This is a peer-reviewed, accepted author manuscript of the following article: Mulholland, R. H., Vasileiou, E., Simpson, C. R., Robertson, C., Ritchie, L. D., Agrawal, U., Woolhouse, M., Murray, J. LK., Stagg, H. R., Docherty, A. B., McCowan, C., Wood, R., Stock, S. J., & Sheikh, A. (2021). Cohort profile: early pandemic evaluation and enhanced surveillance of COVID-19 (EAVE II) database. International Journal of Epidemiology, 50(4), 1064-1074. https://doi.org/10.1093/ije/dyab028

Cohort Profile: Early Pandemic Evaluation and Enhanced Surveillance of COVID-19 (EAVE II) database

Word count: 2796 (including titles and references)

Authors: Rachel H. Mulholland[1], Eleftheria Vasileiou*[1], Colin R. Simpson[1,2], Chris Robertson[3,4], Lewis D. Ritchie[5], Utkarsh Agrawal[6], Mark Woolhouse[1], Josephine L.K. Murray[4], Helen R. Stagg[1], Annemarie B. Docherty[1], Colin McCowan[6], Rachael Wood[1,7], Sarah J. Stock[8] and Aziz Sheikh[1]

Author affiliations: [1]Usher Institute, The University of Edinburgh, Edinburgh UK; [2]School of Health, Wellington Faculty of Health, Victoria University of Wellington, Wellington, New Zealand; [3]Department of Mathematics and Statistics, University of Strathclyde, Glasgow UK; [4]Public Health Scotland, Glasgow and Edinburgh, UK; [5]Centre of Academic Primary Care, University of Aberdeen, Aberdeen, UK; [6]School of Medicine, University of St. Andrews, St Andrews, UK; [7]Centre for Brain Sciences, Centre for Population Health, University of Edinburgh, Edinburgh, UK; [8]MRC Centre for Reproductive Health, University of Edinburgh, Edinburgh, UK

Corresponding author: Dr Eleftheria Vasileiou, Usher Institute, The University of Edinburgh, NINE Edinburgh BioQuarter, Edinburgh, UK, EH16 4UX, eleftheria.vasileiou@ed.ac.uk, +44 07732961139 (corresponding author)

Key Features

- The Early Pandemic Evaluation and Enhanced Surveillance of COVID-19 (EAVE II) database creates a national, real-time prospective cohort utilising Scotland's health data infrastructure, to describe the epidemiology of COVID-19, patterns of healthcare utilisation and outcomes, and insights into the effectiveness and safety of vaccines and treatments for COVID-19. As far as we are aware, EAVE II is the first national end-to-end clinical surveillance platform for COVID-19 predominantly using routinely available data.
- This study contains all 5.4 million individuals registered with a GP in Scotland from 23 February 2020, covering 98-99% of the Scottish population. These primary care records are linked to other data sources from out-of-hours, community, emergency and secondary care, in addition to data on registrations and mortality, laboratory testing, self-reported, and enhanced surveillance.
- These data will be updated throughout the course of the pandemic. Participants who die or permanently leave Scotland (and deregister from general practices) will drop-out of the cohort.
- Combining these rich data sources together provides a wealth of information on the natural history of the condition and patients journeys across Scotland's National Health Service (NHS).
- Data will be hosted in Scotland's National Safe Haven within the electronic Data Research and Innovation Service (eDRIS) of Public Health Scotland (PHS). Applicants must submit an enquiry to the corresponding author.

Why was the cohort set up?

In December 2019, a novel coronavirus COVID-19 emerged from Wuhan, China and was soon declared as a pandemic by the World Health Organization (WHO) on the 11 March 2020 [1]. The United Kingdom (UK) soon followed suit and implemented a national lockdown on the 23 March 2020. As of 09 December 2020, according to WHO, this highly infectious virus has infected more than 67 million people and led to over 1.5 million deaths across the world [2]. There is a growing body of evidence on the epidemiology of the condition, risk factors for poor outcomes and the effects of interventions [3-9].

The rapid generation of robust data is crucial to monitor, understand and mitigate the effects of COVID-19. The Early Pandemic Evaluation and Enhanced Surveillance of COVID-19 (EAVE II) database creates a national, real-time prospective cohort utilising Scotland's health data infrastructure, to describe the epidemiology of COVID-19 infection, patterns of healthcare utilisation and outcomes, and insights into the effectiveness of and safety of vaccines and treatments for COVID-19 [10].

This work builds on an established cohort for seasonal and pandemic influenza vaccine and anti-viral assessment in Scotland EAVE (Early Estimation of Vaccine and Anti-Viral Effectiveness) [11, 12]. EAVE is a dormant pandemic protocol that is part of the National Institute for Health Research (NIHR) Pandemic Preparedness Research Portfolio and a platform for previous studies on influenza vaccine and antiviral assessment [12-16].

Who is in the cohort?

This prospective baseline cohort study contains all 5.4 million individuals registered with a General Practice (GP) in Scotland from 23 February 2020 which, according to

the National Records of Scotland (NRS) 2019 mid-year estimates, covers around 98-99% of the Scottish population [10, 17]. A map of the baseline EAVE II cohort by National Health Service (NHS) Health Board shows that most of the cohort are based in the central belt of Scotland (Figure 1).

A summary of the baseline population by sex, age grouped (as of 23 February) deprivation using the Scottish Index of Multiple Deprivation (SIMD) [18] and Scottish Government Urban Rural Classification [19]. SIMD is a measure of deprivation built on seven domains and is unique to Scotland, with lower quintiles representing the most deprived areas [18].

These primary care records are linked to other data sources from out-of-hours, emergency and secondary care. There are additional linkages to other datasets such as laboratory testing data, registration and mortality data, self-reported data and enhanced surveillance data such as the COVID-19 Clinical Information Network (CO-CIN). This is done using the Community Health Index (CHI), the unique identifier provided by NHS Scotland. It is allocated to all residents in Scotland registered with a GP and for all patients that receive care in Scotland, even if they are non-Scottish residents [10]. Summary of these data sources are given in **Table 2** with a data flow diagram on how they are linked together in **Figure 2.**

This cohort therefore consists of specific groups of interest that are used in EAVE II sub-studies such as the COVID-19 in Pregnancy in Scotland (COPS) [37] and for investigating ethnic and social inequalities in COVID-19.

How often have they been followed up?

The baseline GP records will be updated on a biannual to 3-month basis, if possible. The first update in early 2021 will contain COVID-19 specific GP codes that were

created during the pandemic and were therefore missed in the initial extract. This will capture information on COVID-19 related appointments, vaccinations, therapies and vaccination induced adverse effects. Information on influenza will also be included to facilitate analyses on the effectiveness of and safety of COVID-19 specific and pre-existing vaccines, therapies and treatments. To facilitate UK-wide research, QCOVID groups will also be added to allow validation of the QCOVID 'living' risk prediction model on the Scottish population [38]. Information on shielded risk groups will also be included to assess the impact of COVID-19 on those most at-risk for severe illness where a 12-month self-isolation was recommended by the UK government on 23 March 2020 [39].

Regular updates on a number of linked datasets to the underlying GP data will be undertaken on a daily, weekly or monthly basis, as available and necessary (see **Table 3**). Those who have transferred GPs within Scotland will stay in the cohort. Participants who die or permanently leave Scotland (and deregister from general practices) will drop-out of the cohort. Characteristics of individuals lost to follow-up compared to those remaining in the cohort will also be provided in the study. Missing data will also be reported for each variable.

What has been measured?

Combining these rich data sources together provides a wealth of information on the natural history of the condition and patients journeys across Scotland's NHS. We provide a high-level summary of key available data in **Table 4.**

What has it found?

Permissions to link these datasets were received in May 2020 and the flow of linked data began in June 2020. The initial GP data extract contained the baseline cohort

and the EAVE II risk groups, which were based on the risk groups for seasonal influenza as research at the time of extract did not know exact risk groups for COVID-19. This includes comorbidities and household characteristics, for example an indicator of living in a care home. This EAVE II risk group dataset contained more individuals than the baseline cohort with over representation in certain populations. This is likely to have resulted in residents being registered at multiple GP practices, people who have left Scotland or visitors. To overcome this, weights were calculated by comparing the age and sex profile in the EAVE II cohort with the age and sex profile for the 2019 NRS mid-year population estimates in Scotland [17]. A summary of the number of EAVE II risk groups using these weights are shown in the supplementary material, along with the individual risk groups (Supplementary Table S1). The following analyses were performed using these weights.

Initial explorations showed that as age increased, lower levels of deprivation using SIMD quintiles slightly increased and the number of risk groups increased (Figure 3). These did not differ substantially between sexes (Figure 3).

Since the first follow-up of COVID-19 outcomes in 1 March to 10 November, there have been a total of 835,803 (15.4%) tested, 57,416 (1.1%) with a positive test (out of the total cohort), 9,847 (0.2%) hospitalised with COVID-19, 5,350 (0.1%) admitted to ICU or died with COVID-19 on the death certificate and 4,726 (0.1%) who have died with COVID-19 on the death certificate within the EAVE II cohort. The proportion of these outcomes split by age and sex for the same time period, show that more elderly residents have been tested with a resulting positive test (Figure 4). Elderly residents, particularly males, are also more represented in the more severe outcomes (Figure 4).

These age profiles were repeated for deprivation levels (using SIMD quintiles), the number of risk groups and the 20 most frequent individual risk groups within the EAVE II study (Supplementary material). This showed that there were higher proportions of positive tests and more severe outcomes in more deprived areas, residents belonging to multiple risk groups and those who had comorbidities (Supplementary material).

The map of the proportion of these outcomes by NHS Health Board demonstrated that despite high rates of testing in more rural areas in the northern and southern parts of Scotland, positive tests were low (Figure 5). The central belt had a higher proportion of positive tests out of the total baseline population and higher rates of more severe COVID-19 outcomes (Figure 5).

All relevant R code scripts for the summary tables and figures will be made available on the EAVE II GitHub page (https://github.com/EAVE-II). This will also contain a data dictionary for the entire EAVE cohort which will be updated when new updates and data linkages are made.

We are currently working on the development of a national risk prediction algorithm to identify risk factors for poor outcomes i.e. hospitalisation and death from COVID-19 [10], and the validation of the QCOVID-19 algorithm [38].

What are the main strengths and weaknesses?

The EAVE II cohort will be widely generalisable to the Scottish population as it contains all individual registered within GP practices in Scotland, with exception of homeless, itinerant or travelling groups, those in prison, those who are institutionalised due to mental health reasons and other reasons. Regularly updating and monitoring this cohort over a long period of time will also be quick and cost

effective, due to the underlying data sources being mainly routinely collected, quality assured and easily linkable using unique CHI numbers. This in turn means insights can be kept up to date with the rapidly evolving pandemic situation. The completeness and coverage both in terms of population, but also breadth of data is also a major strength.

The key limitations are the possibility of some selection biases because of excluded patients, although this is estimated to be under 2% of the Scottish population, and the risk of residual confounding in the context of analytic epidemiological studies. Considerable care will need to be taken when making inferences about the effectiveness of interventions because of non-randomised comparisons.

Can I get hold of the data? Where can I find out more?

Data can be accessed by contacting the corresponding author. For more information on the cohort refer to the published EAVE II protocol [10].

Ethics approval

We obtained approval from the National Research Ethics Service Committee, Southeast Scotland 02. The study findings will be presented at international conferences and published in peer-reviewed journals.

Funding

The original EAVE project was funded by the National Institute for Health Research Health Technology Assessment Programme (project number 13/34/14). EAVE II is funded by the Medical Research Council [MR/R008345/1] and supported by the Scottish Government. This work is supported by BREATHE - The Health Data Research Hub for Respiratory Health [MC_PC_19004]. BREATHE is funded through

the UK Research and Innovation Industrial Strategy Challenge Fund and delivered through Health Data Research UK.

Conflict of interest

Details on competing interests are included in the study's protocol (http://dx.doi.org/10.1136/bmjopen-2020-039097). Remaining co-authors (RHM, CM, UA, RW, ABD, SJS) do not report conflicts of interest. AD and SJS are also funded by Wellcome Trust Clinical Career Development. HRS is supported by the Medical Research Council (MR/R008345/1).

Contributorship

AS conceived the manuscript. RHM and EV led the writing of the manuscript and RHM and CR led the analysis. All co-authors reviewed and contributed to the writing of the manuscript.

References

1. World Health Organization. Archived: WHO Timeline – COVID-19. Available from: https://www.who.int/news-room/detail/27-04-2020-who-timeline---covid-19 (accessed June 26 2020)
2. World Health Organization. Coronavirus disease (COVID-19) outbreak. Available from: https://www.who.int/emergencies/diseases/novel-coronavirus-2019 (accessed June 26, 2020).
3. The Lancet. COVID-19 Resource Centre. Available from: https://www.thelancet.com/coronavirus (accessed 3 July 2020)
4. Oxford Academic. Access to OUP resources on COVID-19, other coronaviruses, and related topics. Available from: https://academic.oup.com/journals/pages/coronavirus (accessed 3 July 2020)
5. Springer Nature. COVID-19 and COVID-19. Available from: https://www.springernature.com/gp/researchers/campaigns/coronavirus?utm_source=bmc&utm_medium=referral&utm_content=null&utm_campaign=Homepage_Coronovirus_Teaser (accessed 3 July 2020)
6. The BMJ. BMJ's Coronavirus (Covid-19) Hub. Available from: https://www.bmj.com/coronavirus (accessed 3 July 2020)
7. PLOS. COVID-19 Pandemic (2019-20). Available from: https://collections.plos.org/covid-19 (accessed 3 July 2020)
8. JAMA Network. Coronavirus Disease 2019 (COVID-19). Available from: https://jamanetwork.com/journals/jama/pages/coronavirus-alert?utm_campaign=COVID19&utm_medium=banner&utm_source=house&utm_content=leader&utm_term=728x90 (accessed 3 July 2020)

9. The New England Journal of Medicine. Coronavirus (Covid-19). Available from: https://www.nejm.org/coronavirus (accessed 3 July 2020)
10. Simpson CR, Robertson C, Vasileiou E et al. Early pandemic evaluation and enhanced surveillance of COVID-19 (EAVE II): protocol for an observational study using linked Scottish national data. BMJ Open 2020;10:e039097.
11. Simpson CR, Beever D, Challen K et al. The UK's pandemic influenza research portfolio: a model for future research on emerging infections. Lancet Infect Dis 2019;19(8):e295-e300.
12. Simpson CR, Lone N, McMenamin J et al. Early estimation of pandemic influenza Antiviral and Vaccine Effectiveness (EAVE): use of a unique community and laboratory national data-linked cohort study. Health Technol Assess 2015;19(79).
13. Simpson CR, Ritchie LD, Robertson C et al. Vaccine effectiveness in pandemic influenza - primary care reporting (VIPER): an observational study to assess the effectiveness of the pandemic influenza A (H1N1)v vaccine. Health Technol Asses 2010;14:313-46.
14. Simpson CR, Ritchie LD, Robertson C et al. Effectiveness of H1N1 vaccine for the prevention of pandemic influenza in Scotland, UK: a retrospective observational cohort study. Lancet Infect Dis 2012;12:696-702.
15. Simpson CR, Lone N, Kavanagh K et al. Seasonal Influenza Vaccine Effectiveness (SIVE): an observational retrospective cohort study – exploitation of a unique community-based national-linked database to determine the effectiveness of the seasonal trivalent influenza vaccine. Health Services and Delivery Research 2013;1(10).

16. Simpson CR, Lone NI, Kavanagh K et al. Evaluating the effectiveness, impact and safety of live attenuated and seasonal inactivated influenza vaccination: protocol for the Seasonal Influenza Vaccination Effectiveness II (SIVE II) study. BMJ Open 2017;7:e014200.
17. National Records of Scotland (NRS). Mid-2019 Population Estimates Scotland. Available from: https://www.nrscotland.gov.uk/statistics-and-data/statistics/statistics-by-theme/population/population-estimates/mid-year-population-estimates/mid-2019
18. Scottish Government. Scottish Index of Multiple Deprivation 2020. Available from: https://www.gov.scot/collections/scottish-index-of-multiple-deprivation-2020/ (accessed June 11, 2020)
19. Scottish Government. Scottish Government Urban Rural Classification 2016. Available from: https://www.gov.scot/publications/scottish-government-urban-rural-classification-2016/pages/6/ (accessed June 11, 2020)
20. NHS National Services Scotland. National Data Catalogue. Prescribing Information System (PIS). Available from: https://www.ndc.scot.nhs.uk/National-Datasets/data.asp?SubID=9#Skip-Navigation (accessed June 10, 2020)
21. Public Health Scotland. Data and Intelligence. Primary Care Out of Hours Services. Emergency Care. Available from: https://www.isdscotland.org/Health-Topics/Emergency-Care/GP-Out-of-Hours-Services/ (accessed June 23, 2020)
22. ISD Scotland Home. Data Dictionary A-Z. SMR00 – Outpatient Attendance. Available from: https://www.ndc.scot.nhs.uk/Dictionary-A-Z/Definitions/index.asp?Search=S&ID=459&Title=SMR00%20-%20Outpatient%20Attendance (accessed June 23, 2020)

23. Scottish Ambulance Service. Available from: https://www.scottishambulance.com/TheService/Default.aspx (accessed June 10, 2020).
24. Docherty AB, Harrison EM, Green CA et al. Features of 20 133 UK patients in hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: prospective observational cohort study. BMJ 2020;369:m1985.
25. ISD Scotland Home. National Data Catalogue. Rapid Preliminary Inpatient Data (RAPID). Available from: https://www.ndc.scot.nhs.uk/National-Datasets/data.asp?SubID=37 (accessed June 23, 2020)
26. National Records of Scotland. Vital Events – Deaths – Background Information. Available from: https://www.nrscotland.gov.uk/statistics-and-data/statistics/statistics-by-theme/vital-events/deaths/deaths-background-information (accessed June 11, 2020)
27. NHS Inform. Test and Protect. Available from: https://www.nhsinform.scot/campaigns/test-and-protect (accessed 02 July 2020)
28. National Records of Scotland. Statistics and Data. Scotland's Census. Available from: https://www.nrscotland.gov.uk/statistics-and-data/census (accessed June 10, 2020).
29. Public Health Scotland. Search criteria for highest risk patients for shielding (v4.0); 11 May 2020. Available from: https://hpspubsrepo.blob.core.windows.net/hps-website/nss/3008/documents/1_covid-19-search-criteria-highest-risk-patients.pdf (accessed June 23, 2020)

30. Public Health Scotland. Data and Intelligence. Scottish Birth Record & Universal Newborn Hearing Screening. Available from: https://www.isdscotland.org/Products-and-Services/Scottish-Birth-Record/ (accessed June 10, 2020).
31. NHS Digital. Birth notification service. Available from: https://digital.nhs.uk/services/birth-notification-service (accessed June 10, 2020).
32. National records of Scotland. Vital Events – Births. Available from: https://www.nrscotland.gov.uk/statistics-and-data/statistics/statistics-by-theme/vital-events/births (accessed June 10, 2020).
33. National Records of Scotland. Stillbirths and Infant Deaths. Available from: https://www.nrscotland.gov.uk/statistics-and-data/statistics/statistics-by-theme/vital-events/deaths/deaths-background-information/stillbirths-and-infant-deaths (accessed June 10, 2020).
34. NHS National Services Scotland. National Data Catalogue. Notification of Abortion Statistics (AAS). Available from: https://www.ndc.scot.nhs.uk/National-Datasets/data.asp?SubID=64 (accessed June 10, 2020).
35. Information Services Division (Public Health Scotland). National Data Catalogue. Child Health Systems Programme – School (CHSP School). Available from: https://www.ndc.scot.nhs.uk/National-Datasets/data.asp?ID=4&SubID=11 (accessed June 11, 2020)
36. Information Services Division (Public Health Scotland). National Data Catalogue. Scottish Immunisation Recall System (SIRS). Available from: https://www.ndc.scot.nhs.uk/National-Datasets/data.asp?ID=4&SubID=12 (accessed June 11, 2020)

37. Stock S.J, McAllister D, Vasileiou E, et al. COVID-19 in Pregnancy in Scotland (COPS): protocol for an observational study using linked Scottish national data. BMJ Open. 2020. 10:e042813. Available from: doi: :10.1136/bmjopen-2020-042813
38. Clift A.K, Keogh R.H, Williamson E, et al. Living risk prediction algorithm (QCOVID) for risk of hospital admission and mortality from coronavirus 19 in adults: national derivation and validation cohort study. BMJ. 2020. 371:m3731. Available form: doi: 10.1136/bmj.m3731
39. Scottish Government. Coronavirus (COVID-19): shielding advice and support. Available from: https://www.gov.scot/publications/covid-shielding/.
40. Air Quality in Scotland. Available from: http://www.scottishairquality.scot/ (accessed June 23, 2020)

Tables

Table 1: Baseline characteristics of the population in the Early Pandemic Evaluation and Enhanced Surveillance of COVID-19 (EAVE II) cohort study (n=5,431,034). Update: 23 February 2020.

Characteristics	Total number of individuals (% of total)
Sex	
Female	2733477 (50.3)
Male	2697557 (49.7)
Age group (years)	
0-4	245423 (4.5)
5-14	574389 (10.6)
15-24	624070 (11.5)
25-44	1479594 (27.2)
45-64	1503617 (27.7)
65-74	563605 (10.4)
75-84	323812 (6.0)
≥85	116524 (2.1)
Deprivation quintile [a]	
1 - Most deprived	1100521 (20.3)
2	1074842 (19.8)
3	1050369 (19.3)
4	1079282 (19.9)
5 - Least deprived	1080775 (19.9)
Urban/rural score [b]	
1 - Large urban areas	1920932 (35.4)
2 - Other urban areas	1959281 (36.1)
3 - Accessible small towns	501557 (9.2)
4 - Remote small towns	257264 (4.7)
5 - Accessible rural	486665 (9.0)
6 - Remote rural	260090 (4.8)

Abbreviations: NHS: National Health Service
Missing values (%):
a Deprivation score not available for 45245 (0.8%) individuals
b Urban/rural score not available for 45245 (0.8%) individuals
c NHS Health Board not available for 45245 (0.8%) individuals
Note: a Deprivation score calculated via the Scottish Index of Multiple Deprivation (SIMD)

Table 2: Details of data sources within the different settings for the Early Pandemic Evaluation and Enhanced Surveillance of COVID-19 (EAVE II) cohort.

Setting	Data Sources	Description
Primary Care	General Practice (GP) data [a]	Data from all patients registered in GP's. GP data (demographic, consultation data – categorised into risk groups, prescribing and categorised measurements) will be extracted using the Enhanced Services Contract Reporting Options (ESCRO) system by the third trusted party Albasoft Ltd [10].
	Prescribing Information System (PIS) [a]	Information on all prescribing relating to all prescriptions dispensed in the community. Prescriptions written in hospitals which are dispensed in the community are also included [20].
	Out-of-hours (OOH) [p]	Data on the services a patient receives for primary care when their registered GP practice is closed [21].
	Scottish Morbidity Record 00 (SMR00) [a]	Relates to all outpatients (new and follow-up) in specialties other than Accident & Emergency (A&E), and Genito-Urinary Medicine [22]
Telephone consultation	National Health Service (NHS) 24 [a]	Deliver telephone and online services across Scotland for initial assessments, which are then passed on to the appropriate services if required. [10]
	COVID-19 Community Hubs and Assessment Centres [p]	A network established by NHS Health Boards in Scotland to provide a direct and rapid route of people with COVID-19 [10]. Data from these centres will derive from National Health Service (NHS) 24 and the COVID-19 Enhanced Surveillance dataset [10].
Secondary Care	Scottish Morbidity Record (SMR) including: ➢ SMR01 [a] ➢ SMR02 [a]	SMR01: Episode-based patient record for all inpatients and day cases discharged from non-obstetric and non-psychiatric specialties in Scotland. This includes Accident & Emergency (A&E) attendances [10].
		SMR02: Episode-based patient record for all inpatients and day cases discharged from obstetric specialties in Scotland [10].
	Scottish Hospital Electronic Prescribing and Medicines Administration (HEMPA) system [a]	Data on prescription and administration of medicines for inpatients from a subgroup of hospitals with HEPMA systems [10].
	Scottish Ambulance Service (SAS) [a]	Scottish database for all patients requiring emergency ambulance services or need support to reach their healthcare appointments due to their medical and mobility needs [23].
	Scottish Intensive Care Society Audit Group (SICSAG) [a]	Scottish database for adult patients admitted to all general Intensive Care Units (ICU) and combined ICU/High Dependency Units (HDU) [10].

Setting	Data Sources	Description
	COVID19 Clinical Information Network / International Severe Acute Respiratory and emerging Infection Consortium (CO-CIN / ISARIC) [p]	Data of the clinical characteristics of patients admitted to hospital with COVID-19 infection in Scotland recruited to CO-CIN/ISARIC [24]. As of 22 June 2020 this comprised of 65% of the hospitalised patients in Scotland.
	Rapid Preliminary Inpatient Data (RAPID) [p]	Contains hospital inpatient admission data, which has been used to predict emergency admissions and bed occupancy [25].
Mortality	National Records of Scotland (NRS) deaths [a]	Data on Scottish death certificates and the cause of death [26].
Laboratory and serology data	Electronic Communication of Surveillance in Scotland (ECOSS) [a]	Surveillance data on laboratory results from microorganisms, infections and microbial intoxications. Contains all reverse transcriptase polymerase chain reaction (RT-PCR) tests carried out in Scotland [10].
	Serology data [a]	All serology data will be provided by the 'Seroprevalence' work carried out and commissioned by the COVID-19 Enhanced Surveillance cell of Public Health Scotland (PHS) [10].
	Genome sequencing data [p]	Positive laboratory RT-PCR swab samples for COVID-19 will also be sent to national sequencing centres where 500 COVID-19 genome sequences will be performed [10].
Self-reported data	Test and Protect data [p]	A service which identifies positive cases of COVID-19 and who they have had close, recent contact with. [27].
	Surveys [p]	Surveys on how people have been affected by COVID-19 in Scotland.
	Census 2011 data [p]	Residents in Scotland are asked to fill in a census questionnaire every 10 years and provide information on their demographic (e.g. ethnicity), socioeconomic, health and other circumstances. NRS will provide data from the latest Scottish Census in 2011 [28].
Derived data	COVID-19 Shielding patient list [p]	Uses a combination of primary and secondary care held in Public Health Scotland to derive groups considered to be at high risk if they contract COVID-19 [29].
Births and pregnancy related data	Scottish Birth Record (SBR) [p]	The SBR is a web-based system developed on the NHSNet to ensure that every baby born in Scotland will have one record, which will act as the foundation for future information collection. The system has been implemented to varying degrees in all Scottish hospitals providing midwifery and/or neonatal care [30].
	NHS live birth notifications [p]	Notification of live births from NHS Board maternity units to child health administration departments [31].

Setting	Data Sources	Description
	NRS births [p]	Record of statutory registration of a live birth (live born baby at any gestation) [32].
	NRS statutory stillbirth registrations [p]	Record of statutory registration of a stillbirth (baby born at ≥24w showing no signs of life) [33].
	NHS antenatal care notifications [p]	Public Health Scotland (PHS) has developed a new national data return as part of the response to the COVID-19 pandemic providing information on women booking for antenatal care with NHS maternity services: for identification of women with ongoing pregnancies in near real-time.
	Abortion Act Scotland (AAS) Notifications [p]	Record of statutory notification of all terminations of pregnancy in Scotland [34].
Vaccine treatment	Child Health Systems Programme – School (CHSP-S) [a]	Facilitates the call/recall of both primary and secondary school pupils for screening, review and immunisation [35].
	Scottish Immunisation Recall System (SIRS) [p]	Data on recorded immunisation in children when scheduled for a vaccination, including children at pre-school age [36].

a Data sources approved as of May 2020
p Data sources pursuing

Table 3: Details on frequency of data linkages

Daily or weekly linkages	Weekly or monthly linkages	Monthly linkages
ECOSS	SMR01	SBR
NHS 24	SMR02	NRS births
SAS	PIS	NRS stillbirths
Serology data	NRS deaths	NHS antenatal care
SICSAG	CO-CIN	AAS
RAPID	OOH	

Abbreviations: AAS: Abortion Act Scotland; CO-CIN: COVID-19 Clinical Information Network; ECOSS: Electronic Communication of Surveillance in Scotland; NHS: National Health Service; NRS: National Records of Scotland; OOH: Out-of-hours; PIS: Prescribing Information System; RAPID: Rapid Preliminary Inpatient Data; SAS: Scottish Ambulance Service; SBR: Scottish Birth Record; SICSAG: Scottish Intensive Care Society Audit Group; SMR; Scottish Morbidity Record

Table 4: Variables captured and their relevant data sources

Category	Variable Group	Specific variables	Source(s)
COVID-19 Outcomes	Testing	Tested; date of positive/negative test; test results; type of test; antibody tests (if available)	COVID-19 Community Hubs and Assessment Centres; ECOSS; serology data; genome sequencing data; Test and Protect data
	Severity	Severity; symptoms; Hospital admission; admitted to ICU; treatment in ICU	NHS 24; COVID-19 Community Hubs and Assessment Centres; SMR; SAS; SICSAG; CO-CIN; RAPID; Test and Protect data; Surveys; SMR01; SMR00
	Mortality	Death; Cause of death	NRS Deaths; SMR
	Treatment	Type of vaccination; date of vaccination	GP data; ECOSS; CHSP-S; SIRS
Potential risk factors	Sociodemographic	Age; Sex; Ethnicity; Country of birth; BMI; Smoking; Employment status; Occupation; Country of Birth; Religion; Tenure	GP data; 2011 Census
	Geographical	Data zone; Socioeconomic status (SES) through Scottish Multiple Deprivation Index (SIMD) [18]; Urban Rural Index [19]; pollution exposure [40]; population density	GP data (use postcode to link to relevant datasets)
	Clinical	Comorbidities including chronic respiratory disease (with chronic obstructive pulmonary disease and asthma as subsets); chronic heart disease; chronic liver disease; chronic kidney disease; chronic neurological disease; diabetes type 1 and 2; conditions or medications causing impaired immune function; pregnancy; asplenia or dysfunction of spleen; obesity; hypertension (subsets controlled/uncontrolled hypertension); tuberculosis; multimorbidity; Charlson Comorbidity Index	GP data; SMR
	Medications	Prescription drugs including asthma (including GINA management steps & oral steroids) and COPD related prescriptions; regular inhalers; COVID/pandemic acute therapies and chronic therapy for long-term sequelae; Statins; Rhinitis therapy; immunotherapy; Diabetes therapy; CVD therapy; antihypertensives; Antibiotics; NSAID; Cox2; Paracetamol; Antiviral prescriptions; Drugs for previous primary care consultations; Polypharmacy; High-risk prescribing	GP data; PIS; HEPMA

Category	Variable Group	Specific variables	Source(s)
	Pregnancy and babies	Pregnancy indicator; Miscarriage, Ectopic pregnancy, Pregnancy termination (incl. date, gestation, grounds); Stillbirth (incl. date, gestation, cause of death); Live birth (incl. date, gestation, sex of baby), Congenital anomaly flag; neonatal outcomes following maternal infection	GP data; SMR; SICSAG; CO-CIN; SBR; NHS live birth; NRS births; NRS still births; NHS antenatal care; AAS

Abbreviations: AAS: Abortion Act Scotland; CO-CIN: COVID-19 Clinical Information Network; ECOSS: Electronic Communication of Surveillance in Scotland; ICU: Invasive Care Unit; NHS: National Health Service; NRS: National Records of Scotland; OOH: Out-of-hours; PIS: Prescribing Information System; RAPID: Rapid Preliminary Inpatient Data; SAS: Scottish Ambulance Service; SBR: Scottish Birth Record; SICSAG: Scottish Intensive Care Society Audit Group; SMR; Scottish Morbidity Record

Figures

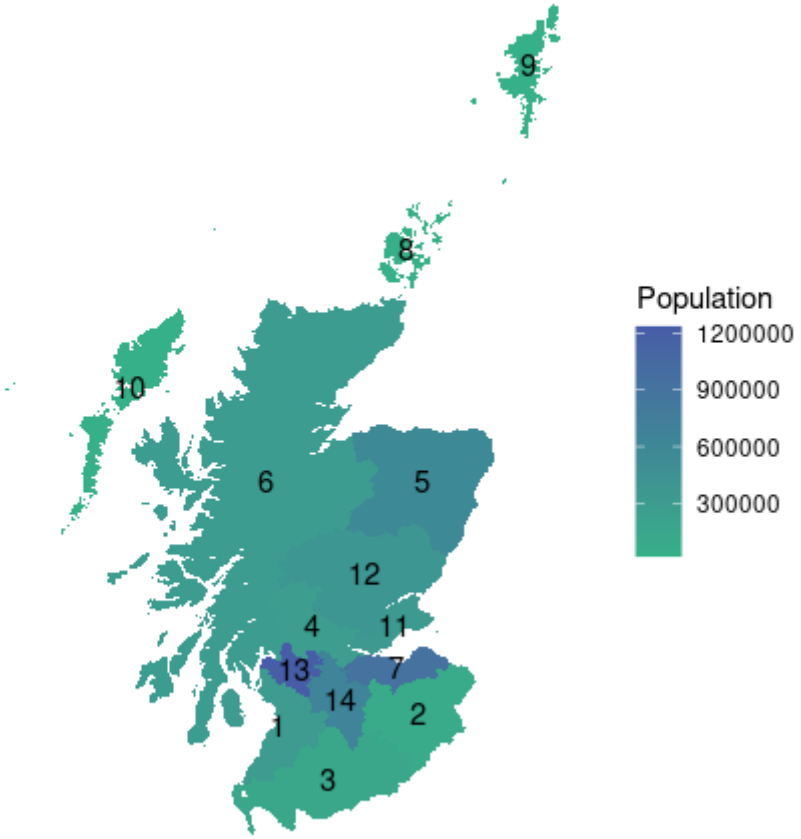


Figure 1: Baseline Early Pandemic Evaluation and Enhanced Surveillance of COVID-19 (EAVE II) cohort population by National Health Service (NHS) Health Board. (1 = NHS Ayrshire and Arran; 2 = NHS Borders; 3 = NHS Dumfries and Galloway; 4 = NHS Forth Valley; 5 = NHS Grampian; 6 = NHS Highland; 7 = NHS Lothian; 8 = NHS Orkney; 9 = NHS Shetland; 10 = NHS Western Isles; 11 = NHS Fife; 12 = NHS Tayside; 13 = NHS Greater Glasgow and Clyde; 14 = NHS Lanarkshire – ordered by Health board code).

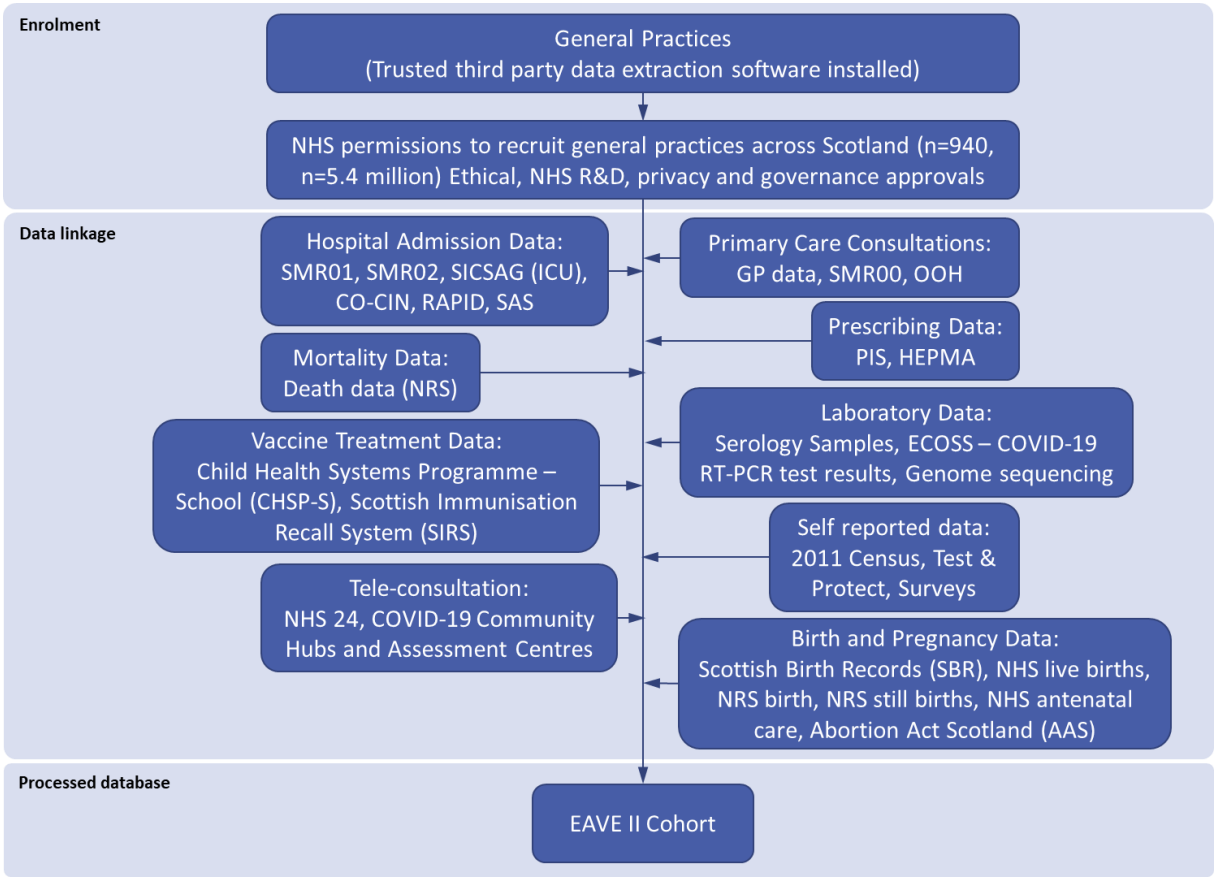


Figure 2: Flow diagram for EAVE II

Primary care consultations (SMR; Scottish Morbidity Record; OOH: Out-of-hours); Hospital Admission (SICSAG: Scottish Intensive Care Society Audit Group; CO-CIN: COVID19 Clinical Information Network; RAPID: Rapid Preliminary Inpatient Data; SAS: Scottish Ambulance Service); Prescribing (PIS: Prescribing Information System; HEMPA: Hospital Electronic Prescribing and Medicines Administration); Laboratory (ECOSS: Electronic Communication of Surveillance in Scotland; RT-PCR: Reverse transcription polymerase chain reaction); Vaccine Treatment (CHSP-S: Child Health Systems Programme – School; SIRS: Scottish Immunisation Recall System); Birth and Pregnancy (SBR: Scottish Birth Record; NRS: National Records of Scotland; AAS: Abortion Act Scotland);

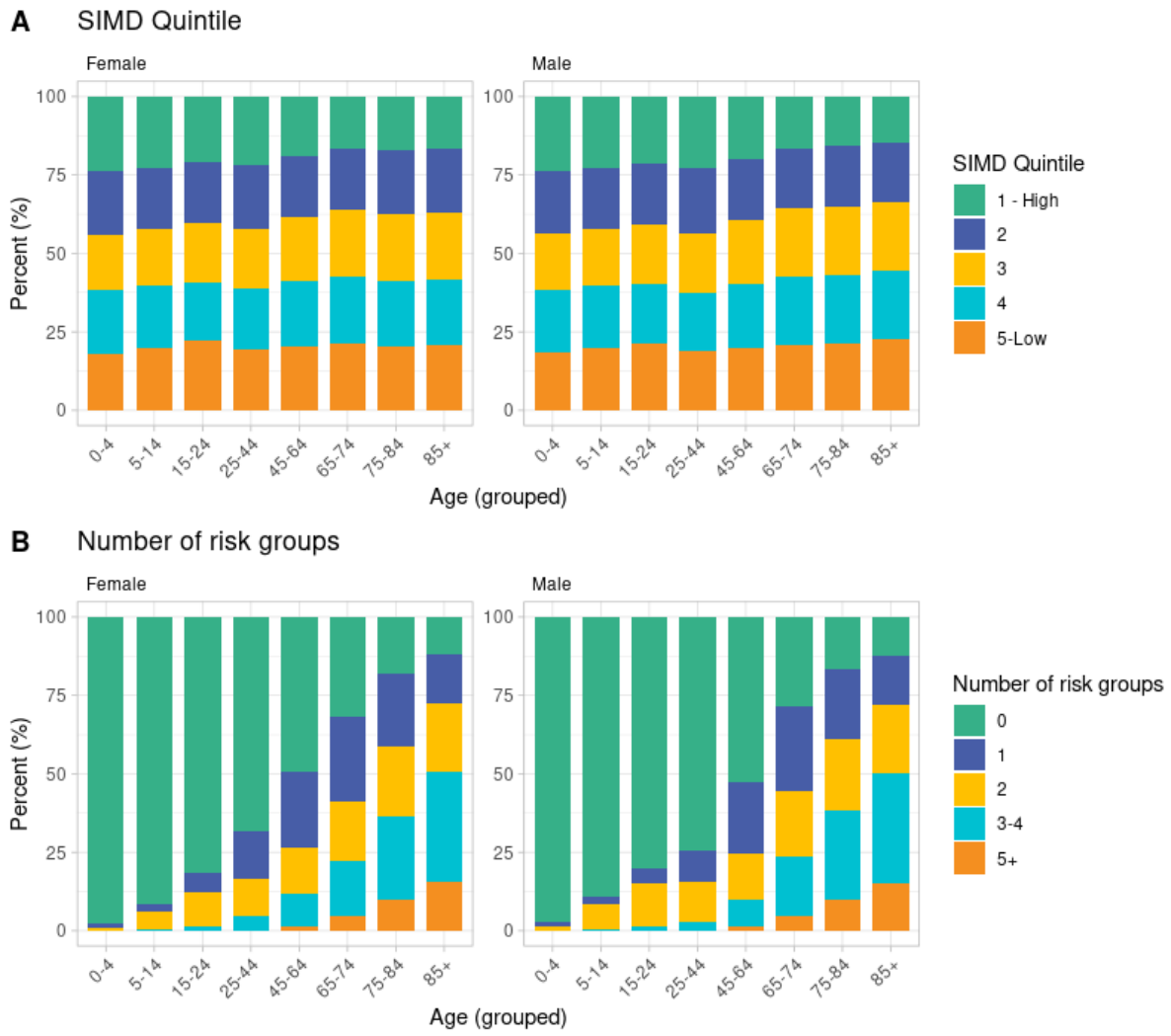


Figure 3: Proportion of baseline cohort in (A) deprivation quintiles (SIMD) and (B) number of risk groups, split by age grouped and sex.

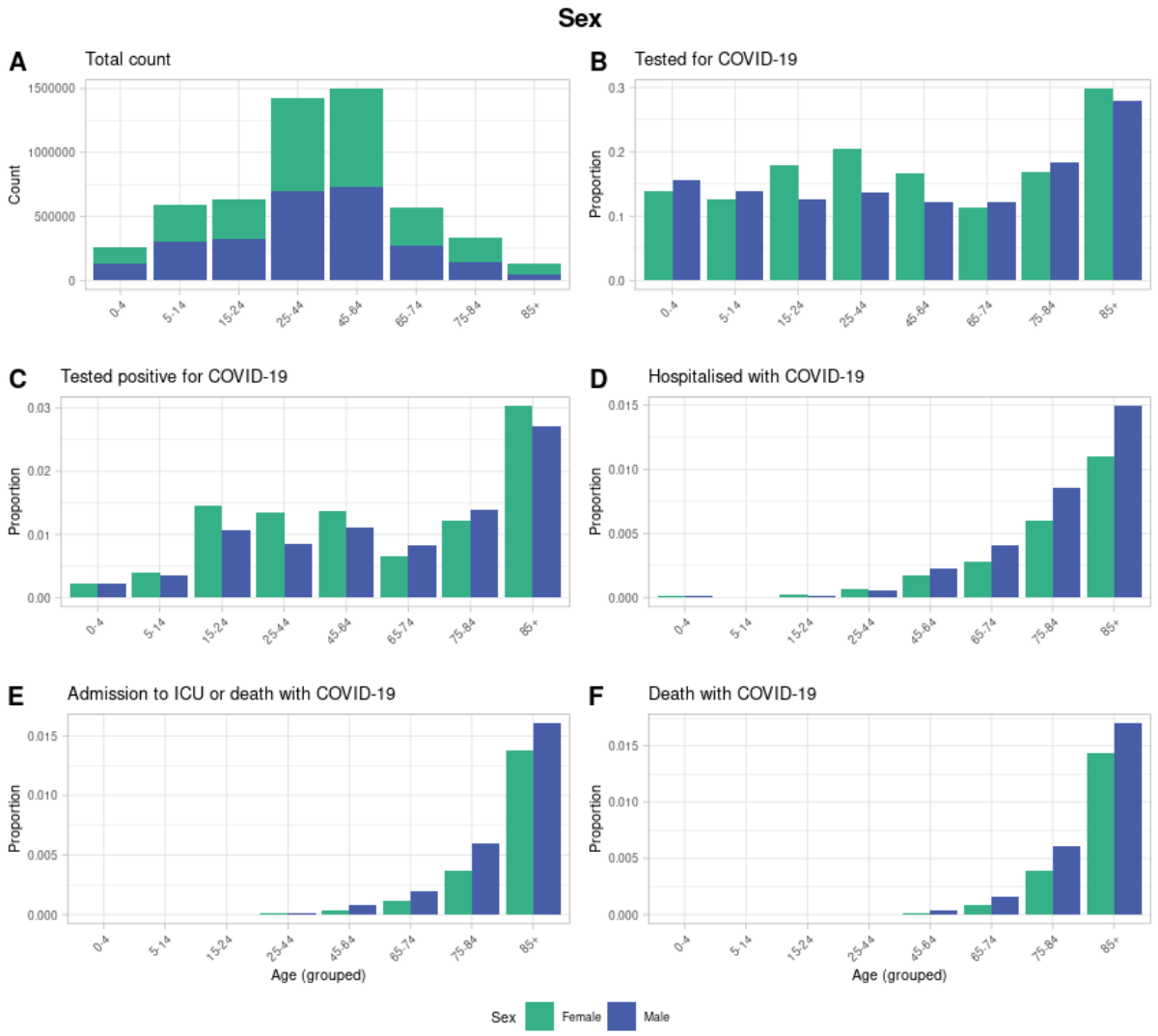


Figure 4: The proportion of the age groups split by sex that have been (A) tested for COVID-19, (B) tested positive for COVID-19, (C) hospitalised with COVID-19, (D) admitted to ICU or died with COVID-19 and (E) died with COVID-19 between 1 March and 10 November. Note: (B) is a proportion out of the total baseline cohort.

Figure 5: Map of Scotland by NHS Health Board for the proportion of the baseline cohort (A) tested, (B) tested positive, (C) hospitalised with COVID-19, (D) admitted to ICU or died with COVID-19 and (E) died between 1 March and 10 November. Note: proportions were out of the total Health Board cohort population and (B) is a proportion out of the total baseline cohort.

Cohort Profile: Pocket Profile

Title: Cohort Profile: Early Pandemic Evaluation and Enhanced Surveillance of COVID-19 (EAVE II) database

Authors: Rachel H. Mulholland[1], Eleftheria Vasileiou[1], Colin R. Simpson[1,2], Chris Robertson[3,4], Lewis D. Ritchie[5], Utkarsh Agrawal[6], Mark Woolhouse[1], Josephine L.K. Murray[4], Helen R. Stagg[1], Annemarie Docherty[1], et al (the complete author list is available in the full version of the profile online)

Cite this as: The full version of this profile is available at *IJE* online and should be used when citing this profile.

Corresponding author: Eleftheria Vasileiou, Usher Institute, The University of Edinburgh, NINE Edinburgh BioQuarter, Edinburgh, UK, EH16 4UX, eleftheria.vasileiou@ed.ac.uk, +44 07732961139

Keywords: COVID-19, coronavirus, pandemic, surveillance, treatment, vaccine

Cohort purpose: Rapid evidence generation of new insights to understand the epidemiology of COVID-19 is essential. Scotland's national health data infrastructure is being utilised to provide a complete population level real-time data platform to understand the epidemiology and healthcare impact of COVID-19, and inform the effectiveness and safety of existing and new (once available) treatments and vaccines.

Cohort basics: The baseline population consists of all 5.4 million individuals registered in 940 general practices (GP) across Scotland from 23 February 2020 linked to other demographic, healthcare and mortality datasets.

Follow-up and attrition: These data will be tracked throughout the course of the pandemic. Those who have transferred GPs within Scotland will stay in the cohort. Participants who die or permanently leave Scotland (and deregister from general practices) will drop-out of the cohort.

Design and measures: This national prospective cohort study links individual patient data across primary care, out-of-hours, accident and emergency, testing, hospitalisation and mortality data, as well as linkages to other demographic and clinical datasets (**Figure 1**). This will enable monitoring of the incidence and prevalence of clinical and laboratory confirmed COVID-19, an appreciation of risk factors for disease and poor outcomes, patterns of healthcare utilisation, and insights into the real-world effectiveness of treatments and vaccines as these become available.

Unique features: This is, as far as we are aware, the first national end-to-end clinical surveillance platform for COVID-19 predominantly generated using routinely available data.

Reasons to be cautious: There is the potential for selection bias as some vulnerable groups (e.g. homeless and asylum seekers) may not be registered with a

GP and are therefore at risk of being excluded from the cohort. There is the possibility of residual confounding when assessing risk factors.

Collaboration and data access: Data will be hosted in Scotland's National Safe Haven within the electronic Data Research and Innovation Service (eDRIS) of Public Health Scotland (PHS). Applicants must submit an enquiry to the corresponding author.

Funding and competing interests: The original EAVE project was funded by the National Institute for Health Research Health Technology Assessment Programme (project number 13/34/14). EAVE II is funded by the Medical Research Council [MR/R008345/1] and supported by the Scottish Government. This work is supported by BREATHE - The Health Data Research Hub for Respiratory Health [MC_PC_19004]. BREATHE is funded through the UK Research and Innovation Industrial Strategy Challenge Fund and delivered through Health Data Research UK Details on competing interests are included in the study's protocol (http://dx.doi.org/10.1136/bmjopen-2020-039097). Remaining co-authors (RHM, CM, UA, RW, AD, SJS) do not report conflicts of interest. AD and SJS are also funded by Wellcome Trust Clinical Career Development. HRS is supported by the Medical Research Council (MR/R008345/1).

Author affiliations: [1]Usher Institute, The University of Edinburgh, Edinburgh UK; [2]Wellington School of Health, Faculty of Health, Victoria University of Wellington, Wellington, New Zealand; [3]Department of Mathematics and Statistics, University of Strathclyde, Glasgow UK; [4]Public Health Scotland, Glasgow and Edinburgh, UK; [5]Centre of Academic Primary Care, University of Aberdeen, Aberdeen, UK; [6]School of Medicine, University of St. Andrews, St Andrews, UK; [7]Centre for Brain Sciences, Centre for Population Health, University of Edinburgh, Edinburgh, UK; [8]MRC Centre for Reproductive Health, University of Edinburgh, Edinburgh, UK

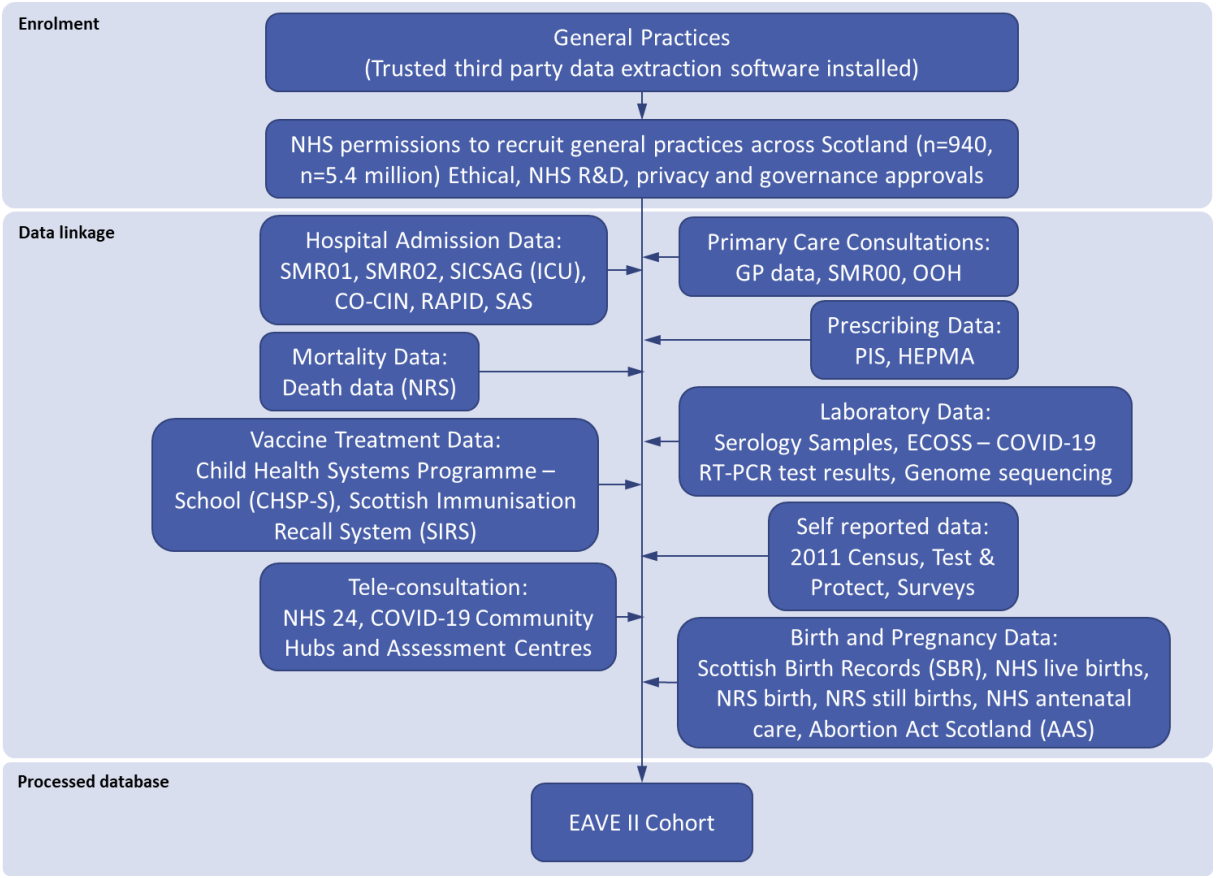


Figure 1: Flow diagram for EAVE II

Primary care consultations (SMR; Scottish Morbidity Record; OOH: Out-of-hours); Hospital Admission (SICSAG: Scottish Intensive Care Society Audit Group; CO-CIN: COVID19 Clinical Information Network; RAPID: Rapid Preliminary Inpatient Data; SAS: Scottish Ambulance Service); Prescribing (PIS: Prescribing Information System; HEMPA: Hospital Electronic Prescribing and Medicines Administration); Laboratory (ECOSS: Electronic Communication of Surveillance in Scotland; RT-PCR: Reverse transcription polymerase chain reaction); Vaccine Treatment (CHSP-S: Child Health Systems Programme – School; SIRS: Scottish Immunisation Recall System); Birth and Pregnancy (SBR: Scottish Birth Record; NRS: National Records of Scotland; AAS: Abortion Act Scotland);

Cohort Profile: Early Pandemic Evaluation and Enhanced Surveillance of COVID-19 (EAVE II) database

Supplementary Material

Abbreviations

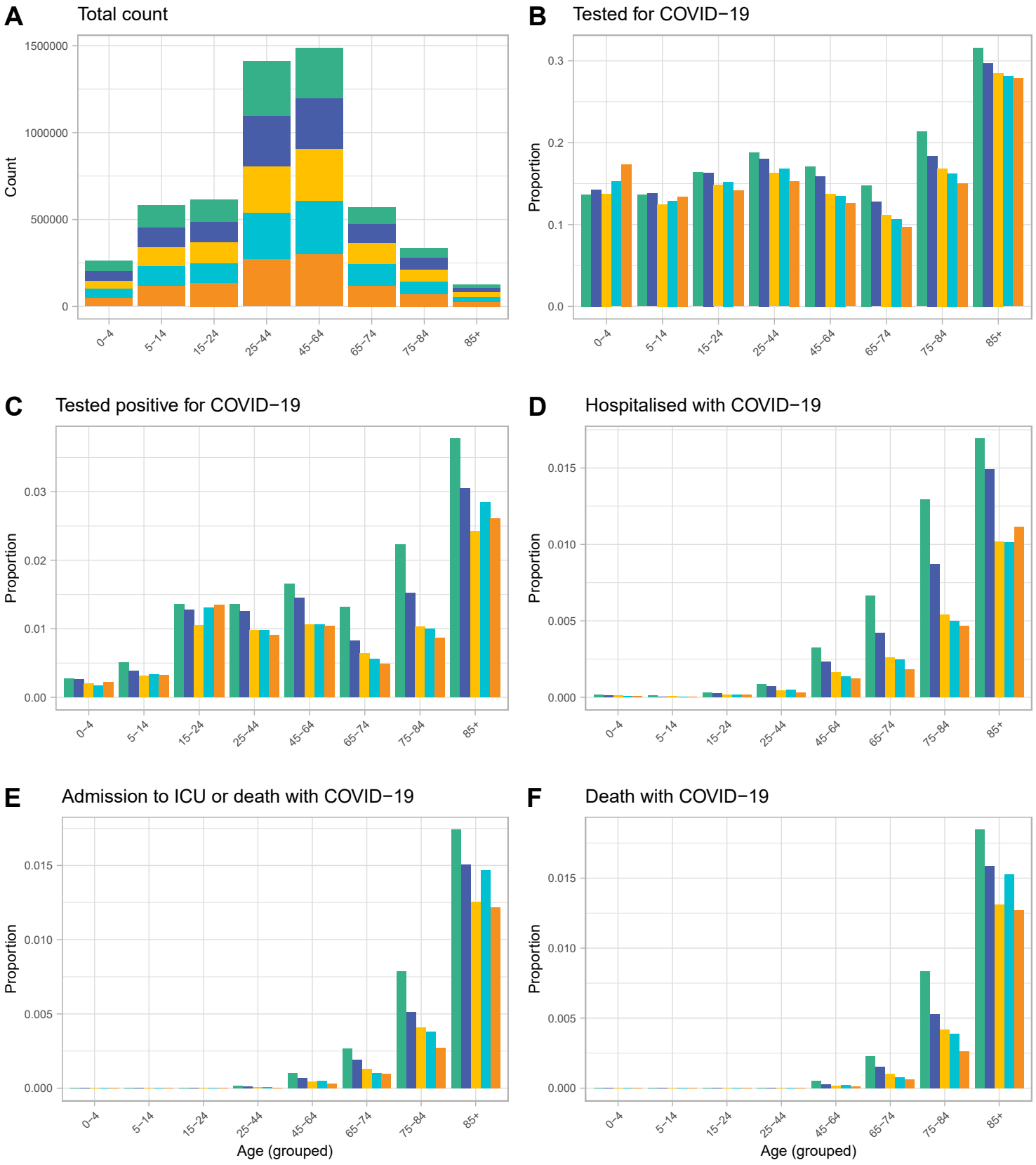
AAS: Abortion Act Scotland; CHI: Community Health Index; CO-CIN: COVID-19 Clinical Information Network; COPS: COVID-19 in Pregnancy in Scotland; EAVE II: Early Pandemic Evaluation and Enhanced Surveillance of COVID-19; ECOSS: Electronic Communication of Surveillance in Scotland; eDRIS: electronic Data Research and Innovation Service; GP: General Practice; ICU: Invasive Care Unit; NHS: National Health Service; NIHR: National Institute for Health Research; NRS: National Records of Scotland; OOH: Out-of-hours; PHS: Public Health Scotland; PIS: Prescribing Information System; RAPID: Rapid Preliminary Inpatient Data; SAS: Scottish Ambulance Service; SBR: Scottish Birth Record; SICSAG: Scottish Intensive Care Society Audit Group; SIMD: Scottish Index of Multiple Deprivation; SMR; Scottish Morbidity Record; UK: United Kingdom; WHO: World Health Organization

Table S1: Baseline risk groups of EAVE II population (N=5,439,030). Update: 23 February 2020

EAVE II Risk Groups	Total number of individuals (% of total)
Number of risk groups	
0	3325499 (61.1)
1	831117 (15.3)
2	735817 (13.5)
3-4	445461 (8.2)
5+	101135 (1.9)
Asthma	613156 (11.3)
Care home	30656 (0.6)
Chronic heart disease	326513 (6)
Chronic kidney disease	177826 (3.3)
Chronic liver disease	90924 (1.7)
Chronic pancreatitis	382 (<0)
Chronic respiratory disease	768781 (14.1)
Dementia	43805 (0.8)
Depression	545715 (10)
Diabetes	298544 (5.5)
Haematological malignancies	20276 (0.4)
Home oxygen	5806 (0.1)
Hypertension	763878 (14)
Immunosuppression	32257 (0.6)
MS and degenerative disease	77273 (1.4)
Myoneural disorders	3386 (0.1)
Nutritional deficiencies	112974 (2.1)
Other suspected malignancy	11844 (0.2)
Peripheral vascular disease	91529 (1.7)
Rheumatological disorders	6420 (0.1)
Social care	7962 (0.1)
Enlarged Spleen/Anaemia	65009 (1.2)
Stroke/TIA	134054 (2.5)
Transplantation	1165 (<0)
Ulcer disease	111755 (2.1)

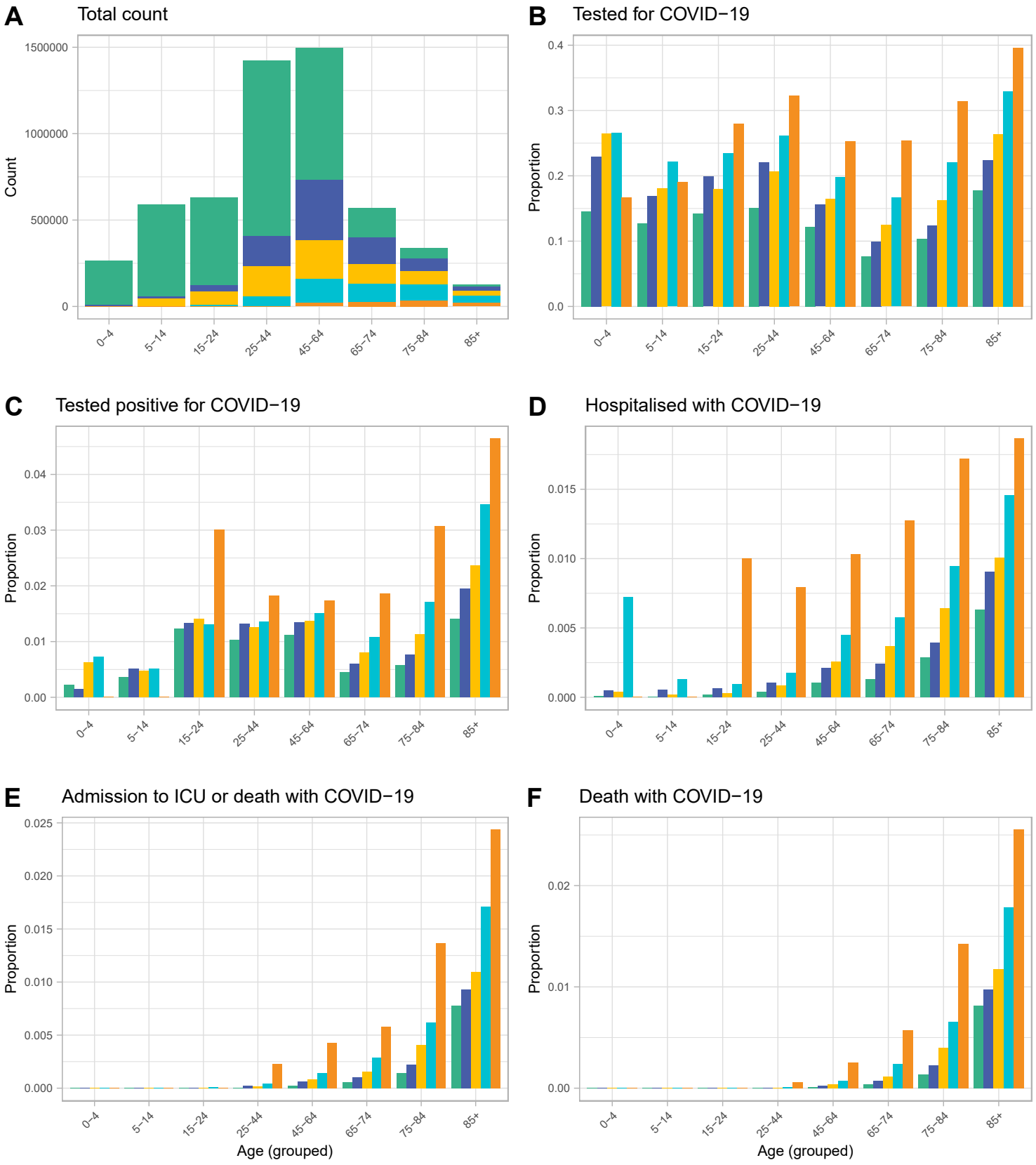
Abbreviations: MS: Multiple sclerosis; TIA: Transient Ischaemic Attack

SIMD Quintiles



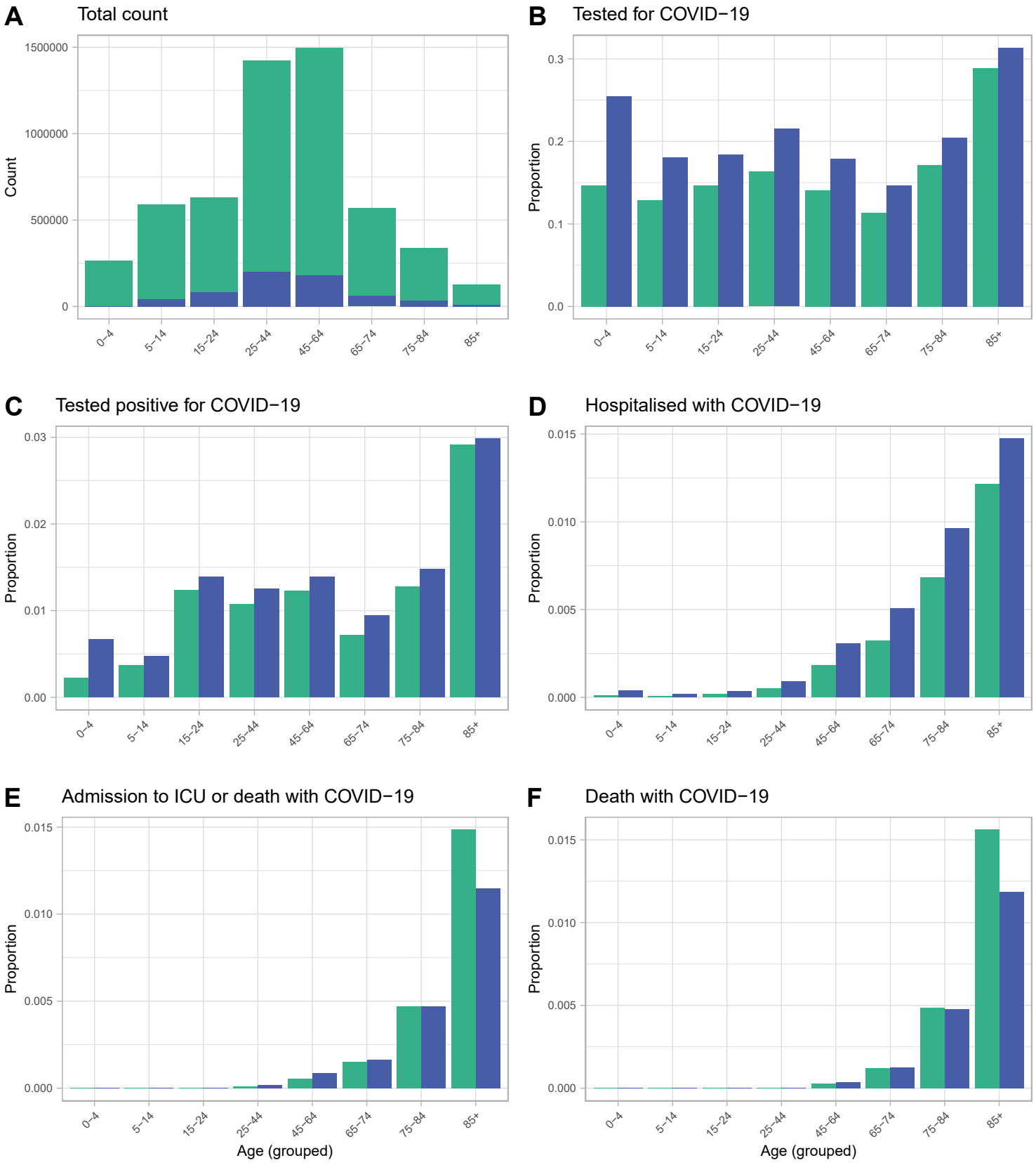
Time frame: 01 Mar to 10 Nov 2020

Number of risk groups



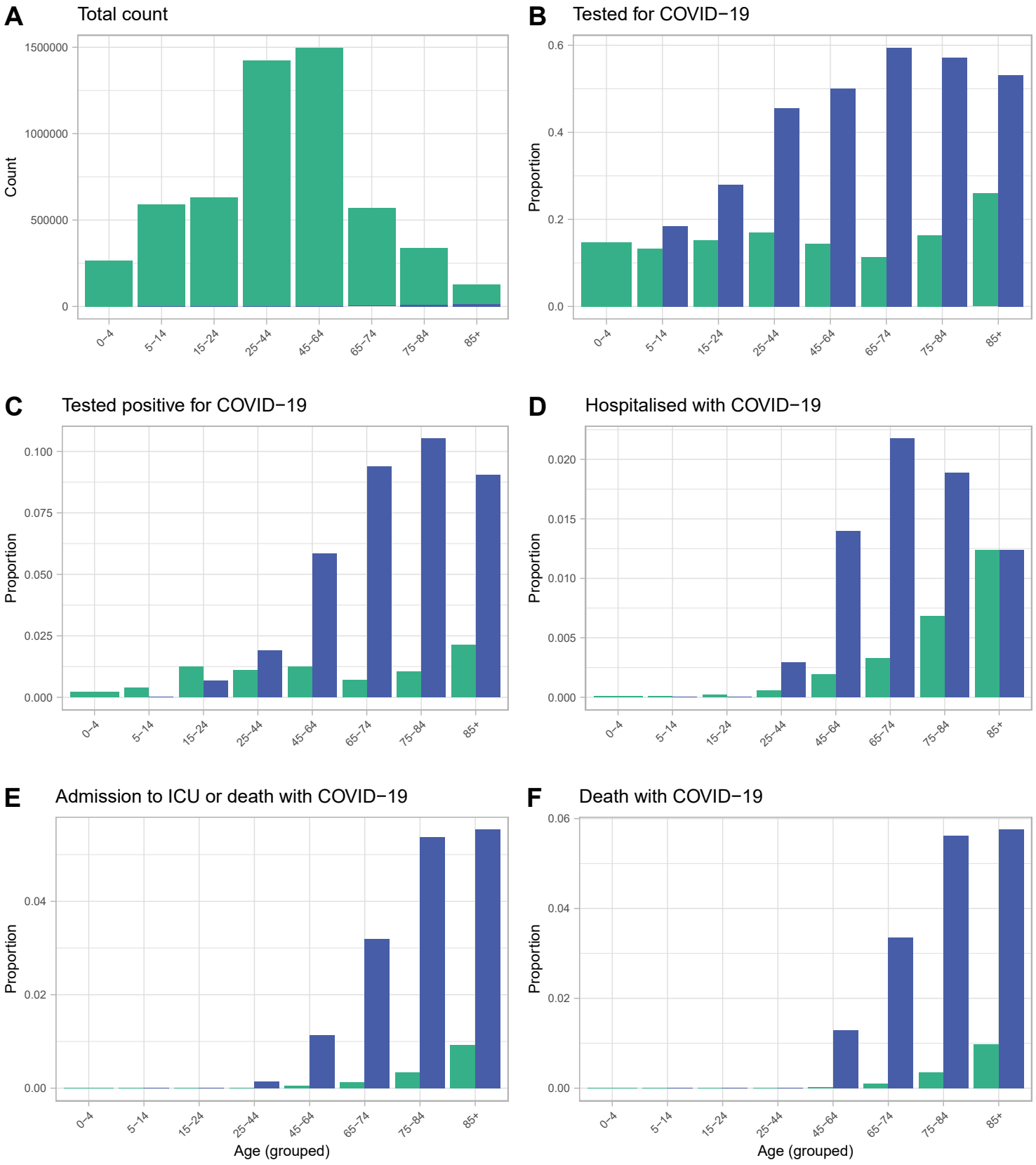
Time frame: 01 Mar to 10 Nov 2020

Asthma



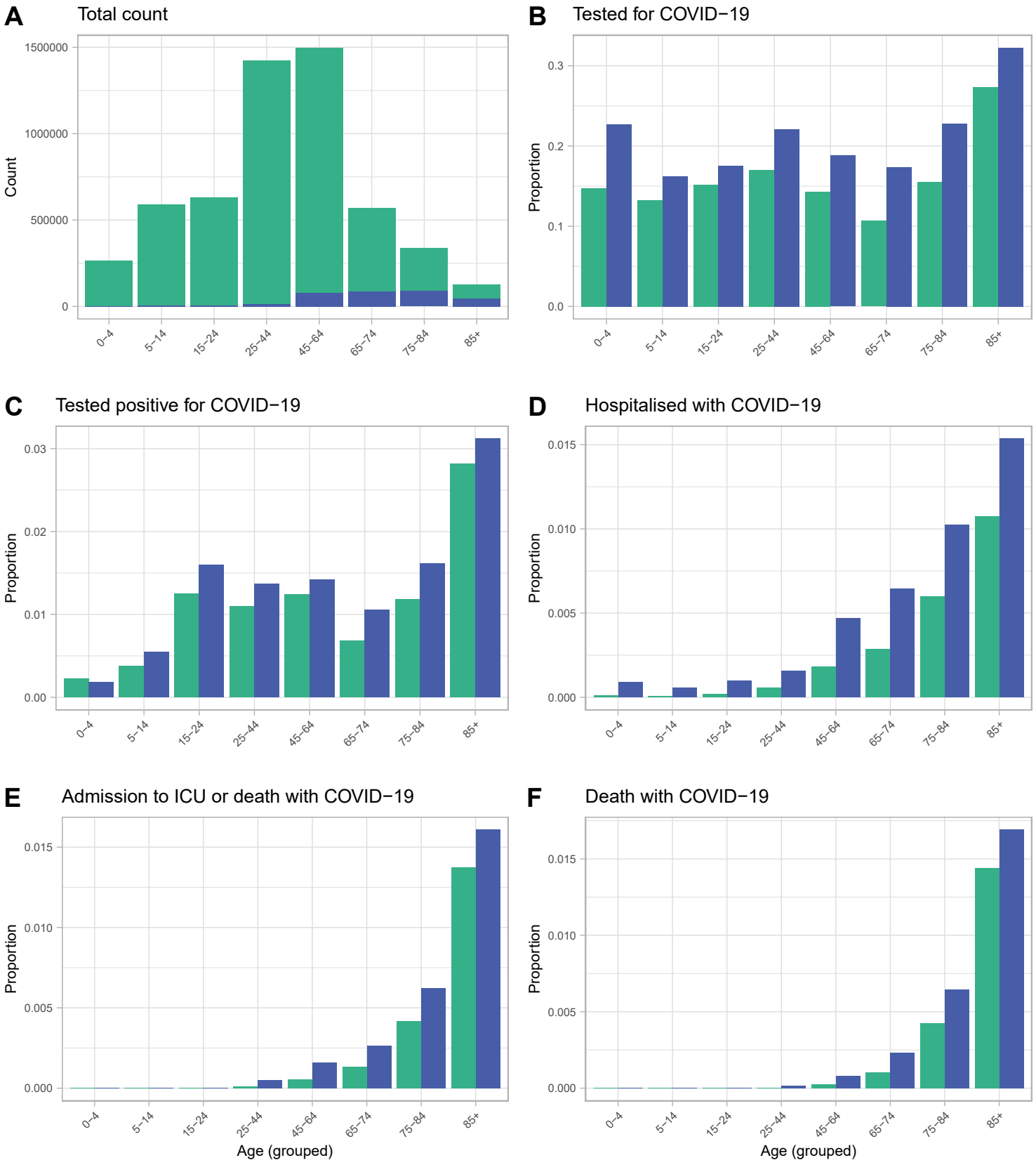
Time frame: 01 Mar to 10 Nov 2020

Care home



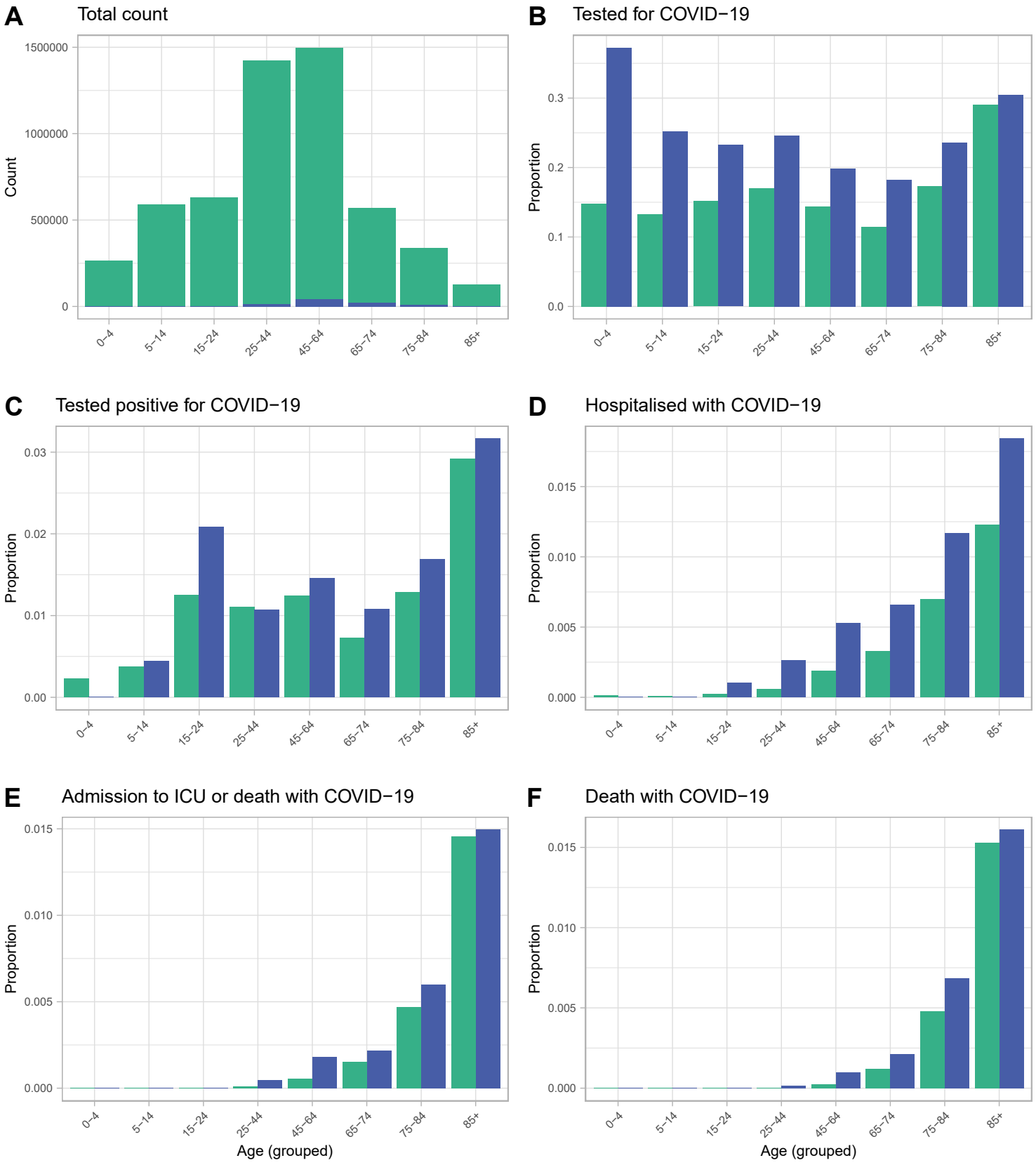
Time frame: 01 Mar to 10 Nov 2020

Chronic heart disease



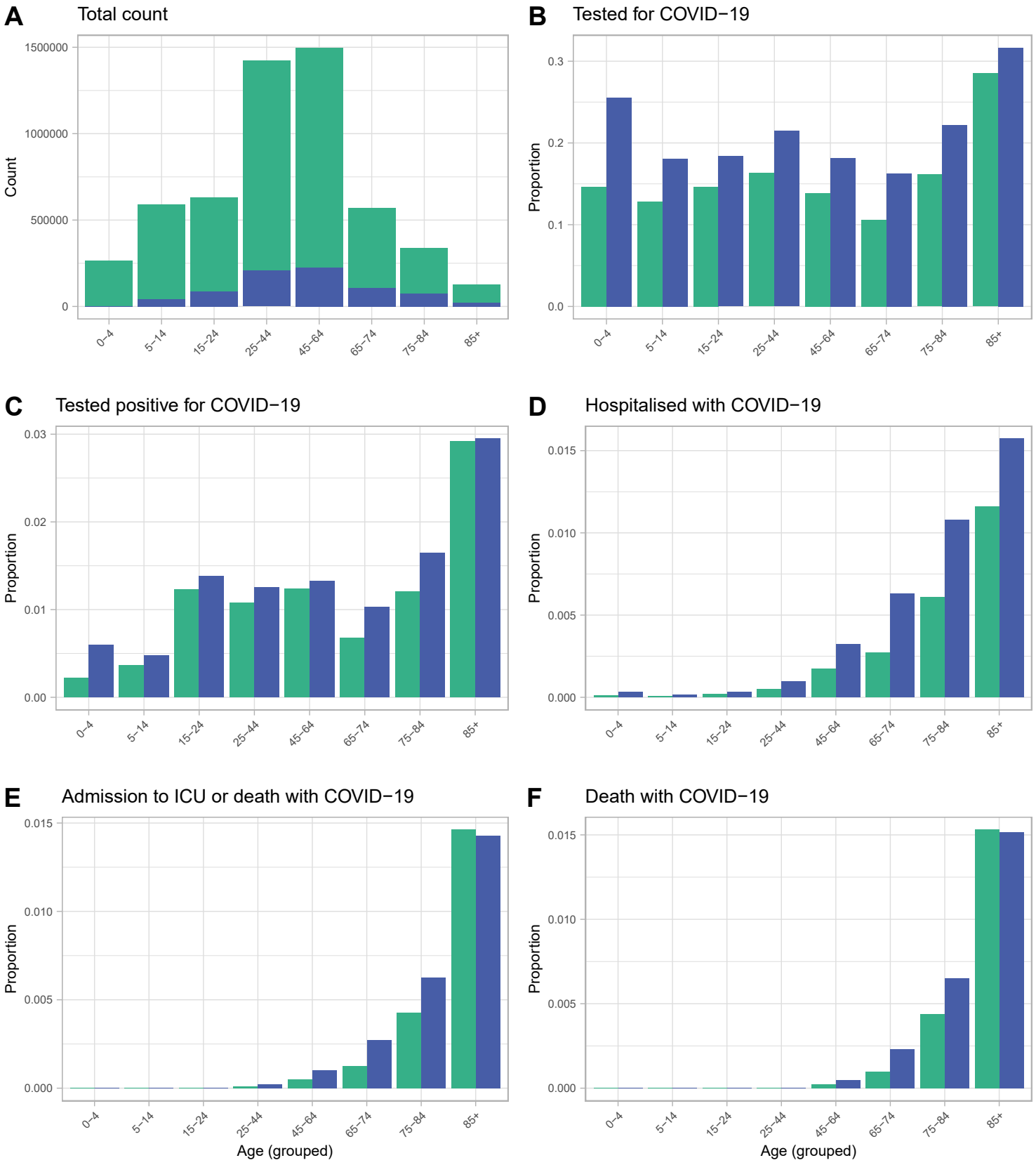
Time frame: 01 Mar to 10 Nov 2020

Chronic liver disease



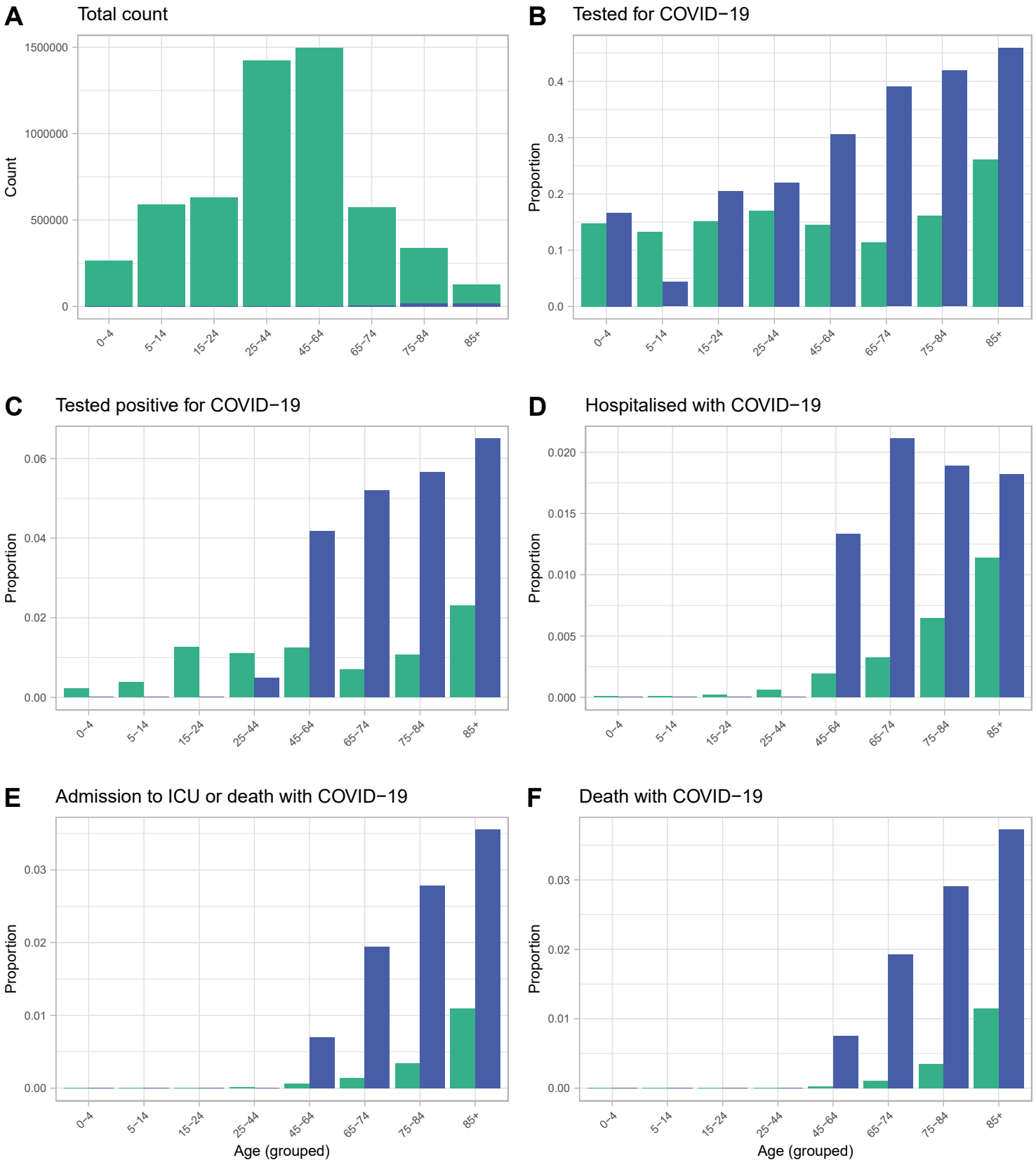
Time frame: 01 Mar to 10 Nov 2020

Chronic respiratory disease



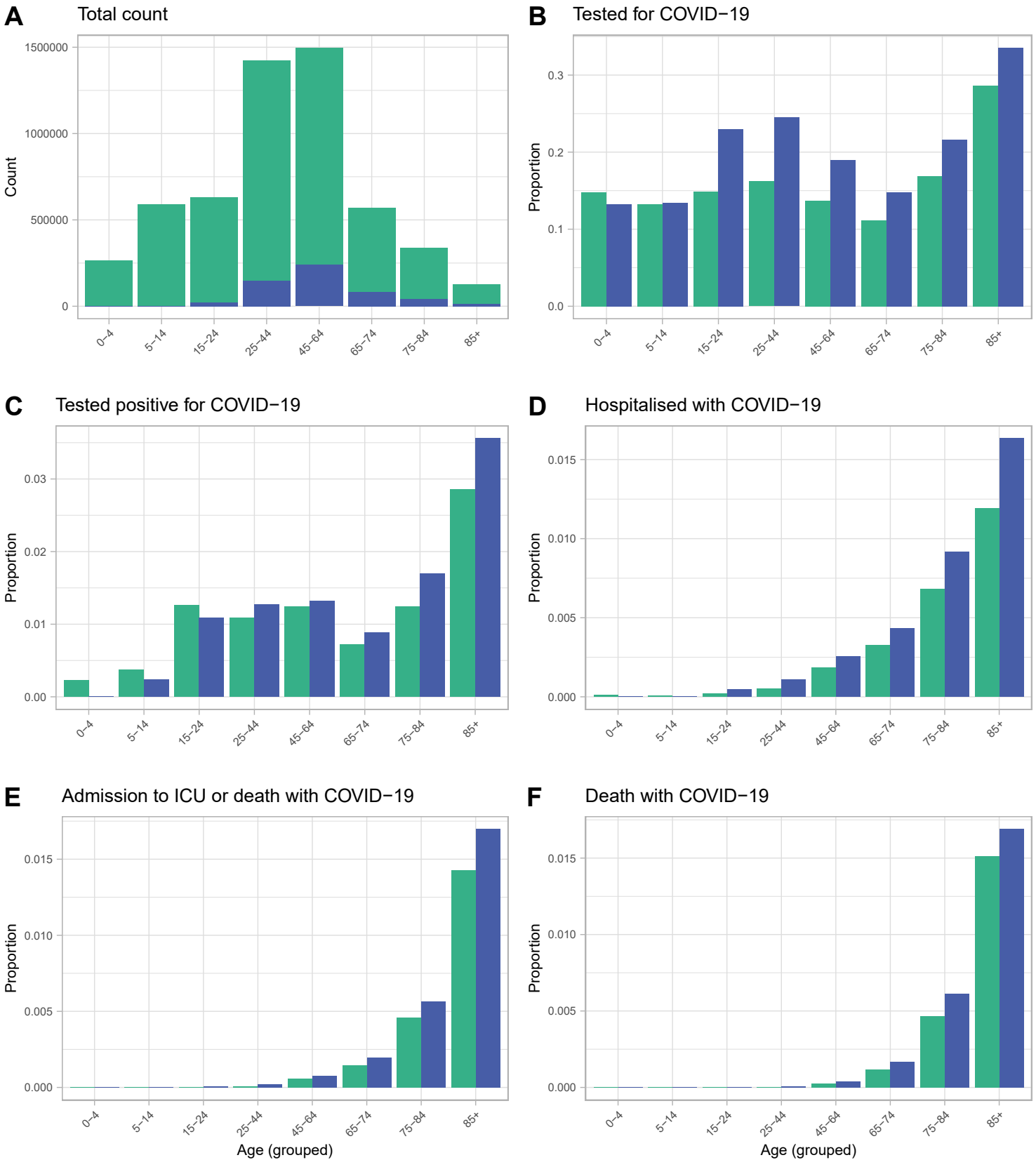
Time frame: 01 Mar to 10 Nov 2020

Dementia



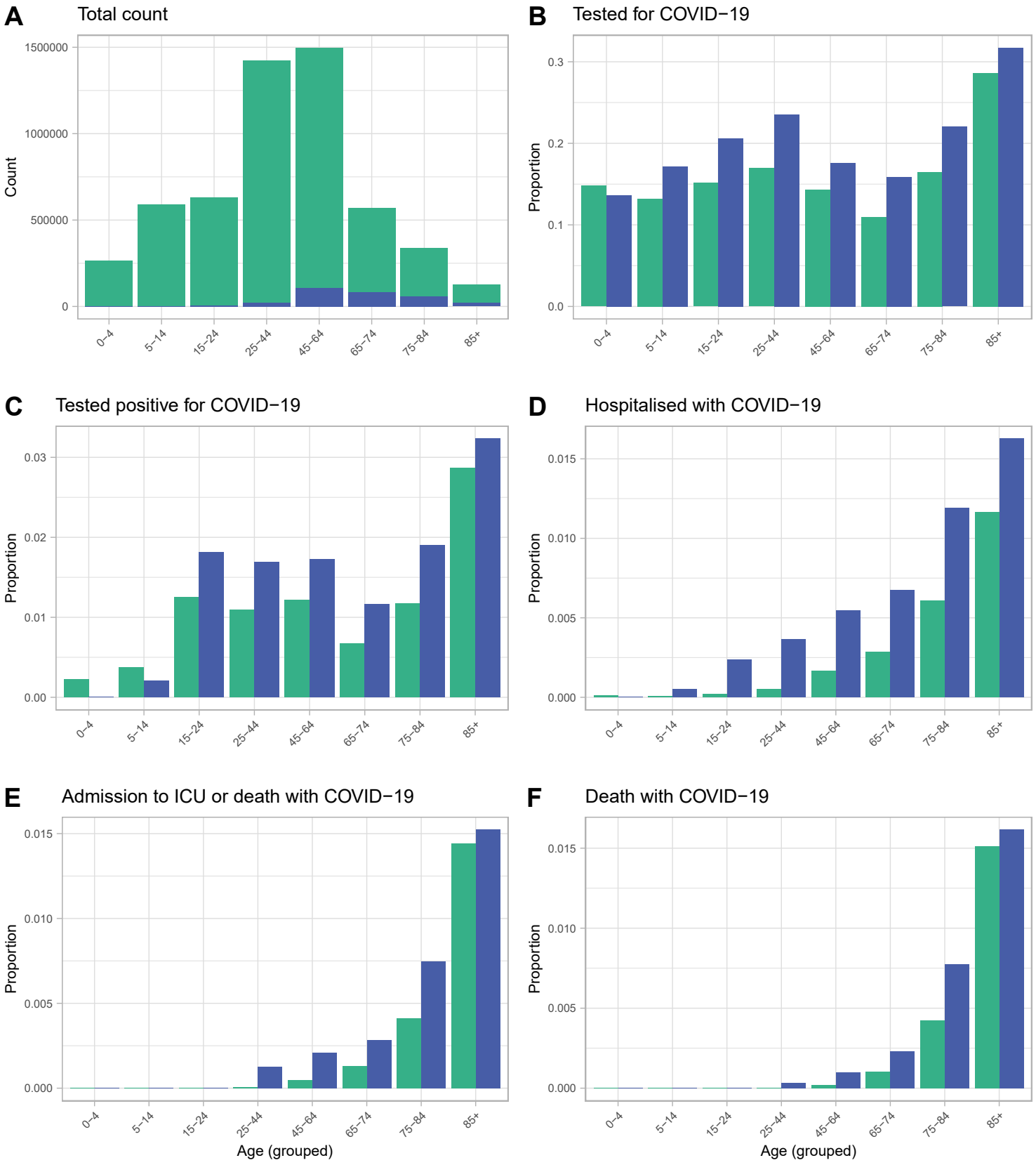
Time frame: 01 Mar to 10 Nov 2020

Depression

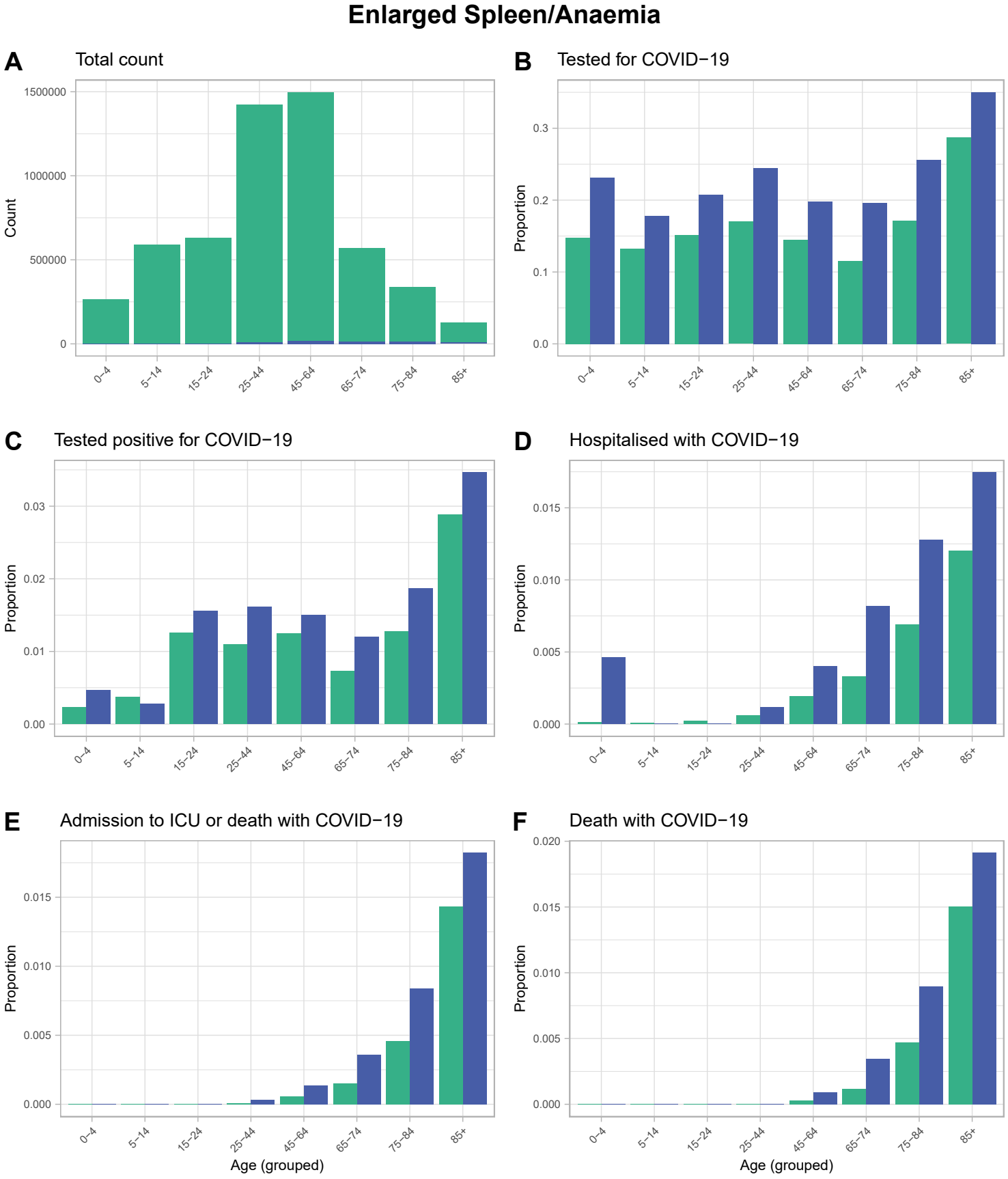


Time frame: 01 Mar to 10 Nov 2020

Diabetes

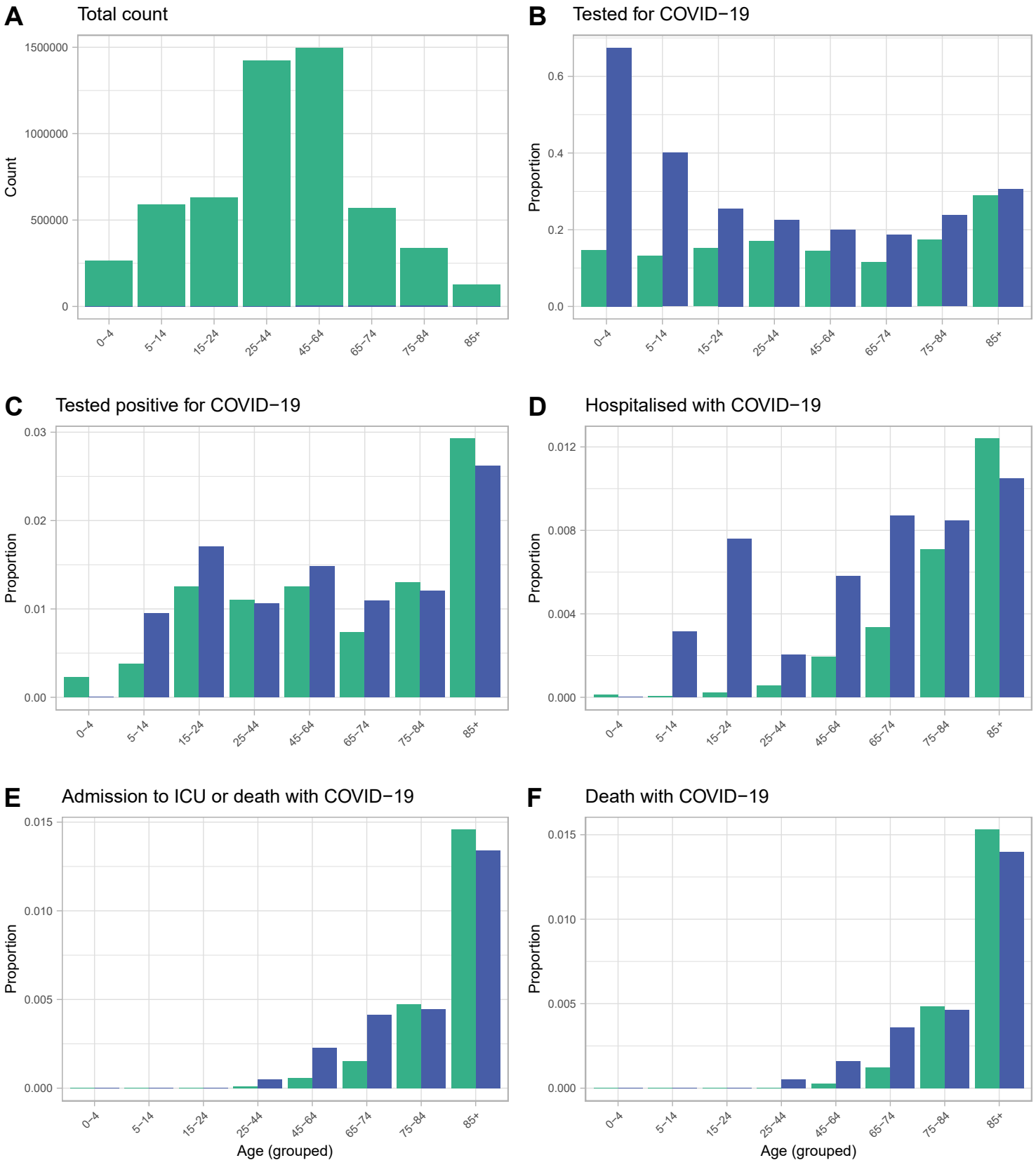


Time frame: 01 Mar to 10 Nov 2020



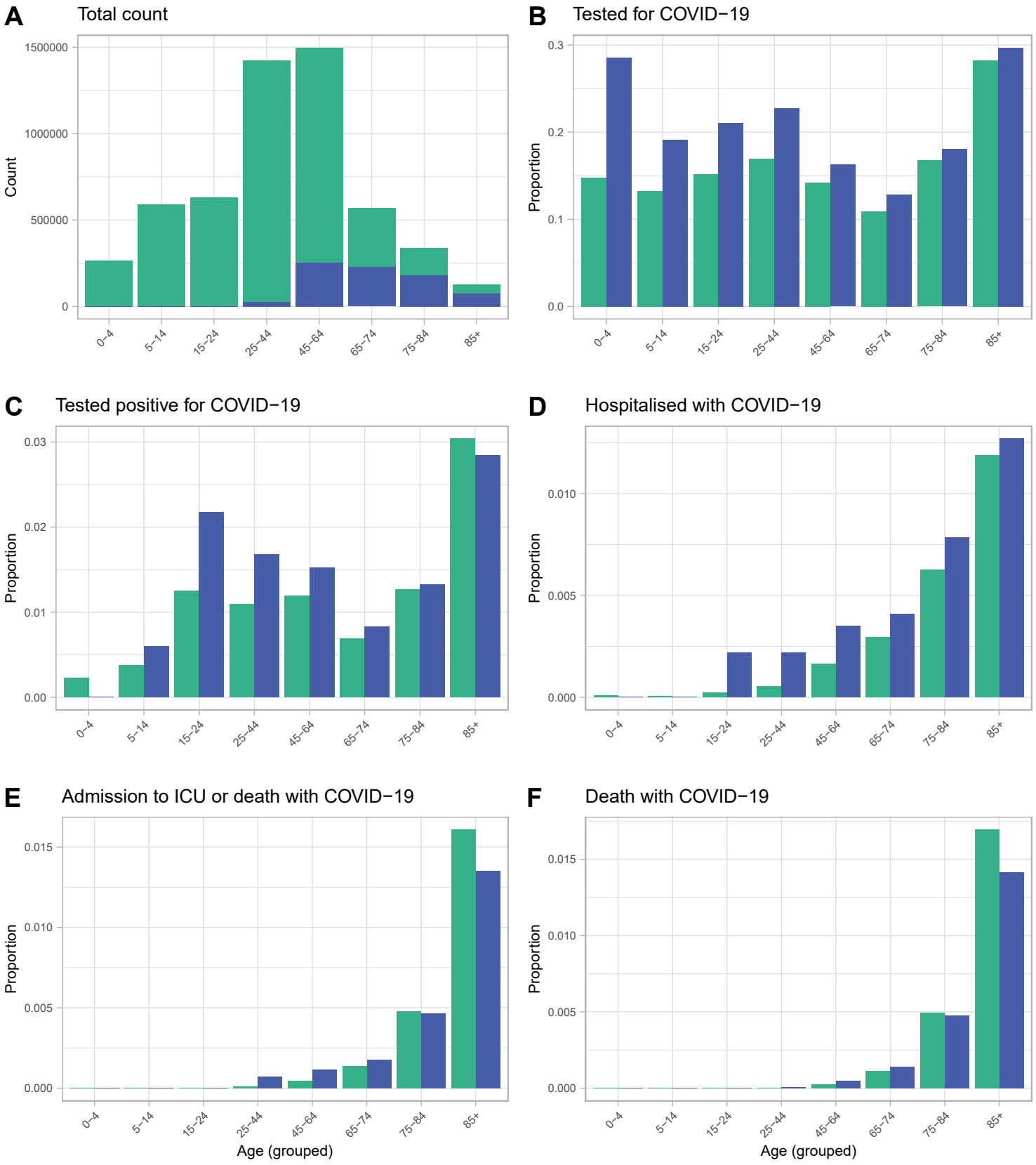
Time frame: 01 Mar to 10 Nov 2020

Haematological malignancy



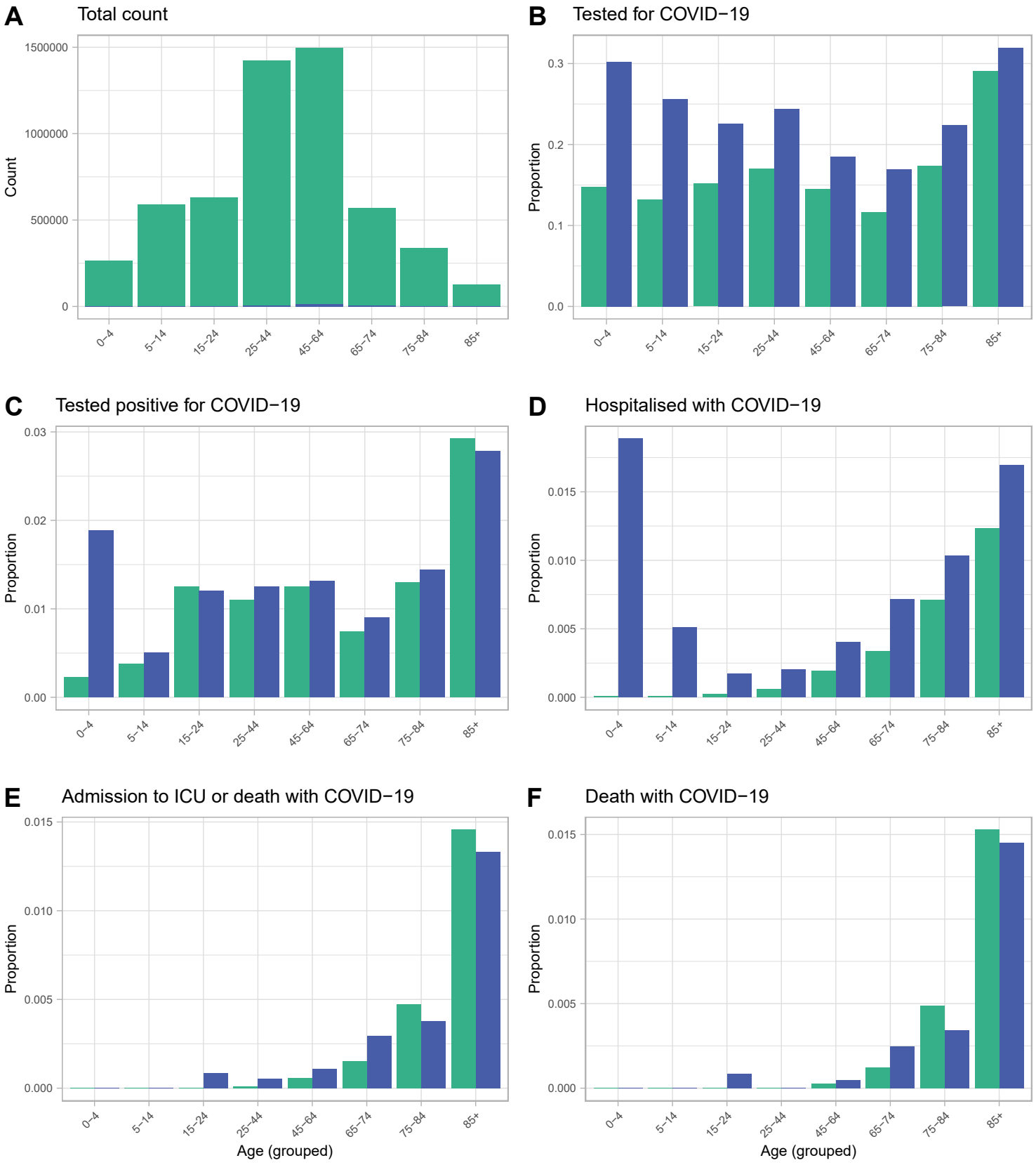
Time frame: 01 Mar to 10 Nov 2020

Hypertension



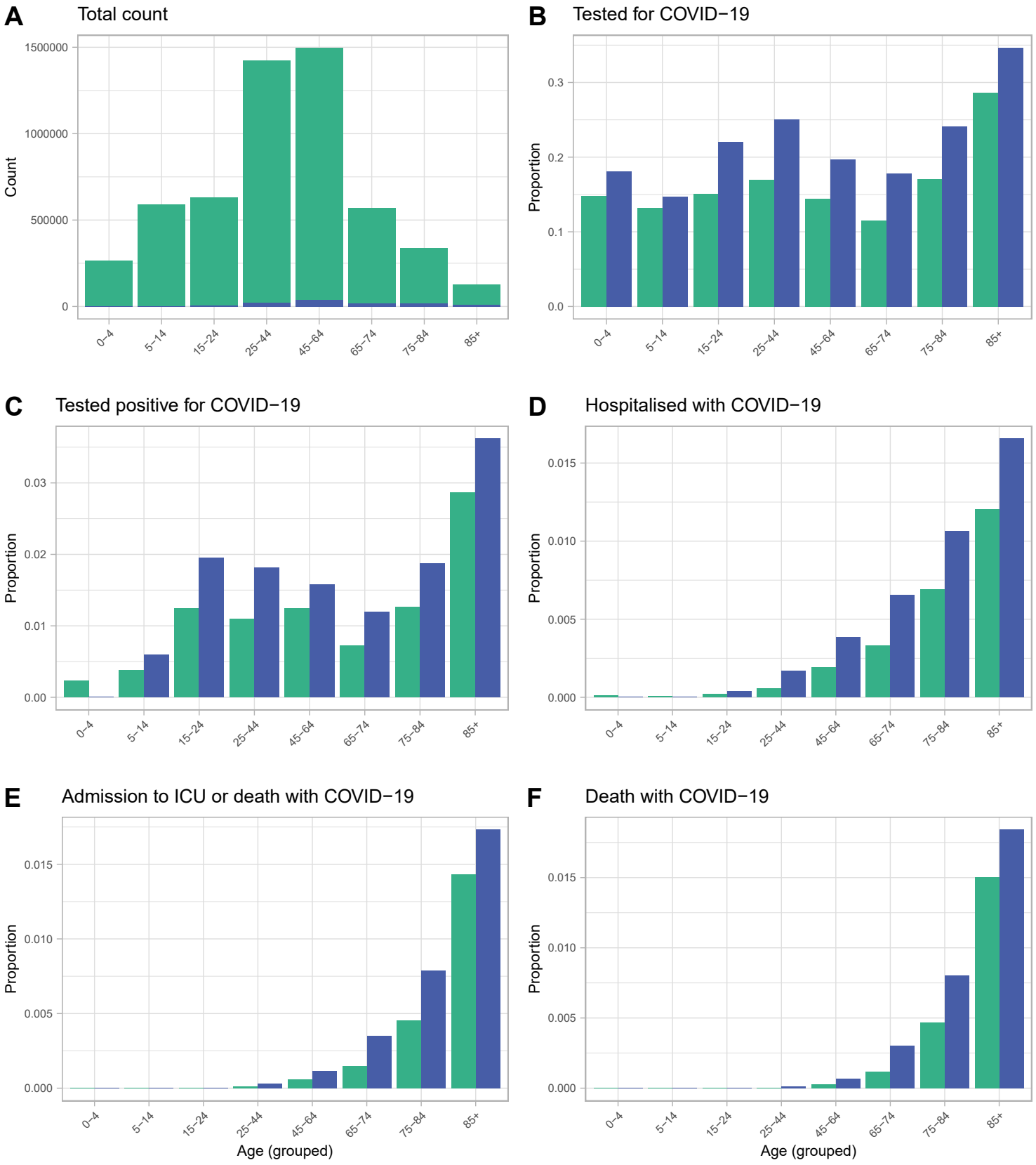
Time frame: 01 Mar to 10 Nov 2020

Immunosuppression



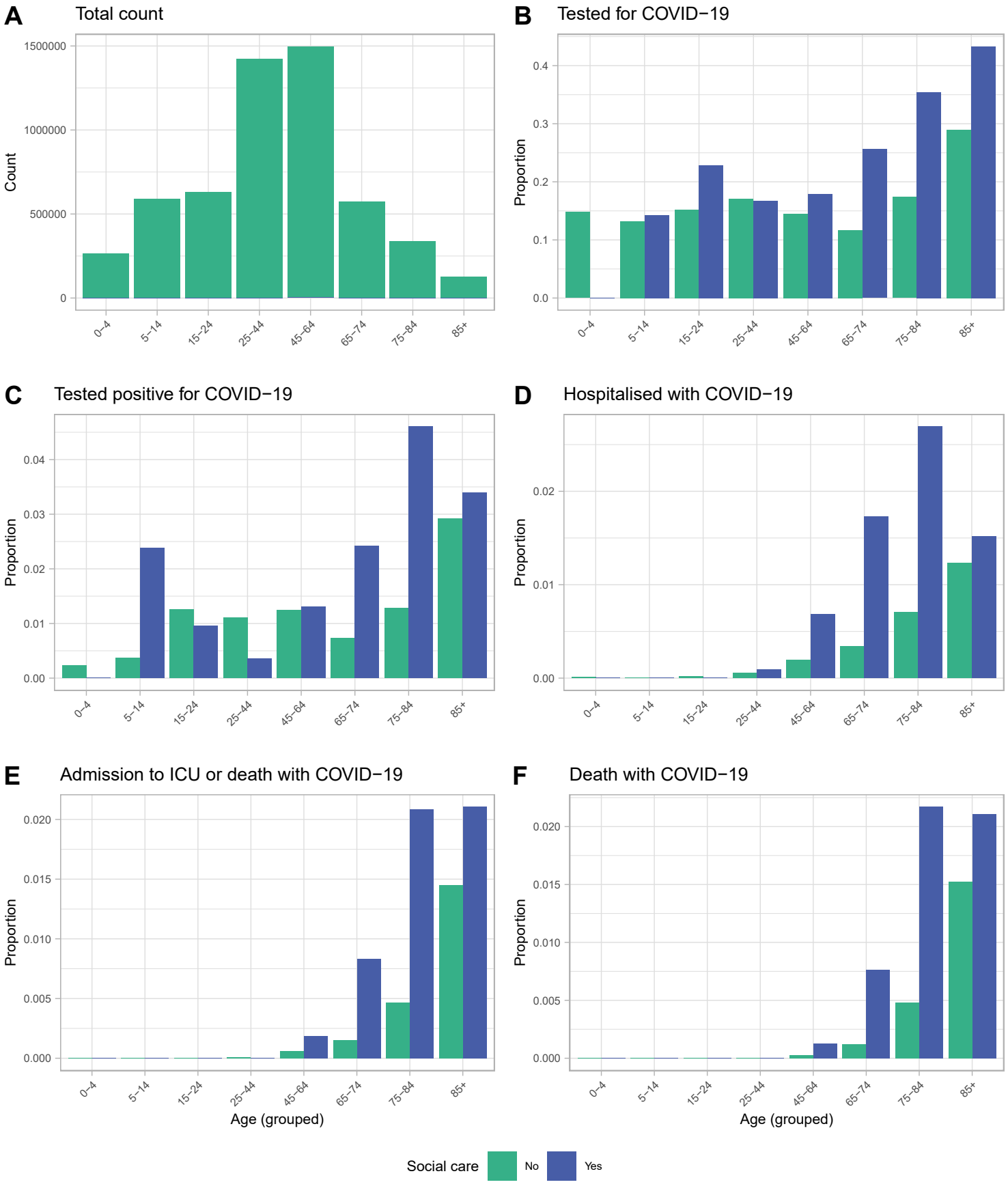
Time frame: 01 Mar to 10 Nov 2020

Nutritional deficiencies



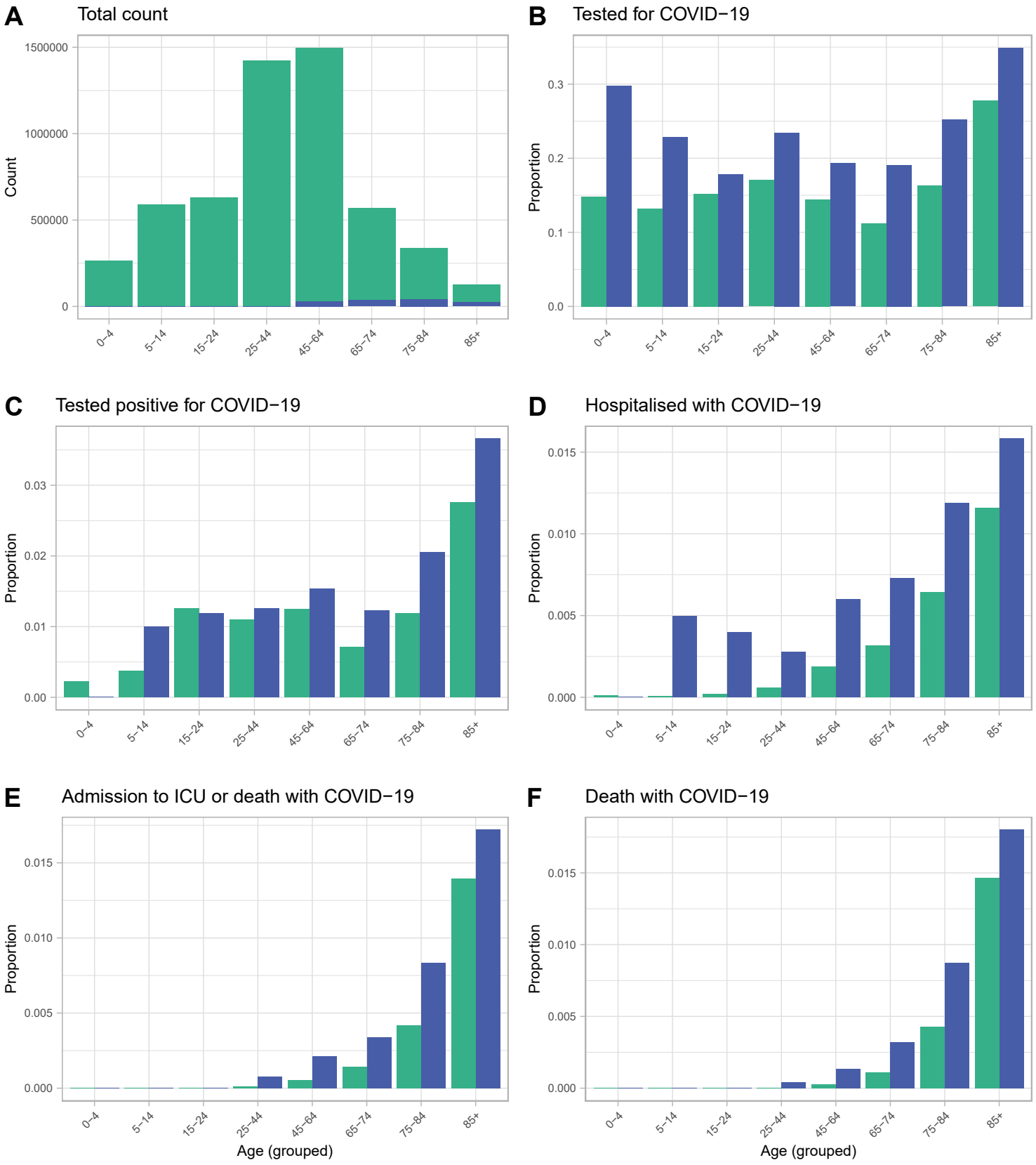
Time frame: 01 Mar to 10 Nov 2020

Social care



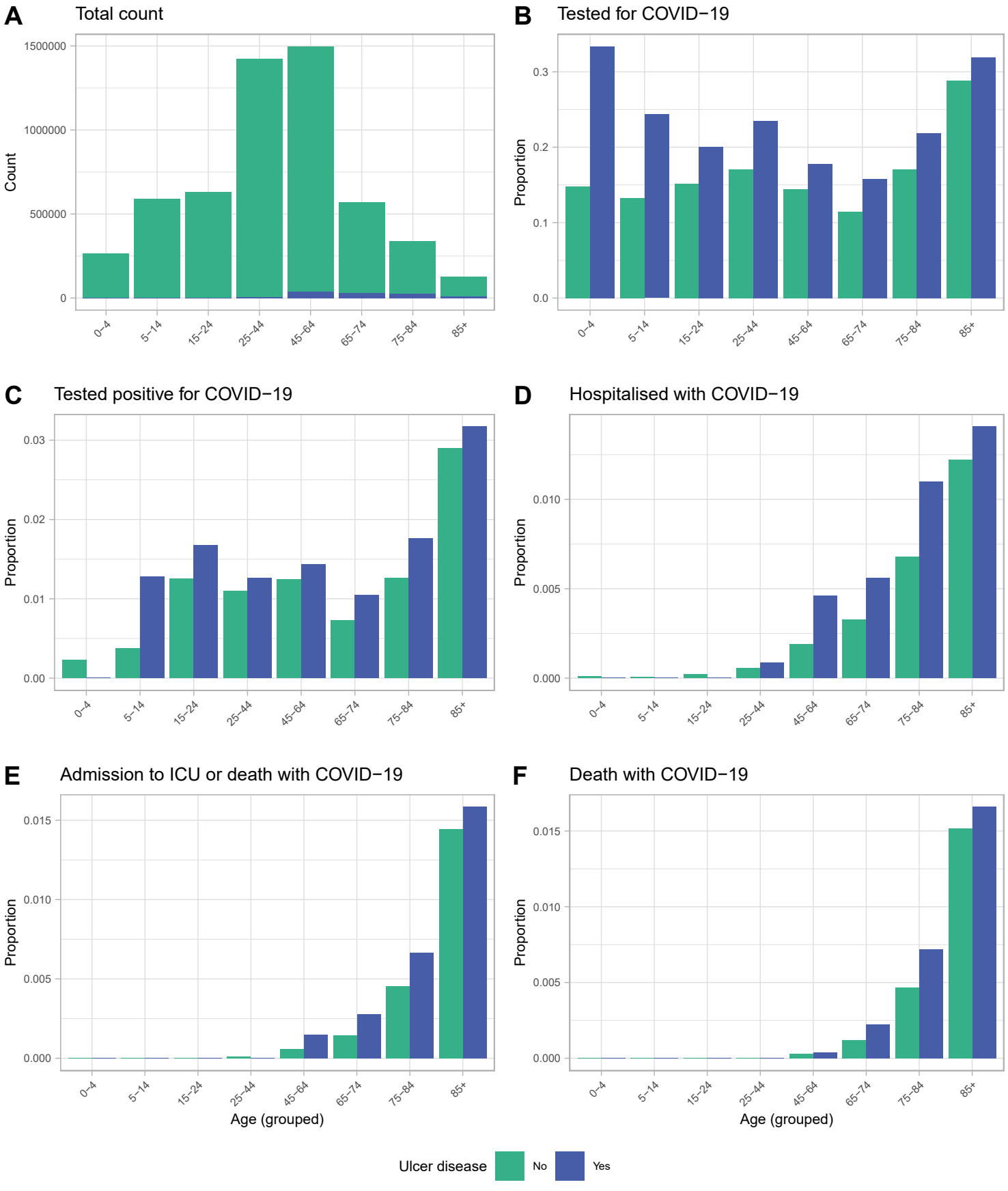
Time frame: 01 Mar to 10 Nov 2020

Stroke/Transient ischaemic attack (TIA)



Time frame: 01 Mar to 10 Nov 2020

Ulcer disease



Time frame: 01 Mar to 10 Nov 2020