Public Health England

Protecting and improving the nation's health

The health and social care costs of a selection of health conditions and multi-morbidities

About Public Health England

Public Health England exists to protect and improve the nation's health and wellbeing and reduce health inequalities. We do this through world-leading science, research, knowledge and intelligence, advocacy, partnerships and the delivery of specialist public health services. We are an executive agency of the Department of Health and Social Care, and a distinct delivery organisation with operational autonomy. We provide government, local government, the NHS, Parliament, industry and the public with evidence-based professional, scientific and delivery expertise and support.

Public Health England
Wellington House
133-155 Waterloo Road
London SE1 8UG
Tel: 020 7654 8000
www.gov.uk/phe
Twitter: @PHE_uk
Facebook: www.facebook.com/PublicHealthEngland

For queries relating to this document, please contact: healtheconomics@phe.gov.uk

Published July 2020
PHE publications
gateway number: GW-1360

PHE supports the UN
Sustainable Development Goals

Authors

Public Health England
Aleksandra Blawat
Emily Green
Annalisa Belloni
Abbygail Jaccard
Vandyck Adomako-Mensah

Imperial College London
Laure de Preux
Sophie Thiébaut
Franco Sassi

UK Health Forum
Laura Webber
Lise Retat

With thanks to:

Nessa Cawte, NHS South, Central & West Commissioning Support Unit

Contents

Glossary

Adjusted clinical groups (ACGs)	The index assigns each individual to a single, mutually exclusive group, based on expected or actual consumption of health services.
Adult Comorbidity Evaluation (ACE)	A scoring (0–3) instrument based on the number and severity of medical comorbidities. The scores range from 0 (no comorbidity) to 3 (severe).
Analysis by number of diseases	Cost analyses based on the number of multi-morbidities rather than index or specific disease.
Body system	Parts of the body which have a specific function, for example: 1. Cardiac (heart only); 2. Vascular (blood, blood vessels and cells, marrow, spleen. lymphatics); 3. Respiratory (lungs, bronchi, trachea below the larynx); 4. ENT (eye, ear, nose, throat, larynx).
Charlson's comorbidity index	The index consists of 22 conditions and is based on the mortality risk for a patient. Each condition is assigned a score of 1, 2, 3, or 6 depending on the risk of mortality. Scores are summed to predict mortality. Variations on this scoring system exist.
Compound comorbidities (CCMs)	Index based on prevalence, odd ratios and costs of diseases.
Cost-per-case	Annual healthcare cost per case, based on sample prevalence in the dataset.
Cumulative illness rating scale	Investigates 14 body systems and rates them using a 5-grade system, where 0 implies no impairment and 5 represents severe life-threatening impairment.
Disease weights or index	A numerical relative importance is given to a disease or group of diseases based on either its contribution to mortality or complexity of care, for example.
Elixhauser index	31 level score determining the level of comorbidity depending on the International Classifications of Diseases (ICD) codes.
Index diseases	May also be termed a 'dominant disease'. This is the primary disease that the individual starts with which increases their chances of getting other diseases or 'co-morbidities'.
Prevalence case	A person in the sample dataset who has been diagnosed with one of the diseases included in the analysis, with diagnosis defined by i) having seen a GP for the condition in 2015 (Definition A) or ii) having seen a doctor for the condition in 2012-2015 (Definition B).
Rx-defined morbidity groups (Rx-MG)	Individuals grouped in terms of their medication use.

Sample prevalence	The proportion of people in the sample dataset who have been diagnosed with one of the diseases included in the analysis (see ‘prevalence case’).

Executive summary

Background

Multimorbidity (MM) is the presence of 2 or more long-term health conditions in a single individual. It impacts an individual's quality of life, mental health and wellbeing, daily function, and often results in greater healthcare utilisation the more co-existing conditions they have (1-4). MM is a big challenge facing the NHS, especially given England's ageing population, with an estimated two-thirds of individuals aged 65 and over having 2 or more long-term conditions (5-6). Yet, little is known about the resource use of these patients despite being the group with the largest impact on the NHS and with the worst health outcomes (7). Existing evidence focuses on specific health conditions and their interactions with other conditions using different methodologies, making comparisons across different conditions difficult.

This work has empirically assessed the impact of multi-morbidity on NHS and social care costs. With the aim of answering the question: is the impact of developing a condition on health and social care costs greater for someone with no prior conditions, or for someone with an existing condition. If patients have multiple conditions, there may be some economies of scale involved with treatment, for example they may be able to discuss multiple queries during a single GP appointment, or in some cases the treatment provided will address multiple conditions. However, treating patients with multi-morbidities could theoretically also be more expensive than treating 2 conditions separately, as patients may be more likely to experience complications.

Methodology

This work considered the individual cost of 11 health conditions with high prevalence in the English population and their most common interactions. These were: chronic obstructive pulmonary disease (COPD), diabetes (types 1 and 2), lung cancer, breast cancer, coronary heart disease (CHD), stroke, hypertension, dementia, liver disease, depression and colorectal cancer.

This project had 2 components: a literature review and an empirical estimation of the costs associated with MM. The literature review was used to inform and establish the methodology used in the empirical estimation.

The empirical estimation used data on primary healthcare, secondary healthcare, and prescriptions usage from 2015 to estimate annual aggregated healthcare costs per patient. We assessed the cost impact of MM in a systematic way by applying advanced

econometric methods to account for the specificities of the data distribution. Our methodology allowed us to attribute healthcare costs to specific conditions.

For social care costs, we calculated the estimated costs using 2 different methodologies.

For the first (preferred) methodology, we used Somerset Symphony data to calculate the 2014/15 social care costs of patients in South Somerset. This is a dataset that combines primary healthcare, secondary healthcare, and social care data. We thus applied the same methodology that was used to calculate primary and secondary healthcare costs.

For the second methodology, we used the estimated health-related quality of life for patients with different conditions and combinations of conditions. We then used a regression ('line of best fit') to estimate their probability of requiring social care. Finally, we used unit cost estimates to arrive at estimated values for the costs of social care for individuals with different diseases.

What this publication adds

Average 'cost per case' estimates for individuals with single conditions or multi-morbidities, each calculated based on the average age of patients with the condition or multi-morbidity of interest.

These average 'costs per case' figures are always higher for individuals with multi-morbidities than individuals with a single condition, as individuals with multi-morbidities tend to be older and additional conditions incur additional costs.

We found that the cost of treating an individual with a multimorbidity is not statistically different than the additive cost of treating 2 individuals, each with one of the conditions, controlling for age and costs unrelated to the condition. As an illustrative example, if it costs £200 to treat a patient with depression and £200 to treat a patient with CHD, we did not find any evidence that it would cost more than £400 to treat a single patient with both depression and CHD (controlling for age and unrelated disease costs).

In numerous cases, when considering healthcare costs, we have found that multi-morbidity is associated with a reduction of the total individual cost compared to the sum of individual costs of patients. For example, a male patient with diabetes and CHD will cost between 77% and 78% (depending on the definition of sample prevalence) of the cost of treating 2 patients, one with diabetes and one with CHD, controlling for age and unrelated costs.

Applying the same methodology for social care costs as for healthcare costs, we did not find any evidence that multi-morbidity is associated with either an increase or a reduction in total individual cost compared to the sum of individual costs of patient, for social care costs. This may be due to the relatively small sample size of the South Somerset data we used to estimate social care costs.

Applying the alternative methodology for social care costs, which estimated social care need based on age and quality of life, we estimated higher social care costs than we found by analysing the South Somerset data. This implies that social care need may be greater than local authority social costs in South Somerset. This may be due to the relative affluence of South Somerset, which would limit the proportion of patients eligible for local authority-funded social care.

Background

The concept of multi-morbidity (MM) is very broad in the literature, varying from a limited to an unlimited set of conditions, and including acute and long-term conditions. It can be captured as a number of conditions, or an index of severity, or a combination of both. Furthermore, it often overlaps with the concept of co-morbidity, which is generally defined as having an increased probability of developing a condition because of the presence of another, 'dominant' or 'index' disease. However, co-morbidity is better suited in the context of a specific health condition under study (the index disease), with the co-morbidity capturing the conditions an individual suffers from in addition to the index health condition (8).

In this project, we refer to MM as the presence of 2 or more long-term medical conditions in one individual (8,9). This definition was used in a recent meta-analysis by Wang et al., 2018 (10), and assumes a differentiation between comorbidity and multi-morbidity. Other definitions of MM in a cost analysis context are described in the review by Sambamoorthi et al., 2015 (11), who defined MM by considering the concepts of index disease (a dominant condition triggering other potential conditions) and weighted scores, defined in the glossary. Weighted scores summarise multiple conditions by various index scores based on, for example, the complexity of care or type of prescriptions (12,13). This method requires assumptions about the weights and is difficult to translate into a different context as no comparison is available. Our definition of MM allows generalisability and comparability with other evidence.

Multimorbidity (MM) impacts quality of life, mental health and wellbeing daily function, and often results in greater healthcare utilisation the more co-existing conditions an individual has (1-4). MM is a big challenge facing the NHS, especially given England's ageing population, with an estimated two-thirds of people aged 65 and over having 2 or more long-term conditions (5-6). Whitty et al (2020) (14) recommend that medical specialisation needs to adapt to the needs of patients with MM, moving away from treating each condition in isolation. Yet, little is known about the resource use of these patients despite being the group with the largest impact on the NHS and with the worst health outcomes (7). Existing evidence focuses on specific health conditions and their interactions with other conditions using different methodologies, making a cross-condition comparison difficult.

Gaining a better understanding of how researchers have tackled the challenge of costing the healthcare utilisation of patients with a MM is achieved by reviewing the methods used in the literature. Work was conducted for the PHE Cost of Air Pollution tool to estimate the cost per case figures of the conditions included in the tool (15). Building on this work, we explored existing methodologies to assess the best way to estimate the cost per case of individual conditions to the NHS and social care. This

exploration aimed to develop the most suitable methodology for costing multi-morbidity by using the richest set of individual healthcare utilisation data comprising of primary and secondary care, as well as prescriptions. Separately, the best source of social care data was used to replicate the methodology and assess the impact of multimorbidity.

The aims of this project were to:

- review the literature and describe the different costing methods used to quantify the costs of MM (part 1)
- calculate the annual cost-per-case figures of a selected range of individual conditions and multi-morbidities (part 2)

Part 1: Literature review of costing methods

Method

A literature review was undertaken with 3 key objectives which were to:

- review the methods used to quantify MM costs, and the strengths and limitations associated with these methods
- inform the method for costing MM in part 2
- extract the existing UK costs of MM, to compare these with our analysis in part 2

Search terms were agreed by the review team and project working group (attended by internal stakeholders and the project team) and then the following 5 databases were searched for articles: PubMed, Cochrane Library, HTA database, NICE Evidence search and NHS Economic Evaluation database. Searches were performed between 2 March 2018 and 8 April 2018. See Table 1 for a summary of the search strategies for each database.

Table 1: Search strategies by database

Database	Search Strategy
Pubmed (without MeSH terms)	((((("cost analysis"[Title/Abstract]) OR spend[Title/Abstract]) OR "cost of illness"[Title/Abstract]) OR "health expenditure"[Title/Abstract]) OR cost[Title/Abstract])) AND (((("multiple chronic"[Title/Abstract]) OR comorbidity[Title/Abstract]) OR co-morbidity[Title/Abstract]) OR multi-morbidity[Title/Abstract]) OR multimorbidity[Title/Abstract])
Cochrane Library, HTA database, NHS Economic Evaluations database	1 (multi-morbidity) OR (multimorbidity) FROM 1998 TO 2018 2 (comorbidity) OR (co-morbidity) FROM 1998 TO 2018 3 (multiple chronic health conditions) OR (long term disease) FROM 1998 TO 2018 4 (health expenditure) OR (cost) OR (cost analysis) FROM 1998 TO 2018 5 #1 OR #2 OR #3 6 #4 AND #5
NICE Evidence Search	(multi-morbidity OR comorbidity) AND "health costs" (multi-morbidity OR comorbidity) AND "cost analysis" Total brought to Endnote Removed duplicates

In addition, a strategy referred to as “reference hopping” was undertaken. This process identified other key documents relevant to the search that are part of the grey literature (such as the Somerset Symphony Project data which integrates social care data with

other health data (16)). Each relevant article identified in the review was searched using PubMed and the ‘similar articles’ section was reviewed to discover further relevant papers.

Limits applied to the search strategies were: English language, Human and 1998 to present day (8 April 2018).

Rationale for the exclusion/inclusion of articles

In the literature, MM has been referred to in multiple ways, such as “multi-morbidity”, “comorbidity” and “multiple diseases”.

For this project, we define MM as the presence of 2 or more long-term (chronic) health conditions in one individual.

Note that our definition of MM does not include disease weights/indexes (a numerical relative importance given to a condition or group of conditions) which is important because weights may reflect complexity of care or mortality rather than the presence of several conditions.

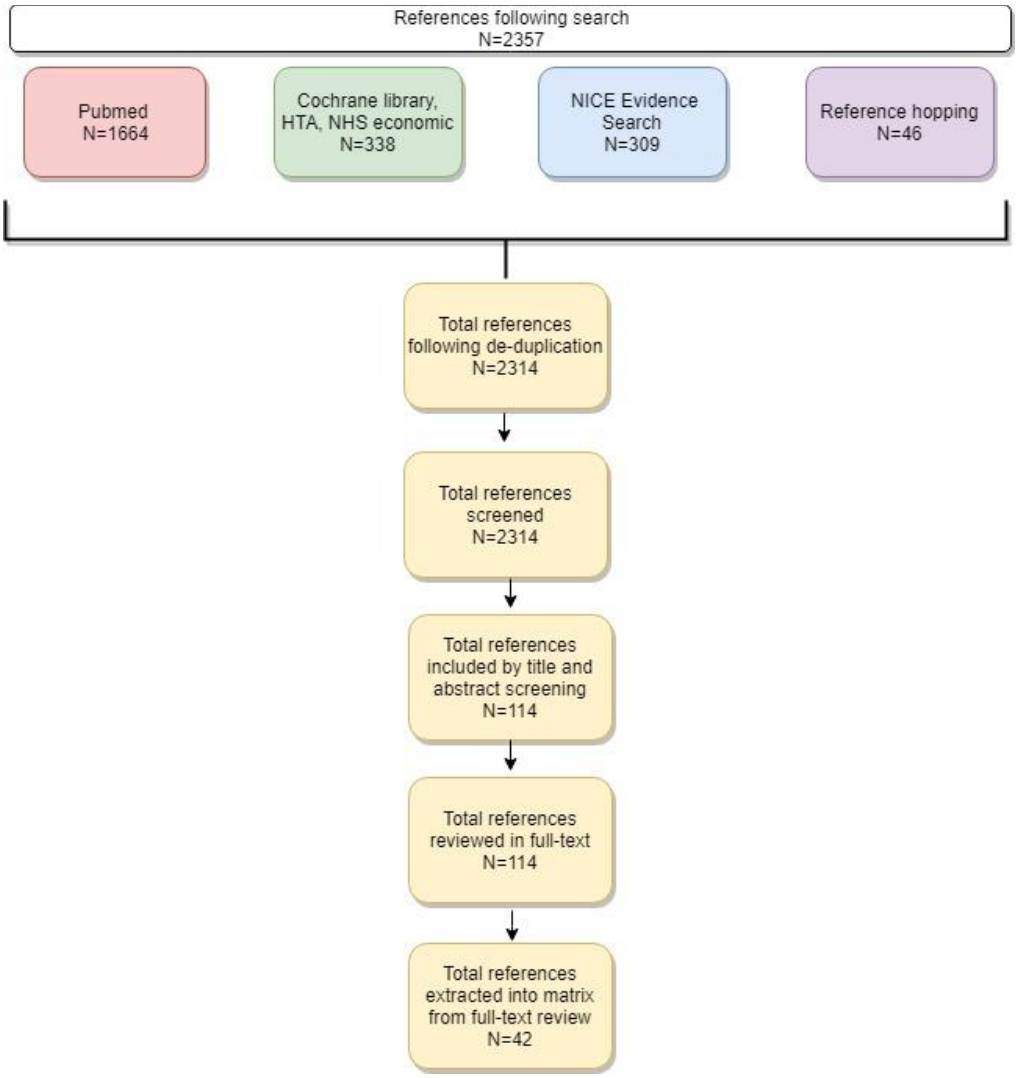
Additionally, comorbidity studies have been excluded from this review as they would require us to focus on a single health condition and its associated conditions. This may result in combined treatment that would underestimate costs. Nevertheless, this term was still included in the literature search as there is often a misunderstanding between the 2 terms. Therefore, our focus defines MM as the presence of 2 or more long-term health conditions, without making any judgment about the most predominant or primary condition (8). Further justification for the exclusion of co-morbidity is detailed in the Technical Appendix.

Finally, papers were excluded if the statistical analysis was not described in sufficient detail to be reproduced (17,18).

Results

A total of 2,357 results were retrieved from the combined searches and exported into Endnote. Removal of duplicates left 2,314 articles. The remaining articles were reviewed by title and abstract, identifying 114 eligible papers from which 42 were included in the final analysis following full text review. Most of the references were excluded because they focused on only one condition or calculated the prevalence of MM rather than cost. Results were summarised by the common methodologies used to calculate costs of MM. Figure 1 shows the procedure for arriving at the 42 papers which were included in the final analysis.

Figure 1: Extraction matrix



Note that our review did not search Econlit, which may have included additional relevant studies.

The literature review was analysed across 6 themes which were: the perspective of the study, the economic resource approach, the epidemiological approach, the type of population, the presence of sensitivity analysis, and the type of statistical analyses undertaken.

One systematic review of 35 studies explored the relationship between MM, healthcare utilisation, and costs of treating older people. It found a positive correlation between the presence of multiple conditions and healthcare costs (19). Many studies reported substantially higher costs with each additional condition, and a nonlinear (curvilinear), near exponential, relationship between the number of conditions and costs (20). Similar results were reported in a more recent systematic review of 26 articles and both studies highlighted the variety of methods used to estimate the MM cost burden (21).

Perspective of the study and types of costs

The 2 main perspectives reported in the literature are[1]:

- healthcare perspective
- societal perspective

The healthcare perspective includes the costs directly incurred by the healthcare provider. The costs typically considered within the healthcare perspective comprise of primary care, secondary care and prescription costs. The societal perspective includes all the costs incurred by society such as productivity losses (for example absenteeism costs).

In England, NICE recommends using a healthcare (NHS and personal social services) perspective only, excluding the wider social costs (22). This perspective was chosen to estimate cost per case figures of health and social care for future modelling work.

There were 33 of the 42 articles identified in the literature that included costs from the healthcare perspective but not in a complete manner since very few articles included all types of healthcare costs such as primary care, secondary care, and prescription costs. The remaining 9 articles adopted a societal perspective.

As per NICE recommendations, the analysis carried out in part 2 therefore provides a complete picture from a healthcare perspective for the following costs: primary care, secondary care, prescription, and social care.

[1] A third perspective exists: the payer perspective but is not reported here because of the rarity of such perspective in the 42 articles included.

Economic resource approach

Two approaches are possible for costing MM: bottom-up and top-down approaches (10, 21).

A bottom-up approach is the estimation of the total cost distribution from the analysis of data on individual healthcare utilisation. For example, the total annual GP cost per patient can be estimated by summing the cost of each GP event by individuals. This analysis is often challenging due to issues around availability of individual level data, which often exists in unstructured datasets. Depending on the level of information available in the data, cost can be aggregated at different organisational levels, or by severity amongst others, and econometric methods can be used to adjust for possible biases in the data collection.

A top-down approach breaks down total healthcare costs into elements (for example the proportion of costs attributable to a specific health condition). This often requires assumptions on the size or share of different conditions' impact on spending. Prevalence or healthcare utilisation is often used to make these assumptions. The main limitation of this approach is the assumption that patients with the same condition have similar healthcare costs (known as 'homogeneity').

Both the top-down and bottom-up methods rely on broad costs being attributed to a single condition. For example, in a bottom-up analysis, a single GP visit may involve a consultation about diabetes as well as related conditions such as cardiovascular disease or dementia. Consequently, difficulties occur when attributing this consultation to a particular condition or group of conditions. Costs need to be applied to both/all conditions, yet the relative shares of the consultation for each condition are very difficult to ascertain.

Thirty-eight out of the 42 studies identified in the literature search used a bottom-up approach. Due to data availability, we took a bottom-up perspective in our analysis in part 2.

Epidemiologic data

Two approaches for identifying patients with conditions/MM are presented in the literature: incidence- and prevalence-based approaches.

An incidence approach considers the occurrence of new cases in a given period, whereas the prevalence approach considers disease burden at one point in time.

The advantage of using a prevalence approach is that it gives an overall picture of disease stages and severity. Conversely an incidence approach implies that the

estimation of costs captures patients newly diagnosed with a condition over a specific period.

Both approaches can be seen in the present literature review with 26 papers taking a prevalence approach and the remaining 16 adopting an incidence approach.

Some estimation approaches included MM indices (see Background section) such as the count method (23), cumulative illness rating scale (24), clinical risk groups model (25), Rx-defined morbidity groups (Rx-MG) (12), or super-additivity (26) as a way to control for severity. More detail about these methods are provided in the Technical Appendix.

Population

Few articles estimated healthcare utilisation and costs for an entire population, or a representative sample of the population. We define a population as a group of individuals with whose demographic characteristics and prevalence of diseases is representative of the population of a country. However, most articles (36 out of 42) analyse cohort studies, with the analysis done on a subset of the population with distinct characteristics, such as having a specific condition of interest or within a specific age range (27, 28). These articles do not allow us to calculate total population costs or average cost-per-case figures of a specific disease or disease combination, as the estimates only apply to the cohort under study. Therefore, part 2 of this project has considered a representative sample of the population for primary and secondary care. For social care, part 2 of this project considered the costs in South Somerset, as nationally representative data was not available.

Sensitivity analysis

Only 3 papers included a sensitivity analysis (29-31). For part 2 it was deemed necessary to carry out a sensitivity analysis on the number of years of history observed when quantifying disease prevalence.

Statistical analyses

Several statistical analyses were used in the literature to derive costs of MM, but there was no consistency between these studies.

A few studies used an ordinary least squares (OLS) regression approach, which requires an unbounded continuous outcome variable. However, patient costs are always positive and therefore the cost distribution is skewed to the right and bounded by 0. In the context of health-care costs, cost distributions are often further skewed

because there tend to be a lot of patients with zero costs and some patients with or without MM might require a lot of healthcare, creating a long right tail.

Skewness, not an issue in itself, can often lead to the violation of regression assumptions which include linearity between the outcome variable and independent variables, normality of the error distribution, independence of errors and homoscedasticity of errors.

Consequently, several more robust approaches to standard OLS regression are available for modelling cost as an outcome:

The typical parametric transformation is the Box-Cox transformation (32), where the logarithmic transformation is chosen. The log of cost data is used to transform the raw cost outcome so that the outcome is unbounded and subsequently the assumptions of linear regression are satisfied (although this may not always solve the latter). The main limitation is the risk that non-adequate transformation or re-transformation will lead to errors, making the method unreliable. For example, if there are lots of zero-cost patients, which is a common phenomenon in cost data, log-transformation does not address the zero-cost issue. Furthermore, back transformation of the estimated coefficients for their interpretation, and their standard errors is a challenge and can lead to errors if done incorrectly. In addition, the model is not suitable for prediction due to the risk of predicting negative estimates.

Another approach to deal with violated regression assumptions which is widely used in the papers reviewed is the Generalised Linear Model (GLM) method. GLMs are a generalisation of linear regressions that solve the problem of transformation and allow non-normal error distributions. The dependent variable is connected to its variables via a link function which can be log, gamma, inverse normal or Poisson, depending on the outcome of interest. In this review of the literature, the Gamma distribution was often used, since it is suitable for strictly positive, continuous outcomes, so GLM datasets are more adapted to cost data modelling (33-35)

Quantile regressions are also used in the literature, to model the median cost patients, as well as the 25% and 75% quantiles of cost. The main benefit of this method is that it provides a range of costs in addition to a central estimate.

Three papers in our review addressed the issue of zero-cost patients by using a two-stage method (16,36,37). The ‘Symphony’ project collected individual-level data on the entire population of the South Somerset population in 2012 (16). The authors estimated the probability of an individual using healthcare in a first step. Then in a second step, the authors estimated average healthcare costs, for those patients who used healthcare. Similarly, Thiébaut et al. (2013) (36) used a logit model to first calculate the probability of consuming healthcare, followed by a GLM model to calculate

expenditures of co-variates, showing which patient characteristics and health conditions were having the biggest impacts on cost.

Note that other cost analysis methods such as non-parametric methods using for example, boot strapping, were not used for costing MM in any of the literature identified in this review.

Informed by this literature, a two-part method was used for the part 2 analysis.

Main results from studies set in the UK

In order to directly compare our results with the literature, we explored studies set in England and the UK. Only costs for the MMs of interest were extracted when available. The MMs of interest are:

- diabetes and coronary heart disease (CHD)/hypertension
- diabetes and depression
- CHD and depression
- diabetes and hypertension
- depression and hypertension
- chronic obstructive pulmonary disease (COPD) and hypertension
- CHD and depression and diabetes

It is difficult to compare the results we obtained in part 2 of this project with those found in the literature due to differences in the methodologies used.

The most similar relevant study identified in the literature review was one by Brilleman et al, (2013) (38). Considering only the conditions that are similar to the MMs of interest selected in this study, the authors found that:

- depression combined with another condition is generally cost-increasing (that is the costs of having a combination of 2 conditions is greater than the sum of having 2 conditions independently)
- chronic kidney disease, COPD and hypertension are cost-limiting conditions (that is the costs of having a combination of 2 diseases are below the sum of having 2 conditions independently)
- the proportion of cost-limiting conditions is greater in older age categories;
- dementia is not cost-increasing when co-occurring with any chronic conditions, but it is cost decreasing when combined with stroke

However, the findings above cannot be directly compared to the findings in part 2 of this project because Brilleman et al. (2013) (38) only considered primary care costs, whereas our analysis considered the overall healthcare costs including secondary care

costs. There were also other, less substantial, methodological differences between the analyses which are that:

- the authors used the Quality and Outcomes Framework (QOF) for the definition of their conditions of interest, while we used MedCODES and READ codes in this study (see the Technical Appendix)
- the authors controlled for a deprivation index
- the authors used an ordinary least squares (OLS) regression which does not take into account the skewed and bounded nature of cost data

The results of other studies identified in the literature review are presented in the Technical Appendix.

Part 2: Calculating the cost of MM in primary and secondary care

This section describes the methodologies used for calculating the costs of MMs in primary, secondary, and social care. The costs in primary and secondary health care (including prescriptions) were calculated separately from the costs in social care due to the availability of data.

Health conditions included

We conditions we studied were:

- COPD
- diabetes (Type 1 and 2)[2]
- lung cancer
- breast cancer
- coronary heart disease (CHD)
- stroke
- hypertension
- dementia
- liver disease
- depression
- colorectal cancer

The associated MedCODES, READ and ICD codes can be found in the Technical Appendix.

The multimorbidities, which were of interest as they are the more common combinations of the included conditions, were:[3]

- diabetes and CHD/hypertension
- diabetes and depression
- CHD and depression
- diabetes and hypertension
- depression and hypertension
- COPD and hypertension

[2] Note we could not differentiate between the types of diabetes in the datasets used such as HES.

[3] They were found common in the CHE report: https://www.york.ac.uk/media/che/documents/papers/researchpapers/CHERP96_multimorbidity_utilisation_costs_health_social%20care.pdf

These 11 health conditions were chosen because they have high prevalence rates and are linked to behavioural risk factors.

Datasets used

The primary care analysis was based on the richest available source of observational primary care data, the Clinical Practice Research Datalink (CPRD). The CPRD includes data on over 20 million patients including over 5 million currently registered and active patients and is representative of the UK population with respect to age, deprivation, gender and ethnicity. It provides information on diagnosis, tests, therapies, referrals and prescriptions at the patient level. For this analysis, the linked CPRD/HES dataset was analysed. The Hospital Episode Statistics (HES) dataset records all secondary care in England provided by an NHS hospital, but in this linked dataset, it is restricted to the patients appearing in CPRD. This analysis jointly studied the primary and secondary care utilisation of patients who appeared in the 2015 CRPD.

HES contains detailed information on the individuals' diagnoses and sufficient information to match the tariffs (the price awarded to commissioners for providing specific treatments) to the different hospital spells, to allow us to cost the healthcare individuals receive. This final dataset of primary and secondary care enabled us to have greater statistical power than previous studies and address a larger range of MM and their different combinations as it contains many individuals.

For social care costs, 2 methods were employed. For Method 1, the South Somerset Symphony Dataset was used. This is an integrated individual-level dataset which includes primary, community, acute, mental health and social care data for the population of South Somerset in 2012 (16).

The data was run through the RISC tool, which groups together several datasets using the pseudonymised NHS number of each individual. The social care cost was provided by the council, and this was also run through the pseudonymisation tool to ensure that the patients were matched with their data, but no patient identifiable information was shared. The final dataset comprises of 469,894 adults over 18, with 229,127 males and 240,767 females. Looking at one year of data (2015) there were 375,381 individuals who had no contact with the health or social care system for any of the conditions of interest and 297,302 when looking over 4 years (2012-15).

Method 2 did not require an individual level dataset but calculated social care costs using a top-down approach based on quality of life scores (QOL). The methodology is based on the same modelling assumptions as in the PRIMEtime model (39). This method was conducted as validation for the first method, to compare different methods for costing social care. Four steps were required (40):

Step 1: The School of Health and Related Research (ScHARR) at the University of Sheffield estimated the probability of a patient receiving residential social care, given their age and QoL using the Adult Social Care Survey (ASCS).

Step 2: The ASCS dataset was used to calculate the average cost to local authorities (LAs) for providing a week's residential care.

Step 3: Condition specific multipliers were applied, which have been estimated by ScHARR using the Personal Social Services Research Unit PSSRU report. These multipliers were used for stroke and dementia.

Step 4: Utility weights for each of the conditions of interest were sourced from the literature. For MMs, the 2 utility weights were multiplied to give a utility weight for the MM combination.

Method used to estimate the costs of multi-morbidity in primary and secondary healthcare

The CPRD extract used comprised of individuals aged 18 years or older in 2015, and this project examined healthcare utilisation in that calendar year. Individuals alive at the beginning of 2015 are included in the analysis, regardless of whether they lived for the entire year. The primary care data was cleaned and grouped into different categories of healthcare providers (for example GP, nurse) and services (for example clinical, home visit). It was then matched to the cost of each service, either based on the duration of the consultation using a pro rata wage (for example GP visit duration), or per unit cost of the service (for example home visit). Prices were sourced from the Unit Costs of Health and Social Care provided by the Personal Social Services Research Unit (PPSRU) which are published annually. Healthcare prices vary every fiscal year. Prices in the fiscal year 2015-16 have been used here, so that variation in cost reflects variation in healthcare utilisation and not a change in prices, similar to what has been done previously in the literature (41). The Technical Appendix provides more detail of the different categories of primary care, and their costing methods.

Prescription data were extracted from CPRD and matched to the cost of that medication. Two price sources were used because each of the single sources included different costs for a subset of the drugs in CPRD. The sources used were the NHS Electronic Drug Tariff and NHS Prescription Cost Analysis (PCA) for 2015 (see Technical Appendix), where costs for drugs were not identified in the former. For the NHS Drug Tariff, we calculated cost per quantity. For the PCA, we used cost per quantity for all liquid formulations, tablets and capsules, and cost per item for other products. We then multiplied the cost per quantity of the products by the quantity ('qty') variable in CPRD. A number of errors in recordings of the 'qty' variable in CPRD were found and corrected before drugs were costed.

Finally, HES episodes were grouped into spells using the “Grouper” (see Technical Appendix for further details) which also provided a main spell Healthcare Resource Groups (HRGs). This is needed when a patient receives multiple treatments during a single hospital stay, as this is the method in which providers are reimbursed. The spell HRGs were matched to the National Tariff in 2015-16. The tariffs are the basis of a spell reimbursement. These costs were aggregated for the year 2015 at the patient level.

It is not possible to associate each contact with health services to specific health conditions, not only because the healthcare provider does not systematically report a diagnosis, but also because doctors and medical professionals may not be able to establish causality between a health event and a specific condition. Furthermore, in the case of MM, a health event requiring contact with health services could be due to one or more health conditions. For example, a retinopathy is a usual complication of diabetes, but patients may suffer from a retinopathy without having diabetes, and patients with diabetes may have had a retinopathy even if they had not suffered from diabetes. Additionally, when diagnosing a MM patient suffering from diabetes and liver disease, the retinopathy could be due to diabetes or a liver malfunction. Therefore, to avoid inaccurately attributing a healthcare event to a specific condition, we compared all annual healthcare utilisation of patients with the condition(s) of interest to patients without these conditions, adjusting for age group and gender. The excess annual healthcare cost (the net cost) is then attributed to the condition(s) the patients are suffering from.

To identify patients with one or more of the conditions of interest, we considered their current and historical diagnosis as evidence of the health condition. Although the patients’ diagnoses may not report systematically to the patients’ long-term condition(s), we assumed that, by using a sufficiently long period of time, we would identify the chronic conditions they are suffering from. Although this may be a reasonable assumption in the case of chronic conditions, it is not obvious what a reasonable period is for the different conditions of interest. A short period, such as a year, would pick up acute conditions, whereas a longer period would allow us to capture individuals suffering from the disease(s) of interest for longer, but for whom the disease was managed well enough so that the patients would require less frequent medical attention. Therefore, we used 2 different periods of time to classify patients into the different disease categories. The 2 different definitions are:

Definition A considered an individual to have a disease if they had the condition reported in their general practice records within the year of study (2015). Definition B considered an individual to have a disease if they had a record of it within the last 4 years (2012-2015). It should be noted that only healthcare utilisation in 2015 was costed since our aim was to calculate an annual cost per case.

It was agreed with the project Steering Group that multiple years of additional diagnosis under Definition B, compared to Definition A, would allow us to capture current conditions that are well managed and do not require regular healthcare visits, without classifying individuals into disease groups that they may have recovered from (for example if depression has not been reported over the last decade, it may be fair to assume that the condition has been resolved). Furthermore, very few GP records reported that a condition had been resolved and we judged this information as unreliable and did not use it in our methodology. Ultimately 4 years was chosen for definition B as the Somerset Symphony dataset is only available for the years 2012-15, and we kept this duration for the analysis using the CPRD data to have comparable estimates for health and social care costs. The implicit assumption made is that any diagnosis made before this four-year period is no longer relevant if not reported again, and the patient had fully recovered from the condition by 2015. While it was decided that definition B was preferable for estimating the cost per case of all individuals with a single health condition or MM, the results of both definitions are presented in this report.

The net healthcare costs were estimated using a two-part cost modelling approach that allowed us to control for disease status (presence/absence of disease), interactions between diseases (MMs), age, and sex covariates. The two-part statistical regression model was specified for total individual expenses and takes account of the probability that an individual used the healthcare system over the year of observation, as well as the expected expenditures related to the individual's overall use, conditional on receiving healthcare. The analysis was repeated twice, once for each definition of disease identification (A and B).

The methodology relied on regression analysis. The dependent variable, that is the annual individual cost, was composed of the combination of primary and secondary healthcare, and prescription costs. For each disease, individuals were dichotomised as either '0' if there was no record of them having each of the diseases of interest and '1' if the disease was reported. Single disease dummy variables were included to control for each morbidity effect. The constant term of the regression model captured the average cost of individuals suffering from none of the conditions considered ('baseline cost'). The additional cost of each disease was captured by the relevant disease dummy coefficient. Therefore, the total 'cost per case' of a patient with each morbidity was computed by combining the constant and the single disease coefficients. The cost of having a MM was calculated by combining the constant, single disease, and the MM (interaction) coefficients, and adjusting for the age and gender of patients with the condition or MM. A positive MM coefficient indicated that the combination of the diseases for one individual was larger than the sum of costs for individuals who each had only one of the diseases (ignoring the 'baseline cost' of these individuals, which represents spending not related to the health conditions of interest). Conversely, a negative MM coefficient indicated that the combination of the diseases for one individual

was less than the sum of costs for individuals who each only had one of the diseases (ignoring the ‘baseline cost’ of these individuals).

For each MM, we reported the average overall ‘cost per case’ which combines the baseline cost, the cost of each morbidity, and the (interaction) MM coefficient. The average cost was computed for each age group and gender and aggregated for each gender weighted by the number of individuals in each age/sex category. Therefore, the reported average cost for, for example, diabetes and CHD is not directly comparable to the average cost for, for example, diabetes and depression as the ages of patients with the different MMs will vary, accounting for some of the difference in average costs between the MMs.

In summary, this approach estimated a ‘baseline cost’, the specific disease costs, and the MM cost, for each gender. However, the final average overall cost of a patient with a MM of interest cannot be directly compared to the combination of the single morbidity costs as it is weighted by the individual prevalence in the data in terms of age.

Two-part model

The econometric treatment of healthcare expenditure requires its distribution to be taken into account, notably the presence of individuals who did not have any consumption of healthcare during the year of study (2015). Although several estimation techniques are available, the analysis of CPRD-HES data characteristics determined the most suitable econometric model. Based on Jones (2000) (42) and Deb and Trivedi (2006) (43), a ‘two-part’ model appeared to be the most appropriate to deal with ‘zero expense’.

The two-part model treats healthcare consumption as 2 statistically independent parts. The first part aims to compute the probability an individual has zero versus positive spending, while the second part computes the level of positive spending. The whole model (that is the conditional mathematical expectation of expenditure) is obtained by multiplying the probability of the individual consuming healthcare in the year by the estimated conditional expenditure. The choice of such a model is also justified by the flexibility it brings to the model, by dividing the mechanism of consumption into nearly independent parts. See Duan et al., (1983) (44), Jones (2000) (42), Manning and Mullahly (2001) (45), Buntin and Zaslavsky (2004) (46) and Deb and Trivedi (2006) (43) for more detail about this methodology.

In general, the average medical cost for any age and gender can be predicted as follows:

$$E[c_i|x_i] = Pr[c_i > 0|x_i] \times E[c_i|c_i > 0, x_i] + Pr[c_i = 0|x_i] \times E[c_i|c_i = 0, x_i] \quad (1)$$

where c_i is the dependent variable of total cost and x_i the set of explanatory variables for individual *i*.

And since expectancy of consumption zero equals zero the model can be written as:

$$E[c_i|x_i] = Pr[c_i > 0|x_i] \times E[c_i|c_i > 0, x_i] \qquad (2)$$

This ensures that predicted costs are representative not only of the people with positive medical costs, but also of all the people with the diagnosed conditions that don't consume healthcare in the year of study. For example, the total predicted cost for a person *i* with diabetes ($NCD_{ij} = 1$) is*: $E(c_i|diabetes=1) = P(c_i>0)*E(c_i|c_i>0, diabetes=1)$*.

The extra cost of a disease can therefore be estimated, for a given gender and age group, as the difference in the predicted costs, conditional on the disease status.

We averaged out the individual costs by age category, gender, morbidity and multimorbidity groups (for example $E\left[c_a|NCD_{ij} = 1, NCD_{ik} = 1, A = a, Gender = g_i\right]$). Finally, we estimated the overall average individual cost by morbidity and MM category, and gender by calculating the weighted average disease cost weighted by the number of individuals in each disease category ($\omega_a|NCD_{ij} = 1, NCD_{ik} = 1,\ A = a, Gender = g_i$), as for example, in the case of the presence of MM such as the non-communicable diseases $NCD_{ij}\ and\ NCD_{ik}$.

$$\bar{c}(M_1 = 1, M_2 = 1, Gender = g_i) = \Sigma^A{}_{a=1} E\left[c_a|NCD_{ij} = 1, NCD_{ik} = 1, A = a, Gender = g_i\right] * \omega_a \quad (3)$$

The Technical Appendix provides more detail about this methodology.

Econometric specification

All models were stratified by gender.

First part: the probability of requiring healthcare

Estimates from the first part give the probability of an individual requiring healthcare. The first part of this two-part model estimator is estimated using a logit regression.

$$Pr[c_i > 0|x_i] = \frac{1}{1+e^{-\alpha x_i}} = \frac{e^{\alpha x_i}}{1+e^{\alpha x_i}} \qquad (3)$$

Second part: the average healthcare consumption of the individuals who required healthcare

Estimates from the second part give expected consumption costs of people who used the healthcare system during the year. The second part is estimated with a Generalized Linear Model (GLM) multivariate Gamma regression with log as the link function.

$$\ln c_i = \beta_0 + \beta_1 \cdot AGEcat_i + \sum_j \beta_j \cdot NCD_j + \sum_{j \neq k} \beta_{jk} \cdot NCD_{ij} \cdot NCD_{ik} + \varepsilon_i \quad (4)$$

Where $AGEcat_i$ is a patient's age categorical variable; $NCD_{i,k} = 1$ iff individual i suffers from illness k, and 0 otherwise. It follows that $NCD_{i,k} \cdot NCD_{i,j} = 1$ iff $NCD_{i,k} = NCD_{i,j} = 1$, that is iff individual i suffers simultaneously from illness k and illness j, and ε_i is an error term. The intercept β_0 represents the predicted medical cost for a person aged 18-39, the reference category.

As an example, the total predicted medical cost for a person aged 55 with none of modelled diseases and with positive costs would be equal to:

$$E(c_i | c_i > 0) = \exp(\hat{\beta}_0 + \hat{\beta}_{50-59}) \quad (5)$$

For a person with diabetes of the same age, the total predicted cost in this sample would be equal to:

$$E(c_i | c_i > 0) = \exp(\hat{\beta}_0 + \hat{\beta}_{50-59} + \hat{\beta}_{diabetes}) \quad (6)$$

For a person with both diabetes and cancer, the total cost in the sample of people with positive costs can be predicted as:

$$E(c_i | c_i > 0; diabetes = 1; cancer = 1) = \exp(\hat{\beta}_0 + \hat{\beta}_{50-59} + \hat{\beta}_{diabetes} + \hat{\beta}_{cancer} + \hat{\beta}_{diabetes*cancer} \quad (7)$$

Assumptions and particular features of the econometric modelling

1. The 2 disease sample prevalence definitions led to variation in the econometric model. For the first part in Definition A, the age categorical dummy was the only covariate, as the disease identification is inherent to positive consumption. A NCD is recorded only if the patient had contact with the healthcare system during the year, therefore, all individuals identified as having a health condition using Definition A will have positive healthcare costs. In the first-part of the model of Definition B, the disease indicators are included as additional regressors.

2. Breast cancer was excluded from the covariates of the male regression, due to low prevalence in the sample.

3. Lung cancer and colorectal cancer were not crossed with other NCDs because of scarcity of cases. All other NCDs interactions were included.

4. A fixed-effect specification with GP practice as the fixed-effect control variable was considered. However, this specification was rejected because of a convergence issue due to over specification of the model.

5. In the first part of the two-part model for both definitions of disease identification, and the second part with sample prevalence as defined by Definition B, age is included as a categorical variable (9 age groups), whereas it is continuous and centred around 0 in the second part for both samples - male and female - for Definition A with NCDs specification. This responds to issues of data dispersion leading to a lack of model convergence (an over-specified regression model considering all included interactions).

Method for calculating social care costs

Two different methods were chosen to derive social care costs.

1. Calculating social care costs of MM using the Somerset Symphony Dataset

The method using the Somerset Symphony data follows the bottom-up costing approach and two-part model adopted for CPRD and HES.

Most of the population in the dataset (457,138) had no recorded social care cost. The model was run on the individual financial year cost from 2014/15, and the years 2011/12, 2012/13, 2013/14 and 2014/15 were used to identify the individuals being affected by the different diseases of interest for Definition B.

2. Calculating social care costs of MM using a top-down approach via quality of life (QOL) scores

This method was added as a validation to compare different methods for costing social care. Four steps are required to derive social care costs of the diseases of interest using the QOL Approach.

First step: Estimating the probability of using social care
The School of Health and Related Research (ScHARR) at the University of Sheffield estimated the probability of a patient receiving residential social care, given their age and QoL using the Adult Social Care Survey (40). They showed that:

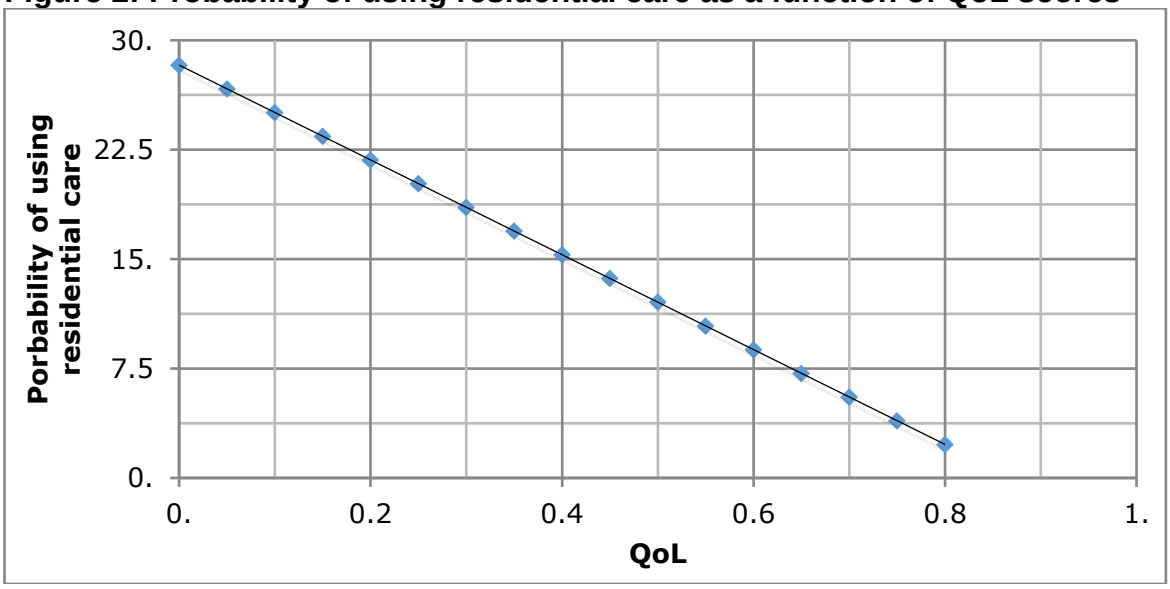
- the probability of residential care is close to zero for patients below 75 years

- the probability of residential care as a function of QoL can be summarised by the linear regression equation (Figure 2)

Probability of using residential care =0.283 - 0.325*QoL

A regression equation is a 'line of best fit' that estimates the probability of using social care given a patient's QoL. It is illustrated on the chart below. For example, you can see that for someone with a QoL of 0.4 (40% of perfect health), the estimated probability of using social care is 15%.

Figure 2: Probability of using residential care as a function of QoL scores



Second step: Estimating the costs of care

From the data from the Adult Social Care Survey 2017, the average cost to local authorities for a residential care user was estimated at a weekly cost of £345.51. We estimated this parameter using STATA ® to analyse the data available from 7 councils out of the 151 in England for which the relevant parameters were available (Halton, Solihull, Derbyshire, Bedford, Bromley, Poole and Brighton).

Third step: Condition-specific multipliers

Condition-specific multipliers estimated by ScHARR using the Personal Social Services Research Unit PSSRU report (47) were that:

- patients with dementia were estimated to have 8.41 times greater care costs than average patients of the same age and QoL
- patients with stroke were estimated to have 5.88 times greater care costs than average patients of the same age and QoL
- a multiplier of 1 was assumed for all the other conditions, following the literature (40)

Fourth step: Multiplication of the results of the first 3 steps for the condition of interest
We sourced utility weights (QoL scores) for each condition to calculate the probability of using social care for each condition. For MMs, we assumed that utilities weights were multiplicative. Table 2 provides the utility weights by condition that were multiplied by steps 1-3.

After calculating the probability of using social care for each condition, we applied the average cost of social care estimated in the second step. We then applied the condition specific multipliers to the estimated costs of stroke and dementia.

Table 2: QoL (EQ-5D) score[4] by disease

Condition	QoL (EQ-5D)	Source
COPD	0.470	Sullivan et al. 2011 (48)
Diabetes	0.661	Sullivan et al. 2011 (48)
Coronary heart disease (CHD)	0.760	Laires et al. 2015 (49)
Stroke	0.713	Rivero-Arias (50)
Hypertension	0.721	Sullivan et al. 2011 (48)
Colorectal cancer	0.676	Sullivan et al. 2011 (48)
Depression	0.600	Turner et al. 2013 (51)
Lung Cancer	0.560	Sullivan et al. 2011 (48)
Breast Cancer	0.750	Sullivan et al. 2011 (48)
Dementia	0.442	Sullivan et al. 2011 (48)
Liver disease	0.620	Sullivan et al. 2011((48)
Diabetes and CHD	0.50236	Multiplied based on the above
Diabetes and Depression	0.3966	Multiplied based on the above
CHD and Depression	0.456	Multiplied based on the above
Diabetes and Hypertension	0.476581	Multiplied based on the above
Depression and Hypertension	0.4326	Multiplied based on the above
COPD and Hypertension	0.33887	Multiplied based on the above

[4] The Qol score varies between a little bit below 0 and 1: 0 representing death and 1 being a state of ideal full health. EQ-5D was developed by the EuroQol Group and is a multi-attribute instrument based on a set of 5 questions related to mobility, self-care, usual activities, pain and anxiety.

Results

Primary care, secondary care, and prescription costs - sample prevalence and costs by condition and MM combination

and Table 4 present the results of healthcare cost estimates for males and females respectively. Each table displays the sample prevalence, mean cost per case in 2014/15, the significance level of the coefficients, as well as the lower and upper limits of 95% confidence intervals. Results using definitions A and B of sample prevalence are presented in the left and right panels respectively.

Definition A always leads to higher average cost estimates than Definition B as it includes only individuals requiring healthcare within the year 2015, thus capturing patients with the acute form of the disease, or those who are more recently diagnosed. Definition B additionally includes individuals who did not necessarily see a doctor within the last 3 years for the condition of interest, and therefore reduces the average cost calculated for 2015.

The mean costs represent the individual annual healthcare expenditures (primary and secondary care, as well as prescription costs) of patients with conditions/MM. The baseline cost and cost of individuals with only one morbidity of interest are always statistically significant. This implies that all the conditions identified impose higher costs, compared to the average cost of individuals without any of the conditions of interest.

Note that the average cost of a person with each condition will include the ‘baseline’ costs, unrelated to their condition. It was not possible in this analysis to control for all conditions, and therefore for example the average cost per case of a patient with diabetes will exclude the costs of other diseases modelled but will include the costs of unrelated or related conditions which have not been modelled.

The age of the average patient with each condition/MM will differ, and therefore no meaningful calculations comparing the figures in the table below can be performed. For example, it is not possible to calculate the additional cost imposed by diabetes by taking the difference in the cost per case for diabetes and the baseline cost.

Average cost per case figures for patients with MM are greyed out where the interaction term is not statistically significant, and therefore the additional costs imposed by having one of the diseases are not affected by whether the patient also has the other disease. This means that the costs of the diseases are ‘additive’. For example, the costs of a patient with depression and COPD is equal to the costs of a patient with COPD plus the additional costs imposed by depression (excluding baseline costs). In other words,

there are no cost savings or additional costs required to treat patients with one of the diseases, who acquire the other, compared to the treatment costs of one of the diseases, for patients who don't have the other.

However, it is not possible to make meaningful comparisons based on the figures in Tables 3 and 4 for the average costs per case of COPD and depression, cost per case of only depression, and cost per case of only COPD. This is because the cost per case for each condition/MM has been weighted by the age distribution of patients with the condition/MM. The age profile of patients with hypertension, patients with diabetes, and patients with both, will differ. This implies that part of the reason for the difference in costs per case is due to age. To control for age, we must look at the coefficients on MM. This analysis can be found in Table 5.

Table 3. Primary, secondary, and prescription cost results for males

MALE	Definition (Definition A)					Definition (Definition B)				
HEALTH CONDITION/MM	**Sample Size**	**Mean cost per case (£, 2015)**		**LL (95%)**	**UL (95%)**	**Sample Size**	**Mean cost per case (£, 2015)**		**LL (95%)**	**UL (95%)**
BASELINE		851	***	835	868		672	***	657	686.66
CHD	4,377	2514	***	2328	2714	4,655	1993	***	1860	2133.74
CHD AND COPD	476	6062.13	*	5171.32	7105.2	502	4460	***	3954	5024.29
CHD AND DIABETES	1,091	3376	***	2999	3800	1,091	2823	**	2586	3079.52
COLORECTAL CANCER	265	6425	***	5315	7765	547	3581.84		3132.29	4075.52
COPD	6,955	3488	***	3274	3715	4,687	3022	***	2816	3240.47
DEMENTIA	1,755	4345	***	3867	4880	1,118	3473	***	3041	3954.84
DEMENTIA AND CHD	110	6793.61		5155.38	8950.47	156	4436	***	3662	5357.31
DEMENTIA AND COPD	193	8296	**	6507	10573	136	6822	***	5495	8419.28
DEMENTIA AND DIABETES	424	5487	**	4549	6617	263	4770.52		3987.84	5679.11
DEMENTIA AND HYPERTENSION	639	8209.74		6964.51	9675.36	1,089	4795.51		4231.97	5422.01
DEMENTIA AND STROKE	38	12282	***	8251	18277	95	6495	***	5114	8185.69
DEPRESSION	9,392	1873	***	1766	1986	34,300	1392	***	1343	1442.83
DEPRESSION AND CHD	132	5471.93		4211.36	7108.99	578	2935.46		2609.72	3294.52
DEPRESSION AND COPD	257	6420.39		5162.22	7984.35	885	4125.94	*	3661.18	4638.63
DEPRESSION AND DEMENTIA	81	7255.1	*	5189.3	10141.21	186	4971.19		4066.56	6041.32
DEPRESSION AND DIABETES	635	3628.82		3088.96	4262.67	2,539	2655.72		2451.74	2873.76
DEPRESSION AND HYPERTENSION	994	5201.1		4527.44	5974.35	5,432	2555.24		2396.18	2723.27
DEPRESSION AND LIVERD	161	8407	***	6192	11413	540	6003.58		5059.98	7089.01
DEPRESSION AND STROKE	45	9777.87		6369.82	15007.89	239	4335.51		3571.91	5226.52

DIABETES	26,288	1870	***	1801	1942	18,119	1606	***	1538	1675.26
DIABETES AND COPD	1,395	4523	***	4040	5064	870	4167	**	3747	4627.35
HYPERTENSION	31,221	2532	***	2444	2622	55,031	1677	***	1631	1723.65
HYPERTENSION AND CHD	3,003	5385.2		4933.16	5877.69	8,449	3326.69		3155.35	3506.09
HYPERTENSION AND COPD	2,126	6587	***	5974	7260	3,587	4452.53		4136.15	4789.56
HYPERTENSION AND DIABETES	8,665	4585.64	*	4330.87	4854.74	18,877	2607.22		2508.57	2709.26
HYPERTENSION AND STROKE	491	9516	***	7808.	11596	1,820	3955	**	3564	4382.21
LIVER DISEASE	861	6795	***	5794	7968	1,660	4016	***	3538	4549.52
LIVER DISEASE AND CHD	23	16510.86		11051.24	24663.85	45	6366	**	5023	8017.95
LIVER DISEASE AND COPD	98	13631	***	9978	18617	129	8395	***	6645	10506.24
LIVERD AND DEMENTIA	20	27900.94		16785.64	46368.32	29	10458.85		7311.1	14709.91
LIVERD AND DIABETES	176	10569.46		8176.27	13661.61	305	6240.53		5205.69	7440.02
LIVERD AND HYPERTENSION	352	9112	***	7370	11265	793	6357.29	*	5514.3	7307.14
LIVERD AND STROKE	12	40782.74		20986.37	79244.15	24	9090.66		6372.88	12682.8
LUNG CANCER	171	4848	***	3966	5926.	171	3297	***	2690	3988.77
STROKE	404	6335	***	5105	7860	1,137	2918	***	2555	3323.8
STROKE AND CHD	30	10872.41		7311.65	16164.35	143	4368	**	3634	5234.93
STROKE AND COPD	50	13606.53	*	9379.77	19734.64	106	7176.49	*	5850.89	8781.57
STROKE AND DIABETES	38	9044.63		6636.65	12324.23	95	4613.63		3901.3	5440.89

Note: The mean cost reports the average cost per case of patients in a specific category in 2015. For patients with conditions/MMs, the average cost per case includes baseline costs. 95% confidence interval is reported for the average cost distribution. Definition A assumes a person had contact with the health system due to their condition in 2015, Definition B included any person who had contact with the healthy system due to their condition between 2012 and 2015. The significance levels are reported for the following thresholds: * significant at 10% (in grey), ** significant at 5%, *** significant at 1%. LL: lower limit; UL: upper limit. Results with less than 5% significance level are greyed out.

Table 4. Primary, secondary, and prescription cost results for females

	Definition A					Definition B				
HEALTH CONDITION/MM	**Sample Size**	**Mean cost (£, 2015)**		**LL (95%)**	**UL (95%)**	**Sample Size**	**Mean cost (£, 2015)**		**LL (95%)**	**UL (95%)**
BASELINE	N/A	1069	***	1055.51	1083.06	N/A	900	***	885.03	914.3
BREAST CANCER	1,187	9124	***	8065.48	10319.7	4,004	3136	***	2921.88	3362.47
BREAST CANCER AND COPD	42	11243	***	7318.95	17268.21	136	5322	***	4329.67	6490.74
BREAST CANCER AND DIABETES	89	11173		8265.89	15100.54	372	4020.9	***	3484.44	4617.83
CHD	2,038	2848	***	2608.15	3109.24	1,889	1961.43	***	1802.75	2130.37
CHD AND BREAST CANCER	11	27354.99		15667.53	47754.89	53	5844.13		4621.86	7296.9
CHD AND COPD	267	5800.88	***	4926.54	6829.55	260	4142.17	***	3649.36	4692.3
CHD AND DIABETES	436	3887.82		3401.66	4443.03	521	2839.55	**	2558.7	3144.3
COLORECTAL CANCER	252	6224.23	***	5223.98	7415.21	538	2955.46	***	2625.08	3310.51
COPD	6,722	3332.15	***	3157.53	3515.92	4,408	2735.35	***	2574.39	2904.39
DEMENTIA	3,416	3243.43	***	3018.75	3484.32	2,040	2518.75	***	2307.52	2744.79
DEMENTIA AND BREAST CANCER	25	9610.36	***	5885.32	15690.8	54	5582.47	**	4253.16	7229.17
DEMENTIA AND CHD	108	5559.05	**	4456.6	6933.17	160	3897.26	***	3333.75	4537.42
DEMENTIA AND COPD	194	6438.21	**	5316.74	7795.10	157	5258.65	**	4470.42	6161.27
DEMENTIA AND DIABETES	584	4163.78	***	3641.26	4760.56	319	3450.29	***	3032.35	3912.11
DEMENTIA AND HYPERTENSION	1,415	6179.27	*	5591.86	6827.36	2,500	3469.31	**	3217.55	3737.06
DEMENTIA AND STROKE	76	7611.6	***	5801.07	9986.79	135	4615.89	***	3847.14	5496.01
DEPRESSION	19,386	2455.72	***	2369.57	2544.86	67,183	1685.5	***	1647.25	1724.51

DEPRESSION AND BREAST CANCER	77	12426.89	*	8681.44	17786.5	839	4021.92	***	3555.83	4535.45
DEPRESSION AND CHD	51	5143.04		4127.15	6408.42	302	3048.17		2731.56	3394.83
DEPRESSION AND COPD	453	6715.05		5803.59	7768.55	1,519	4059.61	**	3738.29	4403.88
DEPRESSION AND DEMENTIA	198	5647.73		4653.39	6854.12	484	3483.18	**	3069.09	3936
DEPRESSION AND DIABETES	1,033	4402.73		3947.03	4910.62	5,634	2757.64	***	2618.95	2902.36
DEPRESSION AND HYPERTENSION	1,741	4799.02		4388.79	5247.36	9,908	2546.94	**	2445.04	2652.11
DEPRESSION AND LIVER DISEASE	136	9580.69	***	7441.45	12333.85	648	5249.74	**	4601.89	5970.7
DEPRESSION AND STROKE	51	9020.57	*	6575.05	12374.55	302	3923.32	*	3392.61	4523.41
DIABETES	21,778	2194.37	***	2122.36	2268.56	19,930	1692.11	***	1632.42	1753.6
DIABETES AND COPD	973	4282.07	***	3849.41	4762.83	671	3734.61	***	3394.09	4102.41
HYPERTENSION	34,268	2566.83	***	2498.33	2636.92	58,091	1640.84	***	1605.15	1677.11
HYPERTENSION AND BREAST CANCER	381	10296.26	***	8484.8	12492.76	1,942	3679.19	***	3369.61	4011.86
HYPERTENSION AND CHD	1,887	5407.29	**	4945.72	5911.17	4,642	3113.42		2940.31	3294.69
HYPERTENSION AND COPD	2,074	5652.53	***	5204.14	6138.84	3,437	3803.37	**	3570.84	4047.78
HYPERTENSION AND DIABETES	7,456	4378	***	4162.73	4602.73	15,955	2493.15		2407.97	2580.81
HYPERTENSION AND STROKE	495	7259	***	6161.26	8551.98	1,672	3819.46		3476.64	4186.46
LIVER DISEASE	804	6646.24	***	5808.11	7604.65	1,527	3722.43	***	3358.05	4118.55
LIVERD AND BREAST CANCER	31	19415.42	***	11688.09	32247.45	77	7835.8	**	5801.85	10382.5

LIVERD AND CHD	17	11703.53		7937.23	17255.35	29	5547.87	**	4388.87	6961.39
LIVERD AND COPD	98	11595.52	***	8856.66	15179.04	124	7244.69	***	5980.68	8738.21
LIVERD AND DEMENTIA	41	10512.26	**	7382.18	14968.25	39	6129.87	***	4722.47	7846.32
LIVERD AND DIABETES	137	10305.36		8222.69	12914.07	285	5583.25		4771.46	6498.35
LIVERD AND HYPERTENSION	379	8535	***	7178.38	10146.92	787	5201.79		4611.7	5849.65
LIVERD AND STROKE	17	24475.26		14623.44	40959.96	29	7780.89	*	5821.39	10280.98
LUNG CANCER	167	4322.57	***	3653.43	5113.81	194	2925.63	***	2471.23	3419.87
STROKE	376	6050.97	***	5044.04	7258.11	918	2695.75	***	2387.37	3033.32
STROKE AND BREAST CANCER	4	110287.1		40795.07	298128.5	28	7642.39		5489.14	10402.23
STROKE AND CHD	21	12446.69		8557.11	18102.08	67	4005.9	***	3312.52	4817.16
STROKE AND COPD	41	11730.93	*	8443.66	16295.96	72	5770.4	**	4773.84	6953.82
STROKE AND DIABETES	61	8630.46		6563	11347.75	157	3911.02	*	3331.23	4568.13

Note: The mean cost reports the average cost per case of patients in a specific category in 2015. For patients with conditions/MMs, the average cost per case includes baseline costs. 95% confidence interval is reported for the average cost distribution. Definition A assumes a person had contact with the health system due to their condition in 2015, Definition B included any person who had contact with the healthy system due to their condition between 2012 and 2015. The significance levels are reported for the following thresholds: * significant at 10% (in grey), ** significant at 5%, *** significant at 1%. LL: lower limit; UL: upper limit. Results with less than 5% significance level are greyed out.

Primary care, secondary care, and prescription costs – interpreting the costs of multimorbidity

As explained in the Methodology section, the predicted costs of a patient with diabetes, aged 50-59, are equal to:

$$E(c_i|c_i > 0) = \exp(\hat{\beta}_0 + \hat{\beta}_{50-59} + \hat{\beta}_{diabetes})$$

For a person with both diabetes and cancer, the total cost in the sample of people with positive costs can be predicted as:

$$E(c_i|c_i > 0; diabetes = 1; cancer = 1) = \exp(\hat{\beta}_0 + \hat{\beta}_{50-59} + \hat{\beta}_{diabetes} + \hat{\beta}_{cancer} + \hat{\beta}_{diabetes*cancer})$$

When examining the impact of multimorbidity, we looked at the 'interaction' term in the regression, which equalled 0 if the patient did not have both diseases, or 1 if they did.

For the majority of combinations, we found that the interaction term was negative, and this finding was statistically significant. This implies that, in the example above, the additional costs imposed by having cancer are smaller if the patient also has diabetes, compared to having cancer for a patient who does not have any condition modelled here. This also implies that for modelling purposes, it may not be robust to simply add the treatment costs of 2 diseases when estimating the costs of a patient with multi-morbidity.

A summary table showing the impact of multimorbidity on costs, for common MMs, is shown below.

Table 5. Summary table of each chronic condition combination and whether costs are less than additive or additive at the one percent significance level

	Definition A		**Definition B**	
	Average cost per case compared to sum of the costs of 2 patients, each with one of the conditions, controlling for age. Note that 'baseline' costs have only been counted once in the comparator group.			
Chronic condition	Men	Women	Men	Women
Diabetes and CHD	77% (n = 1,091)	67% (n = 436)	78% (n = 1,091)	78% (n = 521)
Diabetes and Hypertension	Not significant (n = 8,665)	92% (n = 7,456)	Not significant (n = 18,877)	75% (n = 15,955)
Diabetes and Depression	Not significant (n = 635)	Not significant (n = 1,033)	Not significant (n = 2,539)	Not significant (n = 5,634)
CHD and Depression	Not significant (n = 132)	Not significant (n = 51)	Not significant (n = 578)	Not significant (n = 302)
COPD and Hypertension	109% (n = 2,126)	96% (n = 2,074)	Not significant (n = 3,587)	Not significant (n = 3,437)

At the one percent significance level, the coefficients on MM interactions in the regression are mostly negative, except for COPD and Hypertension. This implies that the cost of each of the conditions is less if a patient also has the other condition. For example, Table 5 shows that a male patient with diabetes and CHD will cost between 77% and 78% (depending on the definition of sample prevalence) of the cost of treating 2 patients, one with diabetes and one with CHD, controlling for age.

Discussion of results for males

The mean annual cost of CHD is £2,514.17 (95% confidence interval: £2,328.41 to £2,714.37) under Definition A, and a mean annual cost of £1,992.99 (95% confidence interval: £1,860.23 to £2,133.74) under Definition B. This is the overall cost imposed by having CHD and no other condition of interest, including the baseline cost. While we have controlled for other conditions, the baseline cost may capture the additional healthcare cost imposed by having other diseases (for example the flu virus or other conditions not included in this analysis). Therefore, care should be taken to interpret the results as average 'costs per case' of individuals with conditions/MMs, rather than the additional treatment costs imposed by the conditions/MMs themselves.

The cost of individuals with CHD and another condition of interest is generally smaller than the combination of the costs of patients each with one of the conditions (less than additive), but the interaction coefficient is not significant at the 1% significance level under Definition A, apart from individuals who suffer from CHD and diabetes. Diabetic patients have an average cost in 2015 of £1,869.97 under Definition A and £1,605.49 under Definition B. Patients with diabetes and CHD have an average 2015 cost of £3,375.84 in Definition A and £2,823.38 in Definition B. The cost of MM is only significantly less than additive in the latter case. Under Definition B, the average cost per case of patients with both CHD and COPD or dementia also suggests some economies of scale with a mean cost of £4,459.93 (95% confidence interval £3,954.22 to £ 5,024.29) for CHD and COPD, and of £4,435.89 (95% confidence interval: £3,661.64 and £5,357.31) for CHD and dementia.

The mean cost of dementia is £4,344.68 (95% confidence interval: £3,867.08 and £4,880.23) under Definition A, and £3,472.82 (95% confidence interval: £3,041.11 to £3,954.84) under Definition B. Patients with both stroke and dementia have a significantly lower cost than the sum of 2 condition costs independently, with a mean cost of £12,281.55 (95% confidence interval: £8,250.69 to £18,277.41) under Definition A, and £6,494.58 (95% confidence interval: £5,113.99 to £ 8,185.69) under Definition B. However, for this specific MM combination, the number of cases is small: 38 in Definition A and 95 in Definition B. This leads to wide confidence intervals. Dementia and COPD patients also have less than additive costs, estimated to be on average £8,295.63 (in 2015), but ranging between £6,507.09 and £10,573.40 under Definition A

(95% confidence interval), and from £5,494.7 to £8,419.28 (95% confidence interval) with a mean cost of £6,821.91 under Definition B.

The mean cost of COPD is £3,488.11 (95% confidence interval: £3,273.97 to £3,715.73) under Definition A and £3,021.95 (95% confidence interval: £2,816.17 to £3,240.47) under Definition B. In addition to the interaction mentioned above, patients with both COPD and hypertension are on average cheaper to treat than the combination of the individual costs in Definition A, costing on average £6,586.59 (95% confidence interval: £5,974.36 to £7,260.34) in 2015.

This list of significant morbidity interactions is not exhaustive, but the interpretation of the remaining results is the same as described above. Note that a large number of morbidity interaction coefficients are not significant, implying that the cost of a patient with MMs is not different from the sum of 2 patients with one of the conditions each (in grey in the table). For example, the mean cost of the combination of liver disease and dementia is £27,900.94 with a 95% confidence interval varying between £20,986.37 and £79,244.15 in Definition A and of £10,458.85 with a 95% confidence interval varying between £7,311.10 and £14,709.91 in Definition B.

The only MM interaction coefficient that has a positive sign was on diabetes and hypertension, under definition A, but the coefficient is only significant at a 10% confidence level. The same coefficient under definition B is negative but not significant suggesting no noticeable effect of having the combination of diseases compared to the sum of the 2 individual costs.

In summary, all the interaction term coefficients significant at the 1% confidence level are negative, and therefore the presence of MMs are at least not greater than additive and often less than additive for individuals of the same age and gender. However, comparison of the overall individual averages as done in Table 5 suggests that this is not the case at the society level. This is because the average cost per case estimates are weighted by the number of individuals in different age groups, and individuals with MMs tend to be older and therefore, more expensive. Cost comparison of individuals in the same age group always suggests that the cost of MM is less than the additive cost of patients with different conditions. However, at the population level, the overall comparison shows that in some cases the presence of MM increases the sum of the costs of individuals with each of the conditions.

Discussion of results for females

For females, the mean cost of CHD is £2,847.86 (95% confidence interval: £2,608.15 to £3,109.24) under Definition A and £1,961.43 (95% confidence interval: £1,802.75 to £2,130.37) under Definition B. Similarly, to males, females with CHD and another condition of interest have costs generally smaller than the combination of the individual costs, but the interaction coefficients are often not significant at the 1% significance

level in Options A and B (for example CHD and breast cancer). Under Definition B, CHD with COPD has a mean cost of £4,142.17 (95% confidence interval: £3,649.36 to £4,692.30), and CHD with dementia has a mean cost of £3,897.26 (95% confidence level: £3,333.75 and £4,537.42).

The mean cost of dementia for females, as observed for most combinations of conditions, is lower than that for males, with an average cost of £3,243.43 (95% confidence interval: £3,018.75 to £3,484.32) under Definition A, and £2,518.75 (95% confidence interval: £2,307.52 to £2,744.79) under Definition B in 2015. Female patients suffering from both stroke and dementia have a mean cost of £7,611.60 (95% confidence interval: £5,801.07 to £9,986.79) under Definition A, and £4,615.89 (95% confidence interval: £3,847.14 to £5,496.01) under Definition B, which is also lower than for males. Individuals with both dementia and COPD also had lower costs than the combination of the 2 individual costs under both options. The mean cost was £6,438.21 (95% confidence interval: £5,316.74 to £7,795.10) under Definition A, and £5,258.65 (95% confidence interval: £4,470.42 to £6,161.27) under Definition B.

The MMs where the interaction coefficient is not significant at the 5% level are greyed out in the tables. For example, the mean cost of the combination of liver disease and diabetes is £10,305.36 with a confidence interval varying between £8,222.69 and £12,914.07 in Definition A and of £5,583.25 with a confidence interval varying between £4,771.46 and £6,498.35 in Definition B. All the significant interaction coefficients are negative, meaning that the cost of a patient with these MMs are less than the sum of the costs of 2 patients, each with one condition each. This may imply the existence of economics of scale when treating individuals with 2 of the conditions of interest.

Social care costs using the Somerset Symphony Dataset

The estimated costs are shown in Tables 6 and 7 below. Most multimorbidity costs are not statistically significant, possibly due to small sample sizes.

From the results, the presence of MM does not appear to affect social care costs at the 1% significance level. While the baseline cost is always significant, this is not always the case for the cost of individuals with only one condition. Depression, diabetes, liver disease, and stroke patients do not have a social care costs higher than the baseline. Only hypertension and dementia have an average cost significantly higher than the baseline.

Most results are similar between genders, except for dementia, which has a higher cost per case for females. This could be because females live longer, and therefore females with dementia may be older. Furthermore, they are likely to outlive male partners, so they are less likely than males to be recipients of informal social care.

Annual costs for females that are significant at a 1% significance level range between £91 and £4,280, while annual costs for males range between £85 and £2,985. These may initially seem relatively low. There are several potential reasons why costs may be higher in reality than estimated here.

The Somerset symphony dataset population is based in South Somerset, a relatively affluent part of the country. Previous research has found a steep socio-economic gradient in social care costs (47% difference between the least and most deprived IMD group) when looking at a similar linked dataset in Kent (52). This is likely to occur for 2 reasons which are that:

- more affluent groups are more likely to be in better health, and therefore have lower social care needs
- the methodology applied only considers local authority funded social care – this is likely to further decrease costs for a relatively affluent sample of the population, who are more likely to self-fund when they have a social care need

Table 6. Cost results for males: social care

MALE	Definition (Option A)					Definition (Option B)				
CONDITION/MM	Sample Size	Mean cost (£, 2014/2015)		LL (95%)	UL (95%)	Sample Size	Mean cost (£, 2014/2015)		LL (95%)	UL (95%)
BASELINE	184,212	139	***	118	164	149,705	91	***	75	110
CHD	1,716	173	***	128	234	2,281	97	*	56	169
CHD AND COPD	193	161		109	237	423	275		143	521
CHD AND DIABETES	415	163		114	234	457	184		88	381
COLORECTAL CANCER	200	134	**	83	216	273	138	**	66	287
COPD	618	188	***	138	256	10,342	129	*	90	184
DEMENTIA	681	822	***	658	1027	595	2985	***	2231	3966
DEMENTIA AND CHD	67	659		446	973	79	2602		1398	4608
DEMENTIA AND COPD	35	691		446	1070	71	3740		2081	6385
DEMENTIA AND DIABETES	107	674		477	952	89	3166		1805	5321
DEMENTIA AND HYPERTENSION	251	548		411	732	387	2579		1862	3535
DEMENTIA AND STROKE	32	963		605	1532	40	2797		1265	5608
DEPRESSION	4,346	103		72	148	8,455	151		101	227
DEPRESSION AND CHD	62	111		63	198	166	204		66	608
DEPRESSION AND COPD	49	168		98	289	1027	237		114	489
DEPRESSION AND DEMENTIA	38	596		356	999	63	4080		2176	7268
DEPRESSION AND DIABETES	238	127		76	213	465	285		124	646
DEPRESSION AND HYPERTENSION	439	125		82	190	1166	268		157	454
DEPRESSION AND LIVERD	55	39		17	88	154	259	*	76	859
DEPRESSION AND STROKE	24	296		155	567	62	2113		840	5012
DIABETES	8,800	169	*	133	215	6,682	144		98	211
DIABETES AND COPD	103	211		138	321	698	252		136	466

HYPERTENSION	15,760	157	***	126	194	20,657	85	***	65	111
HYPERTENSION AND CHD	1,748	189	**	143	251	3,818	186	*	132	263
HYPERTENSION AND COPD	445	180		129	252	2,655	149		99	225
HYPERTENSION AND DIABETES	3,107	161		123	211	5,270	164		117	231
HYPERTENSION AND STROKE	354	288		207	401	826	639		426	951
LIVERD	449	113		63	205	822	180		76	424
LIVERD AND CHD	26	231		110	482	40	428		79	2020
LIVERD AND COPD	26	195		92	414	123	218		49	929
LIVERD AND DEMENTIA	15	530		229	1227	14	1405		184	7612
LIVERD AND DIABETES	69	51		25	102	130	311		84	1116
LIVERD AND HYPERTENSION	155	118		65	212	357	159		61	406
LIVERD AND STROKE	7	212		76	592	12	1485		264	6296
LUNG CANCER	107	145	**	89	236	87	243	***	105	550
STROKE	524	351		247	500	639	727		441	1191
STROKE AND CHD	48	224		138	364	78	1141		526	2349
STROKE AND COPD	20	350	*	200	612	77	1014		391	2472
STROKE AND DIABETES	70	285		177	458	83	1169		495	2626

Note: The mean cost reports the average individual and annual cost of patients in a specific category in 2014/2015. If the MM interaction coefficient is not significant, it implies that the cost of a patient with MM is not different from the sum of 2 patients with one of the conditions each. The 95% confidence interval is reported for the average cost distribution. Option A assumes a person had contact with the health system due to their condition in 2014/15, Option B included any person who had contact with the health system due to their condition between 2011/12 and 2014/15 . The significance levels are reported for the following thresholds: * significant at 10% (in grey), ** significant at 5%, *** significant at 1%. LL: lower limit; UL: upper limit. Results with less than 5% significance level are greyed out.

Table 7. Cost results for females: social care

FEMALE	Definition (Option A)					Definition (Option B)				
CONDITION/MM	Sample Size	Mean cost (£, 2014/15)		LL (95%)	UL (95%)	Sample Size	Mean cost (£, 2014/15)		LL (95%)	UL (95%)
BASELINE	191169	199	***	176	225	147597	91	***	77	109
BREAST CANCER	1190	193	**	130	288	1843	106		53	208
BREAST CANCER AND COPD	19	207		95	452	186	83		21	314
BREAST CANCER AND DIABETES	61	347	*	178	675	99	140		21	850
CHD	879	514	***	410	644	1031	492		324	743
CHD AND BREAST CANCER	23	351		174	709	23	423		62	2169
CHD AND COPD	96	395		283	550	249	632		349	1116
CHD AND DIABETES	164	499		369	675	160	591		270	1240
COLORECTAL CANCER	160	186	***	126	275	252	236	***	136	407
COPD	603	320	***	251	409	12320	117	***	88	156
DEMENTIA	1101	1789	***	1547	2068	942	4131	***	3382	5025
DEMENTIA AND BREAST CANCER	16	797		421	1509	23	2180		641	6193
DEMENTIA AND CHD	74	1572		1186	2084	92	4208		2577	6586
DEMENTIA AND COPD	43	1462		1043	2049	93	4239		2710	6370
DEMENTIA AND DIABETES	112	1707		1325	2198	93	5229		3364	7810
DEMENTIA AND HYPERTENSION	495	1598	**	1347	1896	933	4280	**	3561	5122
DEMENTIA AND STROKE	59	1426		999	2037	80	3797	*	2182	6281
DEPRESSION	8371	95	***	75	121	15760	127		95	170

DEPRESSION AND BREAST CANCER	64	99		42	230	182	39		3	437
DEPRESSION AND CHD	52	383		251	587	123	587		245	1323
DEPRESSION AND COPD	72	190		124	293	2458	241		162	359
DEPRESSION AND DEMENTIA	73	1163		853	1587	143	3893		2602	5674
DEPRESSION AND DIABETES	342	125		87	179	723	396		227	685
DEPRESSION AND HYPERTENSION	783	265	**	205	344	2172	433		319	585
DEPRESSION AND LIVER DISEASE	58	129		63	266	184	286		96	822
DEPRESSION AND STROKE	26	514		332	797	73	1611		792	3109
DIABETES	7074	328		275	393	5736	220		158	307
DIABETES AND COPD	80	275		190	397	714	259		146	454
HYPERTENSION	18261	350	***	308	397	23267	239	***	204	280
HYPERTENSION AND BREAST CANCER	320	264		173	403	729	289		167	497
HYPERTENSION AND CHD	1135	485		400	587	2344	521		407	666
HYPERTENSION AND COPD	382	329		255	425	3227	299	*	226	395
HYPERTENSION AND DIABETES	2712	382		319	458	4382	397		312	506
HYPERTENSION AND STROKE	384	731		575	929	843	1347		1007	1790
LIVER DISEASE	404	143	***	96	214	755	241	**	131	440
LIVERD AND BREAST CANCER	14	33		12	89	33	57	*	4	714
LIVERD AND CHD	13	457		271	768	20	610		132	2219
LIVERD AND COPD	29	133		76	234	119	519	*	205	1270
LIVERD AND DEMENTIA	11	992		563	1747	19	2747		955	6450
LIVERD AND DIABETES	82	89		51	154	93	198		43	860
LIVERD AND HYPERTENSION	169	227		156	330	402	605	*	361	998
LIVERD AND STROKE	-	285		133	611	7	2815		598	8307
LUNG CANCER	95	240	***	162	356	67	226	***	105	474

STROKE	488	746		581	958	543	1316		879	1951
STROKE AND BREAST CANCER	8	474		192	1169	17	1333		252	5166
STROKE AND CHD	25	1018		687	1508	36	1555		570	3701
STROKE AND COPD	13	508		297	869	78	1293		610	2596
STROKE AND DIABETES	44	881		604	1285	53	2324		1051	4720

Note: The mean cost reports the average individual and annual cost of patients in a specific category in 2014/15. If the MM interaction coefficient is not significant, it implies that the cost of a patient with MM is not different from the sum of 2 patients with one of the conditions each. The 95% confidence interval is reported for the average cost distribution. Option A assumes a person had contact with the health system due to their condition in 2014/15, Option B included any person who had contact with the health system due to their condition between 2011/12 and 2014/15. The significance levels are reported for the following thresholds: * significant at 10% (in grey), ** significant at 5%, *** significant at 1%. LL: lower limit; UL: upper limit. Results with less than 5% significance level are greyed out. Sample sizes recorded as – have been small number supressed and represent a value under 5.

Social care costs using the quality life approach

The estimated social care costs using the QoL approach are shown in Table 7. The most expensive costs per case were for patients with dementia and patients with stroke, with mean annual costs of £19,413 and £5,017 respectively.

Table 7. Social Care Costs estimated using the Quality of Life approach

CONDITION/MM	QOL	PROBABILITY OF USING RESIDENTIAL CARE (%)	MULTIPLIER	ANNUAL COSTS PER PERSON WITH DISEASE (£)
COPD	0.47	13.03	1	£2,160
DIABETES	0.661	6.82	1	£1,131
CHD	0.76	3.6	1	£597
STROKE	0.713	5.13	5.9	£5,017
HYPERTENSION	0.721	4.87	1	£807
COLORECTAL CANCER	0.676	6.33	1	£1,050
DEPRESSION	0.6	8.80	1	£1,459
LUNG CANCER	0.56	10.1	1	£1,675
BREAST CANCER	0.75	3.93	1	£651
DEMENTIA	0.442	13.94	8.4	£19,413
LIVER DISEASE	0.62	8.15	1	£1,352
DIABETES AND CHD	0.50236	11.97	1	£1,986
DIABETES AND DEPRESSION	0.3966	15.41	1	£2,556
CHD AND DEPRESSION	0.456	13.48	1	£2,236
DIABETES AND HYPERTENSION	0.476581	12.81	1	£2,125
DEPRESSION AND	0.4326	14.24	1	£2,362
COPD AND HYPERTENSION	0.33887	17.29	1	£2,867

Note that because we have assumed that the utility weights for MMs are multiplicative (see methodology section for further detail), this implies that the costs for patients with MM are less than the costs for patients, each with one of the diseases. This is a direct consequence of our assumption about utilities and should not be interpreted as a ‘finding’ of the analysis.

This methodology produces larger costs than those produced when analysing the Symphony data. For example, this methodology produces and average cost of £807 for individuals with hypertension, while the estimated cost is £85 and £239 for males and females respectively. These costs however, are not comparable as this methodology

produces the average cost of social care for individuals with these NCDs and MMs aged 75 and over, whereas the first approach produces a cost that’s applicable to the whole disease population.

Discussion

In this report, the literature on methods used for costing MM was reviewed. This informed the methods and parameters included in the analyses and confirmed that our methods to estimate the costs of MM were the most appropriate ones.

The main characteristics of our methods were:

- an econometric/bottom-up perspective
- a representative primary and secondary care dataset (CPRD and HES)
- a sensitivity analysis on the number of years of diagnosis history taken into account to identify disease prevalence in the sample (Definitions A and B)
- a two-part method to account for individuals not using healthcare

Our findings showed that the identification of individuals with the conditions of interest over one year (Definition A) always provided higher cost estimates than the consideration of 4 years of diagnosis (Definition B) as it included only individuals who required healthcare for the condition of interest within a year (who could be interpreted as more acute cases). Definition B on the contrary considered additional individuals, who only needed to have seen a doctor within the last 4 years to still have the condition of interest. This assumption brought down the average cost per case estimated based on healthcare utilisation in 2015. No option was superior, and the results should be considered within their context. Definition A exclusively considers the current healthcare users, whereas Definition B comes closer to estimating average cost based on population prevalence. However, due to the limitations of the data, prevalence as defined by Definition B should be interpreted as an estimate of sample prevalence, not population prevalence.

Some care should be taken when using the estimates in this report because sample prevalence in our dataset was determined by identifying patients with specific codes corresponding to conditions. For some conditions, choosing which codes to include and exclude was not straightforward, and we had to make assumptions based on the guidance of the project's Steering Group. There are some limitations to the codes we have used which are:

- we are including a code for secondary malignant lung tumours, which may lead to overstating the cost levels associated with lung cancer patients
- our definition of colorectal cancer includes anal cancer (C21); this is a broader definition than is usually considered to be colorectal cancer by the National Statistics.

Our prevalence figures refer to the sample prevalence of diseases in our dataset. They only include individuals who have required healthcare in the last year ("Definition A" definition), or the last 4 years ("Definition B" definition) for the conditions of interest, and therefore should not be treated as comparable to population prevalence estimates in England. Furthermore, our cost estimates are for 2015 and should be adjusted for inflation if compared to different years. Differences in treatment approaches over the years could also introduce some variation in the costs.

The Market Forces Factor (MFF) is an estimate of unavoidable cost differences between health care providers, based on their geographical location. It is an index with a minimum value of 1 and a maximum value of 1.3. We are underestimating true secondary care costs by about 15% on average, as we could not apply the MFF index. This is because the CPRD does not provide information allowing the identification of single hospitals.

Direct comparison with the literature was not possible because, to our knowledge, no other study has carried out such analyses in England for all of the costs included in the present study (primary care, hospitalisation, prescription, and social care costs). For example, Brilleman et al (2013) (38) only reported primary care cost, and Glynn et al. (2011) (27) only considered patients above 50 years old. While other studies did not include a reference group (that is an average healthcare cost for individuals who have seen the GP but did not have any of the condition of interest). Brilleman and colleagues did find variations in cost-increasing and cost-limiting co-morbidities. For example, they found that in their sample, depression was the most cost-increasing condition, while hypertension was cost-limiting, especially when occurring with other cardiovascular conditions. However, their definitions of disease and the data they analysed differed from the ones used in our analysis.

The University of York (16) used the Somerset Symphony dataset to calculate costs of MM, including both healthcare and social care costs. However, their method was very different to the one applied here since the authors ran an econometric model on the subsample of people with a given disease. This subsampling approach statistically means designing a restricted model, where the intercept of the model is different for each disease. Therefore, it is not possible to compare our results with the ones in this analysis, as the coefficients have a different interpretation.

In our study, the significant interaction term coefficients at the 95% confidence interval were negative, therefore costs of MM were limiting (or not different from the sum of the costs of the 2 conditions when the coefficients were not significant). Drawing on the conclusions of Brilleman and colleagues, this may be because individuals with more than one condition are treated for both conditions in a single consultation, therefore reducing the costs to less than additive compared with if 2 people had a single condition each and 2 consultations would be required to treat them. Alternatively, it

could be due to overlapping treatments or inadequate care. However, these explanations are purely speculative and further work is necessary to substantiate these inferences.

This study has several strengths. Firstly, the datasets used for this study to estimate healthcare costs (HES, CPRD) are large and representative of the English population. Secondly, the robustness of the methodology (two-step method) to estimate the combined costs of disease is the best available and allows us to account for 2 different types of individuals in the population, namely those very likely to require healthcare, and those unlikely to need healthcare over the year of observation.

There are also limitations associated with this study. First, the doctors' diagnosis cannot be systematically associated with a specific condition, often because the cause of the problem may not be known at the time of the visit. As a second-best alternative, the net overall healthcare costs of a patient believed to have a specific condition have been estimated.

Second, despite the richness of the data used in this project, there are still data entries that are not clear or meaningful, such as a 30-second GP appointment. However, this is a known problem in the literature, which has been encountered by other users, and is likely to only impact a small proportion of the overall costs (53).

Third, we did not have much information on the composition of the healthcare costs reference group. We have only focused on 11 health conditions, and the baseline costs will include all healthcare costs unrelated to or not impacted by these conditions.

Fourth, we assumed that the long-term chronic conditions considered in this report were not resolved within the one-year and four-year periods.

Fifth, when calculating social care costs using the QoL method we assumed that QoL was multiplicative which implies that costs of MM are greater than additive.

Sixth, the two-step methodology that was conducted for social care used the Symphony dataset is a small sample of the English population from an affluent area, which may mean the results aren't representative of the whole country.

Finally, we were unable to determine whether any of the costs estimated could be avoided by improving the management of conditions, or whether the costs are unavoidable. Further modelling would be required to better understand the nature and cause of the costs incurred. While some combinations of MM may be unavoidably expensive to treat, it could be the case that for example 2 parts of the health system required to treat the patient do not collaborate efficiently.

There are improvements that could be considered in future work. First, the cost estimates could be broken down by age category as there is likely to be large heterogeneity in the different condition-age cost distribution. In line with previous findings, the average age of patients with MM is higher that patients with a single condition. Therefore, we cannot sum/subtract the costs from each other meaningfully. Further work might disaggregate the costs by age group to explore whether age differences impact the cost-limiting or cost-increasing effect of MM. Note that age was included as a confounder in our analysis.

Further extensions could also consider the cost distribution by the severity of conditions. The current work only controls for age and gender, but one could consider not only different condition types, but also condition-severity based for example on the number of years since the first diagnosis, or sub-categories of condition indicators. Another aspect that could be analysed is the impact of delaying treatment on current healthcare costs. However, this angle requires information beyond the medical records. It may be possible to gain this information from surveys.

The cost distributions could be further investigated to understand what complications of MM drive the costs. For example, a research question examining patients with diabetes and another condition could be to look at the impact of amputations or eye problems. The work could also be expanded to include the impact of lifestyle and genetic factors.

Finally, more work should be done to disentangle inappropriate resource use versus truly expensive conditions. This could be done by matching patients with others with the same types of diagnoses and investigating whether, for the same outcomes achieved, some patients required significantly less care. However, some variation due to unobservable factors would always be present.

References

1. Bayliss EA, Bayliss MS, Ware JE, Jr., Steiner JF. Predicting declines in physical function in persons with multiple chronic medical conditions: what we can learn from the medical problem list. Health Qual Life Outcomes. 2004;2:47.

2. Fortin M, Lapointe L, Hudon C, Vanasse A, Ntetu AL, Maltais D. Multimorbidity and quality of life in primary care: a systematic review. Health Qual Life Outcomes. 2004;2:51.

3. Fortin M, Bravo G, Hudon C, Lapointe L, Dubois M-F, Almirall J. Psychological Distress and Multimorbidity in Primary Care. The Annals of Family Medicine. 2006;4(5):417-22.

4. Condelius A, Edberg AK, Jakobsson U, Hallberg IR. Hospital admissions among people 65+ related to multimorbidity, municipal and outpatient care. Arch Gerontol Geriatr. 2008;46(1):41-55.

5. Bramley D, Moody D. Multi-morbidity - the biggest clinical challenge facing the NHS? 2016. Available from: https://www.england.nhs.uk/blog/dawn-moody-david-bramley/.

6. Barnett K, Mercer SW, Norbury M, Watt G, Wyke S, Guthrie B. Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. Lancet. 2012;380(9836):37-43.

7. Valderas JM. Increasing clinical, community, and patient-centered health research. 2013. 2013;3(2):4.

8. Ording AG, Sørensen HT. Concepts of comorbidities, multiple morbidities, complications, and their clinical epidemiologic analogs. Clinical epidemiology. 2013;5:199.

9. Fortin M, Bravo G, Hudon C, Vanasse A, Lapointe L. Prevalence of Multimorbidity Among Adults Seen in Family Practice. The Annals of Family Medicine. 2005;3(3):223-8.

10. Wang L, Si L, Cocker F, Palmer AJ, Sanderson K. A Systematic Review of Cost-of-Illness Studies of Multimorbidity. Appl Health Econ Health Policy. 2018;16(1):15-29.

11. Sambamoorthi U, Tan X, Deb A. Multiple chronic conditions and healthcare costs among adults. Expert Rev Pharmacoecon Outcomes Res. 2015;15(5):823-32.

12. Kuo RN, Lai MS. The influence of socio-economic status and multimorbidity patterns on healthcare costs: a six-year follow-up under a universal healthcare system. Int J Equity Health. 2013;12:69.

13. Charlson M, Wells MT, Ullman R, King F, Shmukler C. The Charlson comorbidity index can be used prospectively to identify patients who will incur high future costs. PLoS One. 2014;9(12):e112479.

14. Whitty Christopher J M, MacEwen Carrie, Goddard Andrew, Alderson Derek, Marshall Martin, Calderwood Catherine et al. Rising to the challenge of multimorbidity *BMJ* 2020; 368 :l6964

15. Pimpin L, Retat L, Fecht D, de Preux L, Sassi F, Gulliver J, et al. Estimating the costs of air pollution to the National Health Service and social care: An assessment and forecast up to 2035. PLOS Medicine. 2018;15(7):e1002602.

16. Centre for Health Economics University of York. The importance of multi-morbidity in explaining utilisation and costs across health and social care settings: Evidence from South Somerset's Symphony Project. 2014.

17. Walid MS, Robinson JS, Jr. Economic impact of comorbidities in spine surgery. J Neurosurg Spine. 2011;14(3):318-21.

18. Stundner O, Kirksey M, Chiu YL, Mazumdar M, Poultsides L, Gerner P, et al. Demographics and perioperative outcome in patients with depression and anxiety undergoing total joint arthroplasty: a population-based study. Psychosomatics. 2013;54(2):149-57.

19. Lehnert T, Konig HH. [Effects of multimorbidity on health care utilization and costs]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz. 2012;55(5):685-92.

20. Lehnert T, Heider D, Leicht H, Heinrich S, Corrieri S, Luppa M, et al. Review: health care utilization and costs of elderly persons with multiple chronic conditions. Med Care Res Rev. 2011;68(4):387-420.

21. Wang L, Si L, Cocker F, Palmer AJ, Sanderson K. A Systematic Review of Cost-of-Illness Studies of Multimorbidity. Appl Health Econ Health Policy. 2017.

22. 22.NICE. Developing NICE guidelines: the manual. 2014.

23. Orueta JF, Garcia-Alvarez A, Garcia-Goni M, Paolucci F, Nuno-Solinis R. Prevalence and costs of multimorbidity by deprivation levels in the basque country: a population based study using health administrative databases. PLoS One. 2014;9(2):e89787.

24. Nagl A, Witte J, Hodek JM, Greiner W. Relationship between multimorbidity and direct healthcare costs in an advanced elderly population. Results of the PRISCUS trial. Z Gerontol Geriatr. 2012;45(2):146-54.

25. Carreras M, Ibern P, Coderch J, Sanchez I, Inoriza JM. Estimating lifetime healthcare costs with morbidity data. BMC Health Serv Res. 2013;13:440.

26. Cortaredona S, Ventelou B. The extra cost of comorbidity: multiple illnesses and the economic burden of non-communicable diseases. BMC Med. 2017;15(1):216.

27. Glynn LG, Valderas JM, Healy P, Burke E, Newell J, Gillespie P, et al. The prevalence of multimorbidity in primary care and its effect on health care utilization and cost. Fam Pract. 2011;28(5):516-23.

28. Bahler C, Huber CA, Brungger B, Reich O. Multimorbidity, health care utilization and costs in an elderly community-dwelling population: a claims data based observational study. BMC Health Serv Res. 2015;15:23.

29. Pressley JC, Louis ED, Tang MX, Cote L, Cohen PD, Glied S, et al. The impact of comorbid disease and injuries on resource use and expenditures in parkinsonism. Neurology. 2003;60(1):87-93.

30. Pacula RL, Ringel J, Dobkin C, Truong K. The incremental inpatient costs associated with marijuana comorbidity. Drug Alcohol Depend. 2008;92(1-3):248-57.

31. Specogna AV, Turin TC, Patten SB, Hill MD. Hospital treatment costs and length of stay associated with hypertension and multimorbidity after hemorrhagic stroke. BMC Neurol. 2017;17(1):158.

32. Basu A, Manning WG, Mullahy J. Comparing alternative models: log vs Cox proportional hazard? Health economics. 2004;13(8):749-65.

33. Wacker ME, Jorres RA, Schulz H, Heinrich J, Karrasch S, Karch A, et al. Direct and indirect costs of COPD and its comorbidities: Results from the German COSYCONET study. Respir Med. 2016;111:39-46.

34. Wagner CJ, Metzger FG, Sievers C, Marschall U, L'Hoest H, Stollenwerk B, et al. Depression-related treatment and costs in Germany: Do they change with comorbidity? A claims data analysis. J Affect Disord. 2016;193:257-66.

35. Schwab P, Dhamane AD, Hopson SD, Moretz C, Annavarapu S, Burslem K, et al. Impact of comorbid conditions in COPD patients on health care resource utilization and costs in a predominantly Medicare population. Int J Chron Obstruct Pulmon Dis. 2017;12:735-44.

36. Pagano E, Bo S, Petrinco M, Rosato R, Merletti F, Gregori D. Factors affecting hospitalization costs in Type 2 diabetic patients. J Diabetes Complications. 2009;23(1):1-6.

37. Thiébaut S, Barnay T, Ventelou B. Ageing, chronic conditions and the evolution of future drugs expenditure: a five-year micro-simulation from 2004 to 2029. Applied Economics. 2013;45(13):1663-72.

38. Brilleman SL, Purdy S, Salisbury C, Windmeijer F, Gravelle H, Hollinghurst S. Implications of comorbidity for primary care costs in the UK: a retrospective observational study. Br J Gen Pract. 2013;63(609):e274-82.

39. Briggs, A.D.M., Cobiac, L.J., Wolstenholme, J. et al. PRIMEtime CE: a multistate life table model for estimating the cost-effectiveness of interventions affecting diet and physical activity. BMC Health Serv Res 19, 485 (2019) doi:10.1186/s12913-019-4237-4

40. NICE. Methodology for estimating “Wider Societal Benefits” as the net production impact of treatments 2013. Available from: https://www.nice.org.uk/Media/Default/About/what-we-do/NICE-guidance/NICE-technology-appraisals/DH-Documentation-for-Wider-Societal-Benefits.pdf.

41. Canavan C, West J, Card T. Calculating Total Health Service Utilisation and Costs from Routinely Collected Electronic Health Records Using the Example of Patients with Irritable Bowel Syndrome Before and After Their First Gastroenterology Appointment. Pharmacoeconomics. 2016;34(2):181-94.

42. Jones AM. Health econometrics. Handbook of health economics. 1: Elsevier; 2000. p. 265-344.

43. Deb P, Trivedi PK. 14 Empirical models of health care use. The Elgar Companion to Health Economics. 2006:147.

44. Duan N, Manning WG, Morris CN, Newhouse JP. A comparison of alternative models for the demand for medical care. Journal of business & economic statistics. 1983;1(2):115-26.

45. Manning WG, Mullahy J. Estimating log models: to transform or not to transform? Journal of health economics. 2001;20(4):461-94.

46. Buntin MB, Zaslavsky AM. Too much ado about two-part models and transformation?: Comparing methods of modeling Medicare expenditures. Journal of health economics. 2004;23(3):525-42.

47. Curtis, Lesley A. and Burns, Amanda (2017) Unit Costs of Health and Social Care 2017. Report number: https://doi.org/10.22024/UniKent/01.02/65559. Personal Social Services Research Unit, University of Kent, 260 pp. ISBN 978-1-911353-04-1.

48. Sullivan PW, Slejko JF, Sculpher MJ, Ghushchyan V. Catalogue of EQ-5D scores for the United Kingdom. Medical Decision Making. 2011 Nov;31(6):800-4.

49. Laires PA, Ejzykowicz F, Hsu TY, Ambegaonkar B, Davies G. Cost-effectiveness of adding ezetimibe to atorvastatin vs switching to rosuvastatin therapy in Portugal. J Med Econ. 2015;18(8):565–72.

50. Rivero-Arias O, Ouellet M, Gray A, Wolstenholme J, Rothwell PM, Luengo-Fernandez R. Mapping the modified Rankin scale (mRS) measurement into the generic EuroQol (EQ-5D) health outcome. Med Decis Making. 2010;30(3):341–54.

51. Turner N, Campbell J, Peters TJ, Wiles N, Hollinghurst S. A comparison of four different approaches to measuring health utility in depressed patients. Health and quality of life outcomes. 2013;11(1):1.

52. Jayatunga W, Asaria M, Belloni A, George A, Bourne T, Sadique Z. Social gradients in health and social care costs: Analysis of linked electronic health records in Kent, UK, Public Health, Volume 169, 2019, Pages 188-194, ISSN 0033-3506

53. Hobbs, F D Richard et al. Clinical workload in UK primary care: a retrospective analysis of 100 million consultations in England, 2007–14, The Lancet, Volume 387, Issue 10035, 2323 – 2330. 2016