HEALTH ECONOMICS
*Health Econ.* **22**: 1093–1110 (2013)
Published online 12 March 2013 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/hec.2916

# EXPLAINING HEALTH CARE EXPENDITURE VARIATION: LARGE-SAMPLE EVIDENCE USING LINKED SURVEY AND HEALTH ADMINISTRATIVE DATA

RANDALL P. ELLIS[a], DENZIL G. FIEBIG[b], MELIYANNI JOHAR[c], GLENN JONES[c,*] and ELIZABETH SAVAGE[c]

[a]*Department of Economics, Boston University, Boston MA, USA*
[b]*School of Economics, University of New South Wales, Sydney, Australia*
[c]*Economics Discipline Group, University of Technology Sydney, Sydney, Australia*

ABSTRACT

Explaining individual, regional, and provider variation in health care spending is of enormous value to policymakers but is often hampered by the lack of individual level detail in universal public health systems because budgeted spending is often not attributable to specific individuals. Even rarer is self-reported survey information that helps explain this variation in large samples. In this paper, we link a cross-sectional survey of 267 188 Australians age 45 and over to a panel dataset of annual healthcare costs calculated from several years of hospital, medical and pharmaceutical records. We use this data to distinguish between cost variations due to health shocks and those that are intrinsic (fixed) to an individual over three years. We find that high fixed expenditures are positively associated with age, especially older males, poor health, obesity, smoking, cancer, stroke and heart conditions. Being foreign born, speaking a foreign language at home and low income are more strongly associated with higher time-varying expenditures, suggesting greater exposure to adverse health shocks. 

Received 17 May 2012; Revised 1 December 2012; Accepted 25 January 2013

*JEL Classification*: I10; I13; C23

KEY WORDS: health expenditure; health insurance; risk adjustment; panel data

## 1. INTRODUCTION AND BACKGROUND

Understanding patterns of health care spending is of enormous policy and research interest but is often hampered by the available data. Researchers are often forced to choose between large samples of insurance-related administrative data or smaller samples of survey based information, with the former lacking individual demographic and self-reported information and the latter often lacking detailed spending and diagnostic data. In this paper, we exploit the linkage of a large Australian individual survey dataset of the population aged 45 years and over to four years of comprehensive administrative health utilisation data. The administrative records cover hospital admissions, emergency department presentations, claims for out-of-hospital medical services (e.g. general practitioner [GP] and specialist visits, diagnostic testing) and claims for prescription subsidised drugs by each survey respondent.[1] We use this data to calculate individual health expenditure and then identify variations in the individual health expenditures that are due to exogenous changes in health conditions and variations that are

*Correspondence to: Economics Discipline Group, University of Technology Sydney, Sydney, Australia. E-mail: glenn.jones@uts.edu.au

[1]The major excluded services are dental, optometry and allied health, nonsubsidised drugs and outpatient hospital services, which absorb 36% of the total health spending by the entire Australian population (including those under 45 years old). Unfortunately, there are no data on health spending on these services at the individual level that can be merged with our data.

specific to the individual. Exploiting the richness of the survey data and health provider information, we examine the influence of background characteristics and socioeconomic factors, lifestyle, existing health conditions and the use of a regular GP on variations in individual health expenditure. The results not only fill a gap in the understanding of individual health expenditures in publicly funded health systems that often lack individual-level health expenditure data but also improve our understanding of how linking administrative and survey information enhances predictions over time, a key topic for many health plan and provider payment systems.

All Australians have access to a universal public health system. The Australian Medicare program fully covers public inpatient care in public hospitals and heavily subsidises prescription drugs and most private medical care provided both in-hospital and out-of-hospital. General practitioners and specialists are paid on a fee for service basis, and their fees are unregulated. Over the past decade, the share of the public contribution to total health expenditure has been constant at around 70% (Australian Institute of Health and Welfare, AIHW, 2010). In 2009/2010, national total health expenditure was $124 billion, of which $116 billion was recurrent. Of the recurrent expenditure, 31% was spent on public hospitals, 19% on medical services and 14% on pharmaceuticals. Because of the relatively complex structure of health expenditure subsidies, an accurate measure of individual health expenditure can only be obtained through careful integration of various sources of health service use and the application of cost attribution algorithms that take into account the wide variation in patient circumstances (e.g. extended length of stay, transfer to another hospital, admitted after emergency presentation).

For our statistical analysis, we adopt a forward-looking or prospective model of health expenditure in which information from one year is used to predict the following year's health expenditure. The prospective model emphasises systematic variations in health expenditure due to chronic conditions as opposed to acute conditions or expensive one-off events, such as surgery, which are emphasised in a concurrent model in which claims information from one year is used to explain variations in health expenditures from the same (concurrent) period. Furthermore, the prospective model is more useful than the concurrent model for payment or budgeting purposes, as it can forecast the payers' future financial obligations. In addition, from an econometric perspective, the prospective model reduces the problem of reverse causality bias, which arises if health conditions are affected by changes in health expenditure.

The availability of four years of health expenditure data permits a three-year prospective panel sample and allows us to estimate the model using an individual fixed effects approach. An individual's health expenditure is a function of health conditions, use of pharmaceutical and medical services, background and economic characteristics, location, macroeconomic conditions, lifestyle and doctor's characteristics. Many of these potential predictors are constant over time (or slow changing), so their effects are not identified by the standard fixed effects approach. We therefore conduct a two-step analysis. In the first step, we estimate a model that includes both individual specific fixed effects and all of the time-varying predictors. The time-varying portion of the model defines the variation in predicted 'time-varying' health expenditures, while the individual 'fixed effects' capture the variation of health expenditure that is intrinsic (time invariant) to a person. Standard fixed effects estimation provides a decomposition of these two components. In step two, we regress each of these expenditure components on time-invariant predictors.

We find that time-varying factors and fixed effects each explain about half of the variation in individual health expenditures. Acute health shocks that increase spending in any year other than the prediction year (e.g., a broken leg in one year of the sample) increase a person's fixed effect. Since the costs of that broken leg generally are lower the following year, the cost coefficient on the broken leg becomes negative, rather than positive, contrary to usual results. Large fixed effects are associated with age, especially older males, disability, poor health, smoking and chronic conditions, especially cancer, heart disease, stroke and obesity. We do not find strong evidence that high fixed effects are systematically related to the consistent use of a GP nor with the fee setting behaviour of a regular GP, even though GPs are free to set their own fees in Australia. In contrast, large time-varying expenditures are associated with foreign language, low education and low income. Older individuals, obese individuals, those who assessed themselves to be in poor health and those



DOI: 10.1002/hec

with diabetes, depression, breast/prostate cancer and hypertension have both high fixed and time-varying expenditures. These individuals may be characterised as experiencing further health deterioration from an already poor health state and they rely heavily on the hospital sector.

## 2. MODEL

In our prospective model, the dependent variable is next year's expenditure. Let $y_{it}$ be the health expenditure of individual $i = 1, \ldots, n$ in survey year $t = 1, \ldots, T$. Our baseline specification is the following linear model estimated by ordinary least squares (OLS):

$$y_{it} = \beta_0 + x_i'\beta_1 + z_{it-1}'\beta_2 + u_{it}, \tag{1}$$

where $x_i$ is a vector of time invariant (or slow changing) covariates (e.g. background and socioeconomic characteristics, location, GP characteristics); $z_{it}$ is a vector of time-varying covariates, including chronic conditions, various pharmaceuticals and variables capturing medical service use; $\beta_k$ represents conformable parameter vectors to be estimated; and $u_{it}$ is the error term.

Because health expenditures are highly skewed and often include zero observations, reliance on a linear model estimated by OLS may seem inappropriate. There is a large health econometrics literature that is devoted to comparisons of alternative approaches and models designed to deal with these key features of health expenditures; see the survey by Jones (2011) and references therein. There are two broad conclusions that emerge from this research: (i) the best approach is very case specific with no dominant method recommended, but (ii) the linear model estimated by OLS is very competitive. As Jones (2011) concludes:

> *'It is notable that the simple linear model, estimated by OLS, performs quite well across all of the criteria, a finding that has been reinforced for larger datasets than the one used here.'*

Within this context, we proceed by first undertaking a preliminary comparison of methods in order to check whether our base model is in fact a good choice in our application. We compare OLS to a range of more flexible models reviewed in Buntin and Zaslavsky (2004), including log-transformed models and generalised linear models (GLM) assuming gamma and negative binomial distributions, in terms of forecast errors. To account for zero expenditure, we also estimated the two-part version of these models, with a logit model in the first part to estimate the probability of having a positive expenditure. However, perhaps because only 3% of the sample has zero expenditure, the improvement from two-part models was very marginal. We find that OLS is far superior to the other models, having the smallest mean squared and absolute forecast errors, both in and out of sample. Given the large sample size, the out-of-sample performances from all models are similar to their in-sample performances. The log transformation substantially reduces the weight on observations with very high health expenditure, but when retransformed to level (dollar scale), the log model severely overestimates the top expenditure tail; we allowed for varying scales by age groups, number of diagnoses and deciles of predicted expenditure. By taking account the high variance of health expenditure, the GLMs also overestimate the upper tail.

Figure 1 summarises our exercise on model comparison by plotting the relationship between predicted and actual expenditures by the decile of predicted expenditure. OLS predictions are very close to the 45 degree line, indicating that it correctly identifies groups of individuals with low and high expenditures, conditional on covariates (discussed below), and on average correctly predicts expenditures for these groups. In contrast, the predictions from log models and GLMs are less accurate, and for the top decile, they predict way out of the scale. Trimming the data improves fit but creates unnecessary bias by removing individuals who require truly expensive procedures; for example, trimming removes a third of cardiorespiratory arrest cases. Thus, our preferred model is linear model in levels on the untrimmed sample.



DOI: 10.1002/hec

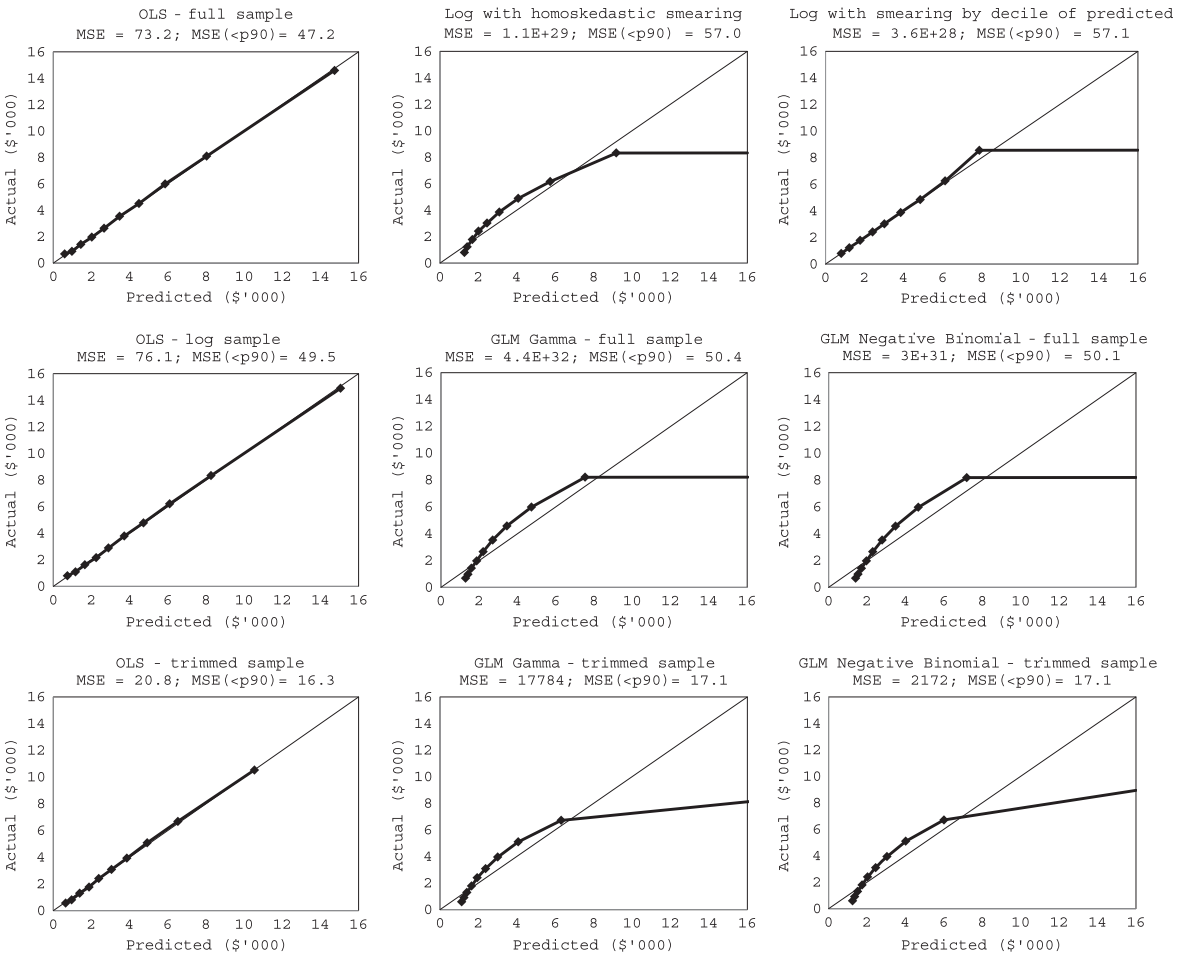


Figure 1. Mean forecast errors by decile of predicted expenditure

To account for the panel nature of our health expenditure data, we estimate linear fixed effect models. Most previous studies that model health care expenditures, such as Buntin and Zaslavsky (2004) and Manning *et al.* (2005), rely on cross-section data, which makes inference difficult because, with cross-section data, it is very hard to control for individual unobservables that may influence health expenditure. Specifically, $u_{it}$ is unlikely to be truly random, for example, due to 'frail' individuals who will always have higher expenditure than others because they are more prone to illness, or individuals who always go to the same doctor who sets high fees. In Equation (1), these systematic variations are captured in the disturbance term and, to the extent they are correlated with the existing set of covariates, will result in biased and inconsistent estimates of the $\beta$'s.

The standard assumption used in the fixed effects model is that $u_{it}$ can be decomposed into individual-specific effects $\alpha_i$ and a random component $\varepsilon_{it}$ as follows:

$$u_{it} = \alpha_i + \varepsilon_{it}. \tag{2}$$

Under this specification, we are able to consistently estimate $\beta_2$, the coefficients of the time-varying parameters, irrespective of any correlation between the covariates included in Equation (1) and time invariant omitted variables. The fixed effects estimator can be thought of as applying OLS to the within transformed



DOI: 10.1002/hec

model (all variables are expressed in terms of deviations from their sample means calculated over time for each individual) or, equivalently, to a model that includes individual specific constants for all $n$ individuals in the data. Since $x_i$ do not vary within an individual, $\beta_1$ will not be identified. However, it is possible to recover estimates for the individual effects. Implementing this procedure, the individual effects are restricted to sum to zero or equivalently are parameterised as deviations from their overall mean, which appears as the estimate of $\beta_0$ in the fixed effects estimation. Therefore,

$$\hat{\alpha}_i = \bar{y}_i - \hat{\beta}_0 - \bar{z}_i'\hat{\beta}_2 \tag{3}$$

where $\hat{\beta}_0$ and $\hat{\beta}_2$ are fixed effects estimates and $\bar{y}_i$ and $\bar{z}_i$ are sample means averaged over the time series observations for each individual.

After estimation, predictions of individual expenditures using the fixed effects results will comprise two components. The first component we term 'time-varying' (TV) expenditure because its predictors vary over time and is defined by

$$\widetilde{y}_{it} = z_{it-1}'\hat{\beta}_2. \tag{4}$$

$\widetilde{y}_{it}$ may be interpreted as changes in expenditure arising from health shocks or other information that changes over our three-year panel. The second component, 'fixed effects' (FE) expenditure, is defined for each individual as $\hat{\alpha}_i + \hat{\beta}_0$ and captures all time invariant effects, including socioeconomic variables, access and provider effects, chronic health problems and long-term effects of preventative care. The two predicted components of expenditure relate to the observed data in the following way:

$$y_{it} = \widetilde{y}_{it} + (\hat{\alpha}_i + \hat{\beta}_0) + \hat{\varepsilon}_{it}. \tag{5}$$

In order to identify the impact of time invariant variables $x_i$ that may contain a considerable amount of information on individuals, we conduct a second stage analysis using the two predicted components of expenditures as the outcome variables in the following auxiliary regressions:

$$\widetilde{y}_{it} = \delta_0 + x_i'\delta_1 + v_{it}, \tag{6}$$

$$(\hat{\alpha}_i + \hat{\beta}_0) = \theta_0 + x_i'\theta_1 + \varpi_i. \tag{7}$$

$x_i$ includes the typical background and socioeconomic characteristics of individuals, but also, we make use of the information on provider identifier and out-of-pocket costs to construct a measure of the strength of the relationship between patient and GP as well as the GP's fee setting behaviour. $x_i$ does not include any element of $z_{it}$ used in the first stage regression. This second set of variables allows us to test the hypothesis that systematically high expenditures could be due to high cost consultations due to high fees or multiple (duplicate) diagnostic tests.

In Equation (6), we seek to explain TV expenditures. Because there is a separate prediction for each time period for each individual, there is a choice as to how this estimation proceeds. In what follows, Equation (6) is estimated only at the survey year in order to better match survey characteristics and time-varying expenditures. Recall that the survey variables are collected only once.

Analogously, Equation (7) provides evidence on the factors associated with FE expenditures. More formally, this procedure produces Hausman and Taylor (1981) estimates of $\beta_1$, in the original specification (Equation (1)). They result from making a stronger assumption about the omitted unobservables, namely, that any correlation between omitted variables is confined to the time-varying variables and are uncorrelated with the



DOI: 10.1002/hec

time invariant variables. Under these conditions, the estimates of $\theta_1$ obtained from Equation (7) are consistent estimates of $\beta_1$ in Equation (1).

We exploit the combination of predicted time-varying and fixed effect expenditures in one further way. Observations are categorised into four mutually exclusive types, namely, observations with (i) large time-varying and large fixed effects (type I); (ii) large time-varying and small fixed effects (type II); (iii) small time-varying and large fixed effects (type III); and (iv) small time-varying and small fixed effects (type IV). For each type, we examine the mix of health services use.

Finally, we discuss the usefulness of survey variables in explaining health expenditure by comparing the predictive power of models with them included as predictors and models with only health shocks as predictors. This exercise informs the reliability of predictions in the absence of panel data or data linkage (i.e. only claims data is available).

## 3. DATA

Our sample is derived from the 45 and Up Study of 267 188 New South Wales (NSW) residents aged 45 and over (45 and Up study collaborators, 2008). Respondents were randomly selected from the Medicare Australia enrolment database, and the survey was collected once between 2006 and 2009. This variation in survey years is part of the data collection design and is not a choice of respondents.[2] The largest collection took place in 2008 (78.9%), with 13.6% in 2006 and 7.5% in 2007. While we would have preferred a more even distribution of survey respondents over time, it is unclear why the survey year would be systematically related to respondents' health over this sample and we place equal weight on all individuals.

New South Wales is the most populous state of Australia. In 2009, it had 7 million residents, or 32% of the total Australian population, and 39% of them are aged 45 and over (45+). Our sample is approximately 10% of this 45+ population. In relation to health expenditure, the 45+ population group incurred 62% of Australia's total health expenditure (AIHW, 2010).[3]

While the survey data is collected only at a point in time, it is linked at the individual level to the following administrative data for each year from 2006 to 2009:

1. NSW Admitted Patient Data Collection, with one record per separation;
2. NSW Emergency Department presentation data, with one record per presentation;
3. Medical Benefits Schedule (MBS) data, with one record per claim; and
4. Pharmaceutical Benefits System (PBS) data, with one record per prescription.

The data linkage to hospital data is performed by the Centre for Health Record Linkage (CHeReL) using a probabilistic matching on first name, surname, date of birth and address.[4] The linkage to the MBS and PBS data is performed by the Sax Institute, which is also the data custodian of the 45 and Up survey. The linked, deidentified data is released under ethics approval.

The first data set is used to price hospital separations. Every separation, in public or private hospital, has an Australian-Refined Diagnostic Related Group (AR-DRG) code. The AR-DRG is a patient classification scheme based on an algorithm of hierarchies of diagnoses and procedures that relates the number and types of patients treated in a hospital to the resources required by the hospital.[5] A cost weight is attached to each AR-DRG measuring its relative cost compared to the average cost of all AR-DRGs; the average cost of all AR-DRGs has a cost weight of 1. Because Australian states manage their own public hospitals, cost weights vary by state. Here,

[2]Details of the survey can be found in http://www.45andup.org.au/

[3]This excludes expenditure non-admitted patients, high-level residential aged care, over-the-counter pharmaceuticals and other health practitioner services.

[4]For details on the linkage procedure and quality, see http://www.cherel.org.au/.

[5]For details on the development of the AR-DRG classification system, see http://www.health.gov.au/internet/main/publishing.nsf/Content/Casemix-1/$File/Final_Report_November_2009.pdf



DOI: 10.1002/hec

we apply the cost weights published in the Costs of Care Standards 2009/10 (NSW Department of Health, 2011). We follow the guidelines in the Standards to adjust the cost weight to the characteristics of each separation. This adjustment depends on hospital type, type of care (overnight, same day, transfer, in mental health unit, nonacute or subacute care units such as rehabilitation), length of stay, ICU hours and the use of ventilation machine. Similarly, for emergency department presentations, the Standards outline variation in emergency department cost by hospital type, triage category (more urgent category is more expensive) and whether the patient is subsequently admitted. Details of the expenditure imputation can be found in Ellis *et al*. (2012).

The MBS and PBS data record the expenditure on subsidised medical services and pharmaceutical items. We aggregate these expenditures on an annual basis and adjust them to constant \$2009 (in December 2009, A\$1 = US\$0.90). Individual annual health expenditure is calculated as the sum of three components: (i) hospital costs (admission and emergency presentation); (ii) charges for out-of-hospital MBS items; and (iii) prices paid to suppliers of out-of-hospital PBS drugs in any given year. We validate these imputed expenditures by their high correlation with the official (AIHW) statistics (0.94).

The hospital and PBS data record diagnoses and drug codes that predict an individual's health expenditure. Each hospital separation record has one primary diagnosis and up to 55 secondary diagnoses. These diagnoses are coded according to the over 25 000 ICD10-AM codes. To help organise this diagnostic information into a more manageable number of diagnosis groups, without loss of information on comorbidities, we use a US-based risk adjustment software called DxCG Risk Solution developed by Verisk Health, which has been extensively applied on US data (e.g., Einav *et al*., 2011; Zhao *et al*., 2005; Ash *et al*., 2001).[6] The software groups diagnoses into 394 'hierarchical condition categories' (HCCs), which group together clinically related diagnoses according to their current and future costs so that cost is primarily driven by the most severe manifestation of a given diagnosis; a new diagnosis that adds a related but less serious medical problem does not increase cost, while unrelated diagnosis contributes cumulatively to cost. From the HCCs, the software offers more aggregated grouping into 117 'related condition categories' (RCCs) and a further 31 'aggregate condition categories' (ACCs). For our purpose, we find that RCCs are preferable to HCCs.[7] The RCCs are not mutually exclusive, e.g. a person with two serious conditions like heart and liver failure will have two RCCs. Similarly, for pharmaceutical information, the software provides a mapping from the most detailed level (seven-digit) Anatomical Therapeutic Chemical (ATC) drug codes into 164 non–over-the-counter RX groups and further combines these into 18 aggregate RX groups (ARX).

Of the original sample, 262 293 respondents (98.2%) are included in our analysis. We exclude respondents who completed the survey outside the data linkage period (2010), have an invalid age, volunteered rather than were randomly selected to participate in the survey and those who died during the study period. Our estimation results are based on a balanced panel. Since the health expenditure data extends for four years, we have 1 049 172 (262 293 $\times$ 4) person-year observations. Given our prospective approach, the first observation year of health expenditure, 2006, is not used in estimation, and the estimation sample size is 786 879 person-years.

Table I reports the summary statistics of health expenditure across years. Less than 3% of the sample has zero expenditure, and spending increases over time. As commonly found in most health expenditure data, spending is skewed to the right with the median much lower than the mean, which is closer to the 75th percentile. The bottom 5% of the sample has less than \$200 in annual health cost while 5% at the top have more than \$15 000 in health expenditure. The coefficient of variation ranges from 2.05 to 2.35, notably lower than in samples from the

[6]The software, which extends the classification system used by the US Medicare program for paying competing health plans, organises ICD-10 diagnosis information into a large number of nonmutually exclusive categories and imposes hierarchies on diseases so that more serious or expensive conditions take precedence over less serious or expensive conditions. Similarly, it groups ATC codes into drug groups based on therapeutic class, active ingredients and doses and strength. The software also performs a number of data-cleaning steps to identify illegal (e.g. coding errors) or invalid (e.g. male pregnancies) diagnoses.

[7]The predictive power gain from using HCCs is small relative to the hundreds of extra parameters to be estimated. Using the RCCs also reduces overfitting in our moderate size sample.



DOI: 10.1002/hec

Table I. Summary statistics of health expenditure (×$1000)

| | 2007 | 2008 | 2009 | All |
|---|---|---|---|---|
| % zero | 0.027 | 0.026 | 0.027 | 0.027 |
| Mean | 4.057 | 4.681 | 5.009 | 4.423 |
| Standard deviation | 8.614 | 9.619 | 11.79 | 9.805 |
| Coefficient of variation | 2.123 | 2.054 | 2.353 | 2.217 |
| Skewness | 9.688 | 8.096 | 11.10 | 10.27 |
| $5^{th}$ percentile | 0.060 | 0.073 | 0.201 | 0.062 |
| $25^{th}$ percentile | 0.565 | 0.661 | 0.652 | 0.598 |
| Median | 1.703 | 1.950 | 1.877 | 1.780 |
| $75^{th}$ percentile | 4.200 | 4.832 | 4.700 | 4.432 |
| 95th percentile | 15.24 | 17.86 | 19.71 | 16.81 |
| *N* | 262 293 | 262 293 | 262 293 | 786 879 |

Note: Sample weighting is used, which reflects the 45+ NSW population by region due to oversampling in regional areas. Values in constant A$2009.

United States (Ash *et al*., 2001), where it is often 3 or more. The skewness measure is also moderate, averaging around 10. These features help explain the higher measures of R-square found in our sample compared with US samples, and may help explain the time pattern of predictiveness we observe below.

Given that we have hundreds of RCC and RX dummy variables, for conciseness, we do not present the full summary statistics for time-varying predictors. Among the most common chronic conditions associated with hospital admission are benign neoplasm, hypertension and gastrointestinal conditions. The most common pharmaceuticals are lipid lowering agents, anti-infectives and ulcer/GERD medications. To measure medical services utilisation, we use the total number of GP and specialist consultations in a year. Specifically, we use a specific MBS item 23, which is for consultation (not for immunisation) less than 20 min by a GP, and item 104 for initial consultation by a specialist excluding an ophthalmologist. On average, individuals have 6.7 GP consultations and 0.6 specialist consultation in a year.

Table II provides the summary statistics of survey variables included in Equations (6) and (7). About 40% of respondents are over 65 years. The very old are well represented in the sample with about 11% over 80 years. There are slightly more female than male respondents, and the majority of respondents are married with tertiary education. Country of birth, foreign language (not English) spoken at home and skin colour are included to capture individual background conditions that may affect health care utilisation. Most respondents are either still working full-time or fully retired. Private health insurance coverage is 65%, higher than the 45+ population average of 55%. In addition to demographics and economic variables, the 45 and Up survey also asked its respondents to self-report lifestyle and ever diagnosed chronic conditions and to provide a self-assessment of their health in general. The self-assessed general health suggests that most respondents are in good health or better. However, the incidence of chronic conditions, which may be regarded as a more objective measure of health, signals health care need. Survey responses indicate that over 30% of the sample report high blood pressure, 28% report skin cancer, 11% report gender-specific or other cancer, 12% report heart diseases and 9% report diabetes.

The presence of a regular GP may influence charging and expenditure. To capture the patient–doctor relationship, we use an indicator variable that takes a value of 1 if the individual has a regular GP who is consulted most of the time. We define an individual as having a regular GP if 85% or more of standard consultations in a given year are from the same GP; for individuals with fewer than seven (the sample mean) consultations a year, having a regular GP is defined as receiving consultations 100% of the time

[8]We do not construct these measures for specialists because patients may go to different specialists for different specialty; furthermore, 88% of patients only have 2 specialist initial consultations.



DOI: 10.1002/hec

Table II. Summary statistics of survey-based and GP variables

| Demographic | Mean | Demographic | Mean | Economic | Mean | Self-reported chronic conditions | Mean | Lifestyle | Mean |
|---|---|---|---|---|---|---|---|---|---|
| Age: 45–19 | 0.137 | Foreign born | 0.121 | Income: <$20K | 0.189 | SAH: excellent | 0.147 | BMI: underweight | 0.013 |
| Age: 50–54 | 0.163 | Foreign language | 0.278 | Income: $20K–<$30K | 0.087 | SAH: very good | 0.352 | BMI: normal | 0.344 |
| Age: 55–59 | 0.170 | Skin: very fair | 0.157 | Income: $30K–<$40K | 0.074 | SAH: good | 0.326 | BMI: overweight | 0.363 |
| Age: 60–64 | 0.146 | Skin: fair | 0.543 | Income: $40K–<$50K | 0.070 | SAH: fair | 0.118 | BMI: obese (I) | 0.146 |
| Age: 65–69 | 0.119 | Skin: light olive | 0.249 | Income: $50K–<$70K | 0.102 | SAH: poor | 0.021 | BMI: obese (II) | 0.042 |
| Age: 70–74 | 0.087 | Skin: dark olive | 0.016 | Income: ≥$70K | 0.254 | SAH: missing | 0.037 | BMI: obese (III) | 0.017 |
| Age: 75–79 | 0.065 | Skin: brown | 0.024 | Income: missing | 0.223 | High blood pressure | 0.356 | BMI: missing | 0.075 |
| Age: 80+ | 0.113 | Skin: black | 0.001 | Full time | 0.352 | Skin cancer | 0.277 | Ever smoke | 0.422 |
| Male | 0.467 | Skin: missing | 0.010 | Part-time/other work | 0.143 | Breast/prostate cancer | 0.059 | Alcohol: none | 0.350 |
| Married | 0.684 | *Region* | | Fully retired | 0.364 | Other cancer | 0.062 | Alcohol: low/moderate | 0.490 |
| Never married | 0.066 | Remote | 0.094 | Disabled | 0.039 | Heart disease | 0.120 | Alcohol: risky (>2/day) | 0.137 |
| Widowed | 0.090 | Outer region | 0.079 | Not in labour force | 0.100 | Stroke | 0.031 | Alcohol: missing | 0.023 |
| Divorced | 0.073 | Inner region | 0.156 | PHI with extra | 0.515 | Diabetes | 0.091 | GP | |
| Separated | 0.028 | Major city | 0.671 | PHI without extra | 0.142 | Asthma | 0.024 | Regular GP | 0.444 |
| Unknown | 0.006 | | | No PHI | 0.343 | Depression | 0.126 | Mean OOP if has regular GP | 1.822 (5.410) |
| Partner | 0.053 | | | Health card | 0.281 | Broken bone | 0.117 | | |
| High school | 0.132 | | | | | | | Mean OOP if has no regular GP | 4.211 (7.584) |
| Certificate | 0.314 | | | | | | | | |
| Trade/ diploma | 0.310 | | | | | | | | |
| University | 0.244 | | | | | | | | |

Note: Sample weighting is used, which reflects the 45+ NSW population by region due to oversampling in regional areas. Figures are sample proportions, except mean out-of-pocket (OOP), regular general practitioner (GP) and other GP which are continuous variables. Standard deviations are provided in parentheses. BMI stands for body mass index, PHI stands for private health insurance, and SAH stands for self-assessed health.





DOI: 10.1002/hec

from the same GP. For the 10% of cases in which there is only one GP consultation in a given year, we assume that this individual does not have a regular GP. The sample mean for having a regular GP is 44%. To measure GP fee-setting behaviour, we use the average out-of-pocket for standard consultations.[8] We make a distinction between the fees of regular GPs and other GPs. When an individual has a regular GP, the fee of that GP is included. For those without a regular GP, the average fee across all consulted GPs is included.

## 4. RESULTS

### 4.1. Fixed effects regression results

We estimate a fixed effects linear regression model based on the specification in Equation (1). We find significant individual heterogeneity, as indicated by strong rejection of the null hypothesis that all $\alpha_i = 0$ ($p$-value $< 0.0005$). The standard Hausman test for comparing fixed effects and random effects models, where the later acknowledges individual heterogeneity but assumes zero correlation between them and covariates, also favours the use of the fixed effects model ($\chi^2_{233} = 164\ 897$, $p$-value $< 0.0005$). The variance of the composite error term in Equation (2) is given by $var(u_{it}) = var(\alpha_i) + var(\varepsilon_{it}) = \sigma^2_\alpha + \sigma^2_\varepsilon$, and their estimates are $\hat{\sigma}^2_\alpha = 52.49$ and $\hat{\sigma}^2_\varepsilon = 59.9$. The ratio $\hat{\sigma}^2_\alpha / \left(\hat{\sigma}^2_\alpha + \hat{\sigma}^2_\varepsilon\right) = 0.47$ indicates that 47% of the variance in health expenditure can be attributed to fixed effects (i.e. variations from the between estimator across individuals); this may also be interpreted as the intraclass correlation coefficient, measuring the correlation between two observations for the same individual.

Figure 2 shows the distribution of (clockwise from top left): observed expenditure ($y_{it}$), predicted expenditure $(\widetilde{y}_{it} + \hat{\alpha}_i + \hat{\beta}_0)$, predicted TV expenditure $(\widetilde{y}_{it})$ and predicted FE expenditure $(\hat{\alpha}_i + \hat{\beta}_0)$. Nothing in the nature of

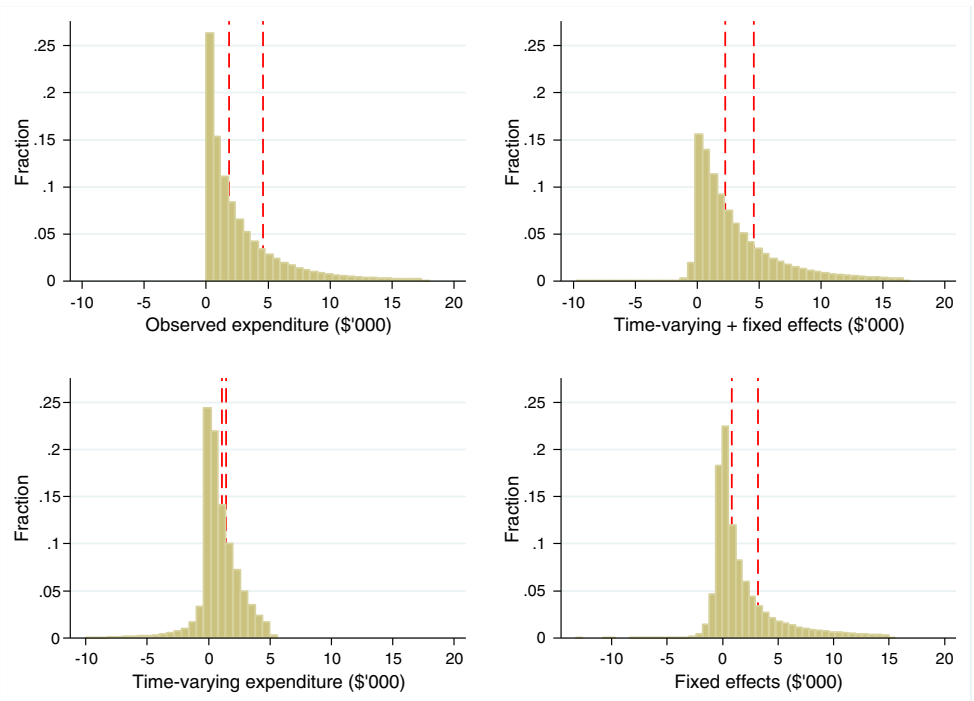


Figure 2. Distribution of observed and predicted health expenditure



DOI: 10.1002/hec

the linear model restricts predictions to be positive, and a few negative predictions (4.9%) appear. Overall predicted expenditure exhibits wider variation than the TV expenditure component with the difference driven by considerable individual heterogeneity. The mean and standard deviation of predicted expenditure are $4423 and $7500, respectively, while the mean of TV expenditure is $1376 and the standard deviation is $2357. Half of the sample is estimated to have TV expenditure between $261 (25th percentile) and $2274 (75th percentile), and in the upper tail of the distribution, 5% of the sample has TV expenditure above $5013 (95th percentile).

A quarter of individuals have negative FE expenditure. The mean is $3047, but the distribution is skewed to the right, indicating the presence of individuals with very large positive fixed expenditures. It can also be seen that the shape of FE distribution drives the shape of the overall predicted expenditures.

Table III summarises the fixed effects regression results. To aid interpretation, instead of presenting the regression coefficients, which would be overwhelming due to the large number of diagnosis and drugs categories, we report the mean TV and FE expenditures for aggregate diagnosis and drugs groups, ACCs and ARXs. Note that the estimation is based on the more detailed groups, RCCs and RXs. The mean expenditures for an ACC (ARX) are given by the expenditures on observations with at least one RCC (RX) within it. For GP and specialist consultations, we present the mean TV and FE expenditures separately for cases below and above the mean consultation level. We also report the corresponding observed expenditure and the number of relevant cases for each aggregate group.

Table 3 presents the expenditure categories for diagnosis and drug groups in ascending order of TV expenditure. We find that ACCs with the smallest (largest negative) TV expenditure tend to be those with the highest FE expenditure. 22 out of 30 ACCs have negative TV expenditure, suggesting that the expenditure impacts of these diagnoses are transitory; diagnosis and treatment this year are not replicated in the following year's expenditure. For example, while those experiencing cardiorespiratory arrest have high FE expenditure reflecting the chronic nature of heart conditions, they have a large negative TV expenditure because the chance of recurrence in the next year is low. Similarly, transplants, complications, infections and injuries are associated with negative TV expenditure and high FE expenditure, as these tend to be one-off health events affecting individuals with already high health expenditures. On the other hand, the health of individuals with diabetes, cancer, stomach disease and eye problems tends to worsen in the subsequent year, reflected in positive TV expenditures. For example, the FE expenditure for diabetes is over $12 000 per year, and people with diabetes tend to see their health expenditures increase by 10% in the subsequent year.

In contrast to diagnoses, the mean TV and FE expenditures for drug categories are all positive; positive FE expenditure suggests that those taking prescription drugs tend to be high users of health services. The mean of TV expenditure ranges between $1555 and $3332 across the ARX groups, which is much narrower than in the case for diagnoses. The top three ARXs with the highest TV expenditure involve small number of cases, followed by diabetes drugs.

The results for GP and specialist consultations are split by frequency, with low frequency associated with low TV and FE expenditures. However, the high frequency cases have very large mean TV expenditure, almost as large as the mean FE expenditure. This finding suggests that a high frequency of visits to a doctor provides a strong indication of much more intensive health treatments, possibly hospitalisation, in the following year. Indeed, the sizes of the mean TV expenditure for high frequency GP and specialist consultations are larger than the mean TV expenditure for any ACC and most ARXs. This is a quite striking result that has not been previously noted. A typical health expenditure model focuses on diagnoses and drugs and neglects doctor consultations.

We now examine the interaction between TV and FE expenditures to further inform the type of health services that are associated with these expenditures. Each individual has three TV expenditures corresponding to each of the three years of data, but the FE expenditure is constant across years. In anticipation of our second-stage analysis in the next subsection, which correlates predicted expenditures with survey information, we use only TV expenditure in the survey year to minimise measurement error due to changes in survey variables.

Figure 3A shows the sample density over the TV–FE expenditures plane. There are four high density regions, suggesting the presence of four types of individuals by high or low TV and FE expenditures. We find



DOI: 10.1002/hec

Table III. Mean TV, FE and observed expenditures by aggregate group

| | TV | FE | Observed | Cases | | TV | FE | Observed | Cases |
|---|---|---|---|---|---|---|---|---|---|
| *Diagnosis groups (ACCs)* | | | | | Obstetric | $657 | $3742 | $4458 | 115 |
| Cardiorespiratory arrest | –$10 101 | $27 791 | $17 804 | 1084 | Diabetes | $830 | $12 133 | $12 962 | 18 417 |
| Transplants, openings, other V-codes | –$6046 | $22 536 | $16 527 | 2189 | Ophthalmic | $1087 | $8294 | $9394 | 19 992 |
| Cognitive disorders | –$4805 | $19 800 | $14 994 | 2724 | Benign/In situ/Uncertain neoplasm | $1090 | $6258 | $7342 | 31 757 |
| Developmental disability | –$4672 | $18 153 | $13 481 | 101 | *Drugs (subset of all RX categories)* | | | | |
| Complications of care | –$4145 | $17 834 | $13 756 | 10 939 | Coagulants and anticoagulants | $1555 | $8129 | $9674 | 91 686 |
| Psychiatric | –$2550 | $17 156 | $14 393 | 4425 | Biologicals | $1814 | $7488 | $9312 | 46 922 |
| Infectious and parasitic | –$2372 | $16 129 | $13 756 | 11 117 | Neurological agents | $1885 | $5985 | $7876 | 140 841 |
| Vascular | –$2256 | $16 319 | $14 033 | 9751 | Antihyperlipidemics | $1965 | $4950 | $6914 | 221 904 |
| Hematological | –$2174 | $17 898 | $15 725 | 8530 | Anti-infectives | $2018 | $5535 | $7556 | 215 536 |
| Pulmonary | –$2090 | $16 662 | $14 561 | 11 588 | GI drugs | $2023 | $5577 | $7595 | 205 794 |
| Cerebrovascular | –$2057 | $15 332 | $13 134 | 4069 | Nutritionals | $2160 | $9139 | $11 280 | 35 009 |
| Injury, poisoning | –$1438 | $12 575 | $11 129 | 8634 | Cardiovascular | $2183 | $5219 | $7396 | 258 705 |
| Nutritional and metabolic | –$1321 | $14 174 | $12 857 | 19 339 | Pulmonary drugs | $2236 | $5251 | $7474 | 90 604 |
| Hepatobiliary | –$1270 | $10 396 | $9081 | 4548 | Analgesics/anti-inflammatories | $2241 | $5519 | $7744 | 198 081 |
| Substance abuse | –$1262 | $15 427 | $14 176 | 2273 | Endocrine/metabolic agents | $2306 | $5297 | $7598 | 129 786 |
| Musculoskeletal and connective tissue | –$871 | $10 311 | $9428 | 35 910 | Dermatologicals | $2453 | $5525 | $7979 | 77 294 |
| Dermatologic | –$706 | $12 236 | $11 513 | 9655 | Additional groups | $2522 | $6648 | $9153 | 25 922 |
| Neurological | –$659 | $12 334 | $11 687 | 9706 | EENT preparations | $2652 | $5412 | $8071 | 111 456 |
| Urinary | –$485 | $14 286 | $13 784 | 18 671 | Diabetes drugs | $2867 | $6624 | $9497 | 46 515 |
| Cardiovascular | –$334 | $12 434 | $12 156 | 41 253 | Genitourinary agents | $2978 | $8590 | $11 568 | 5854 |
| Screening/History | –$96 | $9599 | $9523 | 77 586 | Upper respiratory agents | $3057 | $6378 | $9441 | 5334 |
| Symptoms, signs and ill-defined conditions | –$87 | $10 118 | $10 030 | 46 277 | Immunologic agents | $3332 | $10 023 | $13 367 | 6449 |
| Ears, nose, and throat | $153 | $8900 | $9064 | 7356 | Number of GP consultations (<7) | $688 | $2254 | $2885 | 511 280 |
| Genital | $303 | $6142 | $6446 | 13 546 | Number of GP consultations (≥7) | $2652 | $4519 | $7277 | 275 599 |
| Malignant neoplasm | $572 | $14 054 | $14 628 | 11 810 | Number of specialist consultation (≤1) | $933 | $2711 | $3645 | 670 362 |
| Gastrointestinal | $619 | $7217 | $7849 | 60 592 | Number of specialist consultation (>1) | $3924 | $4980 | $8899 | 116 517 |

Note: FE and TV expenditures are based on fixed effects linear regression on a sample of 786 879 person-years (cases). There are only 30 aggregate diagnosis groups that are relevant to this population group (i.e. no neonatal related diagnoses). The sum of the TV and FE dollars do not sum to exactly the average observed spending for each diagnostic group because the diagnosis groups shown, aggregated condition categories (ACCs), are more aggregated than the related condition categories (RCCs) that were actually used in the regressions.





DOI: 10.1002/hec

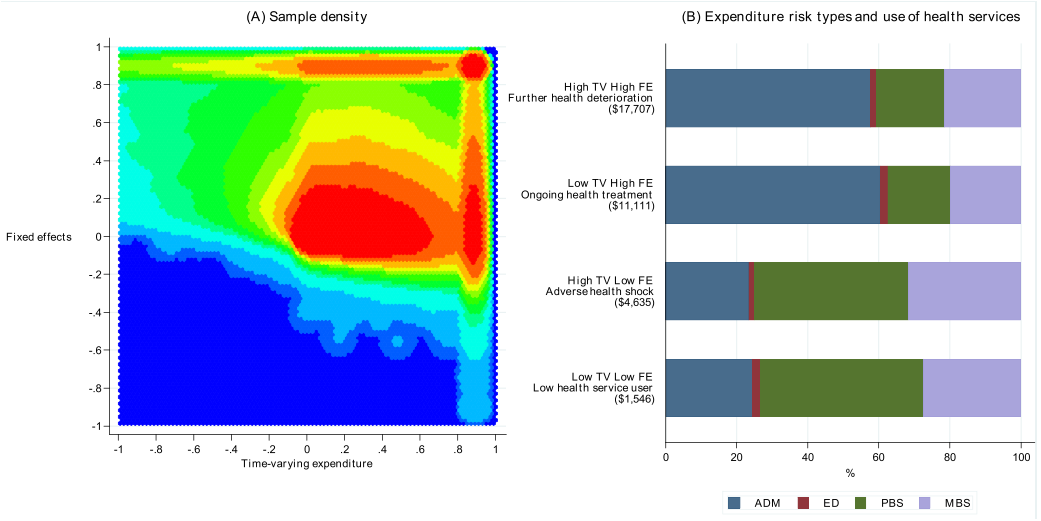


Figure 3. Sample density, expenditure risks and use of health care services

that these four regions are well captured by defining a high expenditure as expenditure in the top 20%. We classify individuals with high TV and high FE expenditures as type I, representing 'further health deterioration'. Type II has low TV and high FE expenditures representing individuals needing 'ongoing treatment'. Type III has high TV and low FE expenditures representing those with 'adverse health shocks', and finally, type IV has low TV and low FE expenditures representing those with 'low use of health services'. Type I has the largest average expenditure ($17 707), and they represent 6.4% of the sample. Type II has the next largest average expenditure of $11 111 for 13.6% of the sample. Type III makes up 13.6% of the sample with $4635 in average expenditure, and type IV makes up the remaining 66.4% with $1546.

Figure 3B shows the mix of health services used by each type. Types I and II's expenditures are high, driven by hospital admission costs. Type II spends the largest share of their expenditure on inpatient costs. This picture is consistent with Table III, where those with at least one hospital episode (i.e. belong to one of the ACCs) have high FE expenditures. These two types have persistently high expenditure, and we can expect heavy reliance on the hospital sector. In contrast, types III and IV's expenditure comprise mainly of out-of-hospital drugs (PBS) and medical services (MBS). Type III may add to the hospital sector burden next year as their health deteriorates.

### 4.2. Variations in components of health expenditure by survey variables

Table IV presents the second-stage estimation results for Equations (6) and (7). We first discuss the results for Equation (6) given in the first column. TV expenditure is higher for older individuals, females, those who were born overseas, speak a foreign language at home, have a health card (which makes a person eligible for public fee subsidies based on a means test), have low education and income, are not in full time employment and live in major cities. Self-assessed health is strongly predictive of TV expenditure. Reported conditions that are positively associated with TV expenditure are hypertension, skin cancer, breast/prostate cancer and diabetes. Lifestyle factor that adds to TV expenditure is obesity. Having a regular GP is associated with higher TV

[9]Our results contrast with Newhouse *et al*. (1989) which finds that about 20-30% of total variation is explained by individual level fixed effects, whereas in our model fixed effects explain about 56% of total variation. As explained below, socio-demographic factors, utilisation, pharmacy, and health conditions explain about half of the total FE in our data, while time invariant survey factors explain about 15% of the FE.



DOI: 10.1002/hec

Table IV. OLS regression of fixed effects and time-varying expenditure

| Dependent variable | TV (×$1000) | FE (×$1000) | | TV (×$1000) | FE (×$1000) | | TV (×$1000) | FE (×$1000) |
|---|---|---|---|---|---|---|---|---|
| Age 50–54 | 0.0612** | 0.0957 | Skin: very fair | 0.008 | 0.0308 | High blood pressure | 0.3761*** | 0.0614* |
| Age 55–59 | 0.1185*** | 0.2403*** | Skin: light olive | –0.0012 | –0.026 | Skin cancer | 0.2434*** | –0.1807*** |
| Age 60–64 | 0.2636*** | 0.6651*** | Skin: dark olive | 0.0623 | –0.0793 | Breast/prostate cancer | 0.0952*** | 1.8827*** |
| Age 65–69 | 0.5694*** | 0.9733*** | Skin: brown | 0.1024** | –0.2197* | Other cancer | –0.0813*** | 2.8643*** |
| Age 70–74 | 0.8105*** | 1.3310*** | Skin: black | 0.0429 | –0.5257 | Heart disease | –0.1998*** | 2.2547*** |
| Age 75–79 | 0.9615*** | 1.9981*** | Skin: miss | 0.1151* | 0.2217 | Stroke | –0.2527*** | 1.7458*** |
| Age 80+ | 0.6784*** | 2.4422*** | Health card | 0.1367*** | –0.0517 | Diabetes | 0.5905*** | 0.8371*** |
| Male | –0.3615*** | 0.0334 | No PHI | –0.0978*** | –0.8832*** | Asthma | 0.571 | 1.0159 |
| Age 50–54 * male | –0.0199 | 0.057 | PHI without extra | –0.0301* | –0.3310*** | Depression | –0.0037 | 0.4829*** |
| Age 55–59 * male | 0.0209 | 0.2732** | Income: <$20K | 0.2491*** | –0.2250*** | Broken bone | –0.0471** | 0.9704*** |
| Age 60–64 * male | –0.0265 | 0.2419* | Income: $20K–<$30K | 0.1561*** | –0.1769** | BMI: under | –0.1519*** | 0.5237*** |
| Age 65–69 * male | –0.0567 | 0.7636*** | Income: $30K–<$40K | 0.0770*** | –0.1415* | BMI: over | 0.0430*** | 0.0283 |
| Age 70–74 * male | –0.0366 | 1.0569*** | Income: $40K–<$50K | 0.0550** | –0.0943 | BMI: obese (30–<35) | 0.0722*** | 0.1543*** |
| Age 75–79 * male | 0.0276 | 1.0896*** | Income: $50K–<$70K | –0.0037 | –0.0504 | BMI: obese (35–<40) | 0.0706** | 0.2845*** |
| Age 80+ * male | 0.0419 | 1.1881*** | Income missing | 0.1368*** | 0.1067* | BMI: obese (40+) | 0.1641*** | 0.4803*** |
| Never married | –0.1278*** | 0.1775** | Part-time/other work | 0.0338* | 0.1298** | BMI: missing | 0.0456* | 0.0996 |
| Widowed | –0.2097*** | 0.5408*** | Fully retired | 0.1844*** | 0.4457*** | Ever smoked | 0.0065 | 0.3379*** |
| Divorced | –0.0616** | 0.1692** | Disabled | 0.1192*** | 3.2773*** | Heavy drinking | –0.029 | –0.1799*** |
| Separated | –0.0898** | 0.0652 | Not in LF | 0.034 | 0.2156*** | No drink | 0.0428*** | 0.2701*** |
| Unknown | –0.1763** | 0.1068 | Remote | –0.3219*** | –0.5767*** | Drink: missing | 0.0985** | 0.1165 |
| Partner | –0.0422 | –0.0015 | Outer region | –0.3803*** | –0.5505*** | Regular GP | 0.0240* | –0.0660* |
| Foreign language | 0.2124*** | –0.4579*** | Inner region | –0.2227*** | –0.3238*** | Mean OOP no regular GP | 0.0065*** | 0.0080*** |
| Foreign born | 0.0302* | –0.2131*** | SAH: very good | 0.0926*** | 0.2775*** | Mean OOP regular GP | –0.0005 | 0.0091*** |
| Certificate | –0.1450*** | 0.0952* | SAH: good | 0.2543*** | 0.9810*** | Year = 2008 | –0.8211*** | –0.3102*** |
| Trade/diploma | –0.1721*** | 0.0374 | SAH: fair | 0.4507*** | 2.7932*** | Year = 2009 | –1.0073*** | –0.2825 |
| University | –0.2288*** | 0.1155* | SAH: poor | 0.2942*** | 6.1148*** | Constant | 1.6188*** | 0.5150*** |
| | | | SAH: unknown | 0.3677*** | 1.8385*** | *N* | 226 823 | 226 823 |
| | | | | | | R-square | 0.0891 | 0.1481 |

Note: FE stands for fixed effects, and TV stands for time-varying expenditures. The number of respondents is fewer than the full sample of 262 293 respondents because some respondents were surveyed in 2006.
*Statistical significance at 5%.
**Statistical significance at 1%.
***Statistical significance at 0.01%.

expenditure, but their fees do not add to TV expenditure. In contrast, for those without a regular GP, for each $1 increase in the average out of pocket, TV expenditure increases by $6.5.

The second column reports the results for Equation (7). Some covariates have opposite effects on TV and FE expenditures. About 15% of variation in FE expenditure can be explained by time invariant survey measures.[9] Those with large FE expenditure tend to be older especially older males; never been married, widowed or divorced; and not in full time employment or out of labour force due to disability. Private health insurance coverage tends to increase FE expenditure, which is consistent with insurees' moral hazard leading to overutilisation of health services. High income is also associated with larger FE expenditure, which may capture taste in investment in health or health technology among the high income individuals. Individuals who regard themselves as being in poor health have more than $6000 more in FE expenditure than an average person. This supports the use of self-assessed health, available in most data sets, as a measure of innate health. Individuals with a heart disease have about $2200 in health expenditure every year, breast/prostate cancer or stroke costs about $1800, other form of cancer costs $2800 and diabetes costs $840. Unhealthy lifestyles are also associated with higher FE expenditure, in addition to the existing chronic condition. The morbidly obese have an extra $285–$480 on top of any cost of chronic conditions, and those who ever smoke has $340 extra. Being underweight and alcohol abstinence are associated with higher FE expenditure, perhaps reflecting consequences of chronic conditions. Those with a regular GP tend to have lower FE expenditure, which is consistent with the cost-saving hypothesis from duplications of diagnostics and tests, and may also suggest direct benefits from building a relationship with the GP; for example, the GP may give health/lifestyle advice more readily and conduct regular check-ups. GP fees increase FE expenditure. This is consistent with the hypothesis that some individuals tend to go to high charging doctors. Interestingly, the size of the effect is larger for regular GP. Patient–doctor interaction may increase other information about the patients to the GP, for instance their income.

Reconciling these results with the types of individuals, survey variables predict that type I tends to be older individuals, those in poor health, obese individuals, and those with breast/prostate cancer and diabetes. This is consistent with the diagnosis and drug information in Table III, where treatment for malignant cancer and diabetes and prescription drugs for diabetes have both high TV and high FE expenditures. On the other hand, sufferers of other cancer, heart disease, stroke and broken bones have high FE but lower TV expenditure suggesting that they are more likely to be type II having ongoing treatments. Type III is characterised by foreign born, foreign language and low income. This group may be a policy target given their vulnerability to large swings in their health expenditure due to adverse health shocks.

### 4.3. Model predictiveness

Our study is unusual in having multiple years of administrative data to merge with survey data that was collected in stages, giving variation in survey year. This gives a nice setup for assessing how well survey information gathered in one year improves the predictive power of the diagnosis, drug and medical service information in the same and different years. Since much of the information gathered in a survey varies only slowly from year to year (e.g. education, location, chronic conditions), it is of use to see whether their contribution to total predictiveness is especially large in the same or subsequent year in which the information is gathered.

Table V presents various R-squares from separate OLS regressions for various subsets of our total sample. Each row in the table corresponds to a different year in which the survey was conducted, while each column corresponds to a different year of the diagnosis, drug and medical service information. Recall that in the one-year prospective model, the outcome variable is one year ahead of the predictors from the administrative data. The first notable feature is that the 2008 data is more predictable than either 2007 or 2009, which we believe is due to the lower coefficient of variation and skewness of spending in that year. The second notable feature is that the predictiveness of spending is higher in 2008 than in other years perhaps because of overfitting on the smaller sample surveyed that year. The predictiveness of the survey information for a given year can be



DOI: 10.1002/hec

Table V. Relation between survey year predictiveness and year of health data

| Expenditure year predicted: | 2007 | 2008 | 2009 | All |
|---|---|---|---|---|
| Survey Year: | | | | |
| 2006 | 0.253 | 0.275 | 0.227 | 0.232 |
| 2007 | 0.346 | 0.294 | 0.234 | 0.256 |
| 2008 | 0.268 | 0.294 | 0.236 | 0.255 |
| All | 0.265 | 0.283 | 0.228 | 0.248 |

The table presents R-squares from independent OLS regressions based on Equation (1) for various subsets of our total sample.
The total sample sizes for survey years 2006, 2007 and 2008 are 106 410 (13.6%), 58 164 (7.5%) and 616 680 (78.9%). Each row in the table corresponds to a different year in which the survey was completed and the administrative data on diagnoses, pharmacy and GP and specialist consultations are used, while each column corresponds to a different year of the health expenditure data.

Table VI. Explanatory power of alternative predictors: Adjusted R-square from various OLS models

| Included variables (number of variables) | No FE | With FE |
|---|---|---|
| [None] | | 0.5631 |
| Age + Gender (15) | 0.0466 | 0.5637 |
| RCC (110) | 0.1117 | 0.5753 |
| RCC + RX | 0.1945 | 0.5769 |
| RCC + RX + Utilisation measures | 0.2382 | 0.5850 |
| RCC + RX + Utilisation measures + Survey | 0.2477 | 0.5850 |
| RCC + RX + Utilisation measures + Survey + Year dummies | 0.2493 | 0.5876 |
| RX (121) | 0.1544 | 0.5679 |
| Utilisation measures (2) | 0.1097 | 0.5721 |
| Survey (76) | 0.1020 | 0.5631 |
| Survey + year dummies [using years before survey sample only] | 0.1170 | |
| Survey + year dummies [using survey year sample only] | 0.1266 | |
| Survey + year dummies [using years after survey sample only] | 0.0907 | |

Note: RCC = DxCG related condition categories, RX = DxCG pharmacy groups. Except in the Age + Gender model, age is included as a Survey variable, which is constant at the survey year. In the model with FE, Survey variables are effectively dropped because of perfect collinearity with the person dummy variable. The first model with FE includes only person dummy variables. The sample size ($N = 786\ 879$) is the same for all rows except for the final three, which use only the years before the survey was conducted ($N = 209\ 310$), the year in which the survey was actually conducted ($N = 262\ 823$) and the years after the survey was conducted ($N = 350\ 746$), respectively.

detected by how much higher the R-square is relative to the average for that survey year. Hence, we see that the 2007 survey responses resulted in an above-average predictiveness for 2007 than the average (0.346 versus 0.256), while for the 2008 survey, the 2008 value is also higher than the average (0.294 versus 0.255). Expectedly, predictiveness is the highest along the diagonal entries, where the survey and administrative data years match. While detectable, the gain from using survey from the same year is modest, suggesting that most of the value of adding survey information is in the FE expenditure rather than TV expenditure. This has important implications on how survey information may be useful as further modifiers for plan payment formulas and provider profiling in the context of risk adjustment models, suggesting that long- rather than short-duration information from surveys is more important.

Table VI examines the predictive power of different types of information by comparing adjusted R-squares from various OLS regressions. For comparison purpose, we also report the adjusted R-squares from OLS regressions with individual dummies, which only identify time-varying predictors. The models containing diagnoses (RCC), drugs (RX) and GP and specialist consultations (utilisation measures) each improve on the modest predictive power of a simple model with just age and gender, which are available in nearly all administrative data and widely used in basic payment formulas. Adding the survey and year dummies increases the predictive power only slightly. The next three rows show that using the drug information alone explains 15% of the variation while utilisation measures alone or survey information alone explains only 10% to 11%. What is noteworthy is that the utilisation measures consist of only two variables. The bottom three rows show that



DOI: 10.1002/hec

the survey information does about as well predicting spending in years one or two years before the year of the survey itself, but not as good for predicting in years after the survey. In comparison, the model with fixed effects shows that time varying information adds at most an additional 0.02 percentage points to predictions: 56% of health spending variation is across individuals, and only an additional 2% explained by time-varying diagnoses, drugs and utilisation measures.

## 5. CONCLUSION

This paper exploits the linkage between a very large survey of the "over 45 years" population and a comprehensive set of health administrative records, to predict prospective health care expenditure for each individual. By isolating the expenditure that is time-invariant or intrinsic to an individual (fixed effects), we find that those with high fixed effects tend to be old, sick and engage in unhealthy lifestyles (having a smoking history and/or being obese). There is evidence that having a regular GP lowers fixed effects, but we find no evidence that fixed effects are correlated with GP fee-setting behaviour. In about 6% of cases, we find high time-varying and high fixed effect expenditures, suggesting further deterioration in health. Together with the 14% who are likely to require ongoing treatments, these cases have relied heavily on the hospital sector. Meanwhile, another 14% of cases may be expected to increase reliance on hospital services for having high time-varying expenditure, possibly due to an adverse health shock. Sociodemographic factors that predict this case include low income and foreign background. The remaining individuals are likely to rely heavily on out-of-hospital services.

At a time when many are looking to predictive models to help inform providers and payers, or even to use for payment innovations (Ash and Ellis, 2012), the results have implications even to the United States or Europe, since they suggest that much of the additional predictive power of survey information is from time-invariant or slow moving variables. Occasional rather than annual surveys may add most of the modest incremental predictive power from this source, and help predict subsequent years nearly as well as the survey year.

## ACKNOWLEDGEMENTS

This research is funded by an ARC Discovery Project grant (DP110100729). It uses data from the 45 and Up Study, which is managed by the Sax Institute in collaboration with major partner Cancer Council New South Wales, and partners the Heart Foundation (NSW Division); NSW Ministry of Health; *beyondblue: the national depression initiative*; Ageing, Disability and Home Care, NSW Family and Community Services; and Australian Red Cross Blood Services. This project was undertaken by the University of Technology Sydney and utilised MBS and PBS data supplied by the Department of Human Services and linked to the 45 and Up Study by the Sax Institute. Data linkage for the hospital admission and emergency department presentations used in the project was undertaken by the Centre for Health Record Linkage. The project has ethics approval from the NSW Population and Health Services Research Ethics Committee. The study's findings are those of the authors' and do not necessarily represent the views of the Department of Health and Ageing, or the Department of Human Services. We thank Paula Lorgelly, Wynand van de Ven and an anonymous referee for their helpful comments.

## REFERENCES

45 and Up Study Collaborators. 2008. Cohort profile: The 45 and Up study. *International Journal of Epidemiology* **37**: 941–947.

Ash AS, Ellis RP. 2012. Risk-Adjusted Payment and Performance Assessment for Primary Care. *Medical Care*. **50**(8): 643–653.

Ash AS, Zhao Y, Ellis RP *et al*. 2001. Finding future high-cost cases: Comparing prior cost versus diagnosis-based methods. *Health Services Research* **26**(6 Part II): 194–206.

Australian Institute of Health and Welfare. 2010. Health system expenditure on disease and injury in Australia, 2004-05. *Health and welfare expenditure series no. 36*. Cat. no. HSE 87. Canberra: AIHW.

Buntin BM, Zaslavsky AM. 2004. Too much ado about two-part models and transformations? Comparing methods of modeling medicare expenditures. *Journal of Health Economics* **23**: 525–542.

Einav L, Finkelstein A, Ryan S, Cullen M. 2011. Selection on moral hazard in health Insurance. SIEPR Discussion Paper No. 10-028. Stanford Institute for Economic Policy Research.

Ellis RP, Fiebig D, Johar M, Jones G, Savage E. 2012. Explaining health care expenditure variation: large-sample evidence using linked survey and health administrative data. Economics Discipline Group Working Paper 2012/1, University of Technology Sydney.

Hausman JA, Taylor WE. 1981. Panel data and unobservable individual effects. *Econometrica* **49**: 1377–1398.

Jones AM. 2011. Models for health care. In *Oxford Handbook of Economic Forecasting*, Clements MP, Hendry DF (eds). Oxford University Press, New York.

Manning WG, Basu A, Mullahy J. 2005. Generalized modeling approaches to risk adjustment of skewed outcomes data. *Journal of Health Economics* **24**: 465–488.

Newhouse JP, Manning WG, Keeler EM, Sloss EM. 1989. Adjusting capitation rates using objective health measures, and prior utilization. *Health Care Financing Review* **10**(3): 41–54.

NSW Department of Health. 2011. Cost of care standards 09/10. No H11/22988.

Zhao Y, Ash AS, Ellis RP *et al*. 2005. Predicting Pharmacy Costs and Other Medical Costs Using Diagnoses and Drug Claims. *Medical Care* **43**(1): 34–43.




DOI: 10.1002/hec