Appendix 3 (as supplied by the authors): *Statistical analysis and data processing*

Morbidity scores were developed using three separate models, one for each outcome, in the 2015 development dataset. The distribution of GP consultations (primary care utilisation) was captured best using a zero-inflated negative binomial model (see below). Mortality and unplanned hospitalisation were modelled using Cox regression. As detailed below, in addition to the extended scores containing all 37 conditions, we constructed a set of simplified primary scores including the most important 20 conditions.

We attempted to model the number of consultations using Poisson, Negative Binomial, Zero Inflated Poisson and Zero Inflated Negative Binomial (ZINB) regression. The model that gave the best fit to the observed distribution of our consultation data was the ZINB regression (Figure A3.1). We used an approach that involved the use of two models on separate groups of patients. The rationale for this approach was that we needed a zero inflated model to account for the excess of zero consultations in the distribution however, the existence of a subsample patients with a high probability of having consultations would make this model fail to run. Our approach consisted of running a negative binomial model in the subset of patients more likely to consult (group A patients) and running the ZINB model in the remaining patients (group B patients). We identified the group of patients more likely to consult based on their prevalent conditions (group A conditions). We started by including the conditions with the highest crude proportions of "Having consulted at least once" in group A. Then, through an iterative process, we identified the remaining conditions to be included in group A which would have very high standard errors in the ZINB model. The following 24 conditions were included in group A: Diabetes, Epilepsy, Asthma, Anxiety or Depression, Hypertension, Painful condition, Parkinson's disease, Prostate disorders, Coronary heart disease, Psoriasis or eczema, Psychosis/bipolar disorder, COPD, Multiple sclerosis, Stroke & TIA, Diverticular disease, Peripheral vascular disease, Chronic kidney disease, Dementia, Migraine, Constipation, Atrial fibrillation, Learning disability, Thyroid disorders. A patient with any of these conditions would be part of the sample for the negative binomial model. Group A patients accounted for 56% of the development sample. In the end, the predicted number of consultations from both models can be used together as if they originated from one single model.

An initial model was built for each outcome including a binary indicator for each of the 37 conditions, age (both as linear and categorical terms) and gender. This model assumed the effect of each condition to be additive and independent of disease burden; in other words, that the effect of each condition is the same regardless of how many and which specific conditions each person had. Given there is reason to doubt this assumption, we compared this initial model with one that attempted to account for the "subadditivity" of conditions by introducing linear and quadratic terms for the count of conditions. Including the count of conditions produced little improvement to model performance (C-index) for each outcome, and the weights for each condition were very similar to those obtained from the main model. Given this small effect and the added complexity, we used the initial model to develop the weights for the extended scores. Our strategy was to create weights that were adjusted for of age and gender, but we purposely did not adjust for other factors as those would be highly context dependent (e.g. UK health care system) and thus the resulting weightings would be less robust and have poorer external generalisability. Of note, we elected to censor age at 95 years, as consultation rates were low above this age suggesting unrecorded deaths in this age group. No attempt was made to reduce the number of conditions through variable selection.

Appendix to: Payne RA, Mendonca SC, Elliott MN, et al. Development and validation of the Cambridge Multimorbidity Score. *CMAJ* 2020. DOI:10.1503/cmaj.190757.

Each model was used to estimate predictions (Average Treatment Effect on the Treated, ATT) which corresponded to the expected difference in outcome in each group with a specific condition compared to the same group if they were not to have the condition. Predictions are expressed in the natural scale of the outcomes (e.g. number of events per person-year) to facilitate interpretation. The multimorbidity scores for each outcome were built as the sum of conditions weighted by these predictions.

To construct the simplified primary scores, we selected the 20 most important conditions according to three criteria: effect size (weights), prevalence, and a combination of effect size and prevalence; this list is shown in table A3.1 below. The ranking of conditions based on each criterion was averaged across outcomes, to construct three sets of conditions common to all outcomes. The shortened list of conditions based on combined effect size and prevalence was considered clinically most relevant to use in further analyses. This decision was also supported by this list also resulting in the best performing score. In addition, a general-outcome multimorbidity score was constructed by averaging the standardised weights of the three simple scores. The resulting general weights were then restandardised; unlike the main weights, these are dimensionless quantities that are associated with approximately a 1 SD increase in each of the three outcomes (i.e. consultations, hospitalisation, mortality).

Table A3.1. Top 20 conditions ordered according to the 3 criteria considered

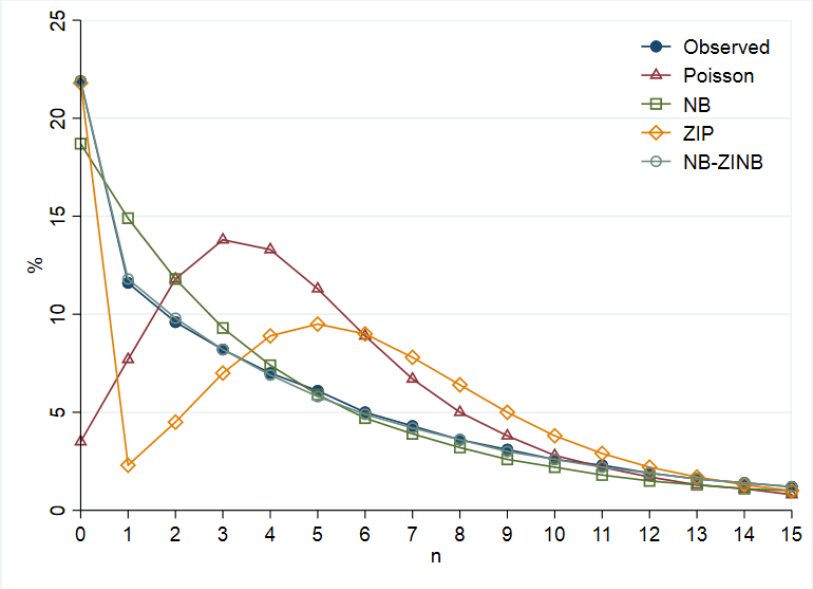
Prevalence	Impact	Prevalence & impact
Hypertension	COPD	Painful condition
Anxiety/Depression	Atrial fibrillation	Anxiety/Depression
Painful condition	Parkinson's disease	Diabetes
Hearing loss	Cancer	COPD
Irritable bowel syndrome	Dementia	Atrial fibrillation
Asthma	Painful condition	Cancer
Diabetes	Heart failure	Constipation
Prostate disorders	Epilepsy	Coronary heart disease
Thyroid disorders	Constipation	Chronic kidney disease
Coronary heart disease	Stroke & TIA	Stroke & TIA
Chronic kidney disease	Multiple sclerosis	Dementia
Diverticular disease	Diabetes	Heart failure
Chronic sinusitis	Bronchiectasis	Hypertension
Atrial fibrillation	Chronic Liver Disease	Alcohol problems
Constipation	Psychosis/bipolar disorder	Epilepsy
Stroke & TIA	Anxiety/Depression	Asthma
COPD	Coronary heart disease	Hearing loss
Connective tissue disorder	Learning disability	Connective tissue disorder
Cancer	Connective tissue disorder	Irritable bowel syndrome
Peptic ulcer disease	Alcohol problems	Psychosis/bipolar disorder

We assessed whether including long-term medication count (defined as the number of unique chemical substances issued at least twice in the 3-month period prior to the index date) would substantially improve model fit. Adding the medication use to the initial 37-condition model slightly improved the C-index only for number of consultations (by 0.007). Importantly, medication count may be endogenous with the number of consultations because of the way consultations are recorded in CPRD. Given this concern, and the small added benefit, medication count was not included in the final models.

Performance of each of the three 37-condition and 20-condition outcome-specific scores, as well as the 20-condition general-outcome score, was independently evaluated at 1-year follow-up in the 2015 (synchronous) dataset, as well as at 1-year and 5-years follow-up in the 2011 (asynchronous) dataset. We examined the performance of each score for predicting each of the three outcomes, and additionally compared performance against the Charlson index. Given that the main goal was to develop weights that reflect patients' multimorbidity burden as opposed to optimize prediction of specific outcomes, model fit was assessed using Harrell's C-index, where 1 represents perfect model fit and 0.5 model performance that is no better than chance alone. Interpretation of the C-index does require a value judgement; we have attempted to be objective in our interpretation by using the definitions provided by Hosmer and Lemeshow (Applied Logistic Regression (2nd ed), Wiley, 2000), where >0.7 is considered acceptable performance, and >0.8 considered excellent performance.

The majority of the analysis was carried out in Stata 15, including data preparation and running the ZINB models; the Cox models were run in R 3.4.2.

Figure A3.1: Probability of the outcome (number of consultations) taking different values, according to different model distributions. The NB-ZINB approach is the one that more closely predicts the distribution of the observed data.



Sample size

Our sample size calculation reflects the aim of developing weights as opposed to hypothesis testing. The sample size was selected to limit the width of a 95% confidence interval for a condition with 2% prevalence to approximately 0.5 on the log-odds scale for a dichotomous outcome such as mortality. Our disease classification was developed from the Cassell work, for which the median prevalence of different morbidities was 2% (range 0.25% to 18.2%). We ran a simple simulation under the null hypothesis (of no association) for an outcome with 1% prevalence, and exposures with 0.25%, 1% and 2%. For a sample size of 300,000 we estimated widths of the confidence intervals (on the log-odds scale) to be 1.19, 0.62 and 0.53 respectively for exposure prevalences of 0.25%, 1% and 2%.

Data cleaning and preparation

An initial examination of the data showed that above 95 years of age there was a marked increase in the number of zero consultations raising suspicion that some of these patients had died without their record being updated. It was therefore decided to exclude patients older than 95 years old at their corresponding index date.

Our initial patient sample (n=606,000) was selected based on the CPRD denominator tables (ie. using CPRD death dates), however a small proportion of discrepancies appeared when comparing CPRD and ONS death dates. This phenomenon has been described previously (Harshfield at al, 2018). A small number of patients (n=237) was found to have died before their corresponding index dates and had to be excluded post-hoc. The final 600,000 patient sample was created from the 606,000 sample (see Appendix 1) after excluding patients with missing deprivation quintiles, patients older than 95 years of age and those who died before the index date.

We considered a patient to have Chronic Kidney Disease (CKD) if the best (highest value) of their last 2 eGFR readings was < 60 mL/min, with eGFR being the estimated Glomerular Filtration Rate. We identified this laboratory test as an entity type with the same name recorded in the CPRD “test” table. eGFR test results are saved as multiple variables resulting in specific values (e.g. “=73”) or ranges of values (e.g. “<60”). First of all, duplicate tests and those with a missing values were excluded. Secondly, test results with ranges that could not be classified were also excluded (e.g. “≤60”, “<90”). Test results with value zero were excluded as they appeared to be missing data rather than an actual zero. We included values associated with units other than “mL/min” because the distributions of different units were all very similar which implied that the data was most likely in the correct unit. A minimum of two tests were required for the patient to be considered having CKD.

Table A3.2. Definitions of morbidities

Morbidities based on Read code ever recorded	
Alcohol problems	Heart failure
Anorexia or bulimia	Hypertension
Atrial fibrillation	Inflammatory bowel disease
Blindness and low vision	Learning disability
Bronchiectasis	Multiple sclerosis
Chronic liver disease and viral hepatitis	Parkinson's disease
Chronic sinusitis	Peptic ulcer disease
COPD	Peripheral vascular disease
Coronary heart disease	Prostate disorders
Dementia	Psychoactive substance misuse (not alcohol)
Diabetes	Rheumatoid arthritis, other inflammatory polyarthropathies & systematic connective tissue disorders
Diverticular disease of intestine	Stroke & transient ischaemic attack
Hearing loss	Thyroid disorders
Morbidities based on prescription in last 12 months	
Constipation (Treated)	≥4 laxative prescriptions
Migraine	≥4 prescription-only medicine anti-migraine prescriptions
Morbidities based on combination of Read code ever recorded and/or prescription in last 12 months	
Epilepsy (currently treated)	Read code AND ≥1 antiepileptic prescription
Asthma (currently treated)	Read code AND ≥1 asthma prescription
Irritable bowel syndrome	Read code OR ≥4 antispasmodic prescriptions
Psoriasis or eczema	Read code AND ≥4 related prescriptions (excluding simple emollients)
Morbidities otherwise defined	
Anxiety & other neurotic, stress related & somatoform disorders OR depression	Read code (depression or anxiety) in last 12 months OR ≥4 anxiolytic/hypnotic prescriptions in last 12 months OR ≥4 anti-depressant prescriptions (excluding low dose tricyclics) in last 12 months
Cancer - [New] Diagnosis in last five years (excluding non-melanoma skin cancer)	Read code [first] recorded in last 5 years
Chronic kidney disease	Highest value of last 2 eGFR readings is < 60 mL/min
Painful condition	≥4 prescription-only medicine analgesics in last 12 months OR (≥4 specified anti-epileptics in last 12 months AND no epilepsy Read code ever recorded)
Schizophrenia (and related non-organic psychosis) or bipolar disorder	Read code ever recorded OR Lithium ever prescribed

Table A3.3. Overlap of Cambridge Multimorbidity Score and Charlson Index conditions

Charlson morbidity	Corresponding Cambridge morbidity
AIDS	***Charlson only***
Cerebrovascular disease	Stroke & transient ischaemic attack
Chronic pulmonary disease	***Combination of*** asthma, bronchiectasis, COPD ***(plus codes for cystic fibrosis and pulmonary fibrosis)***
Congestive heart failure	Heart failure
Dementia	Dementia
Diabetes without complications	***Subset of*** Diabetes
Diabetes with complications	***Subset of*** Diabetes
Hemiplegia	***Charlson only***
Mild liver disease	***Subset of*** Chronic Liver Disease and Viral Hepatitis
Moderate or severe liver disease	***Subset of*** Chronic Liver Disease and Viral Hepatitis
Myocardial infarction	***Subset of*** Coronary heart disease
Peptic ulcer disease	Peptic ulcer disease
Peripheral vascular disease	Peripheral vascular disease
Renal disease	Chronic kidney disease
Connective tissue (rheumatological) disease	Rheumatoid arthritis, other inflammatory polyarthropathies & systematic connective tissue disorders
Cancer (including lymphoma/leukaemia)	***Subset of*** Cancer
Metastatic solid tumour	***Subset of*** Cancer