

# Comparing measures of multimorbidity to predict outcomes in primary care: a cross sectional study

Samuel L Brilleman and Chris Salisbury\*

Academic Unit of Primary Care, School of Social and Community Medicine, University of Bristol, Canynge Hall, 39 Whatley Road, Bristol BS8 2PS, UK.

\*Correspondence to Chris Salisbury, Academic Unit of Primary Care, School of Social and Community Medicine, University of Bristol, Canynge Hall, 39 Whatley Road, Bristol BS8 2PS, UK; E-mail: [c.salisbury@bristol.ac.uk](mailto:c.salisbury@bristol.ac.uk)

Received 24 November 2011; Revised 29 July 2012; Accepted 20 August 2012.

**Background.** An increasing proportion of people are living with multiple health conditions, or 'multimorbidity'. Measures of multimorbidity are useful in studies of interventions in primary care to take account of confounding due to differences in case-mix.

**Objectives.** Assess the predictive validity of commonly used measures of multimorbidity in relation to a health outcome (mortality) and a measure of health service utilization (consultation rate).

**Methods.** We included 95372 patients registered on 1 April 2005 at 174 English general practices included in the General Practice Research Database. Using regression models we compared the explanatory power of six measures of multimorbidity: count of chronic diseases from the Quality and Outcomes Framework (QOF); Charlson index; count of prescribed drugs; three measures from the John Hopkins ACG software [Expanded Diagnosis Clusters count (EDCs), Adjusted Clinical Groups (ACGs), Resource Utilisation Bands (RUBs)].

**Results.** A model containing demographics and GP practice alone explained 22% of the uncertainty in consultation rates. The number of prescribed drugs, ACG category, EDC count, RUB category, QOF disease count, or Charlson index increased this to 42%, 37%, 36%, 35%, 30%, and 26%, respectively. Measures of multimorbidity made little difference to the fit of a model predicting 3-year mortality. Nonetheless, Charlson index score was the best performing measure, followed by the number of prescribed drugs.

**Conclusion.** The number of prescribed drugs is the most powerful measure for predicting future consultations and the second most powerful measure for predicting mortality. It may have potential as a simple proxy measure of multimorbidity in primary care.

**Keywords.** Comorbidity, family practice, mortality, outcome assessment–health care, risk adjustment.

## Introduction

The focus of primary health care in developed countries now largely relates to supporting the management of long-term conditions. As the population ages, an increasing proportion of people are living with multiple health conditions, or 'multimorbidity'. There is a growing body of evidence about the epidemiology and consequences of multimorbidity, and a number of different measures of multimorbidity have been used in these studies.<sup>1–4</sup> Measures of multimorbidity are useful in both epidemiological and experimental studies of interventions in primary care, in order to take account of confounding due to differences in case-mix.

Multimorbidity can be conceptualized and measured in several different ways. Some measures are based on simple counts of patients with multiple health conditions (which raises questions about the range of conditions or diseases that should be included), while other measures differentially weight diseases or use other approaches to take account of the burden of illness on an individual.<sup>5,6</sup> Most commonly used measures of multimorbidity were originally developed and validated among selected patients in hospital.<sup>6,7</sup> The measures were also often developed to demonstrate relationships with a specific outcome, such as mortality,<sup>8</sup> but have since been used in studies with very different outcomes.<sup>6</sup> It cannot be assumed that measures of multimorbidity will be equally valid in populations encountered in primary

care settings, or are able to predict outcomes other than those for which they were developed.

Several previous studies have compared the performance of different measures of multimorbidity to predict a specific outcome (such as mortality) in a specific population (such as people with hypertension). However, few previous studies have directly compared the performance of different measures of multimorbidity in relation to more than one outcome in a general primary care or community-dwelling population. These studies have been restricted to elderly populations<sup>9,10</sup> except for one study that explored the ability of five measures (the Charlson index, Elixhauser index, Chronic Disease Score, number of diagnoses and number of drugs prescribed) to predict mortality and hospitalization in adults living in Saskatchewan, Canada.<sup>11</sup> This study was based on linked administrative data such as billing and prescribing records.

The purpose of this paper is to assess the predictive validity in primary care of commonly used measures of multimorbidity in relation to two different types of outcome: a health outcome (mortality) and a measure of health service utilization (primary care consultation rate). We also wanted to assess whether one measure provided greatest explanatory power in relation to both outcomes or whether different measures were needed for different outcomes.

## Methods

### *Sample*

This study was based on routine anonymised electronic medical records from patients registered with practices contributing to the General Practice Research Database (GPRD), which is now incorporated in the Clinical Practice Research Datalink (<http://www.cprd.com>). The GPRD contains the primary care records of around 5 million patients and is considered broadly representative of the general population in the UK.<sup>12</sup> We used a stratified random sample of 99 997 individuals aged 18 years and over and registered at one of 182 practices on 1 April 2005 and obtained data on all consultations, diagnoses and prescriptions until 31 March 2008. The sample was stratified by age, gender and practice. After excluding eight practices in which data about patient deprivation were unavailable, 95 372 patients from 174 practices were used for analysis.

### *Measures of multimorbidity*

Multimorbidity was operationalized in six ways. The first three multimorbidity indices discussed here used all historic diagnoses up until 1 April 2005.

First, we constructed a simple count of 17 diseases included in the Quality and Outcomes Framework (QOF)<sup>13</sup> that have been classified as chronic in previous research.<sup>1</sup> To identify whether an individual had a

relevant disease we employed Version 11 of the QOF Business Rules.<sup>14</sup>

Second, we used the Charlson index score, in the recent adaptation for use with Read coded data.<sup>15</sup> The Charlson index is a weighted score that includes 17 chronic diseases, where the weighting associated with each disease reflects the strength of its association with inpatient mortality.<sup>8</sup>

The next three multimorbidity measures were constructed using output from the John Hopkins University ACG Case-Mix System, which was developed using administrative data in the USA.<sup>16</sup> The software allows the user to input Read Version 2 diagnostic codes. As our third measure of multimorbidity we used a count of Expanded Diagnosis Clusters (EDCs), which are groupings of diagnostic codes based on disease or organ area. The software identifies 264 separate EDCs, of which 114 were classified as chronic during work previously described.<sup>1</sup> We counted the number of chronic EDCs ever recorded for each individual.

Our fourth multimorbidity measure used Adjusted Clinical Groups (ACGs), which are mutually exclusive categories defined by age, gender and combination of Aggregated Diagnosis Groups (ADGs). ADGs are groupings of diagnostic codes based on clinical dimensions that determine the subsequent expected resource use. Examples of the dimensions used to classify diagnoses into the ADGs are severity of the condition, expected duration of the condition, and the need for specialty care involvement. The ACG classifications used diagnoses recorded over a 1-year period (1 April 2004 to 31 March 2005), as suggested by the software manual.<sup>16</sup> The age range of our sample meant that a total of 68 ACG categories were populated (although the version of the software used identified up to 82 default categories, some of these were age specific). The fifth measure of multimorbidity used Resource Utilisation Bands (RUBs), which are an aggregation of ACGs into six ordinal categories, based solely on expected health care resource demands.

For the sixth measure we used a count of the number of drugs prescribed to an individual in the period between 1 April 2004 and 31 March 2005 as a proxy measure of multimorbidity. In an earlier study from North America in older adults, a similar measure appeared to predict health care utilization better than the Charlson index or ACGs.<sup>10</sup> We counted the number of unique British National Formulary (BNF)<sup>17</sup> codes appearing in the individual's prescription drug data. Each code represents one sub-heading within the BNF and includes drugs that are in the same class (for example thiazide diuretics have one code and loop diuretics have another). Therefore, repeated prescriptions of the same or very similar medication, including different doses or formulations, were only counted once.

### *Analyses*

We investigated the explanatory power of each multimorbidity measure in relation to two outcome

measures: (i) number of consultations and (ii) mortality. Both outcomes were measured over the 3-year period beginning 1 April 2005. Using 3 years of data improved the reliability of estimates, particularly for mortality, which is a rare event. We used the likelihood ratio test (backward selection with a significance level of 0.01) to determine which covariates to include in the baseline model. We considered age, sex, age-by-sex interactions, deprivation and GP practice effects. Age, at 1 April 2005, was categorized into 10-year age bands with the exception of 18–29 and 90+ as lower and upper categories, respectively. Deprivation was categorized into 10 deciles based on the Index of Multiple Deprivation (IMD) 2007, which incorporates seven dimensions of deprivation and relates to the individual's Lower Layer Super Output Area (LSOA).<sup>18</sup>

Each multimorbidity measure was included in the model as a categorical variable to allow for possible non-linearity in the relationship between multimorbidity and the relevant outcome. For each measure we collapsed the uppermost categories to ensure there were enough individuals in each cell to allow estimation of the parameters. Spearman rank correlations were used to assess the level of agreement between the numerical measures of multimorbidity used in the models.

The number of consultations over the 3 years following 1 April 2005 was modelled using Generalised Linear Models (GLMs) assuming a log link function and negative binomial distribution. We included consultations with GPs and practice nurses, both face-to-face and by telephone. For any individual who was not observed for the entire 3-year period due to death or attrition (e.g. transferring out of the practice) we applied a weighting to the observed number of consultations (the weighting was equal to the inverse of the proportion of the 3-year period during which they were observed).

Mortality was modelled using logistic regression models where the outcome was presence, or absence, of death during the 3 years following 1 April 2005. We considered using a Cox proportional hazards regression model; however, the extremely high proportion of censoring in our data (more than 96% of the sample did not die during the three-year period) meant that standard measures of model fit, such as R-squared based measures, would have been unreliable.<sup>19</sup> As age is such a strong predictor of mortality we also ran the mortality regression models using a subset of the sample, namely individuals aged 65 years and over. In this model we included age in years as a categorical variable (i.e. one coefficient estimated for each year rather than each 10-year age band). The results from this analysis were then compared with the results from the logistic regression model using the full sample.

Models were compared using Akaike's Information Criterion (AIC)<sup>20</sup> and Bayesian Information Criterion (BIC),<sup>21</sup> for both of which a smaller value provides an

indication of a better fitting model. The BIC provides a greater penalty for additional parameters in the model than the AIC. We also calculated a deviance-based R-squared measure<sup>22</sup> that may be interpreted as the proportion of uncertainty in the relevant outcome that has been explained by the model.

#### *Software*

We used John Hopkins ACG System Version 8.2 to obtain the EDC, ACG and RUB classifications<sup>16</sup> and STATA Version 11.2 for the statistical analysis.

## Results

### *Summary statistics*

Of the 95 372 individuals in the original sample, 95 188 (99.8%) were included in the complete case analysis after excluding 184 individuals with missing deprivation data. The mean age was 49 years (SD 18 years) and 51% of the sample was female.

**Table 1** shows descriptive statistics for each outcome in relation to age, sex and deprivation. The number of consultations per patient per annum ranged from 0 to 113 with mean, SD and median of 4.3, 5.0 and 3, respectively. Of the 95 372 people in the full sample, 3198 (3.4%) died and 10081 (10.6%) left their practice during the 3-year observation period.

**Figure 1** shows the frequency distribution of each measure of multimorbidity (with the exception of the ACG categories). The vertical axis shows the percentage of the sample at each level or category of multimorbidity. The QOF disease count and Charlson index each show an extremely positively skewed distribution, with the majority of people having a count or score of zero. The EDC count and number of prescribed drugs are also positively skewed, but have a greater number of categories and a more even distribution. In terms of Resource Utilisation Bands very few people were in the high utilisation bands. Of the 68 populated ACG categories, the largest 12 categories contained over 80% of the sample.

**Table 2** provides Spearman rank correlations between the numerical measures of multimorbidity used in the models. There is stronger agreement between the QOF disease count, EDC count and number of prescribed drugs ( $r$  ranging from 0.60 to 0.65) than there is with the Charlson index score ( $r$  ranging from 0.42 to 0.49).

### *Number of consultations*

**Table 3** gives the deviance-based R-squared, AIC and BIC values for the different models predicting the three-year consultation rate. A model containing age, sex, age-by-sex interaction, deprivation and GP practice alone explained 22% of the observed uncertainty. Inclusion of the number of prescribed drugs, ACG category, EDC count, RUB category, QOF disease count,

TABLE 1 *Descriptive statistics on the number of consultations per patient per annum, and the number of deaths, over a 3-year period*

Variable	Total frequency	Number of consultations per patient per annum, mean (SD) [median]	Number of deaths, <i>n</i> (%)
<b>Age category</b>			
18–29	16 099	2.7 (3.4) [2]	27 (0.2%)
30–39	17 338	3.0 (3.7) [2]	33 (0.2%)
40–49	18 059	3.3 (4.1) [2]	89 (0.5%)
50–59	16 220	4.2 (4.5) [3]	232 (1.4%)
60–69	12 469	5.8 (5.5) [5]	411 (3.3%)
70–79	9227	7.5 (6.3) [6]	823 (8.9%)
80–89	4957	7.6 (7.0) [6]	1137 (22.9%)
90+	1003	5.6 (5.9) [4]	446 (44.5%)
<b>Gender</b>			
Male	46 768	3.4 (4.6) [2]	1499 (3.2%)
Female	48604	5.1 (5.2) [4]	1699 (3.5%)
<b>Deprivation</b>			
1 (least)	9625	4.1 (4.6) [3]	236 (2.5%)
2	9417	4.1 (4.7) [3]	288 (3.1%)
3	9654	4.2 (5.0) [3]	298 (3.1%)
4	9390	4.1 (4.6) [3]	298 (3.2%)
5	9512	4.0 (4.8) [3]	271 (2.8%)
6	9537	4.4 (5.1) [3]	322 (3.4%)
7	9554	4.4 (5.2) [3]	341 (3.6%)
8	9464	4.3 (5.0) [3]	335 (3.5%)
9	9541	4.5 (5.2) [3]	369 (3.9%)
10 (most)	9494	4.7 (5.5) [3]	429 (4.5%)
Overall	95 372	4.3 (5.0) [3]	3198 (3.4%)

Total number of subjects with non-missing deprivation data is 95 188. Data source: GPRD which has now been incorporated into the Clinical Practice Research Datalink (<http://www.cprd.com>).

or Charlson index score increased this to 42%, 37%, 36%, 35%, 30% and 26%, respectively. Both AIC and BIC showed the same pattern of support for the models as the deviance-based R-squared values.

### Mortality

Table 4 gives the deviance-based R-squared, AIC and BIC values for the logistic regression models predicting the probability of death within 3 years. The model containing the ACG categories could not be reliably estimated as several of the less populated ACG categories had no observed deaths. A model containing age, sex and deprivation alone explained 28% of the uncertainty. The R-squared value suggested there was very little improvement from including any of the measures of multimorbidity, with gains of no more than 3% over the baseline model. Further investigation showed that a model with age category alone (R-squared = 27.5%, further results not shown) suggested that age was the major contributor to any uncertainty explained in the outcome. Nonetheless, the deviance-based R-squared, AIC and BIC values all showed the Charlson index

score as the highest ranked model. This was followed by a model containing the number of prescribed drugs. When considering ranking of the models beyond that point, there is no clear distinction in R-squared values, nor is there complete agreement between AIC and BIC values.

Fitting the logistic regression models using only individuals aged 65 years and over showed similar results. The model which included the Charlson index score performed best, followed by the number of prescribed drugs. Even using this limited age range, improvements in the deviance-based R-squared values due to the inclusion of a multimorbidity measure were still small (ranging between 1.3% and 3.5%).

## Discussion

### Summary of main findings

This research has demonstrated that all of the multimorbidity measures had moderate predictive validity in relation to consultation rates in primary care, with the number of prescribed drugs having the greatest predictive validity followed by the three ACG based measures. With regard to mortality, none of the measures added greatly to the fit of a model containing demographic variables, but the Charlson index performed best, followed by the number of prescribed drugs. The other diagnosis-based measures of multimorbidity (the ACG-related measures and the number of chronic conditions included in the QOF) added little to the model.

The Charlson index was originally designed to predict mortality in hospitalized patients, and these results confirm that it is similarly valid in the community population. It may not be surprising that the number of drugs predicts consultation rates, since people who consult more often may also be prescribed more drugs, but it is notable that this measure also predicts mortality better than diagnosis-based measures other than the Charlson score. This suggests that this relatively simple measure may be a useful proxy measure for primary care-based studies that are exploring a range of outcomes and need a case-mix measure to account for multimorbidity amongst participants.

### Comparison with previous studies

Our findings build on previous research by enabling the comparison of several different approaches to measuring multimorbidity in the same population, and comparing their predictive validity in relation to both a process and an outcome variable. Our research is also based on a much larger and more reliable dataset than most previous studies.

Several previous studies have confirmed that the Charlson index is a valid measure to predict mortality amongst community populations.<sup>9,10,11,23</sup> Some recent studies have suggested that the Elixhauser index<sup>24</sup> may

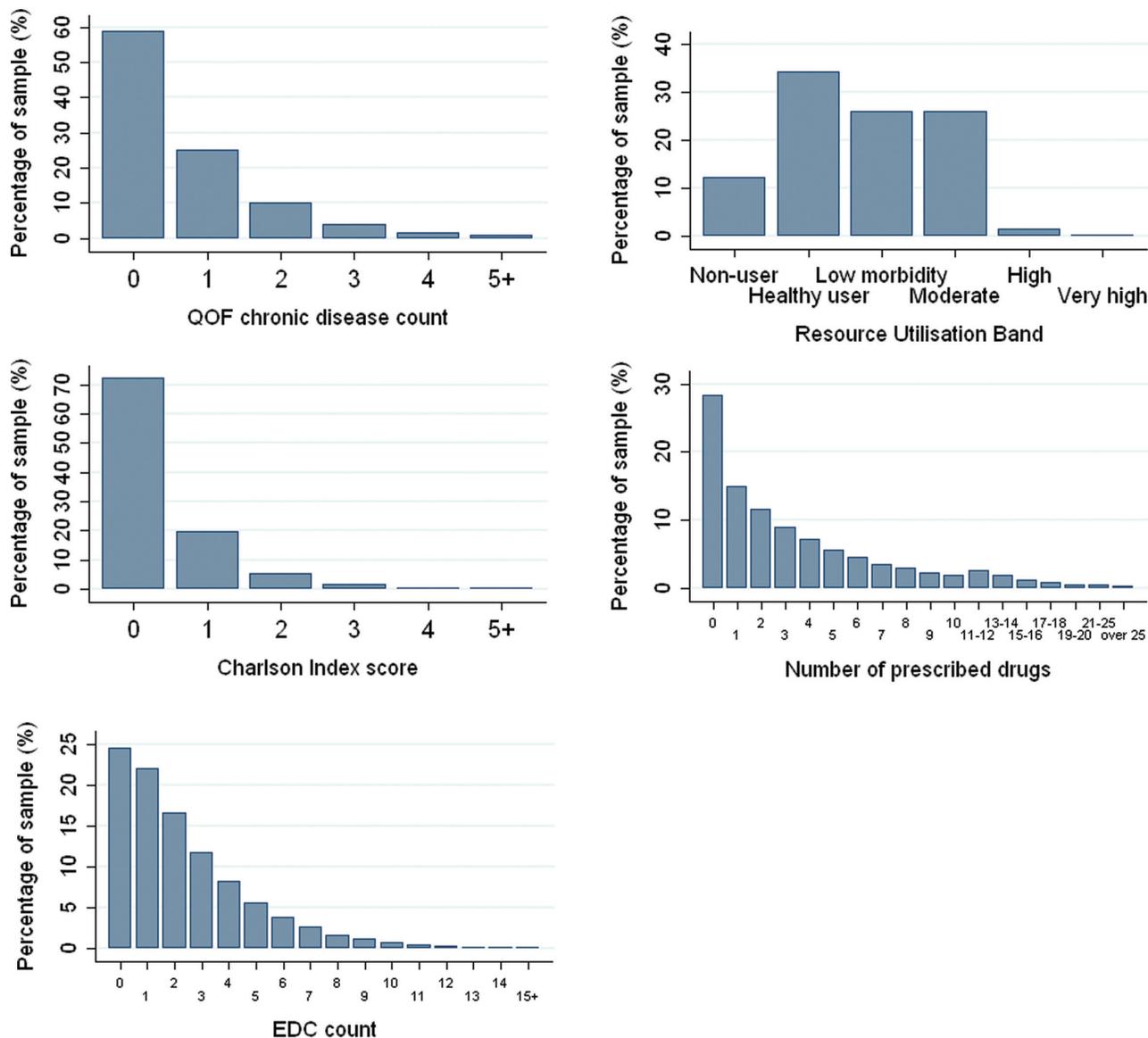


FIGURE 1 Distribution of the QOF disease count, Charlson index score, EDC count, Resource Utilisation Bands, and number of prescribed drugs. Percentage of the sample is shown on the vertical axis.

be superior to the Charlson index<sup>11,25,26</sup> but this is calculated using ICD codes and algorithms have not yet

TABLE 2 Spearman rank correlations for the numerical measures of multimorbidity

	QOF disease count	Charlson index score	Count of EDCs
QOF disease count	1.00	—	—
Charlson index score	0.46	1.00	—
Count of EDCs	0.65	0.49	1.00
Number of prescribed drugs	0.60	0.42	0.65

been developed for use with Read codes, so it was not included in this study.

Our findings are compatible with previous research in a range of populations on prediction of different aspects of health care utilization or health care costs. Both simple counts of number of distinct medications in the previous year and more sophisticated medication-based measures such as the Chronic Disease Score or RxRisk outperform the Charlson index in predicting health care utilization.<sup>10,11,27,28</sup>

*Strengths and limitations*

The GPRD provides a large, reliable and validated dataset of patient records in primary care, for a population that is broadly representative of England. Most previous research has been conducted in administrative databases

TABLE 3 *Model fit statistics for negative binomial regression models predicting 3-year consultation rate*

Model	Deviance-based $R^2$	AIC	BIC
No multimorbidity measure	21.9% (7)	648367 (7)	650373 (7)
QOF disease count	30.2% (5)	636666 (5)	638720 (5)
Charlson index score	25.9% (6)	642855 (6)	644908 (6)
Count of EDCs	36.1% (3)	627617 (3)	629766 (3)
ACG category	37.3% (2)	625798 (2)	628438 (2)
Resource Utilisation Band	34.7% (4)	629809 (4)	631863 (4)
Number of prescribed drugs	41.6% (1)	618632 (1)	620799 (1)

All models include age category, sex, age-by-sex interaction, deprivation (based on IMD2007) and GP practice as fixed effects.

The associated ranking of each model is shown in parentheses, with 1 representing the best performing model and 7 representing the worst performing model.

For AIC and BIC a smaller value indicates better model fit.

TABLE 4 *Model fit statistics for logistic regression models predicting the probability of death over a 3-year period*

Model	Deviance-based $R^2$	AIC	BIC
No multimorbidity measure	28.2% (6)	20079 (6)	20250 (6)
QOF disease count	29.8% (3)	19636 (3)	19854 (3)
Charlson index score	31.3% (1)	19226 (1)	19443 (1)
Count of EDCs	29.8% (3)	19667 (4)	19979 (5)
Resource Utilisation Band	29.5% (5)	19728 (5)	19946 (4)
Number of prescribed drugs	30.9% (2)	19361 (2)	19693 (2)

All models include age category, sex, and deprivation (based on IMD2007) as fixed effects.

The associated ranking of each model is shown in parentheses, with 1 representing the best performing model and 7 representing the worst performing model.

For AIC and BIC a smaller value indicates better model fit.

of uncertain reliability and/or on specific sub-populations. The findings from this study are therefore likely to provide a robust indication of the validity of these measures and be applicable to other primary care research in England and possibly in other developed countries.

There may be limitations due to how the measures were constructed. Both the Charlson index and the ACG measures are based on ICD codes, and although algorithms have been developed to calculate them from Read coded data, few previous studies have used them in this way. The findings may also be sensitive to decisions made in operationalizing measures. For example, we calculated a medication measure based on the number of prescriptions for drugs with different BNF codes, while other researchers have counted the number of

prescription items.<sup>10</sup> Although the number of different drugs (rather than the number of prescription items) may be a better proxy for the number of conditions an individual has been diagnosed with, the number of repeated prescriptions of a drug may be more likely to relate to chronic conditions. Similarly, the ACG software can be used in several ways. For example, we could have included diagnoses recorded over the previous 3 years rather than 1 year in calculating ACG categories or could have based our analyses on ADGs rather than ACGs. Therefore, it may be worthwhile in future studies to investigate the relative performance of different approaches to counting prescriptions or using the ACG software. Finally it should be noted that all measures of multimorbidity (and all epidemiological studies) based on routine data are reliant on whether diseases have been diagnosed and coded.

It is notable that none of the measures used in this study added very substantially to age alone in predicting mortality over the subsequent 3 years. Our findings are likely to be a consequence of the fact that in the population as a whole mortality is a rare event, and this may partly explain the contrast between our findings and previous research, which has been based on elderly populations.<sup>9,10,23</sup> However, we found similar results using a subset of the sample aged 65 years and over. Although multimorbidity measures are commonly used to adjust for case-mix in epidemiological and observational studies, these findings highlight that they have only a limited ability to control for confounding.<sup>27</sup>

It is notable that the number of prescribed drugs over the previous year is not only the most powerful predictor of primary care consultations over the subsequent 3 years but is also (after the Charlson index) the second most powerful predictor of mortality. This is particularly interesting given that the Charlson index was specifically designed to predict mortality, yet has limited additional predictive validity over a simple count of drugs prescribed. The latter measure is simple to use and does not require statistical expertise or proprietary software to calculate. Future research should explore whether this measure predicts other outcomes such as quality of life, self-rated health and functional ability.

## Conclusion

No single measure adds considerably to age alone in predicting mortality amongst the adult population consulting in primary care. All the measures included in this study have weak predictive validity for this purpose but the Charlson index is the most powerful. The number of prescribed drugs is the most powerful measure for predicting future health care consultations and is the second most powerful measure for predicting

mortality. It may have potential as a useful and simple proxy measure of multimorbidity in primary care.

## Acknowledgements

We would like to thank Johns Hopkins University for providing the ACG software. We are also grateful to Sandra Hollinghurst for providing comments on the final text, and Alan A Montgomery for advice relating to the statistical analysis.

## Declaration

**Funding:** National Institute for Health Research (NIHR) School for Primary Care Research Funding Scheme (66). This report presents independent research commissioned by the National Institute for Health Research. The views expressed in this publication are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. This study is based on data from the Full Feature General Practice Research Database obtained under licence from the UK Medicines and Healthcare products Regulatory Agency (MHRA). However, the interpretation and conclusions contained in this study are those of the author/s alone. Access to the GPRD database was funded through the Medical Research Council's licence agreement with MHRA.

**Ethical approval:** Studies based on the GPRD are covered by ethics approval granted by Trent Multicentre Research Ethics Committee, reference 05/MRE04/87.

**Conflict of interest:** none.

## References

- Salisbury C, Johnson L, Purdy S, Valderas JM, Montgomery AA. Epidemiology and impact of multimorbidity in primary care: a retrospective cohort study. *Br J Gen Pract* 2011; **61**: e12–e21.
- van den Akker M, Buntin F, Metsemakers JF, Roos S, Knotnerus JA. Multimorbidity in general practice: prevalence, incidence, and determinants of co-occurring chronic and recurrent diseases. *J Clin Epidemiol* 1998; **51**: 367–75.
- Fortin M, Bravo G, Hudon C, Vanasse A, Lapointe L. Prevalence of multimorbidity among adults seen in family practice. *Ann Fam Med* 2005; **3**: 223–8.
- van den Akker M, Buntinx F, Roos S, Knotnerus J. Problems in determining occurrence rates of multimorbidity. *J Clin Epidemiol* 2001; **54**: 675–9.
- Valderas JM, Starfield B, Sibbald B, Salisbury C, Roland M. Defining comorbidity: implications for understanding health and health services. *Ann Fam Med* 2009; **7**: 357–63.
- Huntley A, Johnson R, Purdy S, Valderas JM, Salisbury C. Measures of multimorbidity and morbidity burden for use in primary care and community settings: a systematic review and guide. *Ann Fam Med* 2012; **10**: 134–41.
- de Groot V, Beckerman H, Lankhorst GJ, Bouter LM. How to measure comorbidity: a critical review of available methods. *J Clin Epidemiol* 2003; **56**: 221–9.
- Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis* 1987; **40**: 373–83.
- Di Bari M, Virgilio A, Matteuzzi D *et al*. Predictive validity of measures of comorbidity in older community dwellers: The Insufficienza Cardiaca negli Anziani Residenti a Dicomano study. *J Am Geriatr Soc* 2006; **54**: 210–12.
- Perkins AJ, Kroenke K, Unutzer J *et al*. Common comorbidity scales were similar in their ability to predict health care costs and mortality. *J Clin Epidemiol* 2004; **57**: 1040–8.
- Quail JM, Lix LM, Osman BA, Teare GF. Comparing comorbidity measures for predicting mortality and hospitalization in three population-based cohorts. *BMC Health Serv Res* 2011; **11**: 146.
- Lawrenson R, Williams T, Farmer R. Clinical information for research; the use of general practice databases. *J Public Health Med* 1999; **21**: 299–304.
- NHS Employers. *Quality and Outcomes Framework Guidance for GMS Contract 2009/10*. [http://www.nhsemployers.org/Aboutus/Publications/Documents/QOF\\_Guidance\\_2009\\_final.pdf](http://www.nhsemployers.org/Aboutus/Publications/Documents/QOF_Guidance_2009_final.pdf) (accessed on 11 September 2012).
- NHS Primary Care Contracting. *QOF Implementation Business Rules v11*. <http://www.Primarycarecontracting.nhs.uk/145.php> (accessed on 11 September 2012).
- Khan NF, Perera R, Harper S, Rose PW. Adaptation and validation of the Charlson Index for Read/OXMIS coded databases. *BMC Fam Pract* 2010; **11**: 1.
- Johns Hopkins Bloomberg School of Public Health. *The Johns Hopkins ACG Case-Mix System Version 8.2*. Baltimore, 2008.
- BNF. *British National Formulary 2011*. <http://bnf.org/bnf/index.htm> (accessed on 11 September 2012).
- Department for Communities and Local Government. *The English Indices of Deprivation 2007*. London: HMSO, 2008.
- Gillespie BW. Use of generalized R-squared in Cox regression. [http://apha.confex.com/apha/134am/techprogram/paper\\_135906.htm](http://apha.confex.com/apha/134am/techprogram/paper_135906.htm) (accessed on 11 September 2012).
- Akaike H. New look at statistical-model identification. *IEEE Transactions on Automatic Control* 1974; **AC19**: 716–23.
- Schwarz G. Estimating dimension of a model. *Annals of Statistics* 1978; **6**: 461–4.
- Cameron AC, Windmeijer FAG. An R-squared measure of goodness of fit for some common nonlinear regression models. *J Economet* 1997; **77**: 329–42.
- Schneeweiss S, Avorn N, Maclure M, Levin R, Glynn R. Consistency of performance ranking of comorbidity adjustment scores in Canadian and U.S. utilization data. *J Gen Intern Med* 2004; **19**: 444–50.
- Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. *Med Care* 1998; **36**: 8–27.
- Southern DA, Quan H, Ghali WA. Comparison of the Elixhauser and Charlson/Deyo methods of comorbidity measurement in administrative data. *Med Care* 2004; **42**: 355–60.
- Chu YT, Ng YY, Wu SC. Comparison of different comorbidity measures for use with administrative data in predicting short- and long-term mortality. *BMC Health Serv Res* 2010; **10**: 140.
- Schneeweiss S, Seeger JD, Maclure M *et al*. Performance of comorbidity scores to control for confounding in epidemiologic studies using claims data. *Am J Epidemiol* 2001; **154**: 854–64.
- Farley JF, Harley CR, Devine JW. A comparison of comorbidity measurements to predict healthcare expenditures. *Am J Manag Care* 2006; **12**: 110–19.