Creating a Modified Version of the Cambridge Multimorbidity Score (CMMS) to Predict Mortality in People Over 16 Years in the English Nationwide General Practice Extraction Service Data for Pandemic Planning and Research (GDPPR) Dataset: Model Development and Validation

Debasish Kar, Kathryn Suzann Taylor, Mark Joy, Sudhir Venkatesan, Wilhelmine Meeraus, Sylvia Taylor, Sneha Anand, Filipa Ferreira, Gavin Jamie, Xuejuan Fan, Simon de Lusignan

Submitted to: Journal of Medical Internet Research
on: January 03, 2024

Disclaimer: © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on it's website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressively prohibit redistribution of this draft paper other than for review purposes.

Table of Contents

Creating a Modified Version of the Cambridge Multimorbidity Score (CMMS) to Predict Mortality in People Over 16 Years in the English Nationwide General Practice Extraction Service Data for Pandemic Planning and Research (GDPPR) Dataset: Model Development and Validation

Debasish Kar[1*] MBBS; Kathryn Suzann Taylor[1*] PhD; Mark Joy[1] PhD; Sudhir Venkatesan[2] PhD; Wilhelmine Meeraus[3] PhD; Sylvia Taylor[3] PhD; Sneha Anand[1] PhD; Filipa Ferreira[1] PhD; Gavin Jamie[1] MBBS, MSc; Xuejuan Fan[1] PhD; Simon de Lusignan[1] MBBS

[1]Nuffield Department of Primary Care Health Sciences University of Oxford Oxford GB
[2]Medical & Payer Evidence Statistics, BioPharmaceuticals Medical AstraZeneca PLC Cambridge GB
[3]Medical Evidence, Vaccines and Immune Therapies AstraZeneca PLC Cambridge GB
[*]these authors contributed equally

Corresponding Author:
Kathryn Suzann Taylor PhD
Nuffield Department of Primary Care Health Sciences
University of Oxford
Radcliffe Primary Care Building, Radcliffe Observatory Quarter
Woodstock Road
Oxford
GB

Abstract

Background: No single multimorbidity measure is validated for use in NHS England's General Practice Extraction Service Data for Pandemic Planning and Research (GDPPR), the nationwide primary care dataset created for coronavirus disease 19 (COVID-19) pandemic research. A single morbidity measure is advantageous when there is a need to adjust for multimorbidity, such as modelling the effectiveness of vaccinations against COVID-19, as including multiple individual morbidities is challenging. The Cambridge Multimorbidity Score (CMMS) is a validated tool for predicting mortality risk. However, the number of Systematised Nomenclature of Medicine clinical terms (SNOMED CT) for the GDPPR dataset is limited and does not define all the conditions used to calculate the CMMS

Objective: To develop and validate a modified version of CMMS using the clinical terms available for the GDPPR.

Methods: We used pseudonymised data from the Oxford-Royal College of General Practitioners Research and Surveillance Centre (RCGP RSC), which has a more extensive SNOMED CT list. From the 37 conditions used in the original CMMS model, we selected conditions either with: (a) high prevalence ratio (? 85%), calculated as the prevalence in the RSC data set as defined by the GDPPR set of SNOMED CT codes, divided by the prevalence as defined by the RSC set of SNOMED CT codes, or (b) conditions with lower prevalence ratio but with high predictive value. The resulting set of conditions was included in Cox proportional hazard models to determine the 1-year mortality risk in a development dataset (n=300,000) and construct a new CMMS model, following the original CMMS, with variable reduction and parsimony, achieved by backward elimination and Akaike information stopping criterion. Model validation involved obtaining 1-year mortality estimates for a synchronous dataset (n=150,000) and 1-year and 5-year mortality estimates for an asynchronous dataset (n=150,000).

Results: The initial model contained 22 conditions and our final model included 17 conditions. The conditions overlapped with those of a modified CMMS, which we previously developed using RSC data and the more extensive RSC SNOMED CT list. For 1-year mortality, discrimination was high in both the derivation and validation datasets (Harrell's C=0.92), and 5-year mortality was slightly lower (Harrell's C= 0.90), and the calibration was reasonable following an adjustment for over-fitting. The performance was similar to that of both the original and previous modified CMMS models.

Conclusions: The modified version of the CMMS can be used on the GDPPR, a nationwide primary care dataset of 54 million people, to predict mortality in people in real-world vaccine effectiveness, pandemic planning, and other research studies. It requires 17 variables to produce a comparable performance with our previous modification of CMMS to enable it to be used in routine data using SNOMED CT.

(JMIR Preprints 03/01/2024:56042)
DOI: https://doi.org/10.2196/preprints.56042

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Creating a Modified Version of the Cambridge Multimorbidity Score (CMMS) to Predict Mortality in People Over 16 Years in the English Nationwide General Practice Extraction Service Data for Pandemic Planning and Research (GDPPR) Dataset: Model Development and Validation

Debasish Kar* (DK) [1], Kathryn S Taylor* (KT) [1], Mark Joy (MJ) [1], Sudhir Venkatesan (SV) [3], Wilhelmine Meeraus (WM)[4], Sylvia Taylor (ST) [4], Sneha N Anand (SA) [1], Filipa Ferreira (FF)[1], Gavin Jamie (GJ) [1], Xuejuan Fan (XF) [1], Simon de Lusignan (SdeL) [1,2]
[1] Nuffield Department of Primary Care Health, University of Oxford
[2] Royal College of General Practitioners of the United Kingdom
[3] Medical & Payer Evidence Statistics, BioPharmaceuticals Medical, AstraZeneca, Cambridge UK
[4] Medical Evidence, Vaccines and Immune Therapies, AstraZeneca, Cambridge, UK
*Joint first author and contributed equally
Corresponding author – Kathryn S Taylor

Abstract

Background

No single multimorbidity measure is validated for use in NHS England's General Practice Extraction Service Data for Pandemic Planning and Research (GDPPR), the nationwide primary care dataset created for coronavirus disease 19 (COVID-19) pandemic research. The Cambridge Multimorbidity Score (CMMS) is a validated tool for predicting mortality risk, with 37 conditions defined by Read Codes. The GDPPR uses the more internationally used Systematised Nomenclature of Medicine clinical terms (SNOMED CT). We previously developed a modified version of the CMMS using SNOMED CT, but the number of terms for the GDPPR dataset is limited making it impossible to use this version.

Objectives

To develop and validate a modified version of CMMS using the clinical terms available for the GDPPR.

Methods

We used pseudonymised data from the Oxford-Royal College of General Practitioners Research and Surveillance Centre (RSC), which has an extensive SNOMED CT list. From the 37 conditions in the original CMMS model, we selected conditions either with: (a) high prevalence ratio (≥ 85%), calculated as the prevalence in the RSC data set but included in the GDPPR set of SNOMED CT codes, divided by the prevalence included in the RSC set of SNOMED CT codes, or (b) conditions with lower prevalence ratios but with high predictive value. The resulting set of conditions was included in Cox proportional hazard models to determine the 1-year mortality risk in a development dataset (n=300,000) and construct a new CMMS model, following the methods for the original CMMS study, with variable reduction and parsimony, achieved by backward elimination and Akaike information stopping criterion. Model validation involved obtaining 1-year mortality estimates for a synchronous dataset (n=150,000) and 1-year and 5-year mortality estimates for an asynchronous dataset (n=150,000). We compared the performance with that of the original CMMS and the modified CMMS that we previously developed using RSC data.

Results

The initial model contained 22 conditions and our final model included 17 conditions. The conditions overlapped with those of the modified CMMS using the more extensive SNOMED CT list. For 1-year mortality, discrimination was high in both the derivation and validation datasets (Harrell's C=0.92), and 5-year mortality was slightly lower (Harrell's C= 0.90) Calibration was reasonable following an adjustment for over-fitting. The performance was similar to that of both the original and previous modified CMMS models.

Conclusions

The new modified version of the CMMS can be used on the GDPPR, a nationwide primary care dataset of 54 million people, to enable adjustment for multimorbidity in predicting mortality in people in real-world vaccine effectiveness, pandemic planning, and other research studies. It requires 17 variables to produce a comparable performance with our previous modification of CMMS to enable it to be used in routine data using SNOMED CT.

Keywords:
Pandemics; COVID-19; Multimorbidity; Prevalence; Predictive Model; Discrimination; Calibration; Systematised Nomenclature of Medicine; Computerised Medical Records; Systems

Introduction

People with multimorbidity, defined by those with two or more long-term conditions (LTCs) [1-6], have complex needs and impose increasing demands on primary care services given the ageing population. Multimorbidity is associated with reduced life expectancy [7], lower quality of life [8] and an increased risk of hospitalisation and death with COVID-19 [9]. In clinical trials, vaccination against COVID-19 showed reduced risk of hospitalisation and death in all groups [10, 11]. However, in real-world studies, people with multimorbidity benefited less from vaccination [12] and were at increased risk of mortality, morbidity and hospitalisation, compared to those without multimorbidity [13]. People with five or more LTCs had a more than four-fold higher risk of severe COVID-19 outcomes than those with less than five LTCs [12].

Developing a single comorbidity measure is challenging [14]. The Charlson Comorbidity Index (CCI) is a commonly used tool to predict mortality over time [15]. However, CCI is based on hospital data, therefore, its applicability to primary care data is limited and not readily implementable [16]. The Cambridge Multimorbidity Score (CMMS) addressed this limitation and is an established measure of multimorbidity in primary care data. The original CMMS used 37 LTCs from routine primary care data in computerised medical records (CMR) to predict the risk of primary care consultations, unplanned hospital admissions and mortality [17]. It was developed and validated using the Clinical Practice Research Datalink (CPRD) [18] from the codes of Read version 2. However, Read version 2 is no longer used in England and is not updated since 2018 [19]. Additionally, the original CMMS model excluded people aged below 21 years, which somewhat restricted its applicability in the general population.

To overcome these limitations, we have already developed and validated a modified CMMS, replacing Read version 2 with the systematised nomenclature for medicine clinical terms (SNOMED CT) [20] and using pseudonymised data from the Oxford-Royal College of General Practitioners (RCGP) Research and Surveillance (RSC) sentinel network of individuals aged 16 years or older [21] Established in 1967, the RSC is an internationally renowned source of primary care data [22]. It has been used for influenza and respiratory disease monitoring for the last 50 years [23]. During the COVID-19 pandemic, with linkage to existing NHS England datasets, RSC data were also used to understand its epidemiology and assess vaccine effectiveness and safety [24-27]. This modified version of the CMMS was used to assess the real-world effectiveness of the Oxford-AstraZeneca Covid-19 Vaccine in England (RAVEN) study, which was run on the RSC using linked data from NHS England [28].

The RAVEN study also used primary care data from the larger, and nationwide General Practice Extraction Service (GPES) Data for Pandemic Planning and Research (GDPPR) data source maintained by NHS England, providing pseudonymised data for over 54 million people in England [29]. GDPPR is linked at the individual patient level to hospital, death, vaccine exposure, and test results. However, whilst substantial, its primary care data collection is incomplete. The primary care data were created from the existing list of conditions comprising 56,319 different SNOMED terms. This is a large number, but it is less than 20% of all SNOMED terms. These covered some clinical conditions well (e.g., diabetes) and some less so (e.g., psychoactive substance disorder). This study aimed to develop and validate a modified version of the CMMS, which could be used for the population aged 16 years and above in this new English NHS nationwide data set (GDPPR).

Methods

UK primary care data

In the UK, each patient registers with a single GP practice. Information about their primary care consultations, prescribing, investigation results, and certified sickness and mortality data are recorded in CMR systems. Each patient has a unique identifier, the NHS number, which allows data linkage with other datasets, including the hospital data, Hospital Episode Statistics (HES), death certificate data provided by the Office for National Statistics (ONS) and the NHS prescribing dataset [30].

Data sources

We used pseudonymised data from the RSC to construct and validate a revised version of CMMS based on the limited set of SNOMED CT codes (we refer to this as the GDPPR-modified CMMS). RSC data are stored in the Oxford Royal College of General Practitioners Digital Informatics Hub (ORCHID) trusted research environment (TRE). The RCGP RSC extracts data from just under 2000 general practices in England [31] and provides a dataset that is representative of individuals in England.

We applied the same analytical approach and the same inclusion criteria as in our previous study, where we developed and validated a new CMMS for RSC using the more extensive list of SNOMED CT codes (we refer to this as the RSC-modified CMMS) [21]. We included people aged 16 years or over on the index date registered with a practice for 12 months or longer. Three separate datasets were sampled from the RSC as described previously (Supplementary Figure S1 and Figure S2): (a) derivation dataset (n=300,000), (b) validation dataset 1 (n=150,000) with the same study start and study end date as the derivation set (synchronous outcome), and (c) validation dataset 2 (n=150,000) with 12 month outcome at a different time point to the derivation dataset (asynchronous outcome), and 60 month outcome occurring at the same time point as the 12 month outcome of the derivation dataset (synchronous outcome), as illustrated in Figure 1. These three datasets were generally comparable in terms of age, sex, number of conditions and follow-up time.

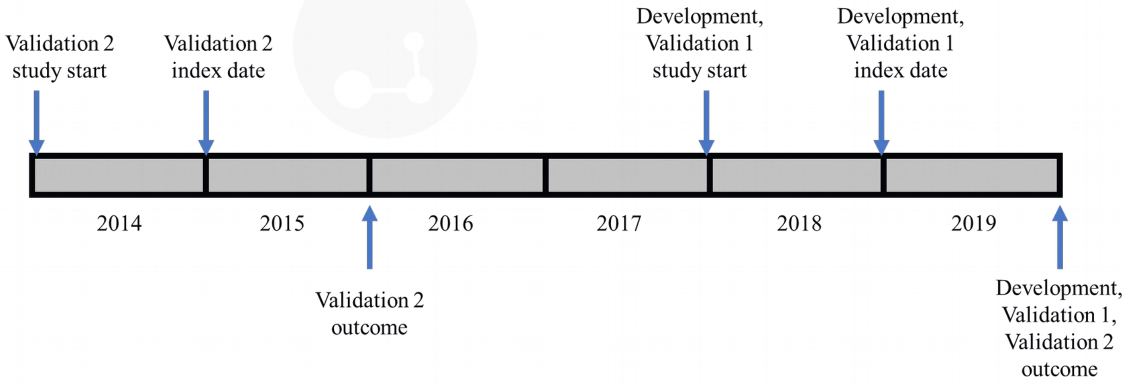

Validation 2 study start
Validation 2 index date
Development, Validation 1 study start
Development, Validation 1 index date
2014
2015
2016
2017
2018
2019
Validation 2 outcome
Development, Validation 1, Validation 2 outcome

Figure 1: Study design for the development and validation of the GDPPR-version of the CMMS, involving three cohorts, with index dates set at 12 months after their respective study

start dates, and synchronous and asynchronous outcomes at 1 year and 5 years (Reproduced with permission - Tsang et al., 2022) [21]

Curating and selecting individual CMMS component variables

In selecting conditions for a new modified CMMS model, we first considered all 37 conditions that were included in the original CMMS development and validation and used the same definitions and / or prescribing [17]. Following the approach we previously developed for the RSC-modified CMMS with SNOMED CT [21], we carefully curated the conditions within the limited set of SNOMED CT codes in the GDPPR. We performed a confirmatory study concerning SNOMED CT coverage in the GDPPR by carrying out a statistical and clinical matching of the conditions between the GDPPR and RSC (Figure 2). A matching percentage for each condition was defined as the prevalence ratio, that is, the prevalence in the RSC data set but included in the GDPPR set of SNOMED CT codes divided by the prevalence included in the RSC set of SNOMED CT codes. Therefore, a ratio less than 100% indicated a higher prevalence of that condition within the RSC dataset using its set of RSC SNOMED CT concepts, compared to the prevalence using the GDPPR set of SNOMED CT concepts on the same dataset. We set an 85% threshold for inclusion in the development of the GDPPR-modified CMMS model unless there was a clinical reason to accept a lower threshold.

As the RSC provides a dataset that is representative of England, we assumed that the actual prevalence of each condition in the RSC dataset is similar to that in the GDPPR dataset, and therefore, the RCS dataset provided a suitable environment to develop the GDPPR-modified CMMS. We developed this GDPPR-modified CMMS in the RSC dataset because it offered a complete set of SNOMED clinical terms for each clinical concept, and we could replicate the reduced dataset within GDPPR and then compare the case finding with each approach.

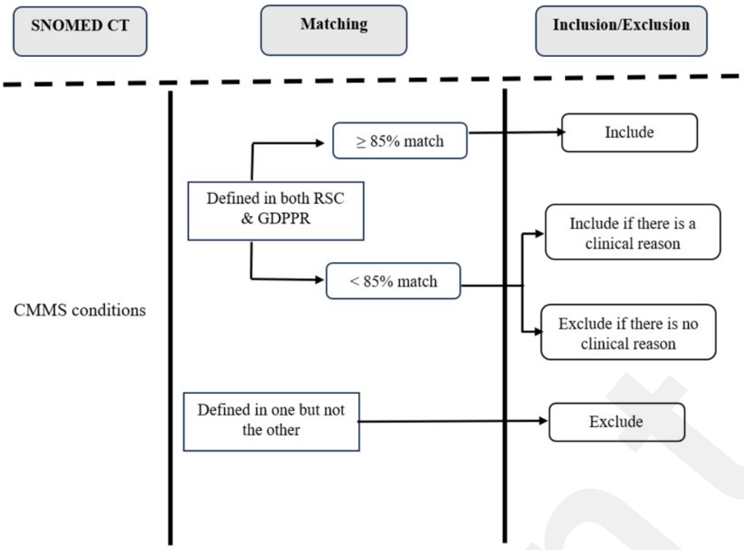

SNOMED CT
Matching
Inclusion/Exclusion
CMMS conditions
Defined in both RSC & GDPPR
≥ 85% match
< 85% match
Include
Include if there is a clinical reason
Exclude if there is no clinical reason
Defined in one but not the other
Exclude

Figure 2: Selection of candidate CMMS conditions for the GDPPR-modified CMMS from the 37 conditions included in the original CMMS and inclusion of conditions based on prevalence defined by both the SNOMED CT lists available in the GDPPR and RSC, or knowledge that the condition is known to have high predictive value

Statistical analyses

Using the previously described method [21], we used time-to-mortality Cox proportional hazards models based on the development data set. We first included the conditions as binary indicators with sex and age (in decades) and a quadratic age term as covariates. We then carried out a variable reduction process via backward elimination and using the Akaike information criteria (AIC) as the stopping criterion [32]. The goal was to be parsimonious with the number of variables needed for implementation. This was carried out using the 'fastbw' function from the rms R package. Model performance evaluation was based on discrimination and calibration. Discrimination was assessed by the pseudo R-squared, Somer's D and Harrell's C [33]. Model calibration was evaluated by plotting a calibration curve and recalibration was carried out by resampling using cross-validation to correct for optimism/overfitting. This was implemented using the 'calibrate' function in the rms R package. The model was developed, and performance was evaluated on the derivation dataset, and we then evaluated the performance of the models on the two validation datasets. All data preparation and analyses were conducted in R version 4.1.0[34], using the following R packages: lme4 (version 1.1-27) [35], lubridate (version 1.7.10)[36], [37]randomizr (version 0.20.0) [37], rms (version 6.2-0) [38], survival (version 3.2-11) [39, 40], tableone (version 0.12.0) [41] and tidyverse (version 1.3.1) [42].

Ethical considerations

This development of the CMMS for use in GDPPR was performed as part of the RAVEN study which received ethical approval (Integrate Research Application Service (IRAS) number 300259), and was approved by the Health Research Authority’s Bromley Research Ethics Committee (REC) reference 21/HRA/1971, on 8th October 2021. NHS England (NHSE) hosts the national safe haven for patient data.

The legal basis for this Regulation 3 of the Health Service (Control of Patient Information Regulations) 2002 [43]. Pseudonymised data extracted from the practices are kept in a secured server at ORCHID, which is an NHS England policy-compliant TRE (Organisation code EE133863-MSD-NDPCHS).

Results

Of the 37 conditions in the original CMMS model, 22 were included in developing the GDPPR-modified CMMS (Table 1). This involved five conditions based on clinical judgement, alcohol problems and chronic liver disease, and viral hepatitis, stroke and transient ischaemic attack, thyroid disorders and dementia which were likely to have a good predictive value, notwithstanding their lower matching proportion (33%, 49%, 81%, 83% and 84% respectively). All but thyroid disorders remained in the model after variable reduction in the final 17-condition model (Table 1).

Table 1: Inclusion of the original CMMS set of 37 medical conditions in the two modified versions of the CMMS, for use in the RSC and for use in the GDPPR, using the sets of SNOMED concepts available for the two respective datasets, ordered by prevalence of these conditions in the RSC dataset

Condition	No of patients in RSC dataset as defined by RSC SNOMED CT concepts (n=7,555,767)	Prevalence ratio (%)	Final RSC-modified model (n=21)	Initial GDPPR-modified model (n=22)	Final GDPPR-modified model* (n=17)
Diabetes	412960	100	x	x	x
Chronic kidney disease	260270	100	x	x	x
Chronis sinusitis	103638	100		x	
Bronchiectasis	34050	99.9		x	
Atrial fibrillation	198949	99.7	x	x	x
Asthma currently treated	542001	99.0		x	
COPD	147191	99.0	x	x	x
Schizophrenia or bipolar	48975	97.9	x	x	x

disease					
Epilepsy	54964	95.9	x	x	x
Parkinsonism	17156	95.8	x	x	x
Constipation	123999	95.7	x	x	x
Hypertension	1280958	94.7		x	
Learning disability	44100	93.4	x	x	x
Heart failure	95828	93.0	x	x	x
Cancer in the last 5 years	168577	91.1	x	x	x
Peripheral vascular disease	38946	87.3	x	x	x
Coronary heart disease	322338	84.7		x	x
Dementia	80156	84.2	x	x	x
Thyroid disorders	420681	83.3		x	
Stroke and transient ischaemic attack	182979	81.0		x	x
Migraine	33719	68.4			
Connective tissue disorder/ Rheumatoid arthritis	153350	51.8			
Chronic liver disease and viral hepatitis	52265	49.3	x	x	x
Painful conditions	810458	43.6	x		
Alcohol problems	190439	33.0	x	x	x
Psoriasis or eczema	63556	26.5			
Inflammatory bowel disease	50370	25.6			
Anxiety or depression	919962	25.1	x		
Disorders of prostate	201559	25.0	x		
Blindness and low vision	88398	17.5			
Diverticular disease of intestine	229536	17.2			
Anorexia	52742	16.7			
Hearing loss	581804	12.9			
Peptic ulcer disease	98487	9.5			
Irritable bowel syndrome	407880	0	x		
Psychoactive substance misuse	92944	0	x		
Multiple sclerosis	15717	0	x		

* after variable reduction

There were few differences in the development and validation datasets in terms of age, sex, number of conditions and follow-time time (Table 2).

Table 2. Descriptive statistics of three datasets sampled from the RSC for deriving and validating a modified version of the CMMS for use in the GDPPR dataset, within the constraints of its limited set of SNOMED CT codes.

	Derivation (2019)	**Validation 1 (2019)**	**Validation 2 (2015)**
Male	247,807 (49.6%)	124,514 (49.8%)	123,541 (49.4%)

Age at index date in years			
mean(SD)	49.03 (1929)	48.11 (19.09)	46.0 (19.34)
Range	16-95	16-95	16-95
65-84 years	103,587 (22%)	48,016 (0.21)	46,834 (0.20)
85 or more years	16,387 (4%)	7,513 (0.03)	7,608 (0.03)
No of conditions			
Mean (SD)	0.72 (1.19)	0.69 (1.17)	0.70 (1.0)
Range	0 to 11	0 to 11	0 to 11
0	309,089 (62%)	157,981 (63%)	161,941 (65%)
1	101,547 (20%)	49,463 (20%)	47,530 (19%)
2 or more	89,364 (18%)	42,556 (17%)	40,529 (16%)
No of deaths in follow-up	5,104	2,392	2,408/11,948
Mean follow-up time† (days)	352.8	351.9	350.4/1,538.0
Total person years*†	482,885.5	240,859.6	239,859.2/1,052,567
Mortality rate (per 1000 person years)†	10.57	9.93	10.04 /11.35

* Calculated as number of person-days divided by 365.25; † 1-year follow-up for Validation 1 and 1- and 5-year follow-up for Validation 2.

The prevalence of the 22 conditions in the model derivation dataset is presented in Table 3. These prevalences and their rankings were generally similar to those reported in the original CMMS study [17], and our previous study [21]. There was only one exception, coronary heart disease, which ranked lower (1.4%; original CMMS study 4.8% and previous RSC study 5.3%).

Table 3. Prevalence in individuals in the RCS data set of the 22 candidate conditions in the GDPPR-modified CMMS model before variable reduction, and weights for the final set of 17 conditions after variable reduction, with conditions ordered by prevalence in the RSC dataset

Condition	**Number (%)**	**Weight**
Hypertension	98849 (19.8)	
Asthma currently treated	36951 (7.4)	
Diabetes	33312 (6.7)	0.2623
Thyroid disorders	28891 (5.8)	
Chronic kidney disease	23145 (4.6)	0.1286
Atrial fibrillation	15041 (3.0)	0.2779
Cancer in the last 5 years	13059 (2.6)	1.1876
COPD	12734 (2.5)	0.6638
Alcohol problems	12132 (2.4)	0.5670
Stroke and transient ischaemic attack	12118 (2.4)	0.2299
Constipation	8698 (1.7)	0.5889
Chronis sinusitis	8195 (1.6)	
Coronary heart disease	21897 (1.4)	0.1201
Heart failure	7163 (1.4)	0.5022
Dementia	5884 (1.2)	0.9815
Epilepsy	4114 (0.8)	0.6714
Schizophrenia or bipolar disorder	3819 (0.8)	0.5621
Learning disability	2857 (0.6)	1.0992

Peripheral vascular disease	2963 (0.6)	0.3519
Bronchiectasis	2563 (0.5)	
Chronic liver disease and viral hepatitis	1890 (0.4)	1.0844
Parkinsonism	1409 (0.3)	0.5339

The model performance of the 22-condition and 17-condition models was almost identical and similar to those in the original CMMS study and the RSC-modified CMMS (Tables 4 and 5).

Table 4. Model discrimination for the final RSC-modified and GDPPR-modified versions of the CMMS, after variable reduction, and compared with a full 37-condition model using RSC data and SNOMED CT.

	37-condition model [15]	**Final RSC-modified 21-condition model [15]**	**Final GDPPR-modified 17-condition model**
Pseudo R-squared	0.153	0.153	0.140
Somers' D	0.851	0.851	0.833
Harrell's C*			
Derivation	0.925 (0.002)	0.926 (0.002)	0.916 (0.002)
Validation 1	0.920 (0.004)	0.921 (0.004)	0.922 (0.003)
Validation 2, 1-year follow-up	0.920 (0.003)	0.920 (0.003)	0.915 (0.003)
Validation 2, 5-year follow-up	0.907 (0.002)	0.907 (0.002)	0.902 (0.001)

* Reported with standard errors

Table 5. Hazard ratios (95 CIs) of the predictors for the final RSC-modified and GDPPR-modified versions of the CMMS, after variable reduction, and compared with a full 37-condition model using RSC data and SNOMED CT.

	37-condition model HR (95% CI) [15]			**RSC-modified 21-condition model HR (95% CI) [15]**			**GDPPR-modified 17-condition model HR (95% CI)**		
Age (10 years)	1.22	1.02	1.47				1.02	1.01	1.04
[Age_(10 years)]^2	1.05	1.03	1.06	1.06	1.06	1.06	1.00	1.00	1.00
Sex=M	1.33	1.23	1.45	1.34	1.24	1.46	1.14	1.08	1.21
Cancer in the last 5 years	3.31	2.99	3.67	3.33	3.00	3.69	3.28	3.06	3.52
Dementia	2.57	2.33	2.84	2.55	2.32	2.82	2.67	2.47	2.88
Alcohol problems	2.17	1.84	2.55	2.21	1.88	2.60	1.76	1.53	2.03
Multiple sclerosis	2.13	1.32	3.44	2.14	1.33	3.46			
Chronic liver disease and viral hepatitis	1.98	1.57	2.49	1.99	1.58	2.50	2.96	2.38	3.68
COPD	1.96	1.76	2.18	2.02	1.83	2.23	1.94	1.80	2.10
Learning disability	1.88	1.14	3.10	1.89	1.15	3.11	3.00	2.18	4.13
Parkinsonism	1.71	1.39	2.11	1.73	1.40	2.13	1.71	1.43	2.04
Heart failure	1.66	1.49	1.85	1.66	1.49	1.84	1.65	1.51	1.80
Epilepsy	1.59	1.25	2.02	1.61	1.27	2.04	1.96	1.62	2.37
Schizophrenia or bipolar disorder	1.59	1.22	2.06	1.62	1.25	2.10	1.75	1.42	2.17
Psychoactive substance abuse	1.57	1.20	2.04	1.57	1.20	2.04			
Painful condition	1.55	1.42	1.68	1.56	1.44	1.69			
Constipation	1.47	1.33	1.62	1.47	1.33	1.62	1.80	1.67	1.95
Atrial fibrillation	1.39	1.27	1.53	1.40	1.27	1.53	1.32	1.23	1.42
Peripheral vascular disease	1.39	1.07	1.81	1.40	1.08	1.82	1.42	1.24	1.63

Anxiety or depression	1.38	1.27	1.50	1.38	1.27	1.50			
Diabetes	1.31	1.20	1.43	1.34	1.23	1.46	1.30	1.21	1.39
Psoriasis or eczema	1.27	1.03	1.57						
Chronic kidney disease	1.24	1.14	1.35	1.24	1.14	1.35	1.14	1.06	1.21
Anorexia or bulimia	1.22	0.66	2.28						
Peptic ulcer	1.13	0.98	1.30						
Bronchiectasis	1.11	0.87	1.41						
Stroke and transient ischaemic attack	1.11	1.00	1.24				1.26	1.16	1.36
Asthma currently treated	1.05	0.93	1.18						
Hypertension	1.04	0.96	1.13						
Thyroid disorder	1.03	0.92	1.14						
Coronary heart disease	1.00	0.91	1.09				1.13	1.05	1.21
Chronic sinusitis	0.98	1.57	2.49						
Rheumatoid arthritis	0.98	0.85	1.12						
Blindness and low vision	0.96	0.84	1.11						
Diverticular disease of intestine	0.92	0.82	1.02						
Hearing loss	0.92	0.95	1.00						
Disorder of the prostate	0.83	0.74	0.93	0.83	0.74	0.93			
Irritable bowel syndrome	0.83	0.71	0.95	0.82	0.71	0.94			
Inflammatory bowel disease	0.65	0.43	0.97						
Migraine	0.59	0.25	1.42						

For 1-year mortality, discrimination was high in both the derivation and validation datasets (Harrell's C=0.92) and for 5-year mortality, it was slightly lower (Harrell's C=0.90). The model calibration displayed underprediction at lower risks (<60%), and the calibration improved with the adjustment for optimism/over-fitting (Figure 3).

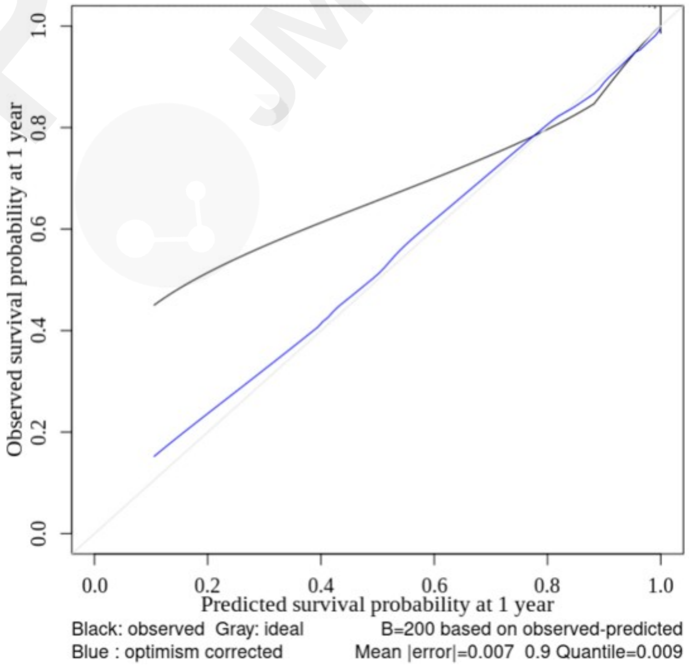


Figure 3. Calibration curve for the final 17-condition GDPPR-modified CMMS model, with and without correction for optimism/over-fitting

Discussion

Principal Findings

In this study, we developed and validated a modified version of a single measure of multimorbidity, CMMS, for use within a national dataset created during the pandemic from all existing primary care data collections, GDPPR and using its limited set of SNOMED CT terms. The initial model included 22 conditions from the set of 37 in the original CMMS model, and the reduced 17-condition model showed an identical performance in predicting mortality to the 22-condition model and a similar performance compared to the original 37-condition CMMS and the previous modification which was based on an extensive SNOMED CT dataset.

Interpretations and Implications

The GDPPR database remains available and is listed in the NHS Data Model and Dictionary [44]. It is now one of the data collections available through the NHS England Secure Data Environment (SDE) [45]. SDE was created as part of NHS England’s Data Saves Lives policy, following the Goldacre report [46].

This new single measure of multimorbidity will help us measure vaccine effectiveness using GDPPR, NHS England's nationwide database. We will use this GDPPR-modified CMMS score in the RAVEN study to build on the existing evidence base [47, 48]. Several observational studies have shown that the effectiveness of vaccination could be suboptimal in people with multimorbidity [49], and thus it is important to be able to explore and adjust for multimorbidity.

This tool may also be useful in a wider range of studies of people with multimorbidity, including vaccine and post-authorisation safety studies. The risk of hospitalisation, admission to intensive care unit (ICU) beds, and mortality in people with multimorbidity are significantly higher than in the general population [50, 51]. In an observational study of hospitalised patients in the UK with COVID-19, the crude mortality in people with multimorbidity, compared to single comorbidity, after adjusting for the demographic factors, was more than double (37.2% versus 17.3%) [52] It was estimated to reduce 63.5% of deaths by prioritising people with multimorbidity [53]. Therefore, people with multimorbidity were prioritised for vaccine rollout [54].

Our previous study showed that reducing the original CMMS variables from 37 to 21 did not compromise mortality predictability in people with multimorbidity [21]. In the current study, we have demonstrated that the number of conditions can be reduced to 17 to match the data available in GDPPR and still can be a very good predictor of mortality.

Comparison with Prior Work

We focused on mortality and this is the outcome most often reported in development studies of comorbidity indices [55]. There are many published comorbidity indices and they vary according to their time of development (and thus the number of modifications), derivation population, conditions (predictors), prediction horizon, outcome predicted and data source [52]. Historically, the mortality indices have been designed for people in hospital, using secondary care coding systems and they have provided predictions of in-hospital mortality and mortality between 6 months and 5 years [55]. The GDPPR-modified CMMS is the latest adaptation to the original CMMS [17] for predicting mortality in primary care. The predictions for the CMMS and its modified versions are for the same prediction horizons of 1 year and 5 years. The previous modification adapted the CMMS to conditions defined by the internationally recognised SNOMED CT coding system [21], as the original CMMS was based on a population in the US and conditions defined by the Read clinical terminology, which is no longer used in England. Both modifications of the CMMS have been developed on English populations with a lower minimum age compared to that for the original version (16 years as opposed to 21 years).

This GDPPR-modified version produces a multimorbidity index for mortality based on conditions defined by the limited SNOMED list of the GDPPR. The RSC provided the development dataset for both modifications of the CMMS. The RCS has a complete set of SNOMED codes. Its dataset includes people registered in a fraction of English general practices, and the assumption of the current study is that the underlying prevalence of each CMMS condition in people in the RSC dataset is the same as those in the larger GDPPR dataset. Whilst the RSC is recruited to be nationally representative, there may inevitably be differences [24].

The GDPPR includes the English primary care data used by the British Heart Foundation's Data Science Centre for their COVID-19 and cardiovascular diseases (CVD-COVID-UK) Consortium's work. Our version of CMMS could be deployed by this and other groups using GDPPR or is underlying primary care (GPES) data[56].

Strengths

The main strength of this study is that it built on our expertise in developing a version of CMMS that could be applied to routine clinical data recorded using SNOMED CT. This terminology is used internationally [21]. We overcame the limitations of the relatively limited number of SNOMED clinical terms in GDPPR and demonstrated that a 17-condition CMMS could run in the GDDPR dataset of 54 million individuals.

Limitations

There are several limitations of this study. We only predicted the mortality risk and did not predict the hospitalisation or ICU admission risk. The conditions were a subset of those included in the original CMMS study, which arose from a review of multimorbidity literature at the time of its development. The nature of disease and treatments, as to population characteristics, will change over time. Hence the new CMMS versions will need to be updated regularly, and this may involve adding conditions that were not included in the original CMMS. Although we split our initial dataset randomly into development and validation datasets, we have performed simple temporal external validation [57]. A more robust form of external validation would involve investigating the generalisability to other countries (geographical validation) or other settings (domain validation), but neither is relevant in this case as we are considering a national dataset, and we have developed the new CMMS using what we assume to be a representative sample of the adult English population. The generalizability could be tested further on other primary care data such as CPRD [18], which provides a database of anonymised health records for another sample of English GP practices.

Conclusions

This latest modification of the CMMS provides a new validated single multimorbidity measure, which was generated through a combination of unique access to data and expertise in validation. The RSC provided nationally representative and comprehensive primary care data. The study team had experience in developing a validated CMMS version to use within SNOMED CT. This combination meant that it was possible to develop and validate a new version of CMMS for use in the national English dataset, the GDPPR. Our previous study showed that reducing the original CMMS variables from 37 to 21 did not compromise mortality predictability in people with multimorbidity [21]. In this study, we have demonstrated that the number of conditions can be reduced to 17 to match the data available in GDPPR and still can be a very good predictor of mortality. Therefore, researchers using this national database, or looking for a further reduced CMMS measure, can utilise this 17-component single measure of comorbidity.

The approach used in this study could also be applied in other contexts. Our approach has been to replicate a validated multimorbidity measure in a smaller, but complete and high data quality sentinel network database, the RSC. Within the RSC we could ensure the model performs as well as one run on complete data [21]. Additionally, developing and validating this reduced CMMS model in the RSC required less processing time. This may make this reduced version more attractive to other users, should processing time be at a premium.

Acknowledgements

We would like to acknowledge that Ruby Tsang prepared the R script that we modified and Julian Sherlock was also involved in preparing the data.

This study was conducted to support the RAVEN study which was funded by AstraZeneca. The sponsors were involved in the SNOMED matching process (SV), and in reviewing and approving the manuscript (SV, WM, and ST).

Data availability statement

The data sets generated during and analyzed during this study are not publicly available. However, the Oxford–RCGP RSC dataset can be accessed by researchers; approval is on a project-by-project basis (https://orchid.phc.ox.ac.uk/index.php/rcgp-rsc). Ethical approval by an NHS Research Ethics Committee/other appropriate approval is needed before any data release. Researchers wishing to directly analyse patient-level pseudonymised data will be required to complete information governance training and work on the data from the secure servers at the University of Oxford. Patient-level data cannot be taken out of the secure network.

The SNOMED CT lists for the 22 candidate variables are available upon request.

Author contributions

SdeL conceived this research project and played a supervisory role. XF was involved in extracting data from the RSC. DK and RW curated the variables. DK compared the SNOMED code lists and DK and SdeL conducted the clinical sign-off of conditions to be included in the analysis. KT, SV, and MJ were involved in the statistical analyses. SA and FF were responsible for managing the project. DK and KT drafted the manuscript. All authors have reviewed and approved the final manuscript.

Conflicts of interest

SdeL is the principal investigator for RAVEN (EUPAS43571) funded by AstraZeneca. SdeL is also the Director of the RCGP RSC, which is included in his academic role at the University of Oxford. He has received research funding through his University from AstraZeneca, GSK, Lily, Moderna, MSD, Sanofi, Seqirus, and Takeda. He has also served as an advisory board member for AstraZeneca, GSK, Sanofi, Seqirus, and Pfizer.

Abbreviations

AIC: Akaike information criteria

CMMS: Cambridge Multimorbidity Score

GDPPR: General Practice Extraction Service Data for Pandemic Planning and Research

CPRD: Clinical Practice Research Datalink

LTC: Long-term conditions

ORCHID: Oxford Royal College of General Practitioners Digital Informatics Hub

RAVEN: Real-world effectiveness of the Oxford-AstraZeneca Covid-19 vaccine in England

RCGP RSC: Oxford-Royal College of General Practitioners Research and Surveillance Centre

RSC: Research and Surveillance Centre

SNOMED CT: Systematised nomenclature of medicine clinical terms

TRE: Trusted research environment

REFERENCES

1. Aubert CE, Schnipper JL, Roumet M, Marques-Vidal P, Stirnemann J, Auerbach AD, et al., Best definitions of multimorbidity to identify patients with high health care resource utilization. Mayo Clin Proc Innov Qual Outcomes, 2020; 4(1): 40-49.DOI: 10.1016/j.mayocpiqo.2019.09.002.
2. Dambha-Miller H, Simpson G, Hobson L, Roderick P, Little P, Everitt H, et al., Integrated primary care and social services for older adults with multimorbidity in England: a scoping review. BMC Geriatrics, 2021; 21(1): 674.DOI: 10.1186/s12877-021-02618-8.
3. Hanlon P, Nicholl BI, Jani BD, Lee D, McQueenie R, and Mair FS, Frailty and pre-frailty in middle-aged and older adults and its association with multimorbidity and mortality: a prospective analysis of 493 737 UK Biobank participants. Lancet Public Health, 2018; 3(7): e323-e332.DOI: 10.1016/s2468-2667(18)30091-4.
4. Smith SM, Wallace E, O'Dowd T, and Fortin M, Interventions for improving outcomes in patients with multimorbidity in primary care and community settings. Cochrane Database Syst Rev, 2016; 3(3): Cd006560.DOI: 10.1002/14651858.CD006560.pub3.
5. Pearson-Stuttard J, Ezzati M, and Gregg EW, Multimorbidity-a defining challenge for health systems. Lancet Public Health, 2019; 4(12): e599-e600.DOI: 10.1016/s2468-2667(19)30222-1.
6. Cassell A, Edwards D, Harshfield A, Rhodes K, Brimicombe J, Payne R, et al., The epidemiology of multimorbidity in primary care: a retrospective cohort study. Br J Gen Pract, 2018; 68(669): e245-e251.DOI: 10.3399/bjgp18X695465.
7. Chudasama YV, Khunti K, Gillies CL, Dhalwani NN, Davies MJ, Yates T, et al., Healthy lifestyle and life expectancy in people with multimorbidity in the UK Biobank: A longitudinal cohort study. PLoS Med, 2020; 17(9): e1003332.DOI: 10.1371/journal.pmed.1003332.
8. Carretero-Bravo J, Ramos-Fiol B, Ortega-Martín E, Suárez-Lledó V, Salazar A, O'Ferrall-González C, et al., Multimorbidity patterns and their association with social determinants, mental and physical health during the COVID-19 pandemic. Int J Environ Res Public Health, 2022; 19(24).DOI: 10.3390/ijerph192416839.
9. Chudasama YV, Zaccardi F, Gillies CL, Razieh C, Yates T, Kloecker DE, et al., Patterns of

multimorbidity and risk of severe SARS-CoV-2 infection: an observational study in the U.K. BMC Infectious Diseases, 2021; 21(1): 908.DOI: 10.1186/s12879-021-06600-y.

10. Graña C, Ghosn L, Evrenoglou T, Jarde A, Minozzi S, Bergman H, et al., Efficacy and safety of COVID-19 vaccines. Cochrane Database Syst Rev, 2022; 12(12): Cd015477.DOI: 10.1002/14651858.Cd015477.
11. Knoll MD and Wonodi C, Oxford-AstraZeneca COVID-19 vaccine efficacy. Lancet, 2021; 397(10269): 72-74.DOI: 10.1016/s0140-6736(20)32623-4.
12. Agrawal U, Bedston S, McCowan C, Oke J, Patterson L, Robertson C, et al., Severe COVID-19 outcomes after full vaccination of primary schedule and initial boosters: pooled analysis of national prospective cohort studies of 30 million individuals in England, Northern Ireland, Scotland, and Wales. The Lancet, 2022; 400(10360): 1305-1320.DOI: https://doi.org/10.1016/S0140-6736(22)01656-7.
13. Lai FTT, Huang L, Chui CSL, Wan EYF, Li X, Wong CKH, et al., Multimorbidity and adverse events of special interest associated with Covid-19 vaccines in Hong Kong. Nat Commun, 2022; 13(1): 411.DOI: 10.1038/s41467-022-28068-3.
14. Ho IS, Azcoaga-Lorenzo A, Akbari A, Black C, Davies J, Hodgins P, et al., Examining variation in the measurement of multimorbidity in research: a systematic review of 566 studies. Lancet Public Health, 2021; 6(8): e587-e597.DOI: 10.1016/s2468-2667(21)00107-9.
15. Fraccaro P, Kontopantelis E, Sperrin M, Peek N, Mallen C, Urban P, et al., Predicting mortality from change-over-time in the Charlson Comorbidity Index: A retrospective cohort study in a data-intensive UK health system. Medicine (Baltimore), 2016; 95(43): e4973.DOI: 10.1097/md.0000000000004973.
16. Drosdowsky A and Gough K, The Charlson Comorbidity Index: problems with use in epidemiological research. J Clin Epidemiol, 2022; 148: 174-177.DOI: 10.1016/j.jclinepi.2022.03.022.
17. Payne RA, Mendonca SC, Elliott MN, Saunders CL, Edwards DA, Marshall M, et al., Development and validation of the Cambridge Multimorbidity Score. Cmaj, 2020; 192(5): E107-e114.DOI: 10.1503/cmaj.190757.
18. CPRD: UK data driving real-world evidence. [Accessed on July 4, 2023]; Available from: https://www.cprd.com/.
19. Retirement of Read Version 2 and Clinical Terms Version 3. . [Accessed on July 4, 2023]; Available from: https://digital.nhs.uk/services/terminology-and-classifications/read-codes.
20. SNOMED CT. [Accessed on July 4, 2023]; Available from: https://digital.nhs.uk/services/terminology-and-classifications/snomed-ct.
21. Tsang RS, Joy M, Whitaker H, Sheppard JP, Williams J, Sherlock J, et al., Development of a modified Cambridge Multimorbidity Score for use with SNOMED CT: an observational English primary care sentinel network study. Br J Gen Pract, 2023; 73(731): e435-e442.DOI: 10.3399/bjgp.2022.0235.
22. Correa A, Hinton W, McGovern A, van Vlymen J, Yonova I, Jones S, et al., Royal College of General Practitioners Research and Surveillance Centre (RCGP RSC) sentinel network: a cohort profile. BMJ Open, 2016; 6(4): e011092.DOI: 10.1136/bmjopen-2016-011092.
23. de Lusignan S, Correa A, Smith GE, Yonova I, Pebody R, Ferreira F, et al., RCGP Research and Surveillance Centre: 50 years' surveillance of influenza, infections, and respiratory conditions. Br J Gen Pract, 2017; 67(663): 440-441.DOI: 10.3399/bjgp17X692645.
24. Leston M, Elson WH, Watson C, Lakhani A, Aspden C, Bankhead CR, et al., Representativeness, vaccination uptake, and COVID-19 clinical outcomes 2020-2021 in the UK Oxford-Royal College of General Practitioners Research and Surveillance Network: Cohort

profile summary. JMIR Public Health Surveill, 2022; 8(12): e39141.DOI: 10.2196/39141.

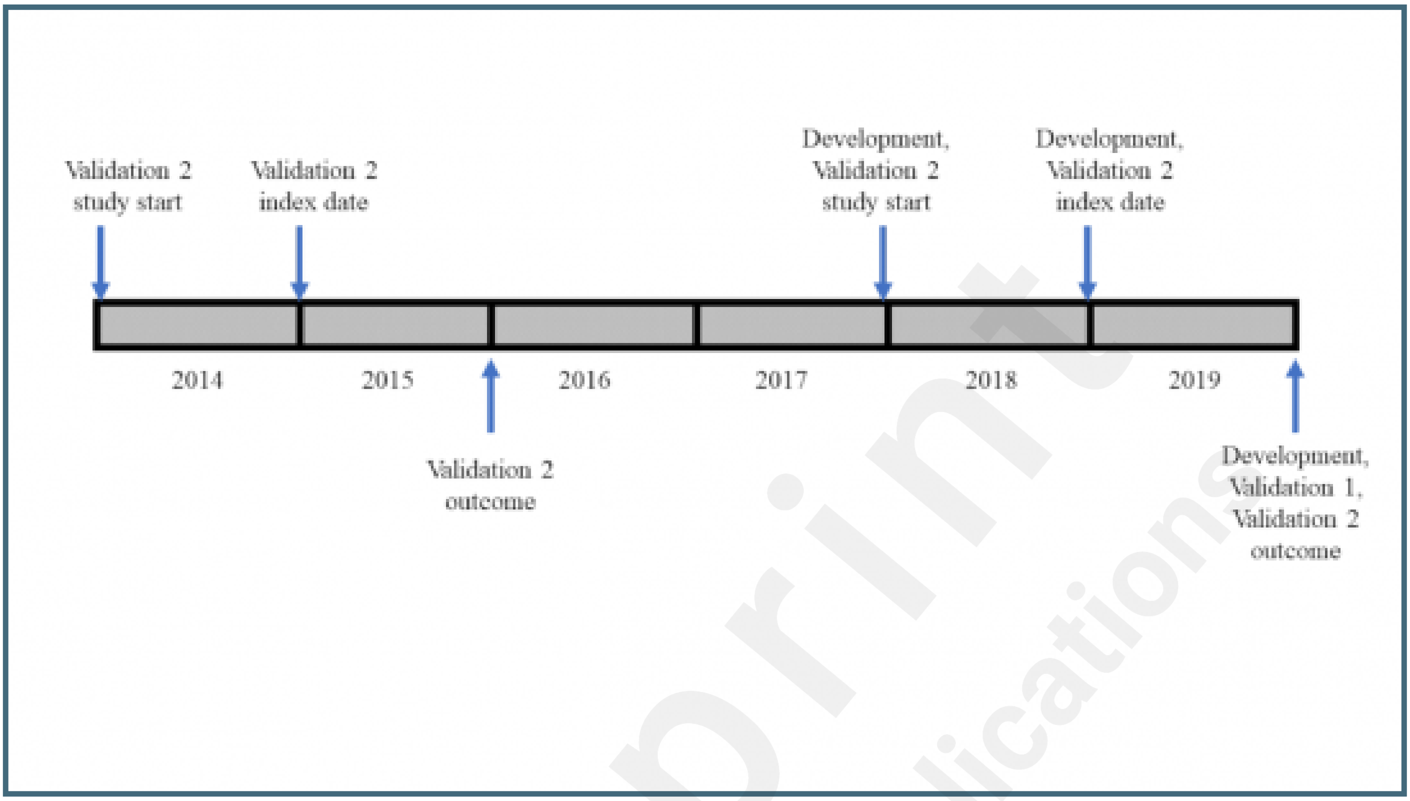
25. de Lusignan S, Dorward J, Correa A, Jones N, Akinyemi O, Amirthalingam G, et al., Risk factors for SARS-CoV-2 among patients in the Oxford Royal College of General Practitioners Research and Surveillance Centre primary care network: a cross-sectional study. Lancet Infect Dis, 2020; 20(9): 1034-1042.DOI: 10.1016/s1473-3099(20)30371-6.
26. Whitaker HJ, Tsang RSM, Byford R, Andrews NJ, Sherlock J, Sebastian Pillai P, et al., Pfizer-BioNTech and Oxford AstraZeneca COVID-19 vaccine effectiveness and immune response amongst individuals in clinical risk groups. J Infect, 2022; 84(5): 675-683.DOI: 10.1016/j.jinf.2021.12.044.
27. Tsang RS, Joy M, Byford R, Robertson C, Anand SN, Hinton W, et al., Adverse events following first and second dose COVID-19 vaccination in England, October 2020 to September 2021: a national vaccine surveillance platform self-controlled case series study. Euro Surveill, 2023; 28(3).DOI: 10.2807/1560-7917.Es.2023.28.3.2200195.
28. Meeraus W, Joy M, Ouwens M, Taylor KS, Venkatesan S, Dennis J, et al., AZD1222 effectiveness against severe COVID-19 in individuals with comorbidity or frailty: The RAVEN cohort study. J Infect, 2024; 88(4): 106129.DOI: 10.1016/j.jinf.2024.106129.
29. COVID-19 General Practice Extraction Service (GPES) Data for Pandemic Planning and Research (GDPPR) [Accessed on July 4, 2023]; Available from: https://digital.nhs.uk/services/data-access-request-service-dars/dars-products-and-services/data-set-catalogue/gpes-data-for-pandemic-planning-and-research-gdppr.
30. de Lusignan S and van Weel C, The use of routinely collected computer data for research in primary care: opportunities and challenges. Fam Pract, 2006; 23(2): 253-63.DOI: 10.1093/fampra/cmi106.
31. de Lusignan S, Jones N, Dorward J, Byford R, Liyanage H, Briggs J, et al., The Oxford Royal College of General Practitioners Clinical Informatics Digital Hub: Protocol to develop extended COVID-19 surveillance and trial platforms. JMIR Public Health Surveill, 2020; 6(3): e19773.DOI: 10.2196/19773.
32. Akaike H, A new look at the statistical model identification. IEEE Transactions on Automatic Control, 1974; 19(6): 716-723.DOI: 10.1109/TAC.1974.1100705.
33. Harrell FE, Jr., Califf RM, Pryor DB, Lee KL, and Rosati RA, Evaluating the yield of medical tests. Jama, 1982; 247(18): 2543-6.
34. R Core Team, R: A language and environment for statistical computing, Vienna, R Foundation for Statistical Computing Available from: https://www.R-project.org/
35. Bates D, Mächler M, Bolker B, and Walker S, Fitting linear mixed-effects models using lme4. Journal of Statistical Software, 2015; 67(1): 1 - 48.DOI: 10.18637/jss.v067.i01.
36. Grolemund G and Wickham H, Dates and times made easy with lubridate. Journal of Statistical Software, 2011; 40(3): 1 - 25.DOI: 10.18637/jss.v040.i03.
37. Coppock A, Cooper J. Randomizr: easy-to-use tools for common forms of random assignment and sampling. R package, version 0.20.0 2019. Available from: https://CRAN.R-project.org/package=randomizr
38. Harrell FE, Regression modeling strategies. R package, version 6.2-0 2021. Available from: https://CRAN.R-project.org/package=rms
39. Therneau T, A package for survival analysis in S. R package, version 3.2-11 2021.
40. Therneau TM, Grambsch PM. Modelling survival data: Extending the Cox model. 2000, New York: Springer.
41. Yoshida K, Bartel A. Create 'Table 1' to describe baseline characteristics with or without propensity score weights. R package, version 0.12.0 2020. Available from: https://CRAN.R-

project.org/package=tableone
42. Wickham H AM, Bryan J, et al, Welcome to the tidyverse. Journal of Open Source Software, 2019; 4: 1686.DOI: 10.21105/joss.01686.
43. Taylor MJ, Legal bases for disclosing confidential patient information for public heath: Distinguishing between health protection and health improvement. Med Law Rev, 2015; 23(3): 348-74.DOI: 10.1093/medlaw/fwv018.
44. NHS data model and dictionary. [Accessed on July 4, 2023]; Available from: https://www.datadictionary.nhs.uk/GPDataSet/index.html.
45. NHS Digital. The NHS England secure data environment. [Accessed on July 4, 2023]; Available from: https://digital.nhs.uk/services/secure-data-environment-servic.
46. de Lusignan S, Leston M, Ikpoh M, and Howsam G, Data saves lives: bottom-up, professionally-led endorsement would increase the chance of success. Br J Gen Pract, 2022; 72(724): 512-513.DOI: 10.3399/bjgp22X720965.
47. Rotshild V, Hirsh-Raccah B, Miskin I, Muszkat M, and Matok I, Comparing the clinical efficacy of COVID-19 vaccines: a systematic review and network meta-analysis. Scientific Reports, 2021; 11(1): 22777.DOI: 10.1038/s41598-021-02321-z.
48. Mills EJ and Reis G, Evaluating COVID-19 vaccines in the real world. Lancet, 2022; 399(10331): 1205-1206.DOI: 10.1016/s0140-6736(22)00194-5.
49. Choi WS and Cheong HJ, COVID-19 vaccination for people with comorbidities. Infect Chemother, 2021; 53(1): 155-158.DOI: 10.3947/ic.2021.0302.
50. Russell CD, Lone NI, and Baillie JK, Comorbidities, multimorbidity and COVID-19. Nature Medicine, 2023; 29(2): 334-343.DOI: 10.1038/s41591-022-02156-9.
51. McQueenie R, Foster HME, Jani BD, Katikireddi SV, Sattar N, Pell JP, et al., Multimorbidity, polypharmacy, and COVID-19 infection within the UK Biobank cohort. PLoS One, 2020; 15(8): e0238091.DOI: 10.1371/journal.pone.0238091.
52. Agrawal U, Azcoaga-Lorenzo A, Fagbamigbe AF, Vasileiou E, Henery P, Simpson CR, et al., Association between multimorbidity and mortality in a cohort of patients admitted to hospital with COVID-19 in Scotland. Journal of the Royal Society of Medicine, 2022; 115(1): 22-30.DOI: 10.1177/01410768211051715.
53. Ioannou GN, Green P, Fan VS, Dominitz JA, O'Hare AM, Backus LI, et al., Development of COVIDVax Model to Estimate the Risk of SARS-CoV-2-Related Death Among 7.6 Million US Veterans for Use in Vaccination Prioritization. JAMA Netw Open, 2021; 4(4): e214347.DOI: 10.1001/jamanetworkopen.2021.4347.
54. Russo AG, Decarli A, and Valsecchi MG, Strategy to identify priority groups for COVID-19 vaccination: A population based cohort study. Vaccine, 2021; 39(18): 2517-2525.DOI: 10.1016/j.vaccine.2021.03.076.
55. Yurkovich M, Avina-Zubieta JA, Thomas J, Gorenchtein M, and Lacaille D, A systematic review identifies valid comorbidity indices derived from administrative health data. Journal of Clinical Epidemiology, 2015; 68(1): 3-14.DOI: https://doi.org/10.1016/j.jclinepi.2014.09.010.
56. Abbasizanjani H, Torabi F, Bedston S, Bolton T, Davies G, Denaxas S, et al., Harmonising electronic health records for reproducible research: challenges, solutions and recommendations from a UK-wide COVID-19 research collaboration. BMC Med Inform Decis Mak, 2023; 23(1): 8.DOI: 10.1186/s12911-022-02093-0.
57. Moons KG, Kengne AP, Grobbee DE, Royston P, Vergouwe Y, Altman DG, et al., Risk prediction models: II. External validation, model updating, and impact assessment. Heart, 2012; 98(9): 691-8.DOI: 10.1136/heartjnl-2011-301247.

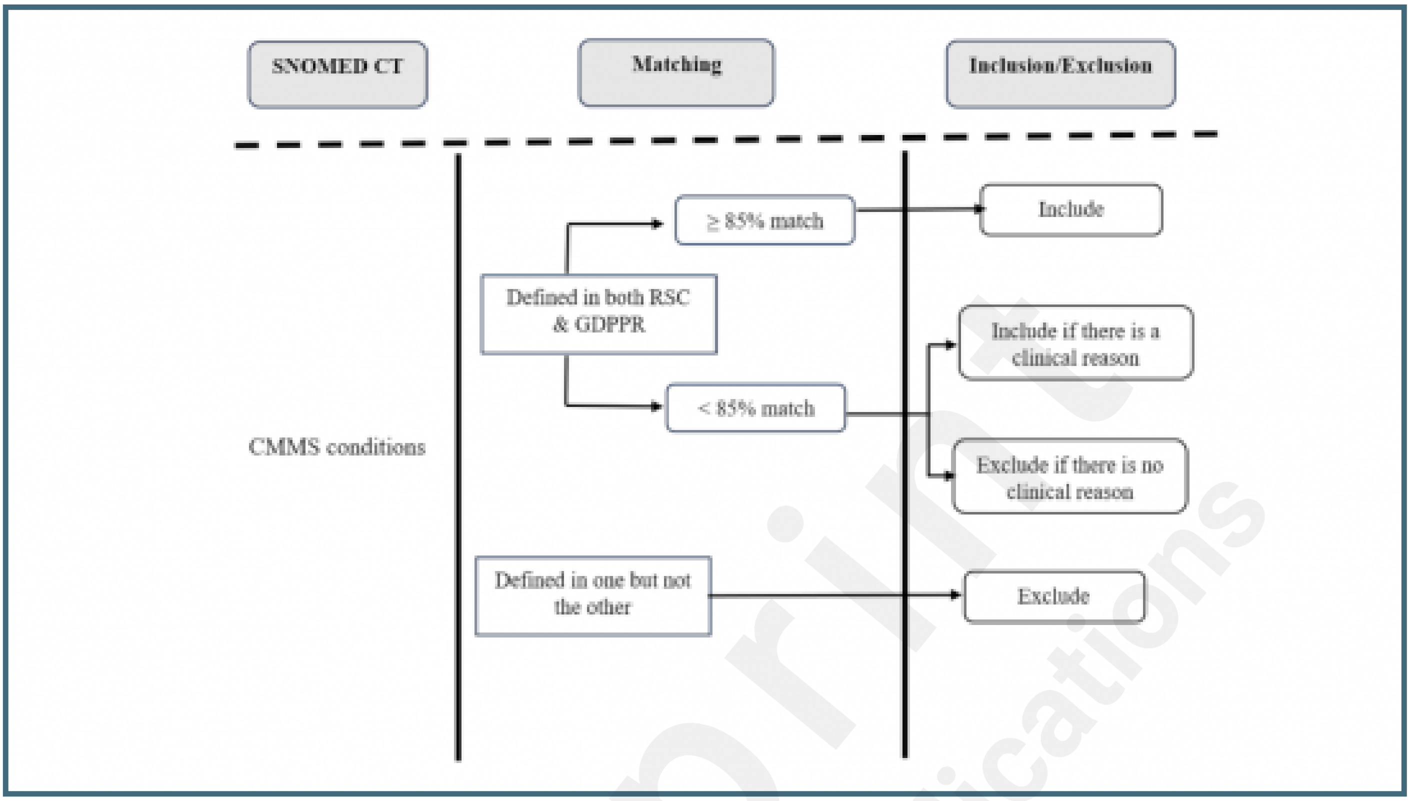
Supplementary Files

Figures

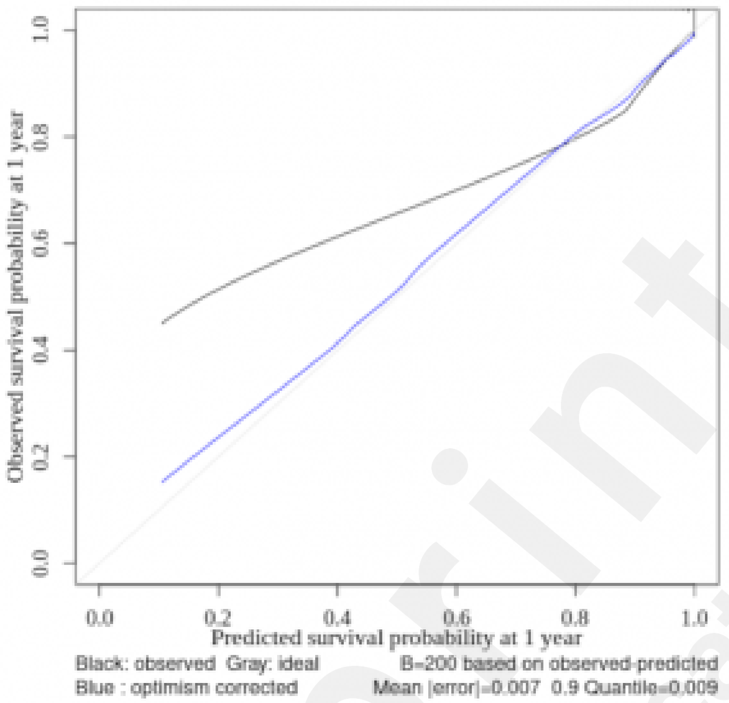
Study design for the development and validation model (Reproduced with permission - Tsang et al, 2022) [15].



Selection of candidate CMMS conditions from the 37 conditions.


SNOMED CT
Matching
Inclusion/Exclusion
CMMS conditions
≥ 85% match
Defined in both RSC & GDPPR
< 85% match
Defined in one but not the other
Include
Include if there is a clinical reason
Exclude if there is no clinical reason
Exclude

Calibration curve for the 17-condition GDPPR-modified CMMS model.



Multimedia Appendixes

Supplementary material.
URL: http://asset.jmir.pub/assets/45c103ba48b87b4f2ca5f9693f7a46de.docx