Effect of Sociodemographic Characteristics on Longitudinal Progression of Multimorbidity: A Multistate Modelling Analysis of a Large Primary Care Records Dataset

Sida Chen (sida.chen@mrc-bsu.cam.ac.uk)

MRC Biostatistics Unit, University of Cambridge

Tom Marshall

Institute of Applied Health Research, University of Birmingham

Christopher Jackson

MRC Biostatistics Unit, University of Cambridge

Jennifer Cooper

Institute of Applied Health Research, University of Birmingham

Francesca Crowe

Institute of Applied Health Research, University of Birmingham

Krish Nirantharakumar

Institute of Applied Health Research, University of Birmingham

Catherine L Saunders

Department of Public Health and Primary Care, University of Cambridge

Paul Kirk

MRC Biostatistics Unit, University of Cambridge

Sylvia Richardson

MRC Biostatistics Unit, University of Cambridge

Duncan Edwards

Department of Public Health and Primary Care, University of Cambridge

Simon Griffin

Department of Public Health and Primary Care, University of Cambridge

Christopher Yau

Nuffield Department for Women's & Reproductive Health, University of Oxford and Health Data Research UK

Jessica K Barrett

MRC Biostatistics Unit, University of Cambridge

Abstract

Background:

Multimorbidity presents a growing public health challenge. While existing research has mainly focused on cross-sectional patterns of multimorbidity, there remains a need to better understand the longitudinal accumulation of diseases and the effects of important sociodemographic characteristics on the accumulation and progression of chronic conditions.

Methods and findings:

We use a spline-based parametric multistate modelling (MSM) framework for investigating the role of several sociodemographic characteristics such as ethnicity, deprivation, gender, and age on the rate of disease transitions along different pathways to multimorbidity accumulation. The model is highly flexible and permits a parallel estimation framework that scales well to very large datasets. Focusing on cardiometabolic, renal conditions and mental conditions, we analysed a dataset of about 13 million English electronic primary care records. Our results revealed complex interactions between temporal trajectories of multimorbidity and the sociodemographic factors, with effects that varied depending on the comorbidity status of the subject. These factors tended to have a larger impact on initial multimorbidity development while a smaller impact on later stages. Mental health as a comorbidity often strengthened the adverse effects of sociodemographic characteristics on disease accumulation.

Conclusion:

Our findings underscore the potential importance of earlier screening and intervention of chronic conditions, particular for mental health conditions, in the higher-risk populations. This may have important implications for the management of multimorbidity.

Background

Multimorbidity or multiple long-term conditions is defined as the co-existence of two or more chronic conditions in an individual [1]. Multimorbidity increases with age and because of the complexity of managing those with multiple health conditions, it is of increasing importance in countries with ageing populations [2]. Existing research indicates that multimorbidity is associated with poorer quality of life, worsened health outcomes, increased healthcare use and higher mortality [3,4]. Multimorbidity develops sequentially, as individuals accumulate diagnoses, but to date, research has mainly described cross-sectional patterns of multimorbidity, primarily focusing on identifying multimorbidity clusters [5–7]. We need to understand how diseases accumulate over time and the effects of potentially important sociodemographic characteristics on the accumulation and progression of diagnoses. This may help identify those at high risk of developing further diseases. A recent systematic scoping review summarised research on longitudinal modelling and analysis of multimorbidity patterns, see [8]. One of these analyses investigated the sequence of occurrence of diseases using a visualisation approach, stratified by deprivation status, age and ethnicity [9]. They reported that the more socially deprived had a greater risk for multimorbidity and that ethnicity was associated with the development of multimorbidity.

The multistate model (MSM) extends the traditional survival model to encompass more than two states, providing a robust framework for modelling longitudinal event history data and making predictions. MSMs have been effectively used in studies of multimorbidity, where states represent the individual's health status. In recent studies, MSMs have been applied to examine the progression of chronic conditions [10] and the impact of lifestyle factors on the accumulation of diseases [11]. In [12] the authors evaluated the effect of the order of disease acquisition on patient mortality using a 17-state MSM, while in [11] a 5-state MSM is used to study the impact of lifestyle factors on cancer and cardiometabolic diseases. MSMs have also been used to analyse socioeconomic inequalities in life expectancy with and without multimorbidity [13], the effects of sociodemographic and clinical factors on different stages of multimorbidity development [14], and the potentially differing impacts of lifestyle risk factors on cardiometabolic multimorbidity [15]. MSMs have also been used to examine the role of midlife factors on different cardiometabolic disease trajectories [16].

Most of the existing MSM models of multimorbidity have included only small numbers of conditions (e.g. three) or represent multimorbidity status as counts of conditions rather than combinations of specific conditions. The resulting MSMs thus permit a relatively simple structure and are specified either using simple parametric models such as the Weibull model, which has limited flexibility, or Cox semi-parametric models, which are incapable of absolute risk quantifications or predictions over time. While some risk factors such as lifestyle factors have been considered, detailed studies on the effects of sociodemographic characteristics on the risk of disease-specific multimorbidity progression remain scarce. In addition, previous analyses are mostly based on cohorts of relatively small sample sizes, which may limit the power of their findings relative to the number of states of the models and subgroups.

Here we propose a flexible parametric multistate modelling framework to investigate the role of several sociodemographic risk factors on the temporal progression of 5 chronic conditions and 33 states, based on a dataset of approximately 13 million English electronic primary care records. We have focused on cardiometabolic, renal conditions and mental conditions because they are common, important conditions and are known to frequently co-occur, especially in socio-economically deprived areas [17]. The considered conditions are cardiovascular disease (CVD), type 2 diabetes (T2D), chronic kidney disease (CKD), mental health problems (MH) and heart failure (HF). Our MSM extends previous works by allowing for more comorbidity states and disease transition types, providing a more detailed characterization of the pattern of progression of these five diseases.

Methods

Data sources

The study investigated the temporal sequence of diagnosed chronic conditions in a large anonymised dataset of electronic primary care records, the Clinical Practice Research Datalink (CPRD) [18]. The CPRD Aurum database contains routinely-collected primary care electronic health records from 19 million patients across the 738 English general practices that use EMIS Web® software [18]. It is representative of the English population in terms of age, gender, regional geographical spread and deprivation. It contains information on individual patient demographics. Clinical observations, diagnoses and treatments are recorded as Read Version 2, SNOMED-CT, and EMIS Web® clinical codes. The full data resource profile has been described elsewhere [18].

Conditions

Five chronic conditions were considered in this study, namely: 1) cardiovascular disease (CVD), including ischaemic heart disease (IHD), peripheral arterial disease (PAD) and stroke or transient ischaemic attack (CVA); 2) type-2 diabetes (T2D); 3) chronic kidney disease (CKD); 4) mental health problems (MH), including depression or anxiety; and 5) heart failure (HF). The presence of a chronic condition was determined by the occurrence of one of the relevant clinical codes (code lists are provided in the supplementary material) in the records, and the date of diagnosis of a condition was taken as the date of the first occurrence of any associated clinical codes. An absence of the code was considered to indicate an absence of the condition.

Sociodemographic characteristics

Sociodemographic characteristics include ethnicity, social and material deprivation, age (calculated from the year of birth) and sex (male/female). Ethnicity was initially classified into five different groups, including White, South Asian (South Asian or British Asian), Black (Black African, Black Carribean or Black British), Mixed and Others (ethnicity codings are provided in the supplementary material) [19]. In our multistate analysis described below, Mixed was merged into Others due to their infrequency in many disease transitions. Social and material deprivation was quantified by the English Index of Multiple Deprivation (IMD), a widely used area-based composite measure of individual-level deprivation [20]. IMD is derived from income, employment, education and living environment and in the CPRD Aurum data, each subject was assigned an IMD decile from 1 (least deprived) to 10 (most deprived). For our analysis, we used IMD quintiles.

Cohort description

Participants were eligible for inclusion in the study if they were registered in a general practice contributing to CPRD between the study start date and the study end date. The study start date was defined as the latest of date the subject turned 18, 01/01/2005, or six months after the patient registration start date. The study end date was defined as the earliest of the patient registration end date, date of death, or the last collection date for practice (11/05/2020). The minimum period of registration ensures time for patient characteristics to be recorded and prior diagnoses to be transferred/recorded. Patients were followed up continuously during the study period. For conditions where code lists were combined the earliest valid code in any code list was used as the date of incidence, and long-term conditions recorded before the study start date were treated as prevalent. For our analysis dataset, we excluded patients with gender classified as indeterminate, those aged >90 years at the study start date and those with missing data on ethnicity or IMD (the vast majority is due to missing information on ethnicity), see supplemental Figure A for more

details on sample selection and supplemental Figure C for results of a sensitivity analysis. The resulting cohort consists of a total of 13.48 million individuals.

Statistical method

To examine the effects of ethnicity, deprivation, age and sex on the rate of different pathways of multimorbidity accumulation, we consider a MSM with a total of 33 states and 112 allowable direct transitions (see Figure 1 for an illustration). The states incorporate all possible combinations of the five conditions a person may have during the process of disease progression and the permitted instantaneous state-to-state transitions were indicated by a direct solid line with an arrow indicating the direction of the transition. The death state is the terminal state and can be directly reached from any other state. Here we made the simplifying assumption that no more than one event per person (diagnosis or death) can happen on the same day (so the number of transitions to be estimated is greatly reduced). It is extremely rare for two or more events to take place on a single day and even the most common transitions on the same day were no more than about 5000 among the 13.47 million subjects (i.e. less than 0.04%) in the cohort. Disease transitions recorded on the same day are also likely to be artificially increased by same-day diagnostic tests or recording of diagnoses, so the accuracy of the timing of these recorded transitions is questionable. To flexibly specify the MSM, we followed the suggestions of [21] to use a separate time-to-event model for each transition, and we used a clock-reset time scale (i.e. the time scale resets to zero after every transition) so that the outcome of interest is the time from one disease diagnosis to the next diagnosis or death. More specifically, we modelled the logarithm of the cumulative transition-specific intensities as a spline function of log time [22], from which age, sex, ethnicity and deprivation adjusted hazard ratios (HR) were estimated, assuming proportional hazards and no interactions between these predictors. For each sociodemographic characteristic, we report hazard ratios (and the associated 95% confidence intervals) adjusted for the other sociodemographic characteristics; e.g. hazard ratios for the ethnic group are adjusted for age, sex and IMD. We were able to conduct inference for such a complex MSM by taking advantage of the factorization property of the model likelihood and parallel computing; see the supplementary file for additional technical details. Model fitting for different transitions is performed in parallel via maximum likelihood estimation using the flexsurv package [23] in the R software environment.

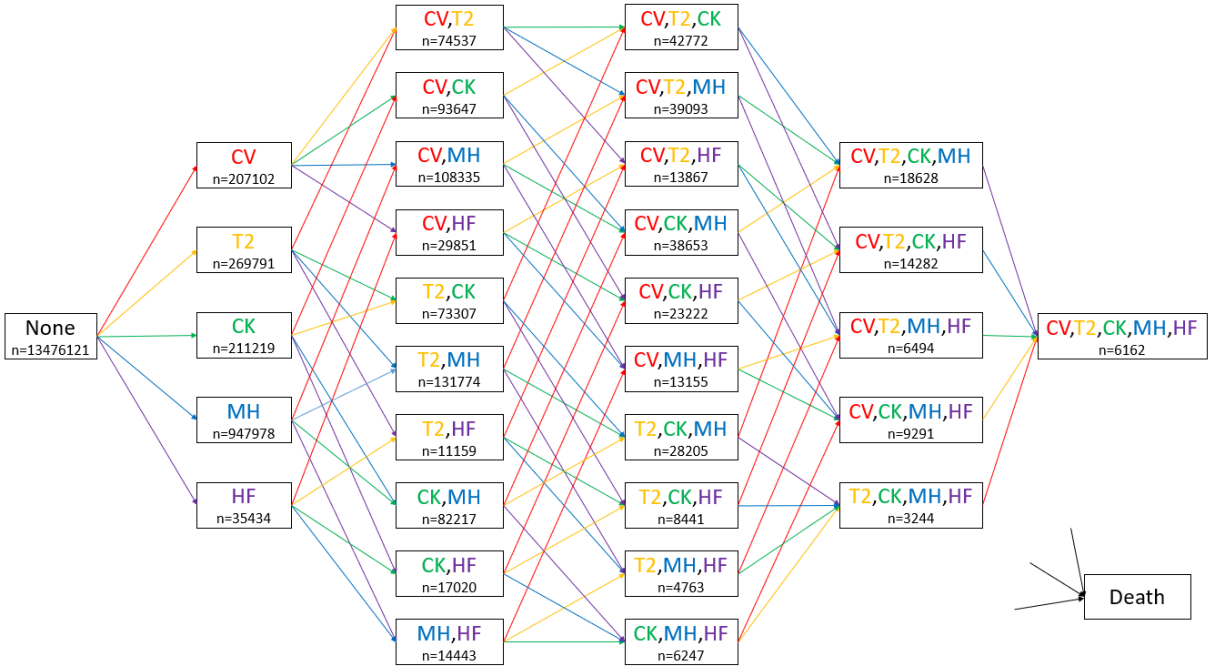


Figure 1: The multistate diagram. The comorbidity states are represented in black boxes where we use red, yellow, green, blue and purple colours to distinguish between cardiovascular disease (CV), type-2 diabetes (T2), chronic kidney disease (CK), mental health (MH) and heart failure (HF), respectively. The transitions are represented using coloured arrows, with colours indicating the disease that is diagnosed when the transition takes place.

Results

Summary statistics

The cohort baseline characteristics are shown in Table 1. At study entry, 78.5% (n=10,577,667) of the eligible participants had no diagnosis of any of the five conditions of interest. By the end of the follow-up, 83.7% (n=8,856,098) of these had not developed any of the five conditions, 13.8% (n=1,462,966) had developed one, 2% (n=209,237) had developed two, 0.4% (n=42504) had developed three, 0.06% (n=6399) four and 0.0044% (n=463) five of the conditions. Five percent of the study population (n=670,753) died during the study period. Detailed summary statistics for each comorbidity state in our MSM are presented in Supplementary Table A. A summary of the incidence rate for each disease transition in our MSM is provided in Supplementary Table B. The incidence rates of transitions to the first morbidity were notably lower compared to the other transitions, and generally, the incidence rates for transitions increased with the number of comorbidities. It should be noted that the relatively lower incidence rates for transitions from MH compared to other disease transitions could be due to the fact that subjects in the MH state were generally younger (this age effect is adjusted in our MSM analysis).

Table 1. Baseline sociodemographic characteristics of study participants

Cohort population		13,476,121
Sex (n (%))	Male	6,315,448 (46.9%)
	Female	7,160,673 (53.1%)
Age at cohort entry (median [IQR])		35 [25, 51]
Ethnicity (n (%))	White	10,753,663 (79.8%)
	South Asian	1,460,750 (10.8%)
	Black	741,504 (5.5%)
	Mixed	257,651 (1.9%)
	Others	262,553 (1.9%)
Index of multiple deprivation (n (%))	1 - least deprived	2,384,659 (17.7%)
	2	2,530,868 (18.8%)
	3	2,620,501 (19.4%)
	4	3,139,349 (23.3%)
	5 - most deprived	2,800,744 (20.8%)
Follow-up period (median (IQR))		4.71 [1.78, 11.28]

Effect of ethnicity on the rate of disease progression

There was no consistent association pattern between ethnicity and disease accumulation as this was also modified by comorbidity status (Figure 2). For CVD diagnosis among different comorbidity groups (panel (a)), the hazard ratios tended to be higher for South Asian but lower for black compared to white ethnicity. For T2D (panel (b)), hazard ratios for South Asian, black and other ethnicities compared to White were higher than 1 for most comorbidity groups, but the differences between each of the non-white and white groups decreased with increasing numbers of comorbidities. Generally, differences were less pronounced

between the four ethnic groups for CKD (panel (c)), HF (panel (d)) and MH (panel (e)) diagnosis among groups with different comorbidities. For a few transitions, people from the three non-white groups had a slightly reduced rate of disease progression compared to the white ethnicity. For MH, the hazard ratios for black compared to white were mostly lower than 1.

Effect of deprivation on the rate of disease progression

For the diagnosis of CVD, T2D, CKD, HF or MH, hazard ratios were greater for the more deprived groups (IMD>1) compared with the least deprived group (IMD=1), with a higher level of deprivation associated with a higher rate of disease transition (Figure 3). The strength of association with deprivation was stronger for T2D (HR between 1 to 2) and HF (HR between 1 to 1.7), while it was weaker for CKD, MH, and CVD. The effect of greater deprivation on all five conditions tended to decrease with increasing numbers of comorbidities.

Effect of gender on the rate of disease progression

Adjusted hazard ratios for men (versus women) were statistically significantly higher for CVD, T2D and HF diagnosis but lower for CKD and MH diagnosis across different comorbidity groups (Figure 4). The difference between men and women for CVD and T2D decreased with increasing numbers of comorbidities (HR for men compared to women was still statistically significantly higher than 1 for most transitions). In contrast, for CKD and HF diagnosis (except for transitions from baseline or MH), there was little difference between men and women regardless of the number of comorbidities. The lower hazards of MH diagnosis for men were slightly reduced with increasing number of comorbidities.

Effect of age on the rate of disease progression

A 10-year increase in age was associated with an elevated rate of CVD (HR between 1 to 2), CKD (HR between 1.3 to 2.5) or HF (HR between 1.2 to 2.4) diagnosis (Figure 5). Adjusted hazard ratios for a 10-year increase in age were greatest for those with a single condition, particularly MH, and approached the null for those with increasing numbers of comorbidities. By contrast, adjusted hazard ratios for age were generally statistically significantly lower than 1 for T2D or MH diagnosis, with little difference across the number of comorbidities.

Comparing the effects of the sociodemographic characteristics

For CVD, CKD, MH and HF diagnosis, a 10-year increase in age, higher deprivation levels, and the sex of the individual (male for CVD and HF, female for CKD and MH) tended to be more strongly associated with an increased rate of diagnosis of a condition compared to ethnic group differences (White as the reference group). Among these four characteristics, the age effect was strongest for CKD and HF. For T2D, belonging to the South Asian ethnic group was associated with the greatest increase in the rate of the diagnosis.

Mental health

The impact of the sociodemographic characteristics on the rate of diagnosis of a condition was generally more significant (in the direction of increasing the transition intensity) for individuals with current diagnoses of MH compared to those with other comorbidities. For age and deprivation (particularly age), this pattern is observed for transitions to all four conditions except MH. Even the slightly protective effect of higher age on transitions to T2D is reversed in the presence of a comorbidity of MH. The adverse effect of being male on transitions to CVD, T2D, or HF was amplified with a MH comorbidity. For ethnicity, the pattern is only observed for transitions to T2D, where the hazard ratio for South Asian ethnicity compared to White was higher for those with a MH diagnosis.

Sensitivity analysis

We conducted a sensitivity analysis to examine the impact of data exclusion on the inferences (here we focused on missing ethnicity as it accounts for the vast majority of missing data). More specifically, we expanded our cohort by including subjects with missing ethnicity, who were assigned to an artificial ethnic group named "missing". We then re-fitted the MSM in the same manner as before. The results, including estimated HRs and associated confidence intervals for each transition, are presented in supplemental Figure C. Upon comparison with Figures 2-5 in the main paper, we observed that there were no significant differences and the key patterns remained unchanged.

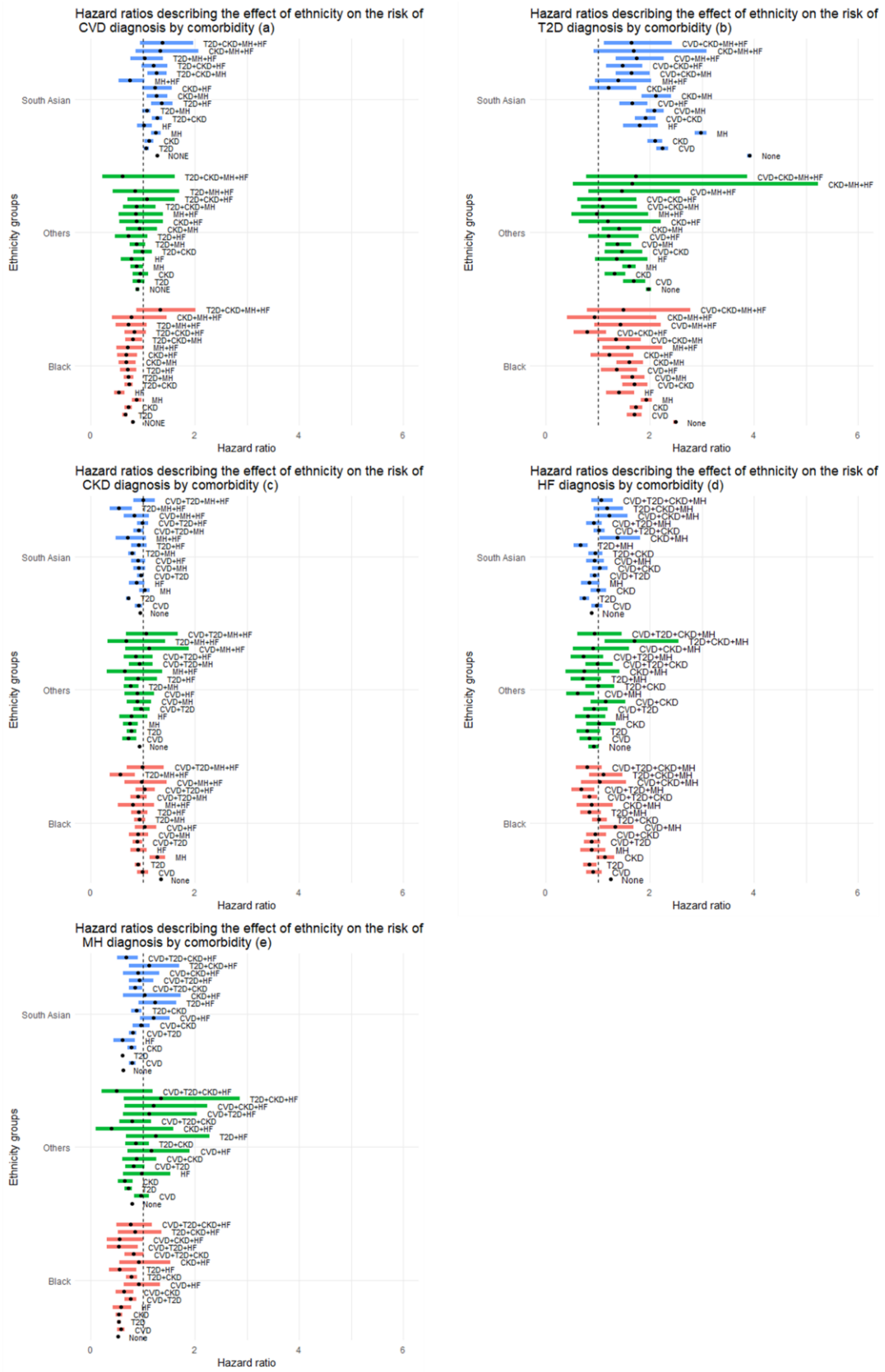


Figure 2. Effect of ethnicity on the rate of disease transition by comorbidity. Estimated hazard ratios are shown in black dots and the associated 95% confidence intervals are represented with coloured bands. White ethnicity is treated as the reference category.

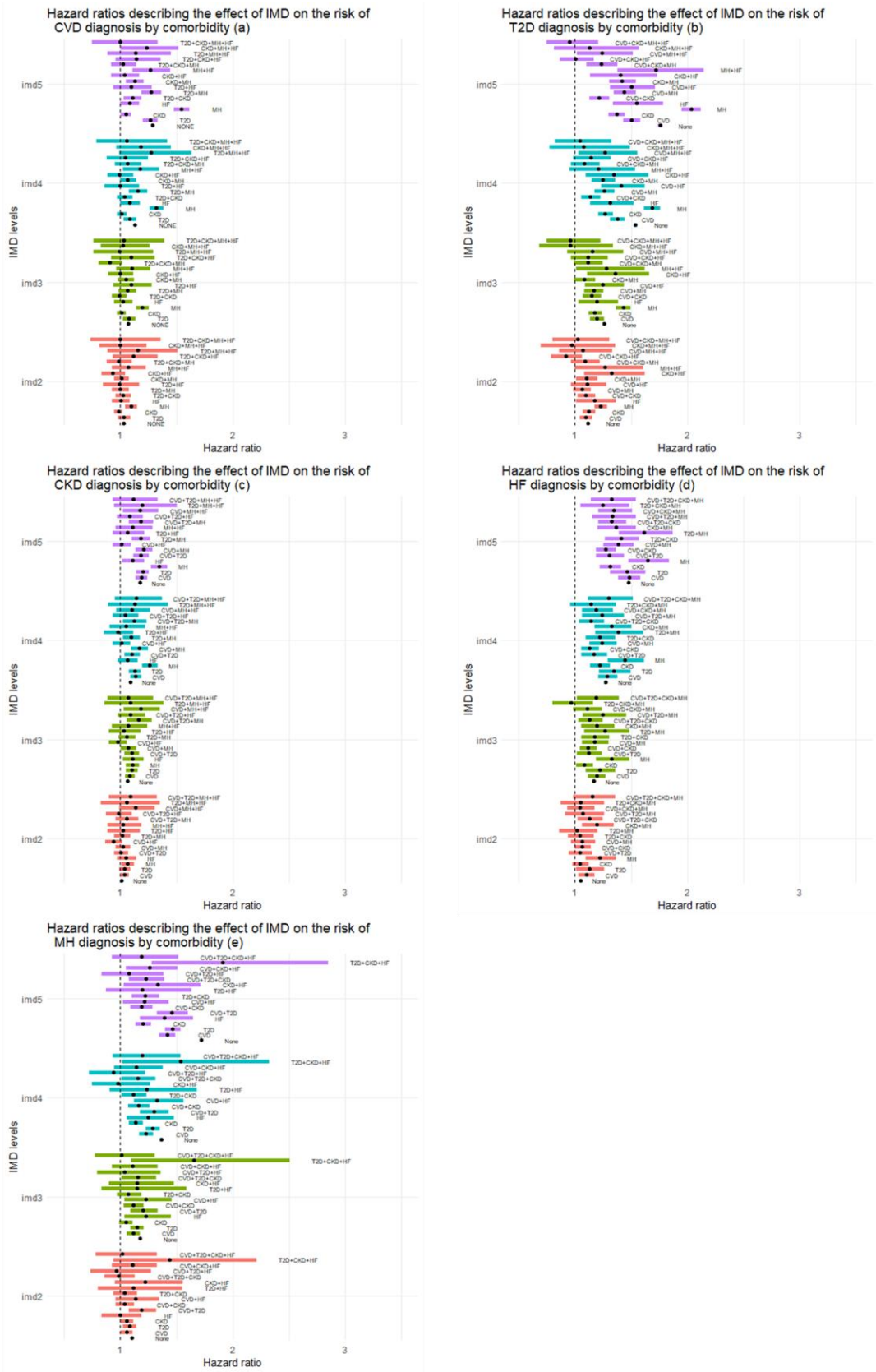


Figure 3. Effect of deprivation on the rate of disease transition by comorbidity. Estimated hazard ratios are shown in black dots and the associated 95% confidence intervals are represented with coloured bands. Quintile 1 is treated as the reference category and represents the least deprived.

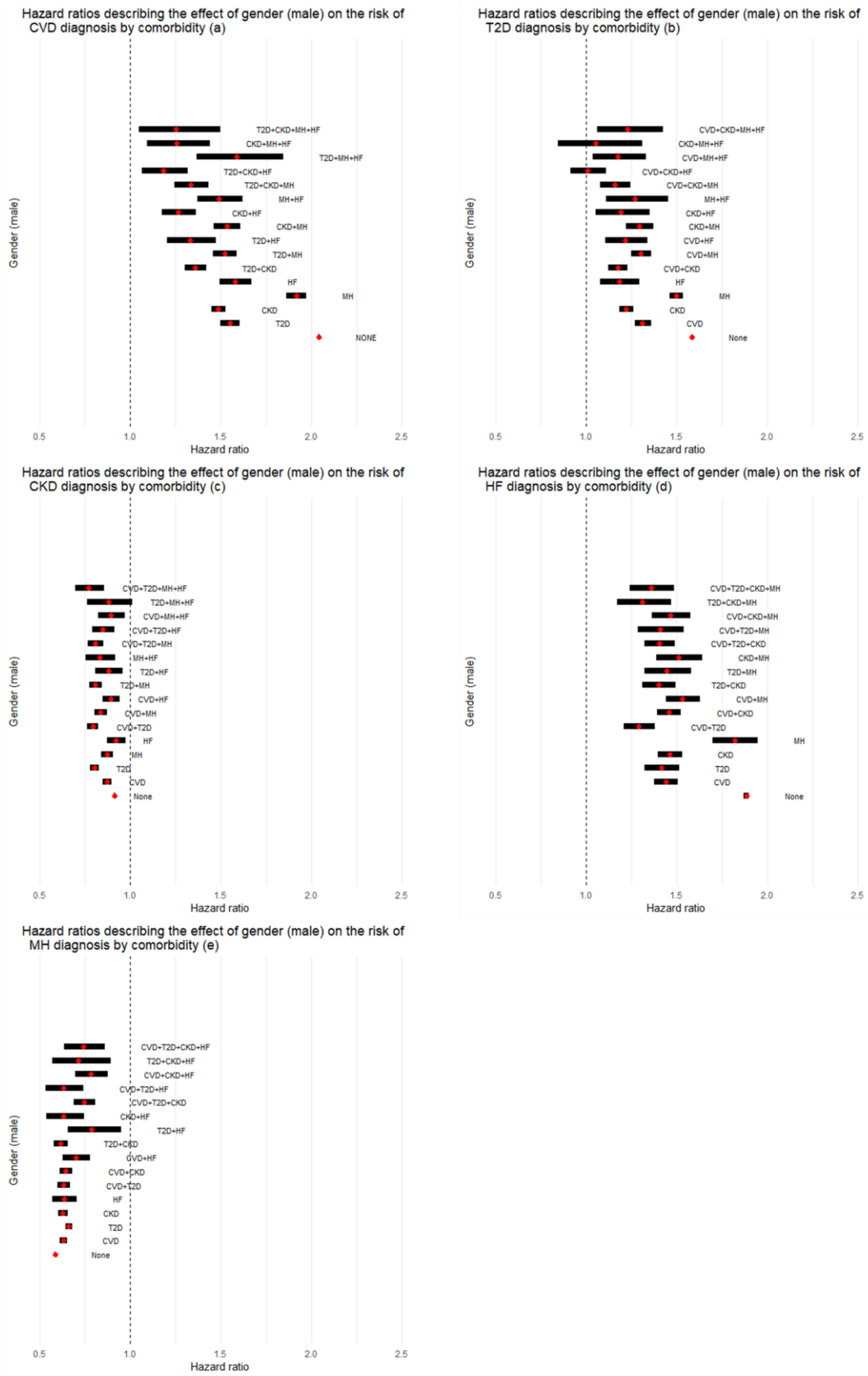


Figure 4. Effect of gender (male) on the rate of disease transition by comorbidity. Estimated hazard ratios are shown in red dots and the associated 95% confidence intervals are represented with black bands. The female gender is treated as the reference category.

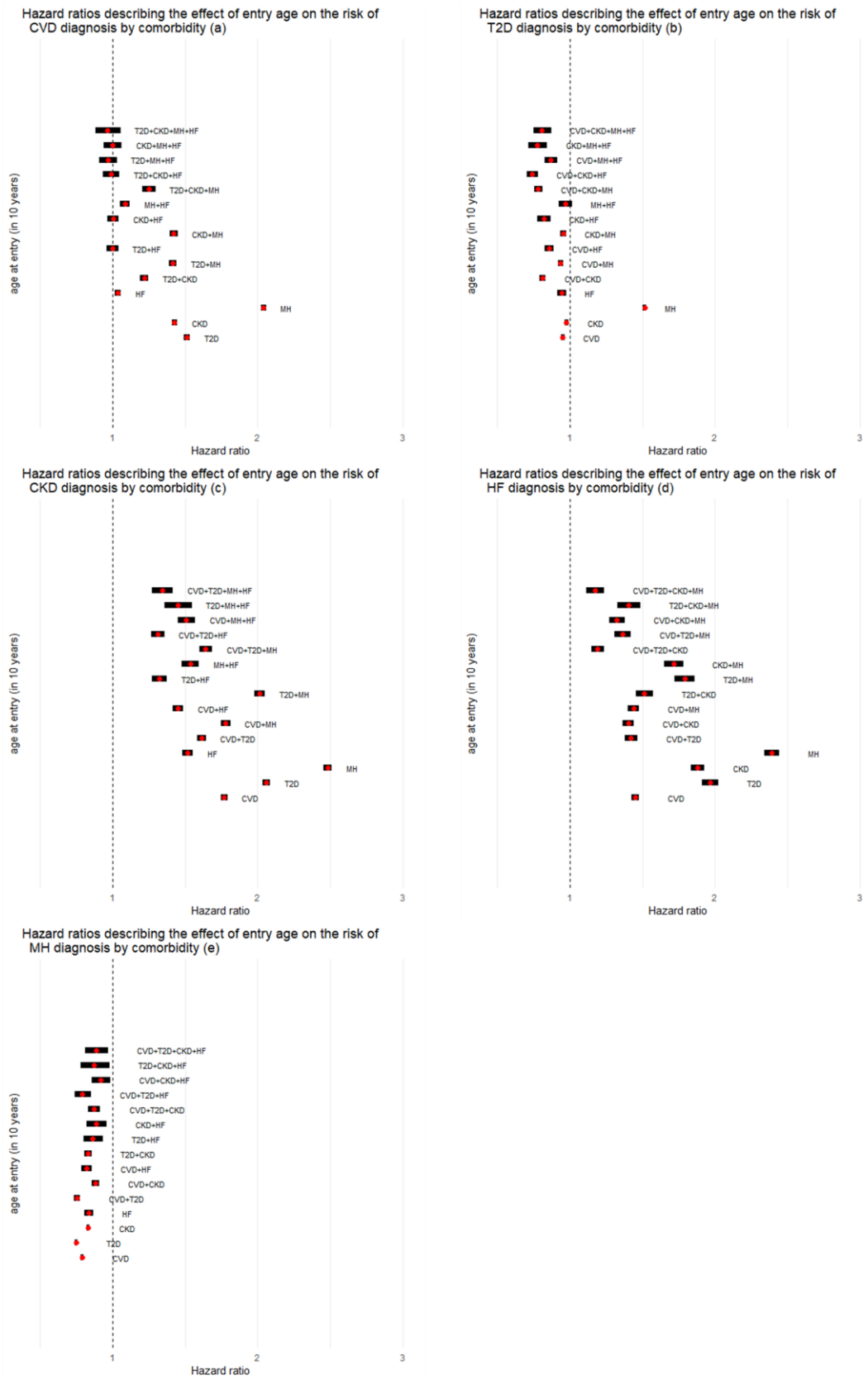


Figure 5. Effect of age at entry to the current state (in 10 years) on the rate of next disease transition by comorbidity. Estimated hazard ratios are shown in red dots and the associated 95% confidence intervals are represented with black bands.

Discussion

Main findings

We analysed the progression of multimorbidity, defined as the combination of five chronic conditions, including CVD, T2D, CKD, MH and HF, in a cohort of more than 13 million English adults. To our knowledge, this is the largest cohort and the highest number of chronic conditions and disease transitions analysed using a flexible multistate modelling framework. The proposed MSM-based approach provides a detailed picture of the effects of sociodemographic characteristics, including ethnicity, deprivation, gender and age, on the rate of disease transitions along different pathways to multimorbidity accumulation.

Our study revealed complex interactions between trajectories of multimorbidity progression and ethnicity, deprivation, gender and age, with effects that varied depending on mental and physical health conditions. Generally, the effect of these characteristics was greater for those with fewer morbidities than those with many comorbidities. This suggests sociodemographic characteristics have a larger impact on initial multimorbidity development while a smaller impact on those with many long-term conditions. Notably, MH, as a co-morbidity, was typically an "outlier" as often exacerbated the effects of sociodemographic characteristics on the transition to the next disease. We also found that the relative magnitude of the effects of these sociodemographic characteristics on the rate of diagnosis of a new condition also depended greatly on the comorbidity status of the subject.

Ethnicity

Compared to the White UK population, a higher risk of diabetes in those of South Asian and Black ethnicity has been observed in other epidemiological research [24]. This research also found higher risk and lower risks of CVD in those of South Asian and Black ethnicity respectively. CKD is underdiagnosed in UK primary care, therefore diagnosis is likely to be influenced by the frequency of diagnostic testing [25]. Rates of testing are strongly associated with the presence of chronic conditions but less strongly with ethnicity [26]. In the past, the estimation of CKD prevalence in Black patients and women may have been overestimated due to adjustments for Black ethnicity and female sex in the calculation of the estimated glomerular filtration rate (eGFR) [27]. Epidemiological research has suggested that depression (the most common mental health condition) is more common in older UK adults of Black and South Asian ethnicity than those of White ethnicity [28]. Our finding of slightly lower rates of mental health problems for those in the Black and South Asian ethnic groups may suggest underascertainment of mental health problems, particularly for these groups.

Deprivation

Our results show that higher levels of deprivation tended to be associated with an elevated risk of disease progression for all conditions and types of transitions. This is in keeping with a previous study, which showed that multimorbidity occurred 10–15 years earlier and physical-mental comorbidity was more than twice as prevalent in the most socioeconomically deprived areas of Scotland compared to the least deprived areas [17]. We found that the effect of deprivation was less strong as the number of comorbidities increased, suggesting that deprivation may have a stronger influence on the initial trajectory of multimorbidity.

Sex

As seen in other studies, men tend to have an increased risk of CVD, T2D, and HF, but a lower risk of MH problems and CKD [29]. Our results echo previous findings and additionally demonstrate that the effect of sex on disease progression is dependent on the underlying comorbidity status. For instance, the association between the male gender and the rate of transition to CVD or T2D was most significant for those with the least comorbidities. For MH, the greater the number of comorbidities, the weaker the gender difference in the rate of MH accumulation. For HF and CKD, the gender effect is less influenced by the level of multimorbidity. Note that our results may be affected by gender disparities in the presentation to healthcare, where women are more likely to present for screening and investigations [30]. On the other hand, those with known health conditions are more likely to be invited for (and maybe more inclined to present for) further monitoring and screening for other health conditions. Thus this gender difference regarding clinical presentation is expected to be smaller with increasing comorbid conditions.

Age

As expected, age was associated with increased risk of CVD, CKD and HF. In addition, we found that age amplified the effect of MH on the likelihood of transition to each of these conditions. Increasing age had a much greater impact on the likelihood of transition to a new condition on those with a single condition than those with multiple conditions. Multimorbidity is increasingly prevalent with increasing age [17].

Mental health

In comparison to physical health conditions such as CVD, T2D, HF, and CKD, MH was a notable outlier in its effect on the rates of disease transition. Our results suggest that the adverse effect of sociodemographic factors on the rate of transitioning to the next condition (CVD, T2D, CKD or HF) tended to be enhanced for trajectories starting with MH compared to those starting with other conditions. The pattern may be partly explained from biological and behavioural points of view. Those with MH as a starting condition tend to fall into the younger age groups (see Supplemental Table A) and mental ill health is associated with unhealthy lifestyle habits such as smoking, alcohol or substance abuse etc [31], which are also known to be important risk factors for cardiometabolic diseases and contribute to the worsening of multimorbidity [32,11]. On the other hand, this adverse impact of MH is attenuated with additional comorbidities, which could be influenced by the additional medical intervention/treatment received by the subject and relevant lifestyle modifications.

Strengths and limitations

This study benefits from a large dataset from primary health care which is broadly representative of the English population. The conditions considered in this study are included in an incentive scheme for the management of chronic conditions, which means they are likely to be more accurately coded [33]. The use of a highly flexible multistate model allows for a detailed characterization of the effect of key sociodemographic characteristics on 112 different transitions with a very good overall fit for the vast majority of the transitions, demonstrating the utility and computational feasibility of the approach.

There are some limitations with the current study. One primary limitation is the possibility of underdiagnosis or ascertainment bias of the onset of chronic conditions. For instance, diseases like CKD and

T2D are often asymptomatic in the early stages, and diagnosis may occur during routine health check/screening or following the diagnosis of a related condition (once diagnosed, general practices actively follow up patients for further disease monitoring). The uptake of such checkups is generally less than 50%, with lower figures for men and some ethnic groups [34]. For CKD, renal function testing is lower in minority ethnic groups [26,35]. For MH, certain groups (e.g. those in the Black or South Asian ethnic groups) are less likely to present to the GP even with symptoms developed [36]. The implication is that the identified date of diagnosis or the transition time between diagnoses of the conditions in the electronic health record data may not reflect the corresponding true onset or transition times. Additionally, the ethnicity categories used in this study crudely group together communities with diverse cultural practices and sociodemographic characteristics, which may mask differences in outcomes within each category. It is also important to note that time-varying factors associated with the disease progression such as treatment information, lifestyle modifications, and change in deprivation status are not available from the present dataset and thus they were not taken into account in the present analysis. Finally, the median follow-up time for the study participants was about 5 years, with a maximum period of about 15 years, which limits our ability to observe complete trajectories for most subjects and gives rise to a very high right censoring rate for most transitions. A longer follow-up period would be desirable for obtaining better inference for transitions in later stages of multimorbidity.

Unanswered questions and future research

Clinical research implications of this analysis are to investigate and better understand the biopsychosocial factors driving progression to multimorbidity. The greatest differences in transition rates to multimorbidity between different groups are in the development of the first few conditions, which suggests the early stages in the pathway to multimorbidity may be particularly important.

The MSM framework used here could permit broader types of inference, such as the probability of transition between two comorbidity states over a given time interval (conditional on sociodemographic characteristics), which can be inferred by simulating trajectories conditional on the fitted model. We could also analyse the effect of disease transition history on subsequent disease progression by representing previous transitions as additional covariates in the model. There is also scope to study the association between concurrently measured longitudinal covariates (e.g. biomarkers) and multimorbidity progression using a joint modelling framework for both the longitudinal marker and the multistate processes [37]. Further methodological research is required to overcome the computational challenge of fitting this model.

Conclusion

Multimorbidity is a growing public health challenge and there is a need of moving from a single disease-oriented to a more individualised and holistic approach. In this study, we investigated the influence of key sociodemographic characteristics on the development and progression of multimorbidity, revealing distinctive association patterns for different levels of multimorbidity and disease transitions. Our results indicate that the effects of these factors tend to be weakened with increasing levels of multimorbidity. The difference of mental health as a comorbidity is also highlighted as it often strengthens the adverse effects of sociodemographic characteristics on disease accumulation. These findings underscore the need for earlier screening and intervention of mental health conditions, especially for those in the higher-risk populations, which may have important implications for the management of multimorbidity.

Author Declarations

Competing interest statement

JB has previously received funding for unrelated work from F. Hoffmann-La Roche Ltd. KN reports grants from NIHR, MRC, Diabetes UK, Vifor, and AstraZeneca, and personal fees from Merck Sharp & Dohme, Sanofi, and Boehringer Ingelheim, outside of and unrelated to the submitted work. TM has previously received funding for unrelated work: from Cancer Research UK and from the Health Research Board (Ireland) for advisory board participation and travel. SJG received payment from Novo Nordisk, Napp and Astra Zeneca for lectures at 6 educational events over the last 3 years. Napp supported SJG's attendance at EASD 2018 and he received payment from Eli Lilly associated with membership of an independent data monitoring committee for a randomised trial of a medication to lower glucose. PK, SR and KN are co-investigators on the HDRUK grant titled Measuring & Understanding Multimorbidity using Routine Data in the UK (MUrMuR-UK). CY has received remuneration for unrelated consultancy services to F. Hoffmann-La Roche and Singula Bio. The remaining authors have no conflicts of interest.

Data availability statement

The data were provided under license by the Clinical Practice Research Datalink (CPRD), and so are not publicly available. Access to CPRD data is subject to protocol approval via CPRD's Research Data Governance (RDG) Process.

Funding statements

This work is funded by the MRC/NIHR grant MR/S027602/1 (BIRM-CAM). TM is supported by the National Institute for Health Research Collaboration Applied Research Collaboration West Midlands (NIHR ARC WM). KN is funded by a Health Data Research UK Fellowship. CY is funded by a UKRI Turing AI Fellowship (EP/V023233/2). JB was funded by MRC unit programme MC_UU_00002/5. CJ was funded by the Medical Research Council programme number MRC_MC_UU_00002/11. PK was funded by MRC unit programme MC_UU_00002/13. SR was funded by MRC unit programme MC_UU_00002/10. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Disclaimers

The views expressed in this publication are those of the authors and not necessarily those of the NIHR, NHS, the UK Department of Health and Social Care or other funding institutions.

Ethics approval statement

The Clinical Practice Research Datalink (CPRD) Research Database has ethics approval from the Health Research Authority to support research using anonymised patient data (East Midlands - Derby Research Ethics Committee, reference 21/EM/0265). Data was approved for this study by the CPRD Independent Scientific Advisory Committee (protocol 19_265).

Acknowledgements

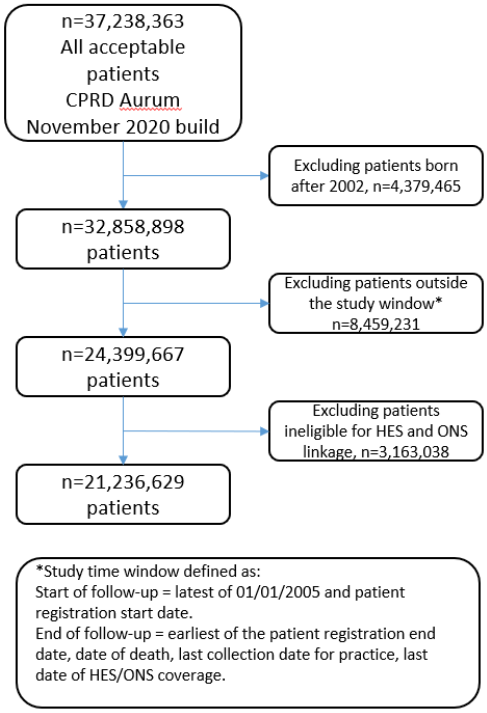
This work used data from the Clinical Practice Research Datalink (CPRD) for ISAC protocol 19_265. The data were provided by patients and collected by the NHS as part of their care and support.

For the purpose of open access, the author has applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising from this submission.

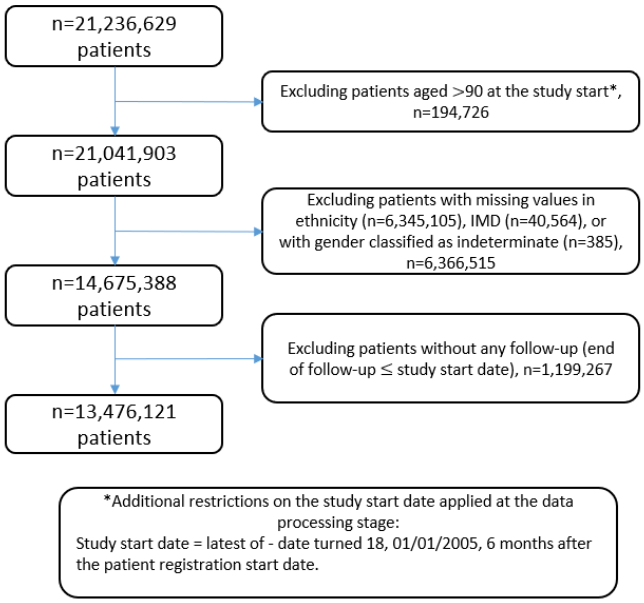
Supplementary file

Supplemental Figure A. Flowchart for sample selection

Extraction stage from CPRD Aurum



Additional criteria applied at the data processing stage



Supplemental Technical Details

The multistate modelling approach

In this section we provide a brief introduction to multistate modelling (MSM), and we refer to [38] and references therein for more in-depth coverage of the topic. A multistate model is built from a mathematical object named a multistate process. A multistate process is a stochastic process $\{X(t),\ t \geq 0\}$ with a finite state space $S = \{1, \dots, N\}$ (here we focus on the setting where t is continuous and $X(t)$ is continuously observed). The law of a multistate process can be fully characterised via the transition intensities, which directly describe the instantaneous dynamic of the process

$$\alpha_{hj}(t; F_{t^-}) = \lim_{\Delta t \to 0} \frac{P(X(t + \Delta t) = j | X(t) = h;\ F_{t^-})}{\Delta t}\ ,\ \ h \neq j \in S,$$

where F_{t^-} represents the history of the process over $[0, t)$. There are two special classes of the multistate processes that have been thoroughly studied. The first is the Markov processes, which relies on the Markov assumption such that risks at a particular time are independent of the history, so that

$$\alpha_{hj}(t; F_{t^-}) = \alpha_{hj}(t).$$

The second class is the so-called clock-reset semi-Markov processes, which relaxes the Markov property by assuming that

$$\alpha_{hj}(t; F_{t^-}) = \alpha_{hj}\big(B(t)\big),$$

where $B(t)$ is the time since entry to the current state h. Therefore, α_{hj} depends on the history only through $B(t)$, and in effect, the time scale after each transition is reset to zero. In the former scenario, the transition probabilities between two states can be derived from the intensities by solving the Kolmogorov differential equations, whereas for the latter case, no explicit formulation exists. In such scenarios, one could estimate transition probabilities from intensities via simulation-based techniques.

A multistate model can be fitted to data representing the evolution of n subjects (defined on the same state space) in a population, each of whom is followed up continuously for a potentially different length of time (may be subject to delayed entry or right censoring). To model the covariate-dependent transition intensities for subject i, $\alpha_{hj}^{(i)}(t; F_{t^-})$, Markov or semi-Markov assumptions are commonly adopted and semi-parametric or parametric proportional hazard regression framework is often used. If we in addition assume that each individual's trajectory, $X_i(t)$, is independent, then it can be shown using counting process theory that the MSM likelihood factorizes as

$$L = \prod_{i=1}^{K} L_i\big(\theta | D^{(i)}\big),$$

where K is the total number of allowed transition types in the multistate process (e.g. if $\alpha_{hj} \neq 0$ then $h \to j$ represents one type of transition), L_i is the likelihood obtained by considering transition type i as a survival problem with data based on all subjects at risk for this transition and θ is the parameter vector consisting of parameters from all transitions. An important consequence of this factorization property is that if no parameters are shared across transitions, i.e. $\theta = (\theta_1, \dots, \theta_K)$, then we could perform maximum likelihood estimation of the MSM by maximising each L_i independently with respect to its parameter set θ_i (see e.g. [39]).

In our analysis we specify the MSM by using a separate time-to-event model for each permitted $h \rightarrow j$ transition. To ensure flexibility, we use the spline-based Royston-Parmar model [22] for the transition-specific intensities, $\alpha_{hj}^{(i)}(u; Z(t))$,

$$log\ (A_{hj}^{(i)}(u; Z)) = s_{hj}(log\ u\ ; \gamma_{hj}) + \beta_{hj}^{T} Z_i\ ,$$

where u refers to the time since entry to the state h, $A_{hj}^{(i)}$ is the cumulative transition intensity associated with $\alpha_{hj}^{(i)}$, $s_{hj}\big(log\ u\ ; \gamma_{hj}\big)$ is a natural cubic spline in $log\ u$ with coefficient vector γ_{hj} (see [22] for more details on the construction of the spline function), Z_i denotes the vector of the exogenous covariates for subject i and β_{hj} is the vector of the associated linear effects. Note that the resulting time-to-event model retains a proportional hazard structure so that elements of β_{hj} are still interpreted as log-hazard ratios. In our context, the covariates are IMD, ethnicity, gender and age (at the time of the previous transition; in 10 years). The first three variables are categorical and are represented via dummy variables with reference groups set to female, IMD=1 (the least deprived group) and White (the dominant ethnic group), respectively. Age is approximated by a piecewise constant process with jumps occurring at the transition times (hence is treated as a time constant covariate in a transition model). A practical issue here is to determine the number and location of the internal knots for the spline term s_{hj}, which would influence the flexibility of the resulting model. We follow the suggestions in [22] to use the Akaike information criterion (AIC) to determine the number of internal knots in the spline from the candidate values of $\{0, \dots, 5\}$. Note that when there are zero knots the model reduces to a Weibull model. Conditional on the knot number, the location of these knots is set to equally spaced percentiles of the logarithm of the uncensored transition times for $h \rightarrow j$ (empirical study indicates that the knot position is not a key driver of model fit).

Model checking

We used the Cox-Snell residuals to graphically examine the goodness of fit of each of our transition models [40,41]. For a fitted time-to-event model, the Cox-Snell residuals are defined as $r_i^{CS} = \hat{A}_i(T_i)$, where $\hat{A}_i$ is the estimated cumulative hazard function for subject i under the fitted model and T_i is the observed event time. If the model fits the data well, we expect the r_i^{CS} to be close to draws from a unit exponential distribution. This diagnostic is motivated by the fact that for any random variable T with survival function $S(t) = P(T > t)$, the transformed random variable $-\ log\ S(T)$ is distributed as an exponential random variable with unit mean. To appropriately take censored event times into account, in practice we estimate the cumulative hazard of the residuals, i.e. $-\ log\ \hat{S}\ \big(r_i^{CS}\big)$, where $\hat{S}$ is the Kaplan-Meier estimate of the survival function associated with r_i^{CS} (for censored event times T_i the corresponding r_i^{CS} are treated as censored as well). A good fit is suggested if a plot of $-\ log\ \ \hat{S}\ \big(r_i^{CS}\big)$ against r_i^{CS} is close to a straight line with an intercept of zero and a slope of one. For our fitted MSM, we computed the Cox-Snell residuals for each transition, and we observed satisfactory results for most of the transitions, see Figure B for selective results. Here we observed a slight lack of fit for transitions MH->MH+T2 and MH->MH+CK. We found that residuals that significantly deviated from the reference line (in red) correspond to subjects who had a long transition time, but their sociodemographic characteristics suggest that the transition should have happened earlier (e.g. diagnosed MH at a high age and with high deprivation levels). The presence of such "outliers" may not be surprising given the complexity of the disease progression mechanisms and the inter-subject heterogeneity in such a large cohort. On the other hand, it should be noted that these outliers only account for an extremely small proportion of the data involved in these 112 transitions. Out of the 112 transitions these two were among the most ill-fitting.

Supplemental Table A. Summary of disease states by sociodemographic characteristics.

		age*	IMD					Gender		Ethnicity				
Disease state	Counts	Median (IQR)	1st Quintile (Least deprived)	2nd Quintile	3rd Quintile	4th Quintile	5th Quintile	Male	Female	White	Asian	Black	Mixed	Others
CV	207102	68.4 [58.7, 77.2]	45889 (22.2%)	44310 (21.4%)	41403 (20%)	38687 (18.7%)	36813 (17.8%)	128239 (61.9%)	78863 (38.1%)	189530 (91.5%)	10563 (5.1%)	4781 (2.3%)	1194 (0.6%)	1034 (0.5%)
T2	269791	58.4 [48.7, 68.3]	44150 (16.4%)	47236 (17.5%)	51680 (19.2%)	62833 (23.3%)	63892 (23.7%)	158576 (58.8%)	111215 (41.2%)	196363 (72.8%)	44772 (16.6%)	21457 (8%)	4157 (1.5%)	3042 (1.1%)
CK	211219	74.3 [65.9, 81.3]	47907 (22.7%)	45631 (21.6%)	43041 (20.4%)	39080 (18.5%)	35560 (16.8%)	85293 (40.4%)	125926 (59.6%)	195653 (92.6%)	6810 (3.2%)	6808 (3.2%)	1333 (0.6%)	615 (0.3%)
MH	947978	37.3 [27.3, 49.8]	166803 (17.6%)	177562 (18.7%)	179870 (19%)	205494 (21.7%)	218249 (23%)	347552 (36.7%)	600426 (63.3%)	835101 (88.1%)	56324 (5.9%)	31355 (3.3%)	14755 (1.6%)	10443 (1.1%)
HF	35434	73.7 [62.9, 82.1]	7283 (20.6%)	7195 (20.3%)	7173 (20.2%)	6960 (19.6%)	6823 (19.3%)	20587 (58.1%)	14847 (41.9%)	32670 (92.2%)	1175 (3.3%)	1221 (3.4%)	221 (0.6%)	147 (0.4%)
CV,T2	74537	69.7 [61.3, 77.3]	12870 (17.3%)	13932 (18.7%)	14573 (19.6%)	16050 (21.5%)	17112 (23%)	50521 (67.8%)	24016 (32.2%)	61380 (82.3%)	8869 (11.9%)	3029 (4.1%)	753 (1%)	506 (0.7%)
CV,CK	93647	79.1 [72.7, 84.6]	20775 (22.2%)	20080 (21.4%)	19178 (20.5%)	17327 (18.5%)	16287 (17.4%)	48526 (51.8%)	45121 (48.2%)	89170 (95.2%)	2543 (2.7%)	1381 (1.5%)	359 (0.4%)	194 (0.2%)

CV,MH	108335	65.8 [56.1, 75.6]	19829 (18.3%)	20968 (19.4%)	21023 (19.4%)	21996 (20.3%)	24519 (22.6%)	51951 (48%)	56384 (52%)	101998 (94.2%)	3701 (3.4%)	1558 (1.4%)	584 (0.5%)	494 (0.5%)
CV,HF	29851	76 [67.2, 83]	5940 (19.9%)	6094 (20.4%)	5962 (20%)	5807 (19.5%)	6048 (20.3%)	20458 (68.5%)	9393 (31.5%)	27995 (93.8%)	1107 (3.7%)	521 (1.7%)	116 (0.4%)	112 (0.4%)
T2,CK	73307	74.5 [67.3, 80.6]	13278 (18.1%)	14230 (19.4%)	14731 (20.1%)	15539 (21.2%)	15529 (21.2%)	34860 (47.6%)	38447 (52.4%)	61431 (83.8%)	6039 (8.2%)	4664 (6.4%)	839 (1.1%)	334 (0.5%)
T2,MH	131774	56.9 [48.2, 66.2]	18727 (14.2%)	21816 (16.6%)	24835 (18.8%)	30130 (22.9%)	36266 (27.5%)	58203 (44.2%)	73571 (55.8%)	109628 (83.2%)	13163 (10%)	6137 (4.7%)	1622 (1.2%)	1224 (0.9%)
T2,HF	11159	73 [63.9, 80.6]	1781 (16%)	2002 (17.9%)	2228 (20%)	2519 (22.6%)	2629 (23.6%)	6943 (62.2%)	4216 (37.8%)	9331 (83.6%)	946 (8.5%)	698 (6.3%)	133 (1.2%)	51 (0.5%)
CK,MH	82217	71.8 [62.5, 80]	17186 (20.9%)	17042 (20.7%)	16515 (20.1%)	15764 (19.2%)	15710 (19.1%)	20981 (25.5%)	61236 (74.5%)	78644 (95.7%)	1616 (2%)	1391 (1.7%)	390 (0.5%)	176 (0.2%)
CK,HF	17020	81.8 [74.9, 87]	3581 (21%)	3600 (21.2%)	3540 (20.8%)	3282 (19.3%)	3017 (17.7%)	7944 (46.7%)	9076 (53.3%)	16066 (94.4%)	409 (2.4%)	423 (2.5%)	80 (0.5%)	42 (0.2%)
MH,HF	14443	70.1 [58.9, 79.8]	2530 (17.5%)	2752 (19.1%)	2849 (19.7%)	2995 (20.7%)	3317 (23%)	6339 (43.9%)	8104 (56.1%)	13762 (95.3%)	288 (2%)	274 (1.9%)	62 (0.4%)	57 (0.4%)
CV,T2,CK	42772	77.4 [71.4, 82.9]	7618 (17.8%)	8224 (19.2%)	8537 (20%)	8897 (20.8%)	9496 (22.2%)	24674 (57.7%)	18098 (42.3%)	36887 (86.2%)	3713 (8.7%)	1603 (3.7%)	385 (0.9%)	184 (0.4%)

CV,T2,MH	39093	66.9 [58.6, 75.1]	5544 (14.2%)	6609 (16.9%)	7195 (18.4%)	8828 (22.6%)	10917 (27.9%)	21282 (54.4%)	17811 (45.6%)	34012 (87%)	3393 (8.7%)	1102 (2.8%)	326 (0.8%)	260 (0.7%)
CV,T2,HF	13867	74.3 [66.8, 80.8]	2135 (15.4%)	2452 (17.7%)	2700 (19.5%)	3099 (22.3%)	3481 (25.1%)	10050 (72.5%)	3817 (27.5%)	11865 (85.6%)	1339 (9.7%)	489 (3.5%)	106 (0.8%)	68 (0.5%)
CV,CK,MH	38653	78.1 [70.8, 84.3]	7634 (19.8%)	7799 (20.2%)	7689 (19.9%)	7577 (19.6%)	7954 (20.6%)	13949 (36.1%)	24704 (63.9%)	37386 (96.7%)	721 (1.9%)	350 (0.9%)	128 (0.3%)	68 (0.2%)
CV,CK,HF	23222	81.5 [75.6, 86.6]	4713 (20.3%)	4788 (20.6%)	4765 (20.5%)	4438 (19.1%)	4518 (19.5%)	13777 (59.3%)	9445 (40.7%)	22115 (95.2%)	631 (2.7%)	335 (1.4%)	90 (0.4%)	51 (0.2%)
CV,MH,HF	13155	73.8 [64.1, 81.6]	2166 (16.5%)	2440 (18.5%)	2587 (19.7%)	2745 (20.9%)	3217 (24.5%)	7389 (56.2%)	5766 (43.8%)	12532 (95.3%)	358 (2.7%)	168 (1.3%)	54 (0.4%)	43 (0.3%)
T2,CK,MH	28205	72.4 [64.6, 79.5]	4703 (16.7%)	5280 (18.7%)	5437 (19.3%)	5951 (21.1%)	6834 (24.2%)	9266 (32.9%)	18939 (67.1%)	25225 (89.4%)	1550 (5.5%)	1083 (3.8%)	224 (0.8%)	123 (0.4%)
T2,CK,HF	8441	79.1 [72, 84.7]	1373 (16.3%)	1586 (18.8%)	1741 (20.6%)	1828 (21.7%)	1913 (22.7%)	4491 (53.2%)	3950 (46.8%)	7257 (86%)	556 (6.6%)	506 (6%)	87 (1%)	35 (0.4%)
T2,MH,HF	4763	69.8 [60.9, 78.2]	658 (13.8%)	774 (16.3%)	921 (19.3%)	1062 (22.3%)	1348 (28.3%)	2346 (49.3%)	2417 (50.7%)	4249 (89.2%)	258 (5.4%)	194 (4.1%)	46 (1%)	16 (0.3%)
CK,MH,HF	6247	80.5 [72.6, 86.2]	1189 (19%)	1315 (21.1%)	1279 (20.5%)	1220 (19.5%)	1244 (19.9%)	1913 (30.6%)	4334 (69.4%)	6033 (96.6%)	109 (1.7%)	80 (1.3%)	13 (0.2%)	12 (0.2%)

CV,T2,CK,MH	18628	76 [69.3, 82.2]	2959 (15.9%)	3354 (18%)	3577 (19.2%)	4052 (21.8%)	4686 (25.2%)	8057 (43.3%)	10571 (56.7%)	16827 (90.3%)	1137 (6.1%)	465 (2.5%)	144 (0.8%)	55 (0.3%)
CV,T2,CK,HF	14282	78.8 [72.6, 84.2]	2295 (16.1%)	2585 (18.1%)	2877 (20.1%)	3024 (21.2%)	3501 (24.5%)	9110 (63.8%)	5172 (36.2%)	12380 (86.7%)	1236 (8.7%)	499 (3.5%)	109 (0.8%)	58 (0.4%)
CV,T2,MH,HF	6494	71.9 [63.9, 79.1]	851 (13.1%)	1045 (16.1%)	1238 (19.1%)	1491 (23%)	1869 (28.8%)	3952 (60.9%)	2542 (39.1%)	5817 (89.6%)	462 (7.1%)	143 (2.2%)	43 (0.7%)	29 (0.4%)
CV,CK,MH,HF	9291	80.8 [74.2, 86.2]	1627 (17.5%)	1795 (19.3%)	1873 (20.2%)	1881 (20.2%)	2115 (22.8%)	4162 (44.8%)	5129 (55.2%)	8977 (96.6%)	189 (2%)	79 (0.9%)	28 (0.3%)	18 (0.2%)
T2,CK,MH,HF	3244	77.2 [69.8, 83.5]	489 (15.1%)	589 (18.2%)	630 (19.4%)	712 (21.9%)	824 (25.4%)	1224 (37.7%)	2020 (62.3%)	2918 (90%)	162 (5%)	117 (3.6%)	30 (0.9%)	17 (0.5%)
All diseases	6162	78 [71, 83.6]	860 (14%)	1075 (17.4%)	1196 (19.4%)	1381 (22.4%)	1650 (26.8%)	3087 (50.1%)	3075 (49.9%)	5548 (90%)	387 (6.3%)	156 (2.5%)	51 (0.8%)	20 (0.3%)

* Age represents age at entry into the disease state

Supplemental Table B. Summary of disease transitions

Summary of CVD diagnosis grouped by comorbidity

Previous Comorbidity		Incidence Rate of CVD	Number of Events
4 conditions	T2D,CKD,MH,HF	0.059	521
3 conditions	CKD,MH,HF	0.052	960
	T2D,MH,HF	0.061	741
	T2D,CKD,HF	0.058	1416
	T2D,CKD,MH	0.025	3370
2 conditions	MH,HF	0.06	2313
	CKD,HF	0.056	2869
	CKD,MH	0.02	8343
	T2D,HF	0.063	1788
	T2D,MH	0.014	8677
	T2D,CKD	0.023	8525
1 condition	HF	0.063	6047
	MH	0.004	18165
	CKD	0.02	23075
	T2D	0.012	15838
0 condition	NONE	0.00046	207,102

Summary of T2D diagnosis grouped by comorbidity

Previous Comorbidity		Incidence Rate of T2D	Number of Events
4 conditions	CVD,CKD,MH,HF	0.023	720
3 conditions	CKD,MH,HF	0.02	375
	CVD,MH,HF	0.025	1101

	CVD,CKD,HF	0.022	1790
	CVD,CKD,MH	0.017	3111
2 conditions	MH,HF	0.023	871
	CKD,HF	0.02	1012
	CKD,MH	0.013	5706
	CVD,HF	0.024	2362
	CVD,MH	0.019	9036
	CVD,CKD	0.016	7512
1 condition	HF	0.021	2012
	MH	0.0055	25540
	CKD	0.013	15174
	CVD	0.017	16224
0 condition	NONE	0.0006	269,791

Summary of CKD diagnosis grouped by comorbidity

Previous Comorbidity		Incidence Rate of CKD	Number of Events
4 conditions	CVD,T2D,MH,HF	0.071	1449
3 conditions	T2D,MH,HF	0.066	812
	CVD,MH,HF	0.054	2332
	CVD,T2D,HF	0.087	3703
	CVD,T2D,MH	0.03	5101
2 conditions	MH,HF	0.048	1844
	T2D,HF	0.081	2306
	T2D,MH	0.014	8813
	CVD,HF	0.062	6117
	CVD,MH	0.021	10103
	CVD,T2D	0.034	11028
1 condition	HF	0.056	5408
	MH	0.0029	13347

	T2D	0.014	19008
	CVD	0.023	21873
0 condition	NONE	0.00047	211,219

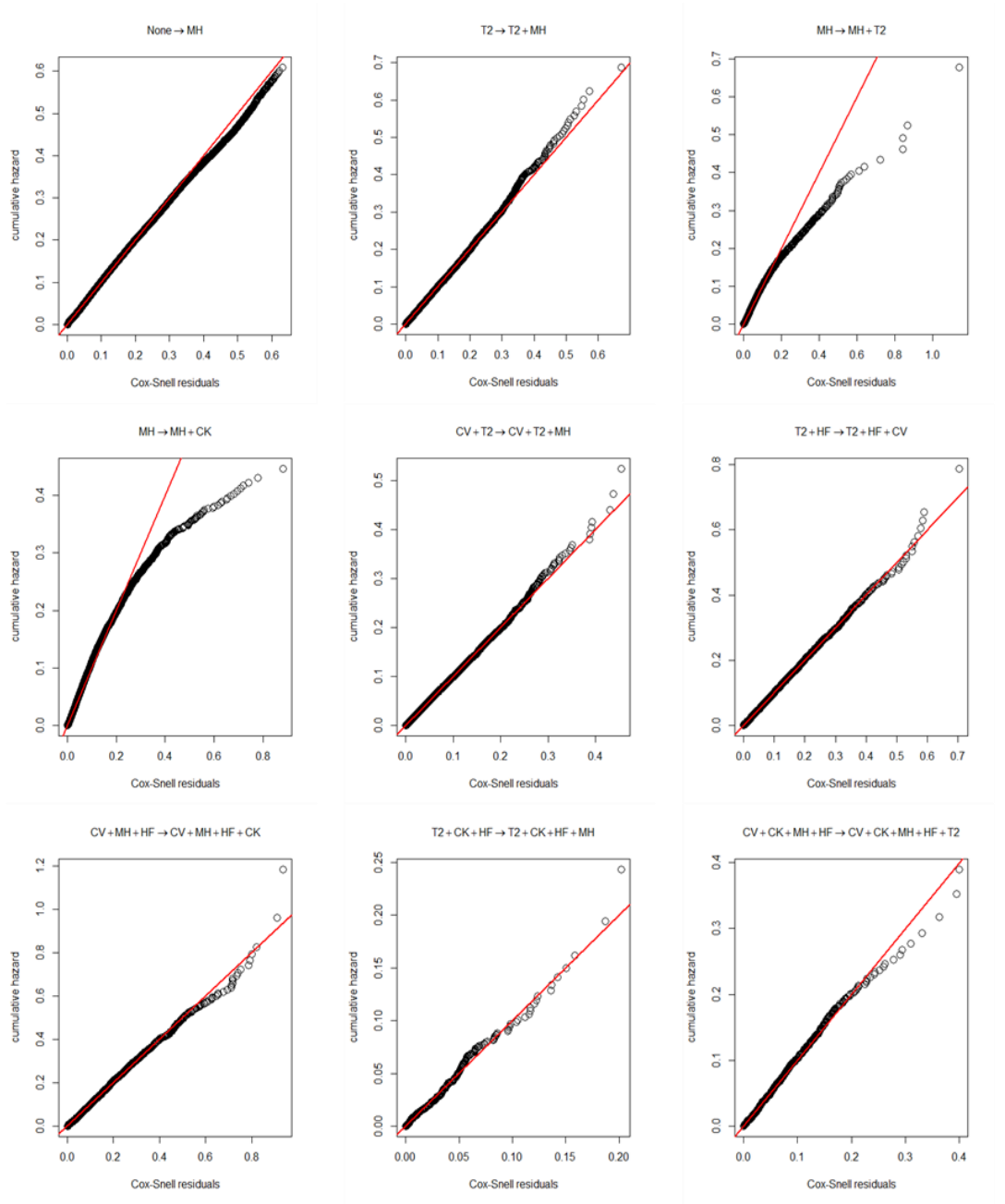
Summary of MH diagnosis grouped by comorbidity

Previous Comorbidity		Incidence Rate of MH	Number of Events
4 conditions	CVD,T2D,CKD,HF	0.015	702
3 conditions	T2D,CKD,HF	0.013	316
	CVD,CKD,HF	0.014	1197
	CVD,T2D,HF	0.016	674
	CVD,T2D,CKD	0.013	2488
2 conditions	CKD,HF	0.012	635
	T2D,HF	0.016	466
	T2D.CKD	0.011	4093
	CVD,HF	0.016	1526
	CVD,CKD	0.013	5780
	CVD,T2D	0.015	4943
1 condition	HF	0.015	1425
	CKD	0.011	12476
	T2D	0.014	18600
	CVD	0.016	15087
0 condition	NONE	0.002	947,978

Summary of HF diagnosis grouped by comorbidity

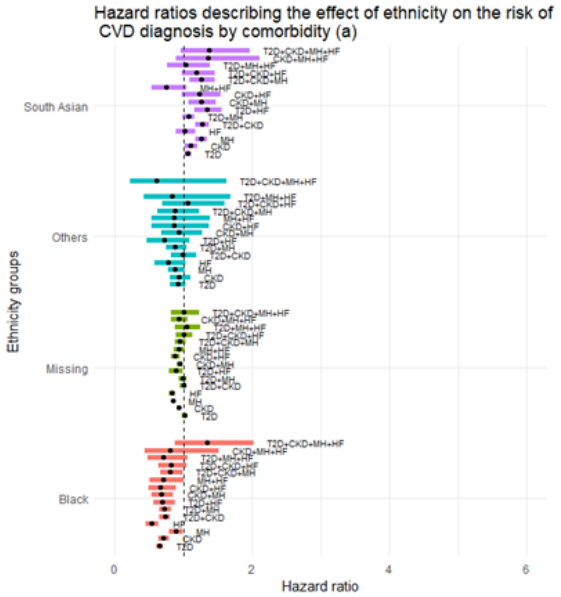
Previous Comorbidity		Incidence Rate of HF	Number of Events
4 conditions	CVD,T2D,CKD,MH	0.024	1924
3 conditions	T2D,CKD,MH	0.01	1321
	CVD,CKD,MH	0.017	3054
	CVD,T2D,MH	0.012	2089
	CVD,T2D,CKD	0.024	4664
2 conditions	CKD,MH	0.0062	2644
	T2D,MH	0.0031	1928
	T2D,CKD	0.0098	3631
	CVD,MH	0.0088	4242
	CVD,CKD	0.018	8085
	CVD,T2D	0.013	4304
1 condition	MH	0.00071	3310
	CKD	0.0066	7707
	T2D	0.0027	3660
	CVD	0.0095	8868
0 condition	NONE	0.000079	35434

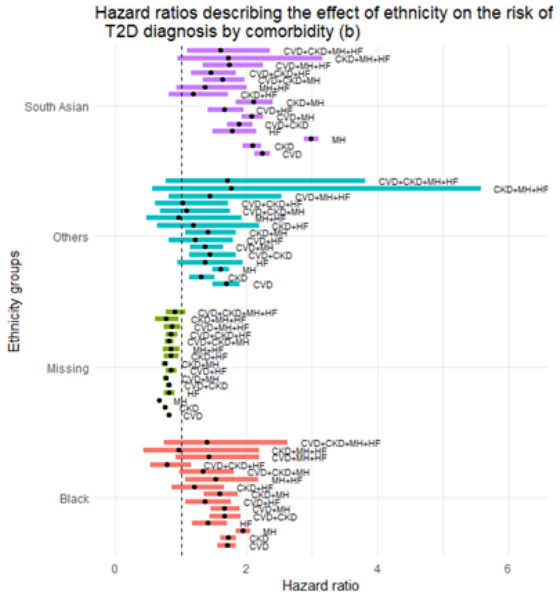
Supplemental Figure B. Plots of the estimated cumulative hazard of the Cox-Snell residuals.

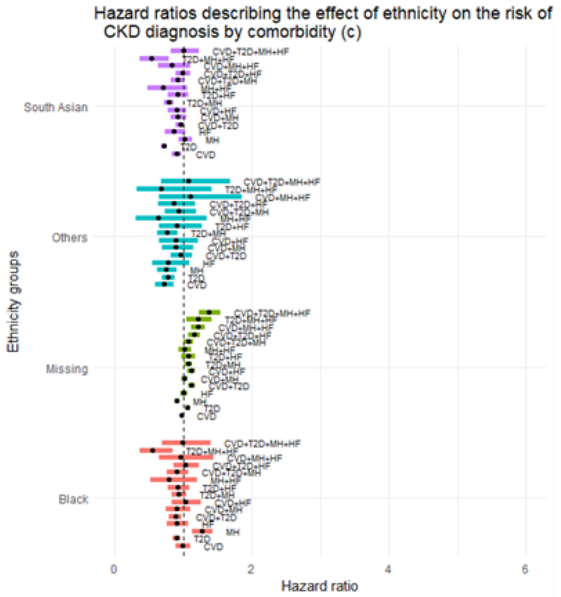


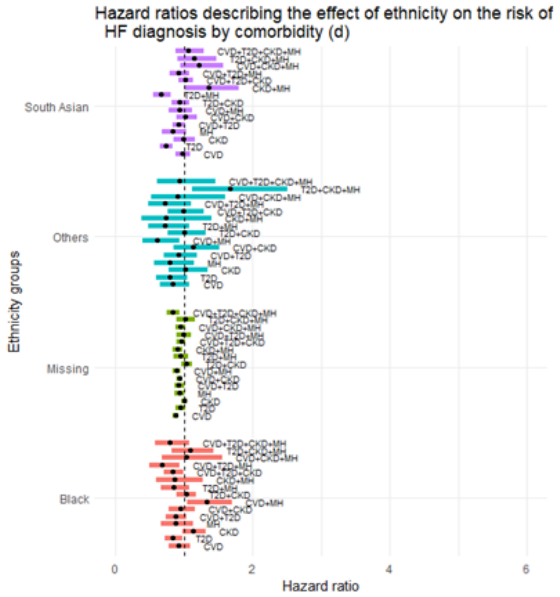
Note: The red line is a reference line with an intercept of zero and a slope of one.

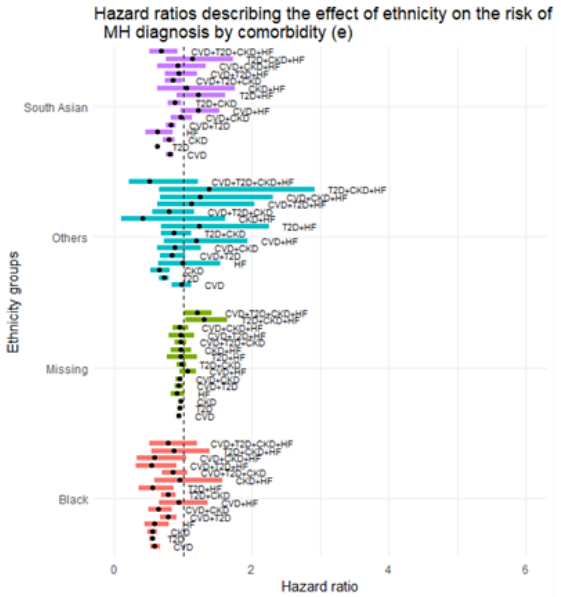
Supplemental Figure C. Effect of the four sociodemographic characteristics on the rate of disease transition by comorbidity. Results are obtained based on the cohort including those with missing ethnicity information. Graphical settings are the same as in Figures 2-5 of the main paper.

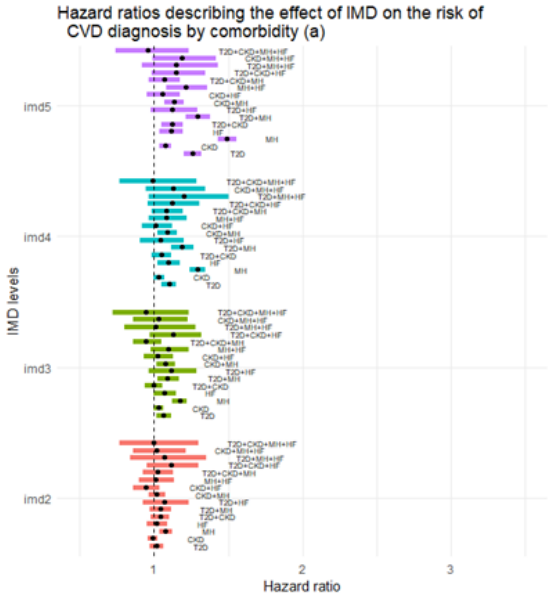

Hazard ratios describing the effect of ethnicity on the risk of CVD diagnosis by comorbidity (a)
Ethnicity groups
South Asian
Others
Missing
Black
Hazard ratio

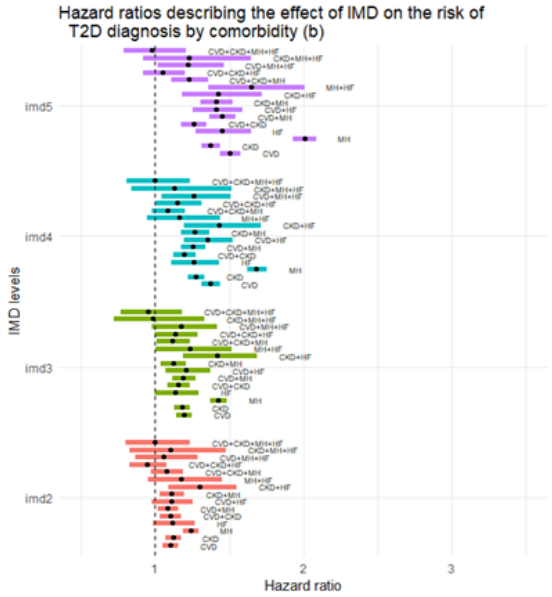

Hazard ratios describing the effect of ethnicity on the risk of T2D diagnosis by comorbidity (b)
Ethnicity groups
South Asian
Others
Missing
Black
Hazard ratio

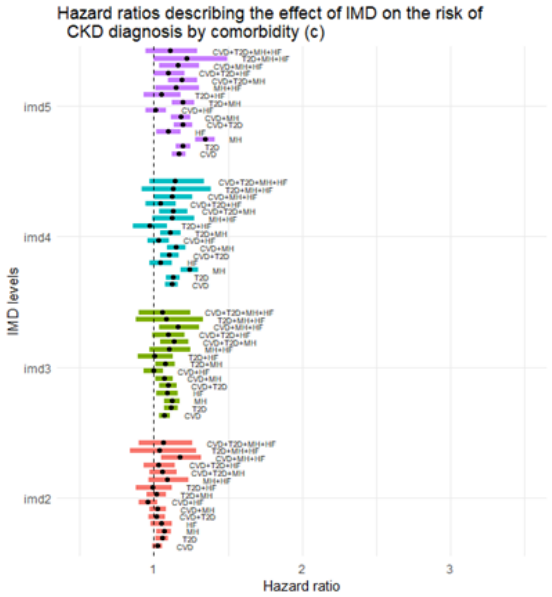

Hazard ratios describing the effect of ethnicity on the risk of CKD diagnosis by comorbidity (c)
Ethnicity groups
South Asian
Others
Missing
Black
Hazard ratio

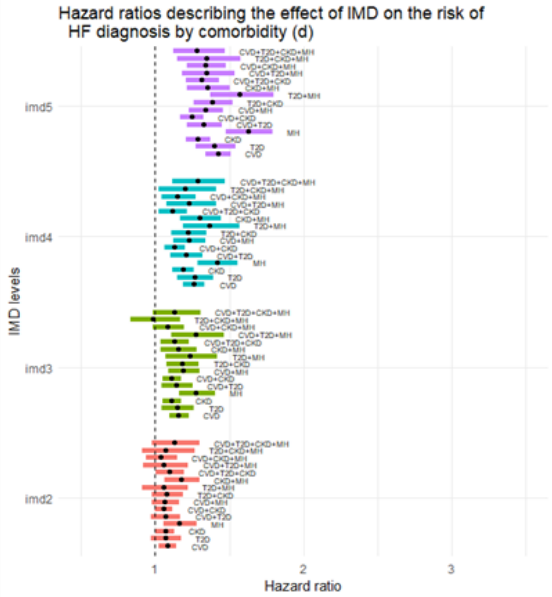

Hazard ratios describing the effect of ethnicity on the risk of HF diagnosis by comorbidity (d)
Ethnicity groups
South Asian
Others
Missing
Black
Hazard ratio

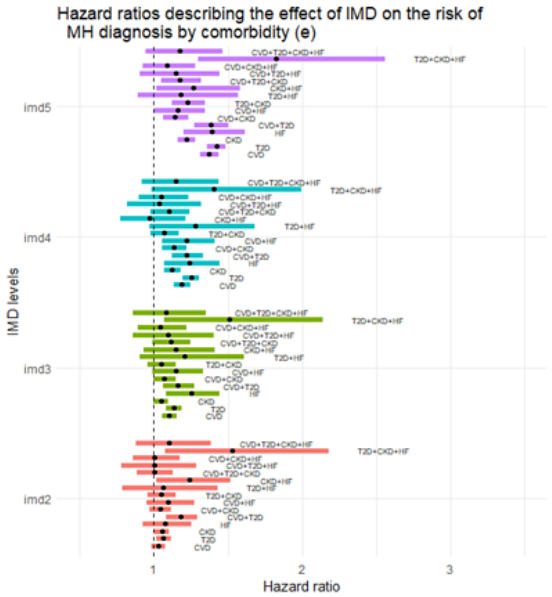

Hazard ratios describing the effect of ethnicity on the risk of MH diagnosis by comorbidity (e)
Ethnicity groups
South Asian
Others
Missing
Black
Hazard ratio

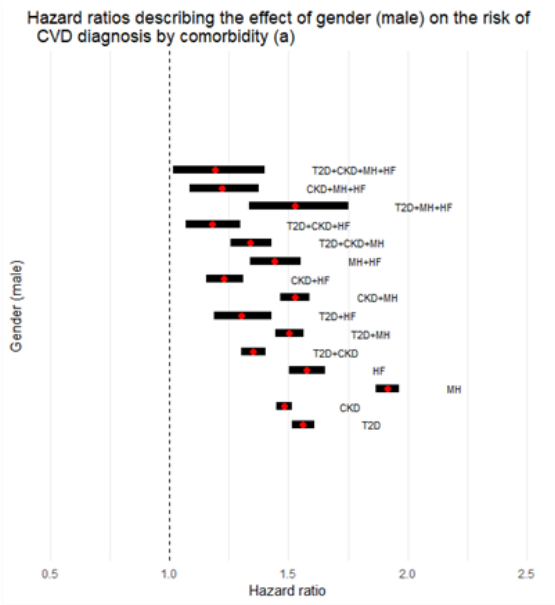

Hazard ratios describing the effect of IMD on the risk of CVD diagnosis by comorbidity (a)
IMD levels
imd5
imd4
imd3
imd2
1
2
3
Hazard ratio

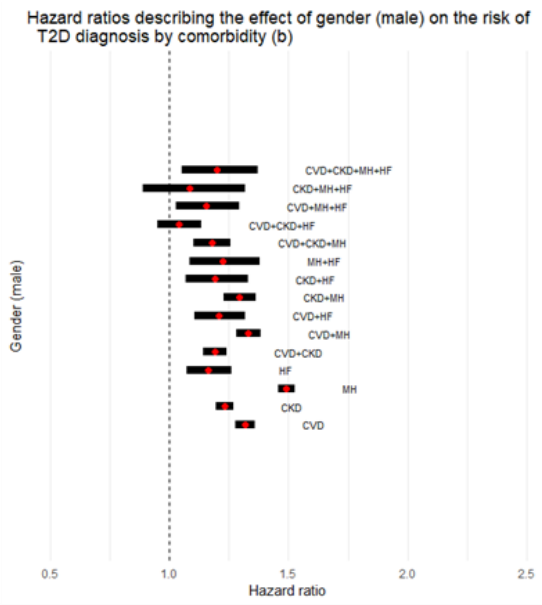

Hazard ratios describing the effect of IMD on the risk of T2D diagnosis by comorbidity (b)
IMD levels
imd5
imd4
imd3
imd2
1
2
3
Hazard ratio

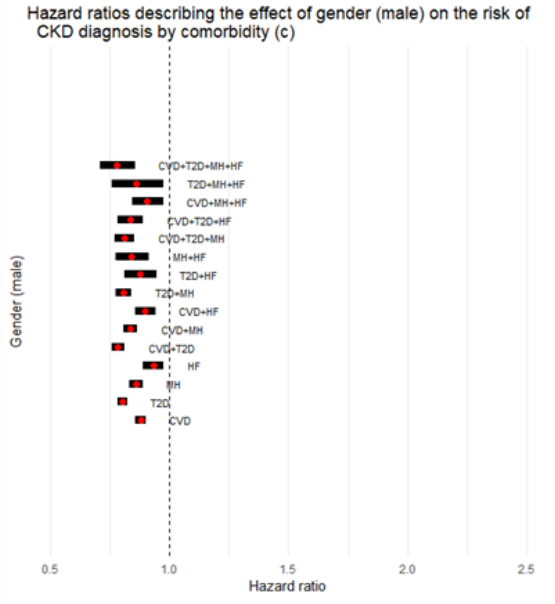

Hazard ratios describing the effect of IMD on the risk of CKD diagnosis by comorbidity (c)
IMD levels
imd5
imd4
imd3
imd2
1
2
3
Hazard ratio

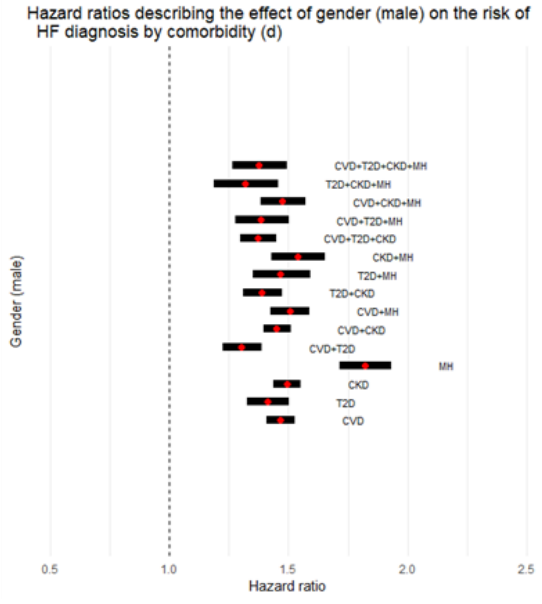

Hazard ratios describing the effect of IMD on the risk of HF diagnosis by comorbidity (d)
IMD levels
imd5
imd4
imd3
imd2
1
2
3
Hazard ratio

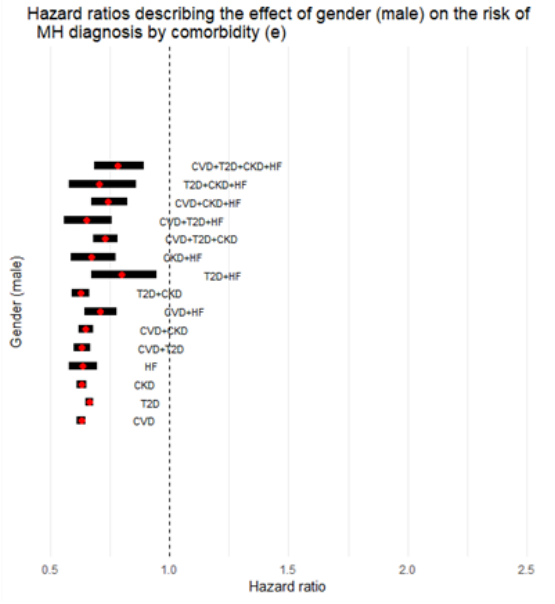

Hazard ratios describing the effect of IMD on the risk of MH diagnosis by comorbidity (e)
IMD levels
imd5
imd4
imd3
imd2
1
2
3
Hazard ratio

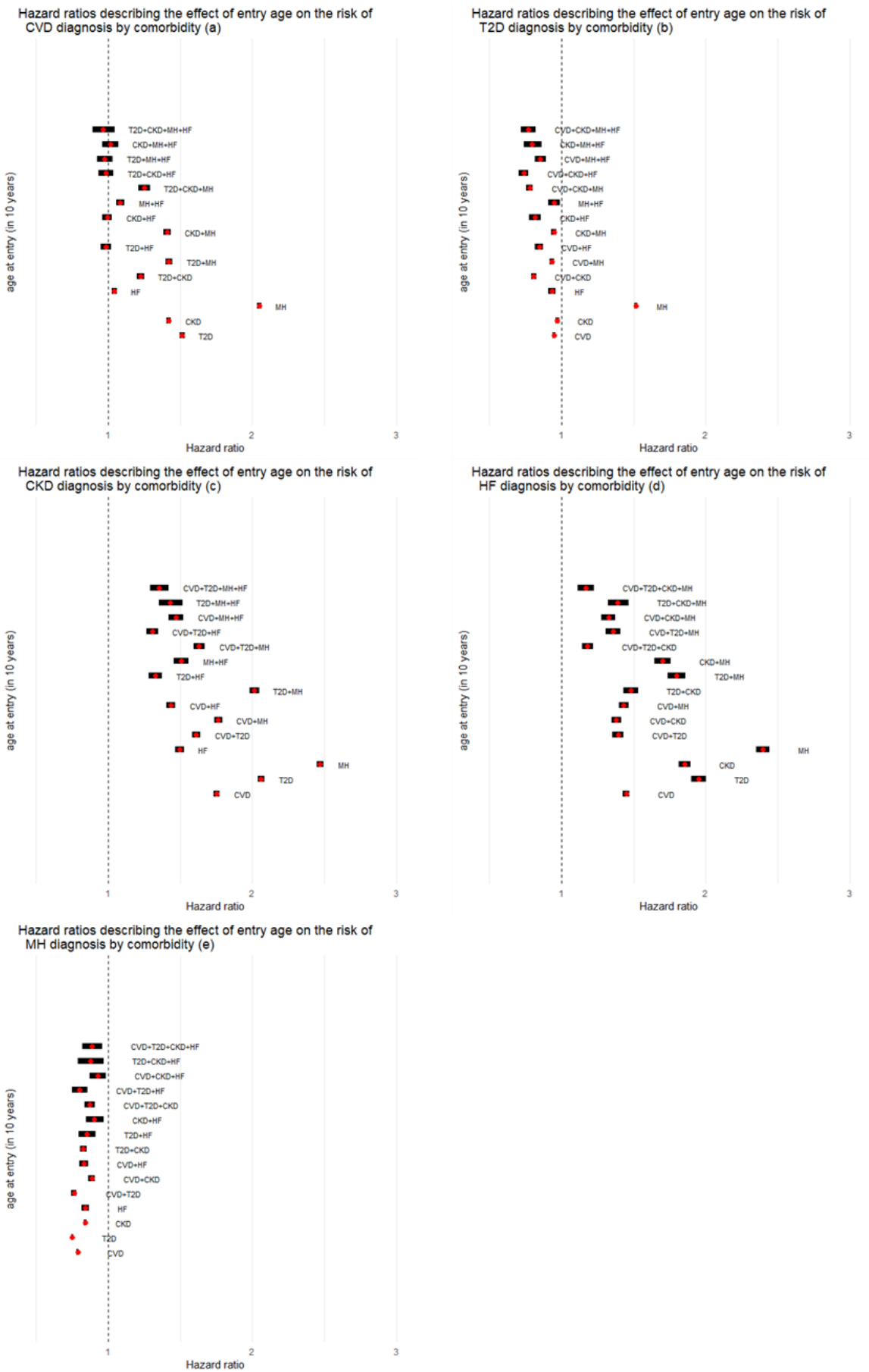

Hazard ratios describing the effect of gender (male) on the risk of CVD diagnosis by comorbidity (a)
Gender (male)
T2D+CKD+MH+HF
CKD+MH+HF
T2D+MH+HF
T2D+CKD+HF
T2D+CKD+MH
MH+HF
CKD+HF
CKD+MH
T2D+HF
T2D+MH
T2D+CKD
HF
MH
CKD
T2D
0.5
1.0
1.5
2.0
2.5
Hazard ratio


Hazard ratios describing the effect of gender (male) on the risk of T2D diagnosis by comorbidity (b)
Gender (male)
CVD+CKD+MH+HF
CKD+MH+HF
CVD+MH+HF
CVD+CKD+HF
CVD+CKD+MH
MH+HF
CKD+HF
CKD+MH
CVD+HF
CVD+MH
CVD+CKD
HF
MH
CKD
CVD
0.5
1.0
1.5
2.0
2.5
Hazard ratio


Hazard ratios describing the effect of gender (male) on the risk of CKD diagnosis by comorbidity (c)
Gender (male)
CVD+T2D+MH+HF
T2D+MH+HF
CVD+MH+HF
CVD+T2D+HF
CVD+T2D+MH
MH+HF
T2D+HF
T2D+MH
CVD+HF
CVD+MH
CVD+T2D
HF
MH
T2D
CVD
0.5
1.0
1.5
2.0
2.5
Hazard ratio


Hazard ratios describing the effect of gender (male) on the risk of HF diagnosis by comorbidity (d)
Gender (male)
CVD+T2D+CKD+MH
T2D+CKD+MH
CVD+CKD+MH
CVD+T2D+MH
CVD+T2D+CKD
CKD+MH
T2D+MH
T2D+CKD
CVD+MH
CVD+CKD
CVD+T2D
MH
CKD
T2D
CVD
0.5
1.0
1.5
2.0
2.5
Hazard ratio


Hazard ratios describing the effect of gender (male) on the risk of MH diagnosis by comorbidity (e)
Gender (male)
CVD+T2D+CKD+HF
T2D+CKD+HF
CVD+CKD+HF
CVD+T2D+HF
CVD+T2D+CKD
CKD+HF
T2D+HF
T2D+CKD
CVD+HF
CVD+CKD
CVD+T2D
HF
CKD
T2D
CVD
0.5
1.0
1.5
2.0
2.5
Hazard ratio


Hazard ratios describing the effect of entry age on the risk of CVD diagnosis by comorbidity (a)
T2D+CKD+MH+HF
CKD+MH+HF
T2D+MH+HF
T2D+CKD+HF
T2D+CKD+MH
MH+HF
CKD+HF
CKD+MH
T2D+HF
T2D+MH
T2D+CKD
HF
MH
CKD
T2D
age at entry (in 10 years)
1
2
3
Hazard ratio
Hazard ratios describing the effect of entry age on the risk of T2D diagnosis by comorbidity (b)
CVD+CKD+MH+HF
CKD+MH+HF
CVD+MH+HF
CVD+CKD+HF
CVD+CKD+MH
MH+HF
CKD+HF
CKD+MH
CVD+HF
CVD+MH
CVD+CKD
HF
MH
CKD
CVD
age at entry (in 10 years)
1
2
3
Hazard ratio
Hazard ratios describing the effect of entry age on the risk of CKD diagnosis by comorbidity (c)
CVD+T2D+MH+HF
T2D+MH+HF
CVD+MH+HF
CVD+T2D+HF
CVD+T2D+MH
MH+HF
T2D+HF
T2D+MH
CVD+HF
CVD+MH
CVD+T2D
HF
MH
T2D
CVD
age at entry (in 10 years)
1
2
3
Hazard ratio
Hazard ratios describing the effect of entry age on the risk of HF diagnosis by comorbidity (d)
CVD+T2D+CKD+MH
T2D+CKD+MH
CVD+CKD+MH
CVD+T2D+MH
CVD+T2D+CKD
CKD+MH
T2D+MH
T2D+CKD
CVD+MH
CVD+CKD
CVD+T2D
MH
CKD
T2D
CVD
age at entry (in 10 years)
1
2
3
Hazard ratio
Hazard ratios describing the effect of entry age on the risk of MH diagnosis by comorbidity (e)
CVD+T2D+CKD+HF
T2D+CKD+HF
CVD+CKD+HF
CVD+T2D+HF
CVD+T2D+CKD
CKD+HF
T2D+HF
T2D+CKD
CVD+HF
CVD+CKD
CVD+T2D
HF
CKD
T2D
CVD
age at entry (in 10 years)
1
2
3
Hazard ratio

References

1. van den Akker M, Buntinx F, Knottnerus JA. Comorbidity or multimorbidity. Eur J Gen Pract. 1996;2: 65–70.

2. Wallace E, Salisbury C, Guthrie B, Lewis C, Fahey T, Smith SM. Managing patients with multimorbidity in primary care. BMJ. 2015;350: h176.

3. Soley-Bori M, Ashworth M, Bisquera A, Dodhia H, Lynch R, Wang Y, et al. Impact of multimorbidity on healthcare costs and utilisation: a systematic review of the UK literature. Br J Gen Pract. 2021;71: e39–e46.

4. Nunes BP, Flores TR, Mielke GI, Thumé E, Facchini LA. Multimorbidity and mortality in older adults: A systematic review and meta-analysis. Arch Gerontol Geriatr. 2016;67: 130–138.

5. Identifying longitudinal clusters of multimorbidity in an urban setting: A population-based cross-sectional study. The Lancet Regional Health - Europe. 2021;3: 100047.

6. Busija L, Lim K, Szoeke C, Sanders KM, McCabe MP. Do replicable profiles of multimorbidity exist? Systematic review and synthesis. Eur J Epidemiol. 2019;34: 1025–1053.

7. Zhu Y, Edwards D, Mant J, Payne RA, Kiddle S. Characteristics, service use and mortality of clusters of multimorbid patients in England: a population-based study. BMC Med. 2020;18: 78.

8. Cezard G, McHale CT, Sullivan F, Bowles JKF, Keenan K. Studying trajectories of multimorbidity: a systematic scoping review of longitudinal approaches and evidence. BMJ Open. 2021;11: e048485.

9. Ashworth M, Durbaba S, Whitney D, Crompton J, Wright M, Dodhia H. Journey to multimorbidity: longitudinal analysis exploring cardiovascular risk factors and sociodemographic determinants in an urban setting. BMJ Open. 2019;9: e031649.

10. Siriwardhana C, Lim E, Davis J, Chen JJ. Progression of diabetes, ischemic heart disease, and chronic kidney disease in a three chronic conditions multistate model. BMC Public Health. 2018;18: 1–14.

11. Freisling H, Viallon V, Lennon H, Bagnardi V, Ricci C, Butterworth AS, et al. Lifestyle factors and risk of multimorbidity of cancer and cardiometabolic diseases: a multinational cohort study. BMC Med. 2020;18: 1–11.

12. Owen R, Lyons J, Akbari A, Guthrie B, Agrawal U, Alexander D, et al. Temporal sequencing in multimorbidity using population-scale linked data for 1.7 million individuals with 20-year follow-up. Research Square. 2022. doi:10.21203/rs.3.rs-1537576/v1

13. Chan MS, van den Hout A, Pujades-Rodriguez M, Jones MM, Matthews FE, Jagger C, et al. Socio-economic inequalities in life expectancy of older adults with and without multimorbidity: a record linkage study of 1.1 million people in England. Int J Epidemiol. 2019;48: 1340–1351.

14. Bisquera A, Turner EB, Ledwaba-Chapman L, Dunbar-Rees R, Hafezparast N, Gulliford M, et al. Inequalities in developing multimorbidity over time: A population-based cohort study from an urban, multi-ethnic borough in the United Kingdom. Lancet Reg Health Eur. 2022;12: 100247.

15. Han Y, Hu Y, Yu C, Guo Y, Pei P, Yang L, et al. Lifestyle, cardiometabolic disease, and multimorbidity in a prospective Chinese study. Eur Heart J. 2021;42: 3374–3384.

16. Singh-Manoux A, Fayosse A, Sabia S, Tabak A, Shipley M, Dugravot A, et al. Clinical, socioeconomic, and behavioural factors at age 50 years and risk of cardiometabolic multimorbidity and mortality: A cohort study. PLoS Med. 2018;15: e1002571.

17. Barnett K, Mercer SW, Norbury M, Watt G, Wyke S, Guthrie B. Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. Lancet. 2012;380: 37–43.

18. Wolf A, Dedman D, Campbell J, Booth H, Lunn D, Chapman J, et al. Data resource profile: Clinical Practice Research Datalink (CPRD) Aurum. Int J Epidemiol. 2019;48: 1740–1740g.

19. List of ethnic groups. [cited 21 Feb 2023]. Available: https://www.ethnicity-facts-figures.service.gov.uk/style-guide/ethnic-groups

20. Ministry of Housing, Communities, Local Government. English indices of deprivation 2019. GOV.UK; 2019. Available: https://www.gov.uk/government/statistics/english-indices-of-deprivation-2019

21. Crowther MJ, Lambert PC. Parametric multistate survival models: Flexible modelling allowing transition-specific distributions with application to estimating clinically useful measures of effect differences. Stat Med. 2017;36: 4719–4742.

22. Royston P, Parmar MKB. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. Stat Med. 2002;21: 2175–2197.

23. Jackson CH. flexsurv: A Platform for Parametric Survival Modeling in R. J Stat Softw. 2016;70. doi:10.18637/jss.v070.i08

24. Tillin T, Hughes AD, Mayet J, Whincup P, Sattar N, Forouhi NG, et al. The relationship between metabolic risk factors and incident cardiovascular disease in Europeans, South Asians, and African Caribbeans: SABRE (Southall and Brent Revisited) -- a prospective population-based study. J Am Coll Cardiol. 2013;61: 1777–1786.

25. Hirst JA, Hill N, O'Callaghan CA, Lasserson D, McManus RJ, Ogburn E, et al. Prevalence of chronic kidney disease in the community using data from OxRen: a UK population-based cohort study. Br J Gen Pract. 2020;70: e285–e293.

26. Feakins B, Oke J, McFadden E, Aronson J, Lasserson D, O'Callaghan C, et al. Trends in kidney function testing in UK primary care since the introduction of the quality and outcomes framework: a retrospective cohort study using CPRD. BMJ Open. 2019;9: e028062.

27. Tomlinson LA, Clase CM. Sex and the Incidence and Prevalence of Kidney Disease. Clin J Am Soc Nephrol. 2019;14: 1557–1559.

28. Williams ED, Tillin T, Richards M, Tuson C, Chaturvedi N, Hughes AD, et al. Depressive symptoms are doubled in older British South Asian and Black Caribbean people compared with Europeans: associations with excess co-morbidity and socioeconomic disadvantage. Psychol Med. 2015;45: 1861–1871.

29. McManus S, Bebbington P, Jenkins R, Brugha T. Mental Health and Wellbeing in England: the Adult Psychiatric Morbidity Survey 2014. McManus S, Bebbington P, Jenkins R, Brugha T, editors. 2016 [cited 18 Dec 2022]. Available: https://openaccess.city.ac.uk/id/eprint/23646/

30. Vaidya V, Partha G, Karmakar M. Gender differences in utilization of preventive care services in the

United States. J Womens Health . 2012;21: 140–145.

31. Jha MK, Qamar A, Vaduganathan M, Charney DS, Murrough JW. Screening and Management of Depression in Patients With Cardiovascular Disease: JACC State-of-the-Art Review. J Am Coll Cardiol. 2019;73: 1827–1845.

32. Bonnet F, Irving K, Terra J-L, Nony P, Berthezène F, Moulin P. Anxiety and depression are associated with unhealthy lifestyle in patients at risk of cardiovascular disease. Atherosclerosis. 2005;178: 339–344.

33. QOF 2021-22. [cited 4 Nov 2022]. Available: https://qof.digital.nhs.uk/

34. Robson J, Garriga C, Coupland C, Hippisley-Cox J. NHS Health Checks: an observational study of equity and outcomes 2009-2017. Br J Gen Pract. 2021;71: e701–e710.

35. Pal K, Horsfall L, Sharma M, Nazareth I, Petersen I. Time trends in the incidence of clinically diagnosed type 2 diabetes and pre-diabetes in the UK 2009–2018: a retrospective cohort study. BMJ Open Diabetes Research and Care. 2021;9: e001989.

36. Weich S, Nazroo J, Sproston K, McManus S, Blanchard M, Erens B, et al. Common mental disorders and ethnicity in England: the EMPIRIC study. Psychol Med. 2004;34: 1543–1551.

37. Rizopoulos D. Joint Models for Longitudinal and Time-to-Event Data: With Applications in R. CRC Press; 2012.

38. Cook RJ, Lawless JF. Multistate models for the analysis of life history data. [cited 22 Oct 2022]. doi:10.1201/9781315119731/multistate-models-analysis-life-history-data-richard-cook-jerald-lawless

39. Ieva F, Jackson CH, Sharples LD. Multi-state modelling of repeated hospitalisation and death in patients with heart failure: The use of large administrative databases in clinical epidemiology. Stat Methods Med Res. 2017;26: 1350–1372.

40. Cox DR, Snell EJ. A general definition of residuals. J R Stat Soc. 1968;30: 248–265.

41. Collett D. Modelling Survival Data in Medical Research. CRC Press; 2015.