The Cost of Ill Health

Heather Holt

NEW ZEALAND TREASURY
WORKING PAPER 10/04

NOVEMBER 2010

THE TREASURY
Kaitohutohu Kaupapa Rawa

New Zealand Government

NZ TREASURY WORKING PAPER 10/04

The Cost of Ill Health

MONTH/YEAR

November 2010

AUTHOR

Heather Holt
formerly of the New Zealand Treasury
1 The Terrace
Wellington 6011
New Zealand

Email heather.holt@bis.gsi.gov.uk

ACKNOWLEDGEMENTS

Thank you to Lisa Meehan, Katy Henderson, Grant Scobie, Caroline Shaw, Kristie Carter, Gerald Minnee and Gary Blick for their advice and comments. I would also like to thank Southern Cross Medical Care Society for allowing me to have access to results from their survey, specifically to Jo Broadhead for dealing with all my enquiries.

The Health Research Council of New Zealand, and Health Inequalities Research Programme of the University of Otago, Wellington, are acknowledged for funding and establishing the SoFIE-Health data utilised in this publication.

NZ TREASURY

New Zealand Treasury
PO Box 3724
Wellington 6140
NEW ZEALAND

Email information@treasury.govt.nz

Telephone 64-4-472 2733

Website www.treasury.govt.nz

DISCLAIMER

Abstract

This paper aims to quantify some of the costs associated with ill health in New Zealand. The main focus is in estimating indirect costs as opposed to direct health care expenditure costs. In particular, it estimates the cost of absenteeism, presenteeism, working less and not working at all owing to ill health. Around 1,196,200 working age, non-students are estimated to contribute to one or more of the components of indirect costs estimated. That is 61.8% of all working age, non-students. Evaluated at the average full-time pay rate, the estimated hours lost equate to $4.127 billion to $11.563 billion in 2004/05; 2.7% to 7.6% of Gross Domestic Product (GDP). The considerable range in the cost estimate is owing to the large range of the presenteeism estimate as a result of having to use a variety of methods and assumptions to obtain estimates. This illustrates what a difficult concept presenteeism is to estimate, and how sensitive estimates are to the assumptions made.

Owing to the assumptions made, the estimate of absenteeism is likely to miss a large group of absenteeism and thus the estimate is likely to be at most a lower bound. Despite this under-coverage, and in line with other research, it seems likely that absenteeism will be generally smaller in size than presenteeism. Working fewer hours, or not working at all, owing to ill health are estimated to affect widely different numbers of people; 458,500 and 42,300 respectively. However, in terms of costs their impact is more similar; $1.442 billion and $1.755 billion respectively.

Taking the estimate of presenteeism nearest the mid-point of the range, indirect costs are estimated to be $7.483 billion; 4.9% of GDP. Presenteeism accounts for 55% of this cost, not working 23%, working less 19% and absenteeism just 3%.

The only component of direct costs estimated is hospital inpatient appointments, owing to data limitations and the particular focus of this study. Around 1,301,700 people are estimated to be affected by hospital inpatient costs or indirect costs. In monetary terms the total cost of the considered components is estimated to be $5.417 to $12.853 billion; 3.6% to 8.5% of GDP.

JEL CLASSIFICATION J22 Labour Supply; I10 General Health

KEYWORDS Health; Productivity; Absenteeism; Presenteeism; Labour Force Participation

Table of Contents

List of Tables

List of Figures

The Cost of Ill Health

1 Introduction

A healthy workforce is an important economic asset. Poor health can impact on the economy in numerous ways. As well as the obvious impact on health care costs, ill health can impact on labour market behaviour; for example, participation, wages, hours worked, productivity and retirement decisions. It is therefore important from an economic growth perspective to understand the relationships between health and labour market behaviour in order to help inform any policy decisions aimed at improving health. In countries such as New Zealand, which have an ageing population, understanding these relationships becomes even more important as more people reach the age at which their health may deteriorate and affect their labour market behaviour (Currie and Madrian, 1999). New Zealand's labour productivity is the subject of ongoing research and debate. One of the motivations behind this paper and a companion piece on labour force participation (Holt, 2010) was to explore the relationship between health and labour market behaviour and productivity. Underlying this issue is the question of the extent to which productivity is affected by barriers to skills utilisation (such as ill health), as opposed to other factors such as poor skills formation.[1] As an initial step towards answering such questions, this paper summarises the results of a cost of illness type study aimed at estimating some of the costs associated with ill health in New Zealand using evidence from the Survey of Family, Income and Employment (SoFIE). It is acknowledged from the outset that it is not possible to use SoFIE to estimate all costs associated with ill health. As such this study aims to estimate the magnitude of just some of the associated costs.

The main focus of this research is estimating some of the indirect costs; that is, the value of potential resources lost. This includes lost output as a result of: being away from work; being less productive at work; working fewer hours; and being out of the labour force completely owing to ill health. Estimating these indirect cost components is inherently difficult. Even with data on current behaviour, it is difficult to predict how behaviour would change if a certain factor, such as an individual's health state, was different. In addition, while the SoFIE data provides a wealth of information on a person's current behaviour, some information required is not available. As such this inevitably requires a number of assumptions. Where assumptions are made, they are based on reviews of other research in the area. The resulting estimates should be interpreted with these caveats in mind. Throughout the results, the strengths and weakness of the estimates will be highlighted

1 It is acknowledged that health itself can impact on skills formation.

and where weaknesses exist, and other data sources are available, these will be presented to help contextualise the estimates from SoFIE.

The only element of direct costs (which are the cost of resources used) that will be included are hospital inpatient treatments. These are estimated from SoFIE despite it being possible to obtain them from Ministry of Health (MoH) published information. SoFIE is used as it enables hospital inpatient appointments to be attributed to people and thus compared with labour market information. This information feeds into the estimates of the indirect costs, ensuring the cost of appointments and those of lost hours are from the same source.

There are relatively few existing cost of illness studies for New Zealand. The ones that do exist generally adopt a top-down approach rather than a bottom-up approach using person-level data. With the availability of micro-level data on health from SoFIE, there is now an opportunity to estimate the cost of illness using a bottom-up approach. This type of micro-level data has the potential to provide information about the characteristics of individuals with various health states and labour market outcomes in a manner that aggregate data does not allow. The current study provides an initial demonstration of the type of cost of illness analysis that can be done using SoFIE. It is therefore important to bear in mind the study's limitations, which will be highlighted throughout this paper, when interpreting the results. This is particularly important when interpreting the results from a policy perspective. This paper is not a review of current health spending; the focus is simply to try to obtain first estimates of costs that are lost in one year as a result of ill health. The paper is not a cost-effectiveness or cost-benefit analysis. As such it does not attempt to assess how changes in current health policies may result in better health outcomes and thus cost savings. Overall, although this study may imply that there is potential for better labour market outcomes if health was improved, it remains silent about the scope to realise this potential, or the possible policy mechanisms to achieve it.

Section 2 provides the theory behind the costs associated with ill health. It then summarises the results of other work done in this area before defining the costs of ill health that can be estimated using SoFIE and those that cannot. Section 3 of this paper describes the data used. The methods and results for direct costs are summarised in Section 4. For each component of indirect costs in turn, Section 5 summarises the method used and then presents the estimates using the defined method. Section 6 brings all the cost estimates together and provides a discussion. Full details of the variables used and the model results can be found in the appendices.

2 Background

This section briefly outlines the theory behind the various costs that may result from ill health before summarising the results of other work done in this area. As discussed above, there are limitations to the analysis that can be done with SoFIE and therefore the aim of this analysis is to use evidence from SoFIE to estimate the magnitude of some of the component costs of ill health. This section therefore then goes on to define the specific costs associated with ill health that will be included in this analysis and gives some examples of the additional costs associated with ill health that are excluded (although this list is not exhaustive).

2.1 Theory

As well as the obvious contribution to health care costs, ill health can also impact on labour market behaviour; for example, participation, wages, hours worked, productivity and retirement decisions. An understanding of the relationship between ill health and labour market behaviour is important for informing policy decisions aimed at improving health and labour market outcomes. There is a vast range of interacting effects between economic status and health. There is some evidence that labour force participation and health are simultaneously determined (Cai and Kalb, 2006). That is, health may affect participation but that participation may also impact on health. For some groups participation was found to be positively related to health, while for others participation was found to have a negative effect. It is acknowledged that this reverse causation, working through various economic measures, exists, however, the implicit assumption in this study is that ill health leads to adverse economic outcomes. Any impacts of participating on health are not accounted for.

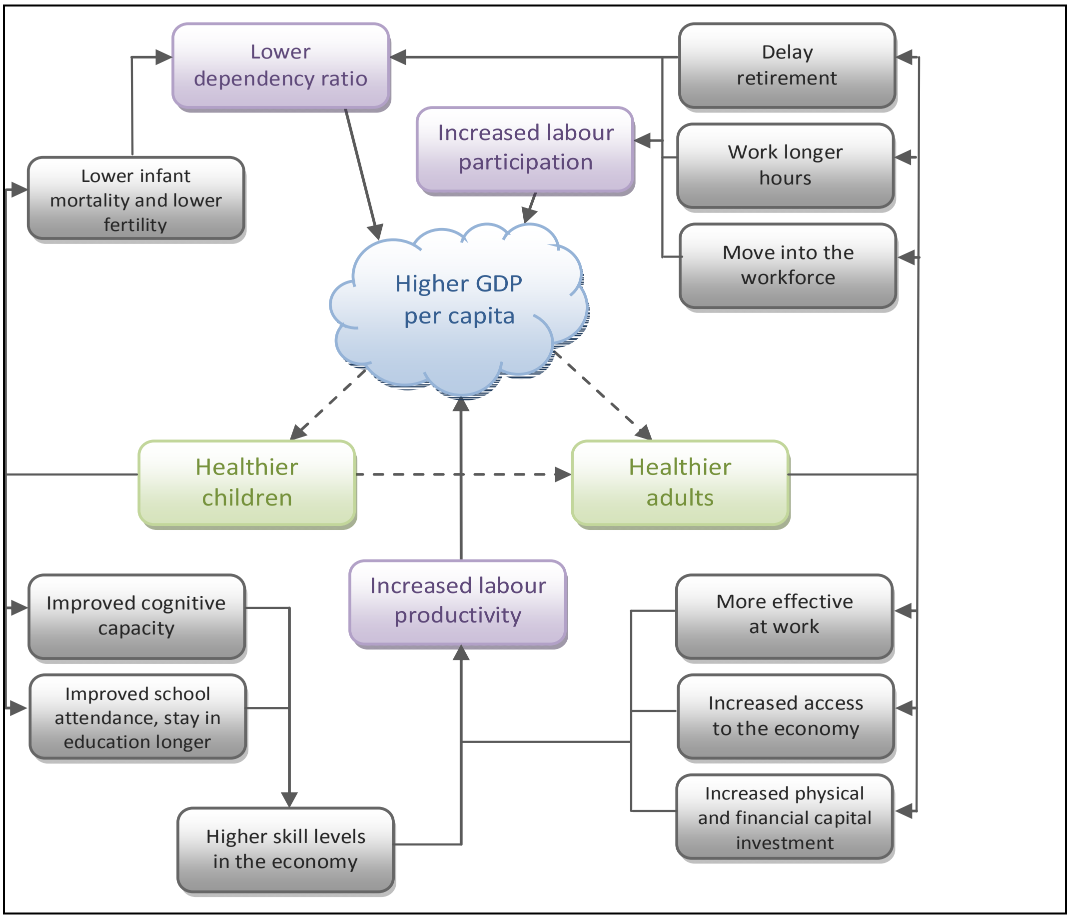
Health is one of the key factors that may affect a person's ability to develop their skills and knowledge. The mix of skills, knowledge and capabilities that a person possesses (their human capital) is positively related to their productivity and the demand for their labour. If poor health is a barrier to the development or use of skills, in youth or adulthood, then improving health will not only lead to a reduction in health care costs, but also result in higher labour force participation and economic output. As people age this relationship becomes even more important. Older people in better health are more likely to maintain their attachment to the labour market than those of poorer health (ie, be less likely to take early retirement or be less likely to reduce the amount of hours worked), again resulting in higher economic output.[2]

Improvements in health are not only likely to increase labour force participation (Holt, 2010; Enright and Scobie, 2010) they are also likely to change the behaviour of those already participating. The number of hours a person is contracted to work in a given period is likely to be impacted by their health. Further, those who participate who have poor health are likely to take more days off work ill (absenteeism) or, if they attend work when unwell, may work less productively (presenteeism). Improvements in health may therefore result in people working more hours than they currently do; being more productive when at work; and taking less days off work for medical appointments or owing to ill health. All this will lead to higher output.

Figure 1 summarises the ways in which improved health is thought to contribute to GDP growth. This lost output, coupled with the costs of treating ill health, may be seen as representing some of the costs resulting from ill health. Any lost output will potentially affect GDP levels and growth. In this paper the lost output is quantified and compared with the actual level of GDP in the period. This quantification assumes that lost output affects not only the individual, but also the GDP output in the economy.

2 In a study on the role of ill health in retirement decisions, Disney *et al* (2003) found evidence that health shocks were associated with an increased chance of leaving the labour market.

Figure 1 – Productivity and economic implications of improved population health



Source: Buchanan, Blick and Isaac (forthcoming)

Note: Increased labour productivity is for all adults (including those who would have had poorer health and their dependants). It includes increased productivity as a result of less time off work (absenteeism); working more; and working more efficiently (ie, owing to less pain or distractions).

2.2 Previous studies

Numerous cost of illness studies have been conducted, although very few of these relate to New Zealand. Where these New Zealand studies exist they often estimate the costs for only a specific disease rather than the cost of ill health as a whole and, owing to a lack of person-level data, often adopt a top-down approach.[3] The inclusion of the health module in SoFIE, and the linking of these responses to hospital inpatient information provides an opportunity to estimate the costs of ill health using a bottom-up approach.

One example of a cost of illness type study for all illnesses in New Zealand that used a bottom-up approach was conducted by Southern Cross Medical Care Society (2009). This study aimed to assess the costs to employers of illness. It therefore only included costs owing to absenteeism and presenteeism, rather than the wider costs to the economy from people not working, working fewer hours or as a result of treatment costs. The results were based on a small online survey of New Zealand workers which asked general lifestyle questions along with a number of health-related and workplace-related questions. Using these results the cost of illness to New Zealand employers was estimated to be over $2 billion a year across the whole workforce.

[3] A top-down approach uses aggregate data to estimate economic costs. The bottom-up approach makes use of more detailed person-level data to compute estimates.

There have been numerous cost of illness studies for countries outside New Zealand using various methodologies based around a bottom-up approach. Many focus on the direct costs, as opposed to trying to estimate more difficult to measure indirect costs, such as lost output. However, those studies that do include such estimates suggest that productivity losses associated with lost workdays (absenteeism) and reduced on-the-job productivity (presenteeism) may be substantially more than the treatment costs of ill health. A recent US study estimated that, in 2003, the combined (direct and indirect) economic impact of selected chronic diseases was US$1.324 billion (around 12.5% of GDP).[4] However, the productivity losses (indirect costs) accounted for around 80% of this estimate; around 10% of US GDP (DeVol and Bedroussian, 2007).

The output lost owing to presenteesim alone is thought to be immense. Some literature suggests that for some diseases these losses can be up to 15 times larger than for absenteeism (Newton, 2000, in DeVol and Bedroussian, 2007). DeVol and Bedroussian draw similar conclusions; that is 79% of the indirect costs are a result of individual presenteeism; that is, around US$828.2 billion. A further US$80.2 billion of the indirect costs are a result of caregiver presenteeism.

Research by Goetzel *et al.* (2004) used the data from a large medical/absence database, along with findings from several productivity surveys, to estimate health, absence, disability and presenteeism cost estimates for certain health conditions. They concluded that presenteeism costs were higher than medical costs in most cases, and represented 18% to 60% of all costs for the 10 conditions.[5]

What is also clear from the literature, however, is that estimates of the economic impact of health vary significantly. This is an indication of the difficulty involved in estimating the indirect costs. The variation can largely be explained by the methodology used to produce the estimates, the assumptions made and also the definition of ill health that is adopted. As an example, another US study estimated that in 2003 labour time lost owing to health reasons was equivalent to lost economic output totalling US$260 billion per year (Davis, K., Collins, S. R., Doty, M. M., Ho, A. and Holmgren, A. L., 2005). This research used a wage-based method to estimate the lost output. This is well below the estimate of US$1,047 billion from the DeVol and Bedroussian study, which used a GDP-based approach; only estimated lost output for those already in the workforce; and covered a narrower definition of ill health – focusing on a selection of chronic diseases. DeVol and Bedroussian (2007) also used their method to estimate the lost output using a wage-based method. Using this method the estimate of the economic impact of selected chronic diseases fall from US$1.324 billion to US$464 billion. Further, unlike the DeVol and Bedroussian results, the study by Davis *et al* (2005). estimated absenteeism to account for a higher proportion of lost output than presenteeism.

More generally there has been much work that indicates a significant relationship between health and labour market behaviour in New Zealand. Recent work by the New Zealand Treasury concluded that health is significantly related to labour force participation, using various health measures and even after accounting for certain types of endogeneity (Holt, 2010). Those people who experienced negative health shocks into fair or poor health, or who have poorer self-rated health were found to be less likely to participate in the labour force. The reduction in the chance of labour force participation associated with poor health appeared to be larger for working full, as oppose to part, time.

---

4 The diseases considered included: cancers, hypertension, mental disorders, heart disease, pulmonary conditions, diabetes and stroke.

5 Results from Geotzel *et al.* (2004) were used to estimate presenteeism in DeVol and Bedroussian (2007).

Another recent piece of research conducted by the Treasury using person-level data focusing on older New Zealanders (aged 55 to 70), also identified a significant relationship between health and labour force participation (Enright and Scobie, 2010). It concluded that, regardless of the various health measures tested, a significant reduction in labour force participation was associated with poorer health status.

3 Data

3.1 Survey methodology

The Survey of Family Income and Employment (SoFIE) is the main data source analysed in this paper. SoFIE is a survey of a nationally representative sample of New Zealand permanent residents in private households. It is conducted by Statistics New Zealand. The core SoFIE survey modules include questions on demographics; dependent children; labour force involvement; education; family and income. All respondents in the original sample are followed over time, even if their household or family circumstances change, forming a longitudinal sample. The survey commenced in 2002 and will continue until 2010. When the present study was undertaken, there were three waves of data available for analysis (SoFIE Waves 1–3 Version 4). Further information on the survey methodology can be found in Appendix A.

3.2 Population and sample of interest

The analysis is based on those people who remain eligible and respond in Waves 1–3 who are aged 15 and over at the end of the reference period in Wave 1, as this is the group that were asked the health module in Wave 3. So they will be aged 17 and over in this analysis. The results are therefore representative of the usual adult resident population of New Zealand who lived in private dwellings on the main islands of New Zealand in 2002/03 and who remain alive and are non-institutionalised by 2004/05.

As with all surveys, not all those approached to take part in the survey agree to participate. In addition, those who initially respond may choose not to respond in subsequent waves of the survey (attrition). While the response rates are good compared with similar surveys, longitudinal response rates were lower for those of fair or poor health compared with those of better health.[6] Statistics New Zealand provides a longitudinal weight (standard longitudinal weight) which accounts for non-response and aligns the composition of the sample with that of the New Zealand population in October 2002 in terms of age, gender and Māori. However, the weights do not completely restore the distribution of people across the health states. The implication of this is that the SoFIE population is likely to be healthier than the population it represents and the New Zealand population more generally. More specifically, those with the most severe cases of the health conditions considered will die or may be institutionalised and therefore are not covered by the survey. Consequently the actual component costs of ill health that are estimated in this study are likely to be higher than the results based on SoFIE suggest. Further information on the limitations and strengths of SoFIE more generally can be found in Appendix A.

6 In part this will be because those of fair or poor health are more likely to die or move into institutions where they are not followed up.

In Wave 3, together with the social and economic questions that are asked in every wave, the original sample members over 15 years old in that wave were asked a detailed set of health questions. At the end of the health module respondents were asked to give permission for their data to be linked to information on hospitalisations, cancer registrations and future death registrations held by New Zealand Health Information Service (NZHIS). The Statistics New Zealand weight does not allow for non-consent. While there was no difference in consent rates by self-rated health, those with a higher number of chronic diseases were significantly more likely to agree to the linkage. All analysis in this paper is based on the linked sample with adjusted longitudinal weights used to realign the sample with the population (adjusted longitudinal weight) as opposed to the weights provided by Statistics New Zealand (standard longitudinal weights).[7] [8]

3.3 Variables considered

The core SoFIE modules include questions on demographics; dependent children; labour force involvement; education; family and income.[9] The main SoFIE health variables used in this analysis are self-rated health, whether activity has been stopped by illness and how often a person's activities have been reduced by their physical or mental health. As well as the detailed information asked in the SoFIE survey, for those matched, consenting SoFIE respondents, inpatient hospital treatment information back to 1990 is available. This information includes the age at discharge; gender; International Statistical Classification of Diseases and Related Health Problems code (ICD code); Diagnosis-related Group (DRG); date of admission; date of discharge; length of stay; and facility type for each hospital episode. The hospital inpatient information does not include records for appointments of less than three hours (ie, outpatient appointments) or information on appointments with primary health care providers such as GPs. It will also not include information on private hospital treatments; treatments at these facilities are only included if the treatment is publicly funded. These health variables, along with the non-health explanatory variables used in some of the models, are defined in Appendix Tables B1, B2 and B3 of Appendix B.

3.4 Coverage

3.4.1 Costs included

In this study "ill" health will be defined to be less than excellent health, which is not the result of injury or pregnancy. That is, excellent health is the base case against which other health states are compared.[10] When estimating the different components of costs a range of questions will be used to define the group in "ill" health. These questions/definitions will be discussed in the method section for each cost component and are defined in Appendix B. This work could potentially be extended in the future to cover injury-related costs.

---

7 More information on the adjusted weights is available from the author.

8 Where possible the same analysis was carried out on the full sample using the standard longitudinal weights and there was little difference in the results.

9 The full questionnaire can be found here:http://www2.stats.govt.nz/domino/external/quest/sddquest.nsf/12df43879eb9b25e4c256809001ee0fe/14d945bb95ab2bbbcc256fb70077b3bb?OpenDocument

10 So people in very good, good, fair or poor health are defined to be in "ill" health if these health states are found to be significantly related to lower participation or hours worked. In fact, the results indicate that there is no evidence to suggest that being in very good health is significantly different being in excellent health in terms of labour force participation or hours worked.

This analysis is a prevalence cost of illness study. That is, it estimates the cost of ill health for a one-year period regardless of when the ill health started.[11] The assumption made is that the cross-section view of the costs at different stages of ill health represents the progression of ill health. As such, the costs may be seen as those that would be saved in a certain period in the absence of ill health rather than the amount saved if ill health was eradicated, as this amount would be much larger.

Cost estimates refer to Wave 3 of SoFIE as this is the period to which the health questions, and the summary labour force information used in the analysis, relate. The reference period for the survey is not the same for all respondents. They are all interviewed between October 2004 and September 2005 and asked usually about the 12 full months prior to the interview date (see Appendix A for more detail). In order to calculate costs for each person the reference period is used as this enables comparisons to be made between labour force participation and health. Despite the differences in the reference period the resulting total annual costs will be referred to as October 2004 to September 2005. All costs are evaluated at current (2004/05) values.

While all the direct (cost of resources used) and indirect (value of resources lost) costs as a result of ill health would ideally be included, owing to data limitations this is not possible. The costs that can be estimated from SoFIE can be seen in Figure 2, along with the groups for which the costs are estimated. These costs will not apply to everyone in the group. For example, a person of working age with ill health may have no hospital inpatient appointments in Wave 3. However, they may work less than they otherwise would as a result of ill health. Similarly, a participant in the labour force may suffer from presenteeism but may have no absenteeism. The direct costs are estimated for those aged 17 and over; however, indirect costs are only estimated for those aged 17 to 64 (working age) who are non-students.[12]

The main focus of this research is on estimating indirect costs. The only component of direct costs (the cost of resources used) that will be included is hospital inpatient treatments. SoFIE will be the main source used to estimate the hospital inpatient costs despite it being possible to estimate them using MoH published information. SoFIE is used as it enables hospital inpatient appointments to be attributed to people and thus linked to labour market information. However, SoFIE does not allow any other direct health care costs to be attributed to specific individuals. This information feeds into the estimates of the indirect costs, ensuring the cost of appointments and those of lost hours are from the same source. Further, it allows more flexibility in estimating the cost of appointments for specific groups and types of treatments than is available using the published MoH information.

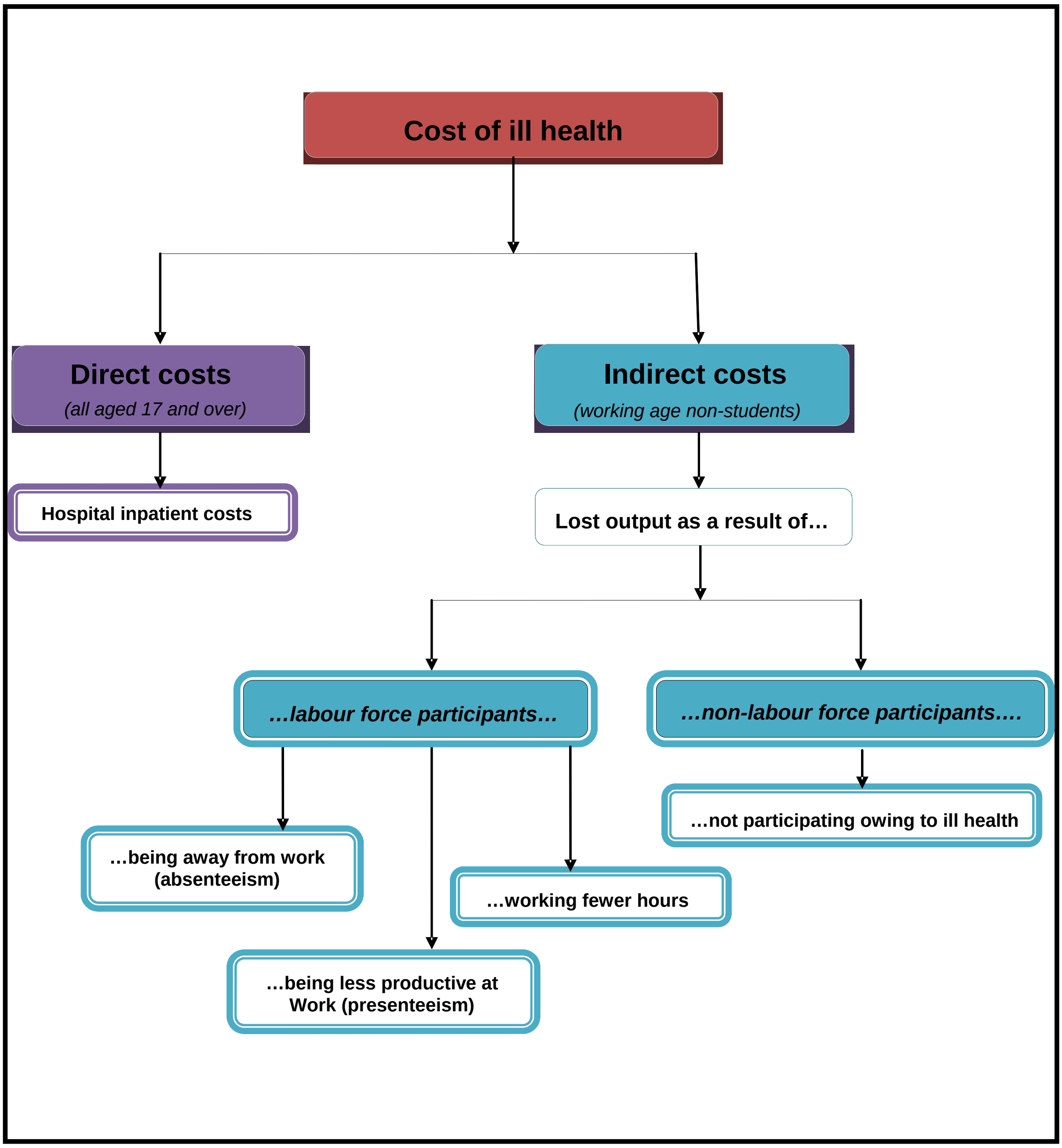
The hospital inpatient costs covered are the medical and surgical costs that are publicly funded (mainly in public hospitals). They include a combination of hospital inpatient, physician inpatient, diagnostic tests, prescription drugs and drug sundries, and medical supplies that are used or carried out within hospitals. At an aggregate level there is little difference in the hospital inpatient costs estimates from SoFIE and those from MoH (see Appendix C).

---

[11] This is as opposed to an incidence-based study which would estimate lifetime costs for diseases that start within the period (Segel, 2006). A prevalence-based study is the best possible type of study as the linked hospital information only goes back to 1990 and the employment information is only known from Wave 1 onwards.

[12] People aged 15 and 16 years are not covered as they were not adults when the survey commenced in Wave 1.

Figure 2 – Possible ill health-related costs that can be estimated using SoFIE



3.4.2 Costs excluded

The figures that can be estimated from SoFIE will be an underestimate of the total cost of ill health as there are some costs that are not captured by SoFIE and therefore not included in analysis. Examples of these excluded costs are:

- Wider health care costs – such as those for primary and community-based care, pharmaceuticals, emergency services, outpatient appointments or non-medical direct costs (for example, transportation and relocation expenses) whether publicly or privately funded. All non-publicly funded hospital costs in private hospitals – as this treatment information is not available to link to SoFIE meaning it is not possible to estimate treatment costs or absenteeism costs as a result of such appointments.
- Costs of training health care providers for a particular illness or capital costs unless these are reflected in the cost of hospital inpatient care – as these costs are often difficult to attribute to a particular disease.
- Intangible costs of pain, such as the perceived cost associated with the loss of quality of life as a result of poor health – as information on the perceived quality of life is not collected in SoFIE.
- Estimates of the value of output lost owing to not being able to undertake housework owing to ill health – as information on the amount of housework undertaken is not collected in SoFIE.
- Costs for those who die from poor health in the period – as respondents have to be alive at the date of interview in order to respond to the survey and consent for their data to be linked to hospital records.[13]
- Costs to employers of having to recruit new employees as a result of staff members leaving owing to ill health (owing to activity limitations or death) – as the behaviour of employers when a staff member leaves, and the associated costs to employers, are not known.
- Indirect costs for those people aged 65 and over or for full-time students – owing to difficulty in predicting labour market behaviour of these groups in the absence of ill health.[14]
- Reductions in the amount of government benefits paid if health improved – as there are problems in the measurement of income in SoFIE and it is therefore difficult to attribute the amount of Sickness Benefits to each person to calculate the amount for the group of interest for the analysis. Also in some cost of illness studies, benefits are seen as a shift in resources, not a use of them (Segel, 2006). To provide some context, in 2005 the value of Sickness Benefits in total for New Zealand was $510 million, Disability Allowances $267 million and Invalid Benefits $1.026 million (The Treasury, 2008).
- The lost output of carers (perhaps family members or friends) of those with ill health (as a result of the carer not working; working less; being absent from work; or being less productive when at work as a result of worrying about their dependants) – as it is not always possible to identify carers in SoFIE.
- Expenditure on private health insurance.

---

[13] A small number of people die after the interview date but within the Wave 3 period, however, for consistency these people are not included.

[14] Output will be lost as a result of ill health for both those over 65 and full-time students who continue to participate in the labour force. Output will also be lost as a result of people over 65 moving out of the labour force earlier than they would have in the absence of ill health. Initial analysis suggests that the amount of hours lost for these excluded groups is relatively small. Around 95% of all hours worked are worked by those of working age, non-students. Expanding the scope of the analysis could be developed in future analysis.

As mentioned earlier, in addition to some costs being excluded, there are also more general limitations of this analysis. In particular, this study is a preliminary investigation into some of the costs of illness in New Zealand. It does not draw any implications for health care policy and does not provide a cost benefit analysis of improving the health of the working age population.

4 Direct costs – method and results

This section describes the method used to estimate one component of the direct cost of ill health; that for publicly funded hospital inpatient appointments. Following the description of the method, basic descriptive information is presented along with the cost estimates based on the method outlined.

As discussed above, the main focus of this research is on estimating indirect costs. The only component of direct costs estimated is hospital inpatient costs. Actual hospital inpatient costs are available from MoH. However, SoFIE is used as it enables hospital inpatient appointments to be attributed to people and thus compared with labour market information. This ensures that the cost of appointments and those of lost hours are from the same source. This information feeds into the later estimates of the indirect costs. Further, the use of SoFIE allows more flexibility in estimating the cost of appointments for specific groups and types of treatments than is available using the published MoH information; by qualification level of individuals is one example.

Hospital inpatient costs represent a small component of known overall health care costs. More information about estimates of hospital inpatient costs from SoFIE compared with those based on published MoH results and details of where these costs sit within Vote Health can be found in Appendix C. As well as publicly funded health care costs there are privately funded health care costs. These privately funded costs are not discussed here.

4.1 Hospital inpatient costs

4.1.1 Method

For people who consent, it is possible to link hospital inpatient appointments to SoFIE responses. Each hospital inpatient case has a Diagnosis-related Group (DRG) code. The DRG codes provide a system whereby hospital treatments are categorised by both a clinical homogeneity and use of a similar hospital resource. As such they form the basis of the calculation used to assign a cost weight, and therefore a cost, to each inpatient hospital admission. Cost weights are used to measure the volume of hospital cases and are the current payment mechanism for inpatient and day cases between District Health Boards (DHBs) and hospitals. Cost weights are computed on a DRG basis using a combination of length of stay, total hours on mechanical ventilation and around 20 procedure and diagnosis codes. Payments between DHBs and hospitals are then calculated based on cost weights multiplied by the national price (a fixed cost multiplier) applicable for that financial year.[15]

The cost weights and DRG codes allow estimation of the cost of each hospital case for SoFIE respondents. Based on the DRG version in place at the time of the appointment a cost weight was linked to each hospital inpatient case. This cost weight was multiplied by

15 The costs are estimated using the national price for the end date of a hospital episode.

the national price applicable at the start date of the treatment.[16] This results in an estimated cost for each hospital appointment. The individual cost for appointments in a specific period can then be weighted to reflect the New Zealand population using the adjusted longitudinal weights from SoFIE and summed to estimate the total costs of inpatient appointments in that period.[17] The formula used can be found in Appendix D.

While this method provides the best estimate of the cost of each appointment it should be remembered that these costs reflect a mix of events only. Applying them to specific events in this way will not reflect the true cost of treatment for many cases. For example, some treatments may be much more complex than average and others will be much less complex, even within the same DRG code.

Hospital inpatient appointments as a result of ill health will be defined as those that are not injury-related or pregnancy-related; that is, ICD codes A00-N99, R00-R94 and Z00-Z99. In order for hospital inpatient appointments to be compared with labour market activity in the later stages of this work, appointments that start in the annual reference period are considered. Owing to the continuous interview period used in SoFIE this means that, while the estimate of costs will be an annual estimate, the annual period will not be the same for each person. There will in fact be 12 annual periods running from October 2003 to September 2005 (see Appendix A for further information). At an aggregate level there is little difference in total hospital cost estimates when the annual reference period is used as opposed to the annual interview period. As such the overall hospital costs will be referred to as costs in Wave 3 (October 2004 to September 2005).

4.1.2 Results

Table 1 shows the estimated inpatient hospital figures from SoFIE. It shows that there were 565,600 inpatient appointments in 2004/05. Around 346,500 people had an inpatient appointment in the same period; that is, 11.9% of all people aged 17 and over, indicating some people had multiple appointments. The cost weighted number of patients was 550,200. The cost of these hospital appointments is estimated to be $1.570 billion.

While the cost is presented for all inpatient appointments, only non-injury or pregnancy-related inpatient appointments are defined in this research to be ill health-related. There were 440,100 inpatient appointments in 2004/05 (77.8% of all inpatient appointments). These appointments were experienced by 267,700 people (9.2% of all people aged 17 and over). The cost of the ill health inpatient appointments is estimated to be $1.290 billion.

---

[16] The national price from July 2004 to June 2005 was $2,854.58 and from July 2005 to June 2006 was $2,949.09.

[17] A description of the adjusted weights can be found in Appendix A.

Table 1 – Publicly funded hospital inpatient costs, aged 17 and over, 2004/05

	Number of appointments (not cost weighted)	Number of appointment (cost weighted)	People affected		Cost $bn
			Count	% of all people	
All inpatient appointments	565,600	550,200	346,500	11.9	1.57
Ill health-related inpatient appointments	440,100	450,400	267,700	9.2	1.29
Of which:					
Working age non-students	232,500	229,500	162,200	7.3	0.66
Students or those over working age	207,600	220,900	105,500	15.5	0.63

Source: SoFIE/NZHIS Wave 3 Version 4, adjusted longitudinal weights, Statistics New Zealand

Note: These are for hospital appointments that start in reference period of interviews held between October 2004 and September 2005.

Self-rated health

Table 2 shows the estimated ill health inpatient hospital figures from SoFIE by self-rated health state. It shows that, as would be expected, those of poor health are the most likely to have had an ill health-related hospital appointment (40.2% of those in poor health have one or more hospital inpatient appointments compared with only 4.3% of those in excellent health). However, owing to the relative sizes of the health state groups, the number of people affected is largest for those in good health; 78,000 people in good health had one or more hospital inpatient appointments (29.1% of those with one or more hospital appointments). The cost of hospital appointments for those in fair or poor health accounts for a higher proportion of the overall cost than the proportion of people affected in the same groups. For example, 8.6% of those people who have one or more ill health-related hospital inpatient appointments are in poor health. This compares with 16.9% of ill health-related hospital inpatient costs being for this group. It indicates that hospital appointments for those in fair or poor health are more numerous than for those in the other health states.

Table 2 – Publicly funded ill health-related inpatient hospital costs by self-rated health, aged 17 and over, 2004/05

	People affected			Cost	
	Count	% of those within the health state	% of those affected	$bn	% of cost
Excellent	43,600	4.3	16.3	0.128	9.9
Very good	71,600	7.1	26.7	0.258	20.0
Good	78,000	12.5	29.1	0.372	28.8
Fair	51,400	25.3	19.2	0.311	24.1
Poor	23,000	40.2	8.6	0.218	16.9
Total	**267,700**	**9.2**	**100.0**	**1.290**	**100.0**

Source: SoFIE/NZHIS Wave 3 Version 4, adjusted longitudinal weights, Statistics New Zealand

Notes:
1. These are for hospital appointments that start in reference period of interviews held between October 2004 and September 2005.
2. GDP for September 2004 to August 2005, the closest period for comparison, is estimated to be $151 billion in current prices.
3. Columns may not sum to totals owing to rounding.

Highest qualification level

Table 3 shows the estimated ill health hospital inpatient figures from SoFIE by highest qualification level. It shows that those with no qualifications are the most likely to have had an ill health-related hospital inpatient appointment (13.3% compared with 5.6% of those with a degree or higher). However, owing to the relative sizes of the qualification groups, the number of people affected is largest for those with post-school vocational qualifications; 99,100 had one or more ill health-related inpatient appointment (37% of those with one or more hospital appointments). There is little difference between the distribution of costs by qualification level and the distribution of people affected by qualification level.

Table 3 – Publicly funded ill health-related inpatient hospital costs by highest qualification level, aged 17 and over, 2004/05

	People affected			Cost	
	Count	% of all people	% of those affected	$bn	% of cost
Degree or Higher	26,000	5.6	9.7	0.082	6.3
Post-school vocational	99,100	9.4	37.0	0.479	37.1
School qualification	64,400	8.0	24.1	0.310	24.0
No qualification	78,100	13.3	29.2	0.416	32.2
Total	**267,700**	**9.2**	**100.0**	**1.290**	**100.0**

Source: SoFIE/NZHIS Wave 3 Version 4, adjusted longitudinal weights, Statistics New Zealand

Notes:
1. These are for hospital appointments that start in reference period of interviews held between October 2004 and September 2005.
2. Columns may not sum to totals owing to rounding.

5 Indirect costs – method and results

This section outlines the methods used to calculate the indirect cost components associated with ill health. Initially some general information on the methods and definitions is presented which is relevant to all components of indirect costs. The methods and related results for estimating the cost of absenteeism, presenteeism, reduced hours and lack of participation in the labour force as a result of ill health are presented in turn. All estimates in this section are for working age (defined as those aged 17 to 64) non-students who are not working overseas.[18]

5.1 General methods

5.1.1 Defining labour force participation

Labour force participation throughout the reference period is not fixed; some people may not be working at the interview date but may have worked at some point in the reference period and thus have non-zero hours information for the annual period. Conversely a person may be working at the interview date but may not have worked throughout the reference period.

Labour market spell data is collected in SoFIE; that is, information on the labour market status throughout each person's annual reference period. This allows labour force participation to be defined in various ways. Table 4 compares two definitions; the number of people participating at the interview date compared with the number participating at any point in the reference period (defined to be one or more weeks of work undertaken). Around 1,866,000 people are estimated to be participating at the interview date; a participation rate of 83.8%. This is the more standard definition of labour force participation, with those who are unemployed being classed as participating in the labour force. As would be expected, a higher number are estimated to participate at any point in the reference period (1,937,100 people), giving a higher participation rate of 87%.[19] Under this definition those who are unemployed for the whole period will be classed as not participating. Despite the differences, 264,800 (11.9%) of people are defined as not participating using both definitions.

18 Students were defined as those who studied full-time for more than nine months in the reference period; reported that they were still at school; or reported that they were economically inactive as a result of being a student.

19 The estimates for the proportion of people participating at the interview date differ from those estimated in Holt (2010) as only the restricted linked sample for Wave 3 is considered here, rather than the average participation rate for the whole sample over Waves 1-3.

Table 4 – Labour market participation at the interview date compared with that for the reference period, working age non-students, 2004/05

	Not participating in reference period		Participating in reference period		Total	
	Count	**%**	**Count**	**%**	**Count**	**%**
Not participating at interview date	264,800	11.9	95,500	4.3	360,300	16.2
Participating at interview date	24,500	1.1	1,841,500	82.7	1,866,000	83.8
Total	**289,300**	**13.0**	**1,937,100**	**87.0**	**2,226,300**	**100.0**

Source: SoFIE/NZHIS Wave 3 Version 4, adjusted longitudinal weights, Statistics New Zealand

The definition of participation is important to prevent miscalculation of the number of hours lost. In this analysis lost hours are being estimated for the reference period, therefore participation is defined to be that throughout the reference period. The estimates of lost hours owing to absenteeism, presenteeism or working fewer hours as a result of ill health will therefore be based on all those who work for one week or more in the reference period, while the estimates of lost hours owing to not working owing to ill health will be based on those who do not work for any weeks in the reference period.[20] In the remainder of this analysis the term participating (and not participating) in the labour force refers to whether a person has undertaken one or more weeks of work in the reference period (or not); however, it should be noted that this definition is different from the more traditional definition of labour force participation. Testing different definitions of labour force participation is an area that could be developed further in future research.

5.1.2 Defining average hours worked

For those defined to be participating in the labour market who are also identified as having absenteeism, presenteeism or as having worked fewer hours owing to ill health, it is assumed that hours have been lost. The total number of hours worked in the reference period is used to determine the amount of hours lost. Statistics New Zealand provides a derived variable indicating the total number of hours worked in the annual reference period.[21] This is derived by Statistics New Zealand using the usual number of weekly hours worked in each labour market spell (or episode). The spell data is then summed to create a total estimate of annual hours worked.[22] As this estimate is based on *usual* rather than *actual* hours worked each week in a job, it should not already exclude the hours of work a respondent missed owing to illness.[23]

Table 5 shows the labour market status of people over the reference period along with the average number of total hours usually worked by health status. The average number of total hours worked falls as health state deteriorates.[24] The total number of hours usually

[20] Voluntary workers, the self-employed and casual workers are defined to be participating. While no income is lost if a voluntary worker is unwell, there would be lost output.

[21] In a few instances this is non-zero while the number of weeks employed or in paid work is zero. The hours for this small group are small and thus not included in this analysis.

[22] For the small group of respondents with gaps in the spell data the total annual hours worked is likely to be an underestimate.

[23] It is not known whether respondents report only paid hours or include unpaid hours, as the SoFIE question does not specify which hours should be included.

[24] A small number of people who report working in one or more weeks in the reference period are missing information on hours worked. Where this occurs, full-time hours (eight hours a day for five days a week) were assumed for the reported number of weeks worked.

worked by working age non-students is 3.76 billion.[25] This is usual annual hours worked. Actual annual hours worked will be lower than this. Actual annual hours would exclude hours not worked owing to illness, but also hours not worked owing to annual leave and unpaid leave etc.[26]

Table 5 – Number and proportion participating in annual reference period, and average and total hours usually worked, by health status, working age non-students, 2004/05

Health state at interview date	Number participating in reference period	Participation rate	Average annual hours usually worked	Total number of hours usually worked (bn)
Excellent	775,600	91.6	1,997	1.55
Very good	702,500	89.5	1,954	1.37
Good	366,800	82.8	1,872	0.69
Fair	78,500	65.3	1,689	0.13
Poor	13,200	42.5	1,446	0.02
Total	**1,937,100**	**87.0**	**1,941**	**3.76**

Source: SoFIE/NZHIS Wave 3 Version 4, adjusted longitudinal weights, Statistics New Zealand

Note: Columns may not sum to the total owing to a small number of people having missing self-rated health information and rounding.

For those who participate at some point in the interview period, the distribution of total hours worked is very different for those who are not participating at the interview date compared with those who are participating. The former group are likely to be people who change labour force status more regularly, or may have moved out of participation well before the interview date. In both cases this means they work for only some weeks of the reference period, therefore working fewer hours in total than those continuously participating.

From the total number of hours usually worked in the reference period the average number of hours worked each week, day and month of the reference period can be evaluated (using a set 365 days, 52.143 weeks and 12 months a year). Using this method creates average hourly measures across all job spells. These hourly figures are the basis for estimating the number of hours lost owing to ill health. This method is used as the specific days of the week, and number of daily hours actually worked, is unknown. Work patterns and arrangements vary substantially, therefore the average daily hours figures allow for weekend working.

For those people who are not participating owing to ill health it is assumed that, in the absence of ill health, they would work full-time; that is eight hours a day for five days a week. This approach is similar to that used in other research (Davis *et al.*, 2005).[27] It is acknowledged that this method will not result in an exact estimate of lost hours. In some cases it may underestimate lost hours and in some cases it may be an overestimate, but these variations should go some way to balancing each other out.

The main problem with this method of estimation is that total *usual* hours worked in the reference period are likely to have been affected by health. While the estimate of hours

[25] The proportion of hours affected reported in later tables is based on this total number of hours worked.

[26] As a result this estimate from SoFIE is above that of actual total hours worked for the whole workforce from the Household Labour Force Survey (HLFS) for a similar period (September 2004 to August 2005) of 3.42 billion.

[27] An alternative approach may have been to evaluate a proportion of these people at full-time hours and a proportion at part-time hours using the distribution of full-time/part-time hours.

should not already exclude time away from work owing to ill health, the number of usual hours contracted to work may already be lower as a result of ill health. The hours lost as a result of this are estimated as one component of the cost of ill health (see Section 5.4); however, the absenteeism and presenteeism estimates are based on the total annual hours usually worked rather than the hours that would usually have been worked in the absence of ill health. This means that the estimates of hours lost owing to absenteeism and presenteeism will be an underestimate. An alternative approach would have been to model the number of hours worked and use the results to predict the number of hours each person would work in the absence of ill health. These hourly estimates could then have been used to calculate absenteeism and presenteeism. This model could also have been used to predict the number of hours that those who would be predicted to participate in the labour force in the absence of ill health would work. However, models to predict the number of hours worked do not account for a high proportion of the variation in the data (only around 26%) and so the error introduced as a result of using this method may be more than that from using reported hours worked. Testing alternative definitions of hours worked is an area that could be developed further in future research.

5.1.3 Evaluating the cost of lost hours

The estimate of lost hours is then evaluated in monetary terms. One way to do this is to apply a given hourly rate; for example, the minimum wage applicable at this time. In March 2005, the mid-point of the interview period, this was $9.50. A limitation of using the minimum wage is that this is likely to understate the cost. An alternative level that could be used is the average full-time hourly pay level from the New Zealand Income Survey, which is specifically set up to collect wage information. Between April and June 2005 this was estimated to be $19.95.[28] In evaluating these lost hours in this way an assumption is made that, in the absence of ill health, there would be demand for these hours to be worked and that the increase in the labour supply would not have impacted negatively on wages.[29] The average full-time wage is used as opposed to the average for all employees, as part-time wages are known to be lower than full-time rates. One reason for this may be owing to ill health, so using the full-time average removes some of the possible ill health effect on part-time pay rates. While the full-time rates from the New Zealand Income Survey will include those whose hourly rate is affected by health, the impact of this is likely to be less than if the individual wage rate or the combined full-time and part-time wage rate had been used. The estimated cost of ill health for October 2004 to September 2005 calculated using this method is represented in later tables in terms of the proportion of GDP this represents. The closest period for comparison is September 2004 to August 2005. GDP for this period is estimated to be $151 billion in current prices.[30]

An alternative method to quantify the lost hours in monetary terms would be to evaluate them for each person at the hourly rate they earn. Where a person is not working, a wage equation, and the characteristic information in SoFIE, could be used to predict the value of pay they would potentially earn if they did participate. Accurately measuring hourly income in surveys is inherently difficult and requires in-depth questioning to ensure it is clear to the respondent what is required to be included and excluded in their answers. While SoFIE collects income information, the collection of this is not its primary purpose. Questions on income are seen as sensitive questions, resulting in a high rate of item non-

28 Source: New Zealand Income Survey. This is the hourly rate from wages and salaries. It is for the June 2005 Quarter (April-June 2005). It does not include wages from being self-employed. Full-time hours are defined as 30 hours or more per week.

29 This assumption can be made if the increase in earnings at the level of the individual translates into that individual spending more, and thus expanding the economy's aggregate demand.

30 Source: Statistics New Zealand.

response. While Statistics New Zealand has imputed values for income where it is missing, the effect of this imputation is unknown. Where people do respond to income questions they often provide hours and income from memory rather than consulting pay records. This, combined with the SoFIE survey questions not specifying what should be included and excluded in the responses, makes deriving accurate hourly rates difficult. As an example, for salaried employees it is not clear whether respondents will include or exclude unpaid hours. Using individual income information from SoFIE to evaluate the lost hours is also problematic as a person who has ill health may have a lower wage rate, partly reflecting lower longer-term productivity or the impact of health on skill development. For these reasons, while this method was considered, it was not used here.

This analysis could be developed in future to determine the impact of using different methods to quantify the lost hours. This could include using different hourly rates from the New Zealand Income Survey for groups within the SoFIE sample, such as occupation or qualification level, or evaluating the cost on a GDP basis rather than a wage basis.

5.2 Absenteeism (labour force participants)

5.2.1 Methods

One of the most visible impacts of ill health for those participating is lost output owing to days absent from work. Research in other countries estimates the cost of absenteeism using the data from administrative sources on employee absence or the results of surveys that directly ask people about the number of sick days they have taken (see Goetzel *et al.*, 2004; Davis *et al.*, 2005; DeVol and Bedroussian, 2007). The number of days away from work owing to illness is not asked in SoFIE. Further, there are limited sources of information of this nature in a New Zealand context. As a result, estimates of absenteeism have to be based on other questions asked in SoFIE and assumptions about the responses to these. As such, absenteeism has two components:

- **Component 1** – Assume that all those who are participating and who have ill health-related hospital inpatient appointments in the reference period have taken time off work to attend hospital. The hours lost per affected person are estimated using average daily hours usually worked multiplied by the length of the hospital appointment in days.[31]
- **Component 2** – Assume that those who are participating and who answer yes to the question: "(Other than anything that resulted from an injury) In the last 12 months, did an illness or health problem stop you doing your usual activities for more than a week?" have taken at least one week off work. The hours lost per affected person are estimated using average weekly hours usually worked.[32]

The total number of hours estimated to be lost owing to absenteeism is simply the sum of these two components. The formula, when these components are evaluated at the average hourly full-time rate, can be found in Figure D1 in Appendix D.

---

31 It is theoretically possible to identify the participation state and hours usually worked at the time of the hospital appointment. However, the format of the spell data and the gaps that exist between spells for some people, make its use difficult. Further, to make the unit record data confidential, start and end dates of hospital appointments have been adjusted by a set amount. Thus relating the hospital appointment to the exact labour market spell may not be possible for a small group of people. Even if the spell was successfully identified, it is not possible to know which days of the week a person works, or how many daily hours they work. For these reasons, and for consistency across the absenteeism and presenteeism estimates (for which the period affected is not known) the summary employment information for the whole reference period was used rather than the individual spell data.

32 The small number of people with missing information for the illness question are assumed to have no Component 2 absenteeism.

There are a number of limitations of using this method to estimate lost hours from being absent from work. Firstly, it is not known what days of the week a person works. Even if they are employed at the time of the hospital appointment, they may arrange the hospital appointment to be for a day on which they do not work, therefore resulting in no hours of work being lost. Averaging the annual hours over the number of days in a year (including weekends), makes the assumption that an "average" day's hours are lost for each day of the hospital appointment. For some who are able to work flexibly and therefore rearrange the hospital appointment to avoid clashes with work, this may be an overestimate; however, for most people, who do not work on every day of the year, this will be an underestimate, thus, to some degree, balancing each other out.

A second issue is that, while hospital inpatient appointments are known to last more than three hours, the amount of work time lost for those hospital inpatient appointments where the number of overnight stays is zero, is hard to ascertain. However, taking into account waiting and transportation time, it is reasonable to assume that at least one day's average hours are lost for this group.

Another issue is the limitation of the assumption made in Component 2. Those reporting that an illness has stopped them doing their usual activities for more than a week may or may not have taken a week off work. This will depend on the nature of their illness and of their work. In contrast, a person reporting that an illness has stopped them doing their usual activities for more than a week may have taken much more time than a week off work. Using these methods, only illnesses that result in absence for more than a week can be estimated (other than hospital inpatient appointments that are less than a week that are included). Many people will be away from work owing to illness for less than a week at a time. In fact it is known from other sources that the majority of employees take less than one week off work in total each year. So this methodology potentially misses a large group of people who take off only a small number of days as well as underestimating absenteeism for those who take off more than a week at a time. This limitation should be remembered when interpreting the results.

A further limitation is that some of those reporting an illness or health problem lasting more than a week may in fact be referring to the hospital appointment being accounted for in Component 1 of the absenteeism estimate. In this analysis both these costs were included for two reasons. Firstly, the absenteeism costs owing to hospital appointments do not account for additional days sick leave required after the hospital appointment as it cannot be ascertained where this would be needed. This inclusion of the crossover allows for this to occur. Secondly, the absenteeism owing to illness or health problems is a week or more; a known lower bound. Therefore the inclusion of this crossover, that is quantified in the next section, allows for affected activity that is more than one week in length.

A broader issue is that it is not known if the respondent took annual leave to attend hospital or when they were sick. If an employee takes annual leave, in theory, hours are not lost. However, as the majority of employees have been with an employer for more than six months they will be entitled to at least five days sick leave so it is sensible to assume they will use this and thus output will be lost.[33]

---

[33] Sick leave entitlement is at least five days after six months' service, increasing by an additional five days after each period of 12-month service. In some cases firms provide this sickness entitlement to employees as soon as employment is undertaken. Sick days can be rolled over to future years to a maximum of 20 days (although employers can allow more than this at their discretion). Further, employees may take unpaid leave to attend hospital appointments. This supports the assumption that output is generally lost as a result of hospital appointments.

This research could be developed further in future with different assumptions being made about the amount of time a person, reporting their activity to have been stopped owing to illness, has had off work.

5.2.2 Results

Overall results

In Table 1 it was found that 267,700 people aged 17 and over had one or more hospital inpatient appointments in the reference period. Table 6 shows that 162,200 people who are working age non-students have one or more ill health-related hospital appointments in the reference period. So working age non-students account for around 60.6% of those with one or more ill health-related hospital inpatient appointments. The remaining 105,500 people with ill health-related hospital inpatient appointments are students or those over working age.

The 162,200 working age non-students with hospital inpatient appointments represent 7.3% of all working age non-students. Therefore the proportion of this group with hospital appointments is lower than the proportion for all those aged over 17 years (7.3% compared with 9.2% in Table 1).

Focusing on working age non-students, the proportion with one or more ill health-related hospital inpatient appointments is higher for those who are not participating compared with those who are. Of those not participating, 13% had one or more ill health-related hospital inpatient appointments, compared with 6.4% of those who are participating. This latter group (124,500 people) is the group estimated to have Component 1 absenteeism. While the proportion of people with ill health-related hospital inpatient appointments is lower for those who are participating, owing to the larger number of people participating, this group represents 76.8% of the 162,200 people who have an ill health hospital inpatient appointment in the period.

Table 6 – Ill health-related hospital inpatient appointments compared with labour force participation in reference period, working age non-students, 2004/05

	Not participating		Participating		Total	
	Count	**%**	**Count**	**%**	**Count**	**%**
No hospital inpatient appointments	251,600	87.0	1,812,600	93.6	2,064,200	92.7
One or more hospital inpatient appointments	37,700	13.0	124,500	6.4	162,200	7.3
Total	**289,300**	**100.0**	**1,937,100**	**100.0**	**2,226,300**	**100.0**

Source: SoFIE/NZHIS Wave 3 Version 4, adjusted longitudinal weight, Statistics New Zealand

Notes:
1. Totals in Tables 6 and 7 not equal owing to non-response and rounding.
2. Columns may not sum to totals owing to rounding.

Table 7 shows that of those participating, 248,100 people (12.8%) had an illness that stopped their activities for more than a week. This is the group estimated to have Component 2 absenteeism. It represents 75.1% of the 330,500 working age non-students who report an illness that stopped their activities for more than a week in the period. The incidence of illness that stopped activities is much higher for those who are not participating. Of those not participating, around 82,400 people (28.5%) reported an illness that stopped their activities for more than a week in the period.

Table 7 – Illness compared with labour force participation in wave, working age non-students, 2004/05

	Not participating		Participating		Total	
	Count	**%**	**Count**	**%**	**Count**	**%**
No illness stopping activities for at least a week	206,900	71.5	1,689,000	87.2	1,895,900	85.2
Illness stopping activities for at least a week	82,400	28.5	248,100	12.8	330,500	14.8
Total	**289,300**	**100.0**	**1,937,100**	**100.0**	**2,226,300**	**100.0**

Source: SoFIE/NZHIS Wave 3 Version 4, adjusted longitudinal weight, Statistics New Zealand

Note: Columns do not sum to totals owing to rounding. A small number had missing information to the illness question. These people are assumed to have no illness stopping activities.

Table 8 is based on all those who have participated in the reference period. It shows 83.7% of people experience no absenteeism (1,621,000 people). It also shows that the crossover between those with absenteeism owing to inpatient hospital appointments (Component 1) and those with absenteeism owing to illness stopping activities for one week or more (Component 2). Of the different components of absenteeism, Component 2 is most common with 12.8% of participating working age non-students experiencing this, compared with 6.4% of people who experience Component 1 absenteeism. Around 56,800 people (2.9% of people) experience both components of absenteeism. In total, around 315,700 have some form of absenteeism; around 16.3% of all those who are participating.

Almost half of those with Component 1 absenteeism have some Component 2 absenteeism (45.6%). This is because people with hospital appointments generally self-report their health to be worse than those without hospital appointments. This is the group who may be referring to the hospital appointment when reporting an illness that affects their activities. However, the majority of people who are participating with Component 2 absenteeism *do not* experience any Component 1 absenteeism (77.1%) so the potential for double counting is not too much of a concern.

Table 8 – Absenteeism owing to hospital inpatient appointments (Component 1) compared with absenteeism owing to illness (Component 2), participating working age non-students, 2004/05

	No Absenteeism Component 2		Absenteeism Component 2		Total	
	Count	**%**	**Count**	**%**	**Count**	**%**
No Absenteeism Component 1	1,621,300	83.7	191,300	9.9	1,812,600	93.6
Absenteeism Component 1	67,600	3.5	56,800	2.9	124,500	6.4
Total	**1,689,000**	**87.2**	**248,100**	**12.8**	**1,937,100**	**100.0**

Source: SoFIE/NZHIS Wave 3 Version 4, adjusted longitudinal weight, Statistics New Zealand

Note: Columns may not sum to totals owing to rounding

Table 9 shows that around 2 million hours are estimated to be lost as a result of hospital inpatient appointments and around 8.2 million owing to an illness stopping activities for one week or more. So in total around 10.3 million hours are estimated to be lost owing to absenteeism; that is 0.3% of all hours usually worked each year. In terms of monetary value, the cost of those lost hours is estimated to be $0.206 billion; which is around 0.14% of GDP for the same period.

Table 9 – Absenteeism, participating working age non-students, 2004/05

	People affected		Hours lost		Cost	
	Number	Prop-ortion	Number (million)	Proportion of all hours	Evaluated at full-time hourly wage ($bn)	% of GDP
Absenteeism owing to:						
Component 1 - Hospital inpatient appointments	124,500	6.4	2.0	0.1	0.041	0.0
Component 2 - Illness stopping activities for one week or more	248,100	12.8	8.2	0.2	0.165	0.1
Total	**315,700**	**16.3**	**10.3**	**0.3**	**0.205**	**0.1**

Source: SoFIE/NZHIS Wave 3 Version 4, adjusted longitudinal weights, Statistics New Zealand

Notes:

1. The total number of people affected is not the sum of the numbers of people affected by each type of absenteeism. This is because a person can be affected by both types of absenteeism.
2. Columns may not sum to totals owing to rounding.

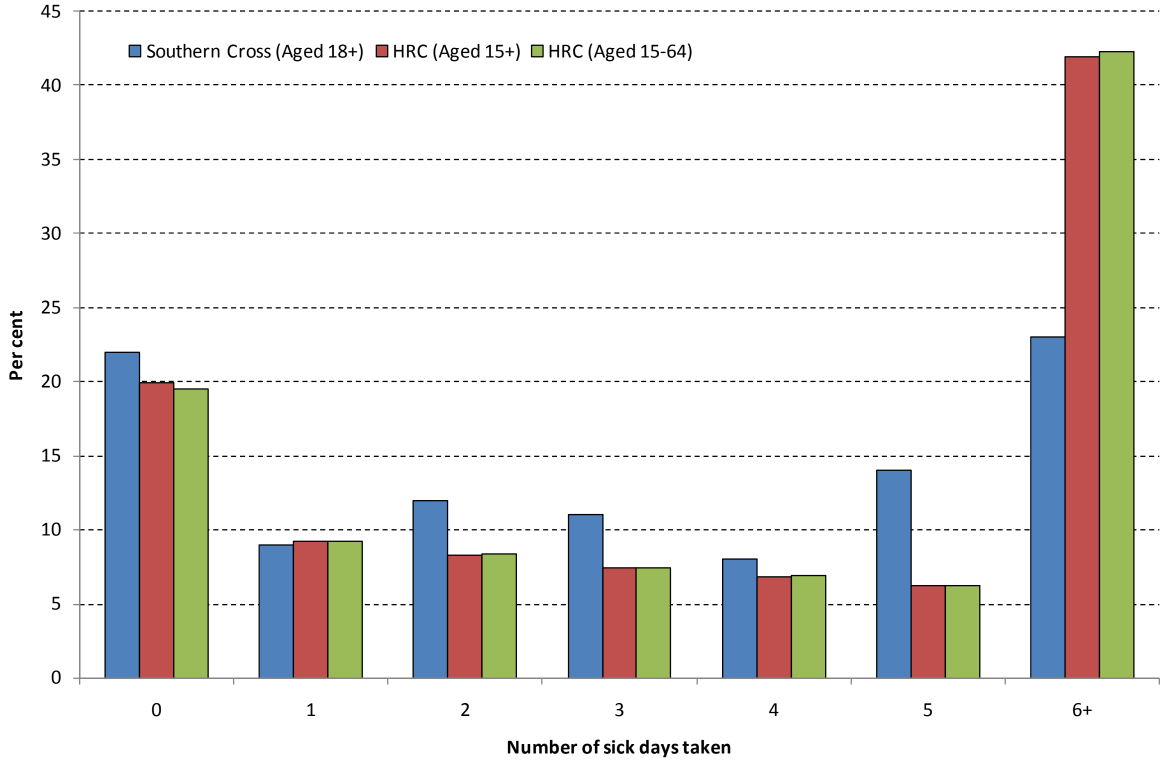
Owing to the two component estimation method used, the distribution of hours lost for those affected by absenteeism is bi-modal. There is an initial peak towards the bottom of the distribution where those with the most common length of hospital appointments and the most standard daily hours are found. There is then a second peak further down the distribution representing the people who have a week off work owing to illness and who work around the average weekly hours. Converting these hours into full-time equivalent days of work lost, using eight-hour days, indicates that around 1,287,250 days of full-time work may be lost owing to absenteeism.

As stated earlier it is known that the majority of people take less than five days sick leave each year. Figure 3 shows the distribution of sick days from recent research by Southern Cross Medical Care Society (2009). In 2008, around a fifth of employees had no “absenteeism” days. This is well below the 83.7% of people estimated from SoFIE to have no absenteeism days. The Southern Cross Medical Care Society research indicates that over half of people aged 18 and over (54%) had at most a week of “absenteeism” days, leaving 46% of people who had more than a week off work. This is well above the 12.8% of people estimated from SoFIE. The average time off work owing to illness was 4.2 days per year (Southern Cross Medical Care Society, 2009). Multiplying these numbers by the number of full-time equivalents suggests that around 5,880,000 full-time days are lost owing to absenteeism.[34] While this includes people over working age, students who also work and absenteeism owing to injury, it is still over four times higher than the estimate from SoFIE.

Another source of sick leave data in the New Zealand context is data from the Human Resource Capability Survey (HRC). This is a collection of administrative data from public service payrolls. It collects information on the number of sick or domestic leave days taken in the public service departments. Figure 3 shows the distribution of sick and domestic leave days for all those aged 15 and over and for those of working age for the year to June 2005 (the period most comparable with that for the SoFIE period being considered). The distribution of sick and domestic leave days is similar for those of working age and for all adults.

[34] The number of full-time equivalent workers is taken to be 1.4 million. This was quoted by Southern Cross in their press release (2009). They obtained this figure from Statistics New Zealand.

Figure 3 – Distribution of sick days



Sources: Southern Cross Medical Care Society (unpublished figures), weighted, Statistics New Zealand, 2008 and Human Resources Capability Survey (unpublished figures), New Zealand, Public Service, year to June 2005

Notes:

1. In the survey conducted by TNS on behalf of Southern Cross Medical Care Society, respondents were asked if they were employed. No definition of employed was given. The figures may therefore include students with part-time jobs, self-employed and those undertaking voluntary work if the respondent considered this as being employed.
2. Results from the annual Human Resource Capability Survey gather anonymous unit record data on all staff in public service departments as at 30 June from public service departments. The figures include domestic leave as well as sick leave.

The distribution of leave days at the bottom of the distribution is similar to that from the Southern Cross research, with around one-fifth of people having no absenteeism. The HRC results indicate that around 20% of people working in public service departments took no sick or domestic leave days in the year to June 2005. Thirty-eight percent of people took at most a week in sick leave, meaning 42% took six or more sick or domestic leave days. This is much higher than the proportion of people estimated to have taken six or more sick days in the Southern Cross research. Assuming that the distribution of sick leave in the public sector is the same as that in the private sector, the HRC data suggests that a total of at least 8,004,803 days may be lost owing to absenteeism and domestic leave. These days may be days taken off work to care for dependants. They also may be days worked by part-time workers (who do not lose a full day) and so cannot be directly compared with those estimated from SoFIE which are full-time equivalent days. For these two reasons it would be expected that the estimated days lost from HRC would be higher than those from Southern Cross or SoFIE estimates.[35] [36] Despite this the estimate is still over six times higher than the SoFIE estimate.

---

35 The number of days lost owing to absenteeism in the public sector has been scaled to cover all those in employment using SoFIE totals. To calculate the total, those with eight or more days are assumed to have lost just eight days. This figure is therefore likely to be lower bound.

36 The HRC results suggest that a smaller proportion of people take at most a week in sick leave than in the Southern Cross Research (around 38% compared with 54%), meaning a higher proportion take six or more days sick or domestic leave (42%

Evidence from the UK Office for National Statistics (in Black, 2008) suggests that around 2.4% of working time is lost in the UK owing to sickness absence, compared with the estimate of 0.3% from SoFIE.

While not identical to SoFIE in terms of the coverage and definitions, these sources suggest that the methodology used to estimate absenteeism from SoFIE misses a large amount of sick leave. This indicates that the estimate from SoFIE is at most a lower bound. There are several possible reasons for this. First, SoFIE does not capture those who take less than five sick days at one time. The method also seems to be underestimating the proportion of people who take five or more sick days based on comparisons with other data sources. These disparities are likely to be owing to a lack of information that directly addresses the subject of interest which necessitates the use of a proxy measure of activity limitations. In addition, even if it is assumed that the activity limitations information is an appropriate proxy to indicate those who have been absent from work sick, the assumption that each person reporting activity limitations for more than five days takes only five days off work will lead to underestimation of the amount of sick days taken.

Self-rated health

Table 10 shows that, as would be expected, absenteeism is related to self-rated health. As self-rated health state deteriorates the proportion of people with absenteeism increases. Around 10.1% of those with excellent health have some ill health absenteeism, compared with almost 69.7% of those with poor health. However, owing to the relative size of the self-rated health groups, around 87% of those with absenteeism are in excellent, very good or good health.

Table 10 – Total absenteeism by self-rated health, participating working age non-students, 2004/05

	Total number of people		Share with absenteeism (%)	Number with absenteeism		Hours lost	
	Count	%		Number	%	Count (mn)	%
Excellent	775,600	40.0	10.1	78,300	24.8	2.4	23.0
Very good	702,500	36.3	15.3	107,600	34.1	3.5	33.8
Good	366,800	18.9	24.2	88,700	28.1	3.1	30.0
Fair	78,500	4.1	40.4	31,700	10.0	1.0	9.9
Poor	13,200	0.7	69.7	9,200	2.9	0.3	3.2
Total	**1,937,100**	**100.0**	**16.3**	**315,700**	**100.0**	**10.3**	**100.0**

Source: SoFIE/NZHIS Wave 3 Version 4, adjusted longitudinal weight, Statistics New Zealand

Note: Columns may not sum to totals owing to rounding and missing self-rated health status.

Table 10 also shows that a higher proportion of those in good, fair or poor health experience absenteeism than the equivalent proportion they represent. Around 23.7% of participating working age non-students classify themselves as being in good, fair or poor health. However, those people account for 41% of those who experience absenteeism. Further, they account for 43.1% of the hours estimated to be lost owing to absenteeism. This suggests that not only are those in poorer health states more likely to experience absenteeism, they also lose more hours.

compared with 23%). The HRC figures include domestic leave, and therefore would be expected to be higher. Further, the HRC data is from payrolls as opposed to being recalled from memory, so may lead to higher, but more accurate, estimates. However, the differences may be owing to the distribution of sick leave in the public sector being different from that in the private sector.

Table 11 decomposes hours lost owing to absenteeism by self-rated health state. It shows that those who consider themselves to be in excellent health lose more hours owing to Component 2 absenteeism than those in poor health. Around 80.4% of all hours lost owing to absenteeism for those in excellent health are a result of Component 2. This compares with 69.2% of hours lost for those in poor health.

Table 11 – Absenteeism by self-rated health, participating working age non-students, 2004/05

	Component 1 - Absenteeism		Component 2 - Absenteeism		Total Absenteeism	
	Hours lost (mn)	**Proportion of hours lost for health state**	**Hours lost (mn)**	**Proportion of hours lost for health state**	**Hours lost (mn)**	**Proportion of hours lost for health state**
Excellent	0.5	19.5	1.9	80.4	2.4	100.0
Very good	0.6	17.2	2.9	82.8	3.5	100.0
Good	0.6	19.4	2.5	80.6	3.1	100.0
Fair	0.3	27.8	0.7	72.2	1.0	100.0
Poor	0.1	30.8	0.2	69.3	0.3	100.0
Total	**2.0**	**19.9**	**8.3**	**80.1**	**10.3**	**100.0**

Source: SoFIE/NZHIS Wave 3 Version 4, adjusted longitudinal weight, Statistics New Zealand

Note: Columns may not sum to totals owing to rounding and missing self-rated health status.

A binomial logistic regression model was used determine the relationship between a set of variables and absenteeism when other variables were held constant. A binomial logistic regression model is suitable as the dependent variable (A) is a binary response variable equal to one for those respondents who have absenteeism and zero for those who do not (the latter was the reference category). More information on logistic regression including the equation can be found in Section 5.5. The aim of this model was to determine relationships between variables and absenteeism, rather than explain all variation in the data.

The results of a logistic regression model can be found in Table E1, Appendix E. The results indicate that those in poorer health states, older workers, females, those who have had an injury lasting more than a week in the past 12 months, those who have deferred going to the their primary healthcare provider in the past 12 months owing to affordability and those who have not collected a prescription in the past 12 months owing to affordability are significantly more likely to experience absenteeism than those in the reference categories, when other variables are held constant. Of the factors that are significantly related to absenteeism, health status is by far the most significant predictor. For example, the odds of experiencing absenteeism for those in poor health are 15 times the odds for those in excellent health. This is in line with the results from US work which found that poor health status was the most significant predictor of missing work (Davis *et al*., 2005).[37] It is interesting that injury is significantly related to absenteeism even when attempts were made to remove injury-related hospital appointments and injury-limiting activities from the way absenteeism is estimated.

Highest qualification level

The average full-time hourly wage rate may overstate the cost if the lost hours are for those in less senior roles. Table 12 looks at the distribution of hours lost owing to

---

[37] In this work the definition of missing work includes domestic leave. The variables included in their model were also different from those in this work.

absenteeism compared with highest qualification level (as a proxy for occupation) in order to determine if this may be an issue. It shows that just under three-fifths of the hours lost are for those whose highest qualification level is post-school vocational or above. This suggests that the use of the average full-time hourly wage is unlikely to be an overestimate of the cost.

Table 12 – Absenteeism by highest qualification level, participating working age non-students, 2004/05

	Total number of people		Share with absenteeism (%)	Number with absenteeism		Hours lost	
	Count	%		Number	%	Count (mn)	%
Degree or higher	373,500	19.3	15.4	57,600	18.2	1.9	18.6
Post-school vocational	766,300	39.6	16.8	128,600	40.7	4.0	39.3
School qualification	500,100	25.8	15.4	77,200	24.5	2.6	24.8
No qualification	297,200	15.3	17.6	52,400	16.6	1.8	17.2
Total	**1,937,100**	**100.0**	**16.3**	**315,700**	**100.0**	**10.3**	**100.0**

Source: SoFIE/NZHIS Wave 3 Version 4, adjusted longitudinal weight, Statistics New Zealand

Note: Columns may not sum to totals owing to rounding.

5.3 Presenteeism (labour force participants)

5.3.1 Methods

Some workers will show up to work even when they do not feel well, perhaps to avoid sick days. As well as creating a heightened risk of injury or spreading of infectious diseases, workers are unlikely to be fully productive, resulting in lost output (Davis *et al*., 2005). Costs associated with such a concept are very difficult to quantify. Ideally there would be survey information about perceived productivity as a result of ill health. Numerous examples based on such questions can be found in US research. For example, estimates of illness-related presenteeism by Davis *et al*. (2005) were based around a survey question which asked the number of days they had been unable to concentrate at work because they were not feeling well or were worried about a sick family member. The estimate of lost output was based on the average earnings of these workers and the assumption that, owing to being unable to concentrate, they were working at half capacity. Similarly, research by Goetzel *et al*. (2004) compared estimates of presenteeism from a number of surveys which asked people, in various ways, how often their work performance had been affected by sickness. They used these results to create an average measure of presenteeism costs. The survey results used included those from surveys specifically undertaken with the aim of quantifying on-the-job productivity losses which result from poor health (Lerner *et al*., 2001). Other US research by the Milken Institute used the number of sick days reported in a US survey and adjusted this estimate using a factor from the Goetzel *et al.* research to estimate presenteeism (DeVol and Bedroussian, 2007). There are no questions in SoFIE that directly ask about the number of hours or days of work where productivity is affected owing to ill health or about the perceived productivity level for such hours. Therefore, to estimate presenteeism costs, assumptions have to be made using other questions asked in SoFIE.

To identify those who may be more productive in the absence of ill health the questions asking whether physical or mental health have interfered with daily activities in the past four weeks are used. For those who participate in the period, it is assumed that if a person's daily activities are limited or performed less well, or with less care, work activities will be affected. These "productivity-related" questions are listed in Figure 4 below.

The responses to each question for those participating were used to determine the proportion of working time in a four-week period where productivity was affected (this scale is similar to one of the surveys considered in the work by Goetzel *et al.,* 2004):

- All of the time – 100% of working time affected
- Most of the time – 75% of working time affected
- Some of the time – 50% of working time affected
- A little of the time – 25% of working time affected
- None of the time – 0% of working time affected.

Figure 4 – List of productivity-related survey questions in SoFIE

During the last four weeks, as a result of your physical health:

- How often did you cut down on the amount of time you spent on your usual daily activities?
- How often did you get less done than you would like?
- How often were you limited in the type of activities you could do?
- How often did you have difficulty doing your usual daily activities; for example, it took extra effort?

During the past four weeks, as a result of any emotional problems such as feeling depressed or anxious:

- How often did you cut down on the amount of time you spent on your usual daily activities?
- How often did you get less done than you would like?
- How often did you do your usual activities less carefully than usual?

Response choices:

- All of the time
- Most of the time
- Some of the time
- A little of the time
- None of the time

Three methods were used to summarise the responses to these questions into a measure of the proportion of hours worked at reduced productivity in a four-week period for each person:

- Method 1 – Maximum – the highest proportion across all questions was taken to be the proportion of hours that were worked at reduced productivity.
- Method 2 – Proportion – assumes each question response carries an equal weight. That is, the proportion of hours is equal to one-seventh of the response to the first question plus one-seventh of the response to the second question etc. This means

that those people with multiple responses of activity being limited have a higher proportion of hours worked at reduced productivity.[38]

- Method 3 – Principal Components Analysis (PCA) – this is a method for aggregating several indicators into a single measure. The method uses components to explain variation in the data. It aims to use as few components as possible to explain as large an amount of variation in the data as possible. This method was used despite the fact that for ordinal variables it means ignoring the discreteness of the variables. Examination of the data indicated that 83% of the variation in the data could be explained using two components. These two components were therefore used to predict the proportion of hours that were less productive for each person.[39] The results from PCA are initially on a standardised scale. These were therefore adjusted to form a continuous measure of the proportion of time that a person worked at reduced productive capacity which lies between zero and 100.[40]

For each method, multiplying the proportion of hours worked at reduced productivity each month by the number of average hours usually worked each month gives the estimated number of hours affected by presenteeism for each person in a month. Multiplying by 12 gives an annual estimate of the hours affected. The total number of hours lost is estimated based on assumptions about the level of reduced productivity owing to ill health.[41]

The level of reduced productivity to assume is difficult to ascertain, given the variation in jobs people undertake and the subjective nature of the questions on role limitations. The level of productivity will not be the same for all respondents even if they have identical responses to the “productivity” questions as it will depend on the job they undertake and the specific illness. The assumptions about level of reduced productivity used in earlier studies vary widely. For example, the US study by Davis *et al*. (2005) assumes workers are half as productive, while research in Australia used results from a survey to estimate that reduced effectiveness when at work owing to chronic pain was 14.2%, much lower than the assumption used in the US study (Access Economics, 2007).[42] When presenteeism is estimated by Southern Cross Medical Care Society (2009) in the New Zealand context, a reduction in productivity of 50% is assumed. As there does not appear to be a consensus approach, in this paper lost hours will be estimated assuming a range of different levels of productivity reduction based around the estimates from other research in this area. The levels of productivity considered will be:

- 85% of full productivity (Assumption 1)
- 75% of full productivity (Assumption 2)
- 50% of full productivity (Assumption 3).

The number of hours lost owing to presenteeism is then estimated using all combinations of the methods and assumptions in turn. That is, for each person the proportion of hours affected each month is multiplied by the number of usual hours worked each month

---

38 The small number of people with missing information for the productivity questions are assumed to have no presenteeism.

39 This was done for all people aged 17 and over; however, the resulting proportions were only used for working age, non-student participants.

40 The small number of people with missing information for the productivity questions are assumed to have no presenteeism.

41 Multiplying the monthly estimate by 12 assumes that workers suffer from presenteeism throughout the year; however, the productivity questions only relate to the four-week reference period. While an individual may not have experienced presenteeism during the rest of the year it is assumed that the four-week period is representative of the rest of the year, and while the same individuals may not experience presenteeism, the same proportion of people will be affected and the characteristics of people affected will be similar.

42 The standard error on this estimate was very large.

multiplied by the level of productivity. This estimate is multiplied by 12 to obtain an annual estimate. The resulting estimates are weighted and summed across all people and evaluated at the average hourly full-time rate. The formula can be found in Appendix D.

One limitation of these productivity-related questions is that it is not possible to split out responses that are a result of injury or pregnancy. As a result some of the loss of productivity estimated may be a result of this. To minimise this, a question asking about how much bodily pain interfered with usual daily activities in the last four weeks was not used as a productivity-related question, as this was more highly correlated with injury.

An alternative approach to estimating the reduced productivity owing to ill health may have been to use differences in wages (ie, when all else is equal, what the difference is in hourly pay between those with and without ill health).[43] However, along with the issues of estimating hourly pay from SoFIE that were discussed in Section 5.1.3, there are a number of problems with using pay as a proxy for productivity. Firstly, it assumes that people's pay reflects their performance which may not be the case. Secondly, even if a relationship between pay and health is established it is difficult to estimate the true magnitude of the relationship as there are lots of unobserved variables that may explain variations in pay. As such the differences in "productivity" that may be attributed to differences in pay may be the result of other unmeasured factors (for example, a year out to gain overseas experience, or a choice to work for lower wages, rather than ill health). Further, the reflection of ill health in pay levels may only exist for longer-term conditions, rather than short-term health problems. The shorter-term conditions, or the period after diagnosis of a longer-term condition, may be the least productive period, as for longer-term conditions it will take time to get the condition under control.

5.3.2 Results

Overall results

Table 13 shows that almost half of those who have participated in the reference period experience presenteeism to some degree for all three of the methods. Table 13 also provides estimates of the number of hours affected by presenteeism using the three different methods. These estimates vary widely depending on the method used. By far the largest estimated number of hours lost comes from Method 1 – Maximum (818 million), while the lowest number is from Method 3 – Principal Component Analysis (242 million). The reason for this result is that in moving from Method 1 through Method 3 the most common number of hours affected becomes lower; in other words the mode of the distribution is pushed to the left.

43 McKee and Suhrcke (2005) suggest that a negative impact of ill health on the wage rate would be expected.

Table 13 – Number of people and number of hours affected by presenteeism using different methods, aged 17 and over, 2004/05

	People affected		Number of hours worked at reduced productivity (mn)
	Number	Proportion	
Method 1 – Maximum	939,200	48.5	818.0
Method 2 – Proportion	937,200	48.4	335.0
Method 3 – PCA	937,200	48.4	242.0

Source: SoFIE/NZHIS Wave 3 Version 4, adjusted longitudinal weights, Statistics New Zealand

Note: Presenteeism information is missing for Method 2 and Method 3 if any component parts of the productivity questions have missing values. This small group of people are assumed to have no presenteeism. This is not the case for Method 1, hence the differences in the number of people affected.

Table 14 shows how the number of hours lost, and thus the cost of the lost hours, varies depending on the reduced level of productivity assumed. The lowest estimate of hours lost comes from Method 3 combined with an assumption of productivity reduction of 15% – 36.3 million hours lost ($0.724 billion).[44] The highest estimate comes from Method 1 combined with an assumption of a 50% reduction in productivity – 409 million hours lost ($8.160 billion). This large range in estimates indicates how difficult such a concept is to quantify. Even using the same method to estimate the number of hours affected, the estimates of hours lost are very sensitive to the assumption made about the level of productivity. For example, considering Method 2, for which the hours affected is between the estimates from the two other methods, the range of hours lost is from 50.3 million to 167.5 million, representing costs of $1.002 billion to $3.342 billion.

Returning to consider all methods and assumptions, Method 1 with Assumption 2 seem to provide an estimate near the mid-point of the range. Using this combination it is estimated that 204.5 million hours are lost; 5.4% of all hours worked. This equates to $4.080 billion; 2.7% of GDP. The further breakdowns provided in this section will be based around these estimates.

Assuming an eight-hour day, the number of full-time days lost owing to presenteeism ranges from 4.538 million to 51.125 million days. The research by the Southern Cross Medical Care Society (2009) found that the average number of days where employees went to work when they were too sick to be fully productive was 11.1 days per year. Multiplying these numbers by the number of full-time equivalents suggests that around 15,540,000 full-time days are lost owing to presenteeism.[45] While this includes people over working age and students who also work, this is within the range of presenteeism estimates from SoFIE.

---

44 This is the number of hours affected, 242 hours, multiplied by 0.15. To evaluate this in monetary terms it is multiplied by $19.95.

45 The number of full-time equivalent workers is taken to be 1.4 million. This was quoted by Southern Cross in their press release (2009). They obtained this figure from Statistics New Zealand.

Table 14 – Presenteeism costs under different productivity assumptions, aged 17 and over, 2004/05

	Hours lost		Cost	
	Number (mn)	Proportion of all hours	Evaluated at full-time hourly wage ($bn)	% of GDP
	Assumption 1 – people with presenteeism are 15% less productive			
Method 1 - Max	122.7	3.3	2.448	1.6
Method 2 - Prop	50.3	1.3	1.002	0.7
Method 3 - PCA	36.3	1.0	0.724	0.5
	Assumption 2 – people with presenteeism are 25% less productive			
Method 1 - Max	204.5	5.4	4.080	2.7
Method 2 - Prop	83.8	2.2	1.671	1.1
Method 3 - PCA	60.5	1.6	1.207	0.8
	Assumption 3 – people with presenteeism are 50% less productive			
Method 1 - Max	409.0	10.9	8.160	5.4
Method 2 - Prop	167.5	4.5	3.342	2.2
Method 3 - PCA	121.0	3.2	2.414	1.6

Source: SoFIE/NZHIS Wave 3 Version 4, adjusted longitudinal weights, Statistics New Zealand

Self-rated health

Table 15 shows that, as would be expected, presenteeism is highly related to self-rated health status. The results presented are for Method 1 with Assumption 2.[46] The proportion of those in each health state suffering from presenteeism increases as self-rated health state decreases. Around one-third of those people who participate who have excellent health have a non-zero amount of presenteeism, compared with over 95% of those in poor health. This supports the use of the productivity questions to determine health reduced productivity. Owing to the relative size of the self-rated health groups, of those who have a non-zero value of presenteeism, over nine-tenths have excellent, very good or good health.

Table 15 also shows that a higher proportion of those in good, fair or poor health experience presenteeism than the equivalent proportion they represent. Around 23.7% of participating working age non-students classify themselves as being in good, fair or poor health. However, those people account for 34% of those who experience presenteeism. Further, they account for 36.4% of the hours estimated to be lost owing to presenteeism being for people in those health states. This suggests that not only are those in poorer health states more likely to experience presenteeism, they may also lose more hours.

46 In Table 15 the columns other than the last two would be very similar for Methods 2 and 3. Further, the final column would be the same irrespective of the assumption combined with Method 1. The main difference in the method and combination assumptions would be in the hours lost column.

Table 15 – Presenteeism costs (Method 1 and Assumption 2) by self-rated health, participating working age non-students, 2004/05

	Total number of people		Share with presenteeism (%)	Number with presenteeism		Hour lost	
	Count	**%**		**Number**	**%**	**Count (mn)**	**%**
Excellent	775,600	40.0	33.0	256,100	27.3	51.8	25.3
Very good	702,500	36.3	51.7	363,400	38.7	78.3	38.3
Good	366,800	18.9	65.2	239,000	25.5	54.0	26.4
Fair	78,500	4.1	86.0	67,500	7.2	17.4	8.5
Poor	13,200	0.7	95.5	12,600	1.3	3.1	1.5
Total	**1,937,100**	**100.0**	**48.5**	**939,200**	**100.0**	**204.5**	**100.0**

Source: SoFIE/NZHIS Wave 3 Version 4, adjusted longitudinal weight, Statistics New Zealand

Notes:
1. Columns may not sum to totals owing to missing self-rated health information and rounding.
2. Method 1 – Maximum. Assumption 2% to 75% of full productivity.
3. The only column that would differ under different assumptions is the penultimate column. The hours lost by health state using Method 2 and Method 3 are distributed in similar ways to Method 1.

A binomial logistic regression model was used to determine the relationship between a set of variables and presenteeism when other variables were held constant. A binomial logistic regression model is suitable as the dependent variable (P) is a binary response variable equal to one for those respondents who have presenteeism and zero for those who do not (the latter was the reference category). More information on logistic regression including the equation can be found in Section 5.5. The aim of this model was to determine relationships between variables and presenteeism, rather than explain all variation in the data.

The results of a logistic regression model can be found in Table E2, Appendix E. The results indicate that those in poorer health states, females, those who have ever smoked, those who drink alcohol, those who have had an injury lasting more than a week in the past 12 months and those who have deferred going to their primary healthcare provider in the past 12 months owing to affordability are significantly more likely to experience presenteeism than those in the reference categories, when other variables are held constant. As with absenteeism, of the factors that are found to be associated with a significant increase in the chance of presenteeism, health state is by far the most significant predictor. For example, the odds of experiencing presenteeism for those in fair health are almost 12 times the odds for those in excellent health. There are also some factors that are found to be associated with a significant reduction in the chance of experiencing presenteeism. These include being older and living in any area of New Zealand other than Auckland (excluding Wellington).

Highest qualification level

Table 16 looks at the distribution of hours lost owing to presenteeism compared with highest qualification level. The results presented are for Method 1 with Assumption 2. The proportion of people with presenteeism is similar across all qualification levels. For example, at most 49.5% of those whose highest qualification level is school qualification have experienced some presenteeism, compared with 47.5% of those who have no qualifications. In terms of the distribution of those who experience presenteeism, almost two-fifths are those whose highest qualification is post-school vocational. A further 18.9% have a degree or higher.

Using the average full-time hourly rate to evaluate the lost hours would be of concern if the majority of hours lost were for those with low or no qualifications, as these people are more likely to be paid below the average full-time rate. However, in total almost three-fifths of those experiencing presenteeism have post-school vocational qualification or above. This suggests that using the average full-time rate to evaluate lost hours is unlikely to be an overestimate of the lost output.

Table 16 – Presenteeism costs (Method 1 and Assumption 2) by highest qualification level, participating working age non-students, 2004/05

	Total number of people		Share with presenteeism (%)	Number with presenteeism		Hour lost	
	Count	%		Number	%	Count (mn)	%
Degree or Higher	373,500	19.3	47.6	177,600	18.9	38.3	18.7
Post-school vocational	766,300	39.6	48.7	373,000	39.7	83.0	40.6
School qualification	500,100	25.8	49.5	247,300	26.3	50.8	24.8
No qualification	297,200	15.3	47.5	141,300	15.0	32.5	15.9
Total	**1,937,100**	**100.0**	**48.5**	**939,200**	**100.0**	**204.5**	**100.0**

Source: SoFIE/NZHIS Wave 3 Version 4, adjusted longitudinal weight, Statistics New Zealand

Notes:
1. Columns may not sum to totals owing to rounding.
2. Method 1 – Maximum. Assumption 2% to 75% of full productivity.
3. The only column that would differ under different assumptions is the penultimate column. The hours lost by qualification level under Method 2 and Method 3 are distributed in similar ways to Method 1.

Absenteeism vs. presenteeism

Table 17 shows that 900,100 of those participating (46.5% of people) suffer from no absenteeism or presenteeism. This is well above the 7% of people found to experience no absenteeism or presenteeism from the Southern Cross Medical Care Society research (2009). Around 5% only experience absenteeism, 37.2% only presenteeism and 11.3% are estimated to experience both. Of the 315,700 people with absenteeism, 218,000 (69.1%) have presenteeism. Of the 939,200 people with presenteeism, 218,000 (23.2%) have absenteeism.

Table 17 – Absenteeism compared with presenteeism (Method 1), participating working age non-students, 2004/05

	No absenteeism		Some absenteeism		Total	
	Count	%	Count	%	Count	%
No presenteeism	900,100	46.5	97,800	5.0	997,900	51.5
Some presenteeism	721,200	37.2	218,000	11.3	939,200	48.5
Total	**1,621,300**	**83.7**	**315,700**	**16.3**	**1,937,100**	**100.0**

Source: SoFIE/NZHIS Wave 3 Version 4, adjusted longitudinal weight, Statistics New Zealand.

Note: Columns may not sum to totals owing to rounding.

Table 18 shows the average hours worked and the average hours lost by presenteeism for those with and without absenteeism. The results presented are for Method 1 with Assumption 2. The average hours usually worked is lower for those with absenteeism than those with and without absenteeism (1,749 hours annually compared with 1,979 hours). Further, the average amount of hours lost to presenteeism is much higher for those with absenteeism than for those with no absenteeism (158 hours compared with 95 hours). These two factors combine and result in those with absenteeism losing a higher proportion of total usual hours worked than those without absenteeism (9% compared with 4.8%). This pattern remains irrespective of the methods and assumptions used and is in line with results from the Southern Cross Medical Care Society research which also found that those with absenteeism experience more presenteeism (Southern Cross Medical Care Society, 2009).[47]

Table 18 – Average hours lost owing to presenteeism by absenteeism (Method 1 and Assumption 2), participating working age non-students, 2004/05

	Average annual hours lost owing to presenteeism	Average usual annual hours worked	Average proportion of hours lost*
No absenteeism	95	1979	4.8
Some absenteeism	158	1749	9.0
Total	**106**	**1941**	**5.5**

Source: SoFIE/NZHIS Wave 3 Version 4, adjusted longitudinal weight, Statistics New Zealand

Note: * This is the proportion based on the average hours lost and hours worked rather than the average of each individual proportion.

5.4 Working fewer hours (labour force participants)

5.4.1 Methods

For those people participating in the labour force, ill health may result in them being contracted to work fewer hours. This is likely to be the case for those who have been in poor health for longer. For those participating in the labour force at any point in the reference period a basic linear regression model was used to estimate the relationship between self-rated health and the total number of hours a respondent worked in the annual reference period while controlling for a set of other variables.

A linear regression model quantifies the relationship between the number of hours worked and self-rated health, while holding all other variables constant. A list of these control variables together with the results is given in Appendix F. Figure 5 shows the form of the equation. In this case, the coefficients indicate the additional number of hours that are worked as a result of having certain characteristics relative to the base category. The coefficients can be positive or negative. For example, the negative coefficient for poor health indicates how many fewer hours are worked owing to being in poor health relative to being in excellent health. Those regression coefficients for the self-rated health indicators (β) which were significant were then used to estimate the additional number of hours that may be worked by each person in the absence of ill health (that is, if they had excellent health). These hours were evaluated at the average hourly full-time rate. The formula can be found in Appendix D.[48] [49]

47 Unpublished results.

48 Estimating a separate model for each gender was considered. However, while the coefficients for each health state for men were larger in absolute magnitude compared with women (ie, health appeared to have a greater relationship with hours than for women), only the coefficients for very good health were found to be significantly different from each other. In terms of estimated impact on hours, there was little difference when using a combined model with interactions between gender and partner and between gender and children, so a combined model was used for simplicity.

49 As in the rest of the report, only Wave 3 data for those who agree for their data to be linked to MoH information is used.

Figure 5 – Form of linear regression model

$$Totalhours_i = H_i^{'}\beta + X_i^{'}\lambda + u_i \qquad i = 1,\ldots,n$$

Where:

$Totalhours_i$ is a continuous variable for the total annual hours usually worked for the i th person.

$\beta, \lambda =$ vectors of regression coefficients.

$H_i =$ a vector of indicators of self-rated health state.

$X_i =$ a vector of explanatory variables.

$u_i =$ error term associated with person i.

One drawback of using a single equation linear regression model to estimate the relationship between health and the amount of hours worked is endogeneity. While giving an initial indication of the possible relationship between the total number of hours worked and health, the standard linear regression model does not account for endogeneity. Endogeneity occurs when the explanatory variables are not exogenous; true exogenous variables are not affected by the outcome variable or by other unobserved characteristics. This is likely to be a problem when trying to estimate the impact of health on hours worked. A fuller discussion of the reasons for this can be found in Holt (2010).

SoFIE is a longitudinal survey and therefore potentially allows more complex modelling techniques to be undertaken to try to account for endogeneity. This entails comparing changes in self-rated health to hours worked. However, the majority of people in SoFIE will not experience an acute health shock in the first three waves of the survey. When health shocks do occur and do not cause the respondent to leave the labour force, it may take time for this to feed through to total usual hours worked, possibly as respondents wait to see if their health improves or while changes in usual hours worked are negotiated with their employers. As a result the number of hours a person usually works in a given period is likely to be impacted, in the main, by their longer-term health.

Despite the availability of a limited number of waves of longitudinal data, the importance of the longer-term health level makes estimating the relationship between health and hours worked using panel models more difficult. Fixed effects panel models consider how changes in health are related to hours worked and there is no way of estimating the impact of the level of health. These models therefore do not provide a full picture. Random effects models do allow for an estimate of both health shocks and health level to be made; however, this requires assumptions to be made about the relationship between the unobserved variables that are correlated with self-rated health. If this assumption is not correct, or if correlation between the unobserved variables and health remains after this assumption has been made, the resulting model will be biased. As such the panel models do not completely meet the needs of this research. For these reasons a basic linear regression model, for which it is possible to obtain an estimate of the level of health, is used to estimate the relationships between health and the number of hours worked. When interpreting the results from the models it should be remembered that the impact estimates are likely to be upper bounds of the true impact as some of the relationship between hours and health status may be the result of other factors that it is not possible to control for in the analysis. Future analysis could attempt to incorporate panel models into the analysis to try to better understand the impact of health on participation.

As discussed in Holt (2010) there are issues with using self-rated health to measure health. One issue is that self-rated health is measured at the interview date and is compared with participation over the reference period. For some people the self-rated health state at the interview date will not be the same as the health state at different points in the reference period. Another limitation is that while self-rated health may be a more current and inclusive measure of health than variables such as the number of chronic diseases a person has been diagnosed with, it is more subjective and, as such, is open to potential bias. Further, self-rated health may not be entirely comparable between respondents. Some respondents may be consistently more optimistic in their health rating and others consistently more pessimistic. In addition the health base respondents use as a comparator when answering this question may change over time. For example, the SoFIE question on self-rated health does not ask respondents to rate their health relative to the health of other people of the same age. Some respondents may compare their health to that of others, but others may compare their current health to their past health. Given that this report focuses on those of working age, this ageing effect appears to be small. A further limitation of using self-rated health as a measure of health is that it is likely to include ill health as a result of injury. Despite these limitations this measure of health is used in this report in the absence of a viable alternative.

5.4.2 Results

Overall results

Full model results along with means and standard deviations for the variables considered can be found in Appendix F. Table 19 shows that if all those participating in the reference period had had excellent health, as opposed to their actual health state, 458,500 people would have worked more hours than they usually did.[50] This is 23.7% of all labour force participants. In total, 72.3 million hours are estimated to be lost; 1.9% of all hours usually worked. This is evaluated to be around $1.442 billion; 0.9% of GDP. This result assumes that, in the absence of ill health, there would be demand for these hours to be worked and that the increase in the labour supply would not have impacted negatively on wages.

Table 19 – Number of people affected and hours lost owing to ill health reducing hours worked, participating working age non-students, 2004/05

	People affected		Hours lost		Cost	
	Count	%	Number (mn)	Proportion of all hours	Evaluated at full-time hourly wage ($bn)	% of GDP
Total	**458,500**	**23.7**	**72.3**	**1.9**	**1.442**	**0.9**

Source: SoFIE/NZHIS Wave 3 Version 4, adjusted longitudinal weights, Statistics New Zealand

Note: Full model results used to derive these estimates can be found in Appendix F.

50 The small number of people with missing self-rated health information are assumed not to work fewer hours owing to ill health.

Self-rated health

Only health states that were found to be significantly different from those for excellent health in terms of hours worked were included in the calculation of hours lost owing to ill health. While the regression coefficient for very good health indicated that people in this health state worked less than those in excellent health, this coefficient was not found to be significantly different. These "lost hours" were therefore not included in the estimate as can be seen in Table 20.

Table 20 indicates that the majority of people who lose hours owing to ill health are in good health (80% of all people affected). However, the proportion of all lost hours for people in this health state is lower at 62.3% of all hours. This indicates that while the number of people affected is lower for those in poor and fair health, the number of hours lost increases as health decreases. So 2.9% of those people affected are in poor health but in terms of hours lost this group represents 8.3% of all hours.

Table 20 – Hours lost owing to ill health by self-rated health, participating working age non-students, 2004/05

	Total number of people		Share who work less (%)	Those who work less		Hours lost	
	Count	%		Number	%	Number (mn)	%
Excellent	775,600	40.0	0.0	0	0.0	0.0	0.0
Very good	702,500	36.3	0.0	0	0.0	0.0	0.0
Good	366,800	18.9	100.0	366,800	80.0	45.0	62.3
Fair	78,500	4.1	100.0	78,500	17.1	21.2	29.4
Poor	13,200	0.7	100.0	13,200	2.9	6.0	8.3
Total	**1,937,100**	**100.0**	**23.7**	**458,500**	**100.0**	**72.3**	**100.0**

Source: SoFIE/NZHIS Wave 3 Version 4, adjusted longitudinal weight, Statistics New Zealand

Note: Columns may not sum to totals owing to rounding and missing self-rated health information. The small number with missing self-rated health are assumed to be in excellent health.

Highest qualification level

Table 21 shows the proportion of people who are estimated to work fewer hours owing to ill health by highest level of qualification. The table only considers the relationship between self-rated health and hours worked; it does not adjust for qualification level. The proportion of hours lost is highest for those with no qualifications. Around 31.6% of those with no qualifications are estimated to work less owing to ill health, compared with 18.8% of those with a degree or higher. However, in terms of hours lost, over half are for people with post-school vocational qualifications or higher. Using the average full-time hourly wage rate to evaluate hours should balance out in terms of those who earn below this level and those who earn above this level.

Table 21 – Hours lost owing to ill health by qualification level, participating working age non-students, 2004/05

	Total number of people		Share who work less (%)	Those who work less		Hours lost	
	Count	%		Number	%	Number (mn)	%
Degree or Higher	373,500	19.3	18.8	70,200	15.3	9.9	13.7
Post school vocational	766,300	39.6	23.6	181,000	39.5	29.1	40.3
School qualification	500,100	25.8	22.7	113,400	24.7	17.5	24.2
No qualification	297,200	15.3	31.6	93,800	20.5	15.7	21.7
Total	**1,937,100**	**100.0**	**100.0**	**458,500**	**100.0**	**72.2**	**100.0**

Source: SoFIE/NZHIS Wave 3 Version 4, adjusted longitudinal weight, Statistics New Zealand

Working less vs. absenteeism and presenteeism

In Section 5.1.2, the issue of using a person's actual usual hours worked to estimate absenteeism and presenteeism was discussed. If the hours a person was contracted to work is affected by ill health then this will lead to absenteeism and presenteeism potentially being underestimated. It is therefore interesting to look at how many people who are estimated to suffer from absenteeism/presenteeism (this can be absenteeism, presenteeism or both) are also estimated to work fewer hours owing to ill health.

Table 22 shows that of those participating, 783,200 (40.4% of people participating) suffer from no absenteeism or presenteeism and do not work any fewer hours as a result of ill health. That leaves 1,153,900 who are estimated to have some absenteeism, presenteeism or work less owing to ill health. Around 35.9% only experience absenteeism/presenteeism, 6% only work fewer hours owing to ill health and 17.6% experience both absenteeism/presenteeism and work fewer hours. It is the latter group, 341,700 people or 17.6%, for whom absenteeism and presenteeism estimates may be underestimated as a result of usual hours worked being used to estimate these rather than the usual hours that would be worked in the absence of ill health.

Table 22 – Absenteeism and presenteeism (Method 1) compared with working less hours, participating working age non-students, 2004/05

	No fewer hours worked		Fewer hours worked owing to ill health		Total	
	Count	%	Count	%	Count	%
No absenteeism or presenteeism	783,200	40.4	116,900	6.0	900,100	46.5
Some absenteeism or presenteeism	695,300	35.9	341,700	17.6	1,037,000	53.5
Total	**1,478,500**	**76.3**	**458,500**	**23.7**	**1,937,100**	**100.0**

Source: SoFIE/NZHIS Wave 3 Version 4, adjusted longitudinal weight, Statistics New Zealand

Note: Columns may not sum to totals owing to rounding.

Table 23 shows the average hours of reduced work by whether a person has experienced absenteeism/presenteeism. The average is much higher for those with absenteeism/presenteeism than for those with no absenteeism or presenteeism. This indicates that not only are those with absenteeism/presenteeism more likely to work fewer annual hours owing to ill health, but the reduction in hours worked is also higher.

Table 23 – Average hours lost owing to presenteeism by absenteeism, participating working age non-students, 2004/05

	Average hours worked less
No absenteeism or presenteeism	17.3
Some absenteeism and/or presenteeism	54.6
Total	**37.3**

Source: SoFIE/NZHIS Wave 3 Version 4, adjusted longitudinal weight, Statistics New Zealand

5.5 Not participating (non-labour force participants)

5.5.1 Methods

For some people ill health may be the reason they do not participate in the labour force. Other studies have employed various methods to identify the group that may work in the absence of ill health. One example involved taking all those who do not work because of a health reason, or those who are not working and have a disability or chronic disease, as those who would potentially work in the absence of ill health (Davis *et al*., 2005). The analysis from SoFIE could have followed a similar approach, however, as well as the information on reason for not working not being available for all respondents; the impact of the health condition on preventing participation may be overestimated if other differences are not controlled for. In other words, even in the absence of the health condition not all of these people will participate owing to other factors. Therefore any estimate using this method would likely be an overestimate, or upper bound, of the impact. A modelling approach as used in Holt (2010), which identified a relationship between health and labour force participation, goes somewhere to control for these differences. This work provided estimates of the additional number of people who may participate in the absence of ill health. The standard binomial logistic regression model used in that report forms the basis for estimating the number of additional labour force participants and therefore the lost hours as a result of these people not working. As in the case of the analysis of labour force participants working fewer hours given in the previous section, this labour force participation analysis assumes that, in the absence of ill health, there would be demand for these hours to be worked and that the increase in the labour supply would not have impacted negatively on wages.

A binomial logistic regression model quantifies the relationship between self-rated health and labour force participation, while holding all other variables constant. The other control variables can be found in the full tables of results in Appendix G. A binomial logistic regression model is suitable as the dependent variable (L) is a binary response variable equal to one for those respondents who are participating and zero for those who were not participating (the latter was the reference category). The form of the equation can be seen in Figure 6. Maximum likelihood estimation was used to estimate the regression coefficients.[51]

Unlike the linear regression model that estimates the actual outcome, the binomial regression model estimates the probability of the outcome (that is, the probability of participating in the labour force). The probability of participating in the labour force (when

[51] Fitting models separately for each gender was considered. This method was rejected as the differences in gender seemed to be explained by the inclusion of gender and gender/partner and gender/children interactions. In fact, once these interactions are included in the model the variable for gender is not significant other than through these interactions.

all other variables are held constant at the mean of the sample) can be calculated for those in each health state. The probability of participating in the labour force can also be calculated for excellent health. The difference between the probability for each health state and that for excellent health is the marginal effect of that health state. Considering poor health, the probability of participating when in poor health minus the probability of participating when in excellent health, gives the marginal effect of poor health. If significant, the marginal effect for each health state is then applied to the number of people in that health state. This gives the additional number who may participate in this health state if they had in fact been in excellent health. The sum across all the significant health states is an estimate of the number who may participate in the absence of ill health (ie, if everyone had excellent health). To evaluate this in terms of lost hours it is assumed that each person would work eight hours a day for 260 days a year. These lost hours are then quantified in monetary terms using the average full-time hourly wage. The formula can be found in Appendix D.[52]

The model estimates the probability of participating for each person. The differences in probability can be used to estimate the additional number of people who may participate; however, this is done at a group level. It is not possible to know exactly which people may move into labour force participation in the absence of ill health. As such the wider characteristics of those who may participate in the absence of ill health (for example, highest qualification level or the presence of hospital inpatient appointments) is not known.

52 As in the rest of the report results are based on only Wave 3 data for those who agree for their data to be linked to MoH information.

Figure 6 – Form of binomial logistic regression model

$$L_i = 1(H_i^{'}\beta + X_i^{'}\lambda + u_i > 0) \qquad i = 1,\ldots,n$$

Where:

L_i is a binary response variable for participation for the i th person, equal to one if participating and zero otherwise.

$1(.)$ is an indicator function that takes the value one or zero according to whether the value in parentheses is true or false.

$\beta, \lambda =$ vectors of regression coefficients.

$H_i =$ a vector of indicators of self-rated health state.

$X_i =$ a vector of explanatory variables.

$u_i =$ error term associated with person i.

$e^{H^{'}\beta + X^{'}\lambda} =$ odds of success.

$$P(L=1 \mid X, H) = \frac{e^{H^{'}\beta + X^{'}\lambda}}{1 + e^{H^{'}\beta + X^{'}\lambda}} = \text{ probability of success.}$$

For similar reasons to those outlined in Section 5.4.1, this analysis will focus on basic logistic regression models rather than using panel models. It should be remembered that the results are therefore likely to form an upper bound of the cost. Again, future analysis could incorporate panel models into the analysis to try to better understand the costs associated with ill health.[53]

Those who are unemployed for the whole of the reference period are classified here as not participating. This is a very small group (around 0.4% of the population and 3.6% of those who are classified as not participating). If these people had been classified as participating then work hours lost would have been estimated in the absenteeism, presenteeism and working fewer hours sections despite them not undertaking any paid work. This section estimates the chance that those who are not participating would participate in the absence of ill health. For those who are continuously unemployed it is really estimating the chance a person would get paid work in the absence of ill health; implying that ill health may be one factor why they are continuously unemployed. The assumption is that, to some degree, health reduces the chance of the long-term unemployed obtaining paid work to the same extent as those who are inactive.

---

53 As well as just focusing on standard logistic regression models, the key differences with the analysis in Holt (2010) are:

- the different definition of participation (one or more weeks worked in the reference period compared with working or looking for work at the interview date)
- the focus on just Wave 3 data for matched consenters to make it comparable with the other analysis in this paper (in Holt, 2010, results were based on combined responses from all three waves)
- the models are weighted (this was not done in Holt, 2010 to make the basic models comparable with the panel models for which weighting was not possible).

5.5.2 Results

Overall results

The full model results, sample means and marginal effects can be found in Appendix G. Table 24 shows that in the absence of ill health (that is, if everyone had excellent health) an additional 42,300 people may participate; this is 14.6% of those not participating, 2.2% of those who are currently participating and 1.9% of all people. If these people had undertaken full-time work for the whole year, 88 million hours have been lost. That is 2.3% of all hours worked in the period. Evaluating this in monetary terms, $1.754 billion has been lost; around 1.2% of GDP.

Table 24 – Hours lost owing to ill health reducing participation, non-participating working age non-students, 2004/05

	People affected		Hours lost		Cost	
	Count	% of people not participating	Number (mn)	Proportion of all hours	Evaluated at full-time hourly wage ($bn)	. % of GDP
Total	**42,300**	**14.6**	**88.0**	**2.3**	**1.755**	**1.2**

Source: SoFIE/NZHIS Wave 3 Version 4, adjusted longitudinal weights, Statistics New Zealand

Note: Models are based on participating and non-participating working age non-students. Full model results can be found in Appendix G. Those with missing self-rated health are assumed to not be affected by ill health.

Self-rated health

Only health states for which participation was found to be significantly different from that for excellent health were included in the calculation of number of people affected and therefore the number of hours lost owing to ill health. While the regression coefficient for very good health indicated that people in this health state were less likely to participate than those in excellent health, this coefficient was not found to be significantly different from excellent health. These “lost hours” were therefore not included in the estimate as can be seen in Table 25.

Table 25 also shows that while the proportion of people who do not work owing to ill health increases as health state decreases, owing to the smaller number of people in the lower health states, the greatest loss of hours (or people affected) are those in good health. For example, 28.9% of those in poor health are estimated not to work owing to poor health; however, of all those who do not work owing to ill health only 21.3% are in poor health. On the other hand, only 3.9% of those in good health are estimated not to work owing to their health; however, of those who don’t work owing to their health, 41.2% are in good health. The distribution of people affected and hours lost over the health states is the same owing to the assumption that all affected people would work full-time for the whole of the year.

Table 25 – Cost of not working by self-rated health, working age non-students, 2004/05

	Total number of people		Prop of all people	Number who don't work owing to health		Hours lost	
	Count	%		Number	%	Number (mn)	%
Excellent	846,300	38.0	0.0	0	0.0	0	0.0
Very good	785,300	35.3	0.0	0	0.0	0	0.0
Good	442,900	19.9	3.9	17,400	41.1	36.2	41.2
Fair	120,200	5.4	13.2	15,900	37.6	33.0	37.5
Poor	31,200	1.4	28.8	9,000	21.3	18.7	21.3
Total	**2,226,300**	**100.0**	**1.9**	**42,300**	**100.0**	**88.0**	**100.0**

Source: SoFIE/NZHIS Wave 3 Version 4, adjusted longitudinal weight, Statistics New Zealand

Notes:
1. Models are based on participating and non-participating working age non-students. Full model results can be found in Appendix G.
2. Results are based on total number in each health state and the marginal effects evaluated at the mean of the whole sample. An alternative is to use the numbers not participating in each health state and the marginal effect evaluated at the means for this group. There was little difference between the two methods.
3. Columns may not sum to totals owing to rounding and missing self-rated health information. The small number with missing self-rated health are assumed to be in excellent health.

Highest qualification level

As discussed in Section 5.5.1 it is not possible to identify the exact people who would participate in the absence of ill health. For this reason a breakdown by highest qualification is not possible.

6 Summary and conclusions

This section first summarises the results of the paper. It then provides some discussion around the results, restating the limitations of the analysis.

6.1 Summary of results

This paper summarises the results of a cost of illness type study aimed at estimating some of the costs associated with ill health in New Zealand using evidence from the Survey of Family, Income and Employment (SoFIE). The focus is to try to obtain first estimates of costs that might be lost in a one-year period as a result of ill health. It is acknowledged from the outset that it is not possible using SoFIE to estimate all costs associated with ill health. As such this study aims to estimate the magnitude of just some of the associated costs (the costs considered are shown in Figure 2).

Table 26 provides a summary of the estimated magnitude of the different components of ill health that can be estimated using the SoFIE data. In monetary terms the total cost of the considered components is estimated to be $5.417 billion to $12,853 billion; between 3.6% and 8.5% of GDP for a similar period.

It is not known which people may participate in the absence of ill health and therefore it is not possible to know whether this group are also those with hospital inpatient costs. To determine the total number of people affected, it is assumed that all of the 42,300 people who may additionally participate in the labour force also experience hospital costs. Under this assumption, 1,301,700 are estimated to be affected by hospital inpatient costs or indirect costs; 44.8% of all those aged over 17 years.[54] [55] Owing to the assumption made, this is likely to represent a lower bound of the number of people affected.

Table 26 – Component costs of ill health: Estimates from SoFIE, 2004/05[1]

	Number of people affected	Work hours lost		Cost	
		Count (mn)	Proportion of total hours worked	Evaluated at full-time hourly wage ($bn)	% of GDP
Direct costs					
Ill health inpatient appointments	267,700	-	-	1.290	0.8
Indirect costs					
Absenteeism	315,700	10.3	0.3	0.205	0.1
Presenteeism	939,200[4]	36.3 – 409.0	1.0 – 10.9	0.724 – 8.160	0.5 – 5.4
Working fewer hours	458,500	72.3	1.9	1.442	0.9
Not being in the workforce	42,300	88.0	2.3	1.755	1.2
Total indirect[2]	1,196,200	206.9 – 579.6	5.5 – 15.4	4.127 – 11.563	2.7 – 7.6
Total [2 3]	**1,301,700**	**-**	**-**	**5.417 – 12.853**	**3.6 – 8.5**

Source: SoFIE/NZHIS Wave 3 Version 4, adjusted longitudinal weights, Statistics New Zealand

Notes:
1. Direct costs are for those aged 17 and over. Indirect costs are for working age non-students.
2. The total number of people affected is not the sum of the individual groups as the groups are not mutually exclusive; that is, some people can appear in both groups.
3. To estimate the total number of people affected it is assumed that the additional number of people who would participate in the absence of ill health are those with hospital appointments.
4. Count is from Method 1 – Maximum.

The main focus of this work is in estimating indirect costs. Around 1,196,200 working age non-students are estimated to suffer from one or more of the components of indirect costs estimated.[56] That is 61.8% of all working age non-students. The range of hours lost as a result of indirect costs is estimated to be 206.9 million to 579.6 million; 5.5% to 15.4% of total hours worked. Evaluated at the average full-time rate these hours equate to $4.127 billion to $11.563 billion; 2.7% to 7.6% of GDP.

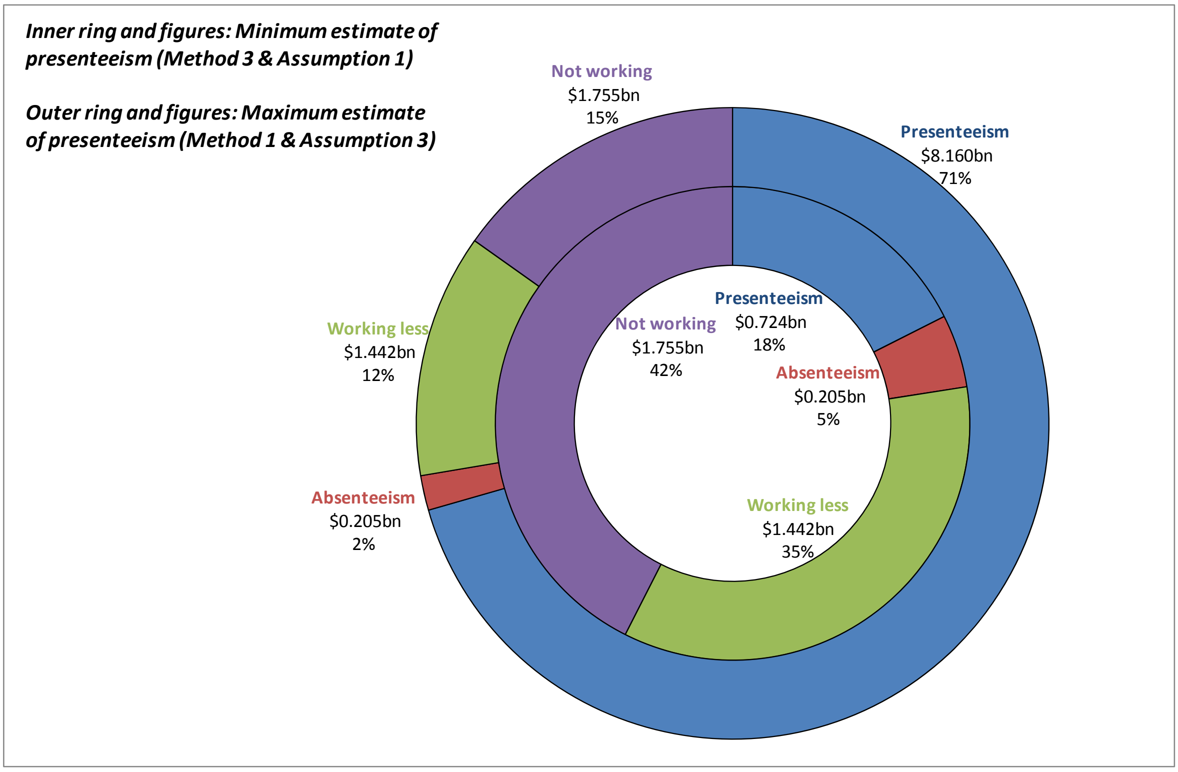
The large range in the estimate of hours lost is a result of the large range of the estimate for presenteeism and comes from using a range of methods about the proportion of hours affected by presenteeism and a number of assumptions about the level of reduced productivity. This illustrates what a difficult concept presenteeism is to estimate, and how sensitive the estimates are to the assumptions made. In terms of hours lost the estimate for presenteeism ranges from 39.3 million to 409.0 million hours; 1.0% to 10.9% of hours worked.

54 Indirect costs are only estimated for working age non-students.
55 Total number affected = number affected by indirect costs + number of students or people aged 65+ with hospital costs.
56 Total number affected by indirect costs = total participants affected by indirect costs + non-participants affected by indirect costs.

In monetary terms, the estimate for presenteeism ranges from $0.724 billion to $8.160 billion; 0.5% to 5.4% of GDP. Figure 7 shows the large impact the different presenteeism estimates have on the distribution of costs over the indirect components. The outer ring and figures show the distribution of indirect costs over the different components when the maximum estimate of presenteeism is used. The inner ring and figures show the distribution when the minimum estimate of presenteeism is used. The estimate of presenteeism ranges from 18% to 71% of total indirect costs.

Irrespective of the method and assumptions used to estimate presenteeism, the estimates of absenteeism are below those for presenteeism. It is estimated that 10.3 million hours were lost owing to absenteeism; 0.3% of all hours worked. In monetary terms these hours equate to $0.205 billion; 0.1% of GDP. This is only 2.5% to 28.3% of presenteeism. Figure 7 indicates that absenteeism accounts for only 2% to 5% of indirect costs. The methods used to estimate absenteeism are known to miss a large group of absenteeism. This under coverage is illustrated when the lost hours are converted into full-time equivalent days and compared with other sources of absenteeism in New Zealand. These figures therefore represent a lower bound of the cost of absenteeism. Despite this under coverage, and in line with other research, it seems likely that absenteeism will be generally smaller in size than presenteeism. Scaling up the estimate of days lost to be in line with results from the Southern Cross research suggests costs of absenteeism may be closer to $1 billion. This is below all but one of the estimates of presenteeism.

Figure 7 – Distribution of indirect costs of ill health from SoFIE using minimum and maximum estimate of presenteeism, working age non-students, 2004/05

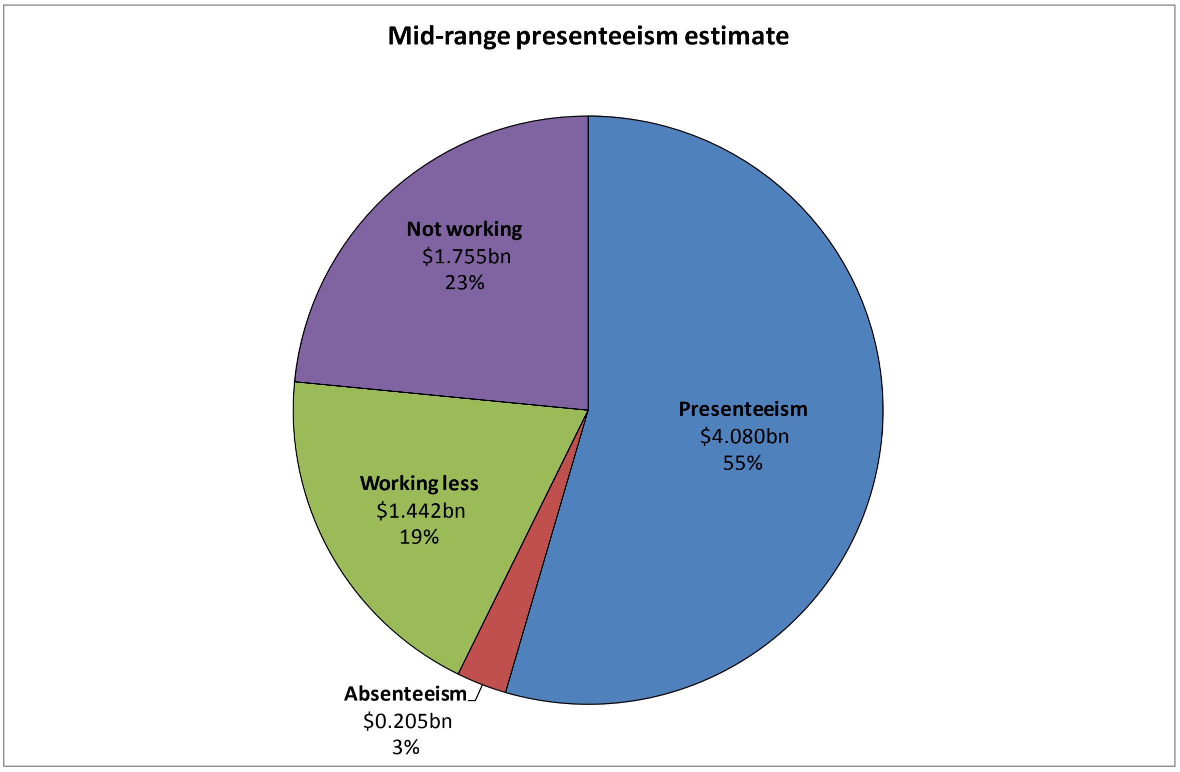


Source: SoFIE/NZHIS Wave 3 Version 4, adjusted longitudinal weights, Statistics New Zealand

Working fewer hours or not working at all owing to ill health are estimated to affect different numbers of people; 458,500 and 42,300 respectively (Table 26). However, in terms of lost hours (or costs) their impact is more similar; 72.3 million hours and 88 million hours respectively (or $1.442 billion and $1.755 billion respectively). Owing to the basic models used to derive these estimates, these figures are likely to represent the upper bounds for these components. In terms of where these costs sit within the overall indirect costs, the proportion represented is heavily reliant on the estimate of presenteeism; together the cost of working less or not working accounts for between 27% and 77% of all indirect costs (Figure 7).

Taking the estimate of presenteeism derived from Method 1 and Assumption 2 (this is closest to the mid-point of the range) the hours lost owing to indirect components are estimated to be 375.1 million; 10.0% of total hours. Indirect costs are estimated to total $7.483 billion; 4.9% of GDP. The distribution of these costs over the different indirect components can be found in Figure 8.

Figure 8 – Distribution of indirect costs of ill health from SoFIE using the estimate of presenteeism closest to the mid-point (Method 1 & Assumption 2), working age non-students, 2004/05



Source: SoFIE/NZHIS Wave 3 Version 4, adjusted longitudinal weights, Statistics New Zealand

The only direct costs included here are direct inpatient hospital costs, as these are the only costs that can be attributed to each respondent in SoFIE and related to labour force status. As such it is clear that the direct costs considered are only a small portion of wider health care costs. With that in mind, around 267,700 people (aged 17 and over) are estimated to have an ill health-related hospital appointment in the period. The cost of these hospital inpatient appointments is estimated to be $1.290 billion; 0.9% of GDP. This is below the majority of the estimated costs of presenteeism and those from working less and not working at all owing to ill health. It should be remembered that this is only one small component of direct costs.

The results presented are point estimates. As these estimates are based on survey data they are subject to sampling error. Sampling error occurs because data is only observed for a sample of the population rather than the whole of the population. Often when survey data is used, 95% confidence intervals are placed around point estimates to give an idea of the accuracy of the results. These confidence intervals reflect the upper and lower bounds between which you can be 95% sure that the true value for the population lies. Appendix H provides some indication of the sampling error associated with these estimates. The sampling variability of the estimates should be borne in mind when results from this analysis are used. While the estimate for presenteeism is a range, this range is a result of the different methods and assumptions used. It does not reflect differences owing to sampling error; in other words, it is not a confidence interval.

As well as the areas of future work that could be undertaken in this area mentioned throughout the report, there is also scope for further work looking at how costs vary for different groups (for example, by gender and by age group).

6.2 Conclusions and limitations

The indirect costs of ill health alone are estimated to be between $4.127 billion and $11.563 billion. To put these figures into context, at the top end of the estimate this is close to the total budget for Vote Health (of around $9.917 billion in 2004/05). The indirect costs covered are not the only indirect costs of ill health. As an example, there will be indirect costs as a result of people caring for sick dependants. These are not included in these estimates.

Alongside the indirect costs there are the direct costs of ill health. One small part of the direct costs, hospital inpatient costs, can be estimated from SoFIE, but this is just one part of a much larger cost. Vote Health is known to have been $9.917 billion in 2004/05. In addition to this cost there are health care costs and treatments that are privately funded.

Focusing on just absenteeism and presenteeism, together these components from SoFIE are estimated to be between $0.929 billion and $8.365 billion. This large range is owing to the varying measures of presenteeism that were employed. The only other known source of data in the New Zealand context is recent research by Southern Cross Medical Care. They estimate these costs to be over $2 billion (Southern Cross Medical Care Society, 2009), within the range of estimates from SoFIE. In line with the Southern Cross results, the SoFIE results suggest that presenteeism is the biggest of the two cost components, although the gap between the two estimates from the Southern Cross research is not as great as for the SoFIE estimates, in part indicating the under coverage of the absenteeism estimates in SoFIE (the Southern Cross estimates are $1.260 billion for presenteeism compared with $0.98 billion for absenteeism). Like the results in the Southern Cross research, results from SoFIE indicate that it is those who take days off sick who are most likely to experience presenteeism. The large estimate of presenteeism is also in line with that found in some US studies (DeVol and Bedroussian, 2007; Newton, 2000, in DeVol and Bedroussian, 2007). The output lost owing to presenteesim alone is thought to be immense. DeVol and Bedroussian estimated that 79% of the indirect costs are a result of individual presenteeism, compared with between 18% to 71% from this analysis.[57]

Looking at the lost hours by qualification level indicates that the lost hours are spread across different qualification levels, rather than being congregated on just a few qualification levels. This indicates that using the average hourly full-time rate to evaluate the hours lost is justified. In future work the costs could be broken down by occupation to better determine if this is the case. The average hourly rate for different occupations could then be used to evaluate the cost rather than the overall average hourly rate.

[57] Note their analysis included caregiver absenteeism and presenteeism.

It is interesting to consider who bears the indirect costs. The employee will bear the costs if they do not work or work less, while if a person is entitled to sick pay, the absenteeism and presenteeism costs are mostly borne by employers. Employers may also have to incur additional costs to hire temporary workers (eg, relief teachers). As a result, the estimates of the cost of ill health owing to not working or working less rest on an assumption that there would be demand for these additional hours to be worked, so the costs result in a loss in GDP rather than just a cost to the individual.

Another interesting finding of this research is a significant relationship between deferring going to the doctors or collecting a prescription owing to affordability in the past 12 months and absenteeism, with those deferring going to the doctors or collecting a prescription significantly more likely to experience absenteeism. Further, those who have deferred going to the doctors in the past 12 months owing to affordability are more likely to have suffered from presenteeism. However, it should be remembered that these results do not imply causality and that the results are based on models that do not control for endogeneity so may be subject to endogeneity bias. However, they do identify an area where further work could be undertaken.

In interpreting the results of this study, it is important to keep in mind a number of limitations, in terms of both the methodology behind its estimates and its implications. This study is a preliminary analysis of some of the costs of ill health. As discussed throughout this paper, there are data limitations owing to the nature of information available in SoFIE. For instance, SoFIE does not ask respondents questions that directly measure the extent of absenteeism and presenteeism, and thus these estimates are based on other related questions in SoFIE and assumptions about the responses to these questions. One component of absenteeism assumes that those who are participating in the labour force who responded that illness or a health problem stopped them from doing their usual activities for more than a week, have taken a week off work. However, this method does not account for those who may have taken less than a week at a time off work during the year. It also does not account for time taken off of more than a week. Other data sources suggest that the methodology used to estimate absenteeism from SoFIE misses a large amount of sick leave. A similar methodology limitation results in the wide range of the presenteeism estimate. This range reflects the several ways in which presenteeism can be measured based on information from SoFIE.

While this study does give an indication of the potential loss in GDP resulting from some of the costs of ill health, value-for-money conclusions *cannot* be drawn from these estimates for the following reasons. First, this research does not comment on the extent to which ill health is amenable to policy interventions. For instance, some of the estimated components rest on an assumption that it would be possible to improve the health status of those with less than excellent health. In the case of some types of illness, this would not be possible. Second, even if it were possible for all working age non-students to achieve excellent health status, this study does not provide an estimate of how much it would cost to achieve this, and therefore, does not offer any information about the value-for-money of policy interventions aimed at improving health status. Moreover, it is important to note that health care funding is driven by an individual's needs rather than an individual's labour market contributions. This study in no way suggests a policy link between an individual's publicly-provided health care and labour market contribution. Finally, this study does not comment on the effectiveness or the economic value of the current stock of health care interventions. Thus, it would be inappropriate to draw policy conclusions about health care interventions and spending based on this research.

References

Access Economics (2007) "The high price of pain: The economic impact of persistent pain in Australia." MBF Foundation in collaboration with the University of Sydney Pain Management Research Institute.

Black, Carol (2008) "Working for a healthier tomorrow: Review of the health of Britain's working age population." London, Presented to the Secretary of State for Health and the Secretary of State for Work and Pensions. <http://www.workingforhealth.gov.uk/Carol-Blacks-Review/Default.aspx>

Buchanan, T., Blick, G., and Isaac, R. (forthcoming) "Chronic disease and economic growth in New Zealand". New Zealand Treasury Policy Perspectives Paper.

Cai, L. and G. Kalb (2006) "Health status and labour force participation: Evidence from Australia." *Health Economics* 15(3): 241-261.

Currie, Janet and Brigette C. Madrian (1999) "Health, Health Insurance and the Labor Market." in O. Ashenfelter and D. Card eds *Handbook of Labor Economics.* (New York and Oxford: Elsevier Science): 3309-3416.

Davis, K., Collins, S.R., Doty, M. M., Ho, A. And Holmgren, A.L. (2005) "Health and productivity among U.S. workers." The Commonwealth Fund.

DeVol, R. and Bedroussian, A. (2007) *An unhealthy America: The economic burden of chronic disease*. (Santa Monica: Milken Institute).

Disney, R., Emmerson, C. and Wakefield, M., (2003) "Ill health and retirement in Britain: A panel data based analysis." The Institute for Fiscal Studies WP03/02.

Enright, J. and Scobie, G. M. (2010) "The effects of health and wealth on the labour supply and retirement decisions of older New Zealanders." New Zealand Treasury Working Paper 10/02.

Goetzel, Ron Z, Stacey R Long, Ronald J Ozminkowski, Kevin Hawkins, Shaohung Wang and Wendy Lynch (2004) "Health, absence, disability, and presenteeism cost estimates of certain physical and mental health conditions affecting U.S. employers." *Journal of Occupational and Environmental Medicine* 46(4): 398-412.

Holt, H. (2010) "Relationship between health and labour force participation: Evidence from SoFIE." New Zealand Treasury Working Paper 10/03.

Lerner, D., Amick, B. C., Rogers, W. H., Malspeis, S., Bungay., K and Cynn, D. (2001) "The work limitations questionnaire." *Medical Care* 39:72-85.

McKee, M. and Suhrcke, M. (2005) "Health and economic transition." *International Journal of Epidemiology* 34(6): 1203-1206.

Ministry of Health (2001) "The burden of disease and injury in New Zealand." Public Health Intelligence, Occasional Bulletin No. 1.

Ministry of Health (2006) "Hospital throughput 2004/05." DHB Funding and Performance.

Segel, Joel E. (2006) "Cost-of-illness studies: A primer." RTI-UNC Center of Excellence in Health Promotion Economics.

Southern Cross Medical Care Society (2009) “A New Zealand study into the hidden costs of unhealthy employees” www.healthybusiness.co.nz/.../633843675534877972CS_A5%20Research%20Publication%2009%20WEB.pdf

The Treasury (2008) “Pre-Election Economic and Fiscal Update.” October 2008.

Appendix A

Survey methodology

When SoFIE commenced in 2002 a total of 15,000 households were approached, of whom around 11,500 (77%) agreed to participate. In the initial interview, data was collected from around 22,000 individuals aged 15 and over. All respondents in the original sample (original sample members) are followed over time, even if their household or family circumstances change, forming a longitudinal sample. In later waves new cohabitants of the sample members are interviewed but asked only a reduced set of questions. These additional sample members are not followed if in future waves they no longer live with the original sample member. For these reasons only original sample members are included in this analysis. All SoFIE interviews are carried out face-to-face using computer assisted interviewing.[58] [59]

Statistics New Zealand provides a longitudinal weight which accounts for non-response and aligns the composition of the sample with that of the New Zealand population in October 2002. Interviews of SoFIE were conducted throughout the year with the sample spread evenly over the 12-month wave period. Each respondent is asked about the previous 12 months (their annual reference period). As a result of this continuous interviewing, there are 12 reference periods in each wave. Some variables collected in each wave of SoFIE, such as age, can be measured at the household interview date or at a point in the reference period. Figure B1, Appendix B shows the relationship between these dates for a hypothetical SoFIE respondent.

At the end of the SoFIE health module respondents were asked to give permission for their data to be linked to information on hospitalisations and cancer registrations held by the New Zealand Health Information Service (NZHIS) back to 1990. Of the 17,615 adults who responded to the survey in all three waves, 13,980 agreed for their information to be linked, which is a consent rate of 79.4%.[60] Statistics New Zealand provides the name and address information of consenters, along with some basic demographic information, such as date of birth, to MoH who uses this to match it to a National Health Information (NHI) number. This was done automatically were possible, with some manual matches made in more difficult cases. The NHI is a unique identifier that is assigned to every person who uses health and disability support services in New Zealand. It is recorded alongside any hospital visits and cancer registrations. Therefore if respondents are matched to a NHI number their hospital and cancer information can then be obtained. Those respondents who have had no contact with the health sector in New Zealand will have no NHI number and therefore will not be matched successfully. MoH estimates that the current coverage of NHI numbers is over 99% of the resident New Zealand population. As a result, this type of non-match would only be expected to be small. MoH believes that the majority of

---

58 Full details of the sampling design for SoFIE can be found here: http://www2.stats.govt.nz/domino/external/pasfull/pasfull.nsf/84bf91b1a7b5d7204c256809000460a4/4c2567ef00247c6acc256fab0082e7fc?OpenDocument . There was no formal oversampling of specific groups; however, stratification was used in the first stage of the sample selection to try to ensure sufficient representation in the survey from specific groups. The strata were defined according to region; urban/rural; high/low Maori population density and other socio-economic variables derived from the most recent census.

59 The full SoFIE questionnaire can be found here: http://www2.stats.govt.nz/domino/external/quest/sddquest.nsf/12df43879eb9b25e4c256809001ee0fe/14d945bb95ab2bbbcc256fb70077b3bb?OpenDocument .

60 More information on the characteristics of consenters and the reweighting method are available on request.

non-matches are likely to be owing to the respondents' details from SoFIE being significantly different from those held with the NHI, therefore preventing successful matching. Of those respondents who gave consent for their data to be matched, around 96% (13,475 respondents) were successfully matched resulting in a matched, consent rate of 76.5%.

While Statistics New Zealand provides weights to account for non-response and to benchmark results to census population estimates these weights do not adjust for non-consent and non-matching. If the weights provided by Statistics New Zealand were used to make estimates of counts from the matched consenters then these would be underestimates. Further other estimates such as means may be biased if the matched consenters differ from the non-matched or non-consenters in important ways. Analysis of the characteristics of consenters indicated consent rates differed significantly by various characteristics. Of these characteristics age, ethnicity and presence of chronic diseases were of particular importance for this work. The standard longitudinal weights were therefore adjusted to account for non-consent and non-matching using the weighting classes of age, ethnicity and the presence of chronic diseases.[61] [62] All analysis in this paper is based on the linked sample with the adjusted longitudinal weights used to realign the sample with the population (adjusted longitudinal weight) as opposed to the weights provided by Statistics New Zealand (standard longitudinal weights). Where possible the same analysis was carried out on the full sample using the standard longitudinal weights to determine the possible impact of using the restricted sample and there was little difference in the results.

Population and sample of interest

The questionnaire is only asked to those aged 15 and over. To ensure there is full information on respondents in all waves, the analysis is focused on those aged 15 and over at the end of the reference period in Wave 1 who remain eligible and respond in all three waves of the survey (adult longitudinal respondents). This is the balanced panel made up of 17,615 respondents in Waves 1–3; an unadjusted attrition rate of 20.5%. Once this is adjusted, to remove those people who move out of the scope of the survey or die, then the adjusted attrition rate is 17.2%. Those over working age or who are full-time students in each wave are excluded from the analysis. The results are therefore representative of the usual adult resident population of New Zealand who lived in private dwellings on the main islands of New Zealand in 2002/03 who are working age non-students. Around three-quarters of the 17,615 adult longitudinal respondents are working age non-students in Wave 1, 2 or 3.[63] [64] Just over 76% of respondents (13,475) consent for their responses to be linked to NZHIS information and were subsequently matched.[65] Statistics New Zealand provided the name and address information of consenters, along with some basic demographic information, such as date of birth, to MoH who used this to match it to a National Health Information (NHI) number. This was done automatically were possible (for 80% of cases), with some manual matches made in more difficult cases. The NHI is a unique identifier that is assigned to every person who uses health

---

61 Gender is not used as a weighting class as there was not a significant difference in matched, consent by gender.

62 More information on the adjusted weight is available from the author.

63 Those respondents with a missing value for any of the variables of interest in a particular wave are excluded from the models for data based on that wave. The number of missing values is small and analysis indicates they appear to be random.

64 Respondents can change status with regard to being a student or moving out of working age over the survey period. Therefore there are not always three responses for each respondent in the analysis even though the balanced panel is the starting point for the analysis (ie, the student/working age values criteria make the panel unbalanced).

65 The few respondents who reported working overseas were removed from this analysis as information on the hours worked was not provided.

and disability support services in New Zealand. It is recorded alongside any hospital visits and cancer registrations.

Figure A1 – SoFIE wave structure

Household is selected for interview – January 2003

Wave 1 (October 2002 to September 2003)

- Household interview date – usually a day in January 2003*
- Annual reference period – January 2002 to December 2002

Wave 2 (October 2003 to September 2004)

- Household interview date – usually a day in January 2004*
- Annual reference period – January 2003 to December 2003

Wave 3 (October 2004 to September 2005)

- Household interview date – usually a day in January 2005*
- Annual reference period – January 2004 to December 2004

* This date could be later if there are problems contacting respondent or arranging an interview; however, even if this moves into February or March the reference period will not change.

Limitations and strengths of SoFIE

The SoFIE data has a few limitations. As with all surveys, there is potential for non-response error – that is, errors because not all potential respondents take part in the survey. Unlike in cross-sectional surveys, non-response in longitudinal surveys has a second element as respondents can also choose whether to respond in each wave. If this non-response (known as attrition) is non-random (that is, the characteristics of those who do respond are systematically different from those who do not) then any inferences based on analyses of the data may be biased. In addition, where longitudinal data is linked to other sources, information is only observed for part of the sample (those who agree to the linkage) and these differences could also be non-random and potentially bias results. While there are differences in the response, consent and matching rates in SoFIE there are no groups of interest that do not contain any respondents. The weights (both the standard weights provided by Statistics New Zealand and adjusted longitudinal weights to take account of non-consenters) go some way to restore the distribution of respondents over the variables of interest and any bias as a result of this should be small when making inferences about the population as a whole.[66] However, it should be remembered that as a longitudinal survey those who are most unhealthy will die or move into institutions where they may not be able to be traced, meaning that the SoFIE population is likely to be healthier than the wider New Zealand population it represents.

A further limitation is that not all variables are available in all waves. An indicator for psychiatric conditions is only available in Wave 3 and an indicator for cancer is only available for the subset of respondents who agreed for their data to be matched to the Cancer Registrations database and were successfully linked. This potentially reduces the

[66] More information on sample attrition and consent in SoFIE and the adjusted longitudinal weights are available from the author.

sample size considerably if only Wave 3 matched consenters are considered. Making an assumption about the presence of psychiatric conditions for Waves 1 and 2 and coding the non-consenters' cancer status as "unknown" rather than missing goes someway to countering this problem, allowing analysis to be undertaken on all three waves rather than the restricted sample.

While SoFIE is a longitudinal survey there are only currently three waves of information. While this provides a wealth of information for variables that do not change very frequently, such as diagnosis of new diseases, modelling the impact of these variables with such a short span of data is difficult.

Lastly, if dependants of respondents have ill health or chronic diseases this may also affect the respondent's labour market participation. The SoFIE questionnaire does not allow "carers" to be identified except when the ill health of a family member is given as a reason for inactivity. In addition, when people do report the ill health of a family member as a reason for inactivity the cause of ill health cannot be identified or attributed to a specific chronic disease or illness. The effect of this on labour market participation is therefore not explored in this analysis.

Despite its limitations, SoFIE collects a wealth of information on respondents over time. This allows a range of labour market transitions, durations and repeat occurrences of respondents to be analysed. It allows comparison of labour market activity and disease presence at more than one point in time. Further, attempts to account for the presence of unobserved variables can be made given that the same respondent is being monitored over time. The linking of SoFIE data to cancer and hospitalisation information adds further depth to the SoFIE data and this additional information is subject to less reporting error than additional questioning of respondents.

While there are differences in response and consent rates by respondent characteristics, for a longitudinal survey of this kind the response and consent rates are high by international standards.

Appendix B

Appendix Table B1 – Health-related SoFIE variables used

Variable name	***Variable categories***	***Notes***
Self-rated health	• Excellent • Very good • Good • Fair • Poor	-
Physical health-related productivity questions	During the last 4 weeks, as a result of your physical health: • how often did you cut down on the amount of time you spent on your usual daily activities? • how often did you get less done than you would like? • how often were you limited in the type of activities you could do? • how often did you have difficulty doing your usual daily activities, for example, it took extra effort?	Response choices: • All of the time • Most of the time • Some of the time • A little of the time • None of the time
Mental health-related productivity questions	During the past 4 weeks, as a result of any emotional problems such as feeling depressed or anxious: • how often did you cut down on the amount of time you spent on your usual daily activities? • how often did you get less done than you would like? • how often did you do your usual activities less carefully than usual?	Response choices: • All of the time • Most of the time • Some of the time • A little of the time • None of the time
Activity stopped for one week owing to ill health	• Yes • No	Based on responses to question "(Other than anything that resulted from an injury) In the last 12 months, did an illness or health problem stop you doing your usual activities for more than a week?"
Smoked	• Yes • No	Whether a person has ever smoked
Drinks alcohol	• Yes • No	Whether drank alcohol in last 12 months
Deferred visit to PCP	• Yes • No	Whether deferred visit to primary care provider in last 12 months owing to affordability
Not collected prescription	• Yes • No	Whether not collected prescription in last 12 months owing to affordability
Injury	• Yes • No	Whether had injury lasting more than a week in last 12 months

Appendix Table B2 – NZHIS health-related variables used

Variable name	*Variable categories*	*Notes*
Hospital inpatient appointment in reference period (main measure used)	• Yes • No	Hospital inpatient information does not include records for those whose appointment is less than three hours (outpatients) or information on appointments with primary health care providers such as GPs. It will also not include private hospital treatments; treatments at these facilities are only included if the treatment is publicly funded.
Hospital inpatient appointment in SoFIE interview period (used briefly for comparisons with MoH figures)	• Yes • No	-
Length of stay for hospital appointments (days)	-	Continuous variable. In the original data this is the number of overnight stays so a day patient would have a length of stay equal to zero. To convert it to length in days one has been added to the original variable. The start and end dates of hospital appointments are altered by a set number of days to prevent disclosure but the length of stay is not affected.
Diagnosis-related Group (DRG)	-	DRG is a system to classify hospital cases into one of approximately 500 groups expected to have similar hospital resource use. DRGs are assigned by a "grouper" program based on ICD diagnoses, procedures, age, sex and the presence of complications or comorbidities. DRGs may be further grouped into Major Diagnostic Categories (MDCs).
International Statistical Classification of Diseases and Related Health Problems code (ICD code)	-	ICD codes classify diseases and a wide variety of signs, symptoms, abnormal findings, complaints, social circumstances and external causes of injury or disease. Every health condition can be assigned to a unique category and given a code, up to six characters long. Such categories can include a set of similar diseases. Hospital inpatient appointments as a result of ill health are defined as those that are not injury-related or pregnancy-related; that is, ICD codes A00-N99, R00-R94 and Z00-Z99.

Appendix Table B3 – Non-health-related SoFIE variables used

Variable name	*Variable categories*	*Notes*
Labour market participation (main measure used)	• Participating • Not participating	One or more weeks worked in the annual reference period. Voluntary workers, self-employed and casual workers are defined to be participating. If a person is unemployed for all weeks in the reference period they are defined to be not participating.
Labour force participation	• Participating • Not participating (inactive)	Labour force participation at the household interview date. Voluntary workers, self-employed and casual workers are defined to be participating. Those who are unemployed at the interview date are defined to be not participating.
Annual total hours usually worked	-	This is derived by Statistics New Zealand using the employment spell data.
Student	• Yes • No	If a respondent is still at school, reported that they were economically inactive as a result of being a student or studied full-time for nine or more months they are classified as students in this analysis.
Gender	• Male • Female	-
Region of residence	• Auckland • Waikato • Wellington • Rest of North Island • Canterbury • Rest of South Island	-
Born in New Zealand	• Yes • No	-
Ethnicity	• NZ/European • Māori • Pacific Islander • Other	Respondents could report more than one ethnicity. Where this occurred, respondents were assigned to a prioritised ethnicity in this order: Māori, Pacific Islander, Other, NZ/European.
Age at interview date	-	Continuous variable.
Aged 50 and over	• Respondent 50 or over • Respondent under 50	Age is at the interview date.
Highest qualification	• No qualification • School qualification • Post-school vocational qualification • Degree or higher	Some respondents reported a fall in qualification level between waves. Where this occurred the highest level of qualification was taken in later waves.
Studying	• No studying undertaken in reference period • Some studying undertaken	Each respondent is defined to have undertaken study if they report one month or more in which they have studied full-time or part-time towards a formal qualification in the reference period.
Partner	• Working partner • Non-working partner • Single	-

Variable name	*Variable categories*	*Notes*
Children	• No dependent children • Child(ren) minimum age <5 • Child(ren) minimum age 5-17	A dependent child is one who is under 18 years and not in full-time employment.
Family economic type	• One person with no dependent child(ren) • One person with dependent child(ren) • Couple with no dependent child(ren) • Couple with dependent child(ren)	A dependent child is one who is under 18 years and not in full-time employment.
Number of years in employment	-	Variable to note number of years in paid employment. Derived from the number of weeks in paid employment in the wave and the number of years reported to be in paid work before the first interview (this is assumed to be before the beginning of the annual reference period). If a respondent has at least one week in paid employment in the wave they are counted as having an additional year in paid employment.
Log household income less personal income	-	Continuous variable which is the log of the consumer price adjusted household income less the consumer price adjusted personal income. Personal income is removed owing to its correlation with labour force participation. There was a small number of respondents with negative personal/household income. This is possible if self-employment income is negative. As the number with negative income was very small these were imputed to be zero. One was added to all values to enable logs to be taken. Income was not adjusted to reflect family size/composition.

Appendix Table B4 – Non-SoFIE variables used

Source	*Variable name*	*Variable categories*	*Notes*
Household Labour Force Survey	Unemployment rate	-	Variable to denote national unemployment rate at the month of the household interview given the continuous interviewing method used in SoFIE.

Appendix C

Comparisons between SoFIE and MoH data

To get a sense of how well SoFIE estimates of hospital inpatient costs represent the true health care costs, the estimates from SoFIE were compared with figures reported by the MoH.

Hospital throughput information from the MoH is available for the period July 2004 to June 2005. Inpatient costs from SoFIE were therefore calculated for this period to allow comparisons to be made. The MoH figures were converted into a cost estimate using the national price for case weighted discharges. At an aggregate level the total cost estimates for all inpatient appointments from both sources was very similar (around $1.5 billion).[67] The comparison was also conducted for the previous period, July 2003 to June 2004. Here the difference between the two sources was slightly larger, indicating the variability in the types of treatments available and, thus, the variability of estimates from SoFIE. Despite these differences, at an aggregate level it seems sensible to use the SoFIE data to estimate inpatient hospital costs for the specific group of interest and be able to link this information back to the labour market information.[68]

Understanding the wider health care costs

Given the size of the Government's health budget this estimation of the cost of hospital treatments may seem small. In order to understand where this cost fits in the overall health budget the total estimated cost of hospital appointments in the July 2004 to June 2005 period for all ages ($2.7 billion) will be compared with information from the Budget 2004 Statement of Appropriations.

The total amount of public funds spent on health is known as Vote Health. In the period July 2004 to June 2005 Vote Health totalled $9.917 billion. Figure C1 shows how this is distributed.

The majority of Vote Health ($7 billion in July 2004 to June 2005) goes to the sector via the District Health Boards (DHBs). The Provider Arm mainly involves the provision of hospital services but it also includes community services, public health services and assessment, treatment and rehabilitation services. The estimated total hospital costs for adults based on the MoH Hospital Throughput information ($2.7 billion) will be funded from this $7 billion. The remaining $4.3 billion is used to fund other services; for example, hospital treatments for those under 15 years old, mental health, further post-hospital follow-up care in the community (for example, rehabilitation services and disability support and older people services).

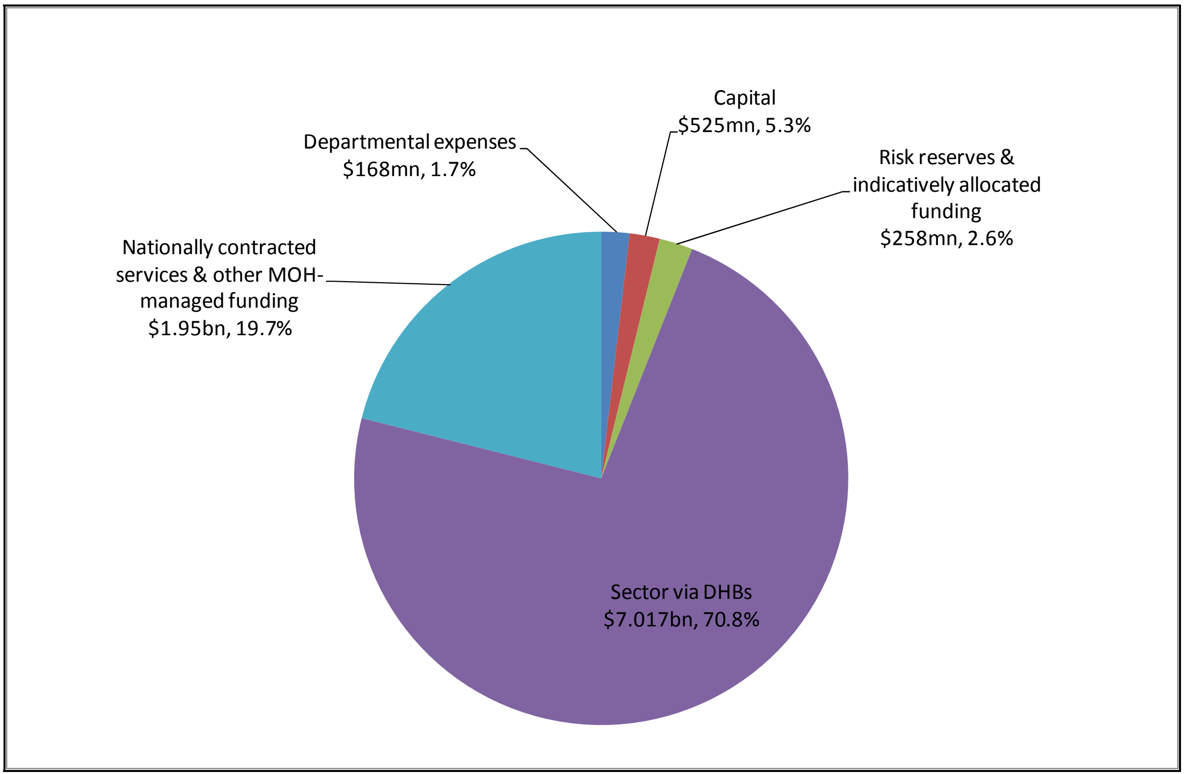
In summary, the figures for hospital inpatient treatments and outpatient treatments are only a small component of overall Vote Health. While it is acknowledged that the SoFIE/NZHIS estimates do not cover all aspects of ill health it is not possible to distribute the other health care costs on a per person basis. It is therefore not possible to ascertain what proportion of the other costs are for ill health rather than pregnancy or injury; what

67 This includes appointments for injury and pregnancy.

68 It should be remembered that the SoFIE estimates are for those aged 17 and over while the MoH estimates are for those aged 15 and over.

proportion are for those aged 15 and over; or similarly whether the costs would remain in the absence of ill health; for example, national campaigns and screenings may still exist in some form in the absence of ill health. Therefore these costs are not considered further in this paper.

Figure C1 – Vote Health overview, July 2004 to June 2005



Source: Estimates from Budget 2004 Statement of Appropriations Vote Health

Appendix D

Figure D1 provides the formulas used to estimate the different component costs of ill health. More explanation of the methods behind these formulas can be found in the main body of the paper.

Figure D1 – Summary of formula used to calculate cost of ill health using SoFIE

Hospital inpatient costs

$$= \sum_{i=1}^{n^{direct}} w_i(\text{Total cost weight of inpatient appointments*National Price})_i$$

Where:

n^{direct} is the total number of people in the sample for direct costs (aged 17 and over).

$i = 1,\ldots,n$ identifies the i th person in the sample.

w_i is the adjusted longitudinal weight for the i th person.

The cost weight is zero if a person has no hospital appointments.

Absenteeism costs = Absenteeism Component 1 + Absenteeism Component 2

Where:

$$\text{Absenteeism Component 1} = pay^{ft} \sum_{i=1}^{n^{indirect}} w_i (h^{daily} * los^{days})_i$$

$n^{indirect}$ is the total number of labour force participants in the sample for indirect costs (participating, working age, non-students).

pay^{ft} is the average full-time hourly rate from the NZ Income Survey ($19.95).

h^{daily} is the average daily hours worked in the reference period.

los^{days} is the length of stay across all hospital inpatient appointments (in days). This is zero if a person has no hospital inpatient appointments.

Figure D1 continued – Summary of formula used to calculate cost of ill health using SoFIE

And:

$$\textbf{Absenteeism Component 2} = pay^{ft} \sum_{i=1}^{n^{indirect}} w_i (h^{weekly} * illness)_i$$

h^{weekly} is the average weekly hours worked in the reference period.

$illness$ is an indicator of whether an illness has stopped activity for at least a week in the last 12 months. For each person this is equal to 1 if the response is yes and 0 otherwise.

$$\textbf{Presenteeism costs} = pay^{ft} \sum_{i=1}^{n^{indirect}} w_i (12 * h^{monthly} * prod^{\%} * prod^{level})_i$$

Where:

$h^{monthly}$ is the average monthly hours worked in the reference period.

$prod^{\%}$ is the proportion of hours worked at reduced productivity, calculated using Methods 1 to 3.

$prod^{level}$ is the level of productivity for those hours worked at reduced productivity, based on Assumptions 1 to 3.

$$\textbf{Cost of working less} = pay^{ft} \sum_{i=1}^{n^{indirect}} w_i hours_i$$

Where:

$hours$ is the additional hours a person would work annually in the absence of ill health. Results are from a regression model. It is non-zero for those affected and zero for all other people.

$$\textbf{Cost of not working} = (pay^{ft} * 8 * 260) \sum_{j=1}^{5} work^{health}{}_j * health_j$$

Where:

$j = 1, \ldots, n$ identifies the health state.

$work^{health}$ is the marginal effect for each health state, evaluated at the sample means. This is zero if the health state is excellent (or not significantly different from excellent). Results are from a logistic regression.

$health$ is the weighted number of people in the equivalent health state.

Appendix E

Appendix Table E1 – Estimated coefficients for absenteeism – logistic regression model – 2004/05

	Coefficient	Standard error	P value	95% confidence intervals	
				Lower	Upper
Self-rated health (base=excellent)					
Very good	0.424***	0.085	0.000	0.258	0.591
Good	0.936***	0.093	0.000	0.754	1.119
Fair	1.626***	0.141	0.000	1.349	1.903
Poor	2.710***	0.303	0.000	2.115	3.304
Sex (base=male)					
Female	0.366***	0.069	0.000	0.232	0.500
Region (base=Auckland)					
Waikato	-0.070	0.127	0.579	-0.319	0.178
Wellington	0.022	0.109	0.839	-0.192	0.237
Rest of North Island	-0.126	0.104	0.225	-0.329	0.077
Canterbury	-0.115	0.105	0.274	-0.322	0.091
Rest of South Island	-0.135	0.112	0.227	-0.354	0.084
Ethnicity (base=NZ/European)					
Māori	0.102	0.104	0.329	-0.103	0.307
Pacific Islander	-0.075	0.211	0.724	-0.489	0.340
Other	-0.142	0.169	0.399	-0.473	0.188
Age at interview date	0.006**	0.003	0.050	0.000	0.012
Highest qualification (base=school qualification)					
Post-school vocational qualification	0.110	0.086	0.202	-0.059	0.278
Degree or higher	0.183*	0.103	0.075	-0.019	0.386
No qualification	-0.052	0.108	0.633	-0.264	0.160
Log other household income	-0.072*	0.040	0.074	-0.150	0.007
Family economic type (base=1 person no dependent children)					
1 person with dependent children	-0.190	0.166	0.251	-0.516	0.135
Couple no dependent children	0.002	0.092	0.987	-0.179	0.182
Couple with dependent children	-0.129	0.087	0.135	-0.299	0.040
Smoked (base=never smoked)	0.087	0.070	0.214	-0.050	0.225
Drinks alcohol (base=yes)	0.154	0.112	0.170	-0.066	0.373
Injury (base=no)	0.444***	0.093	0.000	0.261	0.627

	Coefficient	Standard error	P value	95% confidence intervals	
				Lower	Upper
Deferred visit to PCP (base=no)	0.233**	0.094	0.013	0.049	0.418
Not collected prescription (base=no)	0.430***	0.132	0.001	0.170	0.689
Constant	-1.929***	0.478	0.000	-2.866	-0.992
Model summary statistics					
Number of observations			8,675		
Chi-squared			410.22		
Log-likelihood			-3,599.62		
Pseudo R^2			0.0665		

Source: SoFIE Wave 3 Version 4, adjusted longitudinal weights, Statistics New Zealand

Notes:

1. Based on original sample members who respond in Waves 1-3; who consent for their data to be linked to NZHIS information; are participating; and who are aged over 15 at the end of the reference period in Wave 1. Full-time students and those 65 years of age and over are excluded. All variables were included in the model and significant and insignificant variables or variable categories are kept in for completeness.
2. *Significant at the 90% level. **Significant at the 95% level. ***Significant at the 99% level.
3. The likelihood of experiencing absenteeism was modelled. No absenteeism is the base category.

Appendix Table E2 – Estimated coefficients for presenteeism – logistic regression model – 2004/05

	Coefficient	Standard error	P value	95% confidence intervals	
				Lower	Upper
Self-rated health (base=excellent)					
Very good	0.779***	0.059	0.000	0.663	0.895
Good	1.323***	0.074	0.000	1.178	1.468
Fair	2.479***	0.185	0.000	2.117	2.841
Poor	3.944***	0.774	0.000	2.426	5.461
Sex (base=male)					
Female	0.323***	0.053	0.000	0.219	0.426
Region (base=Auckland)					
Waikato	-0.358***	0.105	0.001	-0.563	-0.153
Wellington	-0.053	0.085	0.532	-0.220	0.114
Rest of North Island	-0.340***	0.078	0.000	-0.493	-0.187
Canterbury	-0.225***	0.080	0.005	-0.382	-0.068
Rest of South Island	-0.202**	0.088	0.021	-0.374	-0.031
Ethnicity (base=NZ/European)					
Māori	-0.020	0.090	0.825	-0.197	0.157
Pacific Islander	0.190	0.163	0.244	-0.129	0.509
Other	-0.134	0.128	0.295	-0.385	0.117
Age at interview date	-0.005**	0.002	0.036	-0.010	0.000
Highest qualification (base=school qualification)					
Post-school vocational qualification	0.013	0.066	0.849	-0.116	0.141
Degree or higher	0.126	0.079	0.110	-0.029	0.280
No qualification	-0.228***	0.085	0.008	-0.395	-0.060
Log other household income	-0.049	0.033	0.133	-0.113	0.015
Family economic type (base=1 person no dependent children)					
1 person with dependent children	-0.149	0.131	0.258	-0.406	0.109
Couple no dependent children	-0.045	0.076	0.549	-0.194	0.103
Couple with dependent children	-0.147**	0.068	0.032	-0.281	-0.013
Smoked (base=never smoked)	0.150***	0.054	0.006	0.044	0.256
Drinks alcohol (base=yes)	0.220**	0.095	0.021	0.033	0.406
Injury (base=no)	0.737***	0.080	0.000	0.579	0.894
Deferred visit to PCP (base=no)	0.540***	0.079	0.000	0.384	0.695

	Coefficient	Standard error	P value	95% confidence intervals	
				Lower	Upper
Not collected prescription (base=no)	0.143	0.126	0.254	-0.103	0.389
Constant	-0.123	0.389	0.751	-0.886	0.639

Model summary statistics	
Number of observations	8,675
Chi-squared	782.73
Log-likelihood	-5,402.30
Pseudo R^2	0.1011

Source: SoFIE Wave 3 Version 4, adjusted longitudinal weights, Statistics New Zealand

Notes:

1. Based on original sample members who respond in Waves 1-3; who consent for their data to be linked to NZHIS information; are participating; and who are aged over 15 at the end of the reference period in Wave 1. Full-time students and those 65 years of age and over are excluded. All variables were included in the model and significant and insignificant variables or variable categories are kept in for completeness.
2. *Significant at the 90% level. **Significant at the 95% level. ***Significant at the 99% level.
3. The likelihood of experiencing absenteeism was modelled. No absenteeism is the base category.

Appendix F

Appendix Table F1 – Mean and standard deviations of variables – linear regression models, participating working age non-students, 2004/05

	Mean	Standard deviation
Usual annual hours worked	1,941.984	864.878
Gender (male=0, female=1)	0.482	0.500
Region (base=Auckland)		
Waikato (=1)	0.097	0.296
Wellington (=1)	0.127	0.333
Rest of North Island (=1)	0.229	0.420
Canterbury (=1)	0.157	0.364
Rest of South Island (=1)	0.111	0.314
Born in NZ (yes=1, no=0)	0.193	0.394
Ethnicity (base=NZ/European)		
Māori (=1)	0.103	0.304
Pacific Islander (=1)	0.043	0.203
Other (=1)	0.067	0.250
Age at interview date	40.878	12.244
Age 50 and over (15-49=0, 50 and over=1)	0.280	0.449
Highest qualification (base=school qualification)		
Post-school vocational qualification (=1)	0.396	0.489
Degree or higher (=1)	0.193	0.395
No qualification (=1)	0.154	0.361
Self-rated health (base=excellent)		
Very good	0.363	0.481
Good	0.189	0.392
Fair	0.041	0.197
Poor	0.001	0.082
Studying (no studying in reference period=0, studying in reference period=1)	0.130	0.336
Log other household income	8.519	4.039

	Mean	Standard deviation
Partner (base=working partner)		
Non-working partner (=1)	0.094	0.292
No partner (=1)	0.307	0.461
Children (base=no children)		
Child(ren) minimum age 0-<5	0.149	0.356
Child(ren) minimum age 5-17	0.272	0.445
Years paid employment	22.248	12.182
Unemployment rate	3.706	0.296
Number of observations	**8,710**	

Source: SoFIE Wave 3 Version 4, adjusted longitudinal weights, Statistics New Zealand

Note: Based on original sample members who respond in Waves 1-3; who consent for their data to be linked to NZHIS information; are participating; and who are aged over 15 at the end of the reference period in Wave 1. Full-time students and those 65 years of age and over are excluded.

Appendix Table F2 – Estimated coefficients for hours worked – linear regression model – 2004/05

	Coefficient	Standard error	P value	95% confidence intervals	
				Lower	Upper
Self-rated health (base=excellent)					
Very good	-21.360	20.581	0.299	-61.704	18.984
Good	-122.732***	26.014	0.000	-173.726	-71.737
Fair	-270.494***	49.993	0.000	-368.493	-172.495
Poor	-454.700***	117.337	0.000	-684.708	-224.692
Sex (base=male)					
Female	-461.142***	34.454	0.000	-528.679	-393.604
Region (base=Auckland)					
Waikato	121.550***	38.333	0.002	46.409	196.692
Wellington	-42.498	27.625	0.124	-96.650	11.654
Rest of North Island	-20.080	28.919	0.487	-76.769	36.608
Canterbury	11.125	28.421	0.695	-44.586	66.836
Rest of South Island	-11.067	29.920	0.711	-69.718	47.584
Born in New Zealand (base=yes)					
No	-50.781*	26.857	0.059	-103.427	1.865
Ethnicity (base=NZ/European)					
Māori	33.602	33.424	0.315	-31.916	99.120
Pacific Islander	96.502*	49.355	0.051	-0.246	193.250
Other	12.389	49.029	0.801	-83.720	108.499
Age at interview date	-27.484***	2.586	0.000	-32.552	-22.415
Aged 50 and over (base=15-49)					
Aged 50 and over	812.546***	311.955	0.009	201.040	1424.052
Highest qualification (base=school qualification)					
Post-school vocational qualification	25.696	23.770	0.280	-20.899	72.291
Degree or higher	207.806***	28.731	0.000	151.487	264.126
No qualification	-62.026**	29.783	0.037	-120.408	-3.643
Studying (base=no studying)	-219.067***	31.893	0.000	-281.585	-156.550
Log other household income	-7.752***	2.788	0.005	-13.217	-2.287
Partner (base=working partner)					
Non-working partner	-113.556**	44.684	0.011	-201.146	-25.965
No partner	-233.905***	38.160	0.000	-308.708	-159.102

	Coefficient	Standard error	P value	95% confidence intervals	
				Lower	Upper
Children (base=no children)					
Child(ren) minimum age 0-<5	8.603	45.541	0.850	-80.669	97.874
Child(ren) minimum age 5-17	-12.768	34.861	0.714	-81.104	55.569
Years paid employment	49.120***	5.146	0.000	39.032	59.208
Years paid employment squared	-0.370***	0.100	0.000	-0.565	-0.175
Unemployment rate	-51.238*	30.803	0.096	-111.620	9.144
Interactions					
Female*Child(ren) minimum age 0-<5	-764.194***	58.740	0.000	-879.338	-649.049
Female*Child(ren) minimum age 5-17	-237.885***	46.749	0.000	-329.525	-146.246
Female*Non-working partner	-28.078	93.140	0.763	-210.654	154.499
Female*No partner	219.626***	44.718	0.000	131.968	307.283
Aged 50 and over*Age	-16.029***	6.077	0.008	-27.941	-4.117
Constant	2881.924***	137.147	0.000	2613.083	3150.765

Model summary statistics	
Number of observations	8,710
F Statistic	77.96
Root Mean Squared Error	746.02
R^2	0.2588

Source: SoFIE Wave 3 Version 4, adjusted longitudinal weight, Statistics New Zealand

Notes:

1. Based on original sample members who respond in Waves 1-3; who consent for their data to be linked to NZHIS information; are participating; and who are aged over 15 at the end of the reference period in Wave 1. Full-time students and those 65 years of age and over are excluded. All variables were included in the model and significant and insignificant variables or variable categories are kept in for completeness.
2. *Significant at the 90% level. **Significant at the 95% level. ***Significant at the 99% level.
3. Total annual hours worked was modelled.
4. The standard error and coefficients from this model are quite variable and in some cases quite high. Various tests for multicolinearity were undertaken but there was no evidence of multicolinearity. The high standard errors seem to be owing to the wide distribution in hours worked and the small number in some of the groups.

Appendix G

Appendix Table G1 – Mean and standard deviations of variables – logistic regression models, all working age non-students, 2004/05

	Mean	Standard deviation
Labour force participation		
(participation=1, not participating=0)	0.871	0.335
Gender (male=0, female=1)	0.515	0.500
Region (base=Auckland)		
Waikato (=1)	0.095	0.294
Wellington (=1)	0.127	0.333
Rest of North Island (=1)	0.231	0.422
Canterbury (=1)	0.153	0.360
Rest of South Island (=1)	0.110	0.313
Born in NZ (yes=1, no=0)	0.201	0.401
Ethnicity (base=NZ/European)		
Māori (=1)	0.110	0.313
Pacific Islander (=1)	0.048	0.214
Other (=1)	0.072	0.259
Age at interview date	41.337	12.530
Age 50 and over (15-49=0, 50 and over=1)	0.297	0.457
Highest qualification (base=school qualification)		
Post-school vocational qualification (=1)	0.383	0.486
Degree or higher (=1)	0.179	0.384
No qualification (=1)	0.179	0.384
Self-rated health (base=excellent)		
Very good	0.353	0.478
Good	0.199	0.399
Fair	0.054	0.226
Poor	0.014	0.117
Studying		
(no studying in reference period=0,		
studying in reference period=1)	0.126	0.332
Log other household income	8.383	4.110

	Mean	Standard deviation
Partner (base=working partner)		
Non-working partner (=1)	0.106	0.308
No partner (=1)		
Children (base=no children)		
Child(ren) minimum age 0-<5	0.160	0.367
Child(ren) minimum age 5-17	0.266	0.442
Years paid employment	21.496	12.490
Unemployment rate	3.707	0.296
Number of observations	**10,120**	

Source: SoFIE Wave 3 Version 4, unweighted, Statistics New Zealand

Note: Based on original sample members who respond in Waves 1-3; who consent for their data to be linked to NZHIS information; and who are aged over 15 at the end of the reference period in Wave 1. Full-time students and those 65 years of age and over are excluded.

Appendix Table G2 – Estimated coefficients for participating – logistic regression model – 2004/05

	Coefficient	Standard error	P value	95% confidence intervals	
				Lower	Upper
Self-rated health (base=excellent)					
Very good	-0.015	0.108	0.890	-0.227	0.197
Good	-0.586***	0.113	0.000	-0.808	-0.364
Fair	-1.353***	0.146	0.000	-1.639	-1.066
Poor	-2.138***	0.218	0.000	-2.565	-1.712
Sex (base=male)					
Female	0.032	0.187	0.863	-0.334	0.398
Region (base=Auckland)					
Waikato	0.306	0.160	0.056	-0.007	0.619
Wellington	-0.129	0.136	0.343	-0.395	0.137
Rest of North Island	0.065	0.126	0.609	-0.183	0.312
Canterbury	0.182	0.136	0.183	-0.085	0.449
Rest of South Island	0.073	0.147	0.621	-0.216	0.361
Born in New Zealand (base=yes)					
No	-0.048	0.127	0.706	-0.297	0.201
Ethnicity (base=NZ/European)					
Māori	-0.187	0.129	0.146	-0.440	0.065
Pacific Islander	-0.095	0.208	0.647	-0.503	0.313
Other	-0.158	0.197	0.422	-0.545	0.228
Age at interview date	-0.141***	0.010	0.000	-0.160	-0.121
Aged 50 and over (base=15-49)					
Aged 50 and over	7.260***	1.152	0.000	5.002	9.518
Highest qualification (base=school qualification)					
Post-school vocational qualification	0.215**	0.105	0.041	0.009	0.421
Degree or higher	0.925***	0.154	0.000	0.622	1.227
No qualification	-0.499***	0.113	0.000	-0.721	-0.277
Studying (base=no studying)	-0.013	0.138	0.924	-0.284	0.257
Log other household income	-0.006	0.011	0.580	-0.028	0.015
Partner (base=working partner)					
Non-working partner	-0.949***	0.198	0.000	-1.337	-0.561
No partner	-0.870***	0.188	0.000	-1.238	-0.501

	Coefficient	Standard error	P value	95% confidence intervals	
				Lower	Upper
Children (base=no children)					
Child(ren) minimum age 0-<5	0.541*	0.313	0.084	-0.072	1.155
Child(ren) minimum age 5-17	-0.190	0.208	0.361	-0.597	0.217
Years paid employment	0.218***	0.016	0.000	0.187	0.250
Years paid employment squared	-0.001***	0.000	0.000	-0.002	-0.001
Unemployment rate	-0.111	0.136	0.416	-0.378	0.156
Interactions					
Female*Child(ren) minimum age 0-<5	-2.850***	0.329	0.000	-3.494	-2.205
Female*Child(ren) minimum age 5-17	-0.501**	0.238	0.035	-0.966	-0.035
Female*Non-working partner	-0.485*	0.289	0.094	-1.052	0.082
Female*No partner	0.142	0.208	0.495	-0.265	0.549
Aged 50 and over*Age	-0.138***	0.022	0.000	-0.181	-0.096
Constant	5.968	0.630	0.000	4.734	7.203
Model summary statistics					
Number of observations			10,120		
Chi-squared			2,936.01		
Log-likelihood			-2,619.53		
Pseudo R^2			0.3591		

Source: SoFIE Wave 3 Version 4, adjusted longitudinal weight, Statistics New Zealand

Notes:

1. Based on original sample members who respond in Waves 1-3; who consent for their data to be linked to NZHIS information; and who are aged over 15 at the end of the reference period in Wave 1. Full-time students and those 65 years of age and over are excluded. All variables were included in the model and significant and insignificant variables or variable categories are kept in for completeness.
2. *Significant at the 90% level. **Significant at the 95% level. ***Significant at the 99% level.
3. The likelihood of participating was modelled. Not participating is the base category.

Appendix Table G3 – Estimated marginal effects – logistic regression model – 2004/05

	Marginal effect	Standard error	P value	95% confidence intervals	
				Lower	**Upper**
Self-rated health (base=excellent)					
Very good	-0.001	0.006	0.890	-0.013	0.011
Good	-0.039***	0.009	0.000	-0.057	-0.022
Fair	-0.132***	0.022	0.000	-0.174	-0.090
Poor	-0.289***	0.048	0.000	-0.384	-0.194

Source: SoFIE Wave 3 Version 4, adjusted longitudinal weight, Statistics New Zealand

Notes:

1. Based on results in Appendix Tables G1 and G2. Other variables are fixed at their mean value.
2. *Significant at the 90% level. **Significant at the 95% level. ***Significant at the 99% level.

Appendix H

Appendix Table H1 shows point estimates and confidence intervals for the number of people affected by each cost component. As an example, an estimated 1,301,700 people are affected by the costs considered. The 95% confidence interval around this figure is (1,254,900; 1,348,400).

Appendix Table H1 – Estimates of number affected and 95% confidence intervals – 2004/05

	Number of people affected		
	Count	**Lower CI**	**Upper CI**
Direct costs			
Ill health inpatient appointments	267,700	252,600	282,800
Indirect costs			
Absenteeism	315,700	299,400	332,100
Presenteeism	939,200[4]	916,600	961,800
Working fewer hours	458,500	419,400	486,200
Not being in the workforce	42,300	26,600	58,000
Total Indirect[1]	*1,196,200*	*1,158,200*	*1,234,100*
Total [2 3]	**1,301,700**	**1,254,900**	**1,348,400**

Source: SoFIE/NZHIS Wave 3 Version 4, adjusted longitudinal weights, Statistics New Zealand

Notes:
1. Direct costs are for those aged 17 and over. Indirect costs are for working age non-students.
2. The total number of people affected is not the sum of the individual groups as the groups are not mutually exclusive; that is, some people can appear in both groups.
3. To estimate the total number of people affected it is assumed that the additional number of people who would participate in the absence of ill health are those with hospital appointments.
4. Count is from Method 1 – Maximum.

Appendix Table H2 shows point estimates and confidence intervals for the number of hours lost by each cost component. Estimates of hours lost are only possible for indirect costs. To estimate the number of hours lost for absenteeism, presenteeism and working fewer hours the number of people affected is assumed to be fixed. To estimate the lost hours for those not in the workforce the lower and upper confidence intervals for the number of people affected are taken. For these people it is assumed that the lost hours are fixed at full-time hours (that is, eight-hour days for 260 days a year). The total hours lost for the indirect components considered are estimated to be 206.9 million to 579.6 million, with a 95% confidence interval (142.2 million to 502.7 million; 271.5 million to 657.3 million). Taking the estimate of presenteeism from Method 1 and Assumption 2 the total hours lost from indirect components are estimated to be 375.1 billion with a 95% confidence interval (304.7 million; 445.3 million).

Appendix Table H2 – Estimates of hours lost and 95% confidence intervals – 2004/05

	Hours lost (mn)		
	Count	**Lower CI**	**Upper CI**
Indirect costs			
Absenteeism	10.3	9.6	11.0
Presenteeism	36.3 – 409.0	34.5 – 395.0	38.2 – 424.0
Working fewer hours	72.3	42.8	101.7
Not being in the workforce	88.0	55.3	120.6
Total indirect	*206.9 – 579.6*	*142.2 – 502.7*	*271.5 – 657.3*

Source: SoFIE/NZHIS Wave 3 Version 4, adjusted longitudinal weights, Statistics New Zealand

Note: Indirect hours lost are for working age non-students.

Appendix Table H3 shows the estimates of costs and the associated confidence intervals by each cost component. To calculate the confidence intervals for the costs, the confidence intervals for the lost hours are used and evaluated at the average full-time hourly rate from the New Zealand Income Survey ($19.95). It is assumed that the average full-time pay rate is fixed (at $19.95), although this itself will be subject to sampling error as it is from the New Zealand Income Survey. The total cost of the components considered is estimated to be $5.417 billion to $ 12.853 billion, with a 95% confidence interval ($3.997 billion to $11.189 billion; $6.827 billion to $14.524 billion). Taking the estimate of presenteeism from Method 1 and Assumption 2 the total estimated costs are $8.773 billion with a 95% confidence interval ($7.239 billion; $10.294 billion).

Appendix Table H3 – Estimates of cost and 95% confidence intervals – 2004/05

	Cost ($bn)		
	Point estimate	**Lower CI**	**Upper CI**
Direct costs			
Ill health inpatient appointments	1.290	1.160	1.410
Indirect costs – Evaluated at full-time hourly wage			
Absenteeism	0.205	0.191	0.219
Presenteeism	0.724 – 8.160	0.688 – 7.880	0.762 – 8.459
Working fewer hours	1.442	0.854	2.029
Not being in the workforce	1.755	1.104	2.407
Total indirect	*4.127 – 11.563*	*2.837 – 10.029*	*5.417 – 13.114*
Total	**5.417 – 12.853**	**3.997 – 11.189**	**6.827 – 14.524**

Source: SoFIE/NZHIS Wave 3 Version 4, adjusted longitudinal weights, Statistics New Zealand

Note: Direct costs are for those aged 17 and over. Indirect costs are for working age non-students.