Supplementary Appendix S2
Guide to implementing the ORCHID-CMMS using SNOMED clinical terms (CT), the Dictionary of medicines and devices (dm+d), and local codes.

Background:
This SNOMED CT version of CMMS was developed using the Oxford-Royal College of General Practitioners (RCGP) Research and Surveillance Centre (RSC) data. The RSC is one of Europe's oldest sentinel systems.[1] The RSC has been participating in sentinel surveillance since 1967 and was recruited to be nationally representative.[2] The RSC has expanded greatly through the COVID-19 pandemic and covers just under a third of the national population.[3]
This method, our code (strictly clinical terms) lists, and weightings. These are free to use, but this is conditional on citing the source peer review publication.

Citation:
Tsang R, Joy M, Whitaker H, et al., Development and validation of a modified Cambridge Multimorbidity Score for use with internationally recognized electronic health record clinical terms (SNOMED CT). BJGP, accepted for publication 7/10/2022

Method:
To implement the SNOMED CT version of CMMS you will need the spreadsheet of the component clinical terms and the two tables listed below. The two tables are included within this supplementary file (**Supplementary Appendix S3**):

1. **Supplementary Appendix S1 Clinical_terms_SNOMED_CMMS_20221015.xlsx ()**
2. **Table S3.1: Variables, type of observation, calculation type, number and aggregation method**
3. **Table S3.2: CMMS weights**

The variable numbers refer to variables developed by the Clinical Informatics and Health Outcomes Research Group (CIHORG) at the Nuffield Department of Health Sciences at University of Oxford.[4] Variables are only curated by clinicians. Most of our clinical data curators are practicing general practitioners but we also have public health specialists and specialists in training. All curated variables are curated by a clinician and quality assured by another, at least one will be a practicing GP as we consider understanding the context of recording clinical data to be crucial. All clinicians involved in data curation are trained how to identify supertypes and subtypes in the Systematized Nomenclature of Medicine (SNOMED) clinical terms (CT). We have an in house tools that enables us to select supertypes, with or without all their subtypes. Variables go into a Themes Access and Dynamic Data Services (TADDS) library within our secure network. We have plans, but to date at which, our TADDS variables will be in the public domain.[5]

The process is as follows:
1. The **S2_Clinical_terms_SNOMED_CMMS_20221015** spreadsheet contains all the clinical terms needed to curate the required variables in your dataset. This is Supplementary Appendix **S2.** There are three worksheets in the workbook:
    a. **SNOMED clinical terms:** This worksheet contains the SNOMED clinical terms (CT) that link to a specific condition that is used in CMMS.
    b. **dm+d Codes:** This worksheet contains Dictionary of Medicines and Devices (dm+d) codes that might either be a component of CMMS or infer that the patient has a given condition.
    c. **EMIS (medication) codes:** This worksheet contains EMIS medication codes that cover areas equivalent to the dm+d listed above.

Table S3.1: Variables, type of observation, calculation type, number and aggregation method

Variable	Condition TADDS ID	Observation or Issued Prescription	Calculation Type	Calculation Number	Aggregate Type
AlcoholProblems	4385	Observation	Ever	1	Any
AlcoholProblems	4392	Observation	Ever	1	Any
AlcoholProblems	4934	Observation	Ever	1	Any
AnxietyOrDepression	4790	Observation	Last12Months	1	Any
AnxietyOrDepression	4791	Observation	Last12Months	1	Any
AnxietyOrDepression	5070	IssuedPrescription	Last12Months	4	Any
AnxietyOrDepression	5151	Observation	Last12Months	1	Any
AnxietyOrDepression	5160	IssuedPrescription	Last12Months	4	Any
AtrialFibrillation	2025	Observation	Ever	1	Any
CancerInTheLast5Years	4778	Observation	Last5Years	1	Any
CancerInTheLast5Years	5059	Observation	Last5Years	1	Any
ChronicKidneyDisease	1500	Observation	eGFR	0	Any
ChronicLiverDiseaseAndViralHepatitis	1056	Observation	Ever	1	Any
ChronicLiverDiseaseAndViralHepatitis	2028	Observation	Ever	1	Any
Constipation	5071	IssuedPrescription	Last12Months	4	Any
COPD	2133	Observation	Ever	1	Any
Dementia	2002	Observation	Ever	1	Any
Diabetes	2047	Observation	Ever	1	Any
Diabetes	2048	Observation	Ever	1	Any
DisorderOfProstate	5052	Observation	Ever	1	Any
Epilepsy	2150	Observation	Ever	1	All
Epilepsy	5069	IssuedPrescription	Last12Months	1	All
HeartFailure	2031	Observation	Ever	1	Any
IrritableBowelSyndrome	5055	Observation	Ever	1	Any
IrritableBowelSyndrome	5068	IssuedPrescription	Last12Months	4	Any
LearningDisability	5054	Observation	Ever	1	Any
LearningDisability	5149	Observation	Ever	1	Any
LearningDisability	5150	Observation	Ever	1	Any
MultipleSclerosis	2032	Observation	Ever	1	Any
PainfulCondition	2150	Observation	Ever	1	AnyX~
PainfulCondition	5065	IssuedPrescription	Last12Months	4	Any
PainfulCondition	5066	IssuedPrescription	Last12Months	4	AnyX
Parkinsonism	5053	Observation	Ever	1	Any
PeriphVascDiseaseLeg	1525	Observation	Ever	1	Any
PsychoactiveSubstanceMisuse	5051	Observation	Ever	1	Any
SchizophreniaOrBipolarDisorder	5029	IssuedPrescription	Ever	1	Any
SchizophreniaOrBipolarDisorder	5047	Observation	Ever	1	Any
SchizophreniaOrBipolarDisorder	5048	Observation	Ever	1	Any

2. **Table S3.1: Variables, type of observation, calculation type, number and aggregation method** contains the variable names, taken from the clinical terms and medication spreadsheet above. The variable names are listed in the table (S3.1) above. Conditions are either derived from clinical observations (e.g. SNOMED CT findings) or from being issued a prescription (e.g. being prescribed an analgesic implies the need for pain relief for a painful condition, which may or may not be recorded in the clinical record). Other logic is applied, for example many diagnoses can be recorded ever – but cancer restricted to the last five years. Relevant prescriptions must have been issued in the last 12 months; we suggest that users set their own index date to match the needs of their study. .

   The calculation number column indicates the minimum number of observation or prescription records that are required for the variable to be met. For example, for the irritable bowel syndrome variable to contribute to a patient's CCMS score they must have four prescriptions related to treating irritable bowel syndrome in the 12 months prior to the index date. If a patient has only three relevant prescriptions in the relevant time period, they would score 0 for that particular variable.

   The chronic kidney disease variable has 0 for the calculation number as the variable construction is a little different to the other variables. First all event records for TADDS identifier (ID) 1500 with an event date on or before the index date are extracted. The minimum estimated glomerular filtration rage (eGFR) value per event date is extracted for each patient. The variable is met if the larger of a patient's two most recent eGFR measurements in that period is less than 60ml/min.

   The aggregate type column specifies how information from the different TADDS IDs should be combined when constructing each variable. 'Any' indicates that the variable would contribute to the overall CMMS score if a patient has observations or prescriptions meeting any the requirements for any of the TADD IDs for that variable, e.g. The Alcohol problems variable would contribute to a patient's CMMS score if they had any observations for ANY of the three TADD IDs for that variable.

   If the aggregate type is 'All' it indicates that all TADDS ID criteria must be met for that variable. For example, for the Epilepsy variable, a patient must have both a diagnosis of epilepsy AND must have been issued with relevant epilepsy medication in the 12 months prior to the index date.

   For the "Painful condition" variable the 'AnyX' and 'AnyX~' notation is used to indicate that prescription records for TADD ID 5066 are only included if the patient does not have observations for TADDS ID 2150.

3. **Table S3.2: CMMS weights** contains the weights applied to the 21 different variables to be used when applying the model. We have not included age or gender, as the balance may change between populations. We commented in the discussion:

   We have opted to use the unadjusted 21-condition model as this would maximise its use in studies of different designs, where researchers can apply their own adjustments for age and sex

   In this table (S3.w below) we list the 21 condition used in the unadjusted model. We include the weightings used for age, the weighting marked as "sex" should be applied to male gender. We used these age-gender adjustments for our broadly nationally representative RSC population.

Table S3.2: CMMS weights

21 included weights	Variable	Weight
1	AlcoholProblems	0.792243
2	AnxietyOrDepression	0.324207
3	AtrialFibrillation	0.334891
4	CancerInTheLast5Years	1.202615
5	ChronicKidneyDisease	0.213652
6	ChronicLiverDiseaseAndViralHepatitis	0.68621
7	Constipation	0.383006
8	COPD	0.702181
9	Dementia	0.938001
10	Diabetes	0.29467
11	DisorderOfProstate	-0.18781
12	Epilepsy	0.477465
13	HeartFailure	0.505245
14	IrritableBowelSyndrome	-0.20368
15	LearningDisability	0.637273
16	MultipleSclerosis	0.761606
17	PainfulCondition	0.445521
18	Parkinsonism	0.546194
19	PeriphVascDiseaseLeg	0.334558
20	PsychoactiveSubstanceMisuse	0.449321
21	SchizophreniaOrBipolarDisorder	0.482469
Not included in the final unadjusted model		
	Age.squared	0.000609
	Sex	0.293827

Clinical terminology updates:
These lists of clinical terms and drug dictionary contents were correct as of autumn 2022. However, there are regular updates of clinical terms, generally six monthly, with emergency release when needed (e.g. COVID-19, when there were serial releases).[6,7]

The CIHORG team update TADDS variables where needed for active projects. Therefore at present we don't offer a readily downloadable update of variables. Users are recommended to update or request updates from us.

Acknowledgement:
Patients and public who allow data sharing, general practices who are members of the Oxford-RCGP RSC for sharing pseudonymised data. EMIS, TPP, InPractice Systems, and Wellbeing for assistance with extraction of pseudonymised patient data.

References:

[1] de Lusignan S, Correa A, Smith GE, Yonova I, Pebody R, Ferreira F, Elliot AJ, Fleming D. RCGP Research and Surveillance Centre: 50 years' surveillance of influenza, infections, and respiratory conditions. Br J Gen Pract. 2017 Oct;67(663):440-441. doi: 10.3399/bjgp17X692645.

[2] Correa A, Hinton W, McGovern A, van Vlymen J, Yonova I, Jones S, de Lusignan S. Royal College of General Practitioners Research and Surveillance Centre (RCGP RSC) sentinel network: a cohort profile. BMJ Open. 2016 Apr 20;6(4):e011092. doi: 10.1136/bmjopen-2016-011092.

[3] Leston M, Elson WH, Watson C, Lakhani A, Aspden C, Bankhead CR, Borrow R, Button E, Byford R, Elliot AJ, Fan X, Hoang U, Linley E, Macartney J, Nicholson BD, Okusi C, Ramsay M, Smith G, Smith S, Thomas M, Todkill D, Tsang RR, Victor W, Williams A, Williams J, Zambon M, Howsam G, Amirthalingam G, Lopez-Bernal J, Hobbs FR, de Lusignan S. Representativeness, vaccination uptake and COVID clinical outcomes 2020-21 in the UK's Oxford-RCGP Research and Surveillance Network: cohort profile. JMIR Preprints. 04/05/2022:3914 URL: https://preprints.jmir.org/preprint/39141

[4] Clinical Informatics and Health Outcomes Research Group (CIHORG). University of Oxford. URL: https://www.phc.ox.ac.uk/research/clinical-informatics-and-health-outcomes

[5] Oxford Royal College of General Practitioners (ORCHID). Access to ORCHID data. URL: https://orchid.phc.ox.ac.uk/index.php/orchid-data/

[6] de Lusignan S, Williams J. To monitor the COVID-19 pandemic we need better quality primary care data. BJGP Open. 2020 Jun 23;4(2):bjgpopen20X101070. doi: 10.3399/bjgpopen20X101070.

[7] de Lusignan S, Liyanage H, McGagh D, Jani BD, Bauwens J, Byford R, Evans D, Fahey T, Greenhalgh T, Jones N, Mair FS, Okusi C, Parimalanathan V, Pell JP, Sherlock J, Tamburis O, Tripathy M, Ferreira F, Williams J, Hobbs FDR. COVID-19 Surveillance in a Primary Care Sentinel Network: In-Pandemic Development of an Application Ontology. JMIR Public Health Surveill. 2020 Nov 17;6(4):e21434. doi: 10.2196/21434.