

---

## 4 Data and variables

This chapter describes the data and variables used to estimate the wage model described in chapter 3, and the indirect approach that was developed to estimate the effects of the COAG target health conditions on wages.

### 4.1 Education and health variables

The database for the regression analysis uses five waves of data from the Household, Income and Labour Dynamics in Australia (HILDA) survey. The data were pooled to form a large, cross-sectional dataset. The construction of the dataset is explained in more detail in appendix B.

The dependent variable is the natural logarithm of hourly wages, derived from gross wage or salary income (from all jobs) and average hours worked per week. Hourly wages are preferred to weekly or annual income because income measures are influenced by the wage rate and hours worked. The wage rate is an indicator of individuals' productivity, while the hours worked relates to individuals' participation in the labour market.

One factor that complicates the analysis is the prevalence of casual employment. Casual employees generally do not receive sick leave or other leave, but are paid a loading as compensation. This may lead them to report higher wages than permanent employees with similar characteristics performing similar jobs. Unfortunately, casual loadings were not available with which to adjust the hourly wages of this group. This may understate the extent to which ill health reduces their productivity (relative to permanent employees).

In total, there are 29 explanatory variables used in the selection equation and 28 in the wage equation.

#### Education variables

The HILDA survey includes questions on the respondents' level of education. For this project, education is represented by four dummy variables that indicate the highest level of education attained (degree or higher; diploma or certificate; year 12;

---

and year 11 or below). These variables are relatively straightforward and, for the purposes of this modelling, are considered a reliable indicator of the level of educational attainment.

## **Health variables**

The HILDA survey also includes questions on the individual's health status. However, the data in HILDA are not ideal for the purposes of this project, and the health variables are less straightforward than the education variables.

Two types of health variables were considered. The first option was to use binary variables to indicate the presence of each of the health conditions. It was concluded that the binary variables did not adequately reflect the health status of HILDA survey respondents (see appendix B). Therefore an alternative technique was developed using general measures of physical and mental health to impute the effects of the conditions on wages.

### *General physical and mental health summary scores*

HILDA includes an internationally-used self-completion questionnaire called the SF-36. Responses to this questionnaire are used to assign to each respondent two summary scores, known as the 'physical component summary' (PCS) and 'mental component summary' (MCS) scores. These scores range from zero to 100 and reflect the reported general physical and mental health of the respondent. The summary scores are included as explanatory variables in the model to indicate the effects of general physical and mental health on wages. Using results from other studies, it was possible to estimate the average effect of each of the target conditions on PCS and MCS scores (box 4.1).

As an alternative to the PCS and MCS scores, the model could have used another set of self-reported health variables reported in the HILDA survey (a five-point scale of 'poor' to 'excellent'). The PCS and MCS were preferred for a number of reasons:

- There is a range of studies (Australian and overseas) that estimate the effects of the target conditions on the PCS and MCS scores. Using the PCS and MCS scores in combination with these earlier studies it is possible to estimate the effects of the target conditions on wages.
- The PCS and MCS scores are continuous variables, and are therefore more flexible for this analysis than the discrete variables based on health status.

---

#### **Box 4.1 Estimating the effects of illness using PCS and MCS scores**

Each of the target conditions has an effect on the physical and mental health of sufferers. A review of Australian and overseas literature was used to estimate the average effect of the conditions on physical health (PCS) and mental health (MCS). The preferred estimates are described in greater detail in appendix B. Based on the literature review, the average effects of the conditions on general physical and mental health were estimated to be:

- Cardiovascular disease: reduction of 3.3 points in PCS and 2.1 points in MCS
- Diabetes: reduction of 3.5 points in PCS and 1.0 points in MCS
- Cancer: reduction of 3.6 points in PCS and 0 points in MCS
- Arthritis: reduction of 4.5 points in PCS and 1.5 points in MCS
- Mental illness: reduction of 3.9 points in PCS and 13.9 points in MCS
- Major injury: reduction of 9.9 points in PCS and 4.3 points in MCS.

*Sources:* Alonso et al. (2004); Surtees et al. (2003); Productivity Commission estimates.

## **4.2 Developing a two-stage process for estimating the effects of the target conditions**

Although using binary variables that reliably indicate the presence of the target health conditions would be the best approach to the research question, suitable binary variables were not available. Instead, an alternative approach was devised that uses the PCS and MCS scores for general health to estimate the effects of the target conditions on wages. This approach involves three steps:

1. Estimate the effect of a change in general health on wages.
2. Estimate the effects of each of the target conditions on general health.
3. Combine the estimates from steps one and two to estimate the effect of each of the target conditions on wages.

There is an academic literature that contains estimates of the effects of different diseases on the PCS and MCS scores (for example, Alonso et al. 2004; Surtees et al. 2003; Ware and Kosinski 2001). These estimates can then be fed into the wage model to indirectly estimate the effects of the conditions on wages.

For example, Alonso et al. (2004) estimated that cardiovascular disease was associated with a 3.3 point reduction in the PCS and a 2.1 point reduction in the MCS (out of 100). The PCS and MCS scores have statistically significant effects on

---

wages, so the coefficients on these scores can be used to indirectly estimate the effects of the conditions on wages. Further detail on this approach is set out in appendix B.

### **Issues arising from the indirect approach**

Using PCS and MCS scores to estimate the effects of the target conditions raises a number of potential problems (see also appendix B):

- It is necessary to assume that the PCS and MCS accurately reflect people's health (including the presence of the target conditions). A number of researchers in Australia and overseas have shown that the PCS and MCS are reliable indicators of the effects that specific health conditions have on general physical and mental health.
- Frijters and Ulker (2008) raised concerns about generalising from the results of one survey-based measure of health (such as the PCS) to another (such as chronic illness). However, personal communication with Frijters suggested that the approach taken for this paper is reasonable for estimating the effects of policy changes.
- There is a limited literature on the effects of the target health conditions on PCS and MCS scores, and most of it is from overseas. Overseas literature might be less relevant to Australia because of different labour market and health policies as well as the incidence of the health conditions.
- There is a wide variation in the estimates of the effects of some conditions on PCS and MCS scores. For example, Chittleborough et al. (2005) estimated the effects of diabetes on Australian PCS and MCS scores. They found that diabetes is associated with a PCS that is 7.8 points lower, and an MCS that is 0.1 points lower than for people with normal glucose levels. Alonso et al. (2004) reported the effects of diabetes in six European countries, Japan and the United States, finding average reductions of 3.5 and 1.0 points, respectively.
- The approach assumes that co-morbidities (that is, people suffering from two or more conditions) have an additive effect on PCS and MCS scores. (Regressions using the Survey of Disability, Ageing and Carers dataset show that this assumption is justified empirically.)
- The approach assumes that the PCS and MCS are linear (for example, that a reduction in physical health from 75 to 72 is equivalent to a fall from 51 to 48).

Although these concerns are real and need to be acknowledged in reporting the results, this approach is judged the best means currently available to estimate the effect of chronic illnesses on wages, given the available data.