Risks of and risk factors for COVID-19 disease and development of a risk prediction model in people with diabetes in Scotland: A cohort study

Stuart J McGurnaghan[c,a], Amanda Weir[c], Jen Bishop[c], Sharon Kennedy[c], Luke AK Blackbourn[a], David McAllister[d], Sharon Hutchinson[f], Thomas M Caparrotta[a], Joseph Mellor[b], Anita Jeyam[a], Joseph E O'Reilly[a], Sarah Wild[b], Sara Hatam[a], Andreas Höhn[a], Marco Colombo[a], Chris Robertson[e], Nazir Lone[b], Janet Murray[c], Elaine Butterly[d], John Petrie[l], Brian Kennon[m], Rory McCrimmon[k], Robert Lindsay[l], Ewan Pearson[k], Naveed Sattar[l], John McKnight[1], Sam Philip[o], Andrew Collier[f], Jim McMenamin[c], Alison Smith-Palmer[c], David Goldberg[c], Paul M McKeigue[c,b], Helen M Colhoun[c,a,h], Public Health Scotland COVID-19 Health Protection Study Group[c], Scottish Diabetes Research Network Epidemiology Group[i]

[a]*Institute of Genetics and Molecular Medicine, College of Medicine and Veterinary Medicine, University of Edinburgh, Western General Hospital Campus, Crewe Road, Edinburgh EH4 2XUC, Scotland. HC - AXA Chair in Medical Informatics and Epidemiology*
[b]*Usher Institute, College of Medicine and Veterinary Medicine, University of Edinburgh, Teviot Place, Edinburgh EH8 9AG, Scotland. PM - Professor of Genetic Epidemiology and Statistical Genetics. NL - Clinical Senior Lecturer in Critical Care*
[c]*Health Protection Scotland (Public Health Scotland), Meridian Court, 5 Cadogan Street, Glasgow G2 6QE*
[d]*Institute of Health and Wellbeing, University of Glasgow, 1 Lilybank Gardens, Glasgow G12 8RZ. DM - Senior Clinical Lecturer in Public Health*
[e]*Department of Mathematics and Statistics, University of Strathclyde, 16 Richmond Street, Glasgow G1 1XQ. CR - Professor of Public Health Epidemiology*
[f]*School of Health and Life Sciences, Glasgow Caledonian University. SH - Professor of Epidemiology and Population Health*
[g]*Information Services Division, NHS Scotland, UK*
[h]*Department of Public Health, NHS Fife, Kirkcaldy, UK*
[i]*The Scottish Diabetes Research Network Epidemiology Group, UK*
[j]*Diabetes Epidemiology Unit, University of Dundee, Dundee, UK*
[k]*Division of Molecular and Clinical Medicine, University of Dundee, Dundee, UK*
[l]*Institute of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow, UK*
[m]*Queen Elizabeth University Hospital, Glasgow, UK*
[n]*Western General Hospital, NHS Lothian Edinburgh, UK*
[o]*Grampian Diabetes Research Unit, Diabetes Centre, Aberdeen Royal Infirmary, Aberdeen, UK*

Abstract

Background: The objectives were to i) ascertain the cumulative risk of fatal or CCU treated COVID-19 in those with diabetes and compare it to those without diabetes ii) among those with diabetes to investigate risk factors for, and build a cross-validated predictive model of, fatal or CCU treated COVID-19.

Methods: In the total population of Scotland we ascertained all persons who had developed fatal or critical care unit-treated COVID-19 (hereafter F/CCU-COVID-19) between 1st March and July 31st 2020 from the nationwide virology, critical care unit, hospital discharge and register of deaths databases. Among those with F/CCU-COVID-19, diabetes status was ascertained by linkage to the national diabetes register. The cumulative incidence of F/CCU-COVID-19 in those with and without diabetes was compared using logistic regression. Among those with diabetes, data on potential risk factors for F/CCU-COVID-19 were obtained from diabetes register and other linked health administrative databases. Among those with diabetes we tested association of these factors with F/CCU-COVID-19 and constructed a prediction model using stepwise regression and 20-fold cross-validation.

Findings: 1082 (0.3%) of all those with diabetes in Scotland (n=319 349) developed F/CCU-COVID-19; the age and sex adjusted odds ratio was [OR] 1.395 95% CI: 1.304-1.494, p<0.001 overall compared to the risk in total population of Scotland without diabetes (n=4081 cases in 5 143 951 persons). The OR was 2.396 (1.815-3.163) in type 1 and 1.369 (1.276-1.468) in type 2 diabetes. Of the 1082 persons with diabetes who developed F/CCU-COVID-19, 90% were ≥60 years old. Among those with diabetes, adjusted for age, sex, diabetes duration and diabetes type, those who developed F/CCU-COVID-19 were more likely to be male, live in residential care, or live in a more deprived area, have a condition already listed as a COVID-19 risk condition such as heart or lung disease, have retinopathy or reduced renal function, have worse glycaemic control, have a prior diabetic ketoacidosis or hypoglycaemia hospitalisation in the past five years, be on more diabetes and other type of drugs, and to have been a smoker (all p<0.001). Those with F/CCU-COVID-19 were less likely to be on an antihypertensive and had lower systolic blood pressure than those without F/CCU-COVID-19 (both p<0.001). The relationship with BMI was J-Shaped. The cross-validated predictive model of F/CCU-COVID-19 in those with diabetes, retained 11 factors in addition to age, sex, diabetes type and duration and had a C-statistic of 0.85 (95% CI 0.83,0.86).

Interpretation: Overall risks of F/CCU-COVID-19 are substantially elevated in those with type 1 and type 2 diabetes compared to the background population. The risk of F/CCU-COVID-19, and therefore the need for special protective measures, varies widely among those with diabetes but can be predicted reasonably well using prior clinical history.
Funding: none

Research in Context

Evidence before this study

We searched PubMed and the META database for studies examining risks of COVID-19 associated with diabetes that had appropriate comparator populations and for studies exploring among those with diabetes what risk factors predict COVID-19 using the terms (COVID-19 OR "novel coronavirus" OR SARS-CoV-2) AND diabetes, through 5th October 2020. Case-series have reported a high prevalence of diabetes among those hospitalised or a high test-positivity rate for diabetes among those tested. However, diabetes is common, so to quantify the risk ratios for COVID-19, comparison to the background population is needed. Only four such studies have reported; these found relative risks of relative risks of 2.04 for type 2 diabetes and 3.5 for type 1 diabetes for COVID-19 hospitalised death and 1.9 and 2.4 for COVID-19 hospitalisation and hospitalised death respectively for all diabetes in the others. The range of potential determinants explored in these studies has been limited.

Added value of this study

In this study, we captured the data encompassing the first wave of the epidemic i.e. from March 1st 2020 when the first case was identified to 31st July 2020 on which date infection rates had dropped sufficiently that shielding measures were officially terminated. Including CCU treated and out of hospital deaths from COVID-19 for the first time, as well as hospitalised deaths, we have shown that the risk of F/CCU-COVID-19 is increased 2.4-times in type 1 diabetes and 1.4-times in type 2 diabetes. For the first time we have shown that those with recent admissions history for hypoglycaemia and diabetic ketoacidosis have an increased risk of severe or fatal disease. Ever smokers also had increased risks. Prior specific comorbidities including heart disease, liver disease, and chronic lower respiratory disease increased risk. We have shown for the first time that being exposed to more drug classes and having more prior hospital admissions are markers of risk. A risk prediction model achieved a C-statistic of 0.85. We have provided a Shiny app to give the reader a sense of how individual risk factor profiles in those with diabetes translate into elevated risks compared to those without diabetes.

Implications of all the available evidence

During phases of the SARS-CoV-2 pandemic, when the effective reproduction number is high, those people with diabetes most at risk may warrant special protection measures. A risk prediction score based on past history can usefully identify those with diabetes most at risk and we provide an example of such a score.

Background

Initial case series of people hospitalised with COVID-19 in several countries found over-representation of people with diabetes[1–8]. More than a quarter of those admitted for COVID-19 in the UK had diabetes[9].

Just four studies, three from the UK, have compared risks in defined populations with and without diabetes and all found increased risks in those with diabetes for in-hospital and total deaths[10–14]. Guidelines accordingly describe all those with diabetes as being at elevated risk[8,15] but it is likely that among those with diabetes some are at very high risk warranting special protection measures and that others not at much more risk than the background population. As we enter the second wave of the epidemic a greater understanding of variation in COVID-19 risk in those with diabetes is needed to tailor protection measures and inform vaccine strategies.

Only one study has explored determinants of risk of COVID-19 among those with diabetes to any extent with Black and South Asian ethnicity, lower socioeconomic status, poorer glycaemic control and prior cardiovascular disease reported to increase risks[13,14]. BMI was the only predictor of outcome of hospitalised COVID-19 beyond age, sex and diabetes duration in a large French case series[16].

In this study we have used data from the first wave of the epidemic in Scotland i.e. from March 1st 2020 when the first case was identified to 31st July 2020 on which date infection rates had dropped sufficiently that shielding measures were officially terminated. For the total population of Scotland we aimed to i) compare the cumulative risk of fatal or critical care unit treated (F/CCU) COVID-19 in all those with and without diabetes ii) among those with diabetes, to ascertain which factors were associated with F/CCU-COVID-19 and iii) to build a cross-validated risk prediction model. Our focus was on F/CCU-COVID-19 as rates of testing positive or being hospitalised with COVID-19 are biased due to selective testing and hospitalisation policies.

Methods

The participants were the total population of Scotland (n=5 463 300) including all those with diabetes nationwide (n=319 349) alive three weeks before the start of the epidemic in Scotland (estimated as Feb 7th 2020). The study period was from March 1st 2020 to the 31st July 2020.

COVID-19 case ascertainment

For the total population of Scotland evidence of any detected COVID-19 was defined as having had a positive RT-PCR test for SARS-CoV-2 or a hospital discharge code for COVID-19, or a COVID-19 code (U071 or U072) anywhere on the death certificate. The databases used were the Electronic Communication of Surveillance in Scotland (ECOSS) database that captures all NHS virology testing, the RAPID database of daily hospitalisations, the Scottish Morbidity Records-01 of hospital discharges and the National Records of Scotland (NRS) death registrations data. All such health related databases in Scotland are linkable as they all use the Community Health Index (CHI) unique identifier. The CHI database also yielded data on age, sex, residential postcode, and residential care home status. For all cases whether critical care had been provided was obtained by linkage to the Scottish Intensive Care Society and Audit Group (SICSAG) database. Critical care included all admissions to intensive (ICU), high dependency (HDU) or combined ICU/HDU care. Fatal COVID-19 was defined based on there being a U071 or U072 code anywhere on the death certificate, OR any death within 28 days of testing positive. These are the official death definitions used by National Register of Scotland and Public Health Scotland.

Diabetes database

To identify diabetes status among all those COVID-19 we linked the COVID-19 data to the national diabetes register (Scottish Care Information (SCI)-Diabetes) and its associated research platform. As described elsewhere[17,18], the diabetes research database has >99% coverage of all those with a diabetes diagnosis in Scotland. Inception into SCI-Diabetes occurs when a diagnostic code for diabetes is assigned in primary or secondary care across Scotland. Nightly uploads of key data items from primary, secondary and community clinical care into a federated national health system (NHS) database follows. These data include diabetes type and clinical measurements such as body mass index (BMI), blood pressure, laboratory results, smoking history and annual screening retinopathy grade that we have used in our analysis. Regular extracts from this database are linked to other datasets including hospitalisations (Scottish Morbidity Record (SMR) 01), dispensed prescriptions (Prescribing Information System-PIS), renal registry, deaths and other routine datasets, using the CHI number then anonymised and imported into the research platform. Records for all persons alive in the register at the start of the epidemic were used in this analysis (n=319 349). See Supplement Methods for a detailed description of key variables from the database used in the analysis.

Demographic data

All those with diabetes were assigned to which quintile of the residential postcode-based indicator, the Scottish Index of Multiple Deprivation (SIMD), they belonged[19]. Self-assigned ethnic group and residential care home status were captured from SCI-diabetes and CHI databases respectively.

Other morbidity and drug prescribing

From the diabetes research platform, all hospital discharge codes from SMR01 over the past five years and Anatomical Therapeutic Chemical (ATC) codes[20] from PIS data for the past three years were extracted and used to define co-morbid conditions and prior drug exposures. We derived history of a specific list of conditions and drug classes that have been included as risk conditions for COVID-19 by public health agencies, hereafter termed "listed conditions"[15] (see Supplement Methods for details of the codes used).

Statistical methods

Cumulative incidence of F/CCU-COVID-19 in those with and without diabetes between March 1st 2020 and July 31st 2020

For calculating cumulative incidence (risk) we used the age and sex specific counts of F/CCU-COVID-19 in those with and without diabetes over the study period. The age and sex distribution of those with diabetes as of three weeks before the first observed positive test nationally was available from SCI-Diabetes (n=319 349). To obtain the at risk population for the population without diabetes we used the most recent publicly available one-year age and sex band counts of the total Scottish population (Mid 2019 n=5 463 300) available from National Records of Scotland[21]. We assumed these counts pertained at the start of the epidemic. From this, we subtracted the numbers alive in the diabetes register in each sex specific age band to give the population without diabetes (n=5 143 951). We summarised the relative difference in cumulative incidence of F/CCU-COVID- up to 31 July 2020 in those with and without diabetes by sex as the odds ratio from a logistic regression model using one-year age band and sex specific counts of cases and denominators.

Excess deaths

For the population with diabetes, the weekly counts of total deaths for the at risk population in each of the past five years was available from the SCI-Diabetes research. We plotted the total number of deaths per week in 2020 in those with diabetes along with the weekly average for the same week over the period 2015-2019 with the difference representing excess deaths.

Associations of risk factors for F/CCU-COVID-19 among those with diabetes

Using the SCI-Diabetes research platform we first described sociodemographic variables, the "listed conditions"[15] and potential vascular and diabetes specific risk factors in those with diabetes who did and did not develop F/CCU-COVID-19. Association of each risk factor with F/CCU-COVID-19 was then reported adjusting for age, sex, diabetes duration and type of diabetes using logistic regression. In total, associations of 35 variables were tested as listed in Table 2. We report p-values unadjusted for multiplicity. Global P-values were calculated using a likelihood ratio test, comparing models with and without the variable using R stats[22] 'drop1' function. For regression, missing variables were imputed using chained equations assuming data were missing at random using the 'Amelia'[23] package (see supplement for details).

Construction of risk prediction model for F/CCU-COVID-19 among those with diabetes

Using the same data on 35 covariates, we constructed a multivariable risk prediction model of F/CCU-COVID-19 among those with diabetes. Age, sex, diabetes type, diabetes duration were fitted simultaneously as the baseline model. The 35 covariates and interaction terms for age:sex, age:diabetes type, sex:diabetes type were made available for selection into the final model. We used the *mfp* package in R[24,25] to first ascertain whether any of the continuous variables should be fitted with any additional polynomial terms because of departure from linearity (see supplement for details). We then used stepwise regression, alternating between forward and backward steps, implemented in the R function stats::step, so as to maximize the Akaike Information Criterion (AIC), selecting any additional potential factors as being predictive of F/CCU-COVID-19. The predictive performance of the base model and then the final model were evaluated by 20-fold cross validation with performance calculated over all test folds as the C-statistic and also as the expected information for discrimination using the *wevid* package (see supplement for additional detail)[26].

The COVID-age for a person with diabetes can be defined as the age at which the risk of COVID-19 in a non-diabetic individual of the same sex equates to the risk in the diabetic individual under study. This can be derived from the final risk model in those with diabetes and the modelled risks in those without diabetes (see supplement methods). To enable a user to calculate this COVID-age for a person with diabetes and a given set

of characteristics, we generated a Shiny application located at https://diabepi.shinyapps.io/covidrisk/ (Online; Accessed 11th November 2020). The purpose of the Shiny app to give the reader a sense of how individual risk factor profiles in those with diabetes translate into elevated risks compared to those without diabetes (see supplement for further details). This study is registered as an International Standard Randomised Controlled Trial number ISCRTN45562523.

Results

Cumulative incidence (risk) of F/CCU-COVID-19 in those with and without diabetes 1 March to 31st July 2020

Risk and excess death in those with diabetes

Among those 319 349 persons in the population with diabetes at the start of the epidemic, by 31st July 2020 2724 had any evidence of COVID-19. Of the 2724 persons with any evidence of COVID-19, 1082 had developed F/CCU-COVID-19 (0.3% of all those with diabetes) of whom 963 died. More details of case ascertainment and severity ascertainment is given in Supplement (Supplement Figure 1 and associated Supplement Results text).

Table 1 shows the distribution of risk of F/CCU-COVID-19 in the population with diabetes. The risk of F/CCU-COVID-19 increased with age. Just 30 of 1082 with F/CCU-COVID-19 were below age 50 years (all of these were age >20 years). The overall risk was 0.4% in males and 0.3% in females. Overall, 0.1% of those with type 1 and 0.4% of those with type 2 diabetes developed F/CCU-COVID-19.

Figure 2 shows the total deaths among those with diabetes in every 7 day-window since 1/1/2020 plotted against the average number of deaths in that same seven day interval in 2015-19. The difference in these two is excess deaths during the period. The grey zone depicts how much was attributable to COVID-19. Total deaths exceeded the average from early March and returned to the average by early June. Altogether in this first wave of the epidemic between March 1st and July 31st 2020 there were 1228 excess deaths in comparison to the average for this period in the preceding years. Of these, 963 (78.45%) were due to COVID-19.

Comparison of risk in those with and without diabetes

In the total population of Scotland without diabetes (n=5 143 951) 4081 (0.08%) developed F/CCU-COVID-19 in the same period (Supplement Table 1). Figure 1 shows the risk of F/CCU-COVID-19 in those with and without diabetes by sex. The diabetes associated increase in risk was apparent in both sexes and at all age bands.

Adjusted for age and sex in a logistic regression, as of 31st July 2020, diabetes was associated with an odds ratio of 1.395 (95% CI 1.304-1.494, p<0.001) for F/CCU-COVID-19 with similar odds ratio found in males and females (Supplement Table 2). The OR was 2.396 (1.815-3.163) in type 1 and 1.369 (1.276-1.468) in type 2 diabetes. There was a statistically significant interaction between diabetes and age on risk of F/CCU-COVID-19 (p<0.001) with the OR being 2.494 (2.032- 3.061) for those age 0-59, 1.764 (1.457-.2.136) for those 60-69 years and 1.327 (1.227-1.434) in those aged 70 upwards for example. When the analysis was limited to various time points since the start of the epidemic the OR associated with diabetes was highest at end of March 1.770 (1.566-2.002) and fell to 1.446 (1.343-1.557) by end of Aril and then to the July 31st value of 1.395 (1.304-1.494).

Associations of risk factors with F/CCU-COVID-19 among those with diabetes

Table 2 shows the crude unadjusted characteristics of those with diabetes who did and did not develop F/CCU-COVID-19. Supplementary Tables 3 and 4 show these data by type of diabetes. Table 3 shows the odds ratios for F/CCU-COVID-19, 95% confidence intervals and p-values for age, sex, diabetes type and duration and then for each other risk factor separately adjusted for age, sex diabetes type and duration. The comments below are based on these adjusted data. Data for continuous variables divided into categories including missingness are given in Supplementary Tables 5 and 6.

Sociodemographic factors

As shown in Table 3 older age, male sex and longer diabetes duration were all associated with statistically significant increased risk of F/CCU-COVID-19. Adjusted for these factors, the type of diabetes was not associated with any statistically significant difference in risk. Living in a residential care home was associated with very large and statistically significant increased risk of F/CCU-COVID-19 (OR 16.570, 95% CI 14.326, 19.165 p<0.001). Sociodemographic (SIMD) quintile showed a strong statistically significant gradient in risk falling from the most to least deprived quintile. There was no statistically significant variation in risk by ethnic group. Note however, that the prevalence of non-White ethnicities in this diabetes population (Table 2) is too low, commensurate with the background general population of Scotland, to have any power to detect ethnic variation in COVID-19 risk among those with diabetes.

Co-morbid conditions and clinical factors

As shown in Table 3, the number of prior hospitalisations for hypoglycaemia, diabetic ketoacidosis and for reasons other than these over the past five years was strongly associated with F/CCU-COVID-19. Each of the co-morbid conditions listed as risk conditions for COVID-19 showed strong, highly statistically significant associations with F/CCU-COVID. Risk increased with increasing HbA1c. There was no statistically significant linear relationship between BMI and disease (OR 1.002, 95% CI 0.991, 1.013, p=0.706, Table 3). However, the multivariable fractional polynomials analysis revealed evidence for a statistically significant non-linear J-shaped relationship with BMI (See Supplement Figure 2 for a plot of the relationship from the *mfp* analysis). Those who developed F/CCU-COVID-19 had statistically significantly lower systolic blood pressure than those who did not. Being on any antihypertensive was associated with a statistically significantly lower risk of F/CCU-COVID-19 (OR 0.801, 95% CI 0.705, 0.909 p=0.001). More detailed exploration of type of antihypertensives (see Supplement Tables 5 and 6) showed that the point estimate for the OR for each antihypertensive subclass was below 1, other than for the rarely used "centrally-acting" class. In those who developed F/CCU-COVID, eGFR was statistically significantly lower and prevalence of albuminuria was higher than those who did not. Having retinopathy was statistically significantly associated with developing F/CCU-COVID-19. Having a history of smoking was associated with a significantly increased risk of F/CCU-COVID-19 though risk was not statistically significantly higher in current versus never smokers.

Statistically significant increased risks were found in recipients versus non-recipients of several drug classes including non-steroidal anti-inflammatory drugs, proton-pump inhibitors and anti-coagulants. The more diabetes drug subclasses used in the past three years, the greater the risk of F/CCU-COVID-19 disease. Having been on insulin or sulphonylureas was associated with the highest risks (Supplement Table 6). As shown in Table 3 the number of different types of drugs other than those used for diabetes that the person had been exposed to in the past three years was strongly associated with F/CCU-COVID.

Distribution of characteristics in those with and without F/CCU-COVID-19 by type of diabetes is given in Supplement Tables 3 and 4. Broadly the same pattern of associations is found in both type 1 and type 2 diabetes. The main differences were that DKA and hypoglycaemia admission rates, and the differences between those with and without F/CCU-COVID-19 were greater, for type 1 diabetes than type 2. A sensitivity analysis restricted to the fatal cases showed the same pattern of associations as was found for F/CCU-COVID-19.

Risk prediction model for F/CCU-COVID-19 among those with diabetes

Table 4 shows the final set of covariates retained in the stepwise selection model, on top of age, sex, diabetes type and duration which were entered as fixed covariates. The multivariable fractional polynomials analysis indicated that the association with number of hospital admission was best entered into the model as log (admissions) and that terms for both BMI and log BMI should be included. Note that selection is based on the AIC not p-values. Terms for interactions between age and type of diabetes, sex and type, and age by sex were not selected. The C-statistic for the baseline model containing just age, sex, diabetes type and duration was 0.76 (95% CI:0.75-0.76). The cross-validated stepwise model retained a further 11 factors and had a C-statistic of 0.85 (95% CI:0.83-0.85) (Supplement Figure 3). The cross-validated model was well calibrated (Supplement Figure 4) and the Hosmer Lemeshow test was not statistically significant at p=0.378.

The Shiny app that uses the prediction model and the risk data in the population without diabetes to return the COVID-age can be found at https://diabepi.shinyapps.io/covidrisk/ (Online; Accessed 11th November 2020).

Discussion

Risk of F/CCU-COVD-19 in diabetes

This report highlights the elevation in risk of F/CCU-COVID-19 with diabetes. The elevation in risk relative to the population without diabetes adjusted for age and sex was higher for type 1 (2.4 times) type 1 than for type 2 (1.4-times). This greater elevation in type 1 diabetes is likely accounted for by longer duration of diabetes since in the older age bands cumulative incidence was higher in type 1 than type 2 and since, among those with diabetes, no significant difference in risk by type was found once age, sex and diabetes duration were adjusted for. The lower overall age distribution in type 1 than type 2 diabetes and the strong association of older age with risk however meant that overall a lower proportion of those with type 1 than type 2 developed F/CCU-COVID-19. Although there were no cases of F/CCU-COVID-19 under age 20 years in those with diabetes, above this age an elevation in the risk associated with diabetes was apparent.

In terms of absolute risk three people with diabetes in every 1000 developed F/CCU-COVID-19 up to July 2020. The impact on the weekly deaths was clearly discernable and peaked in early April. Of note the excess was not all explained by COVID-19 designated deaths; this may represent under-ascertainment of COVID-19 deaths but may also reflect the knock on effect on health services of the epidemic[27].

We focused on F/CCU-COVID-19 rather as the probability of being tested or hospitalised for any given level of symptoms could easily vary by diabetes status leading to observation bias. Another important potential bias might be termed "at risk bias". Diabetes was named early in the epidemic as a moderate risk condition. Thus people with diabetes may have adopted social distancing measures more stringently than those without diabetes which could bias the odds ratio downwards. Indeed, consistent with this, when the analysis was limited to various time points since the start of the epidemic the OR associated with diabetes was highest at end of March 1.770 (1.566-2.002) falling to 1.395 (1.304-1.494) by end July. Depletion of the most susceptible early in the epidemic could also cause the odds ratio to fall over time this since the susceptibility is higher in those with diabetes. However, we have no direct data to prove these potential explanations.

There are few other studies with general population denominators allowing the relative risk of COVID-19 in those with diabetes relative to the background population to be estimated. In the OpenSAFELY study primary care records in England were linked to death certification records. The age, sex-adjusted OR for COVID-death associated with diabetes was 1.6 for those with an HbA1c $<$ 58 mmol/mol and was 2.6 for those with HbA1c above this level. Type of diabetes was not differentiated[11]. An analysis of the UkBiobank study diabetes was associated with a relative risk of 1.91 for COVID-19 hospitalisation[28]. In the English National Audit cohort[13,14] risk ratios of COVID-19 death for type 1 and type 2 diabetes were 3.51 and 2.03 and were attenuated in Whites (3.06 for type 1 and 1.91 for type 2). In a matched case control study of the total population of Scotland from earlier in the epidemic and not including cases derived solely from hospital admissions (as these weren't yet available) we reported slightly higher conditional odds ratios of 2.75 (1.96-3.88) for type 1 and 1.60 for type 2 (1.48-1.74)[29]. All these studies are therefore consistent in finding elevations in risk for type 1 and type 2 diabetes in the same range. However, the OR for diabetes will vary somewhat depending on the stage of the epidemic and with the ethnicity distribution as well as whether out of hospital deaths are captured. Studies that do not capture out of hospital deaths may preferentially omit older cases and will report a higher summary OR since the OR varies with age.

Risk factors association with F/CCU-COVID-19 in diabetes

We found that risk of F/CCU-COVID-19 rose steeply with age and was higher in males as has been reported in many other populations[30]. More than a third of those with diabetes who developed F/CCU-COVID-19 lived in residential care homes emphasizing the critical importance of protecting such vulnerable individuals during the remainder of this epidemic. There was a strong socioeconomic gradient.

We showed that adjusted for age, sex, diabetes duration those who developed F/CCU-COVID-19 on average had worse profiles for almost every clinical measure we examined; they were more likely to have other co-morbidities and evidence of diabetic microvascular disease (with more impaired renal function and retinopathy). On average they had worse glycaemic control, were more likely to have had a prior diabetic ketoacidosis or hypoglycaemia hospitalisation and other hospitalisations in the past five years. They were on more diabetes and non-diabetes drugs. We also found strong associations with NSAIDS and proton pump inhibitors which are among the most commonly prescribed drugs and are often markers of polypharmacy. They were more likely to have smoked. We found a J shaped relationship with BMI. Surprisingly, although strong associations of hypertension with COVID-19 have been reported elsewhere[5–7], we found that those who developed F/CCU-COVID-19 had slightly lower blood pressures than those who did not and that being on anti-hypertensives was associated with a lower risk than not being on any antihypertensive. Among the specific anti-hypertensive drug classes thiazides and angiotensin II receptor antagonists/ blockers had the lowest odds ratios for F/CCU-COVID-19.

Similar associations of age, sex diabetes duration, socioeconomic status, prior CVD, renal status, blood pressure and glycaemic control with COVID-19 death were found in the English National Audit study[13,14]. In that study non-white ethnicity was found to be associated with COVID-19 death whereas in Scotland the prevalence of non-White ethnic groups is too low to allow detection of any ethnic differences in COVID-19; just 2.9% of those with diabetes are known to be of South Asian origin and 0.5% of Black origin. We found that being on antihypertensive drugs was associated with a lower risk but they found an increased risk. However, that higher risk was driven by a large effect in South Asian and Mixed ethnicity groups and was not seen in whites or other ethnic groups. The U-shaped association with BMI in the English Audit Study was stronger than the J-shaped relationship we found. This difference is likely also driven by the different ethnic mix in the studies since the relationship of higher BMI to higher risk was most apparent in those of non-White ethnicity in the English Audit Study. The increased risk at lower BMI including underweight in both studies likely reflects comorbid effects related to F/CCU-COVID-19 that are associated with weight loss. Given the elevation in BMI among people with type 2 diabetes, it would not be surprising to see such comorbid effects resulting in the nadir of the curve being around the average BMI of 30kg/m^2 as was found here. In the English Audit Study, as in ours, ex-smokers were at increased risk but they reported that current smokers were at reduced risk, but we did not. This difference may reflect that their smoking effects were reported adjusted for all other variables including possible mediators such as CVD. The extensive data on other factors we examined were not evaluated in the English Audit Study.

Such minimally adjusted associations that we have reported are useful as a prelude to building the predictive model of F/CCU-COVID-19 discussed further below. They are also useful for suggesting possible causal mechanisms. Thus the SIMD differential may be partly mediated through higher levels of smoking, and worse glycaemic control and onward effects on CVD and other co-morbidities but may also relate to other unmeasured factors determining infection such as overcrowding or occupation. However much more extensive modelling methods[31] are needed to infer causality for each of the associated factors and is out of scope here. Such methods are especially needed to understand drug associations which are hugely susceptible to confounding by indication. Meanwhile it is worth considering which of the associations we report, if causal, would be modifiable. Improved protection in residential care homes, smoking cessation, improved glucose control, reduction of BMI, optimised management of comorbidities, medication reviews of polypharmacy are possible interventions to reduce risk suggested by these analyses but require formal analysis. It is interesting that the data suggest protective effects of anti-hypertensives but this also requires more detailed causal analyses.

Prediction model and estimation of COVID-age

We obtained a reasonable predictive accuracy in our multivariable model with a C-statistic of 0.85; thus faced with a case and non-case pair the model would correctly assign the case as being at higher risk 85% of the time. This level of predictive accuracy disproves the notion all those with diabetes are of similar risk. The variables retained variables in the model are those that are most predictive and not necessarily causal; some of the most valuable predictors include the number of prior hospital admissions in past five years, number of diabetes drugs and number of non-diabetes drugs which were not evaluated in other diabetes COVID-19 studies.

The absolute risk of F/CCU-COVID-19 will mostly reflect the stage of the epidemic and the current effective R number in the population[32]. Accordingly we have produced the Shiny app to convert the absolute risk score produced by the prediction model to the COVID-age i.e. the age at which the same absolute risk was observed in a person of the same sex in the population without diabetes at the same stage of the epidemic. This concept of COVID-age is becoming increasingly used in occupational health[33] and is more interpretable than scores that produce absolute risks such as the QCOVID score[34]. COVID-age should be less susceptible to the prevailing R than absolute risk but we will monitor the need the need to recalibrate the underlying models as the pandemic unfolds.

Limitations and advantages

Some key strengths of our study are the total population coverage, the inclusion of out-of-hospital deaths, the inclusion of those who might have died without CCU, the much more extensive exploration of potential prior risk factors than conducted previously and the development of the shiny app for COVID-age. Limitations of our study are potential biases noted above and that we, as others, lack quality control data on assignation of COVID-19 deaths. Furthermore we lack the detailed clinical data needed to define severe cases according to the World Health Organisation criteria[35] or to capture all possible co-morbidities. Another limitation is that we had to make an assumption age and sex bands population numbers won't have changed much between mid-2019 and the start of the epidemic in Feb/March 2020. This is however a very reasonable assumption: between 2018 and 2019 the overall change in the Scottish population size was just +0.5% with no change in those aged 75+ years[21]. An important limitation is that we have not been able access any other datasets in which to externally validate the risk prediction model. Therefore, its presentation here is primarily to facilitate an understanding of the magnitude of increase in risk that occurs with different risk factor combinations that is not easily intuited from looking a table of odds ratios associated with specific risk factors or markers. We also hope that it serves as an illustration to those in other countries of an approach they might usefully adopt that could help those with diabetes and their clinicians to make shielding decisions during the rest of the epidemic. It is likely that our data are relevant to many developed settings but in developing countries the background mix of other infectious and non-infectious diseases among those with diabetes may vary considerably. Also, for many countries, the low prevalence of non-White people in our population means that potentially important ethnicity effects that may pertain are not represented in our model.

Policy Implications

Both type 1 and type 2 diabetes are associated with substantial increases in the risk of COVID-19 disease compared with the risks seen in people of the same age in the background population. However, it is important to consider the absolute number of persons with diabetes in our population that have developed severe or fatal disease; 3 in 1000 persons have had fatal or CCU treated disease. We have shown that among those with diabetes risk of severe disease varies widely and is predictable. This should inform shielding policies and vaccine prioritisation strategies. The Shiny app has been provided for illustrative purposes only to allow a greater understanding of how a prediction model broadly translates into COVID-age in individuals with diabetes. External validation, regulatory approval and appropriate licensing would be required before this could be used in clinical practice.

Declarations

Information governance

This research was conducted with approval from the Public Benefit Privacy Protection Panel (PBPP ref. 1617-0147), originally set up under PAC 33/11, with approval from the Scotland A Research Ethics Committee (ref. 11/AL/0225). All datasets were de-identified before analysis.

Data Sharing

NHS Data governance rules do not permit us to secondarily share the data directly. However, *Bone fide* researchers can apply to the Scottish Public Benefits and Privacy Protection Committee for access to these data.

Conflicts of interest

SH reports grants from Health Protection Scotland, during the conduct of the study. TMC reports grants from Diabetes UK Grant: 18/0005786, outside the submitted work. SW reports other from Novo Nordisk, other from Gilead, outside the submitted work. CR reports grants from Public Health Scotland, grants from UKRI, during the conduct of the study; and Member of Chief Medical Officer of Scotland Scientific Advisory Group for COVID19 Member of SPI-M a subgroup of the UK Scientific Advisory Group for Epidemics Member of MHRA Advisory Group for Vaccine Safety. JP reports personal fees from Merck KGaA, non-financial support from Merck KGaA, personal fees from Novo Nordisk, personal fees from Boehringer-Ingelheim, grants from Janssen, personal fees from Biocon, non-financial support from Astra Zeneca, all outside the submitted work. RM reports personal fees from NovoNordisk, personal fees from Sanofi Aventis, outside the submitted work. NS reports personal fees from Amgen, personal fees from AstraZeneca, grants and personal fees from Boehringer Ingelheim, personal fees from Eli Lilly, personal fees from Novo Nordisk, personal fees from Pfizer, personal fees from Sanofi, outside the submitted work. JMK reports personal fees from NAPP pharmaceuticals, other from Novonordisk, other from Eli Lilly, other from Boehringer, other from Medimmune Ltd, other from Glaxo-Smith Kline, during the conduct of the study. HMC reports grants and personal fees from Eli Lilly and Company, during the conduct of the study; grants from AstraZeneca LP, other from Novartis Pharmaceuticals, grants from Regeneron, grants from Pfizer Inc, other from Roche Pharmaceuticals, other from Sanofi Aventis, grants and personal fees from Novo Nordisk, all outside the submitted work. All other co-authors don't report any conflict of interest.

Contributions

HMC, PMK and SJM created the concept and design of the manuscript. HMC, PMK, SJM and AW developed the first draft of the paper. EB and TMC conducted the literature review. And JB, SK, LAK, DMA, SH, SW, CB, NL, EWB, JM, JP, BK, RM, RL, EP, NS, JMK, SP, AC, JMM, ASP, DG contributed to the data acquisition. SJM, AW, TMC, JM, AJ, JOR, SH, AH, MC, PMK and HMC conducted the data analysis. All co-authors contributed to critically revising the manuscript for important intellectual content and all co-authors approved the final manuscript.

Acknowledgements

We thank all staff in critical care units who submitted data to the SICSAG database, the Scottish Morbidity Record Data Team, the staff of the National Register of Scotland, the Public Health Scotland Terminology Services, the HPS COVID-19 Laboratory & Testing cell and the NHS Scotland Diagnostic Virology Laboratories, and Nicola Rowan (HPS) for coordinating this collaboration.

Funding

There was no specific funder for this study.

Data validity

Stuart McGurnaghan and Helen Colhoun had full access to the data reported in this paper which they analysed and take responsibility for its validity.

References

[1] Guo W, Li M, Dong Y, Zhou H, Zhang Z, Tian C, et al. Diabetes is a risk factor for the progression and prognosis of COVID-19. Diabetes/Metabolism Research and Reviews 2020;36:e3319. doi:10.1002/dmrr.3319.

[2] Wan S, Xiang Y, Fang W, Zheng Y, Li B, Hu Y, et al. Clinical features and treatment of COVID-19 patients in northeast Chongqing. Journal of Medical Virology 2020. doi:10.1002/jmv.25783.

[3] Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. The Lancet 2020;395:497–506. doi:10.1016/S0140-6736(20)30183-5.

[4] Yang J, Zheng Y, Gou X, Pu K, Chen Z, Guo Q, et al. Prevalence of comorbidities and its effects in patients infected with SARS-CoV-2: A systematic review and meta-analysis. International Journal of Infectious Diseases: IJID: Official Publication of the International Society for Infectious Diseases 2020;94:91–5. doi:10.1016/j.ijid.2020.03.017.

[5] Rossi PG, Marino M, Formisano D, Venturelli F, Vicentini M, Grilli R, et al. Characteristics and outcomes of a cohort of COVID-19 patients in the Province of Reggio Emilia, Italy. PLOS ONE 2020;15:e0238281. doi:10.1371/journal.pone.0238281.

[6] Richardson S, Hirsch JS, Narasimhan M, Crawford JM, McGinn T, Davidson KW, et al. Presenting Characteristics, Comorbidities, and Outcomes Among 5700 Patients Hospitalized With COVID-19 in the New York City Area. JAMA 2020;323:2052–9. doi:10.1001/jama.2020.6775.

[7] Arentz M, Yim E, Klaff L, Lokhandwala S, Riedo FX, Chong M, et al. Characteristics and Outcomes of 21 Critically Ill Patients With COVID-19 in Washington State. JAMA 2020;323:1612–4. doi:10.1001/jama.2020.4326.

[8] Centers for Disease Control. Coronavirus Disease 2019 (COVID-19). Online; Accessed 11th Nov 2020;https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-at-higher-risk.html.

[9] Docherty AB, Harrison EM, Green CA, Hardwick HE, Pius R, Norman L, 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. doi:10.1136/bmj.m1985.

[10] Mancia G, Rea F, Ludergnani M, Apolone G, Corrao G. Renin–Angiotensin–Aldosterone System Blockers and the Risk of Covid-19. New England Journal of Medicine 2020. doi:10.1056/NEJMoa2006923.

[11] Williamson EJ, Walker AJ, Bhaskaran K, Bacon S, Bates C, Morton CE, et al. Factors associated with COVID-19-related death using OpenSAFELY. Nature 2020;584:430–6. doi:10.1038/s41586-020-2521-4.

[12] Bello-Chavolla OY, Bahena-López JP, Antonio-Villa NE, Vargas-Vázquez A, González-Díaz A, Márquez-Salinas A, et al. Predicting mortality due to SARS-CoV-2: A mechanistic score relating obesity and diabetes to COVID-19 outcomes in Mexico. The Journal of Clinical Endocrinology and Metabolism 2020. doi:10.1210/clinem/dgaa346.

[13] Holman N, Knighton P, Kar P, O'Keefe J, Curley M, Weaver A, et al. Type 1 and Type 2 Diabetes and COVID-19 Related Mortality in England: A Cohort Study in People with Diabetes. SSRN Electronic Journal 2020. doi:10.2139/ssrn.3605226.

[14] Barron E, Bakhai C, Kar P, Weaver A, Bradley D, Ismail H, et al. Associations of type 1 and type 2 diabetes with COVID-19-related mortality in England: a whole-population study. The Lancet Diabetes & Endocrinology 2020;8:813–22. doi:10.1016/S2213-8587(20)30272-2.

[15] NHS UK. Who's at higher risk from coronavirus (COVID-19). Online; accessed 11th Nov 2020;https://www.nhs.uk/conditions/coronavirus-covid-19/people-at-higher-risk/whos-at-higher-risk-from-coronavirus/.

[16] Cariou B, Hadjadj S, Wargny M, Pichelin M, Al-Salameh A, Allix I, et al. Phenotypic characteristics and prognosis of inpatients with COVID-19 and diabetes: The CORONADO study. Diabetologia 2020:1–16. doi:10.1007/s00125-020-05180-x.

[17] Livingstone SJ, Levin D, Looker HC, Lindsay RS, Wild SH, Joss N, et al. Estimated life expectancy in a Scottish cohort with type 1 diabetes, 2008-2010. JAMA 2015;313:37–44. doi:10.1001/jama.2014.16425.

[18] McAllister DA, Read SH, Kerssens J, Livingstone S, McGurnaghan S, Jhund P, et al. Incidence of Hospitalization for Heart Failure and Case-Fatality Among 3.25 Million People With and Without Diabetes Mellitus. Circulation 2018;138:2774–86. doi:10.1161/CIRCULATIONAHA.118.034986.

[19] Scottish Government. Scottish index of multiple deprivation. Online; Accessed 11th Nov 2020;https://www.gov.scot/collections/scottish-index-of-multiple-deprivation-2020/.

[20] World Health Organisation. WHO anatomical therapeutic chemical classification. Online; accessed 11th Nov 2020;https://www.whocc.no/atc_ddd_index/.

[21] National Records of Scotland. Scottish mid-year population estimates. Online; accessed 11th Nov 2020;https://www.nrscotland.gov.uk/statistics-and-data/statistics/statistics-by-theme/population/population-estimates/mid-year-

population-estimates.

[22] R Core Team. R: A language and environment for statistical computing. Online; Accessed 11th Nov 2020;https://www.R-project.org/.

[23] Honaker J, King G, Blackwell M. **Amelia** II: A Program for Missing Data. Journal of Statistical Software 2011;45. doi:10.18637/jss.v045.i07.

[24] Ambler, original by Gareth and Benner, modified by Axel. Mfp: Multivariable Fractional Polynomials. Online; Accessed 11th Nov 2020;https://CRAN.R-project.org/package=mfp.

[25] Royston P, Ambler G, Sauerbrei W. The use of fractional polynomials to model continuous risk variables in epidemiology. International Journal of Epidemiology 1999;28:964–74. doi:10.1093/ije/28.5.964.

[26] McKeigue P. Quantifying performance of a diagnostic test as the expected information for discrimination: Relation to the C-statistic: Statistical Methods in Medical Research 2018. doi:10.1177/0962280218776989.

[27] Our World in Data. Excess mortality during the Coronavirus pandemic (COVID-19). Online; accessed 11th Nov 2020;https://ourworldindata.org/excess-mortality-covid.

[28] Hamer M, Gale CR, Batty GD. Diabetes, glycaemic control, and risk of COVID-19 hospitalisation: Population-based, prospective cohort study. Metabolism: Clinical and Experimental 2020;112:154344. doi:10.1016/j.metabol.2020.154344.

[29] McKeigue PM, Weir A, Bishop J, McGurnaghan S, Kennedy S, McAllister D, et al. Rapid Epidemiological Analysis of Comorbidities and Treatments as risk factors for COVID-19 in Scotland (REACT-SCOT): A population-based case-control study. MedRxiv 2020:2020.05.28.20115394. doi:10.1101/2020.05.28.20115394.

[30] Haitao T, Vermunt JV, Abeykoon J, Ghamrawi R, Gunaratne M, Jayachandran M, et al. COVID-19 and Sex Differences. Mayo Clinic Proceedings 2020;95:2189–203. doi:10.1016/j.mayocp.2020.07.024.

[31] Hernán MA, Robins JM (2020). Causal inference: What if. boca raton: Chapman & hall/crc. Online; Accessed 11th Nov 2020;https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/.

[32] Anderson RM, Heesterbeek H, Klinkenberg D, Hollingsworth TD. How will country-based mitigation measures influence the course of the COVID-19 epidemic? The Lancet 2020;395:931–4. doi:10.1016/S0140-6736(20)30567-5.

[33] Occupational Medicine. Covid-age online toolkit The Society of Occupational Medicine. Online; Accessed 11th Nov 2020;https://www.som.org.uk/covid-age-online-toolkit.

[34] Clift AK, Coupland CAC, Keogh RH, Diaz-Ordaz K, Williamson E, Harrison EM, 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 (Clinical Research Ed) 2020;371:m3731. doi:10.1136/bmj.m3731.

[35] World Health Organization. Clinical management of COVID-19. Online; Accessed 11th Nov 2020;https://www.who.int/publications/i/item/clinical-management-of-covid-19.

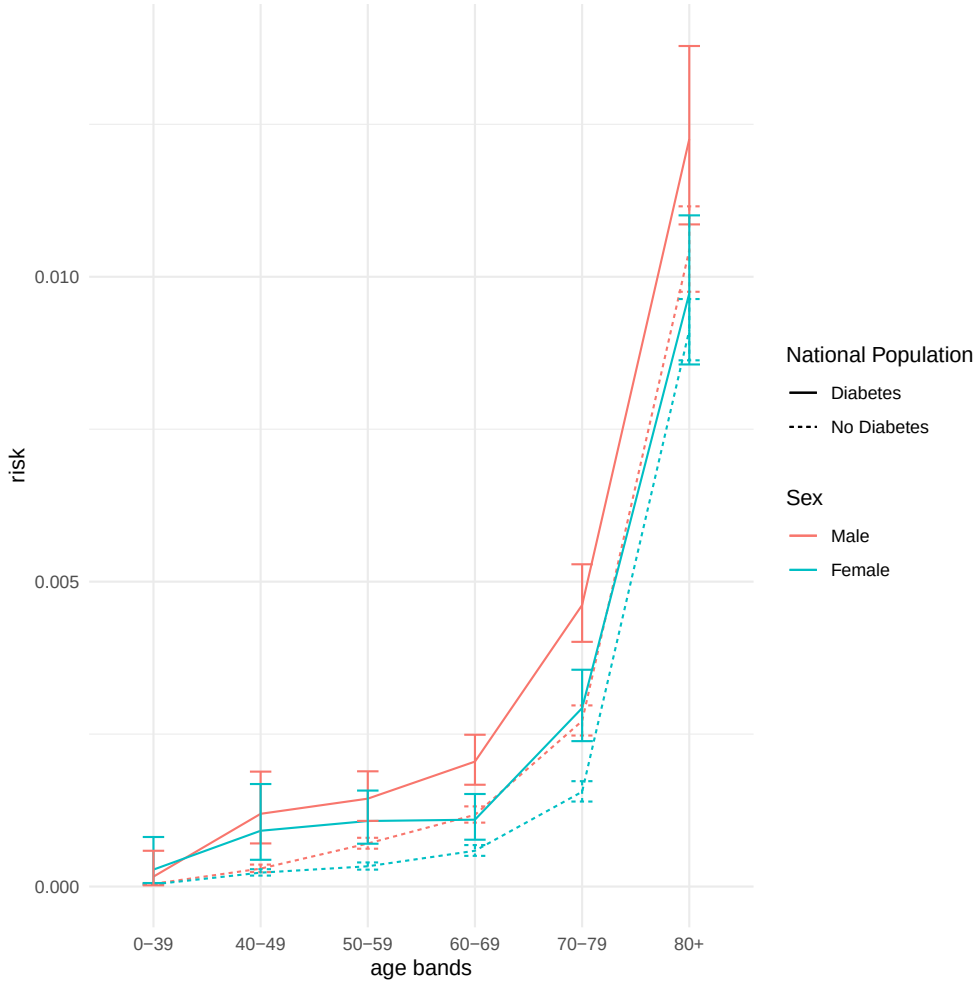


Figure 1: Risk of F/CCU-COVID-19 in the National population of Scotland with and without diabetes by age band and sex by 31st July 2020. High/low bars indicate the 95% confidence interval of the risk

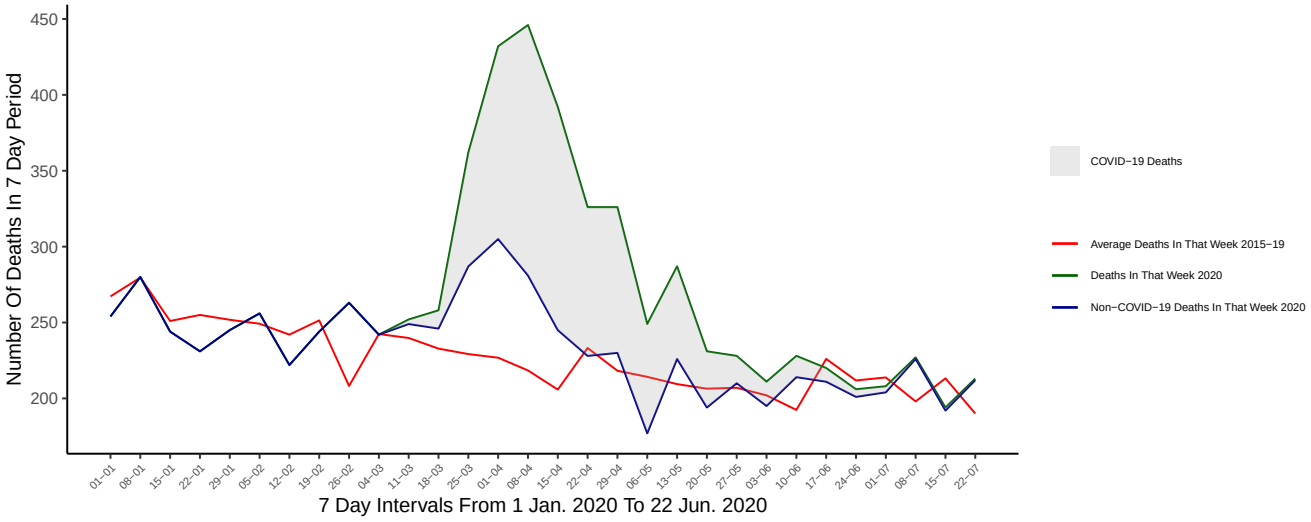


Figure 2: Weekly deaths from from all causes and from causes other than COVID-19 in people with Diabetes in Scotland during 2020 compared with average deaths in that week from 2015 to 2019

Table 1: Cumulative incidence of F/CCU-COVID-19 in the Scottish population with diabetes, by age, sex and diabetes type by 31st July 2020

Age Band (years)	0-39	40-49	50-59	60-69	70-79	80+	Total
All	22264	24863	58438	81606	80909	51269	319349
F/CCU-COVID-19	5(0.0)	25(0.1)	80(0.1)	134(0.2)	306(0.4)	532(1.0)	1082(0.3)
Male	11821	14402	34968	49001	46201	24093	180486
F/CCU-COVID-19	2(0.0)	15(0.1)	54(0.2)	99(0.2)	206(0.4)	281(1.2)	657(0.4)
Female	10443	10461	23470	32605	34708	27176	138863
F/CCU-COVID-19	3(0.0)	10(0.1)	26(0.1)	35(0.1)	100(0.3)	251(0.9)	425(0.3)
Type 1	14732	5747	6333	4486	2227	858	34383
F/CCU-COVID-19	2(0.0)	5(0.1)	10(0.2)	7(0.2)	14(0.6)	13(1.5)	51(0.1)
Type 2	6507	18072	50273	75031	76792	49285	275960
F/CCU-COVID-19	2(0.0)	17(0.1)	68(0.1)	125(0.2)	285(0.4)	511(1.0)	1008(0.4)
Other	1025	1044	1832	2089	1890	1126	9006
F/CCU-COVID-19	1(0.1)	3(0.3)	2(0.1)	2(0.1)	7(0.4)	8(0.7)	23(0.3)

Data are shown in N or N(%)

Table 2: Characteristics of all those with diabetes in Scotland who did and did not develop F/CCU-COVID-19 by 31st July 2020

	No F/CCU-COVID-19	F/CCU-COVID-19	Total diabetes population
Total included	318267(99.66)	1082(0.34)	319349
Sociodemographic			
Age (years)	66.7(56.3,75.8)	79.9(71.4,85.7)	66.7(56.3,75.8)
Diabetes duration (years)	10.5(5.7,16.6)	13.5(8.0,19.2)	10.5(5.7,16.6)
Carehome resident	5897(1.9)	397(36.7)	6294(2.0)
Deprivation index			
Quintile 1 (most deprived)	73188(23.0)	322(29.8)	73510(23.0)
Quintile 2	71102(22.3)	264(24.4)	71366(22.3)
Quintile 3	63401(19.9)	188(17.4)	63589(19.9)
Quintile 4	56203(17.7)	173(16.0)	56376(17.7)
Quintile 5 (least deprived)	46251(14.5)	102(9.4)	46353(14.5)
Unknown	8122(2.6)	33(3.0)	8155(2.6)
Ethnicity			
White	237205(74.5)	870(80.4)	238075(74.6)
South Asian	9218(2.9)	16(1.5)	9234(2.9)
Black	1589(0.5)	5(0.5)	1594(0.5)
Chinese	1205(0.4)	4(0.4)	1209(0.4)
Other	12103(3.8)	30(2.8)	12133(3.8)
Unknown	56947(17.9)	157(14.5)	57104(17.9)
Comorbidities			
Any DKA admission in past 5 years	6623(2.1)	23(2.1)	6646(2.1)
Any hypoglycaemia admission in past 5yrs	5769(1.8)	73(6.7)	5842(1.8)
No. of other hospital admissions in past 5 years	1.0(0.0,3.0)	5.0(2.0,11.0)	1.0(0.0,3.0)
Any heart disease	100482(31.6)	696(64.3)	101178(31.7)
Asthma or chronic lower airway disease	105066(33.0)	504(46.6)	105570(33.1)
Neurological and dementia (excluding epilepsy)	15076(4.7)	232(21.4)	15308(4.8)
Liver disease	3075(1.0)	29(2.7)	3104(1.0)
Immune disease or on immunosuppressants	4078(1.3)	24(2.2)	4102(1.3)
Any listed condition	165813(52.1)	896(82.8)	166709(52.2)
Other clinical measures			
Insulin pump use	4811(1.5)	1(0.1)	4812(1.5)
Flash glucose monitor use	11711(3.7)	6(0.6)	11717(3.7)
HbA1c (mmol/mol)	57(49,70)	58(47,71)	57(49,70)
BMI (kg/m^2)	30(27,35)	29(25,33)	30(27,35)
Systolic BP (mmHg)	134(124,142)	132(122,142)	134(124,142)
Diastolic BP (mmHg)	77(70,82)	74(67,80)	77(70,82)
Total cholesterol (mmol/L)	4(4,5)	4(3,5)	4(4,5)
eGFR (mL/min/1.73m^2)	83(65,97)	64(44,82)	83(65,97)
Albuminuric status			
Normal	131192(41.2)	300(27.7)	131492(41.2)
Micro	55417(17.4)	235(21.7)	55652(17.4)
Macro	11353(3.6)	77(7.1)	11430(3.6)
Unknown	120305(37.8)	470(43.4)	120775(37.8)
Retinopathy			
None	200428(63.0)	618(57.1)	201046(63.0)
Non referable	48624(15.3)	160(14.8)	48784(15.3)
Referable / eye clinic	28170(8.9)	134(12.4)	28304(8.9)
Unknown	41045(12.9)	170(15.7)	41215(12.9)
Tobacco smoking status			
Current smoker	50734(15.9)	111(10.3)	50845(15.9)
Ex smoker	153181(48.1)	679(62.8)	153860(48.2)
Never smoked	111292(35.0)	287(26.5)	111579(34.9)
Unknown	3060(1.0)	5(0.5)	3065(1.0)
Drug exposures			
Lipid lowering	210701(66.2)	806(74.5)	211507(66.2)
Proton Pump Inhibitors	132581(41.7)	582(53.8)	133163(41.7)
NSAIDS	143947(45.2)	698(64.5)	144645(45.3)
Anti-coagulants anti-platelets	112983(35.5)	667(61.6)	113650(35.6)
Antihypertensives (any)	198117(62.2)	713(65.9)	198830(62.3)
Number of ATC level 3 drug classes (excluding DM)	8.0(4.0,12.0)	11.0(8.0,15.0)	8.0(4.0,12.0)
Number of DM drug classes prescribed	1.0(1.0,2.0)	1.0(1.0,2.0)	1.0(1.0,2.0)

Data are shown in N(%) for categorical values and median interquartile range for continuous values

Table 3: Logistic regression of association of characteristics with having F/CCU-COVID-19 in people with diabetes. Diabetes duration was adjusted for age. Sex and diabetes type were adjusted for age and diabetes duration. All other associations were adjusted for age, sex, diabetes duration and diabetes type.

Predictor	Odds Ratio	95%CI	P-Value
Age (years)	1.076	(1.071, 1.082)	<0.001
Sex			(global) <0.001
Male	1 (reference)		
Female	0.705	(0.623, 0.798)	<0.001
Diabetes type			(global) 0.693
Type 2 diabetes	1 (reference)		
Type 1 diabetes	1.087	(0.789, 1.498)	0.610
Other diabetes types	0.869	(0.574, 1.317)	0.509
Diabetes duration (years)	1.016	(1.009, 1.022)	<0.001
Carehome resident	16.570	(14.326, 19.165)	<0.001
Any hypoglycaemia admission in past 5yrs	3.178	(2.480, 4.072)	<0.001
Deprivation index			(global) <0.001
Quintile 1 (most deprived)	1 (reference)		
Quintile 2	0.732	(0.622, 0.862)	<0.001
Quintile 3	0.545	(0.455, 0.653)	<0.001
Quintile 4	0.556	(0.462, 0.669)	<0.001
Quintile 5 (least deprived)	0.379	(0.303, 0.473)	<0.001
Ethnicity			(global) 0.086
White	1 (reference)		
South Asian	0.616	(0.368, 1.033)	0.066
Black	1.770	(0.727, 4.311)	0.209
Chinese	0.784	(0.267, 2.295)	0.656
Other	0.740	(0.513, 1.066)	0.106
Any DKA admission in past 5 years	2.869	(1.846, 4.460)	<0.001
Any hypoglycaemia admission in past 5yrs	3.178	(2.480, 4.072)	<0.001
Ever admitted to hospital in past 5 years	3.307	(2.789, 3.922)	<0.001
Any heart disease	2.425	(2.135, 2.754)	<0.001
Asthma or chronic lower airway disease	1.691	(1.500, 1.907)	<0.001
Neurological and dementia (excluding epilepsy)	3.810	(3.284, 4.421)	<0.001
Liver disease	3.021	(2.082, 4.384)	<0.001
Immune disease or on immunosuppressants	2.334	(1.552, 3.510)	<0.001
Any listed condition	3.167	(2.701, 3.713)	<0.001
Insulin pump use	0.330	(0.046, 2.372)	0.271
Flash glucose monitor use	0.414	(0.176, 0.973)	0.043
HbA1c (mmol/mol)	1.010	(1.006, 1.014)	<0.001
BMI (kg/m^2)	1.002	(0.991, 1.013)	0.706
Systolic BP (mmHg)	0.986	(0.982, 0.990)	<0.001
Diastolic BP (mmHg)	0.994	(0.987, 1.001)	0.074
Total cholesterol (mmol/L)	1.035	(0.974, 1.100)	0.267
eGFR ($mL/min/1.73m^2$)	0.992	(0.988, 0.995)	<0.001
Albuminuric grade			(global) <0.001
Normal	1 (reference)		
Micro	1.352	(1.155, 1.583)	<0.001
Macro	1.922	(1.519, 2.430)	<0.001
Retinopathy grading			(global) <0.001
None	1 (reference)		
Non referable	1.161	(0.975, 1.382)	0.094
Referable or eye clinic	1.672	(1.377, 2.032)	<0.001
Tobacco smoking			(global) 0.001
Never smoked	1 (reference)		
Ex smoker	1.296	(1.126, 1.491)	<0.001
Current smoker	1.133	(0.907, 1.416)	0.270
Any lipid lowering	1.126	(0.981, 1.293)	0.091
Any proton pump inhibitor	1.412	(1.252, 1.593)	<0.001
Any NSAID	1.848	(1.630, 2.097)	<0.001
Any anticoagulants	1.663	(1.466, 1.887)	<0.001
Any antihypertensive	0.801	(0.705, 0.909)	0.001
Number of ATC level 3 drug classes (excluding DM)	1.079	(1.068, 1.091)	<0.001
Number of DM drug classes prescribed	1.139	(1.083, 1.199)	<0.001

Table 4: Stepwise logistic regression of association of characteristics with F/CCU-COVID-19 in people with diabetes.

Predictor	Odds Ratio	95% CI	P-Value
Age (years)	1.044	(1.036, 1.051)	<0.001
Sex			(global) <0.001
Male	1 (reference)		
Female	0.535	(0.470, 0.608)	<0.001
Diabetes type			(global) 0.618
Type 2 diabetes	1 (reference)		
Type 1 diabetes	1.119	(0.806, 1.553)	0.501
Other diabetes types	0.866	(0.567, 1.321)	0.504
Diabetes duration (years)	0.998	(0.990, 1.006)	0.595
Carehome resident	10.828	(9.251, 12.675)	<0.001
Deprivation index			(global) <0.001
Quintile 1 (most deprived)	1 (reference)		
Quintile 2	0.848	(0.718, 1.002)	0.052
Quintile 3	0.619	(0.514, 0.744)	<0.001
Quintile 4	0.656	(0.542, 0.793)	<0.001
Quintile 5 (least deprived)	0.484	(0.385, 0.607)	<0.001
log(No. of other hospital admissions + 1 (5yrs))	1.595	(1.481, 1.717)	<0.001
Neurological and dementia (excluding epilepsy)	1.273	(1.081, 1.499)	0.004
HbA1c (mmol/mol)	1.005	(1.001, 1.009)	0.008
BMI (kg/m^2)	1.091	(1.047, 1.136)	<0.001
log(BMI (kg/m^2))	0.080	(0.022, 0.291)	<0.001
eGFR ($mL/min/1.73m^2$)	0.992	(0.989, 0.995)	<0.001
Systolic BP (mmHg)	0.994	(0.990, 0.998)	0.004
Any antihypertensive	0.792	(0.687, 0.913)	0.001
Number of DM drug classes prescribed	1.065	(1.004, 1.129)	0.036
Number of ATC level 3 drug classes (excluding DM)	1.027	(1.013, 1.041)	<0.001

Age, sex, diabetes duration and type of diabetes were entered as the baseline model. The remaining variables were retained by the stepwise procedure using the Akaike information criterion. The C-Statistic (95%CI) for the base and full models were 0.76(0.75,0.77) and 0.85(0.83,0.86) respectively. The expected information for discrimination Λ was 0.75 bits for the base model and 1.54 for the full model. Model coefficients are provided in supplementary table 7

Supplementary materials

Supplement Methods

Derivation of variables used from the SCI-Diabetes research platform

Clinician assigned diabetes type in SCI-Diabetes was accepted unless contradicted by available data on age at diagnosis and prescription history. Those who were assigned type 1 were reassigned as type 2 if they had a contradictory drug history and age at diagnosis >40 years. A prescription history that contradicts type 1 diabetes was more than 1-year of non-metformin oral diabetes drug use or more than a 1-year interval from diagnosis to insulin. This rule reassigned 10.8% of type 1 diabetes as type 2. Those who were initially assigned as type 2 were reassigned to type 1 only if they have insulin from diagnosis and an age of onset below age 30 years (0.8% were reassigned). Smoking was categorised according to the categories in the electronic health record front end to SCI-Diabetes; never/current/ex or unknown. Albuminuria was defined based on urinary albumin concentration in urines collected at routine clinic visits and categorised as normo-, micro- or macroalbuminuric according to the albumin/creatinine ratio (ACR) falling in the intervals 0–3.39 mg/mmol, 3.39–33.9 mg/mmol, or above 33.9 mg/mmol, based on two out of three most recent consecutive measurements. Retinopathy was defined based on grading of most recent retinal image at screening and categorised as none; non referable which is mild background; referable or at clinic which is pre-proliferative or worse retinopathy or maculopathy or prior referral to eye clinic because of retinopathy.

The Scottish Index of Multiple Deprivation (SIMD) ranks geographic areas by a single index capturing seven domains of deprivation and is summarised in quintiles. Here, for those with diabetes postcode of residence obtained from the CHI database was assigned to the SIMD quintile it belonged and was used as a measure of socioeconomic status. Residential care home status for those with diabetes was available via the care home flag for postal address in the CHI database

ICD-10 Hospital Discharge codes from Scottish Morbidity Record-01 and ATC Drug codes from Prescribing Information System used for deriving "Listed conditions"

Grouping	BNF	ATC regex	ICD10	ICD10 Regex	OPCS4 Code	OPCS4 Regex
Heart disease						
Ischaemic heart disease	2.6.1 Nitrates 2.6.3 Other antianginal drugs	^V03AB22|^C01DA(0[28]|14)|^C05AE0[12]|^D03AX0[78]|^C01DX16|^C01EB1[78]	I20-I25	^I20|^I21|^I22|^I23|^I24|^I25 Collapsed: ^I2[0-5]	K40-K44 CABG K49 K59 PTCA	Simple regex: ^K40|^K41|^K42|^K43|^K44|^K49|^K59 Collapsed: ^K(4[0-49]|59)
Other heart disease	2.3 Anti-arrhythmic drugs	^C01BA0[1-3]|^C01BB0[1-3]|^C01BC0[34]|^C01BD0[127]|^C01DB01|^C01EB10|^D04AB01|^N01BB02|^RA02AD02	I00-102 I05-I09 I10-I15 I26-I28 I30-I52	^I00|^I01|^I02|^I05|^I06|^I07|^I08|^I09|^I10|^I11|^I12|^I13|^I14|^I15|^I12|^I12|^I12|^I30|^I31|^I32|^I33|^I34|^I35|^I36|^I37|^I38|^I39|^I40|^I41|^I42|^I43|^I44|^I45|^I46|^I47|^I48|^I49|^I50|^I51|^I52 Collapsed: ^I0[01256789]|^I1[0-5]|^I2[6-8]|^I3[0-9]|^I4[0-9]|^I5[0-2]	K57 L16 L18-L23 L25-L27 L45-54 L56-L58 L60 L62 L63 L29-L31 X09-X11	Collapsed: ^K57|^L1[689]|^L2[0123567]|^L4[56789]|^L5[012346789]|^L6[023]|^L29|^L3[01]|^X09|^X1[01]
Nervous system other than G00-G09, G40-G46 and G50-G59	4.9 Drugs used in parkinsonism and related disorders 4.11 Dementia	^G02CB03|^G04BE07|^M03AX01|^N04AA0[124|^N04AC01|^N04AB02|^N04BA0[12]|^N04BC0[245679]|^N04BD0[123]|^N04BX0[124]|^N04BB01|^N04BX02|^N04BA03|^N07XX0[268]|^N05AK01|^N06BX03|^N06DA0[234]|^N06DX01	G10-G14 Systemic atrophies primarily affecting the central nervous system G20-G26 Extrapyramidal and movement disorders G30-G32 Other degenerative diseases of the nervous system G35-G37 Demyelinating diseases of the central nervous system G60-64 Polyneuropathies G70-G73 Diseases of myoneural junction and muscle G80-G83 Cerebral palsy and other paralytic syndromes	^G10|^G11|^G12|^G13|^G14|^G20|^G21|^G22|^G23|^G24|^G25|^G26|^G30|^G31|^G32|^G35|^G36|^G37|^G70|^G71|^G72|^G73|^G80|^G81|^G82|^G83|^G90|^G91|^G92|^G93|^G94|^G95|^G96|^G97|^G98|^G99 Collapsed: ^G1[0-4]|^G2[0-6]|^G3[012567]|^G6[		

			G90-G99 Other disorders of the nervous system	0-4]|^G7[0-3]|^G8[0-3]|^G9[0-9]		
Asthma and chronic lower respiratory disease						
	3.1, 3.2, 3.3	^A0(3AB02|7E(A0[67]|B))|^C01CA2|^D(07AC1[357]|11AA)|^R0(1A(A(03)|C01|D0[1589]|X03)|3([AC]|B[A-C]|D([AC]|X07)))|^S01(EA01|GX0[14])	J45 = Asthma J46 = Status asthmaticus	^J45|^J46 Collapsed: ^J4[56]		
			J40 = Bronchitis, not specified as acute or chronic J41 = Simple and mucopurulent chronic bronchitis J42 = Unspecified chronic bronchitis J43 = Emphysema J44 = Other chronic obstructive pulmonary disease J47 = Bronchiectasis J60-J70 Lung diseases due to external agents J80-J84 Other respiratory diseases principally affecting the interstitium J85-J86 Suppurative and necrotic conditions of lower respiratory tract J90-J94 Other diseases of pleura J95-J99 Other diseases of the respiratory system G473 = Sleep apnoea	J40|^J41|^J42|^J43|^J44|^J47|^J60|^J61|^J62|^J63|^J64|^J65|^J66|^J67|^J68|^J69|^J70|^J80|^J81|^J82|^J83|^J84|^J85|^J86|^J90|^J91|^J92|^J93|^J94|^J95|^J96|^J97|^J98|^J99|^G47\.?3 Collapsed: ^J4[012347]|^J6[0-9]|^J70|^J8[0-6]|^J9[0-9]|^G47\.?3		

Chronic liver disease including hepatocellular ca						
	Nothing specific available		C22.0 hepatocellular carcinoma I85.0 Oesophageal varices I98.3 Oesophageal varices K70.2 Alcoholic fibrosis and sclerosis of liver K70.3 Alcoholic cirrhosis of liver K70.4 Alcoholic hepatic failure K71.7 Toxic liver disease with fibrosis and cirrhosis of liver K72.0 Acute and subacute failure of the liver K72.1 Chronic hepatic failure K72.9 Hepatic coma K73 Chronic hepatitis, not elsewhere classified K74.0 Hepatic fibrosis K74.2 Hepatic fibrosis with hepatic sclerosis K74.3 Primary biliary cirrhosis K74.4 Secondary biliary cirrhosis K74.5 Biliary cirrhosis, unspecified K74.6 Other and unspecified cirrhosis of liver K76.7 Hepatorenal syndrome R18 Ascites	^C22\.?0|^I85\.?0|^I98\.?3|^K70\.?2|^K70\.?3|^K70\.?4|^K71\.?7|^K72\.?0|^K72\.?1|^K72\.?9|^K73|^K74\.?0|^K74\.?2|^K74\.?3|^K74\.?4|^K74\.?5|^K74\.?6|^K76\.?7|^R18		
Chronic kidney disease						
CKD 4 or worse	Nothing specific available		N18.4 Chronic kidney disease stage 4 N18.5 Chronic kidney disease stage 5 Z490 = Care involving dialysis Z491 = Care involving dialysis Z492 = Care involving dialysis	^N18\.?4|^N18\.?5|^Z49\.?0|^Z49\.?1|^Z49\.?2|^Z94\.?0|^Z99\.?2	M011 Autotransplantation of kidney M012 Allotransplantation of kidney from live donor	^M01\.?1|^M01\.?2|^M01\.?3|^M01\.?4|^M01\.?5|^M01\.?8|^M01\.?9

			Z940 = Kidney transplant status Z992 = Dependence on renal dialysis		M013 Allotransplantation of kidney from cadaver nec M014 Allotransplantation of kidney from cadaver heart-beating M015 Allotransplantation of kidney from cadaver non-heart-beating M018 Other specified transplantation of kidney M019 Unspecified transplantation of kidney	
Immunodeficiency and immunosuppression						
	5.3.1 hiv 8.2 drugs affecting immune response	^J05A([GR]|E(0|10)|F0[1-79]|X(0[7-9]|1[128]))|^V03AX|^D11A(H01|X14)|^L01(BB04|XC0[24]|XC1[57])|^L01XE34|^L03AB((0[34578])|(1[013]))|^L03AC01|^L03AX(03|13|14)|^L04AA(0[145689]|10|23|27|28|31|3442)|^L04AC(0[128]|11)|^L04AD0[12]|^L04AX0[1247]|^N07XX09|	B20 –B23 = Human immunodeficiency virus [HIV] disease	^B20|^B21|^B22|^B23|^B24		

		^S01XA18|^S01XA23				
			D80-D89 Certain disorders involving the immune system	^D8[0-9]		
Any lipid drug						
	2.12 Lipid-regulating drugs	^A06AC01|^C10				
Any anticoagulant or antiplatelet drug						
	2.8 Anticoagulants and protamine 2.9 Antiplatelet drugs	^B01AA|^B01AB|^B01AC0|^B01AC1[36]|^B01AC2[24]|^B01AE0[27]|^B01AF0[123]|^B01AX0[56]|^G04BX15|^N02BA01|^S01XA14|^V03AB14|^V03AB37				
Any anti-hypertensive drug						
	2.5.1 Vasodilator antihypertensive drugs	^B01AC11|^C02DA01|^C02DB02|^C02DC01|^C02KX0[1-5]|^D11AX01|^G04BE0[38]|^V03AH01				
	2.5.2 Centrally-acting antihypertensive drugs	^C02AB01|^C02AC0[125]|^N02CX02				

	2.5.3 Adrenergic neurone blocking drugs	^C02CC0[24]				
	2.5.4 Alpha-adrenoceptor blocking drugs	^C02CA0[124]|^C04AB01|^C04AX02|^G04CA03				
	2.5.5.1 Angiotensin-converting enzyme inhibitors	^C09A|^C09B				
	2.5.5.2 Angiotensin-II receptor antagonists	^C09C|^C09D				
	2.5.5.3 Renin inhibitors	^C09XA02$				
	2.2.1 Thiazides and related diuretics	^C03AA|^C03BA				
	2.6.2 Calcium-channel blockers	^C01EB15|^C05AE03|^C08				
Any NSAID						
	10.1.1 Non-steroidal anti-inflammatory drugs	^C01EB16|^M01AA01|^M01AB0[1258]|^M01AB1[16]|^M01AB55|^M01AC0[1256]|^M01AE0[123459]|^M01AE1[147]|^M01AE5[12]|^M01AG01|^M01AH0[12356]|^M01AX0[14]|^M02AA07|^M02AA1[035]|^N02AJ08|^N02BA1[01]|^N02BE01|^R02AX01|^S01BC0[34]				
Any proton pump inhibitors						
	1.3.5 Proton pump inhibitors	^A02BC				

Missing data imputation

We used the Amelia II package in R to impute missing values for continuous and categorical variables.[1] This package uses an EMB (bootstrapped expectation maximization) algorithm for imputation under the assumption that data are missing at random. Here we used it to generate 10 imputed datasets and for each individual the mean of the imputed values across these 10 datasets was then used in subsequent analyses including the means of the indicator (dummy) variables for categorical variables.

Selection of the functional form for continuous variables in the stepwise prediction model

We cannot assume that treating continuous variables as linear in regression will maximize prediction. We therefore used the R package *mfp* to find the functional form for continuous variables being entered into the stepwise prediction model.[2,3] Mfp uses a backwards elimination step with an adaptive algorithm to select the best fractional polynomial transformation for each continuous variable. The resulting transformed variables were then made available to the *step* procedure.

Further details of stepwise regression for building the prediction model for F-/CCU-COVID-19

For building the predictive model of F/CCU-COVID-19 we used stepwise regression, alternating between forward and backward steps, implemented in the R function stats::step to select across the 35 variables listed in Table 2 as well as terms for age:sex, age:diabetes type, sex:diabetes type. For variable selection the *step* function maximizes the Akaike Information Criterion (AIC)[4]. The AIC is a trade off between the goodness of fit of a model and its simplicity or sparsity. We set a stringent penalty of $4p$ *-2* log likelihood where p is the number of parameters.

Assessment of model predictive performance

The predictive performance of the base model and then the model incremented with diabetes specific factors was evaluated by 20-fold cross validation. The dataset was split randomly into 20 equal sized disjoint test folds. For each test fold, the remaining observations constitute the training fold. The predictive performance of the model learned on each training fold was then evaluated on the corresponding test fold. The predictive performance was summarised over the entire dataset as the C-statistic and as the expected information for discrimination (lambda) using the wevid package[5,6].

Sensitivity analyses of the final predictive model

We tested whether using a co-morbidity index (the Charlson Index)[7] rather than the individual co-morbidities we considered improved prediction. In addition we evaluated whether an improved predictive performance was obtained using a LASSO penalized regression but not forcing any baseline covariates into the model. We also tested whether modelling only fatal cases resulted in any change in the variables selected or predictive performance.

Derivation of COVID-age

The risk prediction model for F/CCU-COVID-19, when applied to an individual with diabetes with any given set of covariate values, outputs the logit of the probability (i.e. the log odds) of F/CCU-COVID-19 by 31st July 2020. For the Scottish population without diabetes, we fitted a generalised additive model (GAM) with a logistic link function and binomial likelihood to the one-year age and sex-specific counts of F/CCU-COVID-19. This GAM model returns the logit of probability of having developed F/CCU-COVID-19 by 31st July 2020 in the non-diabetic population at any given age and sex and given the risk factor distribution in that population. The COVID-age for a person with diabetes is then defined as the age at which the logit of the probability of disease in a non-diabetic individual of the same sex equates to the logit of probability in the diabetic individual under study. To enable a user to calculate this COVID-age for a person with diabetes and a given set of characteristics, we generated a Shiny application located at https://diabepi.shinyapps.io/covidrisk/ (Online; Accessed 11th November 2020). The purpose of the Shiny app to give the reader a sense of how individual risk factor profiles in those with diabetes translate into elevated risks compared to those without diabetes.

[1]https://cran.r-project.org/web/packages/Amelia/vignettes/amelia.pdf Online; Accessed 11th November 2020

[2]https://cran.r-project.org/web/packages/mfp/mfp.pdf Online; Accessed 11th November 2020

[3]Sauerbrei W, Royston P (1999) Building multivariable prognostic and diagnostic models: transformation of the predictors by using fractional polynomials. Journal of the Royal Statistical Society (Series A) 162: 71–94. Online; Accessed 11th November 2020

[4] https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/step Online; Accessed 11th November 2020

[5] https://cran.r-project.org/web/packages/wevid/index.html Online; Accessed 11th November 2020

[6]McKeigue P. Quantifying performance of a diagnostic test as the expected information for discrimination:Relation to the C-statistic: Statistical Methods in Medical Research 2018. doi:10.1177/0962280218776989.

[7]Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis. 1987;40(5):373-83. doi: 10.1016/0021-9681(87)90171-8. PMID: 3558716. Online; Accessed 11th November 2020

Supplemental results

Source of ascertainment of the cases

Supplement Figure 1 shows the source of ascertainment of the cases; Of all those with F/CCU-COVID-19, there had been a positive test in 71% (768/1082), 3.6% (39/1082) were ascertained from hospital discharge without a prior positive test and 25% (275/1082) were ascertained only from a COVID-19 mention on death certification. Almost all those with any evidence of COVID-19 were evaluable for F/CCU-COVID-19 status in that there were only 26 of those declared as not having had CCU or died that were still within 28 days of diagnosis by time of data extraction. Of those defined as having had a COVID-19 death, just 8/963 were people who died within 28 days of diagnosis but did not have a COVID-19 mention on death certificate.

Variable missingness

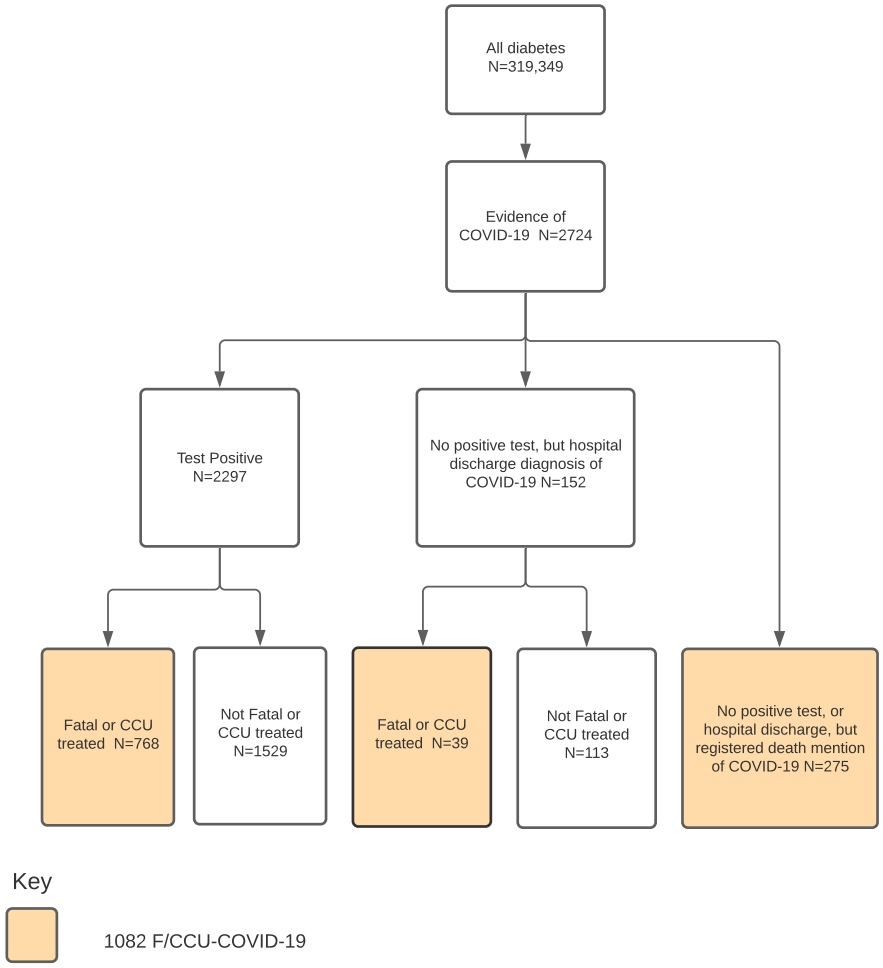
Only ethnicity (14.5%) and albuminuria (37.8%) had substantial missingness (Table 2). For continuous variables missingness was as follows: HbA1c 6%, BMI 10%, SBP 7%, DBP 7%, Total cholesterol 9% BMI 10.2% (Supplement Table 4). BMI was the only variable with substantial differences in missingness between those who did (17.4%) and did not (10.2%) develop F/CCU-COVID-19.

Prediction model sensitivity analyses

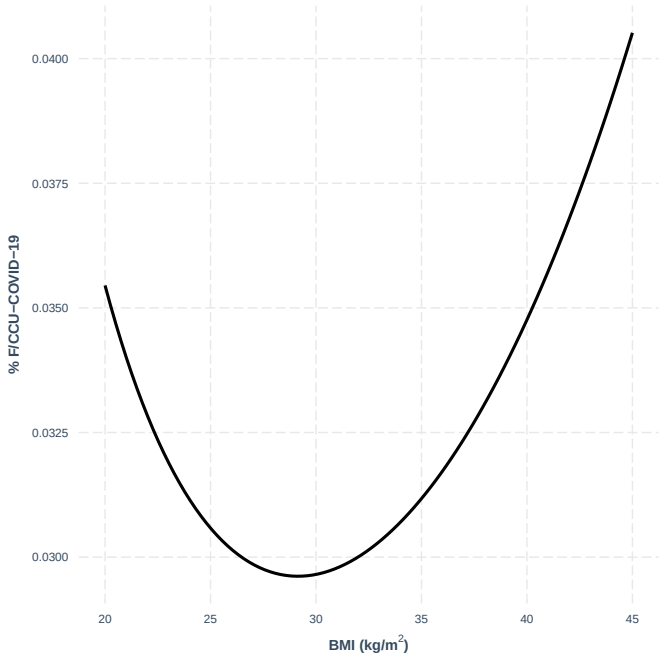
Using LASSO regression gave a similar prediction performance to the *step* procedure (C-statistic 0.836) but was much less sparse than the stepwise model therefore not the preferred model. Feeding the Charlson comorbidity index, rather than each individual listed conditions, to the model did not improve prediction performance (C-statistic 0.841) and was also less sparse. When the modelling was restricted to fatal cases only a similar pattern of associations was seen for each variable, the same variables were retained in the prediction model and it had a similar predictive performance (C-statistic 0.87).

Glossary

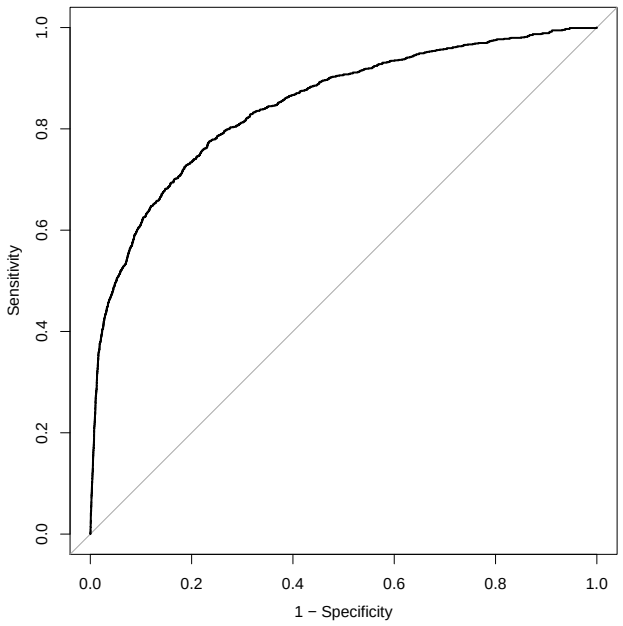
SCI-Diabetes : Scottish Care Information Diabetes system of clinical and network applications
NSAIDs : Non steroidal anti-inflammatory drugs
SGLT2 inhibitors : Sodium glucose transport 2 inhibitors
PPI : Proton pump inhibitors
ATC : Anatomical theraputical classification system
GLP1 : Glucagon-like peptide 1 agonists
DPP4 : Dipeptidyl peptidase-4
TZD : Thiazolidinediones
RRT : Renal replacement therapy
BMI : Body mass index
HbA1c : Glycated haemoglobin
eGFR : Estimated glomerular filtration rate



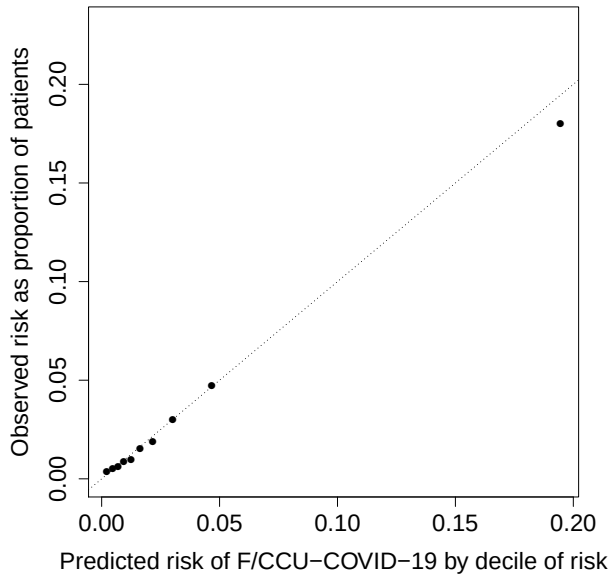
Supplementary Figure 1: Study flow diagram of ascertainment of F/CCU-COVID-19 as of 31st July 2020



Supplementary Figure 2: Relationship of BMI with F/CCU-COVID-19 as modelled by the 'mfp' R package showing the estimated risk of F/CCU-COVID-19 on the y-axis for BMI values on the x-axis



Supplementary Figure 3: ROC curve detailing sensitivity and specificity of the F/CCU-COVID-19 predictive model in people with diabetes (C-statistic 0.85)



Supplementary Figure 4: Calibration plot showing predicted versus observed risk of F/CCU-COVID-19 by decile of risk

Supplementary Table 1: Risk of F/CCU-COVID-19 in the National population of Scotland age with and without diabetes by age band and sex

Sex	National Population	0-39	40-49	50-59	60-69	70-79	80+	Total
All	No Diabetes (Total)	2576253	654864	733937	562011	395025	221861	5143951
All	No Diabetes F/CCU-COVID-19 (N %)	104 (0.00%)	170 (0.03%)	374 (0.05%)	484 (0.09%)	817 (0.21%)	2132 (0.96%)	4081 (0.08%)
All	Diabetes (Total)	23106	26026	60276	82126	79782	48033	319349
All	Diabetes F/CCU-COVID-19 (N %)	5 (0.02%)	28 (0.11%)	78 (0.13%)	137 (0.17%)	310 (0.39%)	524 (1.09%)	1082 (0.34%)
Male	No Diabetes (Total)	1298091	316511	348638	262021	174951	82305	2482517
	No Diabetes F/CCU-COVID-19 (N %)	58 (0.00%)	93 (0.03%)	246 (0.07%)	308 (0.12%)	475 (0.27%)	859 (1.04%)	2039 (0.08%)
	Diabetes (Total)	12317	15091	36077	49291	45274	22436	180486
	Diabetes F/CCU-COVID-19 (N %)	2 (0.02%)	18 (0.12%)	52 (0.14%)	101 (0.20%)	209 (0.46%)	275 (1.23%)	657 (0.36%)
Female	No Diabetes (Total)	1278162	338353	385299	299990	220074	139556	2661434
	No Diabetes F/CCU-COVID-19 (N %)	46 (0.00%)	77 (0.02%)	128 (0.03%)	176 (0.06%)	342 (0.16%)	1273 (0.91%)	2042 (0.08%)
	Diabetes (Total)	10789	10935	24199	32835	34508	25597	138863
	Diabetes F/CCU-COVID-19 (N %)	3 (0.03%)	10 (0.09%)	26 (0.11%)	36 (0.11%)	101 (0.29%)	249 (0.97%)	425 (0.31%)

Supplementary Table 2: Logistic regression of those with diabetes versus those without diabetes developing F/CCU-COVID-19 by 31st July 2020

Group	DM F/CCU-COVID-19	DM no F/CCU-COVID-19	No DM F/CCU-COVID-19	No DM no F/CCU-COVID-19	OR	2.5%	97.5%	P-Value
All	1082	318267	4081	5139870	1.395	1.304	1.494	<0.001
Male	657	179829	2039	2480478	1.449	1.325	1.584	<0.001
Female	425	138438	2042	2659392	1.339	1.205	1.487	<0.001
All Aged < 60	111	109297	648	3964406	2.494	2.032	3.061	<0.001
All Aged 60-70	137	81989	484	561527	1.764	1.457	2.136	<0.001
All Aged 70+	834	126981	2949	613937	1.327	1.227	1.434	<0.001
Other diabetes	23	8983	4081	5139870	1.262	0.836	1.905	0.268
Type 1 diabetes	51	34332	4081	5139870	2.396	1.815	3.163	<0.001
Type 2 diabetes	1008	274952	4081	5139870	1.369	1.276	1.468	<0.001

Associations were adjusted for age and sex, with the exception of sex, which was adjusted for age

Supplementary Table 3: Characteristics of all those with type 1 diabetes in Scotland who did and did not develop F/CCU-COVID-19 by 31st July 2020

	No F/CCU-COVID-19	F/CCU-COVID-19	Total diabetes population
Total included	34332(99.85)	51(0.15)	34383
Sociodemographic			
Age (years)	44.5(29.7,58.3)	71.4(56.4,80.6)	44.5(29.7,58.3)
Diabetes duration (years)	19.5(9.6,30.6)	27.0(17.1,39.8)	19.5(9.6,30.6)
Carehome resident	194(0.6)	20(39.2)	214(0.6)
Deprivation index			
Quintile 1 (most deprived)	6599(19.2)	13(25.5)	6612(19.2)
Quintile 2	6728(19.6)	13(25.5)	6741(19.6)
Quintile 3	6407(18.7)	8(15.7)	6415(18.7)
Quintile 4	6223(18.1)	5(9.8)	6228(18.1)
Quintile 5 (least deprived)	5768(16.8)	5(9.8)	5773(16.8)
Unknown	2607(7.6)	7(13.7)	2614(7.6)
Comorbidities			
Any DKA admission in past 5 years	5015(14.6)	14(27.5)	5029(14.6)
Any hypoglycaemia admission in past 5yrs	2099(6.1)	15(29.4)	2114(6.1)
No. of other hospital admissions in past 5 years	1.0(0.0,2.0)	7.0(2.5,13.5)	1.0(0.0,2.0)
Any heart disease	4816(14.0)	31(60.8)	4847(14.1)
Asthma or chronic lower airway disease	8680(25.3)	24(47.1)	8704(25.3)
Neurological and dementia (excluding epilepsy)	1378(4.0)	12(23.5)	1390(4.0)
Liver disease	160(0.5)	0(0.0)	160(0.5)
Immune disease or on immunosuppressants	626(1.8)	3(5.9)	629(1.8)
Any listed condition	12155(35.4)	43(84.3)	12198(35.5)
Other clinical measures			
Insulin pump use	4811(14.0)	1(2.0)	4812(14.0)
Flash glucose monitor use	11711(34.1)	6(11.8)	11717(34.1)
HbA1c (mmol/mol)	67(57,79)	69(57,87)	67(57,79)
BMI (kg/m^2)	26(23,30)	27(24,32)	26(23,30)
Systolic BP (mmHg)	130(120,140)	132(124,143)	130(120,140)
Diastolic BP (mmHg)	76(70,82)	72(65,80)	76(70,82)
Total cholesterol (mmol/L)	4(4,5)	4(4,5)	4(4,5)
eGFR ($mL/min/1.73m^2$)	100(83,116)	72(44,87)	100(83,116)
Albuminuric status			
Normal	13076(38.1)	16(31.4)	13092(38.1)
Micro	4300(12.5)	10(19.6)	4310(12.5)
Macro	1121(3.3)	4(7.8)	1125(3.3)
Unknown	15835(46.1)	21(41.2)	15856(46.1)
Retinopathy			
None	11384(33.2)	21(41.2)	11405(33.2)
Non referable	9101(26.5)	11(21.6)	9112(26.5)
Referable / eye clinic	7272(21.2)	14(27.5)	7286(21.2)
Unknown	6575(19.2)	5(9.8)	6580(19.1)
Tobacco smoking status			
Current smoker	6195(18.0)	7(13.7)	6202(18.0)
Ex smoker	12555(36.6)	26(51.0)	12581(36.6)
Never smoked	13869(40.4)	17(33.3)	13886(40.4)
Unknown	1713(5.0)	1(2.0)	1714(5.0)
Drug exposures			
Lipid lowering	13038(38.0)	37(72.5)	13075(38.0)
Proton Pump Inhibitors	8574(25.0)	23(45.1)	8597(25.0)
NSAIDS	11238(32.7)	32(62.7)	11270(32.8)
Anti-coagulants anti-platelets	5909(17.2)	33(64.7)	5942(17.3)
Antihypertensives (any)	12070(35.2)	30(58.8)	12100(35.2)
Number of ATC level 3 drug classes (excluding DM)	5.0(2.0,9.0)	12.0(9.5,14.5)	5.0(2.0,9.0)

Data are shown in N(%) for categorical values and median interquartile range for continuous values

Supplementary Table 4: Characteristics of all those with type 2 diabetes in Scotland who did and did not develop F/CCU-COVID-19 by 31st July 2020

	No F/CCU-COVID-19	F/CCU-COVID-19	Total diabetes population
Total included	274952(99.63)	1008(0.37)	275960
Sociodemographic			
Age (years)	68.4(59.1,76.8)	80.2(72.2,86.1)	68.4(59.1,76.9)
Diabetes duration (years)	10.0(5.5,15.6)	13.3(7.9,18.6)	10.1(5.5,15.6)
Carehome resident	5556(2.0)	367(36.4)	5923(2.1)
Deprivation index			
Quintile 1 (most deprived)	64494(23.5)	302(30.0)	64796(23.5)
Quintile 2	62407(22.7)	246(24.4)	62653(22.7)
Quintile 3	55191(20.1)	178(17.7)	55369(20.1)
Quintile 4	48364(17.6)	163(16.2)	48527(17.6)
Quintile 5 (least deprived)	39189(14.3)	95(9.4)	39284(14.2)
Unknown	5307(1.9)	24(2.4)	5331(1.9)
Comorbidities			
Any DKA admission in past 5 years	1320(0.5)	8(0.8)	1328(0.5)
Any hypoglycaemia admission in past 5yrs	3391(1.2)	52(5.2)	3443(1.2)
No. of other hospital admissions in past 5 years	1.0(0.0,3.0)	5.0(2.0,10.0)	1.0(0.0,3.0)
Any heart disease	93240(33.9)	651(64.6)	93891(34.0)
Asthma or chronic lower airway disease	93237(33.9)	467(46.3)	93704(34.0)
Neurological and dementia (excluding epilepsy)	13246(4.8)	214(21.2)	13460(4.9)
Liver disease	2671(1.0)	27(2.7)	2698(1.0)
Immune disease or on immunosuppressants	3143(1.1)	18(1.8)	3161(1.1)
Any listed condition	148979(54.2)	833(82.6)	149812(54.3)
Other clinical measures			
HbA1c (mmol/mol)	56(48,68)	57(47,70)	56(48,68)
BMI (kg/m^2)	31(27,35)	29(25,33)	31(27,35)
Systolic BP (mmHg)	134(125,142)	132(122,142)	134(125,142)
Diastolic BP (mmHg)	77(70,82)	74(67,80)	77(70,82)
Total cholesterol (mmol/L)	4(4,5)	4(3,5)	4(4,5)
eGFR ($mL/min/1.73m^2$)	82(64,95)	63(45,81)	82(64,95)
Albuminuric status			
Normal	114978(41.8)	279(27.7)	115257(41.8)
Micro	49925(18.2)	221(21.9)	50146(18.2)
Macro	10008(3.6)	72(7.1)	10080(3.7)
Unknown	100041(36.4)	436(43.3)	100477(36.4)
Retinopathy			
None	183660(66.8)	584(57.9)	184244(66.8)
Non referable	38540(14.0)	145(14.4)	38685(14.0)
Referable / eye clinic	20436(7.4)	118(11.7)	20554(7.4)
Unknown	32316(11.8)	161(16.0)	32477(11.8)
Tobacco smoking status			
Current smoker	42650(15.5)	101(10.0)	42751(15.5)
Ex smoker	137003(49.8)	641(63.6)	137644(49.9)
Never smoked	94352(34.3)	262(26.0)	94614(34.3)
Unknown	947(0.3)	4(0.4)	951(0.3)
Drug exposures			
Lipid lowering	193144(70.2)	754(74.8)	193898(70.3)
Proton Pump Inhibitors	119879(43.6)	541(53.7)	120420(43.6)
NSAIDS	128843(46.9)	650(64.5)	129493(46.9)
Anti-coagulants anti-platelets	104534(38.0)	621(61.6)	105155(38.1)
Antihypertensives (any)	181786(66.1)	668(66.3)	182454(66.1)
Number of ATC level 3 drug classes (excluding DM)	8.0(5.0,12.0)	11.0(8.0,15.0)	8.0(5.0,12.0)
Number of DM drug classes prescribed	1.0(1.0,2.0)	1.0(1.0,2.0)	1.0(1.0,2.0)

Data are shown in N(%) for categorical values and median interquartile range for continuous values

Supplementary Table 5: Additional characteristics described in categories and showing missingness and detailed drug exposures of all those with diabetes in Scotland who did and did not develop F/CCU-COVID-19 by 31st July 2020

	No F/CCU-COVID-19	F/CCU-COVID-19	Total diabetes population
HbA1c (mmol/mol)			
<53	116898(36.7)	421(38.9)	117319(36.7)
53-58	40527(12.7)	110(10.2)	40637(12.7)
>58-75	86790(27.3)	289(26.7)	87079(27.3)
>75-85	24963(7.8)	83(7.7)	25046(7.8)
>85	29852(9.4)	126(11.6)	29978(9.4)
Unknown	19237(6.0)	53(4.9)	19290(6.0)
BMI (kg/m^2)			
<20	6262(2.0)	55(5.1)	6317(2.0)
20-25	40604(12.8)	185(17.1)	40789(12.8)
>25-30	91486(28.7)	280(25.9)	91766(28.7)
>30-35	79407(24.9)	209(19.3)	79616(24.9)
>35-40	40444(12.7)	101(9.3)	40545(12.7)
>40	27615(8.7)	64(5.9)	27679(8.7)
Unknown	32449(10.2)	188(17.4)	32637(10.2)
eGFR ($mL/min/1.73m^2$)			
<30 or RRT	7620(2.4)	96(8.9)	7716(2.4)
30-60	50543(15.9)	386(35.7)	50929(15.9)
>60-90	126804(39.8)	438(40.5)	127242(39.8)
>90	111661(35.1)	141(13.0)	111802(35.0)
Unknown	21639(6.8)	21(1.9)	21660(6.8)
Systolic blood pressure (mmHg)			
<130	121838(38.3)	473(43.7)	122311(38.3)
130-160	160235(50.3)	486(44.9)	160721(50.3)
>160	13742(4.3)	64(5.9)	13806(4.3)
Unknown	22452(7.1)	59(5.5)	22511(7.0)
Diastolic blood pressure (mmHg)			
≤80	207412(65.2)	820(75.8)	208232(65.2)
>80	88392(27.8)	202(18.7)	88594(27.7)
Unknown	22463(7.1)	60(5.5)	22523(7.1)
Total cholesterol (mmol/L)			
≤5	222234(69.8)	770(71.2)	223004(69.8)
>5	66874(21.0)	180(16.6)	67054(21.0)
Unknown	29159(9.2)	132(12.2)	29291(9.2)
Drug exposures			
Diabetes drugs			
Insulin	68811(21.6)	270(25.0)	69081(21.6)
Biguanides	194242(61.0)	631(58.3)	194873(61.0)
Sulfonylureas	83950(26.4)	361(33.4)	84311(26.4)
TZD's	13363(4.2)	44(4.1)	13407(4.2)
Acarbose	192(0.1)	1(0.1)	193(0.1)
DPP4 inhibitors	49404(15.5)	184(17.0)	49588(15.5)
GLP1 agonists	17539(5.5)	53(4.9)	17592(5.5)
SGLT2 inhibitors	41432(13.0)	96(8.9)	41528(13.0)
Antihypertensive drugs			
Vasodilator	26915(8.5)	52(4.8)	26967(8.4)
Central acting	3037(1.0)	10(0.9)	3047(1.0)
Alpha adrenergic	19194(6.0)	67(6.2)	19261(6.0)
Angiotensin inhibitor	124024(39.0)	449(41.5)	124473(39.0)
Angiotensin antagonist	47644(15.0)	149(13.8)	47793(15.0)
Renin inhibitor	7(0.0)	0(0.0)	7(0.0)
Thiazides	47158(14.8)	129(11.9)	47287(14.8)
Calcium channel blocker	89539(28.1)	340(31.4)	89879(28.1)

Data are shown in N(%) for categorical values and median interquartile range for continuous values

Supplementary Table 6: Logistic regression of association of additional categories and drug exposures with F/CCU-COVID-19 in people with diabetes. Associations were adjusted for age, sex, diabetes duration and type.

Predictor	Odds Ratio	95%CI	P-Value
HbA1c (mmol/mol)			(global) <0.001
<53	1 (reference)		
53-58	0.808	(0.660, 0.988)	0.037
>58-75	1.073	(0.923, 1.246)	0.359
>75-85	1.302	(1.023, 1.655)	0.032
>85	2.065	(1.682, 2.536)	<0.001
BMI (kg/m^2)			(global) <0.001
<20	2.403	(1.773, 3.259)	<0.001
20-25	1 (reference)		
>25-30	0.801	(0.671, 0.956)	0.014
>30-35	0.928	(0.772, 1.115)	0.425
>35-40	0.966	(0.758, 1.229)	0.776
>40	1.191	(0.888, 1.597)	0.244
eGFR (mL/min/1.73m^2)			(global) <0.001
<30 or RRT	2.372	(1.761, 3.194)	<0.001
30-60	1.320	(1.040, 1.675)	0.022
>60-90	0.980	(0.794, 1.210)	0.849
>90	1 (reference)		
Systolic BP (mmHg)			(global) <0.001
<130	1 (reference)		
130-160	0.655	(0.578, 0.741)	<0.001
>160	0.862	(0.662, 1.121)	0.267
Diastolic BP (mmHg)			(global) 0.776
≤80	1 (reference)		
>80	0.978	(0.836, 1.143)	0.776
Total cholesterol (mmol/L)			(global) 0.119
≤5	1 (reference)		
>5	1.140	(0.969, 1.341)	0.114
Diabetes drugs			
Biguanides	1.051	(0.925, 1.195)	0.446
Insulin	1.964	(1.670, 2.309)	<0.001
Sulfonylureas	1.310	(1.149, 1.492)	<0.001
TZD's	0.898	(0.662, 1.218)	0.488
Acarbose	1.078	(0.150, 7.729)	0.941
DPP4 inhibitors	1.092	(0.929, 1.283)	0.286
GLP1 agonists	1.407	(1.059, 1.868)	0.018
SGLT2 inhibitors	1.124	(0.904, 1.398)	0.293
Antihypertensive drugs			
Vasodilator	0.680	(0.511, 0.905)	0.008
Central acting	1.256	(0.672, 2.349)	0.476
Alpha adrenergic	0.746	(0.582, 0.957)	0.021
Angiotensin inhibitor	0.951	(0.842, 1.075)	0.421
Angiotensin antagonist	0.721	(0.606, 0.857)	<0.001
Thiazides	0.615	(0.512, 0.740)	<0.001

Supplementary Table 7: Predictor variables and coefficients for risk prediction of F/CCU-COVID-19

Predictor	Coefficient
(Intercept)	-8.74364437379780
Age (years)	0.04259134072428
Sex male	0.00000000000000
Sex female	-0.62642258060175
Type 2 diabetes	0.00000000000000
Type 1 diabetes	0.11245287638435
Other diabetes types	-0.14409417993123
Diabetes duration (years)	-0.00209419600062
Carehome resident	2.38217027053367
Deprivation quintile 1 (most deprived)	0.00000000000000
Deprivation quintile 2	-0.16490767973292
Deprivation quintile 3	-0.48028099889677
Deprivation quintile 4	-0.42229894439757
Deprivation quintile 5 (least deprived)	-0.72620380743347
log(No. of other hospital admissions + 1 (5yrs))	0.46662994412763
Neurological and dementia (excluding epilepsy)	0.24132804884094
HbA1c (mmol/mol)	0.00507982471972
BMI (kg/m^2)	0.08689021484269
log(BMI (kg/m^2))	-2.53147951308967
eGFR ($mL/min/1.73m^2$)	-0.00770941564453
Systolic BP (mmHg)	-0.00567062648956
Any antihypertensive	-0.23298518580490
Number of DM drug classes prescribed	0.06281305073017
Number of ATC level 3 drug classes (excluding DM)	0.02632838543546