CHERE
CENTRE FOR HEALTH ECONOMICS
RESEARCH AND EVALUATION

UTS

UTS:CHERE

CHERE
CENTRE FOR HEALTH ECONOMICS RESEARCH AND EVALUATION

The importance of comorbidity and multimorbidity in determining health care costs: an analysis of the cost amplifications associated with morbidity interaction variables

Working Paper 2018/01

April 2018

A report by the Centre for Health Economics Research and Evaluation

UTS
UNIVERSITY OF TECHNOLOGY SYDNEY

About CHERE

CHERE is an independent research unit affiliated with the University of Technology Sydney. It has been established since 1991, and in that time has developed a strong reputation for excellence in research and teaching in health economics and public health and for providing timely and high quality policy advice and support. Its research program is policy-relevant and concerned with issues at the forefront of the sub-discipline.

CHERE has extensive experience in evaluating health services and programs, and in assessing the effectiveness of policy initiatives. The Centre provides policy support to all levels of the health care system, through both formal and informal involvement in working parties, committees, and by undertaking commissioned projects. For further details on our work, see www.chere.uts.edu.au.

Authors:

Thomas Longden[1], Chun Yee Wong[1,2], Phil Haywood[1], Jane Hall[1] and Kees van Gool[1]

[1] Centre for Health Economics Research and Evaluation, University of Technology Sydney, Australia

[2] International University of Japan, Minami Uonuma-shi, Niigata-ken, Japan

Contact details:

Dr Thomas Longden
Centre for Health Economics Research and Evaluation (CHERE)
University of Technology Sydney
PO Box 123, Broadway NSW 2007

Ph: + 61 2 9514 4727
Fax: + 61 2 9514 4730
Email: thomas.longden@chere.uts.edu.au

Keywords: healthcare costs, morbidity, multimorbidity, simultaneous quantile regression

JEL Classification: I10; C31; C52

This research was completed using data collected through the 45 and Up Study (www.saxinstitute.org.au). The 45 and Up Study is managed by the Sax Institute in collaboration with major partner Cancer Council NSW; and partners: the National Heart Foundation of Australia (NSW Division); NSW Ministry of Health; NSW Government Family & Community Services – Carers, Ageing and Disability Inclusion; and the Australian Red Cross Blood Service. We thank the many thousands of people participating in the 45 and Up Study. The linked Medicare Benefits Scheme and the Pharmaceutical Benefits Scheme data were supplied to the 45 and Up Study by the Commonwealth Department of Human Services. The linked Admitted Patient Data Collection and Emergency Department Data Collection extracts were supplied by the New South Wales Ministry of Health. We also thank participants of the 38th Australian Health Economics Society Conference held in Fremantle, Western Australia for valuable comments and suggestions. The usual disclaimers apply.

Table of Contents

Abstract

Risk adjustment and payment systems rely on an accurate understanding of the drivers of healthcare costs. A related concern is whether certain cases of multimorbidity coincide with notably higher costs of healthcare. With the utilisation of administrative data for over 250,000 Australian residents, we investigate whether there are specific combinations of morbidities that are associated with amplified healthcare costs. Chi-square automated interaction detection is used to identify the key interactions of morbidities that drive healthcare cost across four healthcare sector categories (i.e. medical services, pharmaceutical, hospital and total healthcare). Using simultaneous quantile regressions, the performance of models that include morbidity interaction variables is assessed at multiple points of the distribution (i.e. the 10^{th}, 50^{th}, 90^{th}, 95^{th} and 98^{th} percentiles). Key drivers of risk included in these models include age, body mass index (BMI), individual morbidity groups and imminent death. Our work builds on previous studies that have tended to focus on aggregate expenditure or primary healthcare costs but have not assessed how model performance differs across the distribution of costs. For all four categories of healthcare costs we find that a morbidity dummy and interaction variable model performs notably better than a model that uses a count variable of the number of coexisting morbidities. This is especially the case for the right tail of the cost distributions.

1 Introduction

Redesigning healthcare interventions to better manage cases where more than one morbidity prevails has been identified as a key challenge for healthcare systems and, accordingly, this has fuelled interest in coordinated and integrated care programs (Guthrie et al. 2008, Barnett et al. 2012, Haggerty 2012, Smith et al. 2012, Tinetti, Fried, and Boyd 2012). The focus on comorbidity (i.e. two morbidities) and multimorbidity (i.e. three or more morbidities) has been, in part, motivated by the success of treating mid-life health conditions and concerns about the adequacy of healthcare systems that have been largely configured to treat individual diseases (Barnett et al. 2012, Salisbury 2012, Banerjee 2015). While there has been a broad discussion of multimorbidity and the consequences this has on managing the healthcare system, the literature that focuses on estimating the cost of healthcare has tended to treat morbidity as individual diseases and it is rare for these models to account for cost amplifications associated with morbidity interactions. In this paper, our aim is to identify morbidity interactions that are related to the cost of healthcare and confirm whether the inclusion of these interactions improves the performance of models of healthcare costs. Using simultaneous quantile regression, we perform this assessment for multiple points of four cost distributions with an allowance for key drivers of risk, specifically age, body mass index (BMI), individual morbidity groups and imminent death.

Refinements in the modelling of healthcare costs have widespread applications and will assist in making accurate adjustments for risk that are relevant to the development and revision of payments systems (including primary care capitation payments, hospital reimbursement based on case-mix weighted payments and the risk-equalisation of insurance premiums). Our focus on the importance of morbidity interaction variables is also motivated by previous studies that include morbidities within models of healthcare cost in an individual manner (Ellis et al. 2013, Buchmueller and Johar 2015, Jones, Lomas, and Rice 2015, Jones et al. 2016); studies that found that models with a count variable of coexisting morbidities performed better than models with complex multimorbidity measures (Brilleman et al. 2014, Islam et al. 2014); and recent papers that associated morbidity interaction

variables with limited improvement in the performance of risk adjustment models (Buchner, Wasem, and Schillo 2017, van Veen et al. 2017). It should be noted that a count variable of the number of coexisting morbidities and morbidity interaction variables are distinct and capture different relationships. A key distinction is the equal weighting of all morbidities when using a count variable, as opposed to, making an allowance for certain combinations of morbidities that have a simultaneous influence that is not additive (i.e. interaction variable).

The importance of morbidity interactions remains ambiguous. For example, Buchner, Wasem, and Schillo (2017) used a large dataset of nearly 2.9 million individuals residing in Germany to identify morbidity interactions using a regression tree analysis; and while they found a marginal improvement in models of risk adjustment, they concluded that the current approach of disregarding interactions was not problematic. In contrast, there are a range of studies that have found that multimorbidity is related to the cost of healthcare (Schoenberg et al. 2007, Lehnert et al. 2011, König et al. 2013, Sambamoorthi, Tan, and Deb 2015, van Veen et al. 2017). It is important to note that van Veen et al. (2017) used a similar methodology and found similar results to those in Buchner, Wasem, and Schillo (2017); however, they concluded that interaction terms can improve the prediction of healthcare costs for both the overall population and specific subgroups in the population.

The focus of this analysis is to investigate the importance of interactions between morbidities for accurately modelling aggregate healthcare costs, as well as, costs in three different healthcare sectors. The use of administrative data for over 250,000 Australian residents allows us to focus on the costs associated with multiple categories of healthcare sectors; specifically, medical services, pharmaceuticals, admitted patient care as well as the aggregate of these costs. At issue is the discrete treatment of morbidities and the use of count variables that do not account for the complexity that comorbidity and multimorbidity presents to the healthcare system. We expect that these complexities lead to increased costs of healthcare. Substantial increases in costs may be driven by a range of factors, such as a need to incur multiple hospitalisations to treat and monitor a patient's condition,

extended lengths of stay or the coincidence of a combination of conditions that make treatment more complex, intensive or risky. Identifying costly morbidity interactions has the potential to inform the development or prioritisation of policies that aim to prevent the escalation of costs associated with multimorbidity. In relation to multimorbidity, it has been noted that the system of specialist referrals is inefficient and ineffective (Wolff, Starfield, and Anderson 2002, Starfield, Shi, and Macinko 2005, Burgers et al. 2010). It has also been found that clinical and prescribing guidelines tend to provide limited advice to clinicians on how they should prioritise recommendations across multiple distinct diseases (Boyd et al. 2005, van Weel and Schellevis 2006, Van Spall et al. 2007). In accordance with this, it was noted in Barnett et al. (2012) that patients with multimorbidity may be prescribed several drugs that are recommended by individual disease-specific guidelines, but the simultaneity of certain conditions may result in the overall drug burden being difficult for patients to manage and potentially harmful (Barnett et al. 2012).

Previous studies have not examined the importance of multimorbidity in a model of healthcare costs at different points of the distribution. Our analysis will address this gap in the literature by using simultaneous quantile regression to examine model performance across multiple points of four cost distributions. This technique is employed to estimate the effects of morbidity, comorbidity and multimorbidity variables for patients who incur low (i.e. the 10th percentile), median (50th percentile), high (i.e. the 90th percentile) and extremely high (i.e. the 95th and 98th percentiles) healthcare costs. It should be noted that the focus on performance at different points of the distribution is, in part, motivated by the findings of Eckardt et al. (2016) as they found ‘two diametrically opposed cost trends’ that were related to the number of co-occurring diseases. It is also motivated by the work of de Meijer et al. (2013) and Jones, Lomas, and Rice (2015) who recommend further research on healthcare costs that explore the full distribution of healthcare costs and goes beyond a focus on estimating the mean.

We make a distinction between cases where patients have two morbidities (i.e. comorbidity) and three or more morbidities (i.e. multimorbidity) to distinguish between the complexity of the interactions in terms of the number of coexisting morbidities. It should be noted that there are different definitions of multimorbidity that are commonly used within the literature (Diederichs, Berger, and Bartels 2011, van den Akker et al. 2001, Le Reste et al. 2013, Harrison et al. 2014). Some studies specify that the medical conditions should be limited to chronic diseases, however, we follow the definition prescribed by van den Akker, Buntinx, and Knottnerus (1996), which includes chronic diseases, acute diseases and other medical conditions.

In this paper, we will use a statistical interaction detection approach to identify morbidity interaction variables that are related to high/low healthcare costs. With the use of Chi-square automated interaction detection (CHAID), we are able to exhaust a very large number of possible interactions to identify the morbidity interaction groups that are associated with each category of healthcare cost. The possible number of interaction variables that can built using 60 morbidity dummy variables (with 0 and 1 as unique classifications) and a large order of interactions is extremely large. Identifying the relevant interactions for four categories of healthcare without an automated technique is unlikely to result in an exhaustive range of interaction variables. Our argument in favour of using interaction detection in this manner is similar to that used by Belloni et al. (2014) when discussing variable and functional form selection with high dimensional data. Note that high dimensional data are data that have a large number of variables relative to the sample size (Belloni, Chernozhukov, and Hansen 2014a). Note that Einav and Levin (2014), Varian (2014) and Athey (2017) have also advocated for the use of automated data driven techniques within empirical economics. The use of the CHAID approach has been chosen as it is specifically designed for categorical variables, produces complex tree structures (as it allows for more than two splits) and reduces the likelihood of over-fitting (i.e. being biased towards selecting predictors with more categories) by filtering out variables that are not related to the dependent variable using a statistical association test (van Diepen and Franses 2006, Loh 2011). This approach allows us to develop morbidity interaction variables that are specified in

relation to deciles of a specific category of healthcare cost. Previous studies have been critiqued for using multimorbidity measures that were not designed with costs in mind (Lapi et al. 2015).

2 Methodology

Our methodology is described in three parts. The first section outlines the model of healthcare costs and specifies the procedure used to estimate costs for multiple points of the distribution. The second section outlines the interaction detection technique used to identify potentially important interaction variables. This includes a discussion of the need to distinguish between data mining as a 'fishing' exercise and as a variable specification tool applied to a well-defined set of variables. The third section outlines our use of the coefficient of determination to assess improvements in model performance within specific sub-samples based on the number of morbidities and whether the individual dies in the year of interest or the next year.

2.1 Specification of a model of annual healthcare cost

The first model of cost (Model 1 – M1) is specified in a similar manner to that in Brilleman et al. (2014) with differences in morbidity profiles captured using a count variable of the number of coexisting morbidities. The second model (Model 2 – M2) is built using a model specification similar to that used by Ellis et al. (2013), Brilleman et al. (2014) and Jones, Lomas, and Rice (2015) with morbidity included as distinct dummy variables. The third model (Model 3 – M3) that we estimate is expected to improve upon model two by including morbidity interaction variables. Equation one defines the model that determines the level of annual non-zero healthcare costs. Specified as subscripts are the four types of healthcare, h, that we focus on with costs, C, specified as an annual cost for each individual, i. The cost categories, h, are medical services, pharmaceuticals, admitted patient care and the aggregate of these costs. X_i is a vector of patient characteristics that are expected to influence the annual cost of each type of healthcare. For each type of healthcare β_h are the vectors of parameters that we will estimate and ε_{hi} are the disturbance terms.

$$C_{hi} = X_i\beta_h + \varepsilon_{hi} \quad (1)$$

We estimate the effects of the explanatory variables for different points of the conditional distribution using simultaneous quantile regression. The quantile regression model that we estimate for the τth percentile of each type of healthcare cost is given by

$$Q_\tau(C_{hi}) = X_i\beta_{h\tau} \quad (2)$$

where τ=0.1, 0.5, 0.9, 0.95, 0.98. This coincides with estimates for the 10th, 50th, 90th, 95th and 98th percentiles and allows us to confirm whether improvements in model performance are associated with low, median, high and extremely high healthcare costs. The equations for these five percentiles are estimated simultaneously using the sqreg command in Stata and the variance–covariance matrix of the estimators is obtained using bootstrapping.

Corresponding to the specification in the aforementioned papers, the variables included in the X_i vector of explanatory variables includes morbidity, demographic, economic, lifestyle and health status dummy variables. Note that Table 1A in the appendix contains the specification of the explanatory variables included in the three models of healthcare cost. In addition to the variables used in previous papers, we also include dummy variables that capture whether a patient dies during the year of observation or the next. This is motivated by research that has found that time to death is a notable driver of healthcare costs (Zweifel, Felder, and Meiers 1999, Werblow, Felder, and Zweifel 2007, Felder, Werblow, and Zweifel 2010). An additional difference to past studies of healthcare costs is the use of a morbidity index that simultaneously uses hospital diagnoses and ambulatory drugs dispensation to specify the morbidity grouping. The Diagnosis and Drug Morbidity Grouping (DDMG), which was developed by Halfon et al. (2013), is used as morbidities can be specified for all patients and cost distributions no matter whether a hospitalisation occurred or not. The DDMG specifies sixty morbidity classifications using the International Classification of Diseases (ICD) codes and the Anatomical, Therapeutic, Chemical (ATC) classification system. This provides sixty dummy

variables for morbidities that are included in models two and three. Within the third model we include dummy variables of morbidity interactions that capture comorbidity and multimorbidity health states. The first model is distinct as it does not have sixty morbidity dummy variables, but contains a count variable of the number of coexisting morbidities that is also constructed using the DDMG.

2.2 Identifying morbidity interactions

To identify the combinations of interactions to include in the multimorbidity healthcare cost model we decided that a statistical interaction detection technique would be most suitable. This was based on having an extremely large number of possible interactions between 60 morbidity variables and a limited number of studies that assess the importance of morbidity interactions with respect to multiple categories of healthcare costs. The specific technique that we use is the Chi-square automated interaction detection (CHAID) procedure that was first specified in Kass (1980) and developed into a command within Stata by Luchman (2013). CHAID is a recursive partitioning algorithm that builds a decision tree structure of split independent variables based on their statistical significance in relation to a dependent variable (Luchman 2013). We have chosen CHAID as the procedure within this paper as it is specifically designed for categorical variables, produces complex tree structures (as it allows for more than two splits) and reduces the likelihood of over-fitting (i.e. selecting predictors with more categories) by filtering out variables that are not related to the dependent variable using a statistical association test (i.e. a Chi-square specification). Refer to van Diepen and Franses (2006) and Loh (2011) for a discussion of these features of the CHAID procedure and how they differ in comparison to other classification and regression tree procedures.

As the underlying recursive estimation procedure utilises ordered logit estimation, we have used the deciles of each type of healthcare cost as the dependent variables. After the procedure is complete we obtain a set of interaction groups that is based on a large number of possible combinations of the sixty morbidity dummy variables (coded to match the DDMG) that result in interaction groups with more than 1000 individuals. Note that this constraint eliminates interaction groups that are only relevant to

a few individuals and is imposed to prevent a large number of interactions being specified. In contrast, Buchner, Wasem, and Schillo (2017) set a similar threshold to 30 individuals and then relied on a minimal R^2 improvement per split of 0.0001 as it is the level used in the German risk adjustment model. van Veen et al. (2017) set this threshold to be between 415 and 862 individuals based on the smallest risk class size within the Dutch risk equalisation model that is the basis of their paper.

It should be noted that our argument in favour of using interaction detection in this manner is similar to that used in Belloni, Chernozhukov, and Hansen (2014a) when discussing variable and functional form selection with high dimensional data. In that paper, Belloni, Chernozhukov, and Hansen (2014a) used a Least Absolute Shrinkage and Selection Operator (LASSO) approach to determine the variables and functional form selection for use in models of government seizure of private property, the effect of legalised abortion on crime and the effect of institutions on output. With hundreds of thousands of individuals in this data set, our data is high-dimensional in nature as the possible number of interaction variables using 60 morbidity dummy variables (with 0 and 1 as unique classifications) and a large order of interactions is extremely large. For example, with 120 morbidity classifications and 4 way interactions, the possible number of interaction variables is approximately 9 million. For a larger order of interactions, the possible number of interaction variables increases considerably.

Upon discussing wide-ranging attitudes towards data mining, Belloni, Chernozhukov, and Hansen (2014a) stated that “data mining methods are relevant for learning about economic parameters where they are motivated, for example, by a desire to control properly for confounding variables” (Belloni, Chernozhukov, and Hansen 2014a). They note that research where the choice of variables is based on economic intuition alone, even though there were a large number of variable specifications that were possible, often resulted in the authors “wondering whether the correct variables and functional forms were chosen” (Belloni, Chernozhukov, and Hansen 2014a). With sixty morbidity groups and a very large number of interactions possible, we argue that our use of interaction detection should be deemed to be consistent with the approach prescribed in Belloni, Chernozhukov, and Hansen (2014a).

Nevertheless, it is important to distinguish between data mining as a 'fishing' exercise used to choose a model specification that is solely based on goodness of fit and as a tool to investigate the complex relationship between cost and morbidity. Our use of the CHAID procedure is also motivated by a limited theoretical basis for identifying an exhaustive set of multimorbidity groups that are relevant to cost amplifications for four different categories of cost.

2.3 Evaluating model performance within sub-samples

When assessing model performance between Models 1, 2 and 3, we will focus on the adjusted R-squared statistics to indicate how the models compare at an aggregate level for each percentile of the conditional distribution estimated. In order to evaluate whether including the morbidity interaction terms in the quantile regressions improves the fit of Model 3, we compute the coefficient of determination for the actual cost and estimated cost for the entire sample and selected sub-samples (defined below). An improvement in model performance coincides with a coefficient of determination that is higher when the morbidity interaction variables are included in the calculation of estimated cost. To measure this, we calculate the coefficient of determination for five samples. These are: i) the entire sample with non-zero healthcare costs; ii) those patients with two morbidities (comorbidity subsample); iii) those with three or more morbidities (multimorbidity subsample); iv) those who died in the year of interest (Death in 2010 subsample); and v) those who died during the next calendar year (Death in 2011 subsample). For each sub-sample, denoted by k for which k=1,…,5, we estimate two sets of costs for each of the four categories of healthcare costs using the estimated coefficients obtained from the third set of quantile regressions (i.e. Model 3). One set of cost estimates are computed with the relevant morbidity interaction terms or death dummy variables included in the computation of the estimated cost of healthcare, $\hat{C}$, while the other set of estimates is computed without the morbidity interaction variables or death dummy variables, $\tilde{C}$. The two coefficients of determination, i.e. the squared correlation coefficients, for each set of estimated healthcare costs and the actual costs are specified as

$$\hat{R}^2_{kh} = \left[\frac{\sum_{i=1}^{n}(\hat{C}_{khi} - \overline{\hat{C}}_{kh})(C_{khi} - \overline{C}_{kh})}{\sqrt{\sum_{i=1}^{n}(\hat{C}_{khi} - \overline{\hat{C}}_{kh})^2}\sqrt{\sum_{i=1}^{n}(C_{khi} - \overline{C}_{kh})^2}} \right]^2 \quad (3)$$

$$\tilde{R}^2_{kh} = \left[\frac{\sum_{i=1}^{n}(\tilde{C}_{khi} - \overline{\tilde{C}}_{kh})(C_{khi} - \overline{C}_{kh})}{\sqrt{\sum_{i=1}^{n}(\tilde{C}_{khi} - \overline{\tilde{C}}_{kh})^2}\sqrt{\sum_{i=1}^{n}(C_{khi} - \overline{C}_{kh})^2}} \right]^2 \quad (4)$$

with *k* being the index of the sub-sample of interest and *h* the type of healthcare cost. For each sub-sample that is conditional on having the determining factor, *k*, it is expected that $\hat{R}^2_{kh} > \tilde{R}^2_{kh}$ when the inclusion of the variables provides additional information and enhances the estimation of the healthcare cost for that subsample group. Note that the case of $\hat{R}^2_{kh} < \tilde{R}^2_{kh}$ can occur and coincides with a reduction in the accuracy of predicting the healthcare cost for that subsample group.

3 Data

The data used in this study is part of the Sax Institute's 45 and Up Study. Note that this is the same data source as that used in Ellis et al. (2013). The baseline survey (conducted between 2006 and 2009) is linked to administrative health data that provides rich detail on an individual's utilisation of medical services, pharmaceutical prescriptions and hospitalisations. The 45 and Up Study recruited more than 267,000 people aged 45 and over in New South Wales (NSW), which is located on the east coast of Australia. Prospective participants were randomly sampled from the Department of Human Services (DHS) enrolment data base, which provides near complete coverage of the population. People aged greater than 80 years old and residents of rural and remote areas were oversampled (45 and Up Study Collaborators 2008). About 18% of those invited participated in the survey and these participants accounted for almost 11% of the NSW population aged 45 years and over. The baseline survey has been linked to data from the Medical Benefits Scheme (MBS), the Pharmaceutical Benefits Scheme (PBS), the NSW Admitted Patient Data Collection (APDC) and the Emergency Department Data Collection (EDDC). The MBS dataset covers the services funded by Medicare, Australia's universal publicly funded system for medical services, which includes visits to GPs and specialists, as well as diagnostic tests and procedures.

The 45 and Up Study was approved by the University of New South Wales Human Research Ethics Committee (HREC). Participants' consent was sought before linking the administrative data to the survey data. Linkage of the 45 and Up baseline survey data to the MBS and PBS data provided by the Department of Human Services was completed by the Sax Institute using a unique identifier. Linkage to the administrative data from the APDC and EDDC was conducted by the Centre for Health Record Linkage (CHeReL) with ethics approval from the NSW Population and Health Services Research Ethics Committee (PHSREC).

The linkage of the 45 and Up Study to administrative data allows us to match the annual costs of healthcare to demographic variables and self-reported health status. These annual costs are Medicare funded medical services (Med. Serv.), pharmaceutical (Pharma), hospital (Hosp.) and total healthcare costs (Total). These total healthcare costs are not all inclusive[1] as they are the aggregation of the medical services, pharmaceutical and hospital costs. Medical services costs are the expenditure on costs related to the MBS and include out-of-pocket costs. Pharmaceutical costs are the prices paid to the suppliers of PBS drugs. Both the MBS and PBS data report the costs that are directly used in the analysis. In the case of hospital costs, our administrative data includes information on the Australian Refined Diagnosis Related Groups (AR-DRG) codes for each hospital separation. This information is used to derive a cost per hospital episode based on the National Hospital Cost Data Collection (NHCDC) average cost estimates produced by the Independent Hospital Pricing Authority for public hospital patients (IHPA 2013). The Hospital Casemix Protocol (HCP) average cost estimates are used for private hospital patients (Dept of Health 2015, 2016).

To account for varying complexities in the treatment of individual hospital episodes of care, we have adjusted the average costs associated with each AR-DRG code for each individual's length of stay when it differed from the average length of stay used in the NHCDC and HCP. Equation five shows

[1] The main types of healthcare costs that are excluded include dental services, optometry, allied health, non-subsidised pharmaceuticals and other outpatient hospital services. Note that there are no individualised data on these services that can be linked to the 45 and Up Study.

the computation that we have made for each episode of hospital care to obtain the adjusted cost of hospitalisation ($AdjC_{de}$) for each diagnosis, d (which matches the AR-DRG code), and episode of care, e. This involved identifying the cost for each additional or lesser day of hospital stay in comparison to the average length of stay ($ALOS_{de}$). We did this by identifying the expenses that tend to be upfront costs and essential services that are likely to be incurred once or at the beginning of a hospital admission (EC_{de}). This includes the cost associated with critical care, operation rooms, emergency departments, special procedure suites and prostheses. These essential service costs, EC_{de}, are deducted from the total average cost of hospitalisation for each AR-DRG (C_{de}) and this is divided by the average length of stay used in the cost schedules for public and private hospitalisations to obtain a daily add-on cost estimate. We obtain the adjusted average cost for the hospital patient's episode of care by adding/subtracting the daily add-on cost for each of the days the individual had stayed in the hospital longer/shorter than the average length of stay for that AR-DRG code.

$$AdjC_{de} = C_{de} + (LOS_{de} - ALOS_{de})([C_{de} - EC_{de}]/ALOS_{de}) \quad (5)$$

After establishing a cost associated with hospital admissions, all of the cost and administrative data have been aggregated to create annual variables for the calendar years between 2006 and 2011. In this paper we focus on the data from 2010 as it allows us to include variables that capture whether death is imminent and occurs in the current year or the next year (i.e. 2011). The analysis is conducted on the full cohort of the 45 and Up Study that had non-zero annual costs in the year 2010. As noted in Table 1A, the dummy variables of sixty morbidity groups are coded using the DRG-10 and ATC codes from the APDC and PBS datasets. Table 1A also describes a range of control variables sourced from the 45 and Up survey and these include a 5-point scale of self-reported health status, demographic, economic and lifestyle dummy variables.

4 Results

4.1 Identification of morbidity interactions

This section focuses on the results from the CHAID procedure that has been used to identify the morbidity interactions that are related to healthcare costs. While the results can be represented as a tree structure with a series of branches that are distinguished by zero and non-zero observations of the dummy variables selected by the procedure, we have converted these trees into interaction dummy variables for inclusion in the healthcare cost models. Tables 3A to 6A in the appendix present the morbidity interactions that were identified for four categories of healthcare costs using the CHAID procedure, respectively. As the CHAID procedure identifies interactions based on the existence and non-existence of morbidities, the rows in Tables 3A to 6A present a unique combination of ✓ and ✕ that coincide with whether the interaction group incorporates the existence or non-existence of that specific combination of morbidities. Tables 3A to 6A also includes the number of morbidity groups identified as being relevant to that interaction group and the number of coexisting morbidities, which coincides with the number of times ✓ appears. Note that when assessing the cost estimates associated with these interaction terms in section 4.3, the relevant figures (i.e. Figures 1 to 4) will include the specification of the relevant interaction groups that coincide with Tables 3A to 6A in the appendix.

As the CHAID procedure selected these interactions using an ordered variable of the decile group of cost that each individual was in, an indication of the general costliness of each interaction group is provided in Tables 3A to 6A with the percent of individuals who had that morbidity interaction and were in the top decile of that healthcare cost distribution. The interaction group with the greatest proportion of individuals within the top decile of total healthcare costs (72%) coincided with individuals who had functional disorders of the digestive system, malignant neoplasm and bacterial infection or septicemia but not thrombo-embolic risk or disease (i.e. interaction group t24). The interaction group with the second greatest proportion of individuals within the top decile (69%)

coincides with malignant neoplasm, bacterial infection or septicemia and thrombo-embolic risk or disease (t32). The third greatest proportion of individuals within the top decile (64%) coincides with interaction group t31 and this group is comprised of those individuals who had heart failure, functional disorders of the digestive system, pain, bacterial infection or septicemia and thrombo-embolic risk or disease but not malignant neoplasm.

For pharmaceutical costs (Table 4A), the interaction group with the greatest proportion of individuals within the top decile (73%) coincided with a group of individuals who had diabetes mellitus, hyperlipidemia, reactive airway disease, diseases of esophagus and peptic ulcer as well as thrombo-embolic risk and disease (p48). In the case of medical services costs (Table 5A), the corresponding group (61%) were those who had malignant neoplasm, bacterial infection or septicemia and thrombo-embolic risk or disease (m32). In the case of hospital costs (Table 6A), there was no interaction group that coincided with more than 50% of individuals within the top decile. This implies that high hospital costs are not as well explained as other types of healthcare costs using morbidity interaction groups as explanatory variables. The results reveal that 40% of individuals with heart failure, functional disorders of the digestive system and thrombo-embolic risk or disease had hospitalisation costs that reached the top decile of the hospital cost distribution. Thirty percent of those with functional disorders of the digestive system, malignant neoplasm and bacterial infection or septicemia but not thrombo-embolic risk or disease were in the top decile of hospital costs.

4.2 Model performance with and without morbidity interactions

This section assesses the model performance that occurs with and without the interaction variables. Before focusing on the sub-sample analysis, we will review model performance for the whole sample using the adjusted R-square statistics. Table 1 contains the adjusted R-square statistics for the simultaneous-quantile regressions that were estimated for each type of healthcare cost and model specification. The improvements in model performance are shown using the difference in adjusted R-

square statistics for Model 3 compared to Model 1 and Model 2. Note that assessing improvements in model performance is difficult without a benchmark to assess how large a difference should be for it to be assessed as notable. Note that Buchner, Wasem, and Schillo (2017) deemed an improvement in the overall adjusted R-square of 0.0038 to be a marginal improvement. Upon finding that the inclusion of interaction variables increased the R-square value of 0.2556 by between 0.008 and 0.0178, van Veen et al. (2017) concluded that future studies should include interactions in risk-equalisation models. Based on these studies we use improvements in model performance that are greater than 0.005 as a benchmark. These cases are bolded in Table 1.

When comparing Model 3 to Model 2 the greatest improvements in the explanatory power of the models occurred within the medical services and hospital cost distributions. In the case of medical services, improvements occur across all percentiles. For hospital costs, the improvements coincide with the median, high and extremely high percentiles. Upon comparing Model 3 to Model 2, the estimation of pharmaceutical costs coincided with little improvement across all of the percentiles reviewed, except for the 10th percentile. However, when comparing Model 3 to Model 1 there are large improvements across all costs and percentiles. For example, a 2.6% to 8.9% improvement occurred at the median of the four cost distributions. In contrast to previous findings that morbidity count variables perform well in comparison to multimorbidity measures (Brilleman et al. 2014, Islam et al. 2014), these results show notable improvements in model performance associated with a morbidity dummy and interaction variable model (i.e. Model 3 compared to Model 1). This is especially the case for pharmaceutical costs with improvements in explaining the variations of cost of between 8.3% and 25.5%. Note that the importance of focusing on the entire distribution of costs is reflected in model improvements that are higher for the right tail of all four cost distributions with respect to the median and 10th percentile. For the extremely high points of the total healthcare, medical services and hospital cost distributions (i.e. 95th and 98th percentiles), improvements in explanatory power were between 7.0% and 8.0%. This compares to 1.0% and 2.6% for the lowest point of the cost distributions reviewed (i.e. 10th percentile).

While the adjusted R-square statistics in Table 1 provide an overview of the differences in model performance at an aggregate level and indicates the conditional percentile estimates where overall model performance improved; it is important to assess model performance using appropriate sub-samples to highlight the cases where there were notable improvements that would otherwise be hidden by the aggregation of many different individuals. Table 2 contains the differences in the coefficient of determination for alternative model specifications across selected subsamples for each category of healthcare cost. For the 10^{th} and 50^{th} percentiles of total healthcare costs, multimorbidity interaction variables were associated with the greatest improvement in the estimation of this healthcare cost. For the multimorbidity subsample the improvement was 2.8% for the 10^{th} percentile and 2.0% for the 50^{th} percentile. Notable improvements also coincide with the overall sample (2.2% for the 10^{th} percentile and 1.6% for the 50^{th} percentile) and the imminent death subsamples (between 0.8% and 2%).

For the estimates of extremely high healthcare costs (95^{th} and 98^{th} percentiles), the death dummy variables are associated with a decrease in model performance in relation to predicting total healthcare and hospital costs. So, while the parameter estimates associated with these death dummy variables are large and statistically significant (refer to Table 3 and Tables 7A to 10A for the cost estimates of these variables), the amount of variance in the costs associated with imminent death is also large and a dummy variable is insufficient to capture the wide dispersion of total healthcare and hospital costs that coincide with death. Note that we have not used the time to death variable specified in Felder, Werblow, and Zweifel (2010) and future work will control for the complexities associated with imminent death. Nevertheless, the death dummy variables do coincide with cost improvements for the medical services cost estimates within the imminent death subsamples. For the death in 2010 variable the improvement ranges from 2.9% to 5.7% depending upon the percentile estimate. In the case of death in the next year, the improvement ranges from 2.9% to 4.7%. In some cases, the estimates are negative (i.e. for pharmaceuticals and medical services) and reflects the nature of pharmaceutical and medical service costs being accumulative so that whether individuals died in the early or latter part of

the year has a strong influence on how high costs become. In contrast, hospital costs can be amplified in a relatively short amount of time and the death dummy variables are positive for all percentiles of total healthcare and hospital costs.

The second segment of Table 2 contains the impact of including morbidity interaction variables on the estimates of pharmaceutical costs. The lack of a notable improvement is in line with the overall sample adjusted R-squared statistics shown in Table 1 for the comparison between Model 2 and Model 3. As noted in the introduction, prescribing practices have been found to be inapt at tackling the rise of multimorbidity. Limited improvements in estimating pharmaceutical costs through the inclusion of morbidity interaction variables is consistent with prescribing guidelines and practices that focus on individual diseases. In many cases, especially for the 10^{th} percentile, there was a negative difference in the coefficients of determination and the model did not improve with the inclusion of morbidity interaction variables.

The third segment of Table 2 reports the impact of including morbidity interaction variables on the estimates of medical services costs. Most of the improvements associated with multimorbidity variables are concentrated in the 10^{th} and 50^{th} percentiles. While the inclusion of death variables decreases model performance for the overall sample at the 90^{th}, 95^{th} and 98^{th} percentiles, the inclusion of death variables improves the model estimates for the imminent death subsamples. As out-of-pocket costs are included in the medical services costs and doctors in Australia are free to set their own fees, medical services costs also reflect the willingness to pay for services by individual patients and exhibit considerable variation. In particular, wealthier patients typically pay higher prices for similar services than poorer patients do (Johar et al. 2016). For this reason, the morbidity interaction variables may not be capturing important drivers of high medical service costs. Rather, higher costs may be associated with the higher fees charged by esteemed specialists, or the greater willingness to pay for health care services by wealthier sections of the community. Improvements for the imminent death

subsamples suggest that in the last year of life there is an increased intensity of treatment by specialists dealing with health issues that have a notable risk of death. However, negative estimates for some percentile estimates show that timing matters and that there is great variability in medical services costs in the last two years of life. Another reason for the importance of imminent death is likely to be the intensity of the use of services related to at-home care for the terminally ill and the provision of palliative care outside of hospital in the final weeks of life.

The last segment of Table 2 contains the impact of including morbidity interaction variables on the estimates of hospital costs and in this case the improvements associated with multimorbidity variables occur for the 10th and 50th percentiles. The inclusion of death variables reduces the performance of model estimates across all of the five samples and this reduction becomes larger at the right tail of the hospital cost distribution. This is likely to be related to the large variation in the costliness of the last year of life, which will depend on the cause of death and/or the types of treatments and services used within hospital or as part of at-home care.

4.3 Cost estimates associated with the morbidity interaction and death variables

This section focuses on the cost estimates associated with the interaction groups for the 50th, 90th, 95th and 98th percentiles of each category of healthcare cost. These cost estimates are separated into the components attributed to the regression's intercept, the sum of costs associated with each of the relevant individual morbidity groups and the cost amplification that coincides with the morbidity interaction variables[2]. Note that Tables 7A to 10A in the appendix provide the parameter estimates used to calculate the costs estimates presented in Figures 1 to 4. Figure 1 contains the cost estimates

[2] In some cases, the cost estimate associated with the interaction variable is negative and for these cases we have adjusted the cost associated with the intercept (or individual morbidity groups) so that the correct aggregate cost is shown on the y axis. The interpretation of interaction variables with a negative cost estimate should be conducted in relation to the other components of cost (specifically the intercept). This then captures cases where the interaction does not tend to coincide with extremely high costs in comparison to other morbidity classifications. This is evident as the incidence of negative estimates tends to occur for the 90th, 95th and 98th percentile estimates but not the 50th percentile estimates.

for total healthcare costs across four percentiles (i.e. the 50th, 90th, 95th and 98th percentiles). Notable differences in the cost amplification associated with the morbidity interaction groups[3] occur for each percentile reviewed. In the majority of cases, we observe that the largest cost amplifications tended to coincide with multimorbidity groups rather than comorbidity groups. For the median of total healthcare costs, these were interaction groups t24, t31 and t32. Note that for t24, the cost amplification was 60% of the total sum of the estimated costs and 1.6 times larger than the individual morbidity group cost component. The equivalent numbers for t31 and t32 are 59% (1.6 times larger) and 56% (1.4 times larger), respectively. While the costs associated with the morbidity interactions tend to increase for the higher percentiles, their magnitude in comparison to the sum of the costs associated with the individual morbidity estimates tends to diminish. Nevertheless, the largest interaction cost estimate for total healthcare costs was over $19,000 and this coincided with the 90th percentile estimate of multimorbidity group t31. However, this interaction cost estimate was smaller than the individual morbidity component of the estimated costs. Multimorbidity group t31 are those individuals who had heart failure, functional disorders of the digestive system, pain, bacterial infection or septicemia and thrombo-embolic risk or disease but not malignant neoplasm. Note that the aforementioned cases of cost amplification (i.e. t24, t31 and t32) correspond to the three interaction groups with the highest proportion of individuals within the top decile of total healthcare cost.

In contrast to the cases of total healthcare and hospital costs, the cost amplification that occurs for pharmaceutical and medical services costs are relatively moderate. Figure 2 and 3 contain the cost estimates for pharmaceuticals and medical services costs. For the 50th percentile, the largest amplification of cost in terms of the relative size of the morbidity interaction estimate is 62% for pharmaceutical costs and 54% for medical services costs. These cases are attributed to multimorbidity groups p46 (with a cost amplification of $801) and m32 (with a cost amplification of $1,344). The

[3] For the morbidity groups that are included and excluded in these morbidity interaction groups we refer the reader to the index listed at the bottom of Figure 1. This corresponds with the classification outlined in Table 3A.

largest interaction cost estimate for pharmaceutical costs occurs at the 98th percentile with the interaction estimate for p43. This corresponds to an increase in pharmaceutical costs of $1608. For medical services costs the largest interaction cost estimate occurs at the 98th percentile with m29 having a cost estimate associated with the interaction variable of $5701.

For hospital costs, the estimates for the 50th percentile show a large cost amplification for multimorbidity group h9 and h18. These cost amplifications are equivalent to $4325 and $5104, respectively. In the case of multimorbidity group h9, this cost amplification is associated with functional disorders of the digestive system, malignant neoplasm and bacterial infection or septicemia but not thrombo-embolic risk or disease. In the case of multimorbidity group h18, this cost amplification is associated with heart failure, functional disorders of the digestive system and thrombo-embolic risk or disease. Within this category of cost, there are a range of cases where the interaction cost estimate is negative, however these need to be considered in relation to the large hospital cost estimates associated with the individual morbidity groups at the extremely high cost points of the distribution.

Table 3 contains the cost estimates associated with the age, BMI and death dummy variables included in the third model. Note that these cost estimates for age, BMI and imminent death have been computed with morbidity dummy and interaction variables included in the model (i.e. M3). The largest death related cost estimates are concentrated in the hospital cost and total healthcare cost models. For pharmaceuticals and medical services, the death related cost estimates are negative or less than a few thousand dollars. From this, it seems that hospital costs are those most amplified with respect to time to death. However, it is of interest to note that the cost estimates related to age are still quite large for the high (i.e. the 90th percentile) and extremely high (i.e. the 95th and 98th percentiles) points of the hospital cost distribution. The estimates related to the death dummy variables are large, however, they were not associated with an improvement in predicting hospital costs for any of the

subsamples. Note that this may be due to correlation with morbidity variables and/or the interaction variables. While we adjusted the DRG costs using an individual's length of stay and this data is the best available at the time of writing, it does not include the heterogeneity that hospital-specific data for the corresponding episodes is likely to include. In future research, we hope to gain access to Independent Hospital Pricing Authority (IHPA) hospital data that captures this heterogeneity and will reveal how large hospital costs do become.

Table 3 also includes the estimates that correspond to BMI dummy variables. Notable differences in cost estimates across the healthcare category and the percentiles used. For total healthcare costs, the largest estimates are associated with those who are obese and the 90th, 95th and 98th percentile estimates. However, for hospital costs the largest cost estimates coincide with underweight individuals and the 95th and 98th percentiles. For those admitted to hospital, being underweight is likely to be related to the type of morbidity they have or the treatment they are undergoing. While the U-shaped relationship between BMI and expenditure found by Buchmueller and Johar (2015) using similar data linked to the 45 and Up study for the period between 2006 to 2009 does occur at least once in the cost estimates of each healthcare category, it does not always appear across all percentiles. Pharmaceutical costs are the category of healthcare where a U-shaped relationship is least notable. This is consistent with the results of Buchmueller and Johar (2015) when they focus on different types of healthcare costs.

5 Conclusion

Within this paper, we have assessed whether morbidity interaction variables improve model performance in comparison to models that only have morbidity dummy variables or morbidity count variables using simultaneous quantile regressions. This was motivated by previous research that found that morbidity count variables outperformed more complicated morbidity measures (Brilleman et al. 2014, Islam et al. 2014). In contrast to these studies, we find that models that include morbidity dummy and interaction variables notably outperform a basic model with a count variable of the

number of coexisting morbidities. This is especially the case for the right tail of the total, medical services and hospital cost distributions (i.e. 90th, 95th and 98th percentiles) and all five points of the pharmaceutical cost distribution. Upon focusing on the conditional estimates of the 10th, 50th, 90th, 95th and 98th percentiles we find that notable cost amplifications coincide with multimorbidity rather than comorbidity.

Our results suggest that further focus on multiple points of healthcare cost distributions will assist in making accurate adjustments for risk that are relevant to the development of payments systems. One of the important findings of this research is that thrombo-embolic risk or disease, malignant neoplasm, hyperlipidemia, diseases of esophagus and peptic ulcer, pain and bacterial infection or septicemia are the morbidities to focus on when investigating multimorbidities as important determinants of costly healthcare. Many of these morbidities are associated with the treatment of acute diseases and, for malignant neoplasm, the management of adverse treatment effects (i.e. bacterial infection after chemotherapy). This implies that a substantial number of patients with very high healthcare costs are suffering from unintended effects of treatment and that the health system's ability to reduce such issues will have a substantial impact on costs.

While the availability of large data sets is becoming more common, studies using this type of data need to adapt and adopt techniques suited to the complexities of 'big data'. Future research should continue to explore the full distribution of healthcare costs and go beyond a focus on estimating the mean (de Meijer et al. 2013, Jones, Lomas, and Rice 2015). As the use of large administrative data is likely to result in data of a high-dimensional nature, automated data driven techniques are expected to be increasingly common in empirical economics. Belloni et al. (2012), Belloni, Chernozhukov, and Hansen (2014a), Belloni, Chernozhukov, and Hansen (2014b), Einav and Levin (2014), Varian (2014), Athey (2017), Buchner, Wasem, and Schillo (2017) and van Veen et al. (2017) are notable

examples that have recently employed or have advocated for the use of automated data driven techniques within the analysis of large data sets.

An important extension of this research will be confirming that similar results hold using other samples and morbidity classifications. In addition, it will be important to establish that similar interactions are found using CHAID and other interaction detection techniques. Epidemiological studies will also be needed to identify the morbidity specific determinants of costly healthcare that are behind the importance of these interaction effects. Our intention upon documenting the full range of interaction groups in Tables 3A to 6A within the appendix was, in part, to facilitate further research that investigates the underlying determinants driving the cost amplifications associated with these interaction groups in different settings.

While we include death dummy variables in our models of healthcare costs, we do not precisely account for an individual's time to death. Nevertheless, our results for medical services and hospital costs indicate that there is more work to be done on disentangling why some people have costly deaths and others do not. Previous research has found that time to death was an important driver of healthcare costs, however these results indicate that a simple variable capturing whether someone dies in the current year or the next does not improve the performance of hospital cost models. It is plausible that differences in the incidence and intensity of at-home care and hospital provided palliative care will be part of the variation in costs during the final year of life.

6 References

45 and Up Study Collaborators. 2008. "Cohort profile: the 45 and up study." *International Journal of Epidemiology* no. 37 (5):941.

45 and Up Study Collaborators. 2011. "The 45 and Up Study Baseline Questionnaire Data Book."

Athey, Susan. 2017. "Beyond prediction: Using big data for policy problems." *Science* no. 355 (6324):483-485.

Banerjee, Sube. 2015. "Multimorbidity—older adults need health care that can count past one." *The Lancet* no. 385 (9968):587-589.

Barnett, Karen, Stewart W Mercer, Michael Norbury, Graham Watt, Sally Wyke, and Bruce Guthrie. 2012. "Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study." *The Lancet* no. 380 (9836):37-43.

Belloni, A., D. Chen, V. Chernozhukov, and C. Hansen. 2012. "Sparse Models and Methods for Optimal Instruments With an Application to Eminent Domain." *Econometrica* no. 80 (6):2369-2429.

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen. 2014a. "High-dimensional methods and inference on structural and treatment effects." *The Journal of Economic Perspectives* no. 28 (2):29-50.

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen. 2014b. "Inference on Treatment Effects after Selection among High-Dimensional Controls†." *Review of Economic Studies* no. 81 (2):608-650.

Boyd, Cynthia M, Jonathan Darer, Chad Boult, Linda P Fried, Lisa Boult, and Albert W Wu. 2005. "Clinical practice guidelines and quality of care for older patients with multiple comorbid diseases: implications for pay for performance." *Jama* no. 294 (6):716-724.

Brilleman, Samuel L, Hugh Gravelle, Sandra Hollinghurst, Sarah Purdy, Chris Salisbury, and Frank Windmeijer. 2014. "Keep it simple? Predicting primary health care costs with clinical morbidity measures." *Journal of health economics* no. 35:109-122.

Buchmueller, Thomas C, and Meliyanni Johar. 2015. "Obesity and health expenditures: evidence from Australia." *Economics & Human Biology* no. 17:42-58.

Buchner, Florian, Jürgen Wasem, and Sonja Schillo. 2017. "Regression Trees Identify Relevant Interactions: Can This Improve the Predictive Performance of Risk Adjustment?" *Health economics* no. 26.

Burgers, Jako S., Gerlienke E. Voerman, Richard Grol, Marjan J. Faber, and Eric C. Schneider. 2010. "Quality and Coordination of Care for Patients With Multiple Conditions: Results From an International Survey of Patient Experience." *Evaluation & the Health Professions* no. 33 (3):343-364. doi: doi:10.1177/0163278710375695.

de Meijer, Claudine, Owen O'Donnell, Marc Koopmanschap, and Eddy van Doorslaer. 2013. "Health expenditure growth: Looking beyond the average through decomposition of the full distribution." *Journal of Health Economics* no. 32 (1):88-105. doi: https://doi.org/10.1016/j.jhealeco.2012.10.009.

Dept of Health. 2015. "Hospital Casemix Protocol: Annual Report 2012-13."

Dept of Health. 2016. "Hospital Casemix Protocol: Annual Report 2013-14."

Diederichs, Claudia, Klaus Berger, and Dorothee B Bartels. 2011. "The measurement of multiple chronic diseases—a systematic review on existing multimorbidity indices." *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences* no. 66 (3):301-311.

Eckardt, Matthias, Christian Brettschneider, Hendrik van den Bussche, Hans-Helmut König, and Group MultiCare Study. 2016. "Analysis of Health Care Costs in Elderly Patients with Multiple Chronic Conditions Using a Finite Mixture of Generalized Linear Models." *Health Economics*. doi: 10.1002/hec.3334.

Einav, Liran, and Jonathan Levin. 2014. "Economics in the age of big data." *Science* no. 346 (6210):1243089.

Ellis, Randall P, Denzil G Fiebig, Meliyanni Johar, Glenn Jones, and Elizabeth Savage. 2013. "Explaining Health Care Expenditure Variation: Large-Sample Evidence Using Linked Survey And Health Administrative Data." *Health economics* no. 22 (9):1093-1110.

Felder, Stefan, Andreas Werblow, and Peter Zweifel. 2010. "Do red herrings swim in circles? Controlling for the endogeneity of time to death." *Journal of health economics* no. 29 (2):205-212.

Guthrie, Bruce, John W Saultz, George K Freeman, and Jeannie L Haggerty. 2008. "Continuity of care matters." *Bmj* no. 337:a867.

Haggerty, Jeannie L. 2012. "Ordering the chaos for patients with multimorbidity." *BMJ* no. 345:e5915.

Halfon, Patricia, Yves Eggli, Anne Decollogny, and Erol Seker. 2013. "Disease identification based on ambulatory drugs dispensation and in-hospital ICD-10 diagnoses: a comparison." *BMC health services research* no. 13 (1):453.

Harrison, Christopher, Helena Britt, Graeme Miller, and Joan Henderson. 2014. "Examining different measures of multimorbidity, using a large prospective cross-sectional study in Australian general practice." *BMJ open* no. 4 (7):e004694.

IHPA. 2013. "National Hospital Cost Data Collection Australian Public Hospitals Cost Report 2010-2011, Round 15."

Islam, M Mofizul, Laurann Yen, Jose M Valderas, and Ian S McRae. 2014. "Out-of-pocket expenditure by Australian seniors with chronic disease: the effect of specific diseases and morbidity clusters." *BMC public health* no. 14 (1):1.

Johar, Meliyanni, Chunzhou Mu, Kees Van Gool, and Chun Yee Wong. 2016. "Bleeding hearts, profiteers, or both: specialist physician fees in an unregulated market." *Health economics*.

Jones, Andrew M, James Lomas, and Nigel Rice. 2015. "Healthcare cost regressions: going beyond the mean to estimate the full distribution." *Health economics* no. 24 (9):1192-1212.

Jones, Andrew Michael, James Lomas, Peter Moore, and Nigel Rice. 2016. "A quasi-Monte Carlo comparison of developments in parametric and semi-parametric regression methods for heavy-tailed and non-normal data: with an application to healthcare costs." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* no. 179 (4):951-974.

Kass, Gordon V. 1980. "An exploratory technique for investigating large quantities of categorical data." *Applied statistics*:119-127.

König, Hans-Helmut, Hanna Leicht, Horst Bickel, Angela Fuchs, Jochen Gensichen, Wolfgang Maier, Karola Mergenthal, Steffi Riedel-Heller, Ingmar Schäfer, and Gerhard Schön. 2013. "Effects of multiple chronic conditions on health care costs: an analysis based on an advanced tree-based regression model." *BMC health services research* no. 13 (1):1.

Lapi, Francesco, Elisa Bianchini, Iacopo Cricelli, Gianluca Trifirò, Giampiero Mazzaglia, and Claudio Cricelli. 2015. "Development and Validation of a Score for Adjusting Health Care Costs in General Practice." *Value in Health* no. 18 (6):884-895. doi: https://doi.org/10.1016/j.jval.2015.05.004.

Le Reste, Jean Yves, Patrice Nabbe, Benedicte Manceau, Charilaos Lygidakis, Christa Doerr, Heidrun Lingner, Slawomir Czachowski, Miguel Munoz, Stella Argyriadou, and Ana Claveria. 2013. "The European General Practice Research Network presents a comprehensive definition of multimorbidity in family medicine and long term care, following a systematic review of relevant literature." *Journal of the American Medical Directors Association* no. 14 (5):319-325.

Lehnert, Thomas, Dirk Heider, Hanna Leicht, Sven Heinrich, Sandro Corrieri, Melanie Luppa, Steffi Riedel-Heller, and Hans-Helmut König. 2011. "Review: health care utilization and costs of elderly persons with multiple chronic conditions." *Medical Care Research and Review* no. 68 (4):387-420.

Loh, Wei-Yin. 2011. "Classification and regression trees." *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* no. 1 (1):14-23.

Luchman, J. N. 2013. "CHAID: Stata module to conduct chi-square automated interaction detection."

Salisbury, Chris. 2012. "Multimorbidity: redesigning health care for people who use it." *The Lancet* no. 380 (9836):7-9. doi: 10.1016/S0140-6736(12)60482-6.

Sambamoorthi, Usha, Xi Tan, and Arijita Deb. 2015. "Multiple chronic conditions and healthcare costs among adults." *Expert Review of Pharmacoeconomics & Outcomes Research* no. 15 (5):823-832. doi: 10.1586/14737167.2015.1091730.

Schoenberg, Nancy E, Hyungsoo Kim, William Edwards, and Steven T Fleming. 2007. "Burden of common multiple-morbidity constellations on out-of-pocket medical expenditures among older adults." *The Gerontologist* no. 47 (4):423-437.

Smith, Susan M, Hassan Soubhi, Martin Fortin, Catherine Hudon, and Tom O'Dowd. 2012. "Managing patients with multimorbidity: systematic review of interventions in primary care and community settings."

Starfield, Barbara, Leiyu Shi, and James Macinko. 2005. "Contribution of primary care to health systems and health." *Milbank quarterly* no. 83 (3):457-502.

Tinetti, Mary E, Terri R Fried, and Cynthia M Boyd. 2012. "Designing health care for the most common chronic condition—multimorbidity." *Jama* no. 307 (23):2493-2494.

van den Akker, Marjan, Frank Buntinx, and J André Knottnerus. 1996. "Comorbidity or multimorbidity: what's in a name? A review of literature." *The European Journal of General Practice* no. 2 (2):65-70.

van den Akker, Marjan, Frank Buntinx, Sjef Roos, and J André Knottnerus. 2001. "Problems in determining occurrence rates of multimorbidity." *Journal of clinical epidemiology* no. 54 (7):675-679.

van Diepen, Merel, and Philip Hans Franses. 2006. "Evaluating chi-squared automatic interaction detection." *Information Systems* no. 31 (8):814-831. doi: https://doi.org/10.1016/j.is.2005.03.002.

Van Spall, Harriette GC, Andrew Toren, Alex Kiss, and Robert A Fowler. 2007. "Eligibility criteria of randomized controlled trials published in high-impact general medical journals: a systematic sampling review." *Jama* no. 297 (11):1233-1240.

van Veen, S. H. C. M., R. C. van Kleef, W. P. M. M. van de Ven, and R. C. J. A. van Vliet. 2017. "Exploring the predictive power of interaction terms in a sophisticated risk equalization model using regression trees." *Health Economics*:n/a-n/a. doi: 10.1002/hec.3523.

van Weel, Chris, and François G Schellevis. 2006. "Comorbidity and guidelines: conflicting interests." *The Lancet* no. 367 (9510):550-551.

Varian, Hal R. 2014. "Big Data: New Tricks for Econometrics." *The Journal of Economic Perspectives* no. 28 (2):3-27.

Werblow, Andreas, Stefan Felder, and Peter Zweifel. 2007. "Population ageing and health care expenditure: a school of 'red herrings'?" *Health economics* no. 16 (10):1109-1126.

Wolff, Jennifer L, Barbara Starfield, and Gerard Anderson. 2002. "Prevalence, expenditures, and complications of multiple chronic conditions in the elderly." *Archives of internal medicine* no. 162 (20):2269-2276.

Zweifel, Peter, Stefan Felder, and Markus Meiers. 1999. "Ageing of population and health care expenditure: a red herring?" *Health economics* no. 8 (6):485-496.

Table 1 – Explanatory power of alternate models – Adjusted R-square

Model		Specification name	Percentile				
			10	50	90	95	98
Total healthcare cost	Morbidity dummy variables and interactions	M3	0.154	0.238	0.379	0.420	0.441
	Difference with respect to M1		**0.026**	**0.039**	**0.076**	**0.078**	**0.078**
	Difference with respect to M2		*0.003*	**0.005**	*0.005*	*0.002*	*0.002*
	Morbidity dummy variables without interactions	M2	0.151	0.233	0.374	0.418	0.439
	Morbidity count variable	M1	0.128	0.199	0.303	0.342	0.363
Pharmaceutical cost	Morbidity dummy variables and interactions	M3	0.213	0.320	0.345	0.377	0.433
	Difference with respect to M1		**0.083**	**0.089**	**0.131**	**0.185**	**0.255**
	Difference with respect to M2		**0.009**	*0.004*	*0.002*	*0.002*	*0.001*
	Morbidity dummy variables without interactions	M2	0.204	0.316	0.343	0.375	0.432
	Morbidity count variable	M1	0.130	0.231	0.214	0.192	0.178
Medical services cost	Morbidity dummy variables and interactions	M3	0.096	0.184	0.280	0.311	0.347
	Difference with respect to M1		**0.012**	**0.026**	**0.057**	**0.070**	**0.080**
	Difference with respect to M2		**0.007**	**0.017**	**0.036**	**0.041**	**0.044**
	Morbidity dummy variables without interactions	M2	0.089	0.167	0.244	0.270	0.303
	Morbidity count variable	M1	0.084	0.158	0.223	0.241	0.267
Hospital cost	Morbidity dummy variables and interactions	M3	0.022	0.105	0.247	0.265	0.281
	Difference with respect to M1		**0.009**	**0.029**	**0.071**	**0.071**	**0.070**
	Difference with respect to M2		*0.002*	**0.008**	**0.017**	**0.028**	**0.040**
	Morbidity dummy variables without interactions	M2	0.020	0.097	0.230	0.237	0.241
	Morbidity count variable	M1	0.013	0.076	0.176	0.194	0.211

Note: we highlight the larger improvements in model performance by bolding the differences in the adjusted R-square statistics that are greater than 0.005.

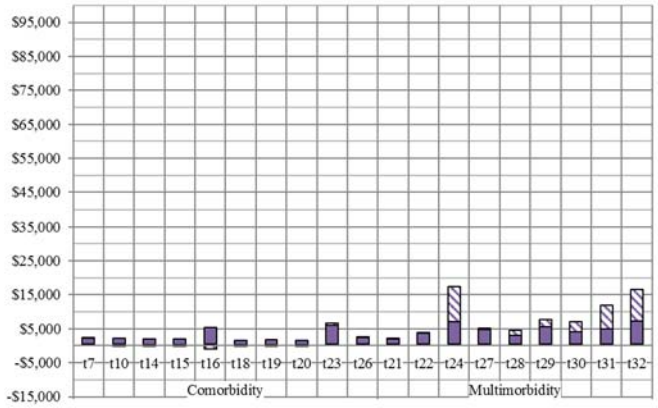
Table 2 – Improvement in the coefficient of determination

Model	Sample	Percentile														
		10			50			90			95			98		
		Comorbidity variables	Multimorbidity variables	Death variables	Comorbidity variables	Multimorbidity variables	Death variables	Comorbidity variables	Multimorbidity variables	Death variables	Comorbidity variables	Multimorbidity variables	Death variables	Comorbidity variables	Multimorbidity variables	Death variables
Total	Overall sample (n = 267,086)	0.000	**0.022**	0.001	0.000	**0.016**	0.000	0.001	0.002	-0.002	-0.000	0.001	**-0.007**	-0.001	-0.000	**-0.015**
	Comorbidity subsample (n = 29,753)	0.003	-	0.001	0.001	-	0.001	0.003	-	0.002	0.003	-	-0.002	0.003	-	**-0.011**
	Multimorbidity subsample (n = 106,032)	0.000	**0.028**	0.002	-0.000	**0.020**	0.003	0.000	0.001	0.001	-0.001	0.001	**-0.005**	-0.002	-0.000	**-0.014**
	Death in 2010 subsample (n = 3,000)	0.002	**0.020**	0.000	0.002	**0.017**	-0.000	0.003	-0.001	-0.004	0.001	-0.002	**-0.011**	0.000	-0.001	**-0.018**
	Death in 2011 subsample (n = 3,399)	-0.000	**0.018**	0.002	-0.000	**0.008**	0.004	0.000	**-0.009**	0.003	-0.001	**-0.005**	-0.005	-0.001	-0.001	**-0.014**
Pharm.	Overall sample (n = 182,484)	-0.003	**-0.010**	-0.004	-0.001	-0.000	-0.002	0.000	0.005	0.002	0.000	0.005	**0.006**	0.000	0.004	**0.006**
	Comorbidity subsample (n = 29,424)	0.000	-	-0.005	0.001	-	-0.004	0.002	-	-0.002	0.001	-	0.001	0.001	-	0.001
	Multimorbidity subsample (n = 105,957)	-0.004	**-0.013**	-0.003	-0.001	-0.002	-0.002	-0.000	0.004	0.001	-0.000	0.003	0.004	-0.000	0.002	0.004
	Death in 2010 subsample (n = 2,622)	-0.003	**-0.011**	**-0.007**	-0.001	-0.001	**-0.008**	-0.000	0.002	-0.005	0.000	0.002	-0.002	0.000	0.001	-0.002
	Death in 2011 subsample (n = 3,022)	-0.002	**-0.007**	-0.003	-0.001	-0.003	-0.004	-0.000	-0.001	**-0.005**	-0.000	-0.000	-0.003	-0.000	-0.000	-0.003
Med. Serv.	Overall sample (n = 253,765)	0.001	**0.012**	**0.007**	0.001	**0.006**	**-0.008**	-0.001	0.003	**-0.042**	-0.001	0.002	**-0.050**	-0.001	0.001	**-0.067**
	Comorbidity subsample (n = 29,552)	0.004	-	**0.011**	0.002	-	**-0.013**	-0.000	-	**-0.022**	0.001	-	**-0.023**	0.001	-	**-0.028**
	Multimorbidity subsample (n = 105,866)	0.002	**0.017**	**0.017**	0.001	**0.009**	**0.014**	-0.001	0.004	0.002	-0.001	0.003	-0.001	-0.002	0.000	**-0.009**
	Death in 2010 subsample (n = 2,783)	0.001	**0.018**	**0.029**	0.001	**0.009**	**0.042**	0.001	**0.010**	**0.054**	0.001	**0.007**	**0.055**	0.001	**0.006**	**0.057**
	Death in 2011 subsample (n = 3,284)	0.000	**0.013**	**0.029**	-0.000	0.004	**0.040**	-0.001	**-0.010**	**0.046**	-0.001	**-0.015**	**0.047**	-0.001	**-0.020**	**0.046**
Hosp.	Overall sample (n = 72,287)	0.004	**0.011**	0.005	-0.002	**0.006**	-0.003	**0.005**	0.003	**-0.006**	0.003	0.002	**-0.016**	0.001	0.001	**-0.023**
	Comorbidity subsample (n = 8,234)	0.000	-	**0.008**	0.002	-	0.002	**0.007**	-	**-0.006**	0.003	-	**-0.025**	0.002	-	**-0.032**
	Multimorbidity subsample (n = 45,352)	0.004	**0.012**	0.003	-0.002	**0.008**	-0.005	**0.006**	0.004	**-0.009**	0.003	0.002	**-0.018**	0.001	0.001	**-0.024**
	Death in 2010 subsample (n = 2,516)	0.002	**0.006**	-0.001	-0.001	0.003	-0.002	-0.001	0.002	**-0.006**	-0.001	0.002	**-0.010**	0.000	0.001	**-0.009**
	Death in 2011 subsample (n = 2,170)	-0.000	0.004	0.002	-0.004	**-0.005**	0.004	0.001	0.000	-0.000	-0.001	0.000	-0.004	-0.000	0.001	**-0.008**

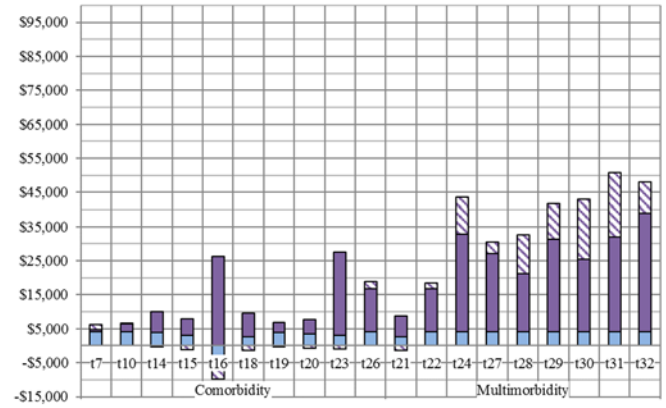
Note: we highlight the larger improvements in model performance by bolding the differences that are greater than 0.005. Large decreases in model performance are highlighted when the differences are less than -0.005.

Figure 1 – Annual total healthcare costs associated with morbidities and morbidity interactions

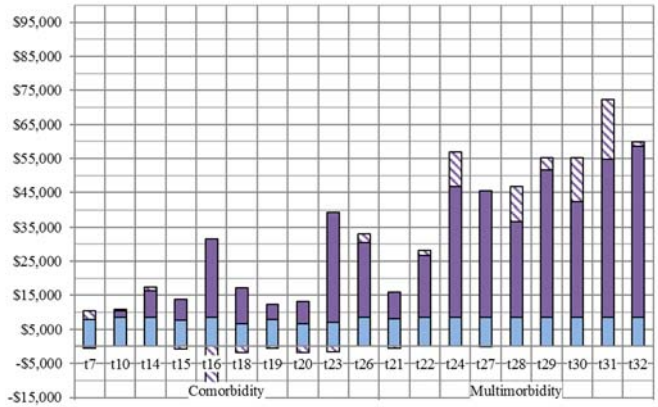
a) 50th percentile



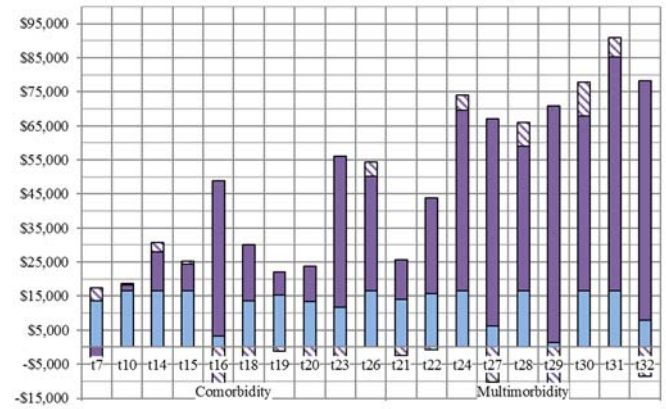
b) 90th percentile



c) 95th percentile



d) 98th percentile



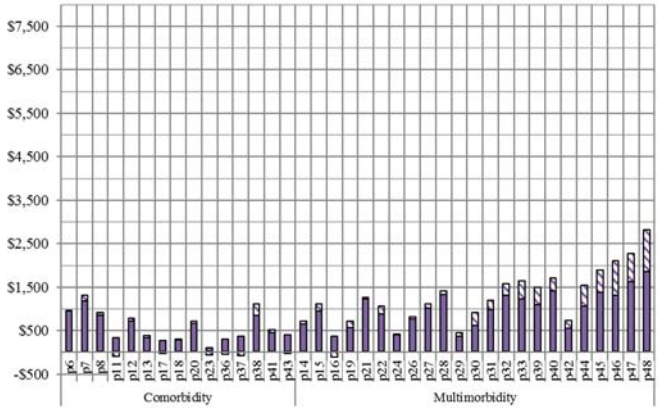
Intercept | Individual morbidities | Morbidity interaction

Morbidity and morbidity interaction groups index

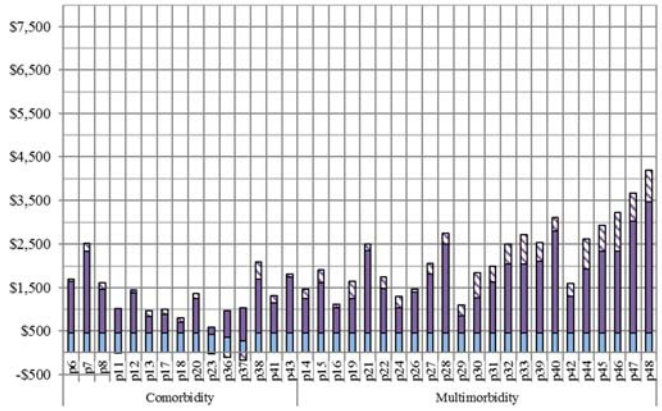
Morbidity groups	Interaction groups
Diabetes (3) Heart failure (13) Hyperlipidemia (29) Functional disorders of the digestive system (43) Ischemic heart diseases (51) Diseases of esophagus and peptic ulcer (53) Malignant neoplasm (54) Pain (56) Bacterial infection or septicemia (58) Thrombo-embolic risk or disease (59)	Comorbidity interaction groups: t7 (3=1, 29=1, 36=0, 53=0, 54=0, 55=0, 56=0, 58=0, 59=0) t10 (29=1, 53=1, 54=0, 55=0, 56=0, 58=0, 59=0) t14 (29=0, 53=1, 54=0, 56=1, 58=0, 59=0) t15 (29=1, 54=0, 56=1, 58=0, 59=0) t16 (54=1, 56=1, 58=0, 59=0) t18 (29=0, 53=0, 54=0, 56=1, 58=1, 59=0) t19 (29=1, 53=0, 54=0, 58=1, 59=0) t20 (53=1, 54=0, 56=0, 58=1, 59=0) t23 (43=0, 54=1, 58=1, 59=0) t26 (51=0, 54=0, 56=0, 58=1, 59=1) Multimorbidity interaction groups: t21 (43=0, 53=1, 54=0, 56=1, 58=1, 59=0) t22 (43=1, 53=1, 54=0, 56=1, 58=1, 59=0) t24 (43=1, 54=1, 58=1, 59=0) t27 (51=1, 54=0, 56=0, 58=1, 59=1) t28 (43=0, 51=0, 54=0, 56=1, 58=1, 59=1) t29 (43=0, 51=1, 54=0, 56=1, 58=1, 59=1) t30 (13=0, 43=1, 54=0, 56=1, 58=1, 59=1) t31 (13=1, 43=1, 54=0, 56=1, 58=1, 59=1) t32 (54=1, 58=1, 59=1)

Figure 2 – Annual pharmaceutical costs associated with morbidities and morbidity interactions

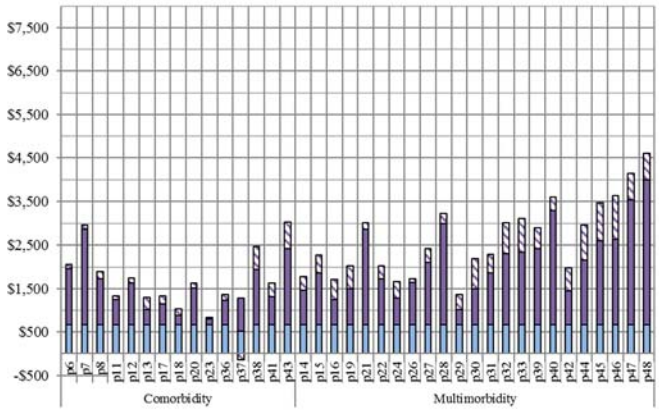
a) 50th percentile



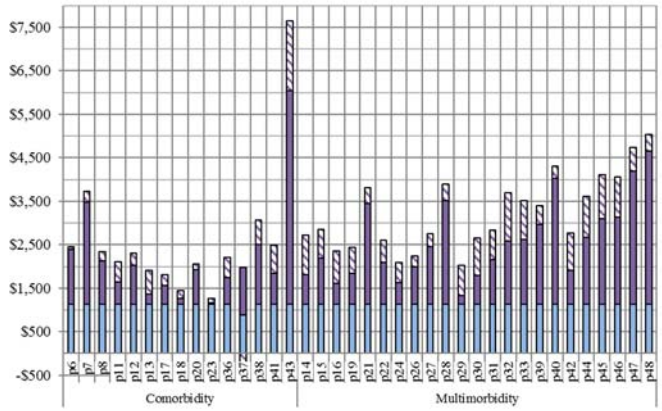
b) 90th percentile



c) 95th percentile



d) 98th percentile



Intercept | Individual morbidities | Morbidity interaction

Morbidity and morbidity interaction groups index

Morbidity groups

Diabetes mellitus (3)
Mood [affective] disorders (6)
Hypertensive disease (11)
Hyperlipidemia (29)
Reactive airway disease (52)
Diseases of esophagus & peptic ulcer (53)
Inflam. polyarthritis & connective tissue disord. (55)
Pain (56)
Bacterial infection or septicemia (58)
Thrombo-embolic risk or disease (59)

Interaction groups

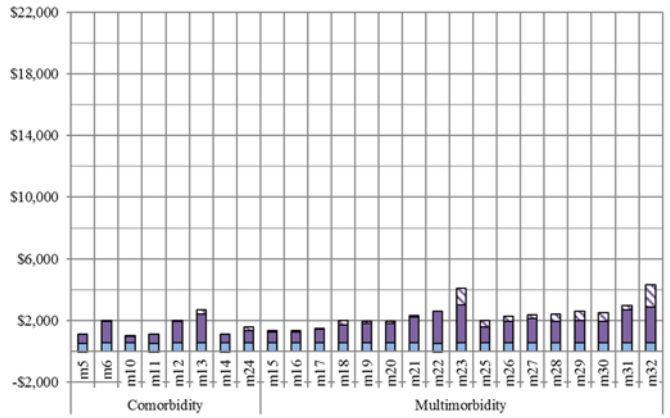
Comorbidity interaction groups:
p6 (3=0, 6=1, 29=1, 53=0, 54=0, 56=0,58=0, 59=0)
p7 (3=1, 29=1, 53=0, 54=0, 56=0, 58=0, 59=0)
p8 (29=1, 53=1, 54=0, 56=0, 58=0, 59=0)
p11 (6=1, 29=0, 53=0, 56=1, 58=0, 59=0)
p12 (29=1, 53=0, 56=1, 58=0, 59=0)
p13 (6=0, 29=0, 53=1, 56=1, 58=0, 59=0)
p17 (29=0, 52=1, 53=0, 56=0, 58=1, 59=0)
p18 (29=0, 52=0, 53=1, 56=0, 58=1, 59=0)
p20 (3=0, 29=1, 53=0, 56=0, 58=1, 59=0)
p23 (6=0, 29=0, 52=0, 53=0, 56=1, 58=1, 59=0)
p36 (11=0, 29=0, 53=0, 56=1, 59=1)
p37 (11=1, 29=0, 53=0, 59=1)
p38 (3=0, 29=1, 52=0, 53=0, 59=1)
p41 (29=0, 53=1, 55=0, 56=0, 59=1)
p43 (29=0, 53=0, 55=1, 59=1)

Interaction groups

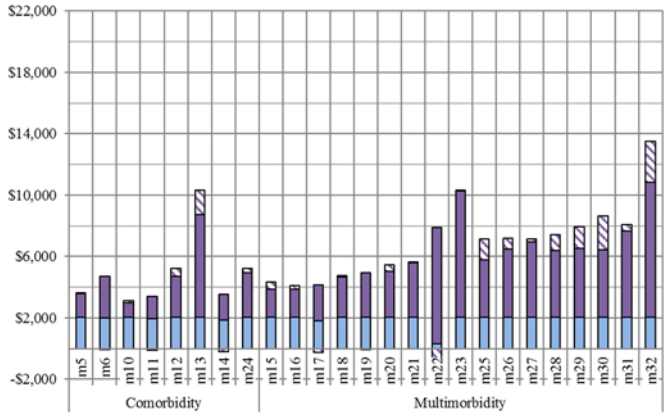
Multimorbidity interaction groups:
p14 (6=1, 29=0, 53=1, 56=1, 58=0, 59=0)
p15 (29=0, 53=1, 56=1, 58=0, 59=0)
p16 (6=1, 29=0, 53=0, 56=1, 58=0, 59=0)
p19 (29=0, 52=1, 53=1, 56=1, 58=0, 59=0)
p21 (3=1, 29=1, 53=0, 56=0, 58=1, 59=0)
p22 (29=1, 53=1, 56=0, 58=1, 59=0)
p24 (6=0, 29=0, 52=1, 53=0, 56=1, 58=1, 59=0)
p26 (3=0, 29=1, 52=0, 53=0, 56=1, 58=1, 59=0)
p27 (3=0, 29=1, 52=1, 53=0, 56=1, 58=1, 59=0)
p28 (3=1, 29=1, 53=0, 56=1, 58=1, 59=0)
p29 (29=0, 52=0, 53=1, 56=1, 58=1, 59=0)
p30 (29=0, 52=1, 53=1, 56=1, 58=1, 59=0)
p31 (3=0, 6=0, 29=1, 52=0, 53=1, 56=1, 58=1, 59=0)
p32 (3=0, 6=1, 29=1, 52=0, 53=1, 56=1, 58=1, 59=0)
p33 (3=0, 29=1, 52=1, 53=1, 56=1, 58=1, 59=0)
p39 (3=0, 29=1, 52=1, 53=0, 59=1)
p40 (3=1, 29=1, 53=0, 59=1)
p42 (29=0, 53=1, 55=0, 56=1, 59=1)
p44 (3=0, 6=0, 29=1, 52=0, 53=1, 59=1)
p45 (3=0, 6=1, 29=1, 52=0, 53=1, 59=1)
p46 (3=0, 29=1, 52=1, 53=1, 59=1)
p47 (3=1, 29=1, 52=0, 53=1, 59=1)
p48 (3=1, 29=1, 52=1, 53=1, 59=1)

Figure 3 – Annual medical services costs associated with morbidities and morbidity interactions

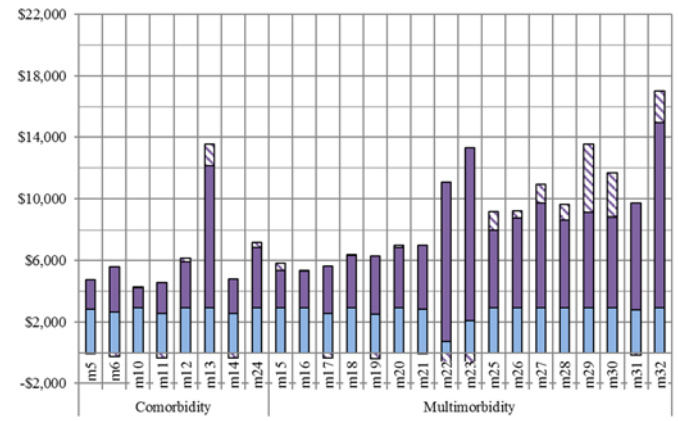
a) 50th percentile



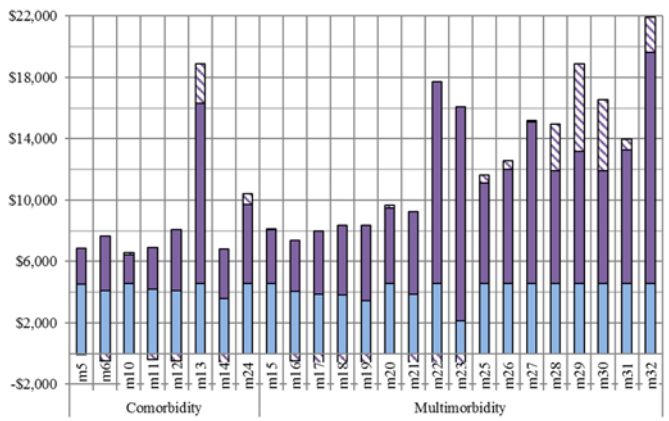
b) 90th percentile



c) 95th percentile



d) 98th percentile



Intercept Individual morbidities Morbidity interaction

Morbidity and morbidity interaction groups index

Morbidity groups

Diabetes mellitus (3)
Other mental and behavioral disorders (7)
Conduction disord. & cardiac arrhythmias (12)
Hyperlipidemia (29)
Eye inflammation (36)
Functional disorders of the digestive system (43)
Ischemic heart diseases (51)
Diseases of esophagus and peptic ulcer (53)
Malignant neoplasm (54)
Inflam. polyarthritis & connective tissue disord. (55)
Pain (56)
Bacterial infection or septicemia (58)
Thrombo-embolic risk or disease (59)

Interaction groups

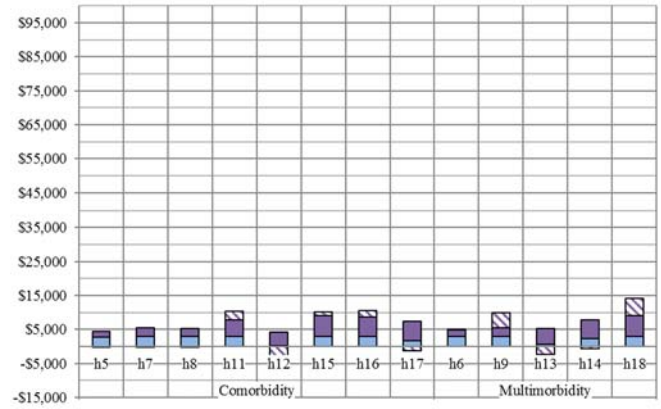
Comorbidity interaction groups:
m5 (36=0, 53=1, 54=0, 56=1, 58=0, 59=0)
m6 (36=1, 54=0, 56=1, 58=0, 59=0)
m10 (29=1, 36=0, 53=0, 54=0, 56=0, 58=1, 59=0)
m11 (36=0, 53=1, 54=0, 56=0, 58=1, 59=0)
m12 (36=1, 54=0, 56=0, 58=1, 59=0)
m13 (54=1, 56=0, 58=1, 59=0)
m14 (7=0, 29=0, 36=0, 53=0, 54=0, 56=1, 58=1, 59=0)
m24 (54=0, 56=0, 58=1, 59=1)

Interaction groups

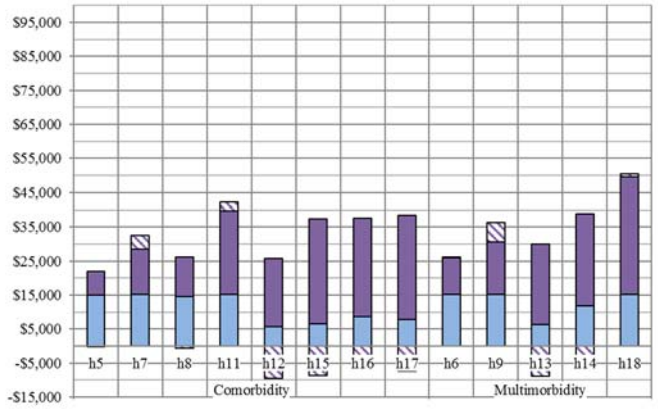
Multimorbidity interaction groups:
m15 (7=1, 29=0, 36=0, 53=0, 54=0, 56=1, 58=1, 59=0)
m16 (29=1, 36=0, 53=0, 54=0, 56=1, 58=1, 59=0)
m17 (3=0, 36=0, 43=0, 53=1, 54=0, 55=0, 56=1, 58=1, 59=0)
m18 (3=1, 36=0, 43=0, 53=1, 54=0, 55=0, 56=1, 58=1, 59=0)
m19 (36=0, 43=0, 53=1, 54=0, 55=1, 56=1, 58=1, 59=0)
m20 (36=0, 43=1, 53=1, 54=0, 56=1, 58=1, 59=0)
m21 (36=1, 54=0, 56=1, 58=1, 59=0)
m22 (43=0, 54=1, 56=1, 58=1, 59=0)
m23 (43=1, 54=1, 56=1, 58=1, 59=0)
m25 (12=0, 36=0, 43=0, 51=0, 53=0, 54=0, 55=0, 56=1, 58=1, 59=1)
m26 (12=0, 36=0, 43=0, 51=0, 53=1, 54=0, 55=0, 56=1, 58=1, 59=1)
m27 (12=0, 36=0, 43=0, 51=1, 54=0, 55=0, 56=1, 58=1, 59=1)
m28 (12=0, 36=0, 43=0, 54=0, 55=1, 56=1, 58=1, 59=1)
m29 (12=1, 36=0, 43=0, 54=0, 56=1, 58=1, 59=1)
m30 (36=0, 43=1, 54=0, 56=1, 58=1, 59=1)
m31 (36=1, 54=0, 56=1, 58=1, 59=1)
m32 (54=1, 58=1, 59=1)

Figure 4 – Annual hospital costs associated with morbidities and morbidity interactions

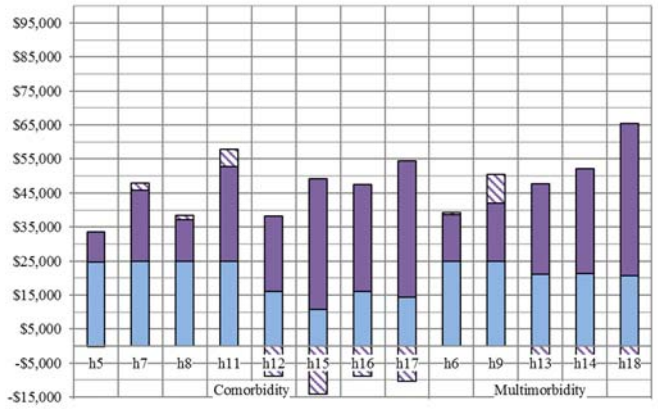
a) 50th percentile



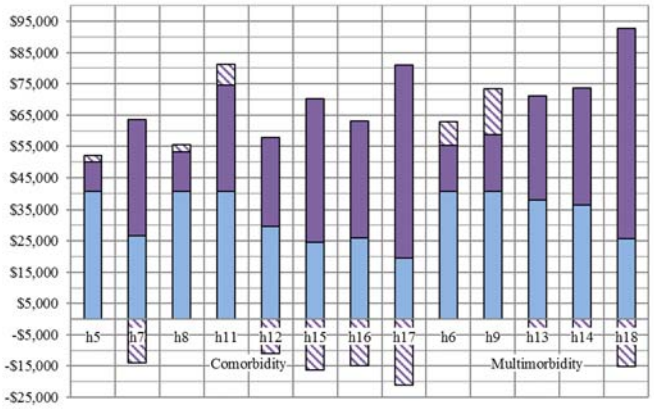
b) 90th percentile



c) 95th percentile



d) 98th percentile



Intercept Individual morbidities Morbidity interaction

Morbidity and morbidity interaction groups index

Morbidity groups	Interaction groups
Heart failure (13) Hyperlipidemia (29) Functional disorders of the digestive system (43) Ischemic heart diseases (51) Malignant neoplasm (54) Pain (56) Bacterial infection or septicemia (58) Thrombo-embolic risk or disease (59)	Comorbidity interaction groups: h5 (13=0, 43=0, 54=0, 56=1, 58=1, 59=0) h7 (13=1, 54=0, 58=1, 59=0) h8 (43=0, 54=1, 58=1, 59=0) h11 (13=0, 29=0, 51=0, 54=0, 56=1, 59=1) h12 (13=0, 29=1, 51=0, 54=0, 58=0, 59=1) h15 (13=0, 51=1, 54=0, 59=1) h16 (13=0, 54=1, 59=1) h17 (13=1, 43=0, 59=1) Multimorbidity interaction groups: h6 (13=0, 43=1, 54=0, 56=1, 58=1, 59=0) h9 (43=1, 54=1, 58=1, 59=0) h13 (13=0, 29=1, 51=0, 54=0, 56=0, 58=1, 59=1) h14 (13=0, 29=1, 51=0, 54=0, 56=1, 58=1, 59=1) h18 (13=1, 43=1, 59=1)

Table 3 – Cost estimates related to age, weight and death variables

Model	Variables	Percentile				
		10	50	90	95	98
Total	Constant (incl. ≥ 75)	13.06	430.50	4,127.58	8,481.36	16,487.08
	Age: 45-49	-16.70	-120.44	-1,435.12	-3,333.33	-7,420.78
	Age: 50-54	-4.74	-74.22	-1,263.12	-3,192.68	-7,461.34
	Age: 55-59	4.49	-15.91	-904.78	-2,472.51	-5,885.02
	Age: 60-64	9.51	24.76	-592.26	-1,949.78	-5,034.04
	Age: 65-69	-1.30	2.63	-728.14	-2,273.99	-5,156.66
	Age: 70-74	15.51	39.39	-555.27	-1,754.92	-3,582.11
	BMI - Underweight	-30.05	-69.08	-66.31	752.28	828.70
	BMI - Overweight	1.72	-0.64	-19.37	-63.28	18.86
	BMI - Obese Class 1	1.98	18.06	287.58	509.11	942.43
	BMI - Obese Class 2	13.03	65.75	683.08	946.46	2,601.98
	BMI - Obese Class 3	25.58	90.73	384.96	793.92	2,431.37
	Death in 2010	116.93	11,625.07	31,878.01	41,305.39	56,024.46
	Death in 2011	246.49	3,653.90	20,486.50	25,279.99	29,953.11
Pharm.	Constant (incl. ≥ 75)	-105.06	17.96	451.53	670.50	1,136.69
	Age: 45-49	-43.04	-97.69	-157.03	-240.97	-419.03
	Age: 50-54	-39.16	-85.01	-139.39	-207.01	-336.31
	Age: 55-59	-39.03	-72.18	-123.44	-191.68	-317.01
	Age: 60-64	-29.14	-50.27	-91.13	-141.29	-248.05
	Age: 65-69	-11.26	-7.91	-27.28	-84.64	-249.29
	Age: 70-74	3.89	25.03	21.48	-22.79	-156.24
	BMI - Underweight	-16.30	-8.37	10.85	35.18	-12.99
	BMI - Overweight	15.50	23.18	32.05	24.61	29.85
	BMI - Obese Class 1	12.36	35.53	51.63	43.42	42.67
	BMI - Obese Class 2	15.22	59.77	67.97	107.11	135.24
	BMI - Obese Class 3	33.11	102.77	147.57	155.95	59.24
	Death in 2010	-517.00	-653.80	-615.62	-696.17	-785.47
	Death in 2011	-1.01	169.78	905.94	1,531.13	3,675.42
Med. Serv.	Constant (incl. ≥ 75)	130.33	564.74	2,048.40	2,910.30	4,567.92
	Age: 45-49	-78.48	-240.13	-568.42	-712.13	-1,115.92
	Age: 50-54	-69.94	-205.68	-485.53	-598.38	-965.10
	Age: 55-59	-55.72	-154.68	-329.90	-375.92	-489.29
	Age: 60-64	-48.18	-113.59	-152.75	-90.88	-178.52
	Age: 65-69	-56.93	-128.67	-159.45	-121.43	-167.91
	Age: 70-74	-43.97	-81.81	-71.12	17.84	-114.84
	BMI - Underweight	-13.00	-11.44	27.86	142.23	118.86
	BMI - Overweight	-5.25	-11.31	-19.97	-76.50	-139.04
	BMI - Obese Class 1	4.56	4.96	84.32	124.41	132.32
	BMI - Obese Class 2	6.04	23.94	155.39	275.36	448.28
	BMI - Obese Class 3	7.44	2.86	90.45	242.69	352.78
	Death in 2010	-435.21	-621.83	-487.91	-232.34	1,083.99
	Death in 2011	-43.09	152.96	1,003.94	2,137.52	2,865.12
Hosp.	Constant (incl. ≥ 75)	1,288.27	3,034.48	15,246.14	24,893.88	40,660.58
	Age: 45-49	-20.22	-296.65	-4,441.49	-7,324.74	-9,326.13
	Age: 50-54	-24.97	-415.69	-4,375.62	-7,922.06	-11,615.05
	Age: 55-59	-17.75	-340.21	-3,388.56	-5,481.91	-7,923.40
	Age: 60-64	-21.05	-225.92	-2,552.23	-3,470.21	-3,125.14
	Age: 65-69	-50.82	-422.86	-2,654.50	-3,696.52	-4,399.41
	Age: 70-74	-44.68	-232.09	-1,563.20	-2,712.41	-1,787.29
	BMI - Underweight	-5.95	-76.29	1,458.13	5,558.97	5,311.61
	BMI - Overweight	-0.16	5.56	-107.77	-181.27	449.88
	BMI - Obese Class 1	4.00	177.79	521.51	526.59	-178.35
	BMI - Obese Class 2	33.44	385.28	1,390.28	2,847.10	516.36
	BMI - Obese Class 3	-3.21	78.15	1,401.62	844.45	-866.49
	Death in 2010	2,710.33	13,081.06	29,974.96	36,836.50	50,140.64
	Death in 2011	625.45	6,687.20	19,441.13	23,185.52	27,840.16

7 Appendix

Table 1A – Specification of the explanatory variables in the models of healthcare costs

Type of variable	Description	Model one	Model two	Model three
Count variable of coexisting morbidities	Count variable of the number of coexisting morbidities using the morbidity classification mentioned below.	✓	×	×
Dummy variables of morbidity groups	Coded using DRG-10 and ATC codes – refer to Halfon et al. (2013) for the full specification. The full list of morbidity groups is included in the appendix in Table 2A.	×	✓	✓
Dummy variables of morbidity interactions that capture multimorbidity	These are a set of dummy variables that capture selected interaction groups from the results of the CHIAD analysis that are unique for each type of healthcare cost. The full sets of interactions identified using the CHAID procedure are included in the appendix in Tables 3A to 6A.	×	×	✓
Dummy variables of self-reported health status	Dummy variables created using the five point scale variable from the baseline questionnaire of the 45 and Up study. For further details refer to 45 and Up Study Collaborators (2011).	✓	✓	✓
Demographic, economic and lifestyle dummy variables	Selected variables included in the baseline questionnaire and described in 45 and Up Study Collaborators (2011). The selection of these variables is based on the models utilised in Ellis et al. (2013) and Jones, Lomas, and Rice (2015). These variables include an individual's demographics, income, labour force status and lifestyle habits (including identifying those who smoke, are underweight or obese and consume high levels of alcohol).	✓	✓	✓
Dummy variables for the case were an individual dies in the same year (2010) or the subsequent year (2011)	Month and year of death are sourced from linked data provided by the NSW Registry of Births, Deaths and Marriages. This has been used to create dummy variables that indicate whether death occurred in 2010 or 2011, respectively.	✓	✓	✓

Table 2A – Morbidity groups and matching morbidity IDs

Morbidity Groups	Morbidity ID
Tuberculosis	1
Human immunodeficiency virus disease	2
Diabetes mellitus	3
Mental and behavioral disorders due to alcohol	4
Schizophrenia, schizotypal and delusional disorders	5
Mood [affective] disorders	6
Other mental and behavioral disorders	7
Alzheimer's disease	8
Multiple sclerosis	9
Epilepsy	10
Hypertensive disease	11
Conduction disorders and cardiac arrhythmias	12
Heart failure	13
Crohn's disease and ulcerative colitis	14
Infections of the skin and subcutaneous tissue	15
Hyperplasia of prostate	16
Inflammatory diseases of female pelvic organs	17
Poisoning by drugs and biological substances	18
Paludism	19
Parasitosis	20
Hepatitis B or C	21
Viral diseases	22
Anemia – nutritional	23
Non nutritional anemia	24
Neutropenia	25
Thyroid disorders	26
Other endocrine, nutritional and metabolic diseases	27
Obesity	28
Hyperlipidemia	29
Mental and behavioral disorders due to opioids	30
Mental and behavioral disorders due to tobacco	31
Migraine	32
Parkinson's disease	33
Conjunctivitis	34
Glaucoma	35
Eye inflammation	36
Disorders of external ear	37
Vertigo	38
Influenza	39
Other acute upper respiratory infections	40
Other diseases of upper respiratory tract – allergy	41
Hemorrhoids	42
Functional disorders of the digestive system	43
Psoriasis	44
Dermatitis	45
Acne	46
Gout	47
Osteoporosis	48
Transplanted organ status	49
Gastroenteritis of presumed infectious origin	50
Ischemic heart diseases	51
Reactive airway disease	52
Diseases of esophagus and peptic ulcer	53
Malignant neoplasm	54
Inflammatory polyarthritis and connective tissue disorders	55
Pain	56
Mycosis	57
Bacterial infection or septicemia	58
Thrombo-embolic risk or disease	59
Hemorrhage risk or disease	60

Note:this list corresponds to the morbidities contained in the DDMG outlined in Halfon et al. (2013)

Table 3A – Morbidity interactions identified using CHAID – Total healthcare costs

Interaction Group	Morbidity groups																Number of morbidity groups identified	Number of co-existing morbidities	% in top decile
	Diabetes mellitus	Mood [affective] disorders	Heart failure	Hyperlipidemia	Glaucoma	Eye inflammation	Functional disorders of the digestive system	Osteoporosis	Ischemic heart diseases	Reactive airway disease	Diseases of esophagus and peptic ulcer	Malignant neoplasm	Inflammatory polyarthritis and connective tissue disorders	Pain	Bacterial infection or septicemia	Thrombo-embolic risk or disease			
t1		×		×	×	×		×		×	×	×	×	×	×	×	12	0	2
t2		×		×	×	×		✓		×	×	×	×	×	×	×	12	0	4
t3		×		×	✓	×				×	×	×	×	×	×	×	11	0	2
t4		×		×		×				✓	×	×	×	×	×	×	10	0	2
t5		✓		×		×					×	×	×	×	×	×	9	0	4
t6	×			✓		×					×	×	×	×	×	×	9	0	2
t7	✓			✓		×					×	×	×	×	×	×	9	2	3
t8						✓					×	×	×	×	×	×	7	0	9
t9				×							×	×	×	×	×	×	7	0	2
t10				✓							✓	×	×	×	×	×	7	2	4
t11												✓	×	×	×	×	5	0	29
t12													✓	×	×	×	4	0	15
t13				×							×	×		✓	×	×	6	0	7
t14				×							✓	×		✓	×	×	6	2	10
t15				✓								×		✓	×	×	5	2	8
t16												✓		✓	×	×	4	2	38
t17				×							×	×		×	✓	×	6	0	6
t18				×							×	×		✓	✓	×	6	2	10
t19				✓							×	×			✓	×	5	2	9
t20											✓	×		×	✓	×	5	2	8
t21							×				✓	×		✓	✓	×	6	3	14
t22							✓				✓	×		✓	✓	×	6	4	25
t23							×					✓			✓	×	4	2	40
t24							✓					✓			✓	×	4	3	72
t25															×	✓	2	0	22
t26									×			×		×	✓	✓	5	2	22
t27									✓			×		×	✓	✓	5	3	38
t28							×		×			×		✓	✓	✓	6	3	33
t29							×		✓			×		✓	✓	✓	6	4	47
t30			×				✓					×		✓	✓	✓	6	4	47
t31			✓				✓					×		✓	✓	✓	6	5	64
t32												✓			✓	✓	3	3	69

Table 4A – Morbidity interactions identified using CHAID – Pharmaceutical costs

Interaction Group	Morbidity groups													Number of morbidity groups identified	Number of co-existing morbidities	% in top decile
	Diabetes mellitus	Mood [affective] disorders	Hypertensive disease	Hyperlipidemia	Glaucoma	Osteoporosis	Reactive airway disease	Diseases of esophagus and peptic ulcer	Malignant neoplasm	Inflammatory polyarthritis and connective tissue disorders	Pain	Bacterial infection or septicemia	Thrombo-embolic risk or disease			
p1		×		×	×	×			×		×	×	×	8	0	2
p2		×		×	✓	×			×		×	×	×	8	0	1
p3		✓		×		×			×		×	×	×	7	0	2
p4				×		✓			×		×	×	×	6	0	3
p5	×	×		✓				×	×		×	×	×	8	0	1
p6	×	✓		✓				×	×		×	×	×	8	2	4
p7	✓			✓				×	×		×	×	×	7	2	10
p8				✓				✓	×		×	×	×	6	2	4
p9									✓		×	×	×	4	0	23
p10		×		×				×			✓	×	×	6	0	4
p11		✓		×				×			✓	×	×	6	2	7
p12				✓				×			✓	×	×	5	2	6
p13		×		×				✓			✓	×	×	6	2	5
p14		✓		×				✓			✓	×	×	6	3	13
p15				✓				✓			✓	×	×	5	3	14
p16		✓		×				×			✓	✓	×	6	3	12
p17				×			✓	×			×	✓	×	6	2	4
p18				×			×	✓			×	✓	×	6	2	6
p19				×			✓	✓			×	✓	×	6	3	10
p20	×			✓				×			×	✓	×	6	2	4
p21	✓			✓				×			×	✓	×	6	3	15
p22				✓				✓			×	✓	×	5	3	13
p23		×		×			×	×			✓	✓	×	7	2	4
p24		×		×			✓	×			✓	✓	×	7	3	7
p25		✓		×				×			✓	✓	×	6	3	12
p26	×			✓			×	×			✓	✓	×	7	3	6
p27	×			✓			✓	×			✓	✓	×	7	4	15
p28	✓			✓				×			✓	✓	×	6	4	25
p29				×			×	✓			✓	✓	×	6	3	11
p30				×			✓	✓			✓	✓	×	6	4	19
p31	×	×		✓			×	✓			✓	✓	×	8	4	11

p32	×	✓		✓			×	✓			✓	✓	×	8	5	23
p33	×			✓			✓	✓			✓	✓	×	7	5	31
p34	✓			✓				✓			✓	✓	×	6	5	44
p35			×	×				×			×		✓	5	0	5
p36			×	×				×			✓		✓	5	2	9
p37			✓	×				×					✓	4	2	11
p38	×			✓			×	×					✓	5	2	10
p39	×			✓			✓	×					✓	5	3	27
p40	✓			✓				×					✓	4	3	34
p41				×				✓		×	×		✓	5	2	9
p42				×				✓		×	✓		✓	5	3	18
p43				×				×		✓			✓	4	2	19
p44	×	×		✓			×	✓					✓	6	3	24
p45	×	✓		✓			×	✓					✓	6	4	41
p46	×			✓			✓	✓					✓	5	4	51
p47	✓			✓			×	✓					✓	5	4	54
p48	✓			✓			✓	✓					✓	5	5	73

Table 5A – Morbidity interactions identified using CHAID – Medical services costs

Interaction Group	Morbidity groups													Number of morbidity groups identified	Number of co-existing morbidities	% in top decile
	Diabetes mellitus	Other mental and behavioral disorders	Conduction disorders and cardiac arrhythmias	Hyperlipidemia	Eye inflammation	Functional disorders of the digestive system	Ischemic heart diseases	Diseases of esophagus and peptic ulcer	Malignant neoplasm	Inflammatory polyarthritis and connective tissue disorders	Pain	Bacterial infection or septicemia	Thrombo-embolic risk or disease			
m1				×				×	×		×	×	×	6	0	2
m2				✓				×	×		×	×	×	6	0	4
m3								✓	×		×	×	×	5	0	6
m4					×			×	×		✓	×	×	6	0	6
m5					×			✓	×		✓	×	×	6	2	9
m6					✓				×		✓	×	×	5	2	24
m7									✓			×	×	3	0	27
m8												×	✓	2	0	18
m9				×	×			×	×		×	✓	×	7	0	5
m10				✓	×			×	×		×	✓	×	7	2	7
m11					×			✓	×		×	✓	×	6	2	10
m12					✓				×		×	✓	×	5	2	27
m13									✓		×	✓	×	4	2	38
m14		×		×	×			×	×		✓	✓	×	8	2	7
m15		✓		×	×			×	×		✓	✓	×	8	3	14
m16				✓	×			×	×		✓	✓	×	7	3	13
m17	×				×	×		✓	×	×	✓	✓	×	9	3	12
m18	✓				×	×		✓	×	×	✓	✓	×	9	4	17
m19					×	×		✓	×	✓	✓	✓	×	8	4	20
m20					×	✓		✓	×		✓	✓	×	7	4	22
m21					✓				×		✓	✓	×	5	3	34
m22						×			✓		✓	✓	×	5	3	38
m23						✓			✓		✓	✓	×	5	4	56
m24									×		×	✓	✓	4	2	24
m25			×		×	×	×	×	×	×	✓	✓	✓	10	3	26
m26			×		×	×	×	✓	×	×	✓	✓	✓	10	4	28
m27			×		×	×	✓		×	×	✓	✓	✓	9	4	30
m28			×		×	×			×	✓	✓	✓	✓	8	4	36
m29			✓		×	×			×		✓	✓	✓	7	4	36
m30					×	✓			×		✓	✓	✓	6	4	40
m31					✓				×		✓	✓	✓	5	4	47
m32									✓			✓	✓	3	3	61

Table 6A – Morbidity interactions identified using CHAID – Hospital costs

Interaction Group	Morbidity Groups								Number of morbidity groups identified	Number of co-existing morbidities	% in top decile
	Heart failure	Hyperlipidemia	Functional disorders of the digestive system	Ischemic heart diseases	Malignant neoplasm	Pain	Bacterial infection or septicemia	Thrombo-embolic risk or disease			
h1					×	×	×	×	4	0	3
h2					×	✓	×	×	4	0	5
h3					✓		×	×	3	0	11
h4	×				×	×	✓	×	5	0	5
h5	×		×		×	✓	✓	×	6	2	6
h6	×		✓		×	✓	✓	×	6	3	10
h7	✓				×		✓	×	4	2	19
h8			×		✓		✓	×	4	2	12
h9			✓		✓		✓	×	4	3	30
h10	×	×		×	×	×		✓	6	0	16
h11	×	×		×	×	✓		✓	6	2	25
h12	×	✓		×	×		×	✓	6	2	10
h13	×	✓		×	×	×	✓	✓	7	3	11
h14	×	✓		×	×	✓	✓	✓	7	4	19
h15	×			✓	×			✓	4	2	19
h16	×				✓			✓	3	2	27
h17	✓		×					✓	3	2	27
h18	✓		✓					✓	3	3	40

Table 7A – Cost estimates related to interaction and selected variables – Total healthcare

	Percentile				
Variables	**10**	**50**	**90**	**95**	**98**
Diabetes (3)	562.37	853.85	62.20	-654.56	-1,962.34
Heart failure (13)	208.51	719.12	6,603.54	12,456.31	17,254.25
Hyperlipidemia (29)	632.50	861.43	484.84	100.49	-854.11
Functional disorders of the digestive system (43)	348.01	1,085.54	4,195.11	5,747.84	8,846.50
Ischemic heart diseases (51)	759.19	2,396.87	10,228.24	15,038.92	27,001.85
Diseases of esophagus and peptic ulcer (53)	404.17	837.62	1,743.67	1,910.36	2,727.98
Malignant neoplasm (54)	1,436.38	4,991.58	22,055.52	27,967.61	36,688.99
Pain (56)	260.74	775.13	4,329.79	6,013.32	8,816.16
Bacterial infection or septicemia (58)	155.01	513.56	2,465.25	4,480.23	7,640.62
Thrombo-embolic risk or disease (59)	570.93	1,409.51	10,241.91	17,620.02	26,054.80
t7 (3=1, 29=1, 36=0, 53=0, 54=0, 55=0, 56=0, 58=0, 59=0)	69.08	320.81	1,651.94	2,424.64	3,738.82
t10 (29=1, 53=1, 54=0, 55=0, 56=0, 58=0, 59=0)	-46.97	-3.30	223.54	465.41	227.01
t14 (29=0, 53=1, 54=0, 56=1, 58=0, 59=0)	-44.06	-45.46	-265.58	1,050.37	2,621.50
t15 (29=1, 54=0, 56=1, 58=0, 59=0)	-7.48	-147.71	-1,125.67	-769.09	675.78
t16 (54=1, 56=1, 58=0, 59=0)	-49.43	-899.92	-6,968.19	-10,807.11	-13,218.64
t18 (29=0, 53=0, 54=0, 56=1, 58=1, 59=0)	-65.84	-203.92	-1,397.72	-1,765.62	-2,769.69
t19 (29=1, 53=0, 54=0, 58=1, 59=0)	41.96	-17.43	-303.95	-619.50	-1,177.83
t20 (53=1, 54=0, 56=0, 58=1, 59=0)	11.13	-114.33	-694.03	-1,726.45	-3,123.01
t23 (43=0, 54=1, 58=1, 59=0)	692.91	748.92	-1,043.69	-1,513.76	-4,800.69
t26 (51=0, 54=0, 56=0, 58=1, 59=1)	193.74	286.87	1,965.04	2,428.46	4,275.61
t21 (43=0, 53=1, 54=0, 56=1, 58=1, 59=0)	171.18	7.51	-1,389.79	-443.34	-2,352.72
t22 (43=1, 53=1, 54=0, 56=1, 58=1, 59=0)	239.94	278.31	1,580.04	1,581.17	-810.74
t24 (43=1, 54=1, 58=1, 59=0)	2,871.10	10,473.58	10,767.15	10,160.47	4,320.23
t27 (51=1, 54=0, 56=0, 58=1, 59=1)	504.50	479.66	3,347.40	-26.80	-10,183.73
t28 (43=0, 51=0, 54=0, 56=1, 58=1, 59=1)	449.69	1,417.24	11,532.32	10,105.81	6,937.52
t29 (43=0, 51=1, 54=0, 56=1, 58=1, 59=1)	620.74	2,165.82	10,386.42	3,574.40	-15,141.26
t30 (13=0, 43=1, 54=0, 56=1, 58=1, 59=1)	799.17	2,832.62	17,538.35	12,874.71	9,864.00
t31 (13=1, 43=1, 54=0, 56=1, 58=1, 59=1)	1,679.52	6,989.78	18,743.12	17,563.37	5,712.70
t32 (54=1, 58=1, 59=1)	2,438.98	9,313.10	9,151.72	1,279.73	-8,634.62
Death in 2010	116.93	11,625.07	31,878.01	41,305.39	56,024.46
Death in 2011	246.49	3,653.90	20,486.50	25,279.99	29,953.11
Intercept	13.06	430.50	4,127.58	8,481.36	16,487.08
Adj. R-square	0.1543	0.2383	0.3789	0.4199	0.4407
N	240,329				

Note: the relevant morbidity IDs are shown in brackets and correspond with those shown in Table 2A.

Table 8A – Cost estimates related to interaction and selected variables – Pharmaceutical

Variables	Percentile				
	10	50	90	95	98
Diabetes mellitus (3)	255.68	560.17	1,109.86	1,362.10	1,528.46
Mood [affective] disorders (6)	163.57	316.87	416.93	446.67	438.43
Hypertensive disease (11)	125.87	200.46	287.59	330.05	551.31
Hyperlipidemia (29)	294.35	610.43	776.77	839.59	821.19
Reactive airway disease (52)	156.07	239.12	418.51	477.90	467.76
Diseases of esophagus & peptic ulcer (53)	148.84	216.44	235.85	218.50	164.23
Inflam. polyarthritis & connective tissue disord. (55)	77.68	182.55	834.64	1,314.46	4,367.85
Pain (56)	83.86	98.06	145.56	124.19	66.47
Bacterial infection or septicemia (58)	54.11	46.55	20.35	2.37	-37.49
Thrombo-embolic risk or disease (59)	120.52	223.81	464.41	433.03	540.97
p6 (3=0, 6=1, 29=1, 53=0, 54=0, 56=0,58=0, 59=0)	36.32	13.01	47.32	99.17	57.18
p7 (3=1, 29=1, 53=0, 54=0, 56=0, 58=0, 59=0)	192.69	121.08	180.52	102.00	250.16
p8 (29=1, 53=1, 54=0, 56=0, 58=0, 59=0)	62.49	72.11	141.34	159.57	211.67
p11 (6=1, 29=0, 53=0, 56=1, 58=0, 59=0)	-73.99	-93.75	-3.82	86.17	467.13
p12 (29=1, 53=0, 56=1, 58=0, 59=0)	74.48	56.75	72.58	103.35	282.35
p13 (6=0, 29=0, 53=1, 56=1, 58=0, 59=0)	-51.99	52.55	134.53	282.27	541.17
p17 (29=0, 52=1, 53=0, 56=0, 58=1, 59=0)	-75.82	-27.01	104.68	179.90	240.89
p18 (29=0, 52=0, 53=1, 56=0, 58=1, 59=0)	-44.57	3.14	95.87	138.41	185.18
p20 (3=0, 29=1, 53=0, 56=0, 58=1, 59=0)	61.72	37.91	110.65	114.22	141.49
p23 (6=0, 29=0, 52=0, 53=0, 56=1, 58=1, 59=0)	-55.22	-61.55	-25.21	32.61	98.77
p36 (11=0, 29=0, 53=0, 56=1, 59=1)	-55.30	-42.35	-102.77	142.10	460.56
p37 (11=1, 29=0, 53=0, 59=1)	-53.87	-77.33	-172.20	-146.21	-259.76
p38 (3=0, 29=1, 52=0, 53=0, 59=1)	181.90	260.18	392.47	521.97	562.94
p41 (29=0, 53=1, 55=0, 56=0, 59=1)	-26.31	63.66	168.06	311.82	652.39
p43 (29=0, 53=0, 55=1, 59=1)	-35.11	-21.28	64.17	621.93	1,608.63
p14 (6=1, 29=0, 53=1, 56=1, 58=0, 59=0)	-28.15	67.18	206.70	309.22	911.01
p15 (29=0, 53=1, 56=1, 58=0, 59=0)	195.19	171.29	302.74	427.53	668.31
p16 (6=1, 29=0, 53=0, 56=1, 58=0, 59=0)	-118.16	-106.41	83.13	471.72	753.33
p19 (29=0, 52=1, 53=1, 56=1, 58=0, 59=0)	0.18	141.52	397.86	540.36	602.62
p21 (3=1, 29=1, 53=0, 56=0, 58=1, 59=0)	134.91	35.07	153.57	143.60	365.51
p22 (29=1, 53=1, 56=0, 58=1, 59=0)	189.14	181.83	265.64	288.46	520.73
p24 (6=0, 29=0, 52=1, 53=0, 56=1, 58=1, 59=0)	-107.53	0.45	263.10	385.02	456.88
p26 (3=0, 29=1, 52=0, 53=0, 56=1, 58=1, 59=0)	75.06	40.81	61.80	95.81	256.38
p27 (3=0, 29=1, 52=1, 53=0, 56=1, 58=1, 59=0)	109.08	107.84	252.33	307.13	298.04
p28 (3=1, 29=1, 53=0, 56=1, 58=1, 59=0)	144.16	86.32	243.62	241.78	373.85
p29 (29=0, 52=0, 53=1, 56=1, 58=1, 59=0)	-54.05	73.22	242.31	348.62	689.25
p30 (29=0, 52=1, 53=1, 56=1, 58=1, 59=0)	42.76	299.83	578.45	693.07	848.35
p31 (3=0, 6=0, 29=1, 52=0, 53=1, 56=1, 58=1, 59=0)	216.60	214.28	358.81	430.88	683.48
p32 (3=0, 6=1, 29=1, 52=0, 53=1, 56=1, 58=1, 59=0)	256.16	273.07	459.29	708.67	1,102.47
p33 (3=0, 29=1, 52=1, 53=1, 56=1, 58=1, 59=0)	358.73	414.83	664.34	778.53	895.54
p39 (3=0, 29=1, 52=1, 53=0, 59=1)	287.65	401.17	432.33	475.36	424.97
p40 (3=1, 29=1, 53=0, 59=1)	289.86	292.61	322.22	286.65	274.05
p42 (29=0, 53=1, 55=0, 56=1, 59=1)	79.12	184.45	300.66	522.19	861.33
p44 (3=0, 6=0, 29=1, 52=0, 53=1, 59=1)	338.84	472.83	693.26	813.56	948.87
p45 (3=0, 6=1, 29=1, 52=0, 53=1, 59=1)	475.51	508.35	588.78	850.23	1,000.73
p46 (3=0, 29=1, 52=1, 53=1, 59=1)	545.13	800.61	880.58	993.75	930.12
p47 (3=1, 29=1, 52=0, 53=1, 59=1)	600.08	641.27	627.23	623.61	546.16
p48 (3=1, 29=1, 52=1, 53=1, 59=1)	784.58	957.76	741.37	607.39	372.06
Death in 2010	-517.00	-653.80	-615.62	-696.17	-785.47
Death in 2011	-1.01	169.78	905.94	1,531.13	3,675.42
Intercept	-105.06	17.96	451.53	670.50	1,136.69
Adj. R-square	0.2125	0.3203	0.3451	0.3768	0.4328
N	162,741				

Note: the relevant morbidity IDs are shown in brackets and correspond with those shown in Table 2A.

Table 9A – Cost estimates related to interaction and selected variables – Medical services

	Percentile				
Variables	**10**	**50**	**90**	**95**	**98**
Diabetes mellitus (3)	139.04	250.01	275.34	332.86	418.80
Other mental and behavioral disorders (7)	71.34	123.20	162.05	185.63	295.09
Conduction disord. & cardiac arrhythmias (12)	224.32	380.82	730.88	1,144.98	2,065.22
Hyperlipidemia (29)	79.24	128.37	123.13	162.96	76.37
Eye inflammation (36)	276.90	1,122.29	1,862.10	1,866.92	2,156.54
Functional disorders of the digestive system (43)	154.62	354.61	641.50	847.16	796.91
Ischemic heart diseases (51)	171.86	524.40	1,185.68	1,761.63	3,981.40
Diseases of esophagus and peptic ulcer (53)	96.01	336.59	673.52	817.39	897.79
Malignant neoplasm (54)	412.62	1,558.08	5,886.63	8,088.99	9,941.93
Inflam. polyarthritis & connective tissue disord. (55)	165.09	375.99	584.25	675.04	809.69
Pain (56)	84.30	262.67	862.55	1,095.39	1,407.79
Bacterial infection or septicemia (58)	88.52	269.58	814.55	1,167.35	1,815.90
Thrombo-embolic risk or disease (59)	118.06	486.87	2,065.08	2,793.45	3,323.25
m5 (36=0, 53=1, 54=0, 56=1, 58=0, 59=0)	-25.24	-62.48	15.80	-65.19	-32.44
m6 (36=1, 54=0, 56=1, 58=0, 59=0)	-4.06	24.55	-51.85	-257.81	-479.20
m10 (29=1, 36=0, 53=0, 54=0, 56=0, 58=1, 59=0)	2.98	14.13	118.32	23.91	138.21
m11 (36=0, 53=1, 54=0, 56=0, 58=1, 59=0)	14.40	-48.45	-120.02	-331.69	-393.25
m12 (36=1, 54=0, 56=0, 58=1, 59=0)	23.52	39.15	490.87	194.25	-478.88
m13 (54=1, 56=0, 58=1, 59=0)	233.61	297.20	1,585.35	1,400.19	2,537.31
m14 (7=0, 29=0, 36=0, 53=0, 54=0, 56=1, 58=1, 59=0)	-15.65	-6.65	-191.79	-364.86	-995.69
m24 (54=0, 56=0, 58=1, 59=1)	121.28	225.92	273.11	297.08	721.75
m15 (7=1, 29=0, 36=0, 53=0, 54=0, 56=1, 58=1, 59=0)	52.82	117.57	425.14	466.82	53.13
m16 (29=1, 36=0, 53=0, 54=0, 56=1, 58=1, 59=0)	22.01	121.72	250.60	28.46	-492.11
m17 (3=0, 36=0, 43=0, 53=1, 54=0, 55=0, 56=1, 58=1, 59=0)	36.47	5.18	-261.00	-367.13	-713.92
m18 (3=1, 36=0, 43=0, 53=1, 54=0, 55=0, 56=1, 58=1, 59=0)	211.81	315.44	67.73	57.44	-758.93
m19 (36=0, 43=0, 53=1, 54=0, 55=1, 56=1, 58=1, 59=0)	55.44	112.97	-29.39	-381.41	-1,128.68
m20 (36=0, 43=1, 53=1, 54=0, 56=1, 58=1, 59=0)	100.49	153.57	402.72	145.25	174.27
m21 (36=1, 54=0, 56=1, 58=1, 59=0)	98.37	103.25	53.26	-51.59	-686.96
m22 (43=0, 54=1, 56=1, 58=1, 59=0)	171.27	-49.80	-1,736.06	-2,188.90	-2,332.21
m23 (43=1, 54=1, 56=1, 58=1, 59=0)	433.46	1,064.92	63.76	-811.74	-2,448.91
m25 (12=0, 36=0, 43=0, 51=0, 53=0, 54=0, 55=0, 56=1, 58=1, 59=1)	186.82	422.21	1,346.45	1,174.50	537.81
m26 (12=0, 36=0, 43=0, 51=0, 53=1, 54=0, 55=0, 56=1, 58=1, 59=1)	190.03	339.84	741.76	406.23	546.32
m27 (12=0, 36=0, 43=0, 51=1, 54=0, 55=0, 56=1, 58=1, 59=1)	229.76	230.16	161.12	1,216.46	86.07
m28 (12=0, 36=0, 43=0, 54=0, 55=1, 56=1, 58=1, 59=1)	268.53	460.43	1,044.81	971.39	3,022.20
m29 (12=1, 36=0, 43=0, 54=0, 56=1, 58=1, 59=1)	380.14	635.50	1,420.89	4,433.85	5,700.94
m30 (36=0, 43=1, 54=0, 56=1, 58=1, 59=1)	274.43	542.38	2,186.65	2,873.15	4,639.93
m31 (36=1, 54=0, 56=1, 58=1, 59=1)	460.10	247.55	440.70	-142.09	681.28
m32 (54=1, 58=1, 59=1)	695.61	1,422.68	2,676.18	2,045.60	2,295.15
Death in 2010	-435.21	-621.83	-487.91	-232.34	1,083.99
Death in 2011	-43.09	152.96	1,003.94	2,137.52	2,865.12
Intercept	130.33	564.74	2,048.40	2,910.30	4,567.92
Adj. R-square	0.0963	0.1837	0.2798	0.3106	0.3470
N	228,545				

Note: the relevant morbidity IDs are shown in brackets and correspond with those shown in Table 2A.

Table 10A – Cost estimates related to interaction and selected variables – Hospital

	Percentile				
Variables	**10**	**50**	**90**	**95**	**98**
Heart failure (13)	95.29	1,643.50	9,694.00	16,413.27	31,702.40
Hyperlipidemia (29)	-2.42	-213.99	-999.29	-1,367.05	-1,838.15
Functional disorders of the digestive system (43)	7.76	269.05	3,684.00	4,942.76	5,412.43
Ischemic heart diseases (51)	294.85	2,020.43	9,868.19	14,863.01	15,968.90
Malignant neoplasm (54)	288.74	1,459.20	7,999.12	7,841.32	7,361.97
Pain (56)	124.56	765.45	3,383.00	4,365.71	4,004.67
Bacterial infection or septicemia (58)	58.74	795.36	3,617.08	4,434.56	5,278.71
Thrombo-embolic risk or disease (59)	224.58	4,091.19	21,015.05	23,517.84	29,890.76
h5 (13=0, 43=0, 54=0, 56=1, 58=1, 59=0)	-41.73	-257.21	-222.60	-98.72	2,204.26
h7 (13=1, 54=0, 58=1, 59=0)	4.08	-57.18	3,881.69	2,242.88	-14,027.48
h8 (43=0, 54=1, 58=1, 59=0)	-9.76	-2.13	-708.88	1,167.39	2,147.30
h11 (13=0, 29=0, 51=0, 54=0, 56=1, 59=1)	2.12	2,391.47	2,626.44	5,172.50	6,690.70
h12 (13=0, 29=1, 51=0, 54=0, 58=0, 59=1)	-146.10	-2,784.36	-9,469.17	-8,834.80	-10,951.02
h15 (13=0, 51=1, 54=0, 59=1)	-7.48	947.35	-8,680.26	-14,056.61	-16,172.65
h16 (13=0, 54=1, 59=1)	390.49	1,946.58	-6,638.91	-8,786.40	-14,814.12
h17 (13=1, 43=0, 59=1)	19.11	-1,300.46	-7,442.80	-10,434.58	-21,157.14
h6 (13=0, 43=1, 54=0, 56=1, 58=1, 59=0)	-55.04	224.63	289.33	633.82	7,601.27
h9 (43=1, 54=1, 58=1, 59=0)	428.82	4,325.21	5,785.55	8,268.86	14,851.54
h13 (13=0, 29=1, 51=0, 54=0, 56=0, 58=1, 59=1)	-77.31	-2,340.71	-8,935.27	-3,732.51	-2,860.28
h14 (13=0, 29=1, 51=0, 54=0, 56=1, 58=1, 59=1)	-61.10	-664.52	-3,498.04	-3,637.96	-4,415.81
h18 (13=1, 43=1, 59=1)	733.10	5,103.90	873.53	-4,252.38	-15,007.17
Death in 2010	2,710.33	13,081.06	29,974.96	36,836.50	50,140.64
Death in 2011	625.45	6,687.20	19,441.13	23,185.52	27,840.16
Intercept	1,288.27	3,034.48	15,246.14	24,893.88	40,660.58
Adj. R-square	0.0224	0.1048	0.2468	0.2646	0.2810
N	64,432				

Note: the relevant morbidity IDs are shown in brackets and correspond with those shown in Table 2°.