Reports to the Joint Committee on Vaccination and Immunisations relating to absolute risks of severe outcomes from COVID-19

27th September 2023

Elizabeth Williamson

Contents

[Note on report 3](#_heading=h.1fob9te)

[OpenSAFELY 4](#_heading=h.3znysh7)

[OpenSAFELY: a new data analytics platform 4](#_heading=h.2et92p0)

[Data within the OpenSAFELY platform 4](#_heading=h.tyjcwt)

[Covariate extraction 4](#_heading=h.3dy6vkm)

[Information governance 5](#_heading=h.1t3h5sf)

[Ethics approvals 5](#_heading=h.4d34og8)

[Availability of code and codelists 5](#_heading=h.2s8eyo1)

[Background 6](#_heading=h.17dp8vu)

[Methods 6](#_heading=h.3rdcrjn)

[Study design 6](#_heading=h.26in1rg)

[Study Population 6](#_heading=h.lnxbz9)

[Outcomes 6](#_heading=h.35nkun2)

[Observation Period 6](#_heading=h.1ksv4uv)

[Covariates 7](#_heading=h.44sinio)

[Statistical Analysis: Analysis 1 7](#_heading=h.2jxsxqh)

[Statistical Analysis: Analysis 2 8](#_heading=h.z337ya)

[Results: Summary of underlying cohort 9](#_heading=h.3j2qqm3)

[Results: Analysis 1 13](#_heading=h.1y810tw)

[Results: Analysis 2 24](#_heading=h.4i7ojhp)

[Discussion 31](#_heading=h.2xcytpi)

[Summary of findings 31](#_heading=h.1ci93xb)

[Limitations 31](#_heading=h.3whwml4)

[JCVI reports 32](#_heading=h.2bn6wsx)

[Availability of code 32](#_heading=h.qsh70q)

[References 33](#_heading=h.3as4poj)

# Note on report

This report provides details and context around the analyses provided to the Joint Committee on Vaccination and Immunisations (JCVI) during the period September 2020 – February 2021.

The presentations were written for the purposes of discussion during meetings and, as such, do not necessarily contain sufficient detail to be understood as standalone documents. I therefore wrote this document post-hoc, during 2023, to provide explanation and context around the analyses that had been presented. These notes represent my own personal reflections and interpretation on the analyses and presentations.

Two key analyses are presented. Both were exploratory analyses, based on the data assembled for an early paper exploring factors associated with COVID-19 related death (Williamson et al, Nature, 2020). These analyses looked at different ways of presenting results in terms of absolute, rather than relative, risks experienced by patients with different comorbidities and demographic characteristics.

1. Re-analysis of Nature paper data presenting risks in absolute terms by age-group and sex.
2. Re-analysis of Nature paper data additionally extending to COVID-19 related hospital admissions, presenting data by ethnicity and continuous age and looking at age 65-yr risk attained.

These analyses were exploratory, to suggest some possible ways in which absolute risks could be presented and thought about.

# OpenSAFELY

## OpenSAFELY: a new data analytics platform

OpenSAFELY is a new data analytics platform that was established to address urgent questions relating to the epidemiology and treatment of COVID-19 in England. OpenSAFELY provides a secure software interface that allows detailed pseudonymised primary care patient records to be analysed in near real-time where they already reside, hosted within the electronic health record (EHR) vendor’s highly secure data centre, to minimise the re-identification risks when data are transported off-site. Other datasets are linked to these data within the same environment using a matching pseudonym derived from the NHS number.

More information can be found on <https://opensafely.org/>.

The data for all analyses presented in this report were accessed, linked and analysed using the OpenSAFELY platform.

## Data within the OpenSAFELY platform

The analyses presented in this report used patient data from general practice (GP) records managed by the GP software provider The Phoenix Partnership (TPP), linked to other datasets including the Office for National Statistics (ONS) death data.

At the time of these analyses, the data within OpenSAFELY was based on 24 million currently registered patients (approximately 40% of the English population) from GP surgeries using the TPP SystmOne electronic health record (EHR) system. SystmOne is a secure centralised EHR used in English clinical practice since 1998, which records data entered by GPs and practice staff during routine primary care. The system is accredited under the NHS approved systems framework for General Practice (NHS Digital, 2020).

A pseudonymised dataset was created from this electronic health record for OpenSAFELY consisting of 20 billion rows of structured data including for example pseudonymised patients’ diagnoses, medications, physiological parameters, and prior investigations.

All OpenSAFELY data processing took place on TPP’s servers. External data providers securely transferred pseudoymised data for linkage to OpenSAFELY. No pseudonymised patient-level data were ever removed from TPP infrastructure. Only aggregated, anonymous, manually checked study results were released for publication.

## Covariate extraction

Information on all covariates used in analyses within this report were obtained from primary care records by searching TPP SystmOne records for specific coded data. TPP SystmOne allows users to work with the SNOMED-CT clinical terminology, using a GP subset of SNOMED-CT codes. This subset maps on to the native Read version 3 (CTV3) clinical coding system that SystmOne is built on. Medicines are entered or prescribed in a format compliant with the NHS Dictionary of Medicines and Devices (dm+d), a local UK extension library of SNOMED. Code lists for particular underlying conditions and medicines were compiled from a variety of sources (more details in Williamson et al, Nature, 2020).

### Information governance

NHS England is the data controller; TPP is the data processor; and the key researchers on OpenSAFELY are acting on behalf of NHS England. This implementation of OpenSAFELY was hosted within the TPP environment which is accredited to the ISO 27001 information security standard and is NHS IG Toolkit compliant (NHS Digital, 2018 and 2019); patient data has been pseudonymised for analysis and linkage using industry standard cryptographic hashing techniques; all pseudonymised datasets transmitted for linkage onto OpenSAFELY are encrypted; access to the platform is via a virtual private network (VPN) connection, restricted to a small group of researchers, their specific machine and IP address; the researchers hold contracts with NHS England and only access the platform to initiate database queries and statistical models; all database activity is logged; only aggregate statistical outputs leave the platform environment following best practice for anonymisation of results such as statistical disclosure control for low cell counts (NHS Digital, 2019).

The OpenSAFELY research platform adheres to the data protection principles of the UK Data Protection Act 2018 and the EU General Data Protection Regulation (GDPR) 2016. In March 2020, the Secretary of State for Health and Social Care used powers under the UK Health Service (Control of Patient Information) Regulations 2002 (COPI) to require organisations to process confidential patient information for the purposes of protecting public health, providing healthcare services to the public and monitoring and managing the COVID-19 outbreak and incidents of exposure (DHSC, 2020). Taken together, these provide the legal bases to link patient datasets on the OpenSAFELY platform. GP practices, from which the primary care data are obtained, are required to share relevant health information to support the public health response to the pandemic, and have been informed of the OpenSAFELY analytics platform.

### Ethics approvals

The studies included in this report were approved by the Health Research Authority (REC reference 20/LO/0651) and by the LSHTM Ethics Board (reference 21863).

## Availability of code and codelists

Codelists used are available for inspection and re-use by the broader research community at <https://codelists.opensafely.org/>. This link also provides detailed information on compilation and sources for every individual codelist.

All code for data management and analysis is archived online. These analyses were undertaken within the repository for the exploration of risk factors associated with COVID-19 death:

<https://github.com/opensafely/risk-factors-research>

More detail can be found at the end of the report.

# Background

At the time of the first report to JCVI, the OpenSAFELY collaborative had already published an analysis in Nature (Williamson, 2020), exploring risk factors associated with COVID-19 related death. All results were presented in terms of relative risks, specifically hazard ratios. For public health and policy-making, it is often crucial to obtain absolute measures of risk. In the setting of an infectious disease, absolute risks of death from the disease among the general population will inevitably depend on the underlying circulating prevalence of the infection. This analysis, therefore, considered the specific question of absolute risks of COVID-19 related death and hospital admission during the first wave of COVID-19. This used the previously published analyses focusing on relative risks, re-expressing them in absolute terms.

# Methods

The data used for this analysis was the same as that used for the exploration of risk factors for severe outcomes from COVID-19 (Williamson et al, Nature, 2020), thus the study design described below is identical to that.

## Study design

We conducted a cohort study using national primary care electronic health record data linked to COVID-19 death data held within the OpenSAFELY platform (see previous section for details). Data were accessed, linked and analysed within the OpenSAFELY platform.

The cohort study began on 1st February 2020, chosen as a date several weeks prior to the first reported COVID-19 deaths and the day after the second laboratory confirmed case; and ended on 25th April 2020. Note that the previously published analyses (Williamson, Nature, 2020) had a slightly longer study period lasting until 6th May 2020. The cohort explores risk among the general population rather than in a population infected with SARS-COV-2. Therefore, all patients were included irrespective of any SARS-COV-2 test results.

## Study Population

The study population consisted of all adults (males and females 18 years and above) currently registered as active patients in a TPP general practice (GP) in England on 1st February 2020. Participants were required to have recorded sex, age, and deprivation status. They also required at least 1 year of prior follow-up within their general practice to ensure adequate capture of baseline patient characteristics.

## Outcomes

For the first analysis, the outcome was death among people with COVID-19, ascertained from ONS death certificate data, where the COVID related ICD-10 codes U071 or U072 were present in the record. At the time of the analysis, ONS data were available to 11th May 2020, but we used an earlier censor date to allow for delays in reporting in the last few days of available data.

The second analysis additionally includes being admitted to ICU for COVID-19.

## Observation Period

Patients were observed from 1st February 2020 and were followed until the first of either their death date (whether COVID-19 related or due to other causes) or the study end date.

For the initial analysis 1, the study end date was set to 25th April 2020, in contrast to the previous study (Williamson et al, Nature, 2020) which used a later study end date of 6th May 2020, due to concerns about incomplete outcome data towards the end of this period. Subsequent updates extended this back to 6th May 2020. For the second analysis, more data was available, thus the study end date was set to 24th June 2020 for COVID-1 death and 24th May 2020 for hospital admissions.

## Covariates

Potential risk factors included: health conditions listed in UK guidance on “higher risk” groups, other common conditions which may cause immunodeficiency inherently or through medication (cancer and common autoimmune conditions); and emerging risk factors for severe outcomes among COVID-19 cases (such as raised blood pressure).

Age, sex, smoking status and body mass index (BMI) measured in kg/m2 were considered as potential risk factors. Where categorised, age groups were: 18-<40, 40-<50, 50-<60, 60-<70, 70-<80, 80+ years. Smoking status was grouped into current, former and never smokers. BMI was obtained from the most recent weight measurement, provided that this was taken within the last 10 years and that the patient was over 16 years old at that time. Obesity was grouped using the World Health Organisation classification of BMI: no evidence of obesity <30 kg/m2, obese I 30-34.9, obese II 35-39.9, obese III 40+.

The following comorbidities were also considered potential risk factors: asthma, other chronic respiratory disease, chronic heart disease, diabetes mellitus, chronic liver disease, chronic neurological diseases, common autoimmune diseases (Rheumatoid Arthritis, Systemic Lupus Erythematosus or psoriasis), solid organ transplant, asplenia, other immunosuppressive conditions, cancer, evidence of reduced kidney function, and raised blood pressure or a diagnosis of hypertension. Full details of what was included in these conditions can be found in Williamson et al (Nature, 2020), along with the codelists used to extract these variables. In some analyses HIV was separated from other immunosuppressive conditions.

Deprivation was considered as a potential upstream risk factor, measured by the Index of Multiple Deprivation (IMD, in quintiles, with higher values indicating greater deprivation), derived from the patient’s postcode at lower super output area level for a high degree of precision. Ethnicity was grouped into 5 categories: White (including British, Irish, other white), South Asian (Indian, Pakistani, Bangladeshi, other Asian), Black (African, Caribbean, other black), Other (Chinese, all other), and Mixed (white and Asian, white and African, white and Caribbean, other mixed).

The Sustainability and Transformation Partnership (STP, an NHS administrative region) of the patient’s general practice was included as an additional adjustment for geographical variation in infection rates across the country. A broader regional classification was also available, grouping the STPs into 9 regions.

### Statistical Analysis: Analysis 1

Within the original analysis of the data from this cohort study, the following analyses were undertaken with results re-presented here for context:

We depicted patient numbers through the study selection process in a flowchart. We estimated a Kaplan-Meier failure function by age group and sex. We fitted a Cox proportional hazards model, with days in study as the timescale, stratified by geographic area (STP), including age (modelled as a spline), sex, BMI, smoking, index of multiple deprivation quintile, and comorbidities listed above. Hazard ratios were reported with 95% confidence intervals.

The current extension to this analysis focused on obtaining absolute risks used the following methods:

We estimated absolute risk experienced over the first 80 days in the cohort study, i.e. from 1st Feb 2020 until 25th April 2020, using a Royston-Parmar parametric survival model. Initial models used 5 degrees of freedom; subsequent updates increased this to 10 to increase the flexibility of the fitted model, and added an initial linear fit to improve convergence. Risks were estimated separately for each sex, age-group and ethnicity group, for a patient who had the following baseline set of characteristics: non-smoker, non-obese, IMD level 3 (mid-level deprivation), mean age within age-group. Each comorbidity was then assessed in turn, assuming no other comorbidity was present. Models adjusted for geographic area, at the broader region level rather than STP. Following the meeting, subsequent analyses ran models for ethnicity groups separately, rather than combine them all together.

In the analyses presented in this report, those with missing BMI were assumed non-obese and those with missing smoking information were assumed to be non-smokers on the assumption that both obesity and smoking would be likely to be recorded if present. Patients with missing ethnicity data were excluded from the analysis. Results are presented only for the White ethnicity group, since the purpose of this report was to illustrate one way of presenting absolute risks rather than present a comprehensive summary. Sensitivity analyses for missing data were run but not reported here.

### Statistical Analysis: Analysis 2

A Royston-Parmar parametric survival model, adjusted for region and including all demographic and comorbidity variables listed above, was fitted. The model was used to predict 80-day risk of COVID-19 related death and hospital admission across ages (20-80 years). Risks were estimated separately for each sex and ethnicity group, for a patient who had the following baseline set of characteristics: non-smoker, non-obese, IMD level 3 (mid-level deprivation), mean age within age-group. Each comorbidity was then assessed in turn, assuming no other comorbidity was present.

Approximate age at which “65-year old risk” is estimated to be attained was obtained by comparing the risk profile for an individual comorbidity across the ages to the mean predicted risk for a 65-year old of the same ethnicity and sex. This was intended to be an illustration of one way of thinking about risks, thus we did not attempt to obtain precise point estimates or put 95% confidence intervals around these estimates, at this stage.

A comorbidity count variable was created, summing the presence of each individual comorbidity listed above. For this purpose, any diabetes and cancer was counted as present, but only severe asthma. The number of comorbidities was grouped into: None, 1, and 2 or more. Additional Royston-Parmar models were fitted including the categorised comorbidity count, rather than individual comorbidities, and used to predict 80-day risks of COVID-19 related death and hospital admission according to number of comorbidities present. These were compared to a reference level of risk experienced by an average person of the same sex and ethnicity who was aged 65.

# Results: Summary of underlying cohort

The results presented in this section are a summary of published results in Williamson et al (Nature, 2020) for context, since the data used in the analysis are almost identical.

Figure 1 shows the flow of patient numbers through the study selection process. Table 1 shows a few selected baseline characteristics. A full description of patients included in the cohort can be found in Williamson et al (Nature, 2020).

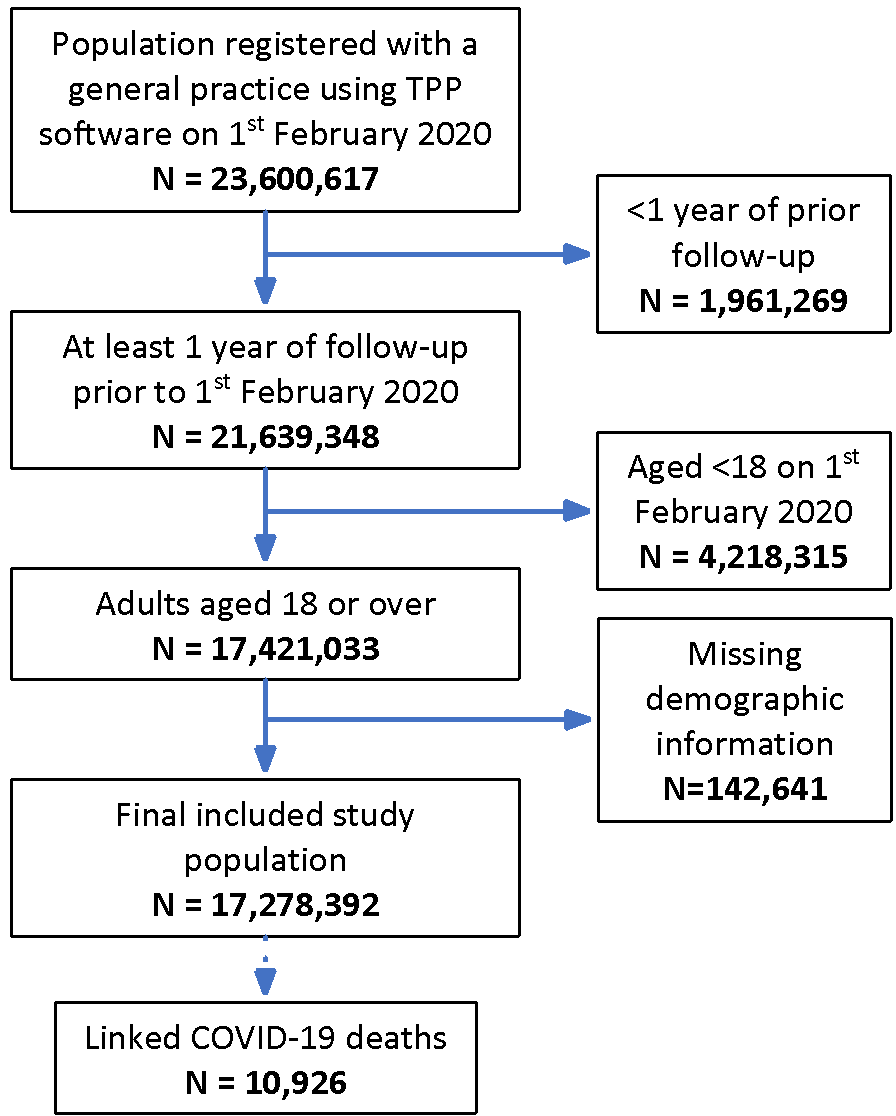


Figure 1. [Published in Williamson et al, Nature, 2020.] Flowchart of patient numbers through study selection process.

Table 1. [Published in Williamson et al, Nature, 2020.] Description of patients included in the cohort study (key demographics only; full details in the Nature paper).

| **Characteristic** | **Category** | **N (column %)** | **N COVID-19-**  **related deaths (%)** |
| --- | --- | --- | --- |
| Total |  | 17,278,392 (100.0) | 10,926 (0.06) |
| Age | 18–39 | 5,914,384 (34.2) | 54 (0.00) |
|  | 40–49 | 2,849,984 (16.5) | 140 (0.00) |
|  | 50–59 | 3,051,110 (17.7) | 522 (0.02) |
|  | 60–69 | 2,392,392 (13.8) | 1,101 (0.05) |
|  | 70–79 | 1,938,842 (11.2) | 2,635 (0.14) |
|  | 80+ | 1,131,680 (6.5) | 6,474 (0.57) |
| Sex | Female | 8,647,989 (50.1) | 4,764 (0.06) |
|  | Male | 8,630,403 (49.9) | 6,162 (0.07) |
| Ethnicity | White | 10,866,411 (62.9) | 7,119 (0.07) |
|  | Mixed | 169,697 (1.0) | 62 (0.04) |
|  | South Asian | 1,022,130 (5.9) | 608 (0.06) |
|  | Black | 339,909 (2.0) | 250 (0.07) |
|  | Other | 320,132 (1.9) | 110 (0.03) |
|  | Missing | 4,560,113 (26.4) | 2,777 (0.06) |

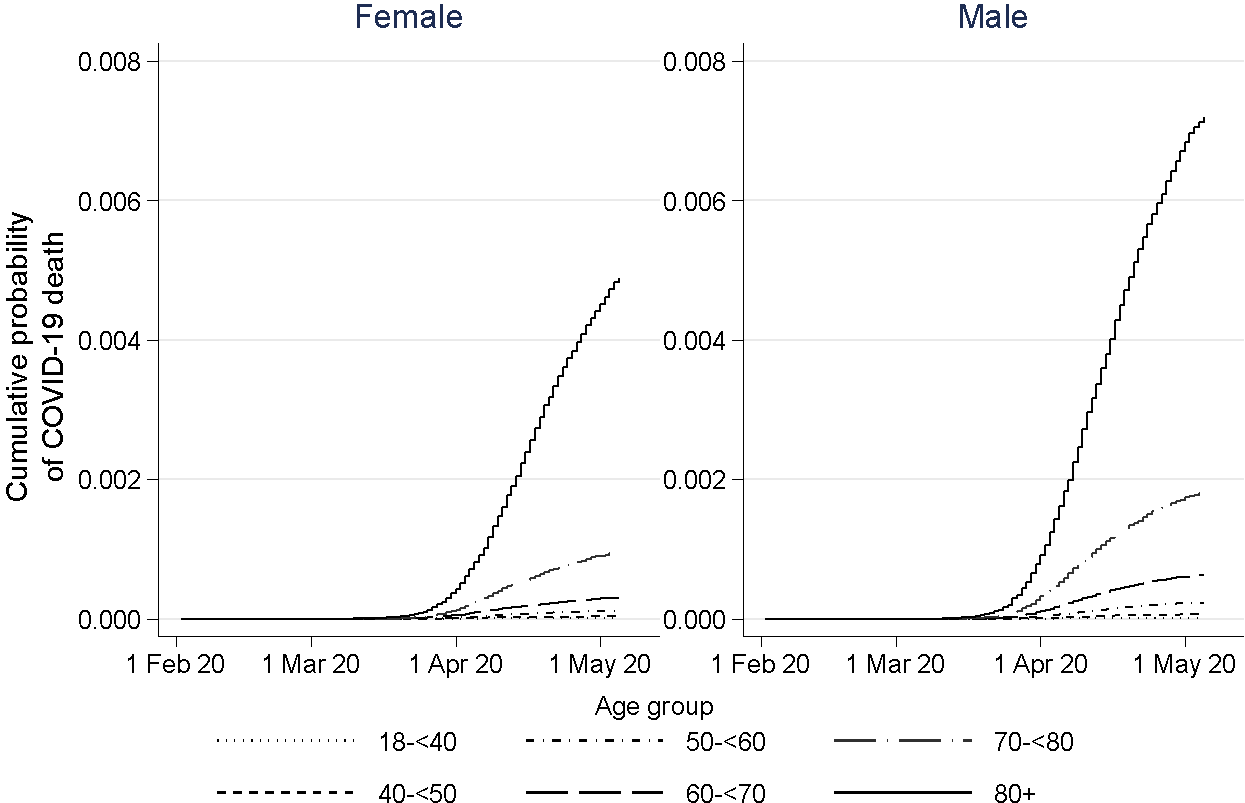


Figure 2. [Published in Williamson et al, Nature, 2020.] Kaplan-Meier estimated failure curves for COVID-19 related death by sex and age-group.

Of the 17,278,392 patients included in the cohort, 10,926 experienced COVID-19 related death during the study period (using the slightly longer study period extending to 6th May 2020). Therefore, in nearly 3 months, 0.06% experienced COVID-19 related death overall.

Figure 3 shows the estimated hazard ratios from a fully adjusted model, as presented in Williamson et al. (Nature, 2020).

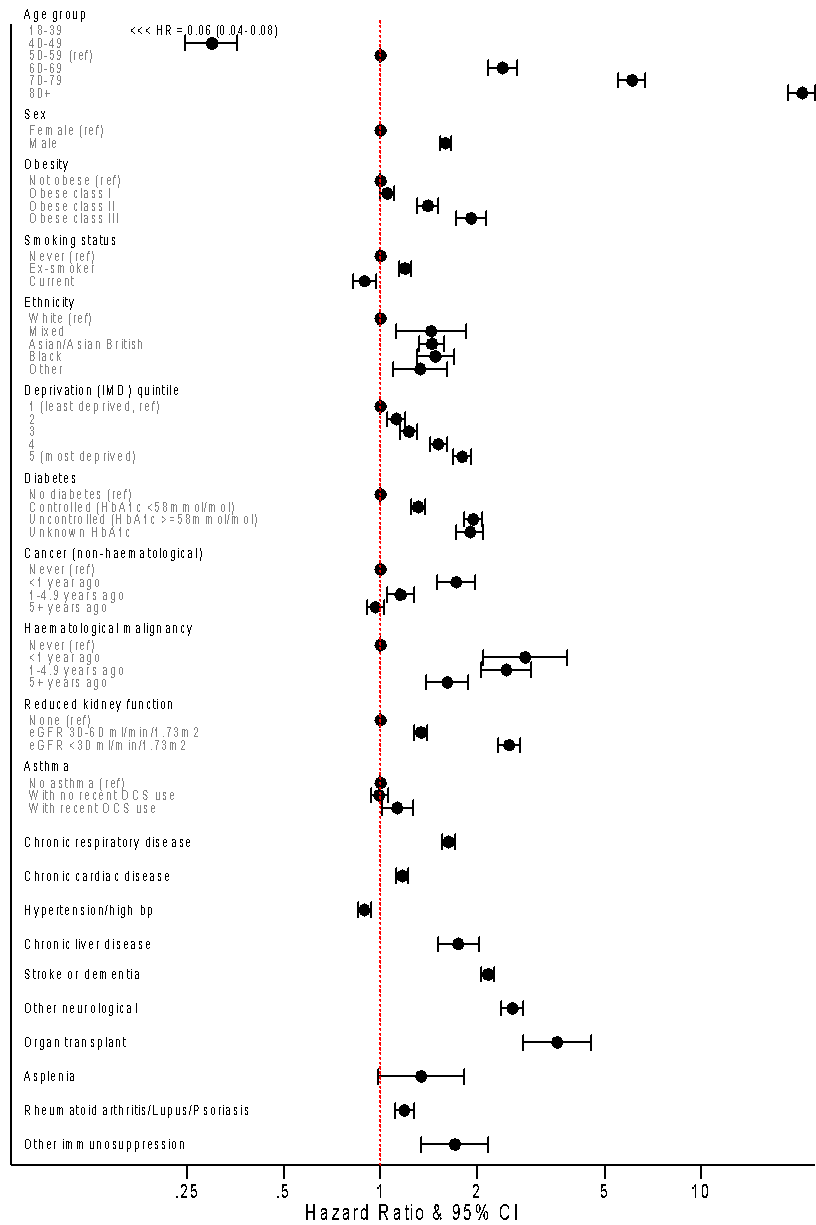


Figure 3. [Published in Williamson et al, Nature, 2020.] Estimated hazard ratios with 95% confidence intervals from a fully adjusted Cox model, stratified by STP (geographical location).

# Results: Analysis 1

The results in this section were created for the JCVI report.

Results are presented in graphical form, by age-group and sex (Figures 4a-4e). These figures show estimated absolute risks of COVID-19 related death, with 95% confidence intervals. The dashed lines show overall risk percentiles, for easier comparison across groups. These lines fall at: 50th percentile 0.004%, 70th percentile 0.01%, 80th percentile 0.03%, 90th percentile 0.06%.Points are colour-coded (in dark green, mustard, orange, red, respectively) to indicate whether they lie above the 50th, 70th, 80th or 90th percentiles.

Overall, these results show that while individual comorbidities appear to have high relative effects (elevated hazard ratios) in isolation, they are not estimated to result in absolute risks above the 80th percentile for patients under 60 years old. The one exception was organ transplant among men aged 50-<60, which had an estimated absolute risk just above the 80th percentile threshold.

These results are limited, in showing risks associated with individual comorbidities. To illustrate differences between this approach and taking into account the fact that patients with a specific comorbidity (e.g. organ transplant) might tend to have high comorbidity burden, Figures 5a and 5b show estimated absolute risks for patients 50-<60 using the individual comorbidity approach taken here (Figure 5a) and plotting the observed absolute risk within the subgroup of patients with that comorbidity (Figure 5b). This figure clearly illustrates the fact that certain comorbidities, particularly organ transplant and very reduced kidney function, experience a very high absolute risk compared to that estimated to be associated only with those comorbidities, reflecting the high comorbidity burden within these patient groups.

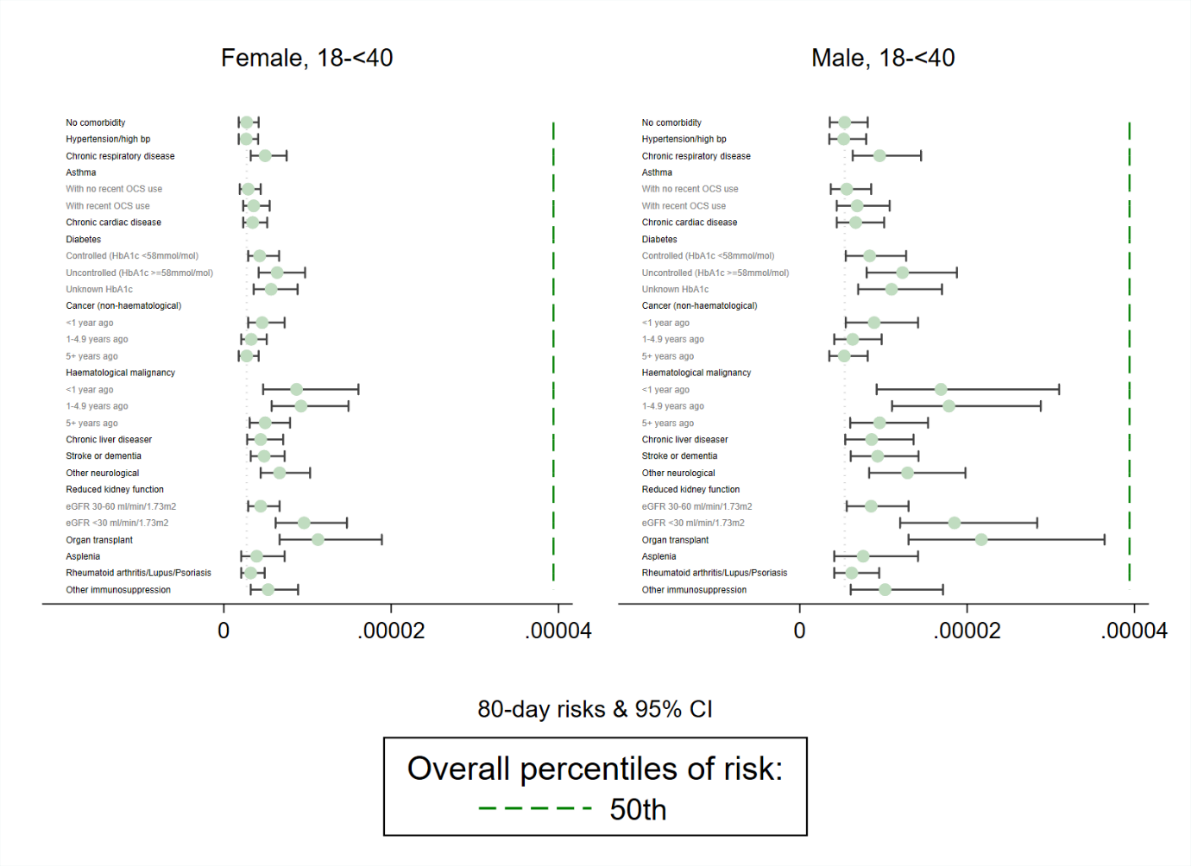


Figure 4a. Estimated 80-day absolute risk of COVID-19 related death for patients 18-<40 years old, by sex, with 95% confidence intervals. Risks are estimated assuming a single comorbidity is present, with a baseline set of demographic characteristics as described in the methods section.

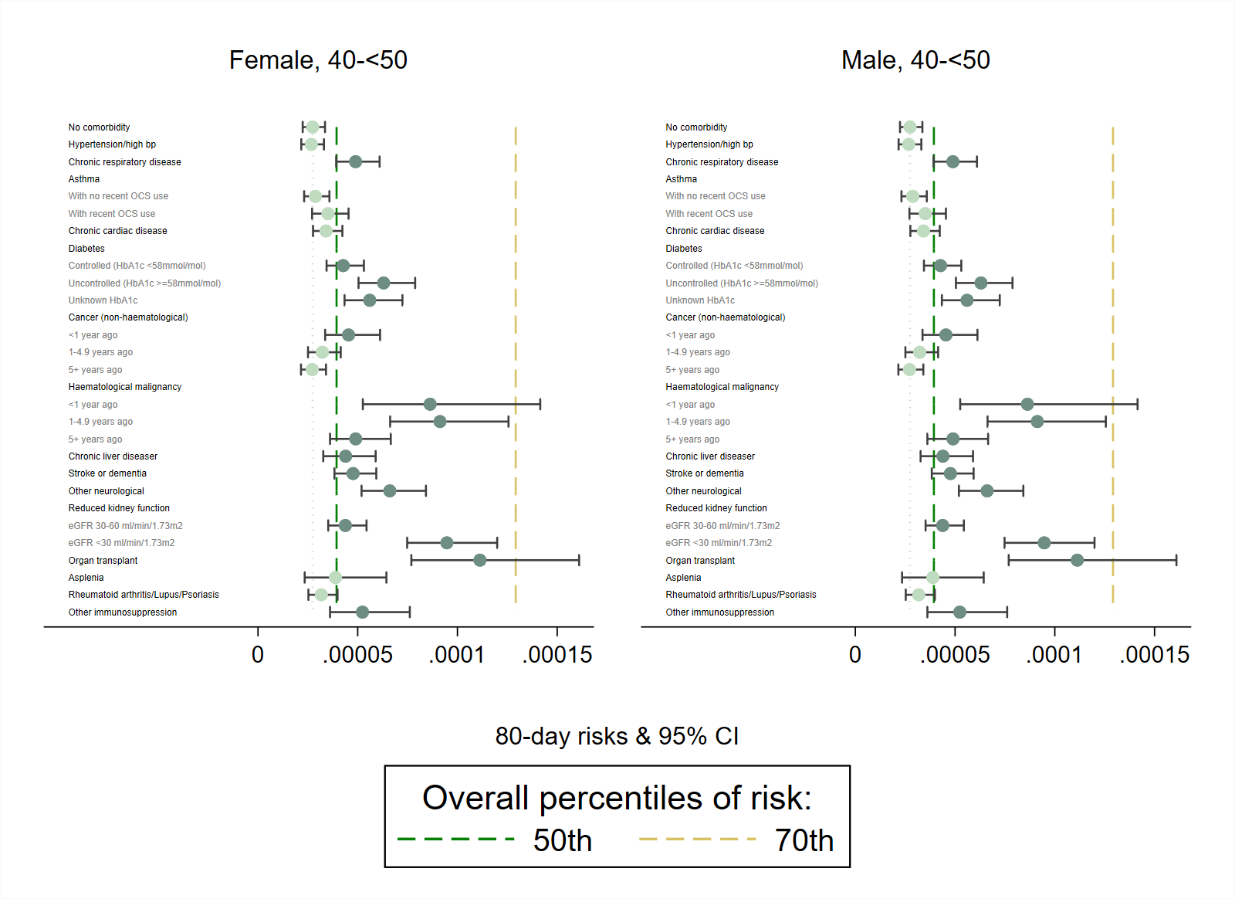


Figure 4b. Estimated 80-day absolute risk of COVID-19 related death for patients 40-<50 years old, by sex, with 95% confidence intervals. Risks are estimated assuming a single comorbidity is present, with a baseline set of demographic characteristics as described in the methods section.

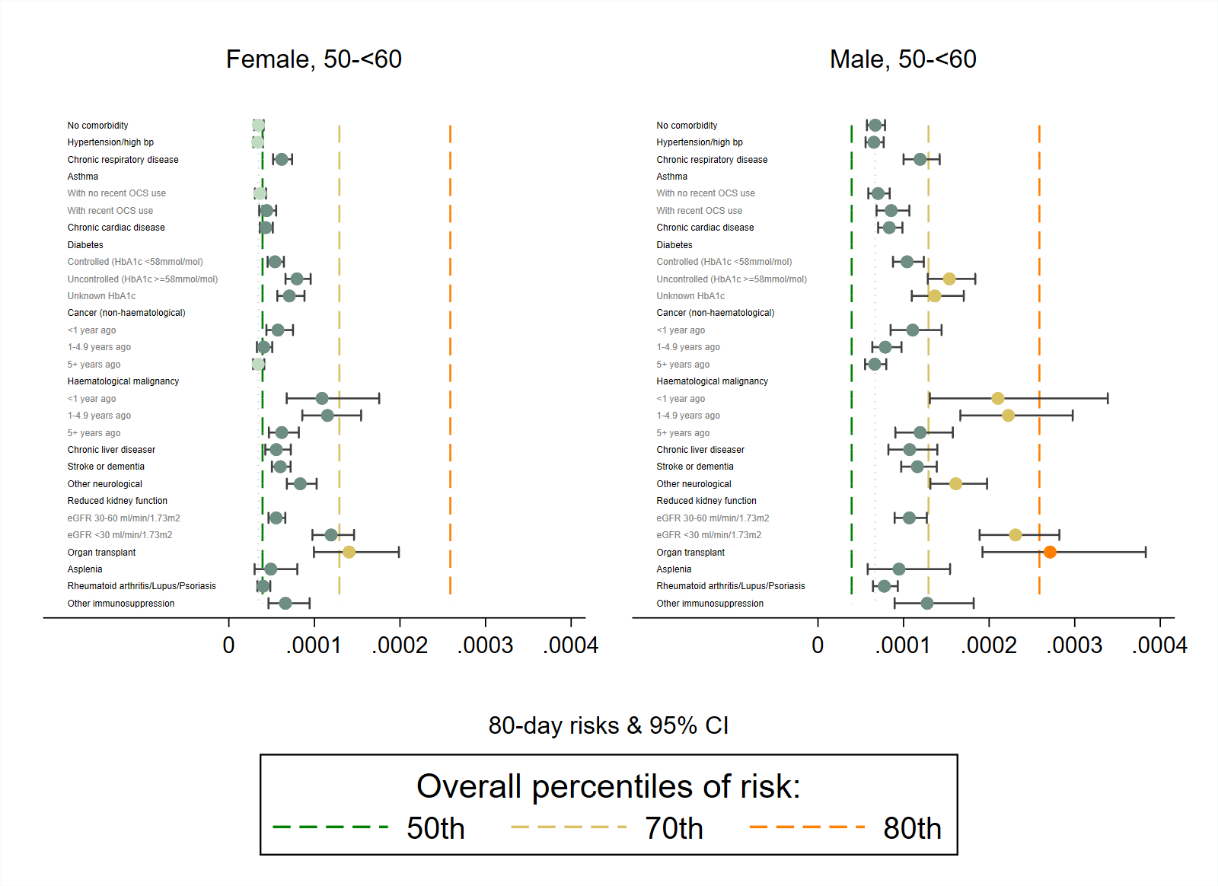


Figure 4c. Estimated 80-day absolute risk of COVID-19 related death for patients 50-<60 years old, by sex, with 95% confidence intervals. Risks are estimated assuming a single comorbidity is present, with a baseline set of demographic characteristics as described in the methods section.

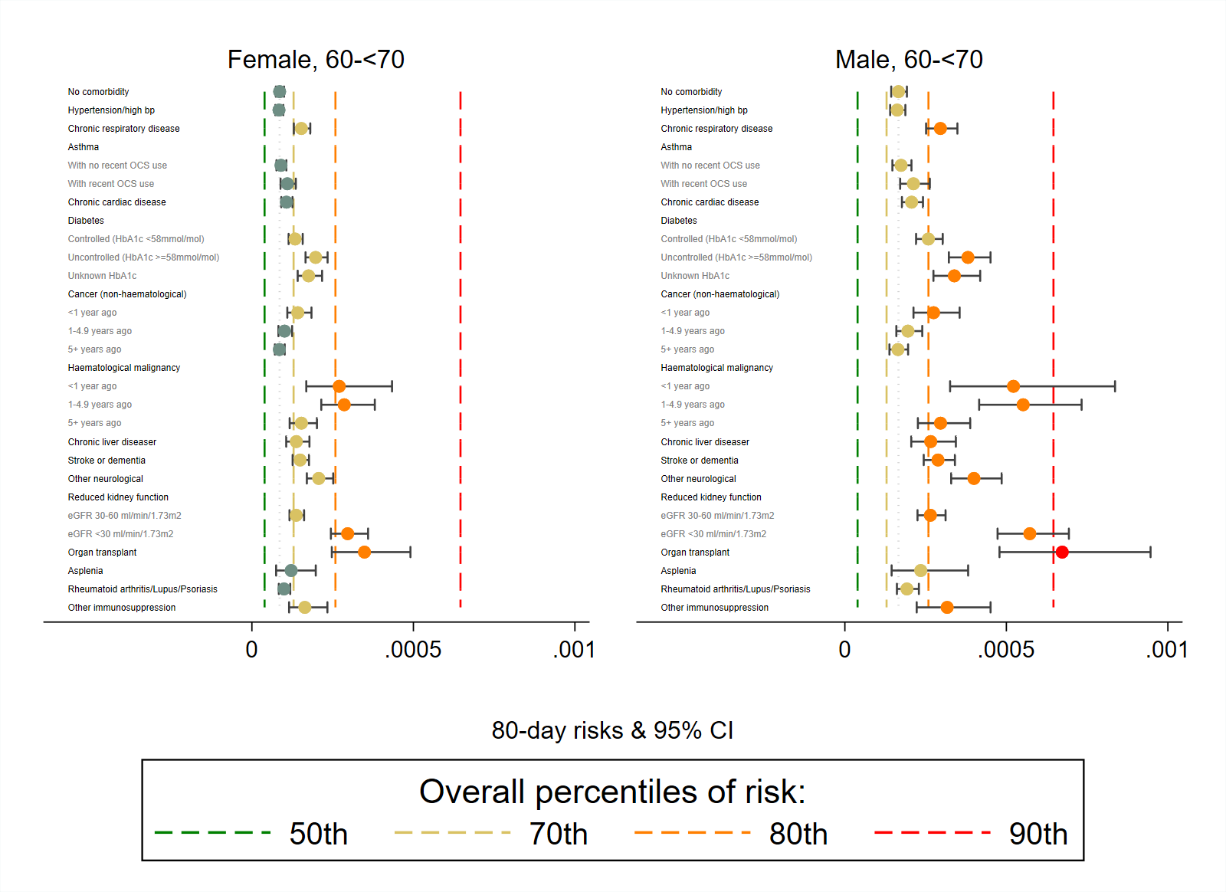


Figure 4d. Estimated 80-day absolute risk of COVID-19 related death for patients 60-<70 years old, by sex, with 95% confidence intervals. Risks are estimated assuming a single comorbidity is present, with a baseline set of demographic characteristics as described in the methods section.

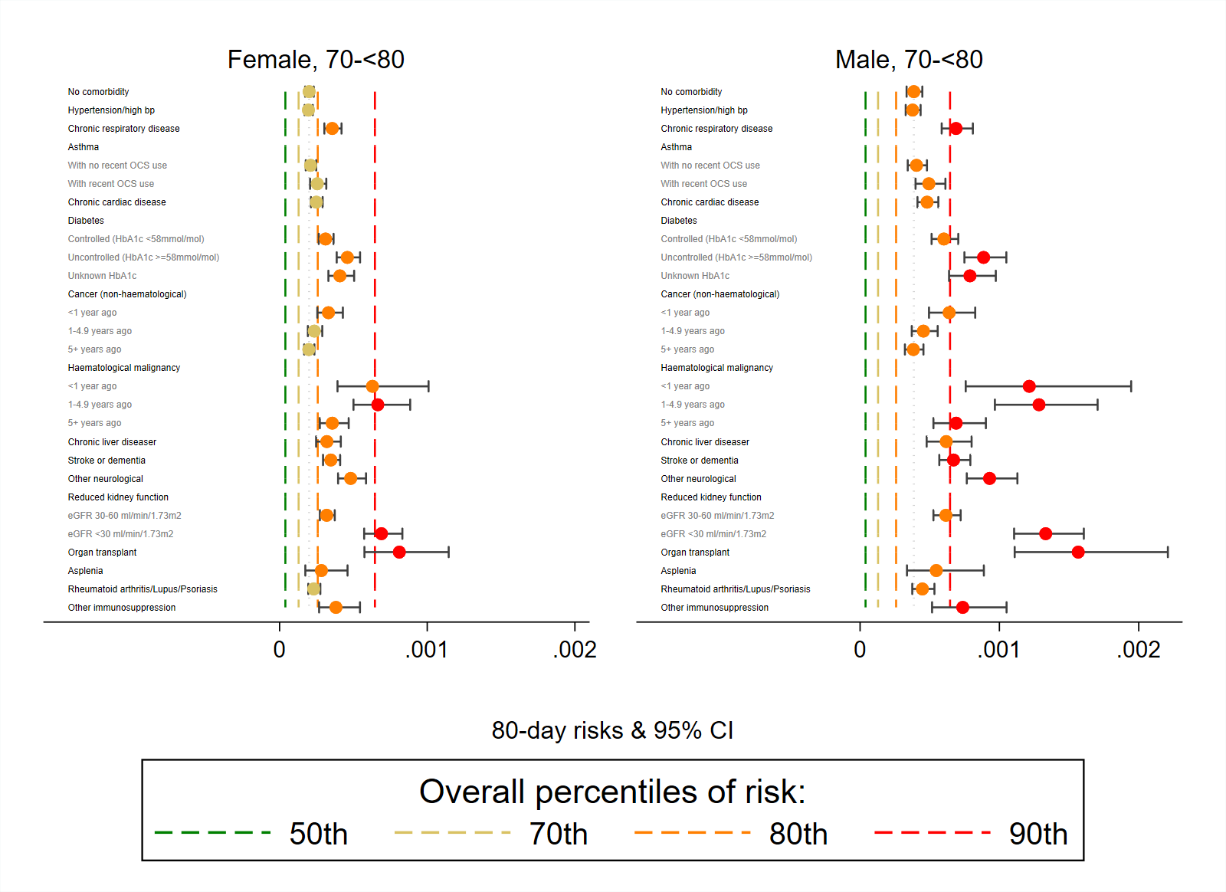


Figure 4e. Estimated 80-day absolute risk of COVID-19 related death for patients 70-<80 years old, by sex, with 95% confidence intervals. Risks are estimated assuming a single comorbidity is present, with a baseline set of demographic characteristics as described in the methods section.

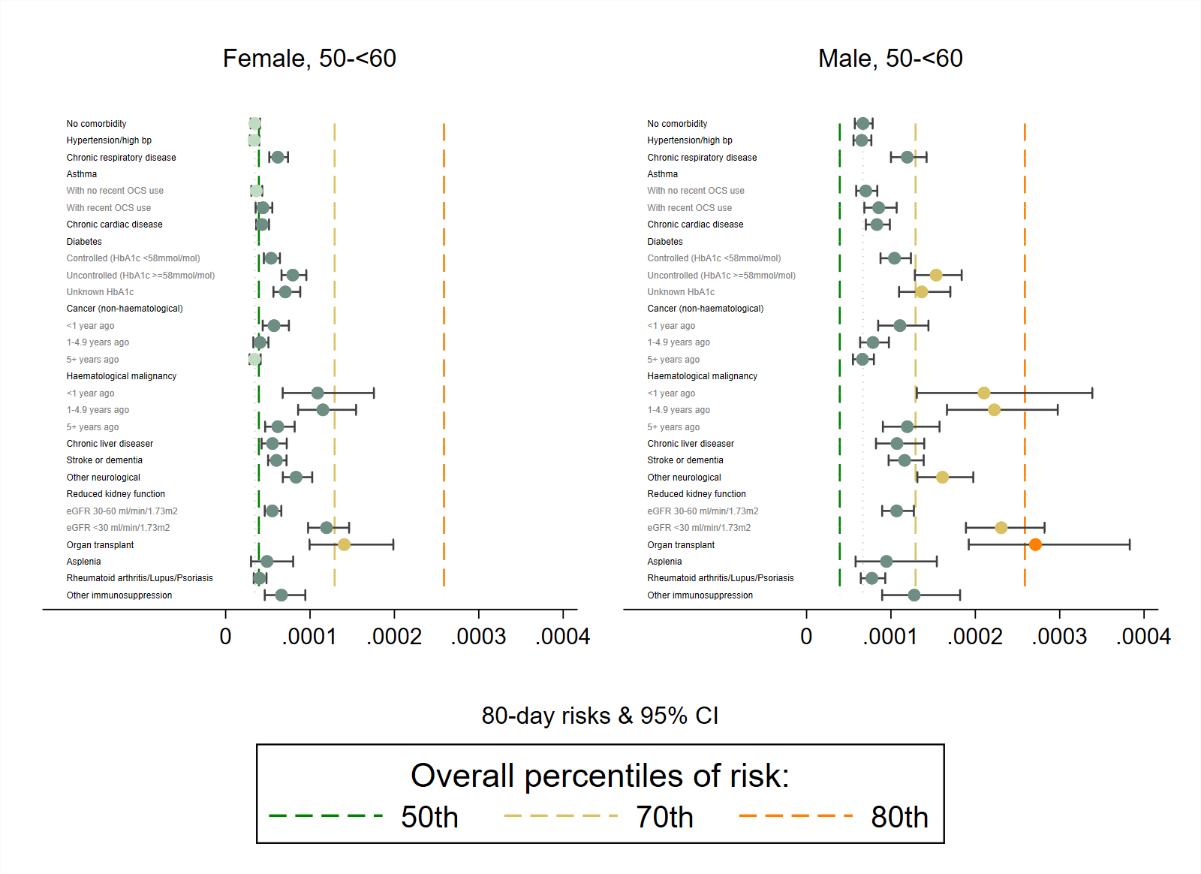


Figure 5a. Reproduction of Figure 4c (Estimated 80-day absolute risk of COVID-19 related death, assuming a single comorbidity is present, with a baseline set of demographic characteristics) for ease of comparison with Figure 5b below.

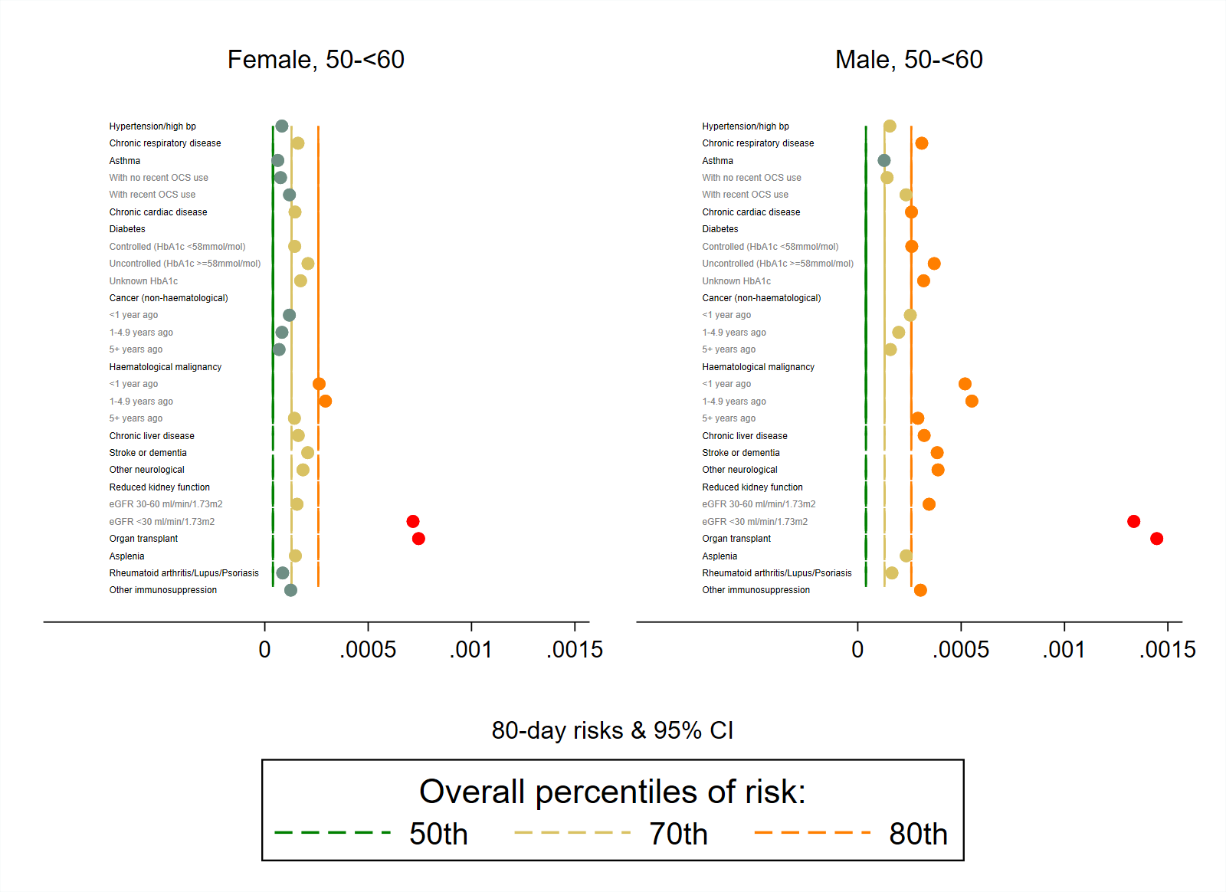


Figure 5b. Observed absolute risks for patients with each comorbidity (irrespective of presence or absence of other comorbidities) plotted for comparison with Figure 5a.

Subsequently, these analyses were re-run as more up-to-date outcome data became available. The updated figures for White ethnicity can be found in Figures 6a-6f below. The patterns that appear are very similar. The main differences are in slight changes in the location of the overall risk percentile estimates, finer age-groups in the older ages and the splitting of HIV from other immunosuppression.

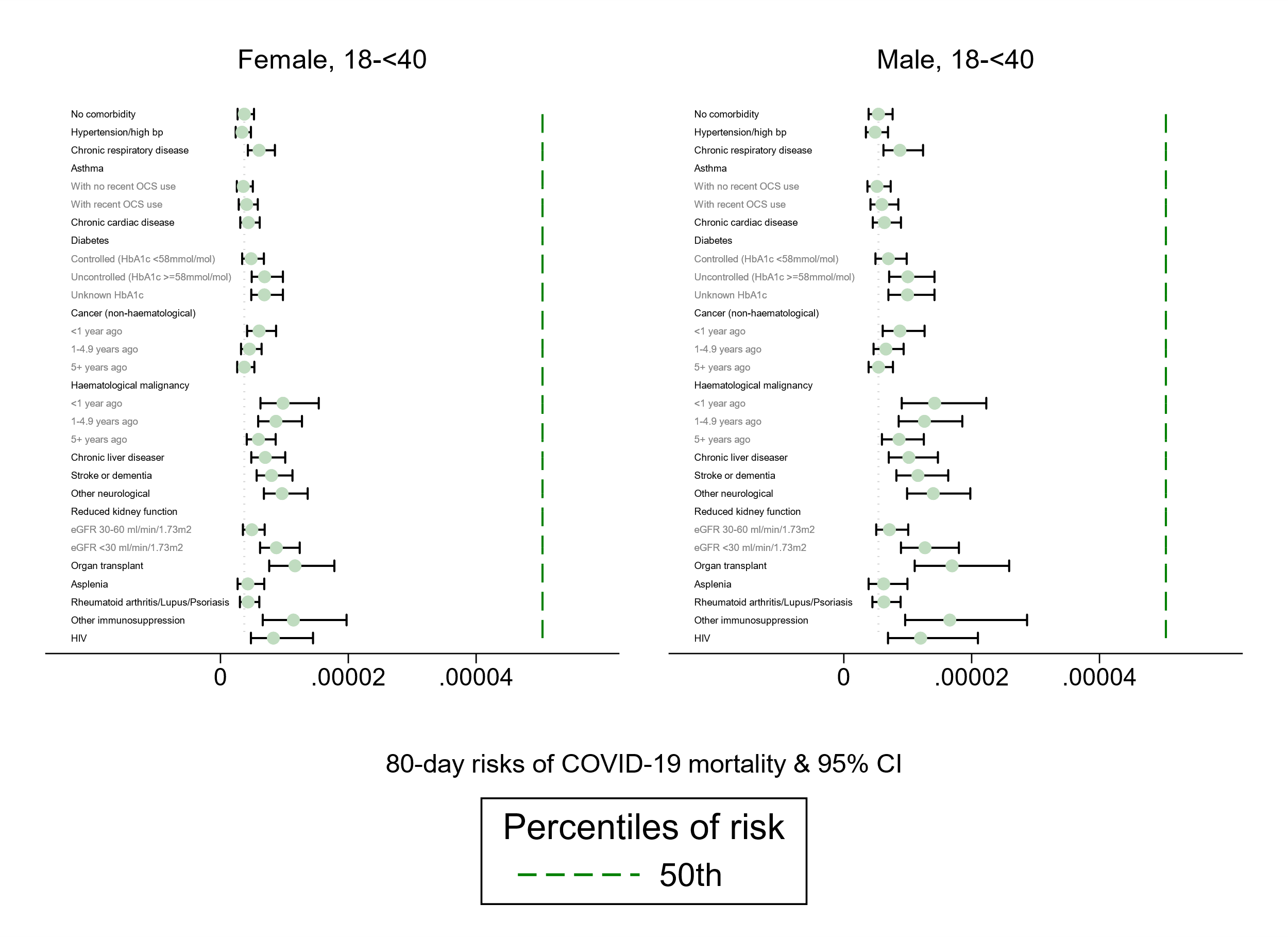


Figure 6a. Estimated 80-day absolute risk of COVID-19 related death for patients 18-<40 years old, by sex, with 95% confidence intervals. Risks are estimated assuming a single comorbidity is present, with a baseline set of demographic characteristics as described in the methods section.

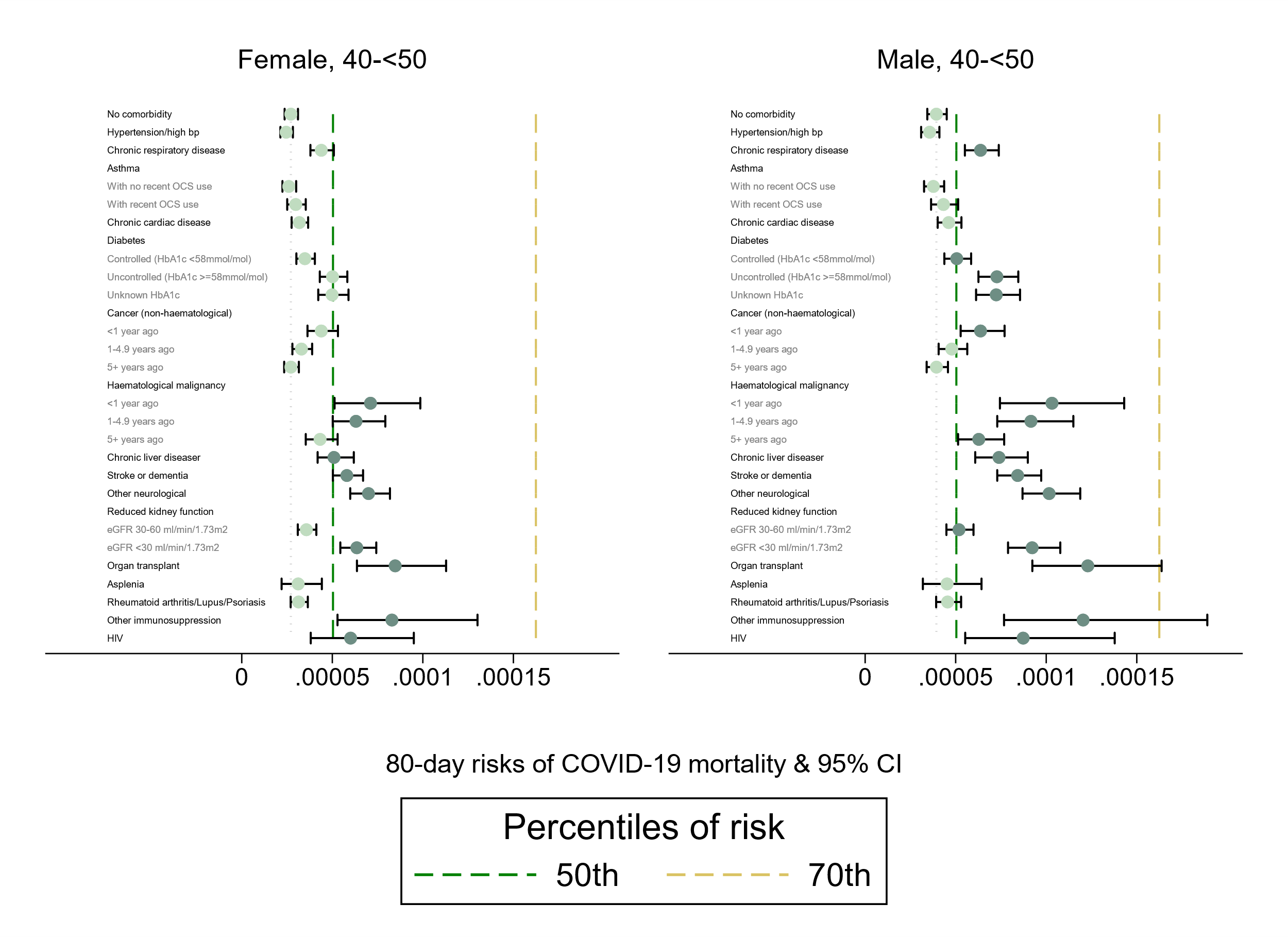


Figure 6b. Estimated 80-day absolute risk of COVID-19 related death for patients 40-<50 years old, by sex, with 95% confidence intervals, for patients in the White ethnicity group. Risks are estimated assuming a single comorbidity is present, with a baseline set of demographic characteristics as described in the methods section.

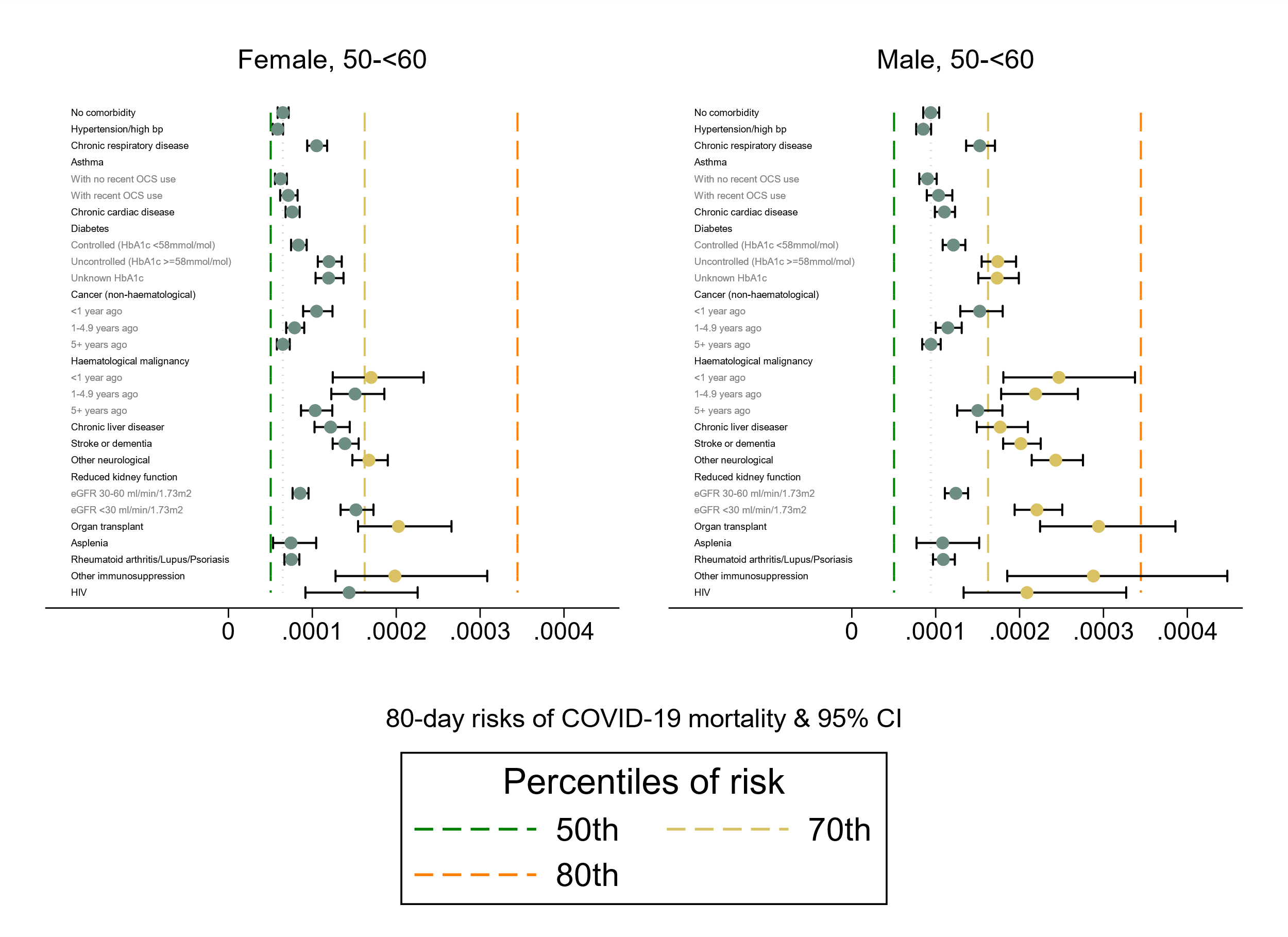


Figure 6c. Estimated 80-day absolute risk of COVID-19 related death for patients 50-<60 years old, by sex, with 95% confidence intervals, for patients in the White ethnicity group. Risks are estimated assuming a single comorbidity is present, with a baseline set of demographic characteristics as described in the methods section.

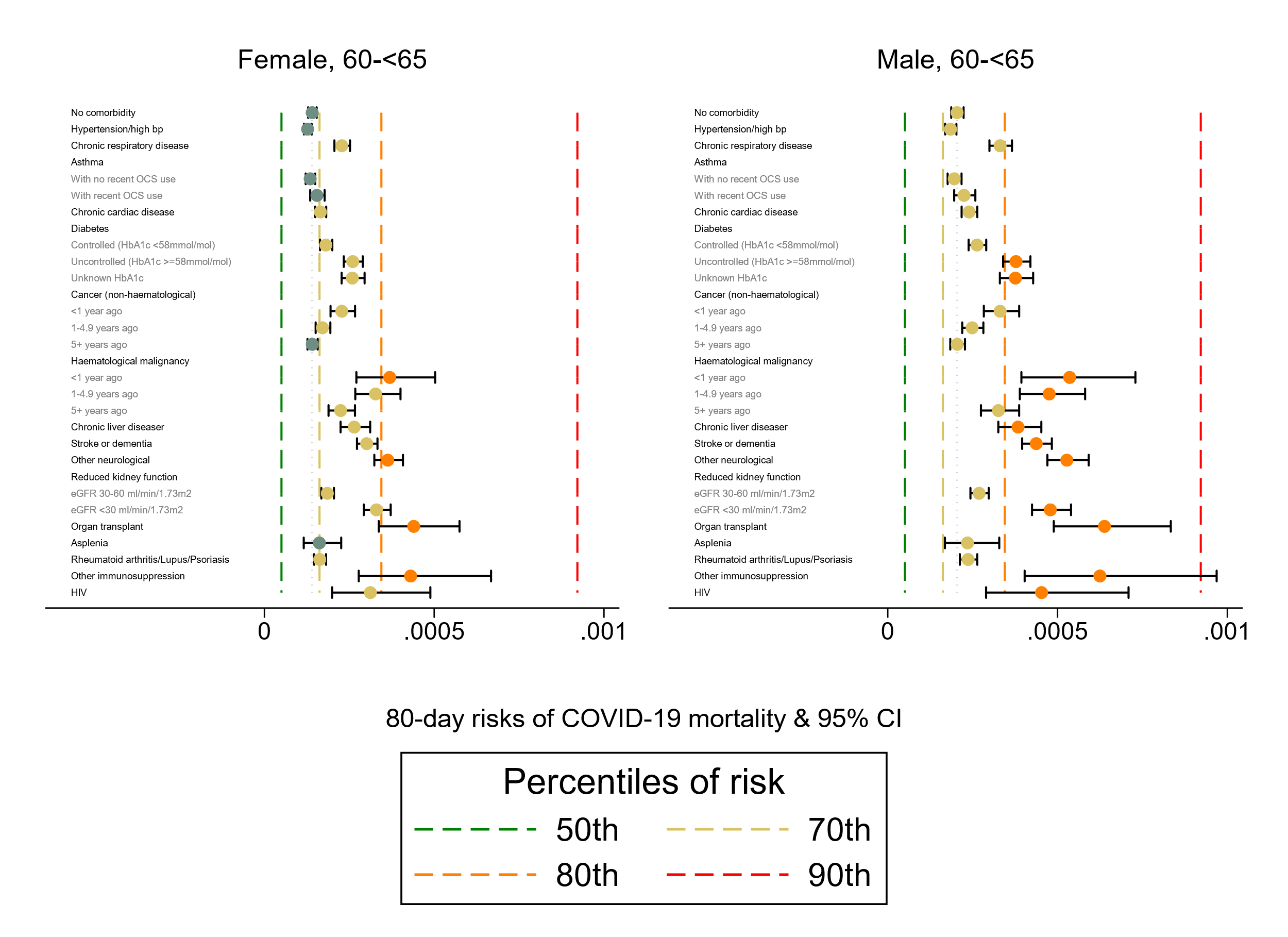


Figure 6d. Estimated 80-day absolute risk of COVID-19 related death for patients 60-<65 years old, by sex, with 95% confidence intervals, for patients in the White ethnicity group. Risks are estimated assuming a single comorbidity is present, with a baseline set of demographic characteristics as described in the methods section.

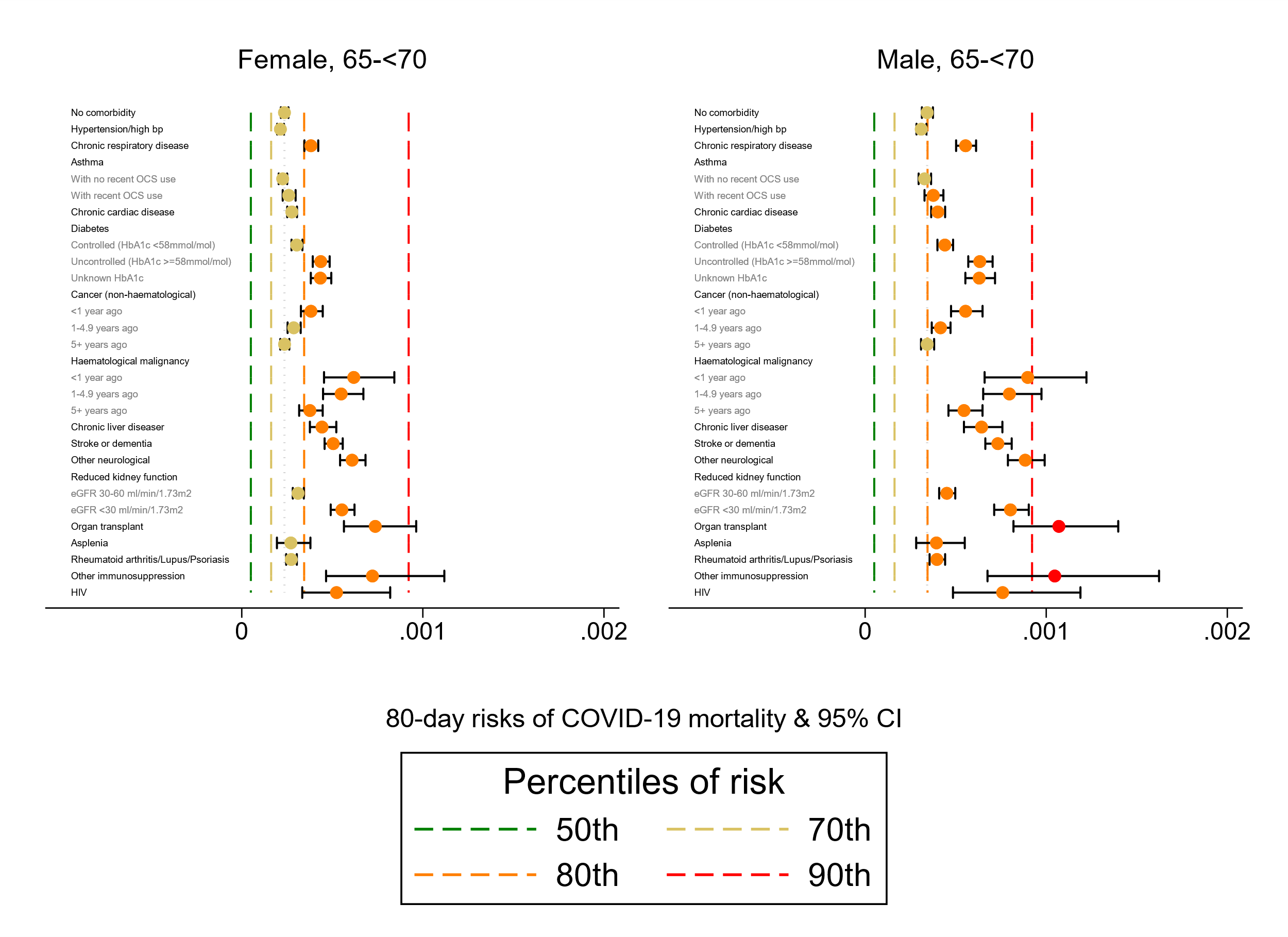


Figure 6e. Estimated 80-day absolute risk of COVID-19 related death for patients 65-<70 years old, by sex, with 95% confidence intervals, for patients in the White ethnicity group. Risks are estimated assuming a single comorbidity is present, with a baseline set of demographic characteristics as described in the methods section.

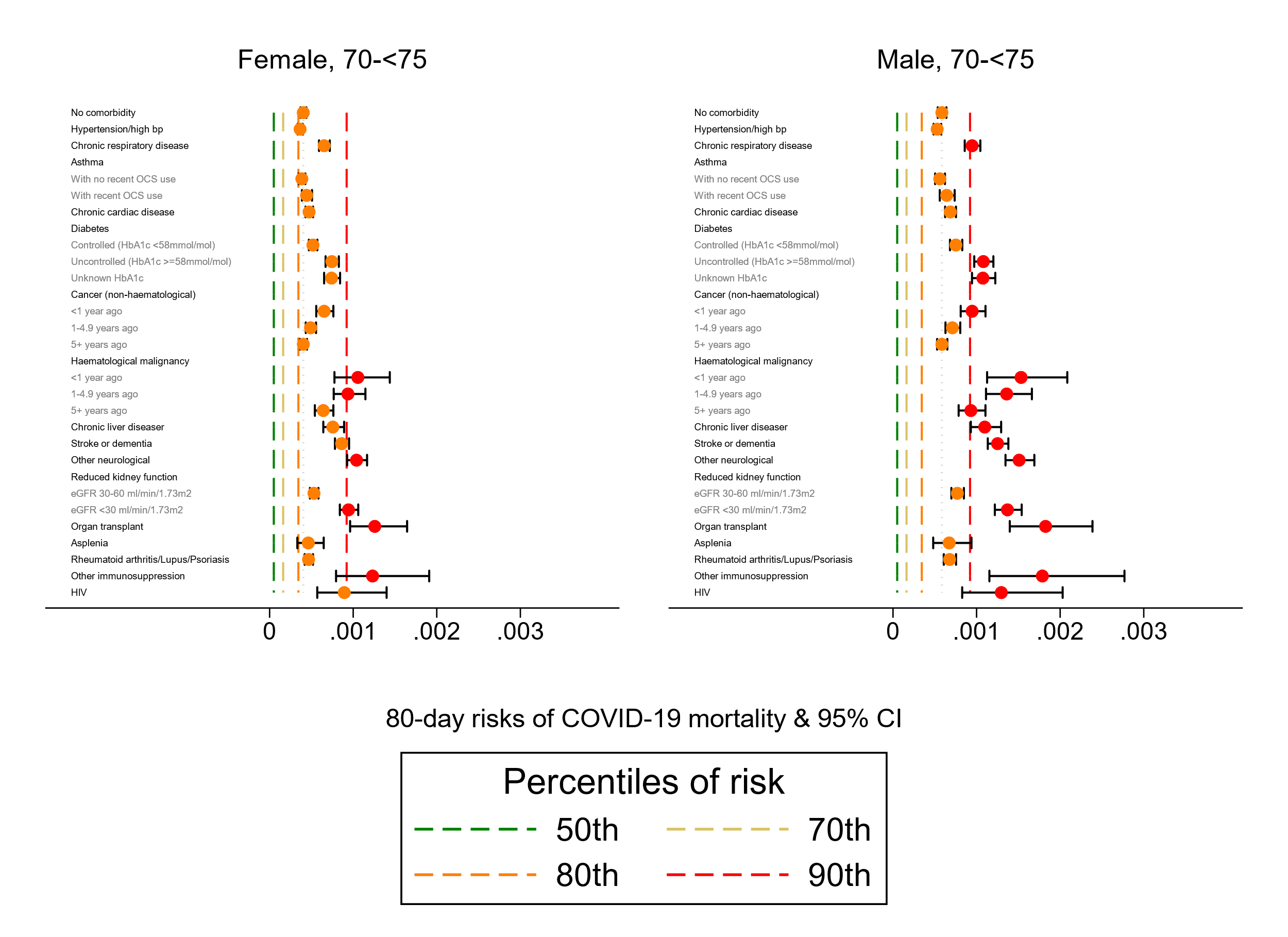


Figure 6e. Estimated 80-day absolute risk of COVID-19 related death for patients 70-<75 years old, by sex, with 95% confidence intervals, for patients in the White ethnicity group. Risks are estimated assuming a single comorbidity is present, with a baseline set of demographic characteristics as described in the methods section.

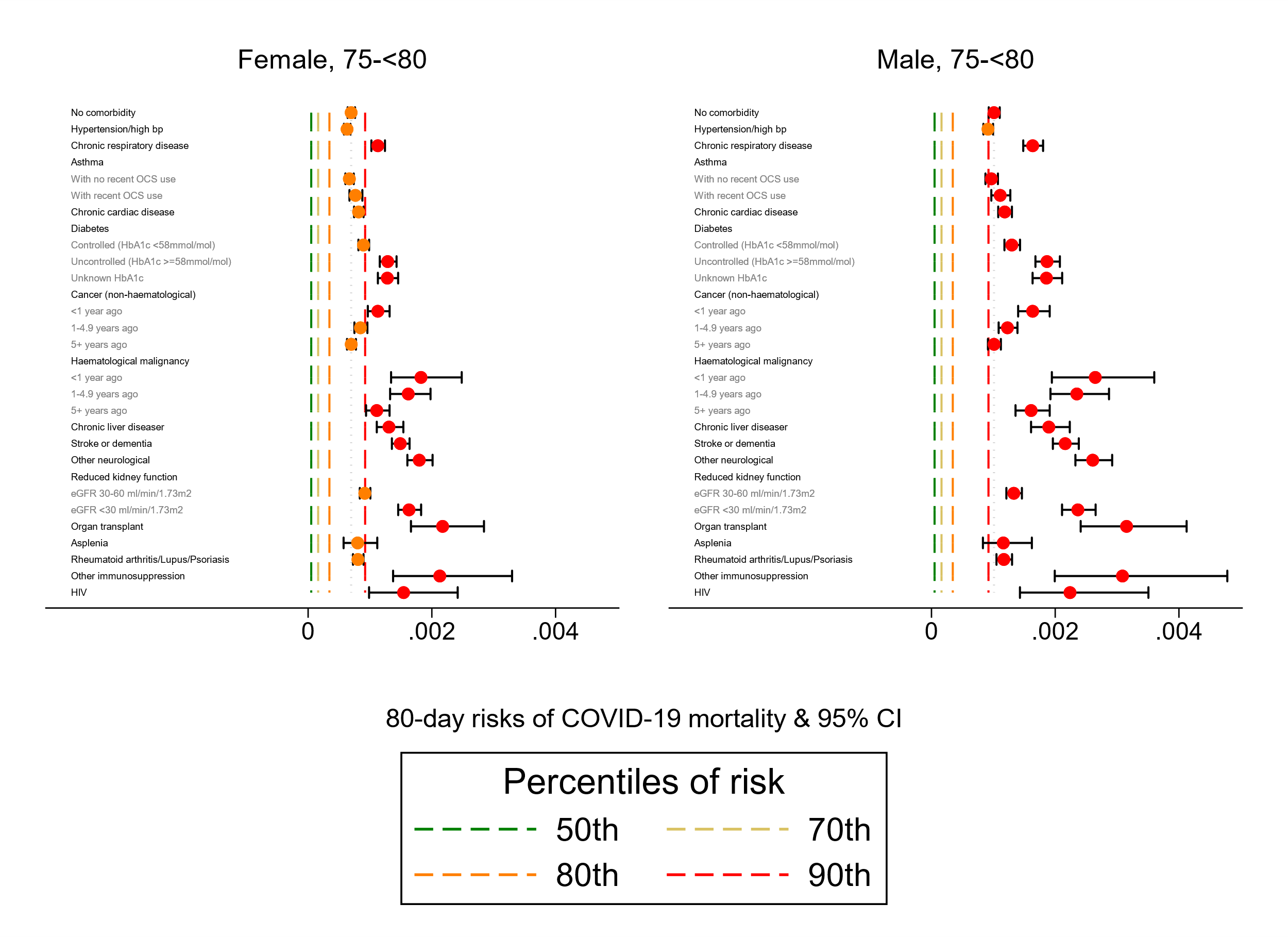


Figure 6e. Estimated 80-day absolute risk of COVID-19 related death for patients 75-<80 years old, by sex, with 95% confidence intervals, for patients in the White ethnicity group. Risks are estimated assuming a single comorbidity is present, with a baseline set of demographic characteristics as described in the methods section.

# Results: Analysis 2

In the extended time-frame (from follow-up to 6th May 2020 to 24th June 2020 for COVID-19 death, and 24th May 2020 for hospital admissions), there were 11,174 COVID-19 related deaths and 21,929 COVID-19 related hospital admissions.

Figures 7a and 7b show the estimated absolute risk profiles over age for people with each individual comorbidity. The dotted flat line represents the reference of mean risk experienced at age 65 for a person of the same sex and ethnicity. Generally, we see higher risks, particularly at older ages, associated with immunosuppression, haematological cancers, organ transplant and poor kidney function. Lower risks were seen for respiratory diseases and asthma, with cardiac conditions appearing somewhere in between.

In contrast to the graph for COVID-19 related mortality (Figure 7a), which shows absolute risks that remain low until around age 60 and then rapidly increase, there is a much greater variation in risk of hospital admission at younger ages with an initial increase around age 30 followed by a slight plateau and then a later increase. For certain comorbidities, such as immunosuppression, the reference level of age-65 risk was attained by patients around 30 years old.

Graphs similar to Figure 7a and 7b showing smaller sets of comorbidities can be found on the relevant Github pages.

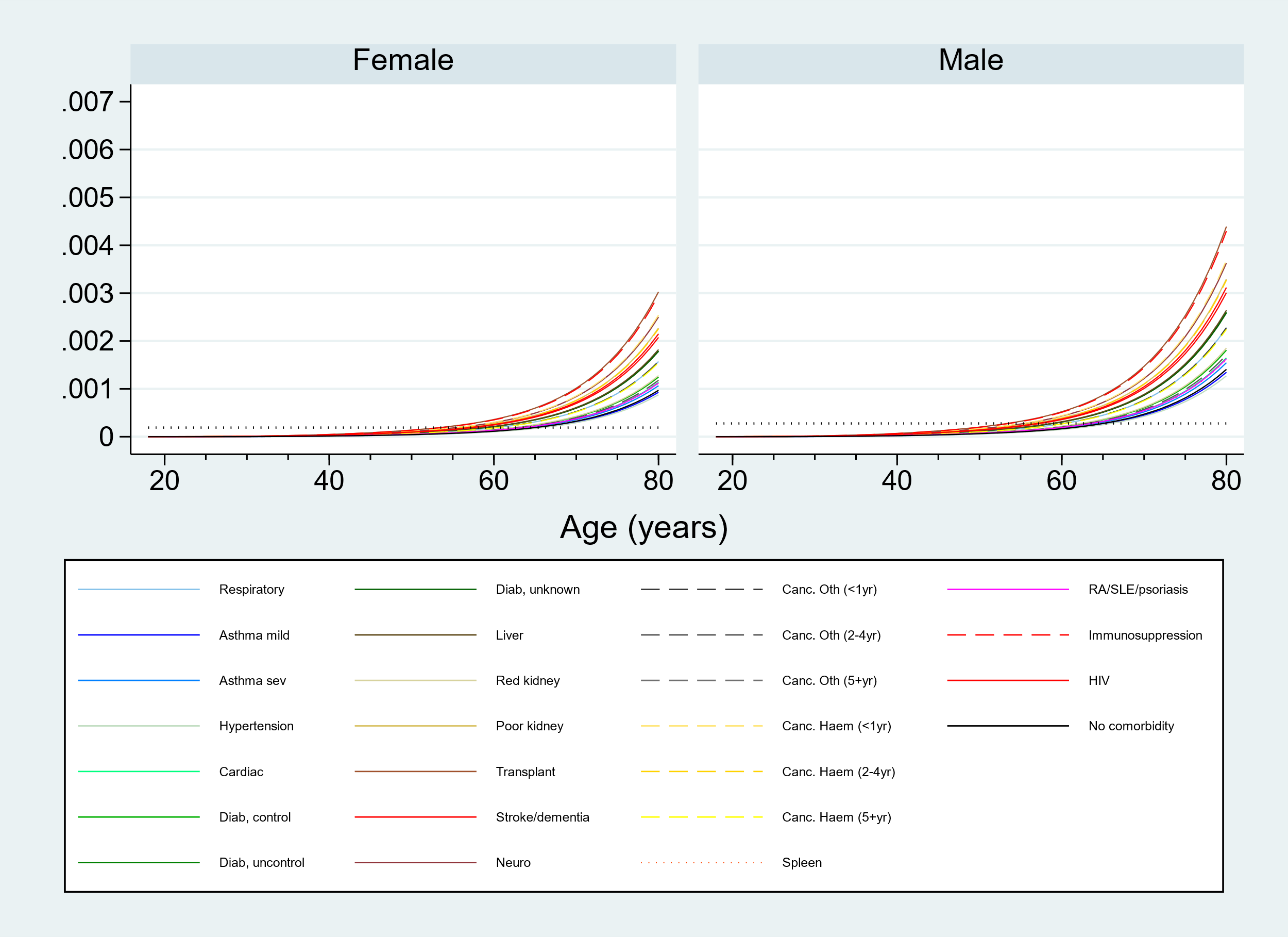


Figure 7a. Estimated 80-day absolute risk of COVID-19 related death by age and sex, for patients in the White ethnicity group. Risks are estimated assuming a single comorbidity is present, with a baseline set of demographic characteristics as described in the methods section. The flat dotted line is a reference, representing the mean risk experienced by people of that sex aged 65.

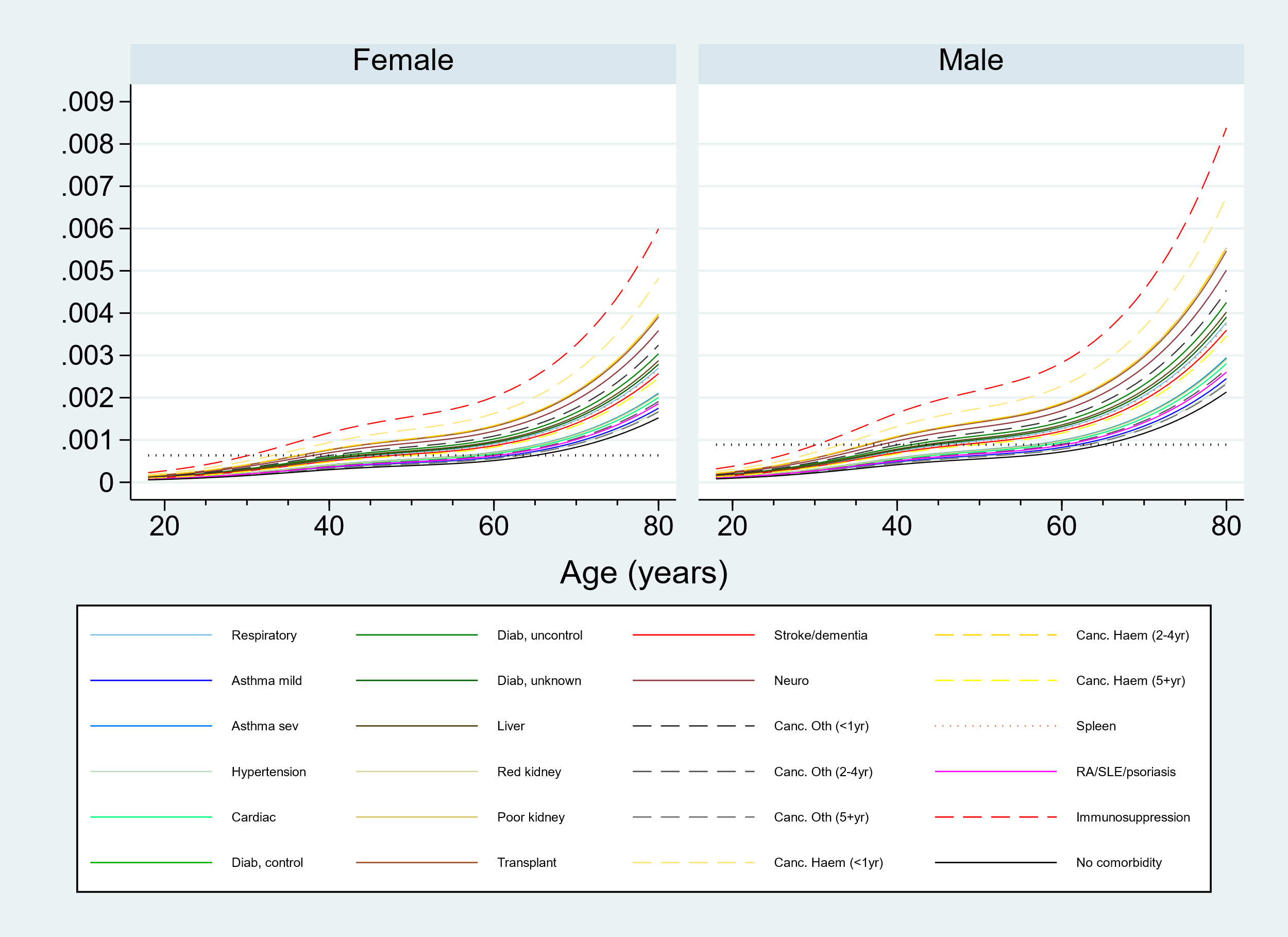


Figure 7b. Estimated 80-day absolute risk of COVID-19 related hospital admission by age and sex, for patients in the White ethnicity group. Risks are estimated assuming a single comorbidity is present, with a baseline set of demographic characteristics as described in the methods section. The flat dotted line is a reference, representing the mean risk experienced by people of that sex aged 65.

Table 2 below shows approximate ages at which age-65 absolute 80-day risk of COVID-19 related mortality and hospital admission was estimated to be experienced by people with individual comorbidities. These numbers illustrate that there is little variability in this age across comorbidities in terms of death, illustrating the huge impact of age on absolute risk, but a much greater variability for hospital admission. Patients in their 30s with immunosuppression, recent haematological cancer diagnosis, poor kidney function or a previous organ transplant, are all estimated to experience the same 80-day risk of COVID-19 hospitalisation as the average 65-year old of the same sex and ethnicity.

Table 2: Approximate age at which “65-year old risk” is estimated to be experienced by people with each individual comorbidity.

|  |  |  |
| --- | --- | --- |
| Comorbidity | Approximate age at which “65-year old risk” is attained | |
|  | COVID-19 related death | COVID-19 related hospital admission |
| Respiratory disease | 60 | 45 |
| Asthma, mild | 65 | 50 |
| Asthma, severe | 65 | 60 |
| Hypertension/High blood pressure | 65 | 63 |
| Diabetes, controlled | 62 | 56 |
| Diabetes, uncontrolled | 56 | 43 |
| Diabetes, unknown | 56 | 43 |
| Cardiac disease | 64 | 55 |
| Organ transplant | 52 | 35 |
| Kidney function, reduced | 61 | 55 |
| Kidney function, poor | 55 | 36 |
| Liver disease | 58 | 44 |
| Stroke/dementia | 57 | 50 |
| Other neurological condition | 55 | 37 |
| Cancer Haematological, <1 year | 55 | 34 |
| Cancer Haematological, 2-4 year | 55 | 35 |
| Cancer Haematological, 5+ year | 60 | 48 |
| Cancer Other, <1 year | 60 | 40 |
| Cancer Other, 2-4 year | 61 | 50 |
| Cancer Other, 5+ year | 65 | 62 |
| Spleen | 62 | 43 |
| RA/SLE/Psoriasis | 63 | 63 |
| Immunosuppression | 57 | 30 |

RA rheumatoid arthritis, SLE systemic lupus erythematosus.

Figures 8a-8d show estimated 80-day risks of COVID-19 related death and hospital admission by age and comorbidity count, by for patients in the White and Black ethnicity groups. Similar graphs for the other ethnicity groups can be found on the relevant Github pages. The flat dotted lines represent the reference level of mean risk experienced at age 65 by people of the same ethnicity and sex.

The comorbidity count includes: respiratory disease, severe asthma, chronic cardiac disease, diabetes, non-haematological cancer (diagnosed in last year), haematological cancer (diagnosed within 5 years), liver disease, stroke, dementia, poor kidney function, organ transplant, asplenia, other immunosuppression

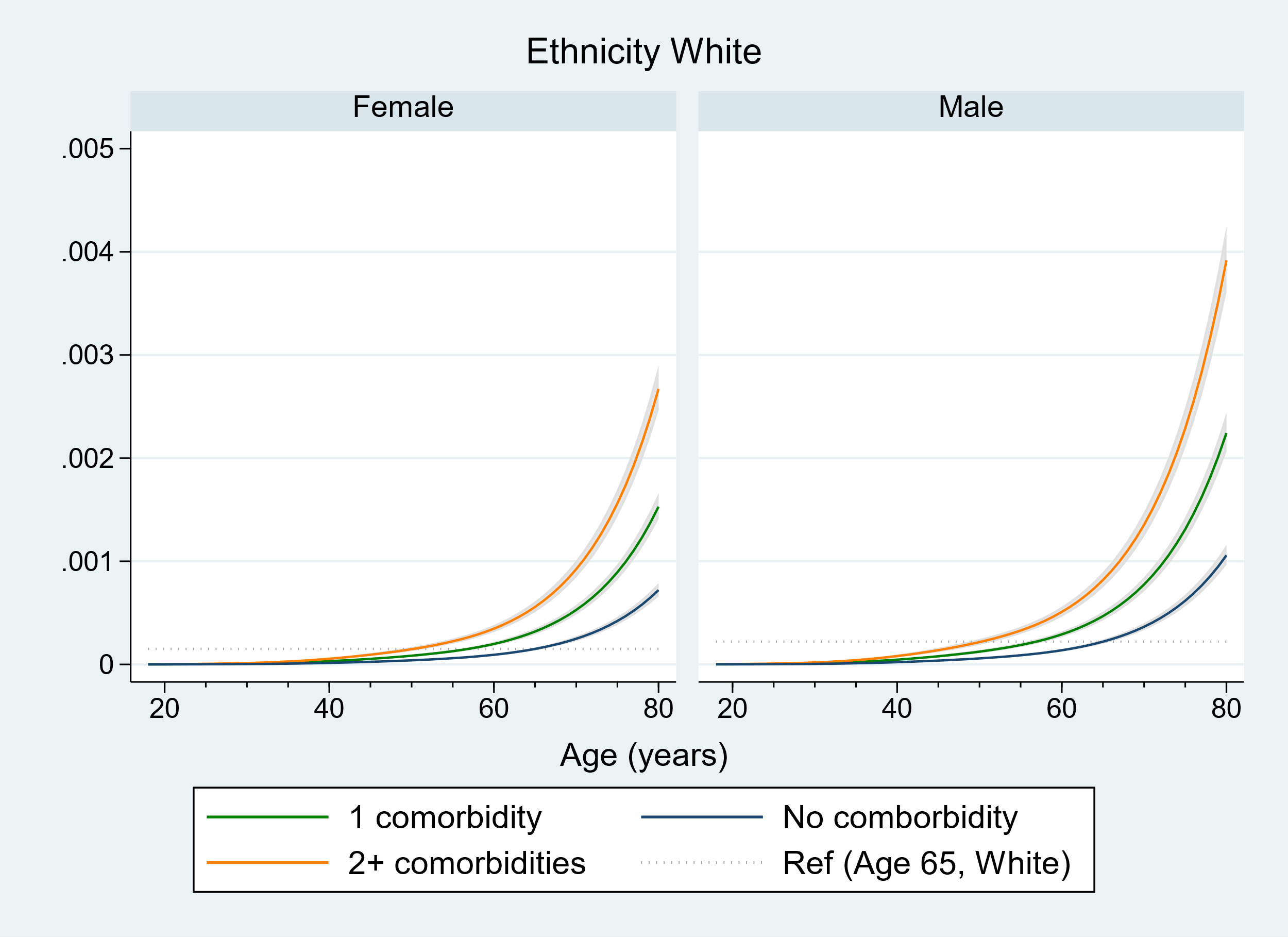


Figure 8a. Estimated 80-day absolute risk of COVID-19 related death by age and sex, for patients in the White ethnicity group. Risks are estimated assuming a baseline set of demographic characteristics as described in the methods section. The flat dotted line is a reference, representing the mean risk experienced by people of that sex aged 65.

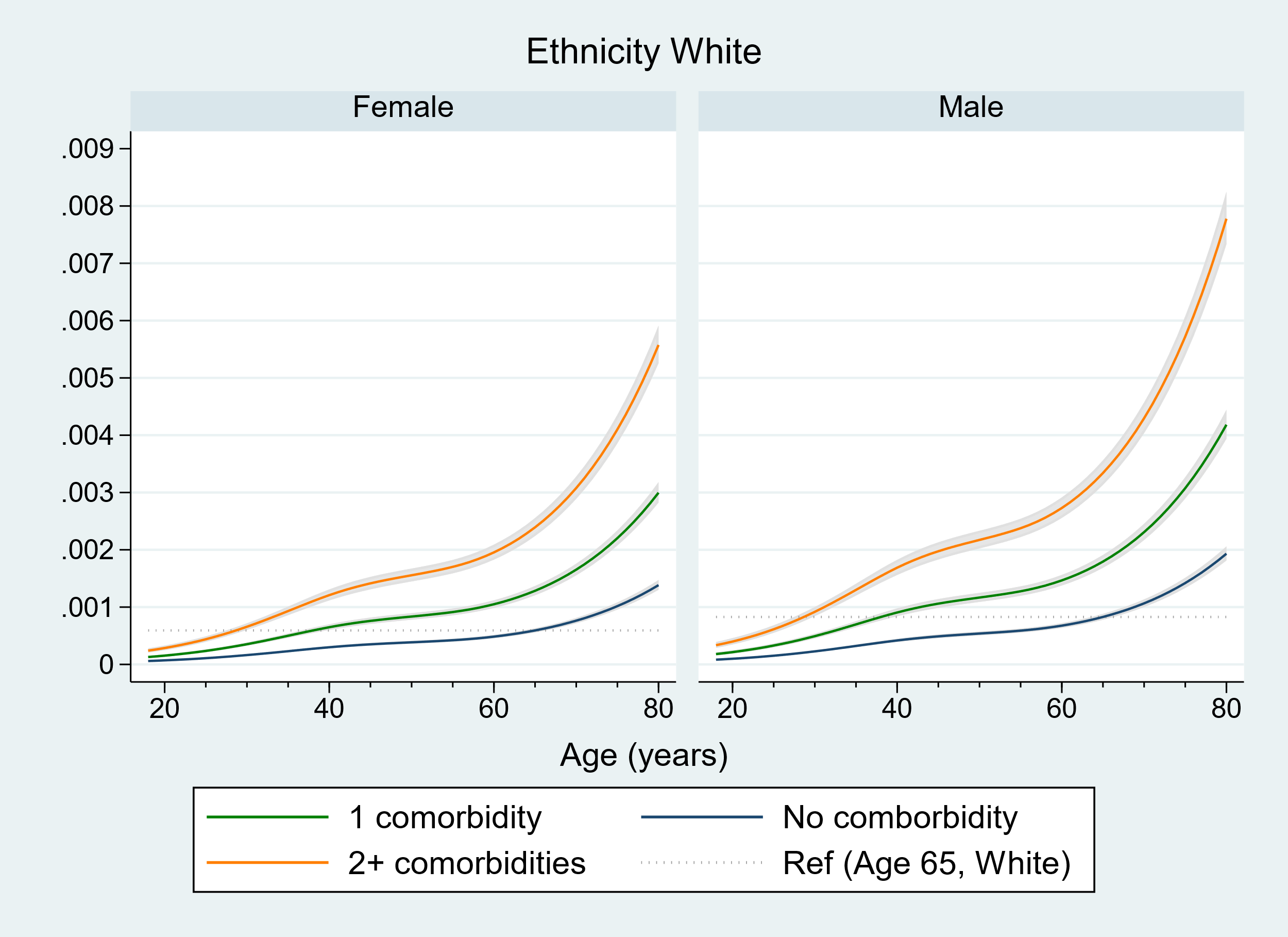


Figure 8b. Estimated 80-day absolute risk of COVID-19 related hospital admission by age and sex, for patients in the White ethnicity group. Risks are estimated assuming a baseline set of demographic characteristics as described in the methods section. The flat dotted line is a reference, representing the mean risk experienced by people of that sex aged 65.

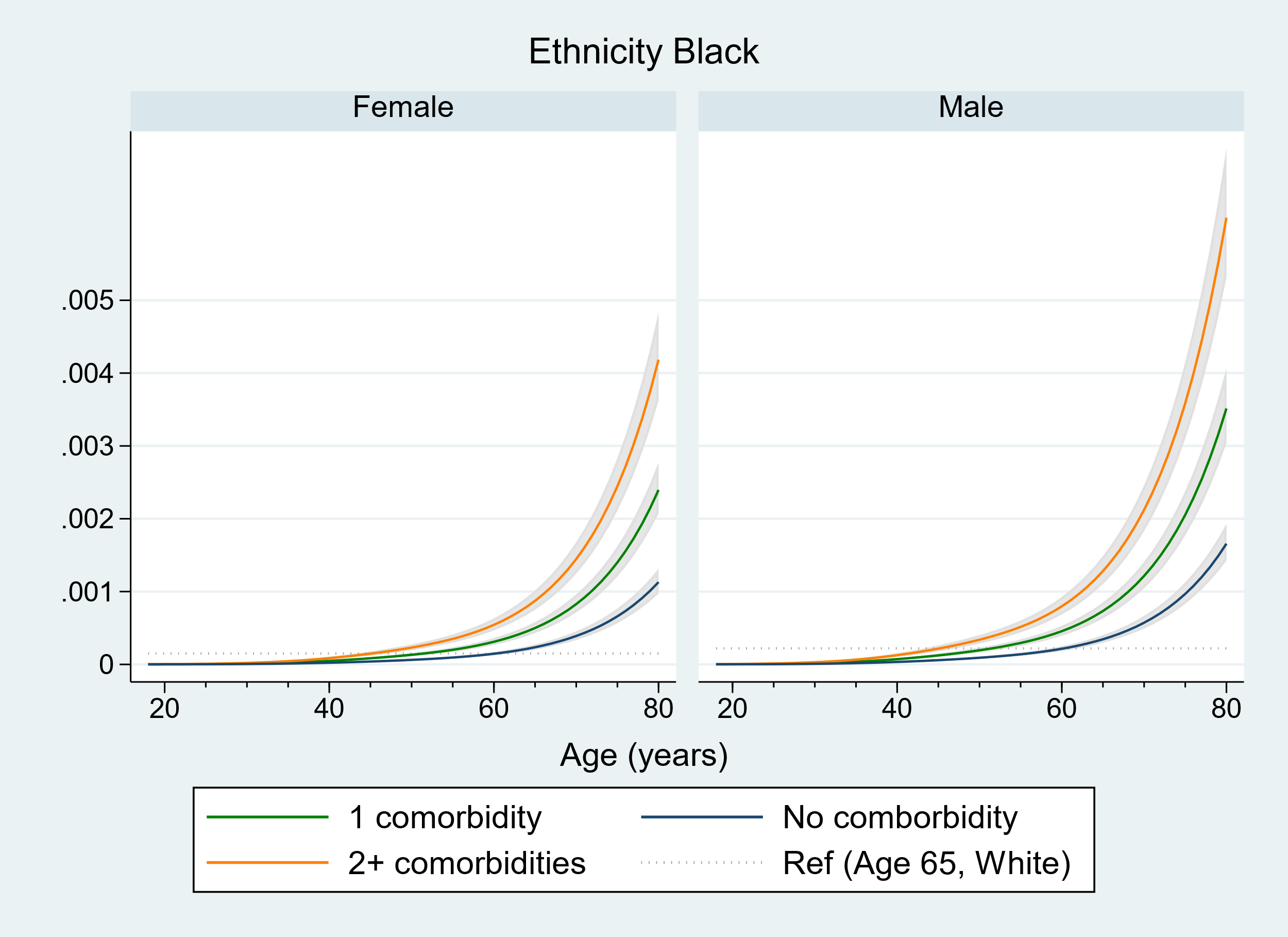


Figure 8c. Estimated 80-day absolute risk of COVID-19 related death by age and sex, for patients in the Black ethnicity group. Risks are estimated assuming a baseline set of demographic characteristics as described in the methods section. The flat dotted line is a reference, representing the mean risk experienced by people of that sex aged 65.

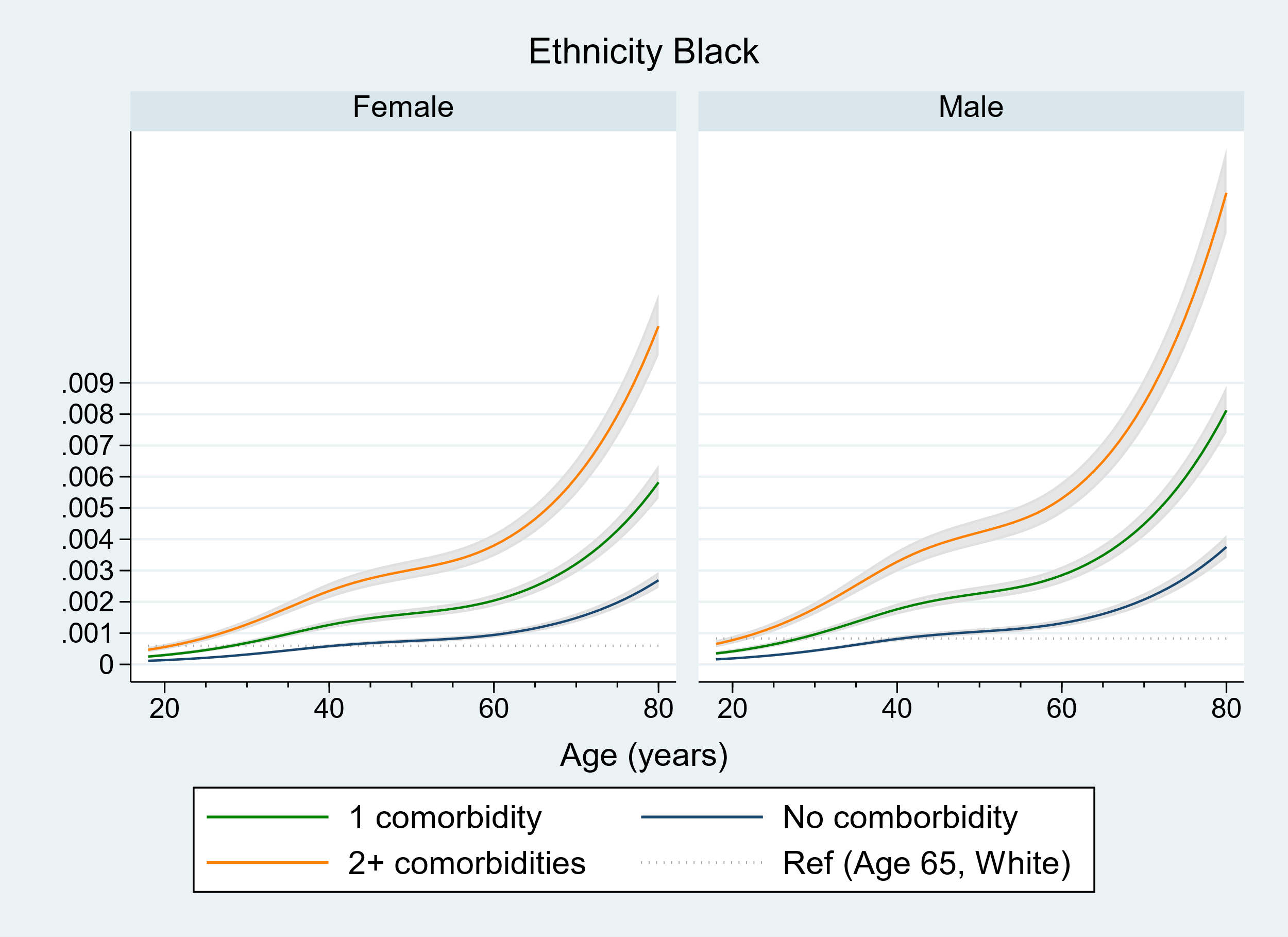


Figure 8d. Estimated 80-day absolute risk of COVID-19 related hospital admission by age and sex, for patients in the Black ethnicity group. Risks are estimated assuming a baseline set of demographic characteristics as described in the methods section. The flat dotted line is a reference, representing the mean risk experienced by people of that sex aged 65.

# Discussion

## Summary of findings

Overall, this secure analytics platform operating across over 23 million patient records for the COVID-19 emergency was used to identify, quantify, and explore factors associated with COVID-19 related death in the largest cohort study conducted by any country to date. By far the biggest risk factor is age. Most comorbidities were associated with increased risk, including cardiovascular disease, diabetes, respiratory disease including severe asthma, obesity, history of haematological malignancy or recent other cancer, kidney, liver, neurological and autoimmune conditions.

Exploring risks on an absolute scale, as opposed to the relative scale giving rise to these conclusions, however, illustrated the fact that while individual comorbidities were estimated to increase risk of COVID-19 related death (on a relative scale) by large amounts, these often did not result in a patient with only this comorbidity having a large absolute risk (defined as being over the 80th percentile of risk overall). In fact, since age is such a strong risk factor, many of the comorbidities with a lower relative risk were sufficient to push an older person with only this comorbidity into the “estimated to have high absolute risk” category, whereas the comorbidities with very high relative risks were still insufficient to result in high estimated absolute risk among younger patients, on the basis of individual comorbidities.

Extending the analysis to consider COVID-19 related hospital admission revealed that unlike absolute risk of COVID-19 death, which remains fairly low for patients with all comorbidities until around aged 60 before a sharp increase, the absolute risk of COVID-19 hospital admission varies widely across different comorbidities for much younger patients before slightly plateauing and then rising again. This means that, for certain comorbidities, patients as young as 30 were estimated to experience the same risk of COVID-19 related hospital admission as the average 65-year old of the same sex.

## Limitations

We are unable to separate out process of infection from process of death once infected, thus these results all reflect the combined processes of becoming infected and experiencing a COVID-19 related death once infected. The results do not take into account changing burden of infection (although the relative risk estimates are geographically stratified to try to mitigate effects of different infection burden in different geographical areas).

Importantly, the individual relative risk estimates (hazard ratios) cannot be interpreted as causal effects, nor can comparisons of absolute risks for different comorbidities be interpreted as causal effects.

Certain factors could not be taken into account. No data were available on occupation or household conditions or composition, at the time of these analyses.

There may be outcome misclassification (e.g. including clinically suspected COVID-19 related deaths), particularly since these data are from the earlier part of the pandemic.

In terms of geography, the cohort data are not fully representative. Due to the locations using the TPP software on which the platform is based, there is relatively little data from London.

# JCVI reports

These analyses were presented to JCVI in a number of presentations from October 2020 to February 2021. Key slide sets are contained in:

* OpenSAFELY\_JCVI\_151020.ppt.
* OpenSAFELY\_JCVI\_ar\_16\_02\_2021.ppt

# Availability of code

All code used to generate the graphs and statistics within these reports can be found in the following public repository:

<https://github.com/opensafely/risk-factors-research>

The table below lists the Stata scripts used to fit the models used in the analyses presented herein and to create the graphs.

|  |  |  |
| --- | --- | --- |
| *Task* | *Do-file used for modellinga* | *Do-file used to create graphsa* |
| Create graphs in analysis 1 | xv2j1\_absolute\_risk\_model.do | xv2j2\_graph\_absolute\_risk\_model.do |
| * Initial analysisb | xj1\_absolute\_risk\_model.do | xj2\_graph\_absolute\_risk\_model.do |
| * Subsequent updateb | xv2j1\_absolute\_risk\_model.do 1 death | xv2j2\_graph\_absolute\_risk\_model.do 1 death |
| Graphs of absolute risk over age, by ethnicity and comorbidity | xv2j3\_absolute\_risk\_model\_fineage.do | xv2j4\_graph\_absolute\_risk\_model\_fineage.do |
| Graphs of absolute risk by ethnicity and comorbidity count | xv2j5\_absolute\_risk\_model\_fineage\_group.do | xv2j6\_graph\_absolute\_risk\_model\_fineage\_group.do |

a These do-files require parameters indicating the outcome (death or hospitalisation) and ethnicity group of interest (entered as numbers)

b In the intervening time between analyses exploring the absolute risks, the data that fed in to the Nature paper (Williamson, Nature, 2020) had been refreshed with more recent outcome data. This was used and cleaned in a subsequent analysis focusing on HIV, so the subsequent do-files all used the analysis dataset from this analysis.

# References

Williamson EJ, Walker AJ, Bhaskaran K, Bacon S, Bates C, Morton CE, Curtis HJ, Mehrkar A, Evans D, Inglesby P, Cockburn J, McDonald HI, MacKenna B, Tomlinson L, Douglas IJ, Rentsch CT, Mathur R, Wong AYS, Grieve R, Harrison D, Forbes H, Schultze A, Croker R, Parry J, Hester F, Harper S, Perera R, Evans SJW, Smeeth L, Goldacre B. Factors associated with COVID-19-related death using OpenSAFELY. Nature. 2020 Aug;584(7821):430-436. doi: 10.1038/s41586-020-2521-4. Epub 2020 Jul 8. PMID: 32640463; PMCID: PMC7611074.

NHS Digital. GP systems of choice. <https://digital.nhs.uk/services/gp-systems-of-choice> (2020).

NHS Digital. Future GP IT systems and services. <https://digital.nhs.uk/services/future-gp-it-systems-and-services> (2020).

NHS Digital. Data security and protection toolkit. <https://digital.nhs.uk/data-and-information/looking-after-information/data-security-and-information-governance/data-security-and-protection-toolkit> (2018).

NHS Digital. BETA – data security standards. <https://digital.nhs.uk/about-nhs-digital/our-work/nhs-digital-data-and-technology-standards/framework/beta---data-security-standards> (2020).

NHS Digital. ISB1523: Anonymisation standard for publishing health and social care data. <https://digital.nhs.uk/data-and-information/information-standards/information-standards-and-data-collections-including-extractions/publications-and-notifications/standards-and-collections/isb1523-anonymisation-standard-for-publishing-health-and-social-care-data> (2019).

Department of Health and Social Care. Coronavirus (COVID-19): notification to organisations to share information. [https://web.archive.org/web/20200421171727/https://www.gov.uk/government/publications/coronavirus-covid-19-notification-of-data-controllers-to-share-information](https://web.archive.org/web/20200421171727/https:/www.gov.uk/government/publications/coronavirus-covid-19-notification-of-data-controllers-to-share-information) (2020).