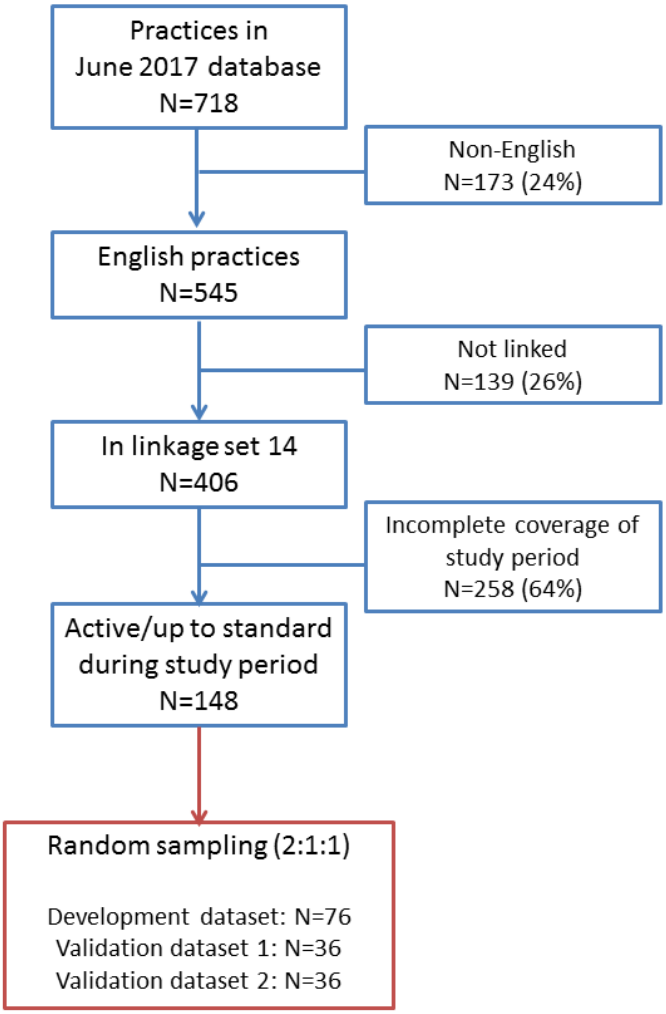
Appendices

Appendix S1

Flowchart for the selection of GP practices.



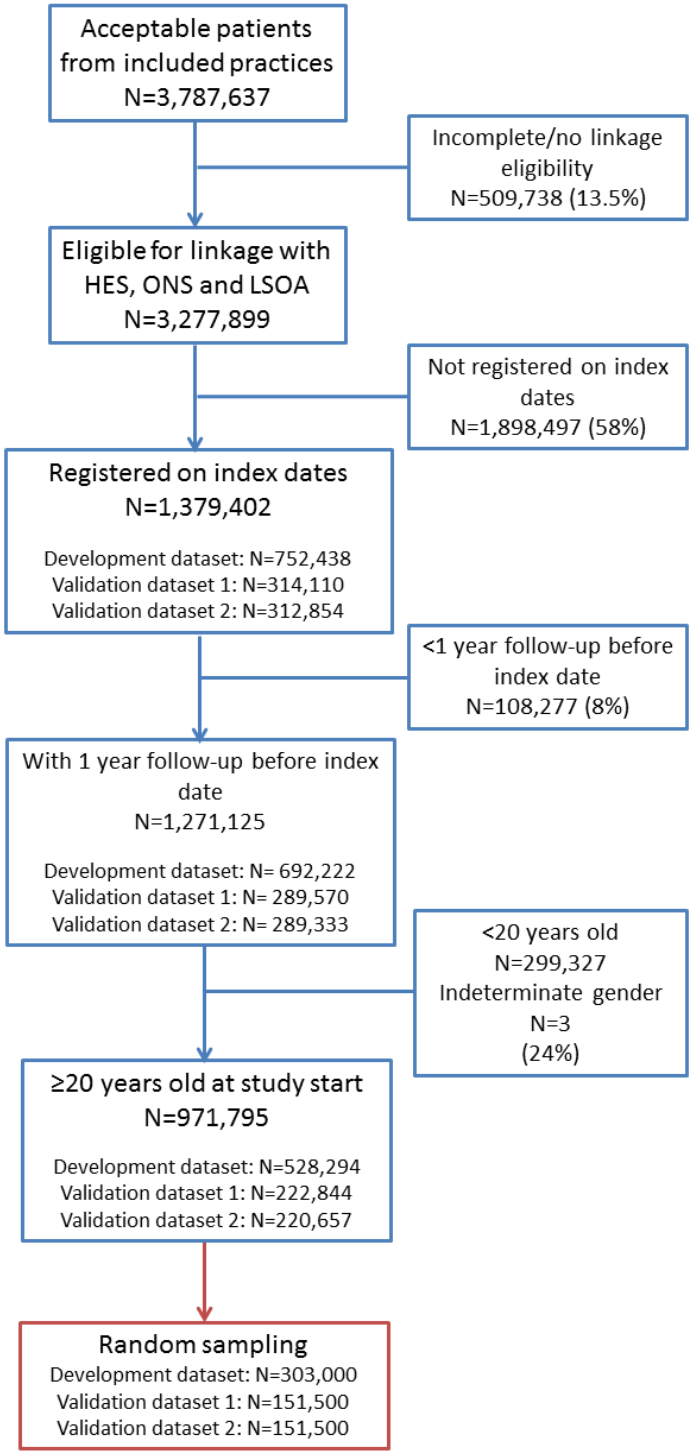
Notes:
i. Study period refers to 01/01/2010 to 31/12/2015. Non-English practices refer to Scottish, Welsh and Northern Irish.
ii. At the allocation stage of practices to the three datasets, in order to minimize variation due to chance, practices were sorted by a measure of mortality not explained by age and gender (i.e. the best linear unbiased predictor from a model including a practice random effect), using a block randomization approach.
iii. We compared included and excluded English practices in terms of list-size, percentage of patients >65 years old and mortality rates, as summarized in the following table. The median list-size for "included" practices is slightly larger than for "excluded" practices, but otherwise the range of values for all three characteristics is comparable to that of the excluded practices.

Practice characteristic[1]	Practices included in study sample (N=148)	English practices excluded (N=377)[2]
List-size (x1,000), median (range)	9.0 (2.2-28.0)	7.7 (1.4-33.5)
Percentage of patients >65 years old, median (range)	17.8 (5.8-39.9)	16.4 (2.9-40.7)
Mortality rate, per 1,000 patient-years (range)	8.8 (2.3-20.9)	8.4 (1.0-24.5)

1. List-size and percentage of patients >65 years old were calculated at one year before each practice last collection date for only those practices where CPRD rated data collection as "up to standard" (UTS) at that time. Mortality rates were calculated in the period between 1 and 2 years before the practice's last data collection date, to avoid changes due to the practice closing.

2. Twenty practices were not UTS one year before last collection date so they were excluded from this analysis, as it required UTS data.

Flowchart for the selection of patients.



Notes: We sampled an additional 1% for each dataset to allow for later exclusion criteria, in particular, missing deprivation quintile and divergence between CPRD and ONS death dates (see Appendix S3).

Appendix S2

Datasets used, relevant time periods and dates.

Dataset	Number of patients	Study start	Index date	Study end	Duration of follow up
Development dataset	300,000	01/01/2014	01/01/2015	31/12/2015	1 year
Validation dataset 1 (synchronous)	150,000	01/01/2014	01/01/2015	31/12/2015	1 year
Validation dataset 2 (asynchronous)	150,000	01/01/2010	01/01/2011	31/12/2011	1 year
		01/01/2010	01/01/2011	31/12/2015	5 years

Note that validation dataset 2 is used to calculate outcomes for both 1 year and 5 years follow up.

Appendix S3

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 S3.1). 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.

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 S3.1 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 S3.1. Top 20 conditions ordered according to the 3 criteria considered

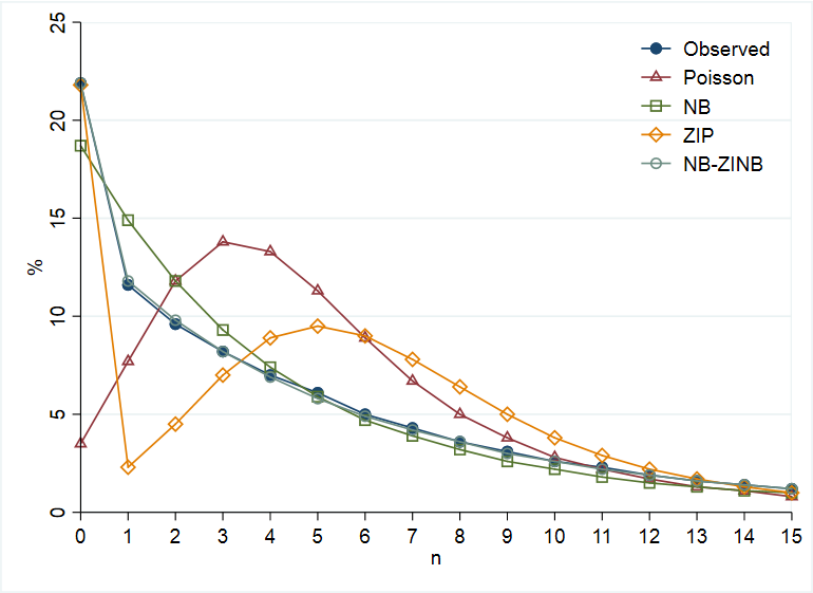
Prevalence	Impact	Prevalence & impact
Hypertension	COPD	Painful condition
Anxiety/Depression	Atrial fibrillation	Anxiety/Depression
Painful condition	Parkinson's disease	Diabetes
Hearing loss	Cancer	COPD
Irritable bowel syndrome	Dementia	Atrial fibrillation
Asthma	Painful condition	Cancer
Diabetes	Heart 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 S3.1: 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 S1) 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 S3.2. Definitions of morbidities

Morbidity	Definition
Alcohol problems	Read code ever recorded
Anorexia or bulimia	Read code ever recorded
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
Asthma (currently treated)	Read code ever recorded AND Any prescription in last 12 months
Atrial fibrillation	Read code ever recorded
Blindness and low vision	Read code ever recorded
Bronchiectasis	Read code ever recorded
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
Chronic Liver Disease and Viral Hepatitis	Read code ever recorded
Chronic sinusitis	Read code ever recorded
Constipation (Treated)	≥4 laxative prescriptions in last 12 months
COPD	Read code ever recorded
Coronary heart disease	Read code ever recorded
Dementia	Read code ever recorded
Diabetes	Read code ever recorded
Diverticular disease of intestine	Read code ever recorded
Epilepsy (currently treated)	Read code ever recorded AND Any antiepileptic prescription in last 12 months
Hearing loss	Read code ever recorded
Heart failure	Read code ever recorded
Hypertension	Read code ever recorded
Inflammatory bowel disease	Read code ever recorded
Irritable bowel syndrome	Read code ever recorded OR ≥4 antispasmodic prescription only in last 12 months
Learning disability	Read code ever recorded
Migraine	≥4 prescription-only medicine anti-migraine prescriptions in last 12 months
Multiple sclerosis	Read code ever recorded
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)
Parkinson's disease	Read code ever recorded
Peptic Ulcer Disease	Read code ever recorded
Peripheral vascular disease	Read code ever recorded
Prostate disorders	Read code ever recorded
Psoriasis or eczema	Read code ever recorded AND ≥4 related prescriptions in last 12 months (excluding simple emollients)
Psychoactive substance misuse (not alcohol)	Read code ever recorded
Rheumatoid arthritis, other inflammatory polyarthropathies & systematic connective tissue disorders	Read code ever recorded

Schizophrenia (and related non-organic psychosis) or bipolar disorder	Read code ever recorded OR Lithium ever prescribed
Stroke & transient ischaemic attack	Read code ever recorded
Thyroid disorders	Read code ever recorded

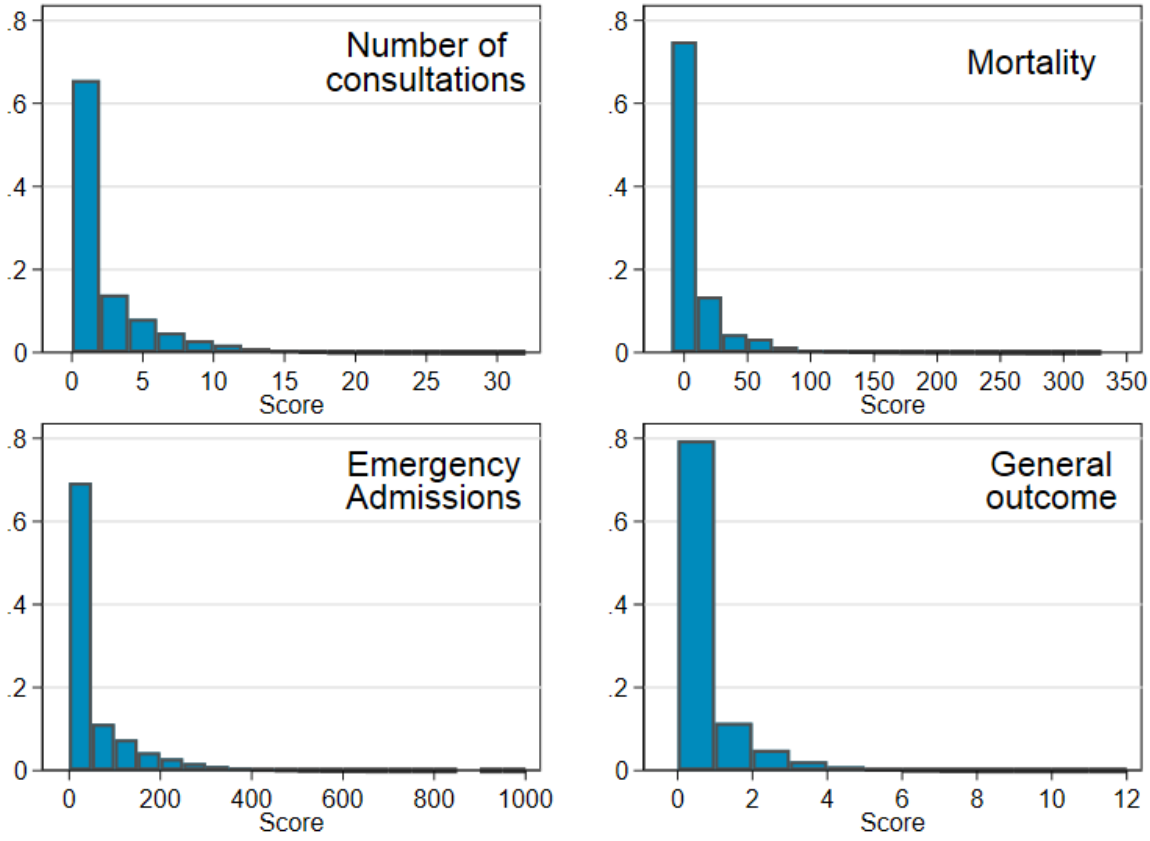
Table S3.3. Overlap of Cambridge Multimorbidity Score and Charlson Index conditions

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

Table S4.1. Descriptive statistics for the Cambridge Multimorbidity Scores (20-conditions) using validation dataset 1. Median and interquartile range for the patient-level sum of weights.

Outcome	Median (IQR)
Consultations	0.66 (0-3.4)
Mortality	0 (0-9.9)
Emergency Admissions	8.6 (0-71.8)
General	0.08 (0-0.75)

Figure S4.1: Histograms showing the distribution of the Cambridge Multimorbidity Scores (20-conditions) in validation dataset 1.



Appendix S5

Prevalence and weights[1] for the extended version of the multimorbidity scores.

	Prevalence[2]	Weight for consultations[3]	Weight for mortality[4]	Weight for emergency admissions[4]	General-outcome weight[5]
Hypertension	19.24	0.83	-1.88	11.29	0.09
Anxiety/Depression	12.85	2.15	6.88	44.33	0.47
Painful condition	11.63	3.41	16.30	82.13	0.87
Hearing loss	11.27	0.96	-3.72	7.73	0.07
Irritable bowel syndrome	7.61	1.71	-0.77	7.46	0.18
Asthma	7.20	1.34	-2.45	21.47	0.18
Diabetes	6.58	3.84	9.83	53.95	0.71
Prostate disorders	6.31	1.26	-10.02	5.13	0.01
Thyroid disorders	5.24	0.93	-0.83	1.24	0.08
Coronary heart disease	4.79	1.49	4.29	68.05	0.46
Chronic kidney disease	4.50	0.97	16.47	51.24	0.51
Diverticular disease	3.24	0.77	-10.09	9.60	-0.02
Chronic sinusitis	2.96	1.11	-0.19	4.88	0.13
Atrial fibrillation	2.72	5.98	22.93	105.78	1.30
Constipation	2.67	3.16	34.58	64.91	1.03
Stroke & TIA	2.55	1.53	20.41	88.15	0.77
COPD	2.46	3.40	42.29	129.18	1.41
Connective tissue disorder	2.33	3.00	0.08	27.45	0.40
Cancer	2.15	2.65	62.28	103.69	1.50
Peptic ulcer disease	1.62	0.53	5.69	17.66	0.20
Alcohol problems	1.60	0.81	11.42	81.19	0.55
Substance misuse	1.19	1.01	2.79	61.41	0.38
Psoriasis or eczema	1.16	1.88	-1.46	22.30	0.25
Blindness and low vision	1.08	0.33	1.16	24.38	0.15
Heart failure	1.04	2.86	42.26	70.44	1.12
Dementia	1.02	1.87	122.92	158.14	2.46
Psychosis/bipolar disorder	0.98	2.22	6.64	71.24	0.58
Epilepsy	0.97	2.05	17.34	107.94	0.85
Inflammatory bowel disease	0.96	2.63	-0.45	49.30	0.44
Peripheral vascular disease	0.88	0.87	15.21	60.09	0.53
Anorexia or bulimia	0.55	0.86	8.54	36.01	0.34
Chronic Liver Disease	0.53	1.27	22.22	77.03	0.72
Migraine	0.51	1.12	-4.04	4.65	0.07
Learning disability	0.47	1.15	10.92	55.75	0.47
Bronchiectasis	0.43	2.69	5.65	84.15	0.66
Multiple sclerosis	0.28	2.18	8.77	94.29	0.69
Parkinson's disease	0.28	3.48	40.46	104.13	1.29

1. Negative weights can be interpreted as reflecting a negative association with the outcome of interest after controlling for other conditions; 2. Based on development dataset; 3. Per person-year; 4. Per 1,000 person-years; 5. Unit change associated a 1 SD change in each of the three outcomes

Appendix S6

Comparison of C-indices (95% confidence intervals) from different model specifications for a NB-ZINB model for number of consultations.

	Development model[1]		Adjusted scores[2]				Unadjusted scores[3]			
	37 conditions	20 conditions	Extended outcome-specific scores	Primary outcome-specific scores	Charlson comorbidity index	Primary general-outcome score	Extended outcome-specific scores	Primary outcome-specific scores	Charlson comorbidity index	Primary general-outcome score
C-index										
Development dataset	0.738	0.736	0.732	0.727	0.695	0.723	0.700	0.690	0.602	0.688
2015, 1-year follow-up	0.737	0.735	0.732 (0.731-0.734)	0.727 (0.725-0.728)	0.691 (0.690-0.693)	0.723 (0.722-0.725)	0.702 (0.701-0.704)	0.692 (0.691-0.694)	0.605 (0.603-0.606)	0.690 (0.689-0.691)
2011, 1-year follow-up	0.729	0.728	0.724 (0.722-0.725)	0.719 (0.717-0.720)	0.686 (0.684-0.688)	0.715 (0.714-0.717)	0.690 (0.689-0.692)	0.681 (0.680-0.683)	0.594 (0.593-0.595)	0.679 (0.677-0.680)
2011, 5-year follow-up	0.750	0.749	0.739 (0.738-0.740)	0.735 (0.734-0.736)	0.709 (0.708-0.711)	0.729 (0.728-0.730)	0.679 (0.678-0.681)	0.669 (0.668-0.671)	0.585 (0.583-0.586)	0.667 (0.665-0.668)

1. Using predictions from the score development model with the conditions as binary indicators, adjusted by age and gender.
2. Using predictions from a model including each score (sum of weights) adjusted by age and gender.
3. Using the score directly, without the use of a model (unadjusted).

Extended scores include all 37 conditions. Primary scores include only the 20 most important conditions according to prevalence and impact.
Confidence intervals are provided only for evaluation models and validation datasets.

Appendix S7

Model output for number of consultations

	37 conditions						20 conditions					
	NB model		Count model		Binary model P(cons>0)		NB model		Count model		Binary model P(cons>0)	
	RR (95% CI)	p-value	RR (95% CI)	p-value	OR (95% CI)	p-value	RR (95% CI)	p-value	RR (95% CI)	p-value	OR (95% CI)	p-value
Atrial fibrillation	1.52 (1.49-1.55)	<0.001					1.52 (1.49-1.55)	<0.001				
Diabetes	1.40 (1.39-1.42)	<0.001					1.40 (1.38-1.42)	<0.001				
Painful condition	1.34 (1.33-1.35)	<0.001					1.35 (1.33-1.36)	<0.001				
COPD	1.30 (1.27-1.32)	<0.001					1.31 (1.28-1.33)	<0.001				
Cancer (in last five	1.28 (1.25-1.31)	<0.001					1.27 (1.25-1.30)	<0.001				
Constipation (Treated)	1.24 (1.22-1.27)	<0.001					1.27 (1.24-1.29)	<0.001				
Rheumatoid arthritis,	1.25 (1.23-1.28)	<0.001	1.68 (1.58-1.78)	<0.001	5.96 (2.19-	<0.001	1.27 (1.24-1.29)	<0.001	1.69 (1.59-1.80)	<0.001	5.83 (2.13-15.96)	0.001
Schizophrenia (and	1.25 (1.21-1.29)	<0.001					1.25 (1.22-1.29)	<0.001				
Epilepsy (currently	1.23 (1.20-1.27)	<0.001					1.25 (1.21-1.28)	<0.001				
Anxiety OR Depression	1.25 (1.23-1.26)	<0.001					1.24 (1.23-1.26)	<0.001				
Irritable bowel	1.18 (1.16-1.20)	<0.001	1.42 (1.38-1.46)	<0.001	3.61 (2.77-4.71)	<0.001	1.20 (1.18-1.22)	<0.001	1.44 (1.41-1.48)	<0.001	3.54 (2.72-4.62)	<0.001
Heart failure	1.18 (1.14-1.21)	<0.001	2.69 (2.04-3.55)	<0.001	2.23 (0.43-	0.343	1.18 (1.15-1.22)	<0.001	2.65 (2.01-3.50)	<0.001	2.29 (0.41-12.91)	0.346
Asthma (currently	1.16 (1.14-1.17)	<0.001					1.16 (1.14-1.17)	<0.001				
Dementia	1.14 (1.11-1.18)	<0.001					1.14 (1.10-1.18)	<0.001				
Coronary heart disease	1.13 (1.11-1.14)	<0.001					1.13 (1.11-1.14)	<0.001				
Stroke & transient	1.12 (1.10-1.14)	<0.001					1.12 (1.10-1.14)	<0.001				
Hearing loss	1.10 (1.08-1.11)	<0.001	1.18 (1.15-1.21)	<0.001	2.53 (2.15-2.97)	<0.001	1.11 (1.09-1.12)	<0.001	1.19 (1.16-1.22)	<0.001	2.53 (2.15-2.98)	<0.001
Alcohol problems	1.07 (1.04-1.11)	<0.001	1.35 (1.26-1.44)	<0.001	1.57 (1.19-2.07)	0.002	1.09 (1.06-1.13)	<0.001	1.41 (1.32-1.51)	<0.001	1.65 (1.25-2.18)	<0.001
Chronic kidney disease	1.08 (1.06-1.09)	<0.001					1.08 (1.06-1.09)	<0.001				
Hypertension	1.08 (1.07-1.09)	<0.001					1.07 (1.06-1.08)	<0.001				
Parkinson's disease	1.29 (1.22-1.36)	<0.001										
Inflammatory bowel	1.25 (1.20-1.30)	<0.001	1.78 (1.66-1.90)	<0.001	4.58 (2.20-9.51)	<0.001						
Multiple sclerosis	1.25 (1.18-1.32)	<0.001										
Bronchiectasis	1.22 (1.16-1.28)	<0.001	1.42 (1.20-1.69)	<0.001	1.92 (0.47-7.77)	0.363						
Psoriasis or eczema	1.19 (1.16-1.22)	<0.001										
Learning disability	1.15 (1.10-1.21)	<0.001										
Migraine	1.13 (1.09-1.18)	<0.001										
Prostate disorders	1.13 (1.11-1.15)	<0.001										
Chronic Liver Disease	1.13 (1.07-1.18)	<0.001	1.29 (1.17-1.43)	<0.001	4.48 (1.57-	0.005						
Psychoactive substance	1.13 (1.09-1.17)	<0.001	1.30 (1.21-1.39)	<0.001	1.48 (1.14-1.91)	0.003						
Anorexia or bulimia	1.11 (1.05-1.17)	<0.001	1.16 (1.06-1.26)	0.001	1.10 (0.56-2.14)	0.782						

Chronic sinusitis	1.11 (1.09-1.13)	<0.001	1.26 (1.21-1.31)	<0.001	3.46 (2.31-5.20)	<0.001						
Thyroid disorders	1.09 (1.08-1.11)	<0.001										
Peripheral vascular	1.06 (1.03-1.10)	<0.001										
Diverticular disease of	1.06 (1.04-1.08)	<0.001										
Peptic ulcer disease	1.04 (1.01-1.06)	0.007	0.57 (0.37-0.88)	0.011	1.75 (1.14-2.70)	0.011						
Blindness and low vision	1.02 (0.99-1.05)	0.200	1.13 (1.02-1.25)	0.021	1.67 (0.91-3.08)	0.100						
Male	0.85 (0.84-0.86)	<0.001	0.65 (0.64-0.66)	<0.001	0.10 (0.09-0.12)	<0.001	0.86 (0.85-0.87)	<0.001	0.66 (0.65-0.66)	<0.001	0.09 (0.07-0.11)	<0.001
Age at index date (10	1.09 (1.07-1.11)	<0.001	1.07 (1.05-1.10)	<0.001	1.14 (1.04-1.25)	0.006	1.10 (1.08-1.11)	<0.001	1.08 (1.05-1.10)	<0.001	1.15 (1.05-1.27)	0.003
Age 21/30	1.26 (1.20-1.33)	<0.001	1.19 (1.11-1.28)	<0.001	0.93 (0.70-1.24)	0.621	1.26 (1.20-1.32)	<0.001	1.19 (1.11-1.28)	<0.001	0.92 (0.69-1.22)	0.552
Age 31/40	1.17 (1.13-1.21)	<0.001	1.11 (1.05-1.17)	<0.001	0.95 (0.77-1.16)	0.588	1.17 (1.13-1.21)	<0.001	1.11 (1.06-1.17)	<0.001	0.93 (0.76-1.14)	0.51
Age 41/50	1.05 (1.03-1.07)	<0.001	1.00 (0.97-1.03)	0.824	1.06 (0.94-1.21)	0.333	1.05 (1.03-1.07)	<0.001	1.00 (0.97-1.03)	0.952	1.05 (0.93-1.20)	0.432
Age 61/70	1.02 (1.00-1.04)	0.093	1.11 (1.07-1.15)	<0.001	1.95 (1.58-2.39)	<0.001	1.02 (1.00-1.04)	0.033	1.11 (1.07-1.15)	<0.001	2.01 (1.62-2.50)	<0.001
Age 71/80	1.07 (1.04-1.11)	<0.001	1.27 (1.20-1.35)	<0.001	3.41 (2.05-5.66)	<0.001	1.09 (1.05-1.12)	<0.001	1.27 (1.20-1.35)	<0.001	3.74 (2.12-6.58)	<0.001
Age 81/max	1.10 (1.05-1.15)	<0.001	1.43 (1.29-1.57)	<0.001	0.75 (0.47-1.19)	0.221	1.11 (1.05-1.16)	<0.001	1.41 (1.28-1.56)	<0.001	0.76 (0.47-1.21)	0.241
incfu			0.57 (0.55-0.60)	<0.001	0.25 (0.23-0.28)	<0.001			0.56 (0.54-0.59)	<0.001	0.27 (0.24-0.30)	<0.001
_cons	3.43 (3.16-3.73)	<0.001	2.55 (2.24-2.90)	<0.001	15.81 (9.25-	<0.001	3.49 (3.21-3.80)	<0.001	2.53 (2.23-2.89)	<0.001	18.33 (10.52-	<0.001

Appendix S8

Comparison of C-indices (95% confidence intervals) from different model specifications for a Cox model for emergency hospital admission.

	Development model[1]		Adjusted scores[2]				Unadjusted scores[3]			
	37 conditions	20 conditions	Extended outcome-specific scores	Primary outcome-specific scores	Charlson comorbidity index	Primary general-outcome score	Extended outcome-specific scores	Primary outcome-specific scores	Charlson comorbidity index	Primary general-outcome score
C-index										
Development dataset	0.750	0.745	0.748	0.744	0.706	0.740	0.742	0.736	0.662	0.735
2015, 1-year follow-up	0.743	0.739	0.742 (0.737-0.747)	0.738 (0.732-0.743)	0.703 (0.697-0.709)	0.735 (0.729-0.740)	0.738 (0.733-0.744)	0.733 (0.728-0.739)	0.660 (0.656-0.664)	0.731 (0.726-0.737)
2011, 1-year follow-up	0.741	0.737	0.739 (0.733-0.744)	0.734 (0.728-0.740)	0.700 (0.694-0.706)	0.732 (0.726-0.737)	0.732 (0.726-0.737)	0.726 (0.720-0.731)	0.651 (0.647-0.656)	0.724 (0.719-0.730)
2011, 5-year follow-up	0.716	0.712	0.712 (0.709-0.715)	0.708 (0.705-0.712)	0.683 (0.680-0.686)	0.706 (0.703-0.709)	0.700 (0.697-0.704)	0.694 (0.691-0.698)	0.623 (0.621-0.625)	0.692 (0.689-0.695)

1. Using predictions from the score development model with the conditions as binary indicators, adjusted by age and gender.
2. Using predictions from a model including each score (sum of weights) adjusted by age and gender.
3. Using the score directly, without the use of a model (unadjusted).

Extended scores include all 37 conditions. Primary scores include only the 20 most important conditions according to prevalence and impact.
Confidence intervals are provided only for evaluation models and validation datasets.

Appendix S9

Model output for the Cox model for emergency hospital admission

	37 conditions		20 conditions	
	HR (95% CI)	p-value	HR (95% CI)	p-value
Epilepsy (currently	2.05 (1.88-2.24)	<0.001	2.17 (1.99-2.37)	<0.001
Alcohol problems	1.86 (1.72-2.02)	<0.001	2.14 (1.99-2.31)	<0.001
Cancer (in last five	1.86 (1.75-1.97)	<0.001	1.87 (1.77-1.98)	<0.001
COPD	1.78 (1.69-1.87)	<0.001	1.84 (1.74-1.94)	<0.001
Schizophrenia (and	1.63 (1.49-1.79)	<0.001	1.73 (1.57-1.89)	<0.001
Painful condition	1.65 (1.60-1.71)	<0.001	1.69 (1.63-1.75)	<0.001
Dementia	1.54 (1.43-1.65)	<0.001	1.53 (1.43-1.64)	<0.001
Atrial fibrillation	1.52 (1.44-1.60)	<0.001	1.51 (1.44-1.59)	<0.001
Anxiety OR Depression	1.45 (1.40-1.50)	<0.001	1.49 (1.43-1.54)	<0.001
Diabetes	1.43 (1.37-1.49)	<0.001	1.45 (1.39-1.51)	<0.001
Stroke & transient	1.40 (1.34-1.48)	<0.001	1.42 (1.35-1.50)	<0.001
Coronary heart disease	1.38 (1.32-1.44)	<0.001	1.40 (1.34-1.46)	<0.001
Constipation (Treated)	1.27 (1.21-1.33)	<0.001	1.31 (1.25-1.38)	<0.001
Chronic kidney disease	1.24 (1.19-1.30)	<0.001	1.25 (1.19-1.31)	<0.001
Asthma (currently	1.23 (1.17-1.28)	<0.001	1.24 (1.19-1.30)	<0.001
Heart failure	1.22 (1.13-1.31)	<0.001	1.23 (1.14-1.32)	<0.001
Rheumatoid arthritis,	1.19 (1.11-1.26)	<0.001	1.20 (1.12-1.27)	<0.001
Irritable bowel	1.09 (1.04-1.15)	<0.001	1.11 (1.05-1.16)	<0.001
Hypertension	1.09 (1.05-1.12)	<0.001	1.08 (1.04-1.12)	<0.001
Hearing loss	1.06 (1.02-1.10)	0.001	1.07 (1.04-1.11)	<0.001
Multiple sclerosis	2.07 (1.75-2.45)	<0.001		
Chronic Liver Disease	1.76 (1.55-1.99)	<0.001		
Learning disability	1.75 (1.49-2.04)	<0.001		
Psychoactive substance	1.69 (1.53-1.86)	<0.001		
Inflammatory bowel	1.58 (1.42-1.76)	<0.001		
Bronchiectasis	1.50 (1.33-1.69)	<0.001		
Parkinson’s disease	1.50 (1.30-1.72)	<0.001		
Anorexia or bulimia	1.45 (1.25-1.69)	<0.001		
Peripheral vascular	1.24 (1.15-1.35)	<0.001		
Psoriasis or eczema	1.18 (1.08-1.30)	<0.001		
Blindness and low vision	1.12 (1.03-1.21)	0.007		
Peptic ulcer disease	1.11 (1.03-1.20)	0.005		
Migraine	1.07 (0.88-1.29)	0.514		
Chronic sinusitis	1.06 (0.98-1.13)	0.147		
Diverticular disease of	1.05 (1.00-1.11)	0.064		
Prostate disorders	1.03 (0.97-1.10)	0.296		
Thyroid disorders	1.01 (0.96-1.06)	0.700		
Male	1.02 (0.98-1.05)	0.332	1.03 (1.00-1.06)	0.079
Age at index date (10	1.27 (1.21-1.33)	<0.001	1.27 (1.21-1.33)	<0.001
Age 21/30	2.21 (1.90-2.56)	<0.001	2.17 (1.87-2.52)	<0.001
Age 31/40	1.51 (1.36-1.68)	<0.001	1.51 (1.35-1.68)	<0.001
Age 41/50	1.12 (1.05-1.21)	0.001	1.12 (1.05-1.21)	0.001
Age 61/70	0.93 (0.86-0.99)	0.032	0.93 (0.87-1.00)	0.059
Age 71/80	1.08 (0.97-1.20)	0.158	1.10 (0.99-1.22)	0.078
Age 81/max	1.51 (1.30-1.74)	<0.001	1.53 (1.32-1.78)	<0.001

Appendix S10

Comparison of C-indices (95% confidence intervals) from different model specifications for a Cox model for mortality.

	Development model[1]		Adjusted scores[2]				Unadjusted scores[3]			
	37 conditions	20 conditions	Extended outcome-specific scores	Primary outcome-specific scores	Charlson comorbidity index	Primary general-outcome score	Extended outcome-specific scores	Primary outcome-specific scores	Charlson comorbidity index	Primary general-outcome score
C-index										
Development dataset	0.921	0.919	0.912	0.911	0.906	0.914	0.866	0.867	0.802	0.881
2015, 1-year follow-up	0.920	0.918	0.912 (0.905-0.918)	0.910 (0.904-0.917)	0.907 (0.900-0.914)	0.913 (0.907-0.920)	0.868 (0.857-0.878)	0.868 (0.857-0.879)	0.804 (0.792-0.815)	0.880 (0.872-0.889)
2011, 1-year follow-up	0.910	0.908	0.901 (0.894-0.908)	0.900 (0.892-0.907)	0.899 (0.892-0.906)	0.902 (0.895-0.909)	0.843 (0.831-0.856)	0.841 (0.829-0.854)	0.781 (0.768-0.793)	0.857 (0.847-0.867)
2011, 5-year follow-up	0.897	0.895	0.890 (0.886-0.894)	0.889 (0.885-0.892)	0.887 (0.883-0.890)	0.891 (0.887-0.894)	0.793 (0.787-0.800)	0.795 (0.788-0.801)	0.742 (0.736-0.748)	0.824 (0.819-0.830)

1. Using predictions from the score development model with the conditions as binary indicators, adjusted by age and gender.
2. Using predictions from a model including each score (sum of weights) adjusted by age and gender.
3. Using the score directly, without the use of a model (unadjusted).

Extended scores include all 37 conditions. Primary scores include only the 20 most important conditions according to prevalence and impact.
Confidence intervals are provided only for evaluation models and validation datasets.

Appendix S11

Model output for the Cox model for mortality

	37 conditions		20 conditions	
	HR (95% CI)	p-value	HR (95% CI)	p-value
Cancer (in last five	3.64 (3.31-4.01)	<0.001	3.60 (3.27-3.96)	<0.001
Dementia	2.54 (2.29-2.82)	<0.001	2.59 (2.33-2.87)	<0.001
COPD	2.34 (2.12-2.60)	<0.001	2.36 (2.13-2.61)	<0.001
Alcohol problems	2.00 (1.61-2.47)	<0.001	2.25 (1.85-2.75)	<0.001
Epilepsy (currently	2.02 (1.64-2.48)	<0.001	2.13 (1.74-2.61)	<0.001
Painful condition	1.61 (1.49-1.75)	<0.001	1.62 (1.50-1.76)	<0.001
Constipation (Treated)	1.59 (1.45-1.75)	<0.001	1.62 (1.47-1.78)	<0.001
Heart failure	1.53 (1.35-1.73)	<0.001	1.55 (1.37-1.75)	<0.001
Schizophrenia (and	1.37 (1.07-1.75)	0.012	1.41 (1.11-1.80)	0.005
Anxiety OR Depression	1.40 (1.28-1.52)	<0.001	1.41 (1.30-1.53)	<0.001
Diabetes	1.39 (1.27-1.51)	<0.001	1.41 (1.29-1.54)	<0.001
Atrial fibrillation	1.40 (1.27-1.55)	<0.001	1.38 (1.26-1.52)	<0.001
Stroke & transient	1.36 (1.23-1.50)	<0.001	1.36 (1.24-1.50)	<0.001
Chronic kidney disease	1.31 (1.20-1.43)	<0.001	1.31 (1.21-1.43)	<0.001
Coronary heart disease	1.09 (0.99-1.19)	0.068	1.09 (0.99-1.19)	0.071
Rheumatoid arthritis,	1.00 (0.87-1.15)	0.973	0.99 (0.86-1.14)	0.867
Hypertension	0.94 (0.87-1.02)	0.118	0.93 (0.87-1.01)	0.082
Irritable bowel	0.93 (0.81-1.07)	0.301	0.88 (0.77-1.01)	0.079
Hearing loss	0.87 (0.81-0.95)	0.001	0.87 (0.80-0.94)	0.001
Asthma (currently	0.85 (0.75-0.96)	0.012	0.84 (0.74-0.95)	0.005
Migraine	0.33 (0.11-1.03)	0.057		
Chronic sinusitis	0.99 (0.82-1.19)	0.879		
Anorexia or bulimia	1.81 (1.26-2.60)	0.001		
Peptic ulcer disease	1.15 (1.00-1.33)	0.058		
Parkinson's disease	1.73 (1.36-2.19)	<0.001		
Diverticular disease of	0.79 (0.70-0.88)	<0.001		
Psychoactive substance	1.32 (0.94-1.86)	0.106		
Chronic Liver Disease	2.56 (1.95-3.35)	<0.001		
Peripheral vascular	1.25 (1.08-1.46)	0.004		
Bronchiectasis	1.16 (0.88-1.54)	0.302		
Blindness and low vision	1.02 (0.87-1.19)	0.796		
Thyroid disorders	0.97 (0.86-1.08)	0.553		
Inflammatory bowel	0.97 (0.70-1.33)	0.843		
Psoriasis or eczema	0.94 (0.74-1.18)	0.580		
Learning disability	2.83 (1.86-4.30)	<0.001		
Prostate disorders	0.78 (0.69-0.88)	<0.001		
Multiple sclerosis	1.67 (1.05-2.66)	0.031		
Male	1.36 (1.26-1.48)	<0.001	1.32 (1.23-1.42)	<0.001
Age at index date (10	2.41 (2.17-2.68)	<0.001	2.36 (2.13-2.62)	<0.001
Age 21/30	1.77 (0.99-3.18)	0.055	1.67 (0.93-3.00)	0.084
Age 31/40	1.26 (0.83-1.90)	0.277	1.22 (0.81-1.84)	0.348
Age 41/50	0.97 (0.74-1.28)	0.845	0.96 (0.73-1.26)	0.773
Age 61/70	0.86 (0.71-1.06)	0.152	0.87 (0.71-1.06)	0.170
Age 71/80	0.76 (0.58-0.98)	0.035	0.76 (0.59-0.98)	0.038
Age 81/max	0.85 (0.60-1.20)	0.352	0.85 (0.60-1.21)	0.362

Appendix S12

P-values from the statistical tests on the equivalence between the C statistics from the Primary outcome-specific scores and Charlson (columns 4 vs 5 and 8 vs 9 of Appendix S6, S8 and S10).

	Adjusted scores			Unadjusted scores		
	Number of Consultations	Mortality	Emergency hospital admissions	Number of Consultations	Mortality	Emergency hospital admissions
2015, 1-year follow-up	<0.001	0.001	<0.001	<0.001	<0.001	<0.001
2011, 1-year follow-up	<0.001	0.562	<0.001	<0.001	<0.001	<0.001
2011, 5-year follow-up	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

Appendix S13

Calibration plots for the Primary outcome-specific scores.