of Scotland and the Scottish Government for their roles as data controllers supporting approved access.

About ADR UK

ADR UK (Administrative Data Research UK), ESRC's flagship data-sharing partnership, enables secure access to linked UK public-sector data for researchers. From crime and justice to education, the economy, and health, this UK-wide partnership is transforming public sector data into research insights and policy evidence to improve lives.

Disclaimer

This work was produced using pseudonymised linked administrative data accessed through the Scottish National Safe Haven. The interpretation and analysis are the responsibility of the authors and do not imply endorsement by Public Health Scotland, National Records of Scotland, the Scottish Government, Ordnance Survey or other data owners/controllers.

The research extracts used for analysis may not exactly reproduce official published statistics, which can use different definitions, update cycles and statistical conventions.

Publication contact

Name: Yusuff Adebayo Adebisi
Email: y.adebisi.1@research.gla.ac.uk



<https://edin.ac/4qT11Lx>

