

Health and Labour Force Participation

Heather Holt

NEW ZEALAND TREASURY
WORKING PAPER 10/03

NOVEMBER 2010



N Z T R E A S U R Y
W O R K I N G P A P E R
1 0 / 0 3

Health and Labour Force Participation

M O N T H / Y E A R

November 2010

A U T H O R

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

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

A C K N O W L E D G E M E N T S

Thank you to Dean Hyslop, Steve Stillman and Katy Henderson for their advice and guidance throughout. Thanks also to Kristie Carter, Tony Burton, Grant Scobie, Gerald Minnee, Ken Richardson and Martin Tobias for providing useful comments.

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.

N Z T R E A S U R Y

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

D I S C L A I M E R

Access to the data used in this study was provided by Statistics New Zealand in a secure environment designed to give effect to the confidentiality provisions of the Statistics Act 1975. A large portion of the analysis in this paper is based on data from the Survey of Family Income and Employment (SoFIE). Statistics New Zealand has initiated a systems review for SoFIE. Therefore data contained in this paper could be subject to change. However, any errors in the analysis are those of the author, not Statistics New Zealand.

The views, opinions, findings and conclusions or recommendations expressed in this Working Paper are strictly those of the author. They do not necessarily reflect the views of the New Zealand Treasury or the New Zealand Government. The New Zealand Treasury and the New Zealand Government take no responsibility for any errors or omissions in, or for the correctness of, the information contained in these working papers. The paper is presented not as policy, but with a view to inform and stimulate wider debate.

Abstract

This paper examines the relationship between health and labour force participation using data from the first three waves of the Survey of Family, Income and Employment (SoFIE) (2002/05). Using various health measures, the results show that health is significantly related to labour force participation, even after accounting for certain types of endogeneity.

The results of the standard regression models including individual chronic diseases indicate that five out of the nine chronic diseases considered have a significant negative relationship with labour force participation once other factors are controlled for. These diseases are: psychiatric conditions (depression, manic depression or schizophrenia); stroke; heart disease; diabetes and high blood pressure. For psychiatric conditions, stroke and diabetes the negative relationship with full-time work is larger than that for part-time work (ie, the chance of working full-time rather than being inactive is reduced more than the reduction in the chance of working part-time rather than being inactive). This suggests that the presence of these diseases is associated not only with lower participation but also with working fewer hours.

Various modelling techniques and a more general measure of overall health (self-rated health) are then used to account for possible endogeneity. The results of these models indicate that poorer self-rated health is associated with a reduced chance of participating in the labour force. The relationship between self-rated health and labour market participation is found to be significant even when time-constant unobserved variables are controlled for and when self-rated health is adjusted to account for possible rationalisation of labour force participation using self-rated health. More specifically, a health shock (measured using adjusted or unadjusted self-rated health) was found to be associated with a reduction in the chance of participating. While the results from all models are in a similar direction, they have different strengths and the preferred estimators are those from the fixed effects model.

Using various assumptions, the model results were used to estimate the impact at the economy level. The point estimates from these models indicate that if there was an improvement in health (ie, no negative health shocks and/or everyone had excellent average health) an additional 12,700 to 66,800 people may participate; that represents a 0.7% to 3.6% increase in the total number of people participating. Based on the limitations of the models discussed in the paper it is more sensible to assume that, if there was an improvement in health, the additional number of people who may participate is likely to be between 5,300 and 38,700; that is, a 0.3% to 2.1% increase in the total number of people participating.

The results do not control for unobserved variables that vary over time. They also do not allow for the “feedback effect”; that is, that participation could influence health. As such, the results do not address causality but only establish relationships between health and participation. Feasible instruments were explored to try to instrument health, thus making it possible to take into account both unobserved variables that change over time and causality, but no suitable instrument was found.

J E L C L A S S I F I C A T I O N J22 Labour Supply; I10 General Health

K E Y W O R D S Health; Chronic Diseases; Labour Force Participation

Table of Contents

Abstract	i
1 Introduction	1
2 Previous studies.....	2
3 Data.....	3
3.1 Survey methodology	3
3.2 Population and sample of interest	3
4 Measurement and methods	4
4.1 Measurement of labour market activity.....	4
4.2 Measurement of health	5
4.3 Modelling the health effect.....	7
5 Chronic diseases.....	15
5.1 Chronic disease and labour market participation	15
5.2 Chronic disease and labour market outcome	22
6 Self-rated health and labour market participation.....	25
6.1 Models used.....	25
6.2 Unadjusted self-rated health.....	27
6.3 Adjusted self-rated health	36
7 Conclusion.....	40
8 Discussion	42
References	50
Bibliography.....	51
Appendix A.....	52
Appendix B.....	57
Appendix C.....	60
Appendix D.....	69
Appendix E.....	79
Appendix F	84
Appendix G.....	88

List of Tables

Table 1 – Chronic disease prevalence: 2002/03 to 2004/05	16
Table 2 – Labour market participation rates by disease presence: 2002/03 to 2004/05	17
Table 3 – Marginal effects by disease presence: 2002/03 to 2004/05	20
Table 4 – Labour market outcome rates by disease presence: 2002/03 to 2004/05	22
Table 5 – Estimated odds ratios for each labour market outcome – pooled multinomial logistic regression – grouped and individual diseases: 2002/03 to 2004/05	24
Table 6 – Estimated marginal effects for each labour market outcome – pooled multinomial logistic regression – grouped and individual diseases: 2002/03 to 2004/05	25
Table 7 – Distribution and participation rates by self-rated health: 2002/03 to 2004/05	27
Table 8 – Labour market outcome rates by self-rated health: 2002/03 to 2004/05	29
Table 9 – Estimated odds ratios for each labour market outcome – pooled multinomial logistic regression – self-rated health: 2002/03 to 2004/05	29
Table 10 – Estimated marginal effects for each labour market outcome – pooled multinomial logistic regression – self-rated health: 2002/03 to 2004/05	30
Table 11 – Changes in self-rated health in consecutive waves: 2002/03 to 2004/05	30
Table 12 – Changes in participation compared with changes in self-rated health in consecutive waves: 2002/03 to 2004/05.....	31
Table 13 – Marginal effects by self-rated health: 2002/03 to 2004/05.....	34
Table 14 – Distribution of objective health measures with self-rated health: 2002/03 to 2004/05	37
Table 15 – Mean and standard deviation of adjusted health measure for sample of interest – self-rated health: 2002/03 to 2004/05	38
Table 16 – Summary of marginal effects from all models: 2002/03 to 2004/05	40
Table 17 – Estimates of annual impact from each binomial model, count (increase in number currently participating) and % (of current number of working age non-students participating in the labour force).....	45
Table 18 – Estimates of annual impact from each multinomial model, count (increase in number in each labour market outcome).....	48
Appendix Table A1 – Definitions of SoFIE variables in participation models	52
Appendix Table A2 – Non-SoFIE variables in participation models	55
Appendix Table A3 – Definitions of additional SoFIE variables in adjusted health model.....	56
Appendix Table D1 – Estimated coefficients for labour force participation – pooled logistic regression model – grouped chronic diseases: 2002/03 to 2004/05.....	69
Appendix Table D2 – Mean and standard deviations of variables – pooled regression models – individual chronic diseases: 2002/03 to 2004/05	71
Appendix Table D3 – Estimated coefficients for labour force participation – pooled logistic regression model – individual chronic diseases: 2002/03 to 2004/05	73
Appendix Table D4 – Estimated coefficients for labour market outcome – pooled multinomial logistic regression model – grouped chronic diseases: 2002/03 to 2004/05	75
Appendix Table D5 – Estimated coefficients for labour market outcome – pooled multinomial logistic regression model – individual chronic diseases: 2002/03 to 2004/05	77
Appendix Table E1 – Mean and standard deviations – pooled regression models –self-rated health: 2002/03 to 2004/05	79
Appendix Table E2 – Estimated coefficients for labour force participation – pooled logistic regression model – self-rated health: 2002/03 to 2004/05	80
Appendix Table E3 – Estimated coefficients for labour market outcome – pooled multinomial logistic regression model – self-rated health: 2002/03 to 2004/05	82
Appendix Table F1 – Estimated coefficients for labour force participation – fixed effects logistic regression model – self-rated health: 2002/03 to 2004/05	84
Appendix Table F2 – Estimated coefficients for labour force participation – correlated random effects logistic regression model – self-rated health: 2002/03 to 2004/05	86

Appendix Table G1 – Estimated coefficients for self-rated health – ordered logistic regression model: 2002/03 to 2004/05	88
Appendix Table G2 – Estimated coefficients for labour force participation – pooled logistic regression model – adjusted health stock: 2002/03 to 2004/05	91
Appendix Table G3 – Estimated coefficients for labour force participation – fixed effects logistic regression model – adjusted health stock: 2002/03 to 2004/05	93
Appendix Table G4 – Estimated coefficients for labour force participation – correlated random effects logistic regression model – adjusted health stock: 2002/03 to 2004/05	95

List of Figures

Figure 1 – Relationship between results from binomial logistic regression – numeric example	9
Figure 2 – Relationship between results from multinomial logit model – numeric example	11
Figure 3 – Estimated odds ratios of participating in the labour force – pooled logistic regression – grouped and individual diseases: 2002/03 to 2004/05	19
Figure 4 – Estimated odds ratios of participating in the labour force – pooled logistic regression – self-rated health: 2002/03 to 2004/05	28
Figure 5 – Estimated odds ratios of participating in the labour force – fixed effects model – self-rated health: 2002/03 to 2004/05	32
Figure 6 – Estimated odds ratios of participating in the labour force – correlated random effects model – self-rated health: 2002/03 to 2004/05	33
 Figure B1 – SoFIE wave structure	58
 Figure C1 – Form of binomial logistic regression model.....	61
Figure C2 – Initial form of the fixed and standard random effects logistic panel model	62
Figure C3 – Equations used in the correlated random effects logistic regression panel model.....	63
Figure C4 – Relationship between true health and measured health in each wave	65

Health and Labour Force Participation

1 Introduction

Health is a key factor in 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 developing or using skills, then improving health could raise labour force participation and economic output. In addition, if poor health reduces the number of hours worked, or lowers productivity when at work, then further output could be lost. The costs of treating poor health and the value of lost output are measures of the economic cost of ill health. A better understanding of the relationships between health and labour market participation is a first step towards estimating these costs.

Chronic diseases are of particular health interest as they are a major component of ill health and deaths in New Zealand, and could place even greater burdens on the health system over time. Furthermore, the incidence of chronic disease is partly driven by lifestyle-related risk factors such as unhealthy diet and tobacco consumption that can potentially be modified. In 2005, around three-quarters of deaths in New Zealand resulted from chronic diseases, a proportion that has been rising in recent years.¹ In countries such as New Zealand, that have an ageing population, understanding this relationship becomes even more important as more people reach the lifestage (note: there are different views about how ageing might affect morbidity as longevity rises) at which their health tends to deteriorate and affect their labour market behaviour (Currie and Madrian, 1999). If both prevalence of, and deaths from, chronic disease continue to rise, there may be significant long-term negative economic impacts arising from increased health care costs and lower labour market participation.

This paper assesses the relationship between health and labour market participation for working age adults in New Zealand. Limited data means there has been little research into the effect of health on labour market participation in New Zealand. However, the inclusion of a detailed health module in the third wave of the longitudinal Survey of Family, Income and Expenditure (SoFIE) has allowed such analysis to be undertaken.

Section 2 of this paper summarises other work done in this area, while Section 3 describes the data used in the paper. Section 4 summarises the methods used, Section 5 reviews the results of the relationship between chronic diseases and labour market participation, Section 6 summarises the results of the relationship between self-rated

¹ Figure based on data from the New Zealand Health Information Service.

health and labour market participation and Section 7 concludes. Section 8 presents estimates of the potential impact at the population level; based on the individual level results. Full details of the variables used, methods and the model results can be found in the appendices.

The paper is not a review of current health policy or spending; the focus is identifying relationships (if any) between health and labour force participation. Where any relationships are established, the paper does not attempt to assess how changes in current health policies may interact with these relationships. For example, the case for investing more resources in managing particular chronic diseases to improve labour market participation would require evidence on: how far such investments might reduce the incidence and prevalence of that disease; and how that, in turn, might affect labour market behaviour. This paper does not address such evidence.

2 Previous studies

Previous research in New Zealand has identified extensive interactions between health and human capital development (Biddulph, F., Biddulph, J. and Biddulph, C., 2003). However, most work has focused on the impact of poor health on the human capital development of young people, rather than the impact of poor health in later life. One health related measure is the presence of a disability. A recent paper using the New Zealand Disability Survey found that all of the six disabilities considered had a negative impact on employment.² In addition, for all disabilities other than hearing, increased severity of the disability was found to reduce the rate of employment (Jensen *et al*, 2005). This work also found that the impact of disability on full-time employment was much larger than for total employment (full-time and part-time).

Another health-related measure is injury. A paper using Statistics New Zealand's Linked Employer-Employee Database (LEED) estimated the effects of injuries on employment (Crichton, Stillman and Hyslop, 2007). Crichton *et al* found that injuries resulting in more than three months of earnings compensation have negative effects on future labour market outcomes; with the magnitude of these effects increasing with injury duration. While disability and injury are possible indicators of health, more direct measures, such as the presence of chronic disease, are better measures of poor health. No New Zealand studies examining the impact of chronic diseases on labour market participation were found.

Interest in the relationship between health and labour market participation is not confined to New Zealand. Literature reviews (Currie and Madrian, 1999; Chirikos, 1993, in Currie and Madrian, 1999) have identified considerable evidence linking health and labour market activity, but wide disagreement on the magnitude of the effect. Numerous papers using US data suggest a strong link between health and labour market participation. In 1989, Stern found that health problems limiting the amount of work that can be done and poor self-rated health reduced the probability of labour market participation. While looking at the relationship between health and retirement in the later part of working life, (Bound, Schoenbaum, Stienbrickner and Waidmann, 1999) found that poorer health lead many older workers to withdraw from the labour force.

² The disabilities considered included vision; hearing; restricted mobility; restricted coordination; learning/memory; and psychological disabilities.

Evidence from the US on the relationship between labour force participation and health is not directly applicable to New Zealand. For instance, those with poorer health in the US may be motivated to participate in the labour force as health insurance is often tied to employment (Cai and Kalb, 2006). As such, a better comparator may be Australia or the UK. A few recent papers using the Australian equivalent of SoFIE (the Household, Income and Labour Dynamics in Australia (HILDA)) have examined the relationship between health and participation. Using data from HILDA, Cai and Kalb (2006) examined the effect of self-rated health on labour force participation for men and women of working age. They found that health was positively associated with participation for four groups (younger males, younger females, older males and older females) even after controlling for the fact that labour force participation may in turn affect health. Further work by Cai (2007) confirmed these findings.

Work by the Australian Productivity Commission examined the impact of chronic diseases on labour market participation (Laplagne, Glover and Shomos, 2007). The chronic diseases considered were cancer, cardiovascular disease, mental/nervous condition, major injury, diabetes and arthritis. They found that absence of chronic diseases can result in substantially greater labour force participation for those affected again even after using different methods to allow for unobserved variables that may affect labour force participation and to allow for the fact that participating in the labour market may in turn affect health. Of the six health conditions considered, mental health or a nervous condition had the largest impact on labour market participation.

Turning to evidence from Britain, work by the Institute of Fiscal Studies, using the British Household Panel survey, examined the role of ill health in retirement decisions (Disney, Emmerson and Wakefield, 2003). They found that deterioration in an individual's self-reported health was strongly associated with movements out of work.

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 B.

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. 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. Those over working age or who are full-time students in each wave are excluded from the analysis.

As with all surveys, not all those approached to take part 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. Statistics New Zealand provides a standard longitudinal weight that 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.

For these reasons the results in this paper reflect the SoFIE population, who are likely to be somewhat healthier than both the population it aims to represent and the New Zealand population more generally. More specifically, those with the most severe health conditions considered may die or be institutionalised, and so are not covered by the survey results used in this analysis. Therefore, the impact of the health conditions considered in this study on labour force participation may 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 B.

4 Measurement and methods

4.1 Measurement of labour market activity

Labour market activity at the household interview date is used for this analysis. Two breakdowns of labour market activity are used: *labour market participation* and *labour market outcome*.

The main focus of the report will be on *labour market participation*; that is:

- participating (working full-time or part-time (including unpaid work) or being unemployed (that is not working but actively looking for work))
- not participating (that is, not working and not looking for work so that the person is economically inactive).³

Labour market outcome is also briefly considered; that is:

- full-time paid or unpaid work (30 hours or more on average in a week)
- part-time paid or unpaid work (less than 30 hours on average in a week)
- unemployed
- inactive.

³ This definition differs from the more standard definition of labour force participation as unpaid workers here are defined to be participating rather than not participating.

4.2 Measurement of health

In Wave 3 of the survey respondents were asked a detailed set of health questions. Hence a respondent's health status could be linked to their current and previous labour market outcomes to see what relationships could be established. Two measures of health are available in all three waves of the survey: the presence of chronic diseases (derived from Wave 3 responses); and self-rated health. Neither provide perfect measures of ill health (the sub-sections below provide further discussion of the problems with each health measure). In a review of the literature, Currie and Madrian (1999) concluded that the effects of health on labour supply are sensitive to the way health is measured, so a range of health measures need to be considered to properly understand the impact of health on labour market status. For these reasons this paper summarises and compares results using each of the available health measures in turn.

4.2.1 Chronic diseases

The health module asked respondents if, before the interview date, they have ever been told by a doctor that they have any of the following eight health conditions:

- asthma
- high blood pressure
- high cholesterol
- heart disease
- diabetes (other than during pregnancy for women)
- stroke
- migraines
- psychiatric conditions (depression, manic depression or schizophrenia).

The inclusion of these eight health conditions on the survey defined the conditions to be considered in this report (with the addition of cancer). They are loosely termed "chronic diseases", a term that has been used by others to refer to similar groups of diseases (DeVol and Bedroussian, 2007). Chronic diseases represent a diverse mix of health conditions. For example, the characteristics of migraines, which are a series of often infrequent brief, acute episodes separated by long periods with no functional loss, are very different from those of cancer. And even cancer covers a large mix of disease characteristics. Some chronic conditions, such as high blood pressure and high cholesterol, are in fact risk factors for diseases. This should be borne in mind when interpreting the results.

As well as the detailed information on each individual disease, a summary variable that indicates the presence of one or more chronic diseases is also used. For people who reported having a particular disease, the age at diagnosis was asked for diseases other than psychiatric conditions. This age of diagnosis was used to estimate the number of years since a disease was diagnosed. The presence of chronic diseases is only asked in Wave 3. For all diseases other than psychiatric conditions, the derived number of years since diagnosis was used to measure its presence in Waves 1 and 2. Diagnosis of mental illnesses (other than depression) almost always have onset in childhood and

adolescence. After analysis of the group who had this disease in Wave 3, all these respondents were assumed to have had the disease in Waves 1 and 2. While this may not be the case for all respondents, the assumption is likely to hold for the majority.

The number of years since diagnosis was also used in combination with the presence of chronic disease information to break those with a disease into two groups. Using asthma as an example, this resulted in a variable with the following categories:

- No diagnosis of asthma
- Asthma diagnosed in the last 5 years
- Asthma diagnosed more than 5 years ago.

While the age of diagnosis variable is useful for estimating the time since the onset of each health condition there are likely to be issues with respondents being able to accurately recall this information, especially if this was some years in the past. This should be borne in mind when assessing the results. This is one of the reasons that the time since diagnosis variables were not disaggregated further.

An additional disease of interest not covered in the SoFIE questionnaire is cancer. SoFIE respondents were asked to give permission for their data to be linked to information on cancer registrations held by the New Zealand Health Information Service. For those respondents who agreed to the data linkage (and were successfully matched), it was possible to construct the same presence and years since diagnosis variables in each wave as for the other chronic diseases covered by SoFIE. These variables will only be available for those in the linked data and are only available back to 1990 so the proportion of the population who have had a cancer diagnosis will be an underestimate. The linked sample is used for descriptive statistics that relate to cancer only.⁴ In the models a “cancer unknown” category was included so the sample size available for analysis was not reduced.

Finally, using diagnosis of a chronic disease is an incomplete indicator of health status, which does not capture the relative severity of respondents’ conditions. At best, this indicator focuses on a particular set of chronic diseases, and is not an encompassing measure of current health. SoFIE respondents are asked if they have ever been told by a doctor that they had the disease (or if they have ever had a cancer registration). A person may have had a disease diagnosis but no longer suffer symptoms. An example would be asthma or migraines, from which respondents may have suffered in their youth, but be symptom free by adulthood. On the other hand, a person may have the disease but not have been diagnosed by a doctor. Hence, this indicator of the disease diagnosis gives no indication of severity, and may not capture all those with a disease. An indication of the severity of such diseases, in terms of the functional losses or activity limitations, would allow better analysis of the relationship between health and labour market participation.

4.2.2 Self-rated health

An alternative health measure available in all three waves is self-rated health. Respondents are asked “*In general how would you rate your health – excellent, very good, good, fair or poor?*” Self-rated health is potentially a more encompassing measure of current health state than presence of chronic diseases as it can include other illnesses

⁴ Where only the linked sample was used, adjusted weights were used to realign the sample with the population (adjusted longitudinal weight) as oppose to the weights provided by Statistics New Zealand (standard longitudinal weights).

as well as chronic diseases and is collected for all respondents. As a result of this wider coverage, there is potential for more changes in health to be observed during the survey period. While this may be a more current and inclusive measure of health, allowing for the fact that a respondent may no longer suffer from symptoms of a chronic disease and including other health factors such as injury and illness, it is more subjective and, as such, may be subject to potential bias.

Firstly, 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. Secondly, with only three waves of data, most respondents are unlikely to experience many dramatic health status changes over this short period; and reported changes may not be true changes (Mathiowetz and Laird, 1994 in Bound *et al*, 1999). In addition, the subjective health baseline respondents use as a comparator when answering this question is ill-defined and may change over time. For example, the SoFIE question on self-rated health does not ask respondents to rate their health relative to 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.⁵ Given that there are only three waves of data, and that this report focuses on those of working age, this ageing effect appears to be small and is therefore not considered further in this work. Finally, even for the same person, self-rated health may be dependent on labour market status. This is considered in detail later in this paper.

4.3 Modelling the health effect

4.3.1 Modelling methods and issues

Standard logistic regressions were the starting point for this analysis. Binomial and multinomial logistic regression models were fitted to the data to quantify the relationship between: the presence of different chronic diseases and labour force status; and self-rated health and labour force status (while holding all other variables constant). The binomial and multinomial models use the available characteristics of people to predict the chance of being in each labour market state. All other characteristics can then be held constant to determine the impact of a small change in one characteristic on the chance of participating. In this cross-sectional analysis, responses in each wave were combined together (pooled) so that each respondent had up to three responses in the data. Standard binomial or multinomial logistic regressions were then fit to this pooled data (these models are hereafter referred to as pooled logistic regressions). This “pooling” maximises the data available for analysis. The correlation between the error term for the same respondent in each wave was allowed for by identifying the people as clusters. Full details of the model and methods used in this paper can be found in Appendix C.

The results of binomial logistic regressions can be presented in two main ways:

- Probability – This is the chance that a respondent with certain characteristics participates in the labour market. In a logit model a *marginal effect* is the relationship between a small change in a variable and the change in the probability of the outcome. As an example, where the characteristic of interest is a binary variable (such as disease present/not present), the difference between the

⁵ In fact, data for all longitudinal respondents indicates a fall in the proportion of those who rate their health as excellent between Wave 1 and Wave 3 of around 5 percentage points and an increase in other health states, possibly indicating the ageing SoFIE population. This occurs despite the fact that those respondents who are most unwell are likely to die or move into institutions.

probabilities of the outcome (participating) for two groups (which share all the same characteristics other than for the binary variable) is known as the marginal effect.

- Odds ratio – This is defined as the ratio of the odds of an event occurring in one group to the odds of it occurring in another group.⁶ For example, the ratio of the odds of participating for those with chronic diseases to the odds of participating for those with no chronic diseases. The odds ratios are equal to the exponential of the coefficient when all other factors are held constant. An odds ratio greater than one indicates a positive effect, whilst one between zero and one indicates a negative effect. It is important to remember that a relative change in odds is not the same thing as a relative change in probabilities. In general, the magnitude of the odds ratios will be larger than that of the marginal effects because they are summarising the results in different ways.

The relationship between probabilities, odds, odds ratios and marginal effects in a binomial logistic regression model can be seen in Figure 1, where the results from the first model described in Section 5.1 are presented. The benefit of using odds ratios is that all other variables can be held constant but a value for these variables does not have to be specified. This is not the case for probabilities (or marginal effects) where the values of the other variables need to be specified (these are usually set at their mean value for the whole sample).⁷ However, the interpretation of marginal effects is more intuitive. For these reasons, both odds ratios and marginal effects are presented here.

⁶ Where the odds is the ratio of the probability of an event occurring to the probability of it not occurring within a group; so the probability of participating to the probability of not participating.

⁷ The marginal effects presented here use this method. Alternative methods include using the means for certain groups (ie, those with chronic diseases) or calculating the person-specific marginal effects and averaging them over the groups of interest. These methods were considered here but, as the differences in the resulting marginal effects using these methods were small, the mean for the whole sample was used.

Figure 1 – Relationship between results from binomial logistic regression – numeric example

When all other variables are fixed at their mean value the **probability** of participating in the labour force for people:

- with a chronic disease = $P_1 = 0.865$
- without a chronic disease = $P_2 = 0.903$.

The **odds** of participating in the labour force for people:

- with a chronic disease = $[P_1/(1-P_1)] = [0.865/(1-0.865)] = 6.40$
- without a chronic disease = $[P_2/(1-P_2)] = [0.903/(1-0.903)] = 9.34$.

That means that people with chronic diseases are 6.4 times more likely to participate in the labour force than not participate, while people without chronic diseases are 9.34 times more likely to participate in the labour force than not.

The **odds ratio** for those with chronic diseases is the ratio of the odds of participating for those with chronic diseases to those without chronic diseases. If this value is less than 1 then the odds of participating are lower for those with chronic diseases compared to those without a chronic diseases:

- Odds ratio = $[P_1/(1-P_1)] / [P_2/(1-P_2)] = 6.40/9.34 = 0.685$
- Percentage change in odds = $(0.685-1)*100 = -31.5\%$.

The **marginal effect** is the difference in the probability of participating for those with chronic diseases compared to those without chronic diseases:

- Marginal effect = $P_1 - P_2 = 0.865 - 0.903 = -0.038$
- Percentage point (ppts) change in probability = $-0.038*100 = -3.8\text{ppts}$
- Percentage change in probability = $(-0.038/0.903)*100 = -4.3\%$.

This leads to the following conclusions:

1. The **odds** of participating (relative to not participating) are 31.4% lower for people with a chronic disease compared to people without a chronic disease.
2. The **probability** of participating in the labour force is 3.8 percentage points lower for people with a chronic disease compared to people without a chronic disease.
3. The **probability** of participating in the labour force is 4.3% lower for people with a chronic disease compared to people without a chronic disease.

Note: These results are derived from Appendix Tables D1 and D2. Probabilities are calculated using the formula outlined in Appendix Figure C1.

While a binomial logistic regression model predicts the chance of participating, multinomial models predict the chance of multiple states (ie, working full-time, part-time, being unemployed or being inactive). As with the binomial logistic regression the results from the multinomial logistic regression can be presented in various ways, including probabilities/marginal effects or odds ratios. However, there is a slight difference in how these are interpreted for the multinomial model which is important to understand. The interpretation of the results is explained below and a numeric example, based on the first multinomial model discussed in Section 5.2, can be found in Figure 2.

- Probability – This is the chance that a respondent with certain characteristics is in each labour market state: that is full-time; part-time; unemployed; or inactive. Each respondent has a probability of being in each of the four labour market outcomes (although the probability for any state can be zero). These four probabilities always sum to one, as a person has to be in one of the four states. The *marginal effect* is the relationship between a small change in a variable and the change in the probabilities of being in each of the four labour market outcomes. As an example, where the characteristic of interest is a binary variable (disease present/no disease present), the difference between the probabilities of being in each labour market outcome (full-time/part-time/unemployed/inactive) for two groups (which share all the same characteristics other than for the binary variable) are known as the marginal effects. The marginal effects sum to zero across each respondent. So if the chance of being in three of the four labour market states increases, then the chance of being in the fourth labour market state must decrease by the same amount. Unlike the odds ratios, the marginal effects are not interpreted relative to a particular labour market category, but need to be interpreted across the labour market states.
- Odds ratio – This is defined as the ratio of the odds of an event occurring in one group to the odds of it occurring in another group. The odds ratios are equal to the exponential of the coefficient when all other factors are held constant. In these results the reference labour market outcome is inactive. Taking part-time as an example, the odds ratios for those with chronic diseases is the ratio of the odds of working part-time (rather than being inactive) for those with one or more chronic diseases to the same odds for those without chronic diseases.⁸ As with the binomial models an odds ratio greater than one indicates a positive effect, whilst one between zero and one indicates a negative effect.

Owing to the differences as to what odds ratios and marginal effects measure, and therefore the different magnitudes of the two measures, it is perfectly plausible for the odds ratio for a specific category to be significantly different from the reference category, but for the marginal effect for the same group to not be significant. When calculating the odds ratio, the baseline odds (the ratio of the probability of an event occurring to the probability of it not occurring) drop out, so the magnitude of the probability is not important in the odds ratio calculation. The test for significance indicates whether the odds ratio (which is not dependent on the baseline odds) is different from one. However, the magnitude of the probabilities is important in testing the significance of a marginal effect. The test here is whether the marginal effect significantly changes the baseline probability. If the base probability for the sample is very small or very large then small marginal effects may not be significant. Another way of thinking about this is that a big sounding odds ratio can easily correspond to a very small sounding difference in marginal effect.

⁸ So the odds are the probability of working part-time to the probability of being inactive.

Figure 2 – Relationship between results from multinomial logit model – numeric example

When all other variables are fixed at their mean value the **probability** of being in each labour force state for people:

- with a chronic disease are:
 - $P_1 \text{Full-time} = 0.663$
 - $P_1 \text{Part-time} = 0.165$
 - $P_1 \text{Unemployed} = 0.023$
 - $P_1 \text{Inactive} = 0.149$
- without a chronic disease are:
 - $P_2 \text{Full-time} = 0.715$
 - $P_2 \text{Part-time} = 0.160$
 - $P_2 \text{Unemployed} = 0.019$
 - $P_2 \text{Inactive} = 0.106$.

Focusing on full-time, the **odds** of being in each labour market state relative to being inactive for people:

- with a chronic disease = $P_1 \text{Full-time} / P_1 \text{Inactive} = [0.663/0.149] = 4.45$
- without a chronic disease = $P_2 \text{Full-time} / P_2 \text{Inactive} = [0.715/0.106] = 6.75$.

That means people with chronic diseases are 4.45 times more likely to work full-time than be inactive, while people without chronic diseases are 6.75 times more likely to work full-time than be inactive.

The **odds ratio** for those with chronic diseases is the ratio of the odds of working full-time (relative to inactive) for those with chronic diseases to those without chronic diseases. If this value is less than 1 then the odds of participating is lower for those with chronic diseases compared to those without a chronic diseases:

- Odds ratio = $[P_1 \text{Full-time} / P_1 \text{Inactive}] / [P_2 \text{Full-time} / P_2 \text{Inactive}]$
 $= 4.45/6.75 = 0.659$
- Percentage change in odds = $(0.659-1)*100 = -34.1\%$.

The **marginal effect** is the difference in the probability of working full-time for those with chronic diseases compared to those without chronic diseases. However, the probabilities for each labour market state are not independent; each person must be in one of the four labour market states, so the probabilities across each group must sum to one; that means the marginal effects across each state must sum to zero:

- Marginal effect:
 - Full-time = $P_1 \text{Full-time} - P_2 \text{Full-time} = 0.663-0.715 = -0.052$
 - Part-time = $P_1 \text{Part-time} - P_2 \text{Part-time} = 0.165-0.160 = 0.005$
 - Unemployed = $P_1 \text{Unemployed} - P_2 \text{Unemployed} = 0.023-0.019 = 0.004$
 - Inactive = $P_1 \text{Inactive} - P_2 \text{Inactive} = 0.149-0.106 = 0.043$

Figure 2 Continued – Relationship between results from multinomial logit model – numeric example

- Percentage point (pps) change in probability of working full-time
 $= -0.052 \times 100 = -5.2$ ppts
- Percentage change in probability of working full-time
 $= (-0.052 / 0.715) \times 100 = -7.3\%$.

This leads to the following conclusions:

1. The **odds** of working full-time relative to being inactive are 34.1% lower for people with a chronic disease compared to people without a chronic disease.
2. The **probability** of people with a chronic disease working full-time in the labour force is 5.2 ppts lower than for those without chronic diseases. Comparing the same groups, the probability of working part-time is 0.05 ppts higher, being unemployed is 0.04 ppts higher and being inactive is 4.3 ppts higher.
3. The **probability** of working full-time in the labour force is 7.3% lower for people with a chronic disease.

Note: These results are derived from Appendix Tables D1 and D4. As described in Appendix C, probabilities are calculated using a variation of the formula outlined in Appendix Figure C1.

One of the common problems encountered when trying to estimate the effect of a variable on a particular outcome is endogeneity. Endogeneity occurs if the value of one of the explanatory variables (for example, health status) is dependent on the value of other unobserved variables or on the outcome variable (in this case, labour market participation). In other words, the explanatory variables are not exogenous; true exogenous variables are not affected by the outcome variable or by other unobserved characteristics. One of the assumptions of the standard logistic regression model is that the explanatory variables are exogenous. If endogeneity is present standard logistic regression models can produce inconsistent and possibly biased (incorrect) regression coefficients. While giving an initial indication of possible relationships between labour force participation and health, the standard logistic regression models cannot account for endogeneity. Endogeneity is likely to be an issue when trying to estimate the impact of health on participation for the following reasons:

1. Previous studies have shown that for some groups, as well as affecting labour force participation, health may in turn be influenced by labour force participation; or labour force participation and health may be simultaneously determined (eg, Cai and Kalb, 2006). For example, being inactive may lead some people to be depressed, while being employed in a stressful role may lead to high blood pressure. Therefore the fact that a model may indicate a relationship between the dependent and explanatory variables does not necessarily mean the explanatory variables cause the outcome (Tabachnick and Fidell, Using Multivariate Statistics 4th Edition, 2001). These problems are referred to in the literature as “reverse causality and simultaneity”.
2. Other factors that are not observed in the data may influence both labour force participation and/or health. An example would be average motivation (Laplagne et

al, 2007).⁹ Someone who is less motivated to participate in the labour force may also be less motivated to take the steps to stay healthy (for example, undertaking exercise). Differences in these unobservables between respondents may explain variation in both health and labour force participation. If they are excluded from the model, the variation in labour force participation will appear to be owing to variation in health and therefore the estimated health effect will be biased. This is a particular kind of “unobserved individual heterogeneity”.¹⁰

3. The way health variables are reported may reflect the respondent's labour force participation. For example, respondents may report their health state to justify their labour market state (eg, someone who is not participating in the labour force may report that their health is poorer than they would report if they were participating). This is referred to as “rationalisation bias or endogeneity”.

The longitudinal design of SoFIE allows more complex modelling techniques to try to account for the types of endogeneity outlined above. In addition to the standard logistic regression models for self-rated health the following methods were considered:

- **Fixed and correlated random effects panel logistic regression** – This technique examines the impact of changes in actual self-rated health on participation taking into account unobserved time constant variables that will vary between people and may influence labour force participation and/or health (time constant unobserved heterogeneity). By looking at how changes in participation relate to changes in other variables between waves the time constant unobserved variables are removed when fixed and random effects models are used.¹¹
- **Standard pooled binomial and multinomial models and fixed and correlated random effects panel logistic regression with an adjusted health measure** – These models adjust health for potential rationalisation bias and account for unobserved factors that do not change over time. First, self-rated health was modelled based on a set of more objective health measures and a set of other health-related variables. An adjusted measure of health stock was then predicted using these models. This adjusted measure of health was then included in all of the previous models.
- **Instrumental variables/simultaneous equations** – These techniques can account for unobserved variables that do and do not change over time, and for reverse causality. Although considered in depth, no successful instrument was found.

More discussion of these modelling methods can be found in Appendix C. Ideally these techniques would also have been applied to individual chronic diseases which (like self-rated health) could suffer from endogeneity. Not all of these diseases will be open to all three types of endogeneity identified above and some diseases are more susceptible to certain types of endogeneity than others. Some literature (Cai and Kalb, 2006; Laplagne *et al*, 2007) suggests that rationalisation endogeneity is less likely for the chronic diseases considered given they are less subjective as they depend on a doctor's diagnosis.

⁹ Motivation is not totally fixed over time as, even with a short period, motivation can vary. However, average motivation will be fixed within a person and is likely to vary across individuals.

¹⁰ Another form of unobserved heterogeneity occurs when the unobserved variables are not related to the other explanatory variables although they do explain a certain amount of variation in labour force participation. Note that this form of unobserved heterogeneity would not bias the coefficient on health.

¹¹ Only binomial panel models were considered. This work could be extended in future to consider multinomial panel models.

However, doctors' diagnoses of diseases may in turn affect labour force participation decisions, even when the symptoms of the disease are mild. Applying the techniques to control for possible endogeneity to individual chronic diseases proved problematic for several reasons. These include: the relatively small numbers of people with each chronic disease; the fact that the presence of chronic disease is slow changing (making it hard to compare changes in participation and disease diagnosis within respondents); and only three waves of SoFIE data were available at the time of the analysis. This means that trying to use panel models to account for possible endogeneity would not be especially effective for chronic diseases until more waves of data are available. Standard logistic regression models are therefore the only models considered for individual chronic diseases despite the possibility of endogeneity bias. For self-rated health, results for standard logistic regression models are reported before the more advanced panel model results to compare with both the models for individual chronic diseases and to the panel models as a way of demonstrating possible endogeneity bias.

The chronic disease questions are considered to be more objective than self-reported health (Bound *et al*, 1999), suggesting that such measures are less likely to suffer from rationalisation bias. However, these more objective measures may not always be good predictors of overall health and the ability to work. As noted in Section 4.1.1, a person may no longer suffer from symptoms of a previously diagnosed disease, while others may suffer from a disease but be undiagnosed. Further, modelling difficulties may emerge as the presence of some of these diseases is likely to be collinear to some degree (owing to co-morbidity or secondary diseases), making the coefficients more difficult to interpret (Bound *et al*, 1999). As an example, diabetes is associated with an increased risk of developing heart disease; as such, heart disease may be a secondary disease. These interactions are complex and therefore difficult to include in the analysis. As a result they are not considered further in this report.

Note that throughout the remainder of this paper, words such as "impact" and "effect" are used to describe relationships but do not denote causation. This should be borne in mind when reading the results. Further, where results of the standard logistic regression models are discussed in this paper potential endogeneity bias should be remembered.

4.3.2 Model variables

The decisions on which variables to include in the models were made based on reviews of the literature and best practice. The following variables were included in the standard logistic regression (cross-sectional models):¹²

- gender
- region
- age (and whether aged 50 or above)¹³
- highest qualification
- study status
- marital status
- place of birth
- ethnicity
- presence of children

¹² Wealth of the respondent and the labour force state of any parents the respondent lived with at age 10 were also considered for inclusion. Wealth was not available in all three waves and the labour force state of parents was not significant in the models once other variables were included.

¹³ Unadjusted age was included. The aged 50 and over indicator was included to pick up a change in participation habits that appeared to occur for men and women around the age of 50.

- household income less personal income
- years in paid employment
- the unemployment rate at the time of the interview.¹⁴

Those variables included in the cross-sectional models that were slow or little changing or that could be directly impacted by changes in health were excluded from the fixed and random effects models (longitudinal models) leaving the following variables: gender (random effects only); region; age (and whether aged 50 or above); marital status; place of birth (random effects only); children; household income less personal income; and the unemployment rate at the time of the interview.

In addition to the variables for the cross-sectional models, the model creating the adjusted health measure included the following variables: total household income (as opposed to household income less personal income); health benefit receipt; housing tenure; and whether a respondent has ever smoked. All these variables are defined in Appendix A, Tables A1, A2 and A3.

5 Chronic diseases

This section explores the relationship between different types of chronic disease and labour market participation. It begins by reporting basic descriptive statistics and then summarises the results from the logistic regression models. The analysis in this section is based on pooled cross-sectional data analysis. As previously mentioned, it should be remembered that words such as “impact” and “effect” are used to describe relationships but do not attempt to denote causation and that the results of the standard logistic regression models are subject to potential endogeneity bias. Full tables of results from the main models, including unweighted means and standard deviations for the variables, can be found in Appendix D where the reference categories are labelled.

5.1 Chronic disease and labour market participation

Table 1 shows the proportion of the sample with various disease diagnoses. The results indicate that around half of the sample has been diagnosed with one or more chronic diseases.¹⁵ Table 1 indicates that the most common disease is asthma with 18.5% of respondents having been diagnosed with this disease at some point. The rarest disease is a stroke with only 1% of respondents having been diagnosed with a stroke. This small disease prevalence is not surprising given that strokes are likely to be quite rare for those of working age, the group being analysed. Further, stroke is one disease that is more likely to result in death for this group. In other words, for some diseases the prevalence is higher than others as a result of being more likely to survive with the disease (survivor bias).

¹⁴ The unemployment rate for the time of the interview was included to reflect the rolling interview period throughout the year.

¹⁵ The true proportion is likely to be slightly higher than this as those for whom the presence of cancer is unknown and who have no other chronic diseases have been assumed to have no chronic diseases.

Table 1 – Chronic disease prevalence: 2002/03 to 2004/05

Disease	Disease prevalence (%)
Any chronic disease	49.5
Asthma	18.5
High blood pressure	14.9
High cholesterol	13.4
Heart disease	2.9
Diabetes	3.0
Stroke	1.0
Migraine	13.4
Psychiatric conditions	9.5
Cancer*	3.5

Source: SoFIE Waves 1-3 Version 4, standard longitudinal weights (*adjusted longitudinal weight), Statistics New Zealand.

Note: Results are for those aged 15-64 and who are not full-time students. Data for all three waves is pooled together to create an average rate.

Table 2 shows the labour market participation rates by disease presence. The observed labour market participation rates are considerably lower for those with a disease diagnosis compared to the overall participation rate. Participation is lowest for those who have suffered from a stroke. About half (54%) of people with a diagnosed stroke participate in the labour market, compared to the average participation rate of 83%, a reduction in the likelihood of participation of 35% (29 percentage points). However, this estimate is subject to a larger error given it is based on a relatively small group. Only 1% of the sample reported ever being told by a doctor they had suffered a stroke.

Table 2 – Labour market participation rates by disease presence: 2002/03 to 2004/05

Disease	Average number participating over 3 waves (count)	Participation rate (%)
Total	1,835,000	82.6
No chronic disease	958,600	85.5
Any chronic disease	876,500	79.7
Asthma	327,500	80.0
High blood pressure	251,800	76.0
High cholesterol	237,500	80.6
Heart disease	40,500	64.0
Diabetes	41,900	63.7
Stroke	11,700	53.8
Migraine	234,200	78.4
Psychiatric conditions	146,700	69.0
Cancer*	59,000	76.4

Source: SoFIE Waves 1-3 Version 4, standard longitudinal weights (*adjusted longitudinal weights), Statistics New Zealand

Notes:

1. See note on Table 1.
2. This is just a crude participation rate. It had not been age standardised.
3. Counts may not sum to totals owing to rounding.

The bivariate analysis in Table 2 above, while interesting, does not control for other factors that may be related to participation. Pooled cross-sectional logistic regressions were used to determine the relationship between disease presence and participation when some other factors were controlled for.

Initially a basic model was conducted including a summary chronic disease indicator (rather than the individual chronic diseases) to determine the overall impact of having chronic disease on participation. Results show that, even after controlling for other variables, the relationship between chronic disease presence and participation is significant (Appendix Table D1). Figure 3 shows that the odds of participating in the labour force are reduced by 31.5% for those with any chronic disease(s).¹⁶ When all variables are fixed at their mean value, the probability of participating is 0.885. This is above the unconditional mean participation rate of 0.827, perhaps because of the more rapid decline in participation for those over 50 years of age which reduces the unconditional average. For those with no chronic diseases, the estimated probability of participating is 0.903, while for those with a chronic disease the estimated probability is reduced to 0.865; a marginal effect of -0.038 (Table 3). This suggests that for an average person, having chronic diseases reduces labour market participation by 3.8 percentage points on average, or 4.3% in a relative sense.

By contrast, the bivariate analysis in Table 2 indicated a difference of 5.8 percentage points. This suggests that other differences in characteristics are important in explaining the lower participation rate of those diagnosed with a chronic disease (Table 2). For example, the odds of participating are lower for: females with young children (this is associated with a reduction in the odds of participating of 90%); those with non-working

¹⁶ The odds of participating for those with one or more chronic diseases are 6.4:1, without disease are 9.3:1, giving an odds ratio of 0.685 = 6.4/9.3.

partners or no partner (75% and 65% reduction respectively); and for females (22% reduction).

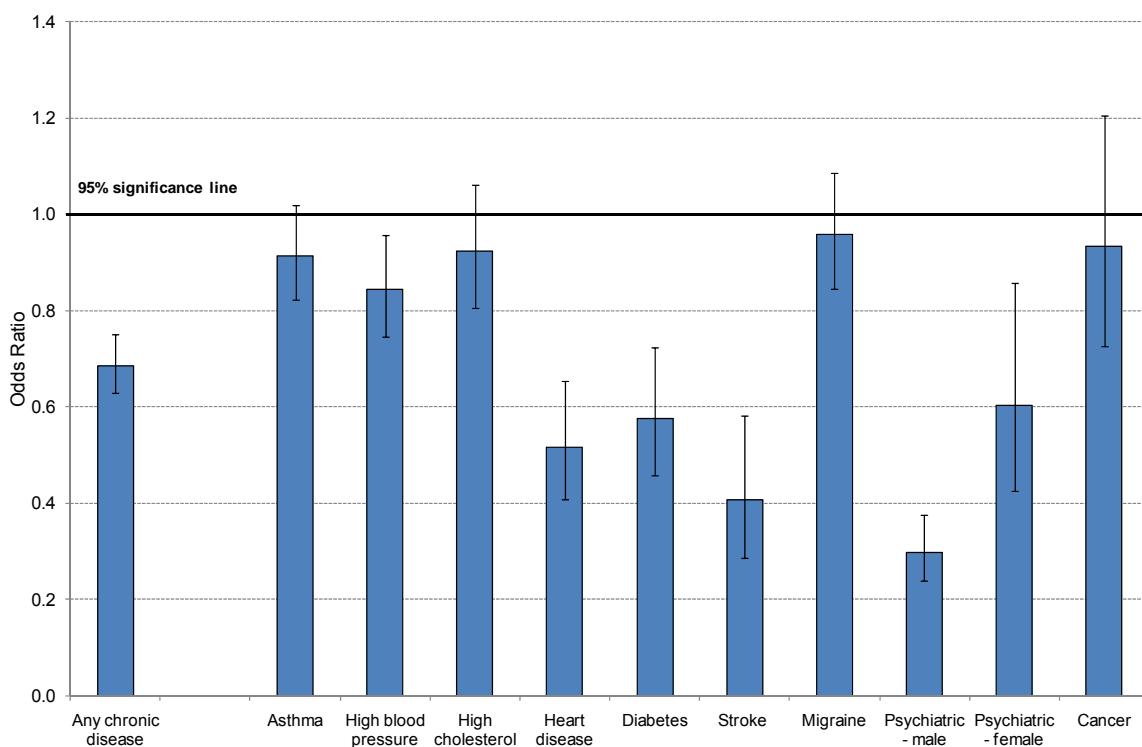
Next, models were considered that included variables for each individual disease, rather than a summary variable indicating disease presence. Figure 3 shows the estimated ratio of the odds of labour market participation for those with each disease to the odds for those without each disease. An odds ratio greater than one indicates a positive effect, whilst one between zero and one indicates a negative effect on the odds of participation for those with each disease. If the vertical line for each bar, showing the 95% confidence interval for the odds ratio, crosses one (indicated by the horizontal 95% significance line), then the chance of participation for those with the disease is not significantly different from those without the disease at the 95% level (once other factors are controlled for). Therefore there was insufficient evidence that those with an asthma, high cholesterol, migraine or cancer diagnosis were any less likely to be participating in the labour market than those without these diseases, once other factors were controlled for. For asthma, migraine and high cholesterol this may be a result of such diseases typically being manageable once identified and therefore not inhibiting labour market participation in many cases.

Having been diagnosed with any of the following diseases (in order of impact from highest to lowest) is associated with a significantly reduced odds of labour market participation compared to someone without the disease, once other factors are controlled for:

- psychiatric conditions (are associated with a 70% reduction in the odds of labour market participation for males and 40% for females)
- stroke (59% reduction);
- heart disease (48% reduction);
- diabetes (42% reduction)
- high blood pressure (16% reduction).

For some of these, the presence of the particular reported condition may not itself be associated with lower odds of participating. Rather, other secondary diseases related to the primary disease may be causing the association. For example, high blood pressure may not be associated with reduced odds of participating, but kidney failure resulting from high blood pressure may. Further, collinearity between these health conditions is not formally investigated here.

Figure 3 – Estimated odds ratios of participating in the labour force – pooled logistic regression – grouped and individual diseases: 2002/03 to 2004/05



Source: SoFIE Waves 1-3 Version 4, unweighted, Statistics New Zealand

Notes:

1. The odds ratios for the summary chronic disease indicator and for individual diseases are derived from different models. Odds ratios for summary chronic disease indicator are derived from Appendix Table D1, while those for individual chronic diseases are derived from Appendix Table D3. The footnotes from those tables apply to this chart.
2. The following factors were held constant: gender; region; age (and whether 50 years of age or above); highest qualification; study status; marital status; place of birth; ethnicity; children; household income less personal income; years in paid employment; and unemployment rate at the time of the interview.

The impact of a disease diagnosis on labour market participation did not differ significantly by gender other than for psychiatric conditions. The presence of this disease was associated with a 70% reduction in the odds of participating in the labour market for men, and 40% for women, and the 95% confidence intervals do not overlap. This substantial and significant difference is in line with work done by the Australian Productivity Commission (Laplange *et al*, 2007).

The results of the model indicate a reduction in the odds of participation of 9% for males with psychiatric conditions relative to females with psychiatric conditions (with an odds ratio of 0.91).¹⁷ This difference by gender may in part be owing to compositional differences between the kinds of men and women who go to the doctor and are diagnosed with psychiatric conditions. A higher proportion of women have been told by a doctor that they suffer from psychiatric conditions (12.8% compared to 6.2%), suggesting that the threshold for men seeking psychiatric help may be higher. Tests indicate that the impact of psychiatric conditions for men is significantly higher than that of heart disease and diabetes, but not significantly different from that for a stroke.

As an illustration of the impact on the probability of participating in the labour force, Table 3 shows the marginal effects on labour market participation as a result of moving from not having a disease to having a disease when all other variables are held at their mean.

¹⁷ These figures are not presented in the chart. They can be derived using the information in Appendix Tables D2 and D3.

The probabilities the marginal effects are based on are derived from Appendix Tables D1, D2 and D3. For instance, when all other variables are fixed at the mean values, the probability of a person participating in the labour market given they have no diabetes diagnosis is 0.890 (which is similar than the average participation probability for all respondents from the model of 0.888). Given a diagnosis of diabetes, the probability is lower at 0.823, giving a marginal effect on participation of -0.067 (shown in Table 3).

Table 3 – Marginal effects by disease presence: 2002/03 to 2004/05

Disease	Marginal effects
Any chronic disease	-0.038***
Asthma	-0.009
High blood pressure	-0.018**
High cholesterol	-0.008
Heart disease	-0.083***
Diabetes	-0.067***
Stroke	-0.123***
Migraine	-0.004
Psychiatric conditions – male	-0.132***
Psychiatric conditions – female	-0.065***
Cancer	-0.007

Source: SoFIE Waves 1-3 Version 4, unweighted, Statistics New Zealand

Notes:

1. The marginal effects for the summary chronic disease indicator and for individual diseases are derived from different models. Marginal effects for summary chronic disease indicators are derived from Appendix Table D1, while those for individual chronic diseases are derived from Appendix Table D3. All marginal effects are calculated holding all other variables at their mean. The footnotes from those tables apply to this table.
2. *Significant at the 90% level. **Significant at the 95% level. ***Significant at the 99% level.
3. These marginal effects are the actual differences in probabilities compared to those without each condition.

The analysis of marginal effects indicates that, in terms of magnitude, the impact of psychiatric conditions is much lower than suggested by the odds ratios. Holding all other values at their mean, the probability of a male with psychiatric conditions participating is 0.797, compared to the probability for a male without psychiatric conditions of 0.929, giving a marginal effect of -0.132. In other words, the labour force participation rate for men with psychiatric conditions is 13.2 percentage points below that for men without psychiatric conditions on average. Similarly, the probability of a female with psychiatric conditions participating in the labour market is 0.812 compared to a probability of 0.878 for females without psychiatric conditions on average, giving a marginal effect of around -0.065. The marginal effect between males with psychiatric conditions and females with psychiatric conditions is -0.015.

The coefficient for cancer indicated a negative relationship with participation, but this relationship wasn't significant.¹⁸ This may reflect the nature of cancer treatment, which is very intensive over a compressed period, or that those with the most severe cases of cancer die. Having cancer diagnosed may result in people of working age taking sick leave for cancer treatment rather than leaving the labour force completely. The result may not hold if full cancer information were available, as those diagnosed with cancer before 1990 are not identifiable but they may have poorer health than those diagnosed later.

¹⁸ In part this maybe owing to the larger error around the estimate owing to cancer information only being known for a restricted sample. Interestingly, when those over working age were included in the model, cancer was found to be significantly related to a reduction in participation.

Interestingly, the coefficient for those respondents who did not agree for their data to be linked to the cancer information (and so were coded as unknown for cancer presence) were significantly less likely to participate in the labour market than those without cancer. This may indicate potential differences in the unobserved characteristics of those who do and do not consent.

Interestingly, the impact of a disease diagnosis on labour market participation did not vary significantly by age. In other words, the reduction in the chance of labour market participation for those with a disease diagnosis was no higher if the respondent was young compared to if they were old.

The non-health related variables indicate that, when all other explanatory factors are held constant, the following groups have lower chance of participating in the labour market: females; those born outside of New Zealand; those who are older; those with no qualifications; those undertaking some form of study; those with non-working partners; and those with higher other household income (relative to the reference categories). Additional years of paid employment is associated with an increase in the chance of participation.¹⁹ For males, having no partner is associated with a reduced chance of participation. This is also true for females but to a lesser extent. Men who have young children are more likely to work than those without children, while men with older children are less likely to work than those without children. For women, having children of any age is associated with a reduction in the chance of participating, with the chance of participating being reduced by the most for those with young children.

Finally, the model of individual diseases was then developed to include, where possible, a variable summarising the presence of the disease and the years since diagnosis. This was done to determine whether more recent diagnoses are associated with higher or lower labour market participation.^{20 21} Of the diseases found to be significantly negatively related to participation (other than psychiatric conditions for which this durational breakdown is not possible), the impact of a more recent diagnosis (in the last five years) of high blood pressure, heart disease or stroke appeared more detrimental than an older diagnosis. For example, the odds of participating for those who have had a stroke in the last five years are reduced by 62%. This compares to a 57% reduction for those who had a stroke five or more years ago. This difference may in part be because the further from the point of diagnosis, the more a person may have recovered. It also may be because the person may no longer be undergoing intensive treatments that prevent them from working, or have learnt how to manage their conditions. Conversely, the difference may also reflect the fact that those who suffer more severe strokes die within five years of being diagnosed and are therefore not included in the data (survivorship bias).

The effect was reversed for diabetes, with a less recent diagnosis being associated with a larger reduction in participation than a more recent diagnosis. The odds of a respondent working who had been diagnosed with diabetes in the last five years were reduced by 29% (which was not significantly different from those with no diabetes diagnosis) while the odds of those with a diagnosis of diabetes five years ago or more participating were 52% lower than those with no diabetes, possibly indicating the progressive nature of diabetes. While providing a possible indication of direction, the coefficients for the two periods of

¹⁹ Over the relevant range (the quadratic peaks in the mid-1980s).

²⁰ Again, the following variables were held constant: gender; region; age (and whether 50 years of age or above); highest qualification; study status; marital status; place of birth; ethnicity; children; household income less personal income; years in paid employment; and unemployment rate at the time of the interview.

²¹ Full model results are available on request.

diagnosis were only found to be significantly different from each for diabetes and heart disease when tested for equality (using a Wald test).

When all those with high cholesterol were considered together, there was insufficient evidence to suggest this group were less likely to participate in the labour force than those without high cholesterol. However, when the period since diagnosis was interacted with high cholesterol, there was a significant reduction in the chance of participating for those who had been diagnosed with high cholesterol five years ago or more, compared to those without high cholesterol. Again, this is possibly owing to the progressive nature of high cholesterol risk.

5.2 Chronic disease and labour market outcome

Not only do those with certain chronic diseases participate less in the labour market, Table 4 shows that those who do participate seem more likely to work part-time than those who have not been diagnosed with a chronic disease. The largest difference is for those who have had a stroke. About a third (33%) of those people who have had a stroke and who are participating in the labour market work part-time, compared to only 19% of all participating respondents. The previous analysis was therefore developed to examine the impact of chronic disease on level of participation, once other factors are controlled for.

Table 4 – Labour market outcome rates by disease presence: 2002/03 to 2004/05

Disease	Labour market outcome (%)			
	Full-time employment	Part-time employment	Unemployment	Total participating
Total	78.4	19.0	2.7	100.0
Any chronic disease	76.3	20.9	2.8	100.0
Asthma	77.6	19.4	3.1	100.0
High blood pressure	76.8	20.7	2.5	100.0
High cholesterol	79.2	18.6	2.2	100.0
Heart disease	76.6	21.6	1.9	100.0
Diabetes	71.7	22.8	5.4	100.0
Stroke	63.7	32.6	3.7	100.0
Migraine	71.5	25.1	3.4	100.0
Psychiatric conditions	68.0	26.9	5.1	100.0
Cancer*	71.5	26.7	1.9	100.0

Source: SoFIE Waves 1-3 Version 4, standard longitudinal weights (*adjusted longitudinal weights), Statistics New Zealand

Note: See footnote on Table 1.

Table 5 summarises the odds ratios from the model. The model with an indicator of any chronic disease presence indicates that, even after controlling for other factors, having a chronic disease is also associated with a larger reduction in the odds of working full-time (relative to being inactive) compared to part-time (relative to being inactive). The odds of a person with one or more chronic diseases working full-time (relative to being inactive) are around 34% lower than those for a person without a chronic disease; however, the odds of a person with one or more chronic diseases working part-time (relative to being inactive) are around 27% lower than those for a person without a chronic disease.

The results of the model including each individual chronic disease indicate that even after controlling for other factors, the presence of diabetes, stroke and psychiatric conditions (which are associated with a significant reduction in the odds of participation) are also associated with a larger reduction in the odds of working full-time (relative to being inactive) compared to part-time (relative to being inactive).²² As an example, the odds of a person with a stroke working full-time (relative to being inactive) are around 67% lower than those for a person without a stroke. However, the odds of a person with a stroke working part-time (relative to being inactive) are only around 39% lower than those of someone without a stroke. The effect for high blood pressure and heart disease (the other two diseases which were found to be significantly related to participation) is the reverse, with the impact of working full-time (relative to being inactive) being less than the impact of working part-time (again relative to being inactive). However, the differences between the effects for full-time and part-time for high blood pressure were not found to be significant at the 95% level.

For those with asthma, high cholesterol, migraine or cancer the odds of being in each of the employment states are not significantly different from those without these diseases (relative to being inactive).²³

²² The full-time and part-time coefficients for diabetes were significantly different from each other at the 95% level, as were those for psychiatric conditions and stroke.

²³ The results for unemployment are subject to large standard errors as they are based on small groups.

Table 5 – Estimated odds ratios for each labour market outcome – pooled multinomial logistic regression – grouped and individual diseases: 2002/03 to 2004/05

Disease	Odd ratios		
	Full-time employment	Part-time employment	Unemployment
Any chronic disease (base=no known chronic disease)	0.659***	0.733***	0.878
Asthma (base=no asthma)	0.923	0.892*	0.962
High blood pressure (base=no high blood pressure)	0.849**	0.833**	0.842
High cholesterol (base=no high cholesterol)	0.916	0.923	0.997
Heart disease (base=no heart disease)	0.539***	0.530***	0.417***
Diabetes (base=no diabetes)	0.497***	0.697***	0.985
Stroke (base=no stroke)	0.327***	0.612**	0.446**
Migraine (base=no migraine)	0.923	0.989	1.257*
Psychiatric conditions – male (base=male no psychiatric conditions)	0.265***	0.472***	0.550***
Psychiatric conditions – female (base=female no psychiatric conditions)	0.531***	0.679	0.958
Cancer (base=no cancer)	0.945	0.935	0.828

Source: SoFIE Waves 1-3 Version 4, unweighted, Statistics New Zealand

Notes:

1. The odds ratios for the summary chronic disease indicator and for individual diseases are derived from different models. Odds ratios for summary chronic disease indicator are derived from Appendix Table D4, while those for individual chronic diseases are derived from Appendix Table D5. The footnotes from that table apply to this table.
2. The following factors were held constant: gender; region; age (and whether 50 years of age or above); highest qualification; study status; marital status; place of birth; ethnicity; children; household income less personal income; years in paid employment; and unemployment rate at the time of the interview.
3. *Significant at the 90% level. **Significant at the 95% level. ***Significant at the 99% level.

Table 6 summarises the marginal effects for each disease. Looking at the result for grouped chronic diseases indicates that an average person with chronic diseases is 5.2 percentage points less likely to be full-time, 0.5 percentage points more likely to be part-time, 0.4 percentage points more likely to be unemployed and 4.3 percentage points more likely to be inactive, than an average person with no chronic diseases.

Turning to the model including the individual chronic diseases shows that, for an average person, having heart disease, diabetes, a stroke or a psychiatric condition is highly significant in reducing the chance of working full-time and increasing the chance of being inactive. For example, for an average person, having a stroke is associated with a 19.9 percentage point decrease in the chance of working full-time, a 5.5 percentage point increase in the chance of working part-time and a 14.5 percentage point increase in the chance of being inactive. So while the odds of working part-time rather than being inactive for those with a stroke are higher than the odds for those without a stroke, the chance of working part-time for those with a stroke is higher than for those without a stroke (ie, some of those with a stroke who do not work full-time work part-time instead).

Table 6 – Estimated marginal effects for each labour market outcome – pooled multinomial logistic regression – grouped and individual diseases: 2002/03 to 2004/05

Disease	Marginal effects			
	Full-time employment	Part-time employment	Unemployment	Inactive
Any chronic disease	-0.052***	0.005	0.004**	0.043***
Asthma	-0.004	-0.006	0.001	0.009
High blood pressure	-0.012	-0.006	-0.001	0.019**
High cholesterol	-0.010	-0.001	0.002	0.009
Heart disease	-0.061**	-0.017	-0.006	0.084***
Diabetes	-0.120***	0.026	0.013*	0.081***
Stroke	-0.199***	0.055*	-0.001	0.145***
Migraine	-0.020*	0.007	0.007**	0.006
Psychiatric conditions – male	-0.185***	0.030*	0.014	0.141***
Psychiatric conditions – female	-0.099***	0.017	0.008	0.074***
Cancer	-0.002	-0.002	-0.003	0.007

Source: SoFIE Waves 1-3 Version 4, unweighted, Statistics New Zealand

Notes:

1. The marginal effects for the summary chronic disease indicator and for individual diseases are derived from different models. Marginal effects for summary chronic disease indicators are derived from Appendix Table D4, while those for individual chronic diseases are derived from Appendix Table D5. The footnotes from those tables apply to this table.
2. The following variables were held at the mean value for the whole sample: gender; region; age (and whether 50 years of age or above); highest qualification; study status; marital status; place of birth; ethnicity; children; household income less personal income; years in paid employment; and unemployment rate at the time of the interview.
3. *Significant at the 90% level. **Significant at the 95% level. ***Significant at the 99% level.

6 Self-rated health and labour market participation

This section explores the relationship between self-rated health and labour market participation. It begins by revisiting the reasons for considering self-rated health and for using the various modelling approaches. Basic descriptive statistics related to self-rated health are then presented, before the results from the corresponding pooled (cross-sectional) models, as considered in the previous section, are summarised. The results of the fixed and correlated random effects (longitudinal) logistic regression models, and the equivalent models using an adjusted measure of self-rated health, are then discussed. Again, words such as “impact” and “effect” are used to describe relationships but do not denote causation. Full tables of results from the main models, including unweighted means and standard deviations for the self-rated health variable, can be found in Appendices E, F and G where the reference categories are labelled.

6.1 Models used

As outlined in Section 4, two measures of health are available in all three waves of SoFIE: chronic diseases; and self-rated health. Given the issues with both of these health measures, and the conclusion of an earlier literature review in the area, it is preferable to consider the relationships between both of these measures and labour force participation. This section therefore begins by reporting basic descriptive statistics related to self-rated

health and then summarises the results from the corresponding pooled (cross-sectional) models as presented in the previous section. The results of the pooled logistic regression model are presented to enable comparison with both the equivalent models for self-rated health and with the subsequent panel models for self-rated health. Where results of the standard logistic regression models are discussed in this paper potential endogeneity bias should be remembered (as explained in Section 4.3.1).

The results of the fixed and correlated random effects (longitudinal) logistic regression models and the equivalent models using an adjusted measure of self-rated health are then presented. These models make use of the longitudinal nature of the data and aim to resolve some of the endogeneity issues identified in Section 4. Ideally these models would have been applied to the models including individual chronic diseases but owing to small numbers in some groups and that the diagnosis of chronic diseases is slow changing this was not possible. Unlike the standard logistic regression results (for which the assumptions may not be satisfied owing to endogeneity, thus possibility resulting in inconsistent (and biased) regression coefficients) the panel models account for some forms of endogeneity, and thus should produce estimates that are consistent and unbiased, if the model assumptions are satisfied.

In addition to the above, the health coefficients from the standard pooled regression, the fixed effects and the correlated random effects models are interpreted differently. The coefficients from the pooled regressions indicate how health levels are related to the chance of participation for a cross-section, while the health coefficients from the fixed and correlated random effects models use longitudinal data to indicate how health shocks are related to participation (although health level is also estimated in the latter model). The fixed effects model attempts to explain variation within (rather than between) respondents over time, making direct comparison of the odds ratios with those from the standard and random effects models problematic.

All three types of models identify a highly significant relationship between health and labour force participation; however, no model is perfect. The best model is found to be the fixed effects model. However, this model is not without its drawbacks. By definition a fixed effects model excludes all those for whom participation does not change over the period from the analysis, meaning there is no estimate of the relationship between health and participation for those continually inactive. Also the fixed effects model focuses on variation in participation for each respondent. This means that only within (rather than within and between) person variation is considered. Finally, there may be other types of endogeneity present that it is not possible to account for using a fixed effects model; for example, unobserved variables that change over time and are related to the explanatory variables. Assuming that this is not the case, the fixed effects model should produce estimates that are consistent (and unbiased).

The cross-sectional pooled regression considers the relationship between health state and participation for all respondents but does not consider within person variation. It is also not possible to control for any types of endogeneity so the results are likely to be biased. The correlated random effects model considers within and between person variation and includes an estimate of the average health level for respondents as well as looking at health shocks. However, if the assumption that the only correlation between health shocks and the unobserved variables that are fixed over time is through average health is not valid, or if average health is itself correlated with unobserved variables, then the coefficients from this model may be biased. Further, other types of endogeneity such as unobserved variables that change over time cannot be accounted for. Owing to the pros and cons of each of the models, and to allow comparisons between the models to be

seen, all of the model results are presented in this section to illustrate the different types of relationships identified between health and labour force participation.

6.2 Unadjusted self-rated health

First, basic descriptives are considered. Table 7 shows the distribution of self-rated health across the population. Around three-quarters of the people consider themselves to be in excellent or very good health. A further 18% feel they are in good health. The remaining 6% feel they are in fair or poor health.

Table 7 – Distribution and participation rates by self-rated health: 2002/03 to 2004/05

Health status	Distribution (%)	Participation rate (%)
Excellent health	41.3	87.7
Very good health	33.9	85.6
Good health	18.4	77.0
Fair health	5.0	56.9
Poor health	1.4	29.1
Total	100.0	82.7

Source: SoFIE Waves 1-3 Version 4, standard longitudinal weights, Statistics New Zealand

Note: Results are for those aged 15-64 and are not full-time students. Data for all three waves is pooled together to create an average rate.

Table 7 also shows that, as with the individual diseases, participation decreases as health declines. Around 88% of those in excellent health participate in the labour market, compared to just 29% of those in poor health.

6.2.1 Standard pooled regression

The odds ratios for the pooled logistic regression model where the chronic disease variables have been replaced by the self-rated health variable are shown in Figure 4. The participation rates for those in excellent health appear to be above those for people in very good health. However, the odds of participating for those of very good health are not significantly different from those of excellent health, once other factors are controlled for. Being in good, fair or poor self-rated health is associated with a reduction in the odds of participating compared to those of excellent self-rated health, by 46%, 76% or 92% respectively.

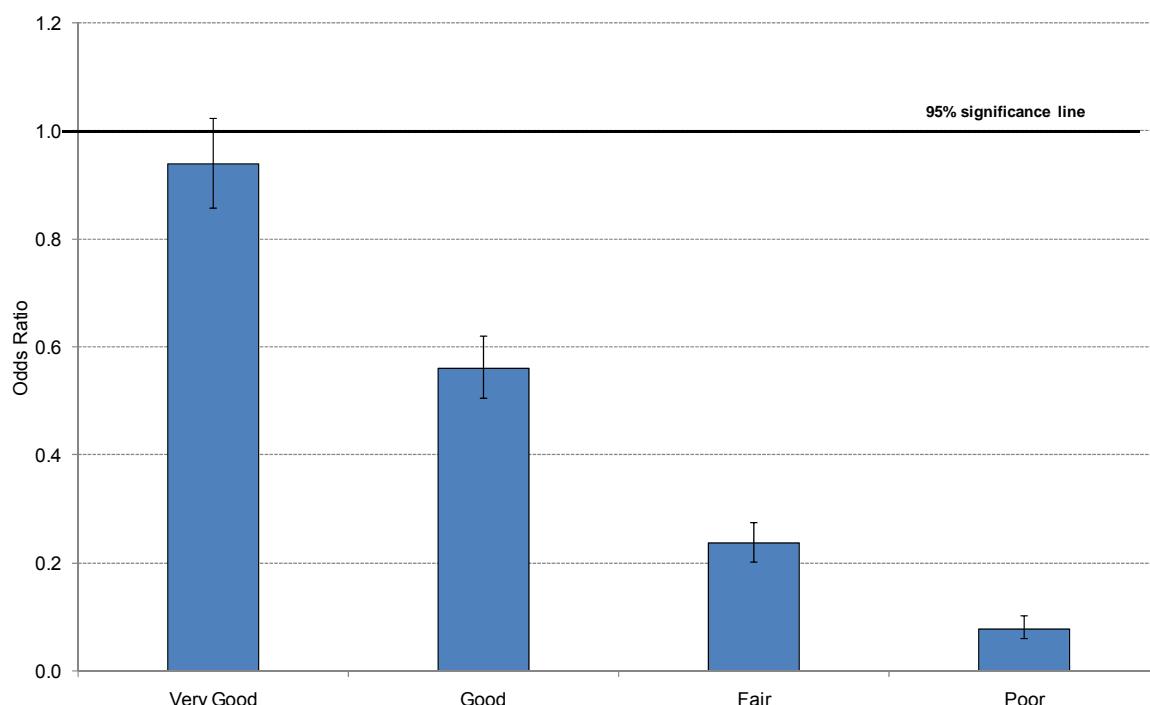
The equivalent marginal effects indicate that being of good, fair or poor health reduces the probability of participating by 6, 22 and 50 percentage points respectively (see Table 13).²⁴ The impact of being in these health states is significantly different from being in excellent health but also the impact of each health state is significantly different from one another (ie, the magnitude of the relationship between being in fair health and participation is less than that between poor health and participation). The R^2 for the self-rated health model is slightly higher than that for the individual diseases (0.3227 compared to 0.3090), suggesting self-rated health explains slightly more of the variation. An alternative test

²⁴ In order for the marginal effects to be comparable to those from the fixed and random effects model they are calculated as if the health states are independent. This means the marginal effects are slightly higher than if independence had not been assumed.

statistic to compare the models is the area under the Receiver Operating Characteristic (ROC) curve.²⁵ As with the R^2 these diagnostics indicate that the model including self-rated health performs slightly better than the model including individual diseases, with the area under the ROC curve of 0.871 and 0.864 respectively.

The only other variable that has odds of participating in the labour force of a similar magnitude to those for fair or poor health is having a young child for females (a reduction in the odds of participating of around 90%). This indicates the relative magnitude of the relationship between fair/poor health and participation.

Figure 4 – Estimated odds ratios of participating in the labour force – pooled logistic regression – self-rated health: 2002/03 to 2004/05



Source: SoFIE Waves 1-3 Version 4, unweighted, Statistics New Zealand

Notes:

1. Odds ratios are derived from Appendix Table E2 and are relative to excellent health. The footnotes from that table apply to this chart.
2. The following factors were held constant: gender; region; age (and whether 50 years of age or above); highest qualification; study status; marital status; place of birth; ethnicity; children; household income less personal income; years in paid employment; and unemployment rate at the time of the interview.

Table 8 indicates that around 18% of those in excellent health are participating in part-time work compared to 31% of those in poor health. As self-rated health decreases, the likelihood of working full-time appears to fall and the likelihood of working part-time to increase. This is consistent with the earlier observation that those who have been diagnosed with a chronic disease are relatively more likely to work part-time.

²⁵ This curve looks at the trade-off between false negative and false positive rates for the model at various cut-off points; in other words, the ROC curve is the representation of the trade-offs between sensitivity and specificity. The larger the area (with the maximum being one) the better the diagnostic test.

Table 8 – Labour market outcome rates by self-rated health: 2002/03 to 2004/05

Health status	Labour market outcome (%)			
	Full-time employment	Part-time employment	Unemployment	Total participation
Total	78.4	19.0	2.7	100.0
Excellent health	80.4	17.5	2.1	100.0
Very good health	78.6	19.1	2.3	100.0
Good health	75.9	20.1	3.9	100.0
Fair health	64.5	28.7	6.8	100.0
Poor health	58.2	31.1	10.6	100.0

Source: SoFIE Waves 1-3 Version 4, standard longitudinal weights, Statistics New Zealand

Note: See footnotes Table 5.

Table 9 shows the odds ratios from a multinomial logistic regression when other factors are controlled for. Even when other factors are held constant, being of good, fair or poor health is associated with a larger reduction in the odds of working full-time (relative to being inactive) as opposed to part-time (relative to being inactive).²⁶ For example, being in fair health rather than excellent is associated with an 83% reduction in the odds of working full-time (relative to being inactive), compared to a 61% reduction in working part-time (relative to being inactive).

Table 9 – Estimated odds ratios for each labour market outcome – pooled multinomial logistic regression – self-rated health: 2002/03 to 2004/05

Health status	Odds ratios		
	Full-time employment	Part-time employment	Unemployment
Very good health	0.925	0.974	1.037
Good health	0.514***	0.626***	0.965
Fair health	0.174***	0.389***	0.537***
Poor health	0.054***	0.139***	0.291***

Source: SoFIE Waves 1-3 Version 4, unweighted, Statistics New Zealand

Notes:

1. These odds are derived from the data in Appendix Table E3. For full footnotes see that table.
2. The following factors were held constant: gender; region; age (and whether 50 years of age or above); highest qualification; study status; marital status; place of birth; ethnicity; children; household income less personal income; years in paid employment; and unemployment rate at the time of the interview.
3. *Significant at the 90% level. **Significant at the 95% level. ***Significant at the 99% level.

Table 10 shows the marginal effects from the same model. The results show that, for an average person, being in any health state other than excellent is associated with a reduced chance of working full-time. For the majority of health states (other than poor health) this reduction in the chance of working part-time is balanced by increases (both significant and not significant) in the chance of working part-time, being unemployed or being inactive. For those in poor health the chance of working part-time is also reduced compared to someone of excellent health. An average person in poor health, compared to an average person in excellent health, is 49.1 percentage points less likely to work full-time, 4 percentage points less likely to work part-time and 51.9 percentage points more likely to be inactive.

²⁶ The coefficients for full-time and part-time were significantly different from each other at the 95% level.

Table 10 – Estimated marginal effects for each labour market outcome – pooled multinomial logistic regression – self-rated health: 2002/03 to 2004/05

Health status	Marginal effects			
	Full-time employment	Part-time employment	Unemployment	Inactive
Very good health	-0.014*	0.005	0.002	0.007
Good health	-0.095***	0.009	0.012***	0.074***
Fair health	-0.308***	0.043***	0.015***	0.250***
Poor health	-0.491***	-0.040**	0.012	0.519***

Source: SoFIE Waves 1-3 Version 4, unweighted, Statistics New Zealand

Notes:

1. These marginal effects are derived from the data in Appendix Table E3. For full footnotes see that table.
2. The following factors were held at the mean value for the whole sample: gender; region; age (and whether 50 years of age or above); highest qualification; study status; marital status; place of birth; ethnicity; children; household income less personal income; years in paid employment; and unemployment rate at the time of the interview.
3. *Significant at the 90% level. **Significant at the 95% level. ***Significant at the 99% level.

6.2.2 Fixed and correlated random effects panel models

The standard pooled logit model considered the impact of the self-rated health state at a given point in time, but, unlike panel models, it is not possible to adjust for any possible types of endogeneity that might exist. The panel models estimate the health effect in a slightly different way than standard cross-sectional logistic regressions: considering changes in health (health shocks) over time. Table 11 shows transitions across the self-rated health state between two consecutive waves. The results indicate that, while the majority of respondents do not change health state between waves, there is some movement both to better health and poorer health between consecutive waves. For example, while around two-thirds of those in excellent health in one wave remain there in the consecutive wave, the remaining third move to poorer health.

Table 11 – Changes in self-rated health in consecutive waves: 2002/03 to 2004/05

	Health status in following wave (t+1)					
	Excellent	Very good	Good	Fair	Poor	
Health status in wave t	Excellent	67.5	24.7	6.6	1.0	0.1
	Very good	27.9	50.1	19.3	2.4	0.3
	Good	12.5	31.8	44.8	9.4	1.4
	Fair	3.4	13.7	34.2	39.5	9.3
	Poor	3.0	5.2	17.2	30.7	43.6

Source: SoFIE Waves 1-3 Version 4, standard longitudinal weights, Statistics New Zealand

Note: Results are for those aged 15-64 and are not full-time students. Data for changes between Wave 1 and Wave 2 and between Wave 2 and Wave 3 are pooled together to create an average rate.

The fixed effects model looks at how changes in the explanatory variables are related to changes in labour force participation, when other unobserved time constant variables such as genetics, are controlled for. Table 12 shows how changes in participation compare with changes in self-rated health between two consecutive waves. The first part of the table is based on those who are participating in Wave 1 or Wave 2 (21,610).²⁷ The percentage indicates the proportion of these who move to not participating in Wave 2 and Wave 3 respectively. So around 4% of those who report their health to be excellent in Waves 1 and 2 or in Waves 2 and 3 respectively move from participating to not

²⁷ Unweighted count.

participating. The proportion moving out of participation is generally higher for those who experience a decline in self-rated health. For example, 16% of those who report their health to be excellent in Wave 1 or Wave 2 but fair or poor in Wave 2 or Wave 3 respectively move out from participating to not participating. The second part of the table shows the reverse of this; that is, those who are not participating in Wave 1 or Wave 2 (4,975).²⁸ Of those who are not participating in Wave 1 and Wave 2 those who experience negative changes to self-rated health are less likely to move into participation. For example, 40% of those who report being in excellent health in two consecutive waves who are not participating in Wave 1 or Wave 2 move into participation in Wave 2 or Wave 3 respectively. For those who report their health changing from excellent to fair/poor between waves, only 34.1% move into participation.

Table 12 – Changes in participation compared with changes in self-rated health in consecutive waves: 2002/03 to 2004/05

		% moving from participating in Wave t to non-participating in Wave t+1 (N=21,610)			
		Health status in following wave (t+1)			
		Excellent	Very good	Good	Fair or poor
Health status in wave t	Excellent	3.8	5.5	6.5	16.2
	Very good	4.3	3.9	6.1	11.5
	Good	6.4	5.3	6.7	14.3
	Fair or poor	S	13.5	8.2	15.1
		% moving from non-participating in Wave t to participating in Wave t+1 (N=4,975)			
		Health status in following wave (t+1)			
		Excellent	Very good	Good	Fair or poor
Health status in wave t	Excellent	40.2	39.1	41.7	34.1
	Very good	36.3	34.5	28.6	20.5
	Good	38.2	33.3	20.2	15.9
	Fair or poor	44.4	40.4	21.4	9.6

Source: SoFIE Waves 1-3 Version 4, standard longitudinal weights, Statistics New Zealand

Notes:

1. Results are for those aged 15-64 and are not full-time students. Data for changes between Wave 1 and Wave 2 and between Wave 2 and Wave 3 is pooled together to create an average rate.
2. Fair and poor are combined owing to small numbers in some of the categories.
3. S - This cell is suppressed as it is subject to sample error too great for most practical purposes.

Of those longitudinal working age non-student respondents in the survey period, around 14% (5,710) experience a change in participation status and have non-missing data in two consecutive waves for the variables of interest. This is the group that are used for analysis in the fixed effects logistic model. Around 20% of these experience a change in self-rated health between two consecutive waves.

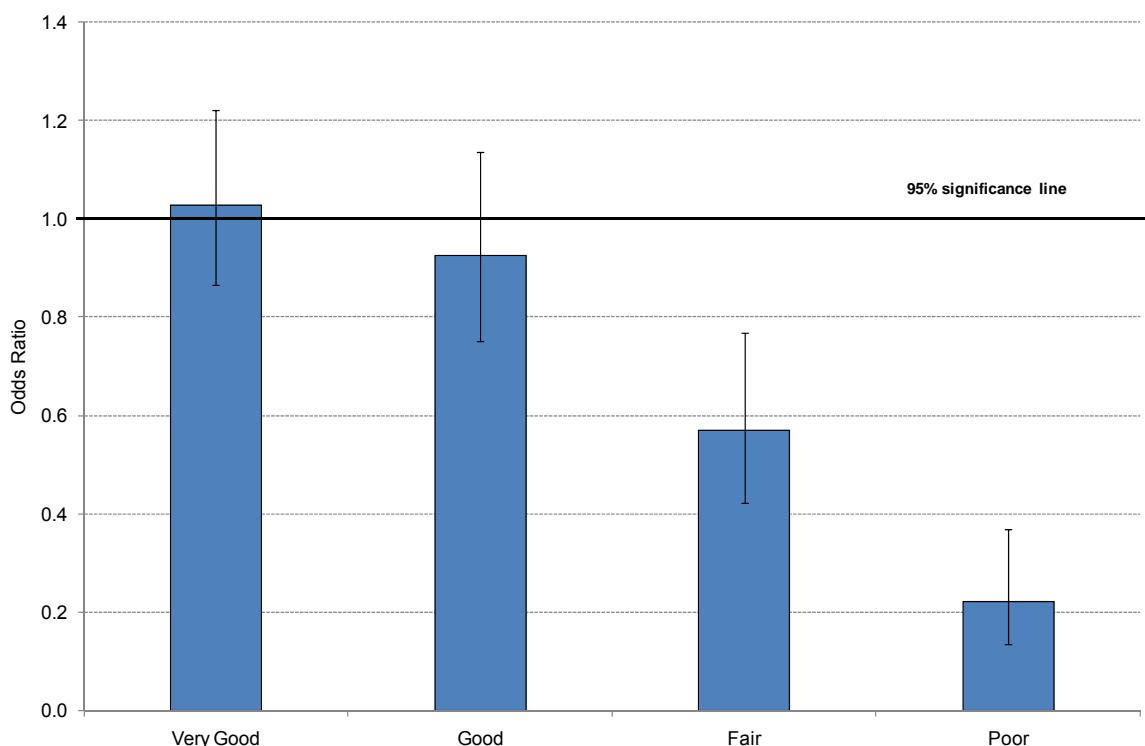
In the fixed effects model, the effect of any variables that are non-time varying over the survey period cannot be estimated. In this case the effect of gender and place of birth on labour force participation are not estimated. Also, following best practice, those variables that are little or slow changing (eg, ethnicity and highest qualification); or which could be impacted on by health changes (eg, studying status and years in paid employment) are excluded from both the fixed and random effects models. Full results are presented in Appendix Table F1. The results for the non-health variables indicate that a movement to the South Island from Auckland is associated with a significant reduction in the chance of participating. A change to having a partner who does not work reduces the chance of

²⁸ Unweighted count.

participation, possibly indicating couples taking early retirement together. For females, having a child is associated with an 88% decrease in the odds of participating.

Figure 5 shows the odds ratios for the self-rated health categories from the fixed effects regression model. The results indicate that there is not a significant relationship between a move into very good or good health from excellent health and the chance of participating. However, a move to fair or poor health from excellent health is associated with a 43% or 78% reduction in the odds of participating respectively for each person (equivalent to the odds ratios of 0.57 and 0.22). It should be remembered that the fixed effects model attempts to explain variation in participation for each respondent; that is, only within, rather than within and between, person variation is considered.

Figure 5 – Estimated odds ratios of participating in the labour force – fixed effects model – self-rated health: 2002/03 to 2004/05



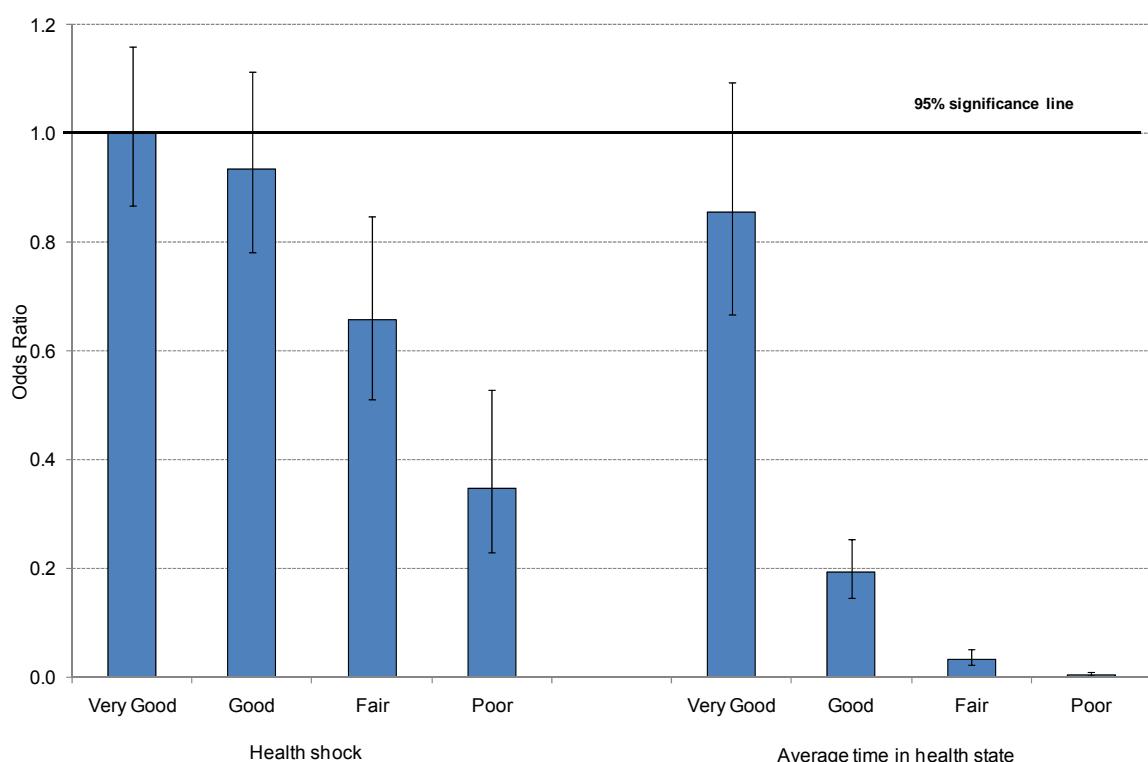
Source: SoFIE Waves 1-3 Version 4, unweighted, Statistics New Zealand

Note: Odds ratios are derived from Appendix Table F1. The footnotes from that table apply to this chart. They are the odds within people as between respondent variation is not considered. As such, they are not directly comparable to the odds from the pooled or random effects models. The following factors were held constant: region; age (and whether 50 years of age or above); marital status; children; household income less personal income; and unemployment rate at the time of the interview.

Finally, the correlated random effects model was estimated. A standard random effects model allows for time constant unobserved variables that are fixed over time but that are uncorrelated with the explanatory variables in the model. The concern here is that health is correlated with the unobservables. If this were not the case then the coefficients for health would not be biased. Using a correlated random effects model it is assumed that the only correlation between the health and the unobservables is through average health and includes a variable indicating average health in a standard random effects model. Full information on the model, including the equation and the assumed relationship, can be found in Appendix C.

Figure 6 summarises the odds ratios for the health shock variables from the correlated random effects model. Full results can be found in Appendix Table F2. Looking at the health shocks indicates that, as in the fixed effects model, only a fair or poor health shock from excellent is significant in affecting participation, reducing the odds of participating by 34% and 65% respectively (slightly lower than the within person odds estimated in the fixed effects model of 43% and 78%). What is more influential is the average time in a health state of a person. Spending more time in good, fair or poor health significantly reduces the odds of participating relative to being in excellent health. Being in good, fair or poor health for all three waves reduces the odds of participating by 80%, 97% and 99% respectively. The model summary statistics indicate that 59% of the total variation is contributed by the panel-level variance component.

Figure 6 – Estimated odds ratios of participating in the labour force – correlated random effects model – self-rated health: 2002/03 to 2004/05



Source: SoFIE Waves 1-3 Version 4, unweighted, Statistics New Zealand

Note: Odds ratios are derived from Appendix Table F2. The footnotes from that table apply to this chart. The following factors were held constant: gender; region; age (and whether 50 years of age or above); marital status; place of birth; children; household income less personal income; and unemployment rate at the time of the interview.

6.2.3 Model comparisons

An alternative way to look at the results is to calculate the marginal effects. The odds ratios for the fixed effects model are not directly comparable with the odds from the pooled or random effects regression because the variation is coming from the variation within individuals. However, an average marginal effect can be computed for the fixed effects model to enable relative comparisons between groups of people with different covariates. These are shown in Table 13. The results of the fixed and correlated random effects models indicate that even after controlling for time invariant unobserved variables, poorer health is still associated with a reduction in the chance of participating (shown by the lower marginal effect for the panel models than the pooled model, consistent with the

results from the odds ratios – a higher ratio of which indicates a lower reduction in the chance of participating). This possibly indicates that there are time constant unobserved variables that should have been included in the standard pooled regression that are positively correlated with health and participation (eg, motivation), and hence the coefficients in the pooled model are systematically overestimated. Further, while the magnitude of the impact of health shocks is lower, they are still significant in reducing the chance of participating when average health state over the period is allowed for (shown by the results of the correlated random effects model).

Table 13 – Marginal effects by self-rated health: 2002/03 to 2004/05

Health status	Marginal effects		
	Pooled regression	Fixed effect model	Random effects model
Very good health	-0.006	0.006	0.000
Good health	-0.065***	-0.018	-0.003
Fair health	-0.222***	-0.127***	-0.019***
Poor health	-0.496***	-0.340***	-0.065***
Average time in very good health	-	-	0.006
Average time in good health	-	-	-0.062***
Average time in fair health	-	-	-0.127***
Average time in poor health	-	-	-0.201***

Source: SoFIE Waves 1-3 Version 4, unweighted, Statistics New Zealand

Notes:

1. Marginal effects are derived from Appendix Tables E2, F1 and F2 holding all other factors at the mean value for the whole sample. The footnotes from those tables apply to this table.
2. For the pooled regression the effect is of being in the health state rather than being in excellent health. For the fixed and random effects models the marginal effects for each health state are the effect of a health shock from excellent into that health state. The final marginal effects for the random effects model are the effect of spending all waves in a health state rather than all waves in excellent health.
3. The marginal effects for the fixed effects model are pseudo marginal effects calculated based on the overall sample mean of the predicted probability of a positive outcome.
4. *Significant at the 90% level. **Significant at the 95% level. ***Significant at the 99% level.
5. These marginal effects assume that the health states are independent. Accounting for the fact that the health states aren't independent reduces the marginal effects slightly.

The results for all three types of models identify a highly significant relationship between health and labour force participation; however, no model is perfect and the type and magnitude of the impact estimated varies. Tests to determine which model is preferred were carried out. A likelihood-ratio test indicated that the proportion of variation from the panel component of the random effects model was significantly different from zero and as such the panel element of the data should not be ignored (thus the standard logistic regression results are likely to be biased). This means that the panel models are preferable to the pooled estimator.

A Hausman test comparing the fixed effects model with the uncorrelated random effects model indicated correlation between the unobserved individual level effects and the other covariates, hence the use of the correlated random effects model (making the assumption that the correlation between the unobserved individual level effects and other covariates is only with health and only through average health).

However, a significant Hausman test comparing the fixed effects and correlated random effects model indicates that the unobserved individual level effects are still correlated with

the covariates in the fixed effects model, even after controlling for the correlation between these unobserved variables and health. This may be: correlation between the unobserved variables and non-health covariates; correlation between the unobserved covariates and health shocks, if the expected value of the unobservables is not equal to a linear function of the average time spent in each health state (which was assumed for the correlated random effects model); or correlation between the average health level variable and the unobserved variables.

This correlation means that the health coefficients (both health level and/or health shocks) from this model may be biased. Further, other types of endogeneity such as unobserved variables that change over time cannot be accounted for. This indicates that the preferred model is the fixed effects model.

However, this model is not without its drawbacks. By definition a fixed effects model excludes all those for whom participation does not change over the period from the analysis, meaning there is no estimate of the relationship between health and participation for those continually inactive. It seems theoretically sensible that some people will be in consistently poor health over the periods considered and not participate as a result of this. These people will not be included in any estimates of impact from this model. Also the fixed effects model focuses on variation in participation for each respondent. This means that only within (rather than within and between) person variation is considered. Finally, there may be other types of endogeneity present that it is not possible to account for using a fixed effects model; for example, unobserved variables that change over time and are related to the explanatory variables. Assuming that this is not the case, the fixed effects model should produce estimates that are consistent (and unbiased).

As the models look at the relationship between health and labour force participation in different ways all the results are informative in their own way. The key result is that a significant relationship between health and participation was identified in all of the models.

6.3 Adjusted self-rated health

In the previous section it was found that there was a significant relationship between health and participation even after accounting for unobserved variables. However, these results may occur owing to respondents using their health status to rationalise their participation; that is, reporting their health to be worse than it actually is to justify the fact that they are not participating. In previous studies, for example Disney *et al* (2003), one approach to try to remove this rationalisation bias from health measures has been to model self-rated health using more objective health related variables. Estimates from such a model have then been standardised and included in models to estimate the relationship between health and labour force participation in place of self-rated health. This approach was therefore used to try to rid the self-rated measure of health in SoFIE of its potential rationalisation bias. Full details of how the adjusted health measure was calculated and used in the models can be found in Appendix C. These results complement the findings in the previous section. The key finding is that, even when self-rated health is adjusted to account for potential rationalisation bias, a highly significant relationship is still found between health and labour force participation. This approach also leads to the fixed effects model being identified as the preferred model. The results strengthen the conclusions made in the previous section, in that it seems that the relationship identified between health and labour force participation is not owing to rationalisation bias.

6.3.1 Calculation of adjusted health measure

The following measures are available for each respondent in every wave of SoFIE: whether a respondent has ever smoked; the presence of each individual chronic disease; and the receipt of a health or illness related benefit.^{29 30} Table 14 shows for each health state the proportion of people who report each health related measure. For example, 38% of those in excellent health have been diagnosed with one or more chronic diseases, compared to 84.8% of those in poor health. It shows that all three measures are correlated to some extent with health. These three measures are also more objective than self-rated health. These variables will therefore be termed “objective health measures”.

While Table 14 shows that only 4%, 7.2% and 14.8% of those who consider their health to be excellent, very good and good, respectively, receive a health related benefit. However, it is important to remember that these groups account for a large proportion of those who receive health related benefits once the relative size of these health states is considered. Table 7 shows that around 94% of the population consider themselves to be in excellent, very good or good health. Combining the figures from Table 14 and Table 7 indicates that around seven-tenths of those receiving a health related benefit consider themselves to be in good, very good or excellent health. This is well below the level for the population as a whole but it is still higher than may have been expected. This highlights the possible issues with the survey questions that measure self-rated health (discussed in Section 4.2.2) and the mis-match between health and disability; for example, a person who is blind may be eligible for a disability benefit (included here within health related benefit) but may consider themselves to be in excellent health. This finding is along similar lines to international evidence that suggests that on average one in three qualified recipients of a

²⁹ The latter assumes that people receiving a health related benefit are less healthy than people who don't. Also note that some illness benefits included are joint income tested so this variable is likely to have a lower correlation with health for those wealthier households.

³⁰ These variables are defined in Appendix Tables A1, A2 and A3.

disability related benefit claim to have no subjectively perceived disability that limits their daily activity (OECD, 2003).

Table 14 – Distribution of objective health measures with self-rated health: 2002/03 to 2004/05

Objective health measure	Self-rated health				
	Excellent	Very good	Good	Fair	Poor
Any chronic disease	38.0	51.4	61.7	77.1	84.8
Asthma	14.6	19.6	21.7	27.4	34.1
High blood pressure	8.2	15.3	22.9	33.3	39.3
High cholesterol	8.8	13.8	18.2	24.6	33.5
Heart disease	0.8	2.2	4.9	11.6	23.6
Diabetes	0.6	2.1	5.9	12.4	21.0
Stroke	0.3	0.7	1.6	4.3	7.8
Migraine	10.2	13.7	17.2	22.0	28.7
Psychiatric conditions	5.2	8.8	14.4	25.6	36.6
Cancer*	2.2	3.8	4.3	7.7	7.9
Smoked	38.5	48.5	55.9	60.3	66.2
Health related benefit	4.0	7.2	14.8	36.5	64.2

Source: SoFIE Waves 1-3 Version 4, standard longitudinal weights (* adjusted longitudinal weight), Statistics New Zealand

Notes:

1. The figures in each cell are the proportion in a certain health state that report the health related measure.
2. See footnotes Table 5.

It was shown previously that chronic disease presence is correlated with participation. There is also correlation between health benefit receipt and labour market participation and a weak correlation between whether a person has ever smoked and labour market participation. However, it is sensible to assume that if true health was measured correctly these objective health measures should only affect participation through this health measure once other factors are controlled for.

Given this relationship, these objective health measures (along with a set of other health related variables) were used to model self-rated health for each year. The results of these models can be found in Appendix Table G1.

Looking at the model results indicates that all of the objective health measures are highly significant in explaining self-rated health. However, overall, the models only explain around 11% of the variation in the data. In terms of interpreting the model results, a higher value of self-reported health means poorer health. This means that positive coefficients on the objective health measure, for example 0.418 for those who have cancer in Wave 1, are associated with an increase in the predicted probability that an individual will be in poor health and a decrease in the predicted probability that they will be in excellent health. With this in mind the largest health impact is seen from those receiving health related benefits (a coefficient of 1.286 in Wave 1) while the most influential health condition is diabetes (a coefficient of 1.086). The least influential health condition is high cholesterol (a coefficient of 0.177). Looking at the non-health coefficients indicates that health is generally predicted to be poorer for those outside Auckland; those born outside of New Zealand; those of non-NZ/European ethnicity; older respondents;

those with no qualifications; and those with no partner relative to the reference categories. Health is generally predicted to be better for females; those with tertiary education; those who are undertaking some form of study; and those with higher household income.

The results of these models were used to create an adjusted health stock. The probability of being in poor health was predicted for each person. This probability was then standardised across all respondents to give a continuous measure of adjusted health status (or adjusted health stock). For all respondents (including those over working age) this adjusted measure therefore had a mean zero and standard deviation of one. As with self-rated health, a higher adjusted health stock indicates poorer health. This is illustrated in Table 15 where the mean and standard deviation of the adjusted health stock are presented for each self-rated health state. The mean and standard deviation of this health stock for those of interest (working age non-students) is less than zero. This is because those with generally poorer health (respondents aged 65 and over) are included in the model to create the adjusted health stock, but are excluded from the analysis to determine the relationship between health and participation. As with unadjusted self-rated health, this was done to ensure the total distribution of the adjusted health measure reflected that of health in the total population.

Table 15 – Mean and standard deviation of adjusted health measure for sample of interest – self-rated health: 2002/03 to 2004/05

Health status	Mean	Standard deviation
Excellent	-0.306	0.171
Very good health	-0.230	0.329
Good health	-0.049	0.708
Fair health	0.522	1.590
Poor health	1.556	2.523
Total	-0.167	0.652

Source: SoFIE Waves 1-3 Version 4, standard longitudinal weights, Statistics New Zealand

Notes:

1. The health measure is derived based on the standardised probabilities of poor health for all longitudinal respondents from the data in Appendix Table G1. For full footnotes see that table. These means and standard deviations are for those of working age who aren't students.
2. The total figures are for the sample from the pooled and random effects regression. For the fixed effects regression the mean was -0.062 and standard deviation 0.750 indicating those who change participation status during the period are slightly less healthy than those who do not change.

6.3.2 Standard pooled regression

The adjusted health stock was then included in the standard pooled logistic regression in place of individual chronic diseases or self-rated health. Full results can be found in Table G2. This model explains a similar amount of variation in the data as the model including the unadjusted self-rated health and the model including the individual chronic diseases (32% compared to 32.3% and 30.9% respectively).

The coefficients of the non-health variables in this model are little changed from those in the pooled models, including chronic diseases or unadjusted self-rated health. Health is still highly significant in affecting participation even after attempting to adjust for possible incorrect measurement of self-rated health. The coefficient for adjusted health indicates that a one unit increase in the level of health (a move to poorer health) is associated with a 57% reduction in the odds of participating. The adjustment of self-rated health results in difficulties interpreting what a unit change in this measure actually means in the real

world. To give an indication of the dispersion of the adjusted health measure for the sample used in analysis, the average adjusted health level was -0.167. The standard deviation was 0.652 indicating that, while a one unit increase in health reduces the odds of participating by around 57%, many respondents will not experience a one unit change in adjusted health. It is therefore more sensible to consider a one standard deviation increase in adjusted health; this is associated with a 42% reduction in the odds of participating. While the categories of self-rated health are subjective and have no definite boundaries, it is easier to relate to a change from excellent to poor health than to a one unit change in the adjusted health stock. However, the fact that this health measure is still significant in impacting on participation illustrates that health is significantly related to participation even allowing for possible rationalisation.

6.3.3 Fixed and correlated random effects panel models

The adjusted health measure was then included in the fixed and correlated random effect models. The results can be found in Appendix Tables G3 and G4 respectively. The coefficients for the non-health variables were similar to the models for unadjusted self-rated health (Tables F1 and F2).

The key thing to note from the fixed effects model (Appendix Table G3) is that a one standard deviation increase in adjusted health stock (so a poorer health shock) is associated with a 31% increase in the odds of not participating. This is in line with what was found when comparing the pooled and fixed effects model using unadjusted self-rated health (again the odds are not directly comparable as the fixed effects model only considers within person variation).

Turning to the correlated random effects model, both the health shocks (a change in adjusted health) and the average level of adjusted health are significantly related to participation. A one standard deviation increase in adjusted health is associated with a 31% reduction in the odds of participating. Further, the higher the average adjusted health state over a period is (ie, the poorer a person's longer term health) the less chance there is they will participate and this impact is larger than that for a health shock (a one standard deviation increase in the average adjusted health stock is associated with a 52% reduction in the odds a person will participate). Again these results are similar to what was found in the correlated random effects model including unadjusted self-rated health. This illustrates that health is significantly related to participation even allowing for possible rationalisation.³¹

As with the unadjusted health models a likelihood-ratio test for the random effects model indicates that the panel variation is significant and thus a panel model is preferred. A significant Hausman test, comparing the fixed effects and uncorrelated random effects model, indicated that the fixed effects estimator should be used instead of the random effects as the unobserved individual level effects were correlated with the other covariates. This correlation remains even after the correlated random effects model is used. This indicates that the preferred model is the fixed effects model.

³¹ Based on the arguments given by Bound *et al*(1999) it may be expected that lagged health might affect current behaviour because transitions may take time. A lagged adjusted health variable was also included in the fixed and correlated random effects model, along with current health, using just two waves of the data to see if a health shock in a previous period was significantly related to participation. However, unlike in Bound *et al*the lagged effect was not found to be significant on top of current health. It should be noted that this relationship might exist but that with only three waves of data may be hard to estimate.

7 Conclusion

This paper has examined the relationship between health and labour force participation. It found that health was significantly related to participation, using various health measures and even after accounting for certain types of endogeneity. Table 16 summarises the marginal effects from all the models considered.

Table 16 – Summary of marginal effects from all models: 2002/03 to 2004/05

Health status	Marginal effects		
	Pooled regression	Fixed effects model	Random effects model
Any chronic disease	-0.038***	-	-
Asthma	-0.009	-	-
High blood pressure	-0.018***	-	-
High cholesterol	-0.008	-	-
Heart disease	-0.083***	-	-
Diabetes	-0.067***	-	-
Stroke	-0.123***	-	-
Migraine	-0.004	-	-
Psychiatric conditions – male	-0.132***	-	-
Psychiatric conditions – female	-0.065***	-	-
Cancer	-0.007	-	-
Very good health	-0.006	0.006	0.000
Good health	-0.065***	-0.018	-0.003
Fair health	-0.222***	-0.127***	-0.019***
Poor health	-0.496***	-0.340***	-0.065***
Average time in very good health	-	-	0.006
Average time in good health	-	-	-0.062***
Average time in fair health	-	-	-0.127***
Average time in poor health	-	-	-0.201***

Source: SoFIE Waves 1-3 Version 4, unweighted, Statistics New Zealand

Note:

1. All other variables in the models are fixed at the mean value for the whole sample.
2. *Significant at the 90% level. **Significant at the 95% level. ***Significant at the 99% level.
3. For the pooled regression the effect is of being in the health state rather than being in excellent health. For the fixed and random effects models the marginal effects for each health state are the effects of a health shock from excellent into that health state. The final marginal effects for the random effects model are the effects of spending all waves in a health state rather than all waves in excellent health.

Results of the standard pooled regression models that included individual chronic diseases indicated that there was insufficient evidence that those with asthma, high cholesterol, migraine or cancer were any less likely to be participating in the labour market than those without these diseases, once other factors were controlled for. In contrast, psychiatric conditions, stroke, heart disease, diabetes and high blood pressure were all associated with significant decreases in participation once other factors are held constant. Further, for psychiatric conditions, stroke and high cholesterol, the relationship with full-time work was higher than that for part-time work (ie, the chance of working full-time was reduced more than the reduction in the chance of working part-time), suggesting that not

only is the presence of these diseases associated with lower participation but it is also associated with working fewer hours.

Psychiatric conditions for males were associated with the largest reduction in the chance of participation. This was the only disease where the relationship with labour force participation was significantly different by gender. When all other variables were fixed at their mean value, being a male with psychiatric conditions reduces labour market participation by 13.2 percentage points compared to that for males without psychiatric conditions. When all other variables were fixed at their mean value, being a female with psychiatric conditions was associated with a reduced labour market participation by 6.5 percentage points compared to that for females without psychiatric conditions. When all other variables were set at their mean level, being a male with psychiatric conditions was associated with a 1.5 percentage point reduction in participation compared to a female with psychiatric conditions. Following psychiatric conditions, for males the diseases that were associated with the largest fall in participation were strokes (a 12.3 percentage point reduction in labour market participation on average), heart disease (8.3 percentage point reduction), diabetes (6.7 percentage point reduction) and high blood pressure (1.8 percentage point reduction). The effect of the presence of disease did not differ significantly by gender, other than for psychiatric conditions.

These pooled regressions did not allow for possible endogeneity and as a result the coefficients may be biased. As the number of chronic diseases of interest diagnosed during the three waves of data available for analysis is relatively small, the paper moved to consider self-rated health. Fixed and correlated random effects models were used to allow for unobserved variables and an adjusted health measure was constructed to allow for possible rationalisation.

Results of the standard pooled regression models for self-rated health indicated that those in good, fair or poor health are significantly less likely to participate than those of excellent health. Being in good, fair or poor health was associated with a reduction in the chance of participating of 6.5, 22.2 and 49.6 percentage points respectively compared to being in excellent health. The only other variable for which the reduction in the chance of participating in the labour force is of a similar magnitude to that for fair or poor health is having a young child for females. This indicates the relative magnitude of the relationship between fair/poor health and participation. As with the individual chronic diseases, being in good, fair or poor health was associated with a larger reduction in the chance of working full-time than that for working part-time.

The fixed and correlated random effects panel models indicated that a negative health shock significantly reduced the chance of participation even when unobserved time-constant factors were controlled for. The coefficients for the fixed and correlated random effects model are higher (therefore the reduction in the chance of participation lower) than the pooled regression, suggesting possible unobserved variables that are correlated with health and participation. In the fixed effects model only a fair or poor health shock was associated with a significant reduction in participation; reducing the chance of participating by 12.7 and 34 percentage points respectively. The coefficients for the correlated random effects model indicate that a health shock to fair or poor health from excellent health significantly impacted on participation, reducing the chance of participating by 1.9 and 6.5 percentage points respectively. Further, even after controlling for the average time spent in each health state, health shocks were still found to be significantly related to participation. Spending all three waves in good, fair or poor health was associated with a 6.2, 12.7 and 20.1 percentage point reduction in the chance of participating.

All models indicate a significant relationship between health and labour force participation; as such the results complement each other. Tests suggested that the preferred model was the fixed effects model. If it is assumed that there are no unobserved variables that vary over time that are correlated with the explanatory variables, then estimates from this model are consistent (and unbiased). However, this model also had weakness and, owing to the slightly different things being estimated in the different models, results from all three models including self-rated health are informative.

An attempt was then made to remove possible rationalisation from the self-rated health variable. Results of the pooled, fixed and correlated random effects regression models using the adjusted health measure complement those from the unadjusted self-rated health models; that is, they indicate a significant relationship between adjusted health and participation above that from possible rationalisation. As with the longitudinal models that use unadjusted health, the impact of adjusted health on participation is reduced when unobserved time-constant variables are taken into account but remains significant.

The results do not control for unobserved variables that change over time. They also do not allow for the “feedback effect”; that is, that participation could influence health. As such, the results do not address causality but simply establish relationships between health and participation. An exploration of feasible instruments was conducted in order to try to instrument health thus making it possible to take into account variables that vary over time and causality, but no suitable instruments were found.

8 Discussion

8.1.1 Impact on the labour force

The results so far have considered the relationship between health and labour force participation at an individual level. For policy purposes, it is helpful to understand the potential impact of these relationships at the population level. While the magnitude of relationship between health and labour force participation is larger for those of poorer health, if the number in poorer health in the population is small then the estimated impact at the population level may not be large. It is important to remember again that, in this section, words such as “impact” and “effect” are used to describe relationships but do not attempt to explain causation.

Table 17 presents the estimated impact of different diseases and health states. These estimates are based on the marginal effects reported in Tables 3 and 13 and the estimated number of working age non-students in each group.³² They therefore provide an indicator of the workforce impact of poor health. The marginal effects were estimated with the other variables set at the whole sample mean; that is, the figures estimate the additional number of people who may participate in the absence of poor health, if they have average values for the remaining characteristics.³³ The error margin around the estimated impact figures only considers error in the marginal effect. The proportion figures in the table illustrate the proportion of the number of participating working age non-

³² While these figures are themselves estimates from SoFIE, and therefore subject to error, they were taken to be fixed in the calculation of the estimated impact.

³³ The marginal effects estimated using group means rather than whole sample means were broadly the same.

students the count represents. The number of working age non-students estimated to be participating on average over the three waves of SoFIE is 1.84 million. Figures for all diseases and self-rated health states/shocks are reported even if they are not significant. The groups for which the number impacted, or the proportion, crosses zero indicate where the impact is not statistically significant. For the level of health this means that there was insufficient evidence to suggest that the chance of labour force participation for those in this health state was statistically different from those in the “best” health state. For the health shocks this means there is insufficient evidence to suggest that a negative health shock into this health state would significantly affect the chance of labour force participation. For those diseases or health states/shocks that are significant, the asterisks indicate the level of significance of the marginal effect. The categories that are not significant are **excluded** where totals are calculated. This is justifiable for the purpose of estimating the potential change in labour force participation that can be associated with the movement of those in these categories to better health, or the prevention of a health shock, as the models found insufficient evidence that there would be one.

As discussed in Section 6, the preferred model is the fixed effects model (for which results are assumed to be unbiased). Results from the standard logistic regression models may be biased owing to possible endogeneity. While the correlated random effects model attempts to account for some types of endogeneity, the results of this model may also be subject to bias. Despite this, impact figures are presented from all of these models to allow comparison of the model results and because the fixed effects model does not allow an estimate of the relationship between a constant health level and labour force participation.

Looking at the grouped chronic disease indicator from the pooled regression model indicates that if this group no longer had chronic diseases an additional 42,200 people may participate. This represents a 2.3% increase in the total number of people participating.

Moving on to consider individual chronic diseases, the table shows that the largest increase in the number of additional participants is for females with psychiatric conditions. This is despite the fact that the odds ratios and marginal effects are estimated to be of greater magnitude for stroke, heart disease, diabetes and males with psychiatric conditions. This illustrates the importance of the size of the group of interest when relating the results to the population as a whole. If no females suffered from psychiatric conditions it is estimated that an additional 9,500 people may participate; which represents a 0.5% increase in the total number of people participating. It should be remembered that as the disease groups are not independent (ie, a person may have diabetes as well as heart disease) the number impacted cannot be summed across all diseases.

The results of the pooled logistic regression for self-rated health in Table 17 illustrate the estimated additional number of people who would participate if they had excellent health, as opposed to the health state listed. So if all those people with good health had excellent health an additional 26,400 people may participate; which represents a 1.4% increase in the total number of people participating. Again, this illustrates that, while the marginal effects and odds ratios are higher for those in fair or poor health, the biggest potential increase in participation comes from those in good health. Overall, an additional 66,800 people may participate if they had excellent health; a 3.6% increase in the total number of people participating.

As explained previously, there may be unobserved variables that impact labour force participation and/or health. The logistic regressions do not account for this. Despite this, the estimates from the pooled models give an indication of the possible impact of health on participation. To try to control for unobserved time-constant variables, panel models were used. Their interpretation is slightly different from the pooled models. The results for the fixed effects model for self-rated health in Table 17 illustrate the additional number of people who may participate in the absence of negative health shocks.³⁴ That is, if during an annual period there were no negative health shocks, an additional 12,700 people may participate; which represents a 0.7% increase in the total number of people participating. While the coefficients and odds ratio from this model reported earlier were those for a health shock from excellent to a lower health state, other health shocks are possible and these health shocks are accounted for in these figures. For example, if there were no health shocks into poor health (from any of the higher health states) then an additional 5,200 people may participate.

³⁴ It should be remembered that the fixed effects model only considers within, rather than within and between, person variation. Despite this, in order to estimate the impact at the population level, the results of the model are assumed to be the same as for the population as a whole.

Table 17 – Estimates of annual impact from each binomial model, count (increase in number currently participating) and % (of current number of working age non-students participating in the labour force)

Health status	Count		%	
	Point estimate	95% CI (lower; upper)	Point estimate	95% CI (lower; upper)
Grouped chronic diseases – pooled regression				
Any chronic disease	42,200***	(32,200; 52,200)	2.30	(1.75; 2.84)
Individual chronic diseases – pooled regression				
Asthma	3,700	(-800; 8,300)	0.20	(-0.04; 0.45)
High blood pressure	5,800***	(1,300; 10,300)	0.32	(0.07; 0.56)
High cholesterol	2,400	(-1,800; 6,600)	0.13	(-0.10; 0.36)
Heart disease	5,300***	(3,000; 7,600)	0.29	(0.16; 0.41)
Diabetes	4,400***	(2,200; 6,600)	0.24	(0.12; 0.36)
Stroke	2,700***	(1,300; 4,000)	0.15	(0.07; 0.22)
Migraine	1,300	(-2,500; 5,100)	0.07	(-0.14; 0.28)
Psychiatric conditions – male	8,900***	(6,600; 11,600)	0.49	(0.36; 0.63)
Psychiatric conditions – female	9,500***	(2,500; 18,100)	0.52	(0.14; 0.99)
Cancer	500	(-1,500; 2,600)	0.03	(0.08; 0.14)
Self-rated health – pooled regression				
Very good health	4,700	(-1,900; 11,300)	0.26	(-0.10; 0.62)
Good health	26,400***	(21,000; 31,800)	1.44	(1.14; 1.73)
Fair health	24,900***	(21,200; 28,600)	1.36	(1.16; 1.56)
Poor health	15,500***	(13,500; 17,400)	0.84	(0.74; 0.95)
Total (exc. insignificant)	66,800	(55,800; 77,800)	3.64	(3.04;4.24)
Self-rated health – fixed effects				
Very good health shock	-1,500	(-10,800; 7,800)	-0.08	(-0.59; 0.43)
Good health shock	4,500	(-7,100;16,200)	0.25	(0.03; 0.47)
Fair health shock	7,600***	(2,500;12,600)	0.41	(0.31;0.52)
Poor health shock	5,200***	(2,700;7,600)	0.28	(0.22; 0.34)
Total (exc. insignificant)	12,700	(5,300; 20,200)	0.69	(0.29; 1.10)
Self-rated health – random effects				
Very good health shock	0	(-1,300;1,300)	0.00	(-0.07;0.07)
Good health shock	600	(-700; 1,800)	0.03	(-0.04; 0.10)
Fair health shock	1,100***	(300; 1,800)	0.06	(0.02; 0.10)
Poor health shock	900***	(300, 1,500)	0.05	(0.02; 0.08)
Average time in very good health	-4,500	(-2,500; 11,400)	-0.25	(-0.14; 0.62)
Average time in good health	25,300***	(20,900; 29,600)	1.38	(1.14; 1.61)
Average time in fair health	13,900***	(12,000; 15,700)	0.76	(0.65; 0.86)
Average time in poor health	6,000***	(5,200; 6,900)	0.33	(0.28; 0.38)
Total (exc. insignificant)	47,100	(38,700; 55,500)	2.57	(2.11; 3.02)

Source: SoFIE Waves 1-3 Version 4, Statistics New Zealand

Notes:

1. These estimates are calculated using the marginal effects in Tables 3 and 13 from unweighted models and the weighted count participation estimates. Standard longitudinal weights are used other than for cancer where the adjusted weights are used. Data is for 2002/05 period but estimate of impact is for the annual average over this period.
2. Groups may not sum to totals owing to rounding.
3. The totals include only significant estimates.
4. *Significant at the 90% level. **Significant at the 95% level. ***Significant at the 99% level.
5. For the pooled regression the impact is the number of additional people participating if all participants had excellent health/no chronic disease(s). For the fixed and random effects models the impact is the number of additional people participating if there were no negative health shocks in a year (this is from any higher health state into the health state mentioned). The final marginal effects listed for the random effects model illustrate the impact of having excellent health in all waves rather than a proportion of time in the lower self-rated health state listed.

The impact figures for the fixed effects models are much lower than those estimated from the pooled models. This is owing to the differences in what is being estimated; that is the relationship between labour market participation and health shocks rather than the level of health. Using the fixed effects model it is not possible to estimate the impact of the current level of health; it is just the impact for those whose health deteriorates that can be estimated.³⁵ Given that in an annual period not everyone experiences a health shock, the number impacted is smaller.

The estimates for the health shocks for the random effects model are calculated in the same way as for the fixed effects model; that is, the numbers represent the increase in the number of people participating if they had not had a negative health shock to the state listed. However, this model also takes into account the average level of health. These estimates illustrate the additional number of people who may participate if their average level of health had been excellent in the last three years rather than being at the stated level for some period in recent years. That is, if those who had spent some time in good health in recent years had instead been in excellent health, an additional 25,300 people may participate. These results illustrate that the level of health rather than health shocks, is much more influential in the relationship with labour force participation. In total the results of the random effects model for self-rated health indicate that if there were no negative health shocks and the average level of health in previous periods had been excellent then an additional 47,100 people may participate.³⁶

Tests of the models indicated that the fixed effects model was preferred in a statistical sense. This was because the logistic models did not account for time-constant unobserved variables, and as such the results are likely to be biased, and the correlated random effects model indicated that, even after including the average health level over the period, there still may be correlation between the unobserved variables and health, again meaning the results may be biased. However, it is important to remember that while the fixed effects model appeared to be the best model, the lack of inclusion of an estimate for the impact of health level is a large drawback. From a theoretical perspective it seems sensible that the level of health will be related to labour force participation. The fixed effects model indicates that a health shock in a specific period is significantly negatively related to participation. In the same period some people will be continuously inactive owing to poor health. It is not possible to estimate the significance or impact of this from the fixed effects models. Some of this group will have experienced a health shock at some point but, as they did not have their health shock in the period of consideration, this cannot be accounted for. While the results of the other models may be biased, they, along with the results of other international research, suggest the impact of the level of health may be non-zero. Given this, the lower confidence interval for the estimated impact figures from the fixed effects model could be seen as a lower bound for the

³⁵ The model takes into account changes in health but only negative health changes are considered here.

³⁶ It should be noted that these estimates do not take into account the relationship between health shocks and average health. Average health is thought of as average health in the period before the health shock period.

estimate of the true impact of overall health (shocks and level) on labour force participation. It is known that the pooled regression results are likely to be biased as they do not account for unobserved variables that may explain variations in health. The impact estimates for this model are therefore likely to be too high. While the estimates from the random effects model may still be biased they provide an intermediate model, which is an improvement on the pooled model but not the fixed effects model. Owing to the potential bias, and owing to the fact that the relationship between the average health state and health shocks for the same individual are not accounted for in these impact estimates, it seems sensible to take the lower confidence interval of the random effects model as the upper bound for the impact of health on participation.

The point estimates from these models indicate that if there was an improvement in health (ie, no negative health shocks and everyone had excellent average health) an additional 12,700 to 47,100 people may participate; that represents a 0.7% to 2.6% increase in the total number of people participating. Based on the discussion above it is more sensible to assume that, if there was an improvement in health, the additional number of people who may participate is likely to be between 5,300 and 38,700; that is, a 0.3% to 2.1% increase in the total number of people participating.

It is important to remember that all of these impact figures are likely to be an underestimate of the impact of health on labour force participation for the population of New Zealand as a whole. One reason for this is that the SoFIE population is healthier than the population it aims to represent owing to those of poorer health being less likely to respond over time (see Section 3.2 for further explanation). Another reason is that the estimates are for those of working age only. They therefore do not account for the fact that improvements in health may result in those over working age participating in the labour force for longer. Further, reduced labour force participation is unlikely to be the only factor related to poorer health. Poor health will also result in lost output owing to people being away from work ill (absenteeism) and owing to lower productivity when at work (presenteeism). Health also impacts on educational development and skill usage. These “costs” are not considered here.

Table 18 provides the same estimates of impact as Table 17 but for the multinomial models. These estimates are based on the marginal effects from Tables 6 and 10 along with the number estimated to be in these groups. The results illustrate where the increase in full-time working comes from. In the main, the increase is a result of a decrease in inactive people, but there is often a reduction in the number who work part-time or who are unemployed. As an example, consider the chronic disease indicator. If the group with one or more chronic diseases no longer had these diseases it is estimated that an additional 57,600 people may work full-time. The majority (47,700) of these people move from being inactive to working full-time. However, 5,500 of these people move from part-time employment to working full-time and 4,400 move from being unemployed. It is estimated that, on average over the three waves of SoFIE, 1,436,800 people of working age worked full-time; 348,700 worked part-time; 49,500 were unemployed; and 385,400 were inactive. The increasing number of people who may work full-time in the absence of chronic disease therefore represents a 4% increase in the number of people who work full-time, with falls of 1.6% in the number of people working part-time, 8.9% in the number who are unemployed and 12.4% in the number who are inactive.

**Table 18 – Estimates of annual impact from each multinomial model, count
(increase in number in each labour market outcome)**

Disease	Full-time employment	Part-time employment	Unemployment	Inactive
Grouped chronic diseases – pooled regression				
Any chronic disease	57,600***	-5,500	-4,400**	-47,700***
Individual chronic diseases – pooled regression				
Asthma	1,600	2,500	-400	-3,700
High blood pressure	4,000	2,000	300	-6,300**
High cholesterol	2,900	300	-600	-2,700
Heart disease	3,900**	1,100	400	-5,300***
Diabetes	7,900***	-1,700	-900*	-5,300***
Stroke	4,300***	-1,200*	0	-3,100***
Migraine	6,000*	-2,100	-2,100**	-1,800
Psychiatric conditions – male	12,500***	-2,000*	-1,000	-9,600***
Psychiatric conditions – female	14,400***	-2,500	-1,200	-10,700***
Cancer	200	200	200	-500
Self-rated health – pooled regression				
Very good health	10,600*	-3,800	-1,500	-5,300
Good health	38,700***	-3,700	-4,900***	-30,100***
Fair health	34,500***	-4,800***	-1,700***	-28,000***
Poor health	15,300***	1,200**	-400	-16,200***
Total	99,100	-11,000	-8,400	-79,600

Source: SoFIE Waves 1-3 Version 4, unweighted, Statistics New Zealand

Notes:

1. These estimates are based on the marginal effects from Tables 6 and 9 from unweighted models and the weighted count participation estimates. Standard longitudinal weights are used other than for cancer where the adjusted weights are used. Data is for 2002/05 period but estimate of impact is for the annual average over this period.
2. Asterixes indicate the impact is significantly different from zero when all other variables are evaluated at the mean for the sample.
*Significant at the 90% level. **Significant at the 95% level. ***Significant at the 99% level.
3. Counts may not sum to totals and rows may not sum to zero owing to rounding.
4. The totals include only significant estimates.

8.1.2 Concluding remarks

In drawing conclusions from these results it should again be remembered that it was not possible to identify whether health had impacted on labour market status, or vice versa. Further, it is known that there are already interventions to help people with some of these conditions and the efficacy of further intervention may be limited. Nevertheless, there are a number of tentative conclusions that can be drawn from these results.

Firstly, in considering whether tackling chronic diseases would increase participation in the labour market, it is psychiatric illnesses where there is potential to have the greatest impact. As shown in Table 17, an additional 18,400 people may participate in the labour market in the absence of psychiatric conditions, which represents a 1% increase in the total number of people participating. Considering the point estimates, this is over three times more than the potential increase from any of the other conditions considered. Again, it must be remembered that these results are based on basic models which do not

determine causality and do not control for unobserved factors that may explain some of the variation in labour force participation that is attributed to these conditions.

Secondly, by far the greatest impact on numbers in the labour market appears to come from people being in good or fair health rather than excellent health. In other words, interventions for basically healthy people may have a greater impact on labour market participation, than attempts to help those with the poorest health. However, in terms of health shocks, the only significant potential impacts are in the absence of health shocks from excellent health into fair or poor health.

Finally, in the absence of ill health, by far the greatest change in status is from not working to working full-time. While, in the main, the results indicate some movement from working part-time to working full-time with improvements in health, this movement is often not found to be significant. This is a particularly striking result when the change is from good or very good to excellent health, and suggests that part-time work may not be a common substitution for full-time work for those in these health categories.

Another interesting finding briefly noted in the report is around health related benefits. While, as would be expected, the proportion of those receiving a health related benefit increases with decreasing self-rated health, around seven-tenths of those receiving a health related benefit consider themselves to be in excellent, very good or good health. Owing to the survey question (as discussed in Section 4.2.2) and the differences between health and disability (for example, a person who is blind may be eligible for a benefit but consider themselves to be in excellent health), it is perfectly plausible for a person to be eligible for a health related benefit and to self-rate their health to be good or above. However, this does highlight an area where further work could be undertaken to better understand the reasons for this result.

References

- Biddulph, F., Biddulph, J. and Biddulph, C. et al (2003), *The complexity of community and family influences on children's achievement in New Zealand: Best evidence synthesis*. (Wellington: Ministry of Education).
- Bound, J., Schoenbaum, M., Stinebrickner, T. R. and Waidmann, T. (1999), "The dynamic effects of health on the labour force of older workers", *Labour Economics*, No 6: 179–202.
- Cai, L., (2007), *The relationship between health and labour force participation: Evidence from a panel data simultaneous equation model*, Melbourne Institute Working Paper Series Working Paper No. 1/07, February.
- Cai, L. and Kalb, G., 2006, *Health status and labour force participation: Evidence from Australia*, Health Economics, Vol 15: 241–261.
- Crichton, S., Stillman, S. and Hyslop, D. (2007), *Returning to work from injury: Longitudinal evidence on employment and earnings*, Statistics New Zealand, December.
- Currie, J. and Madrian, B. C. (1999), *Health, health insurance and the labour market*, In *Handbook of labour economics*, vol. 3, Ashenfelter O., Card D. (eds) Elsevier Science BV: Amsterdam, 1999: 3310–3415.
- DeVol, R. and Bedroussian, A. (2007), *An unhealthy America: The economic burden of chronic disease*, Milken Institute, October.
- Disney, R., Emmerson, C. and Wakefield, M. (2003), *Ill health and retirement in Britain: A panel data based analysis*, IFS.
- Freese, J. and Scott Long, J. (2006), *Regression models for categorical dependent variables using stata: Second edition*. Strata Press.
- Jensen, J., Sathyandra, S., Rochford, M., Jones, Davina, Krishnan, V. and McLeod, K. (2005), *Disability and work participation in New Zealand: Outcomes relating to paid employment and benefit receipt*, Ministry of Social Development, June.
- Laplagne, P., Glover, M. and Shomos, A. (2007), *The effects of health and education on labour force participation*. Australian Productivity Commission.
- OECD (2003), *Transforming disability into ability: policies to promote work and income security for disabled people*. Campus Verlag.
- Stata (2007), *Stata statistical software: Release 10*, StatCorp LP, College Station, Texas.
- Stern, S. (1989), *Measuring the effect of disability on labour force participation*, The Journal of Human Resources, Vol 24, No. 3: 361-395.
- Tabachnick, B.G. and Fidell L. S. (2001), *Using multivariate statistics*, 4th Edition.
- Wooldridge, J. M. (2006), *Introductory econometrics: A modern approach*, 3rd edition.

Bibliography

- Cai, L. and Kalb, G. (2004), *Health status and labour force participation: Evidence from the HILDA data*, Melbourne Institute Working Paper No. 4/04, March.
- Carter, K., Hayward, M. and Richardson, K. (2008), *SoFIE-Health baseline report*. University of Otago, Wellington.
- Carter, K., Hayward, M. and Richardson, K. (2008), *SoFIE-Health data and processing systems*. University of Otago, Wellington.
- Davis, K. et al (2005), *Health and productivity among U.S. workers*, The Commonwealth Fund, August.
- Department of Labour (2006), *45 Plus: Choices in the labour market*, November.
- Hsiao, C. (2003), *Analysis of panel data*: Second edition.
- MBF Foundation (2007), *The high price of pain: The economic impact of persistent pain in Australia*, November.
- Richardson K., Carter K., and Hayward M. (2008) *SoFIE-Health data and processing systems. SoFIE-Health report 1*. University of Otago, Wellington.
- Statistics New Zealand (2005), *Survey of family, income and expenditure estimation specifications*, November. Statistics New Zealand.
- Statistics New Zealand (2005), *Survey of family, income and expenditure imputation specifications*, November. Statistics New Zealand.
- Statistics New Zealand (2005). *Survey of family, income and expenditure questionnaires*. Statistics New Zealand.
- Statistics New Zealand (2008), *Data laboratory output guide*, March. Access Economics, Sydney.
- The Treasury (2008), *Social mobility: Annexes for Treasury report*, June. Treasury
- Wooldridge, J. M. (2002), *Econometric analysis of cross section and panel data*. MIT Press

Appendix A

Appendix Table A1 – Definitions of SoFIE variables in participation models

Variable name	Variable categories	Notes
Labour market participation	<ul style="list-style-type: none"> • Participating • Not participating (inactive) 	Labour force participation at the household interview date.
Labour market outcome	<ul style="list-style-type: none"> • Full-time employment • Part-time employment • Unemployed • Inactive 	Labour market activity at the household interview date. Hours are the average weekly hours a respondent worked whilst employed in the annual reference period. Full-time hours are 30 hours or more. Unemployed is not employed but actively looking for work. Inactive is not employed and not looking for work.
Gender	<ul style="list-style-type: none"> • Male • Female 	-
Region of residence	<ul style="list-style-type: none"> • Auckland • Waikato • Wellington • Rest of North Island • Canterbury • Rest of South Island 	-
Born in New Zealand	<ul style="list-style-type: none"> • Yes • No 	-
Ethnicity	<ul style="list-style-type: none"> • 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	<ul style="list-style-type: none"> • Respondent 50 or over • Respondent under 50 	Age is at the interview date.
Highest qualification	<ul style="list-style-type: none"> • 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.
Chronic disease indicator (grouped chronic disease)	<ul style="list-style-type: none"> • No chronic disease • One or more chronic diseases 	An indicator of chronic disease presence. Those for whom cancer is unknown who do not report any other chronic disease are assumed to not have a chronic disease.
For each chronic disease	<ul style="list-style-type: none"> • No chronic disease • Chronic disease 	The eight chronic diseases covered in SoFIE are asthma, high blood pressure, high cholesterol, heart disease, diabetes, stroke, migraine and psychiatric conditions (depression, manic depression or schizophrenia). The question on

Variable name	Variable categories	Notes
		presence of diseases is only asked in Wave 3. Other than for psychiatric conditions and for a small number of cases where data was missing, disease presence in earlier waves was derived using the presence of disease and the age at diagnosis. As the question was on the age of diagnosis rather than the year, the variables created are not exact. The age at diagnosis for psychiatric conditions is unknown. As disease diagnosis in the survey period is likely to be small, after preliminary analysis of this group it was decided to assume that those with psychiatric conditions in Wave 3 have psychiatric conditions in Waves 1 and 2.
For each chronic disease (excluding psychiatric conditions)	<ul style="list-style-type: none"> • No chronic disease • Chronic disease diagnosed 0 to <5 years ago • Chronic disease diagnosed 5 or more years ago 	Derived using chronic disease diagnosis, the age at the household interview date and, where present, the age of disease diagnosis. This is a proxy for the number of years since diagnosis, as we only know the age at diagnosis not the actual date. The age of diagnosis is not asked for psychiatric conditions (depression, schizophrenia, manic depression) so this variable is not available for this disease.
Self-rated health	<ul style="list-style-type: none"> • Excellent • Very good • Good • Fair • Poor 	-
Studying	<ul style="list-style-type: none"> • 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. If a respondent was 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 and excluded from the analysis.
Partner	<ul style="list-style-type: none"> • Working partner • Non-working partner • Single 	-
Children	<ul style="list-style-type: none"> • No dependent children • Child(ren) minimum age <5 • Child(ren) minimum age 5 to 17 	A dependent child is one who is under 18 years and not in full-time employment.

Variable name	Variable categories	Notes
Benefit	<ul style="list-style-type: none"> • No government benefits in reference period • Government benefits received in reference period 	These include ACC, student allowance payment, IRD payment, Veteran Pension Fund and WINZ benefit payment . It also includes the small number of respondents under 65 who receive NZ Superannuation payments.
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.
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 A2 – Non-SoFIE variables in participation models

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.
Cancer registration data	Cancer	<ul style="list-style-type: none"> • No cancer • Cancer • Cancer unknown 	This variable indicates a cancer registration prior to the interview date; determined by the age at the registration compared with the age at the interview date. Cancer information is unknown for non-consenters and non-matched consenters.
Cancer registration data	Cancer by age diagnosed	<ul style="list-style-type: none"> • No cancer • Cancer diagnosed 0 to <5 years ago • Cancer diagnosed 5 or more years ago • Cancer unknown 	Derived using the age at diagnosis from the cancer registration data and the age at the household interview date from SoFIE.

Appendix Table A3 – Definitions of additional SoFIE variables in adjusted health model

Variable name	Variable categories	Notes
Total household income	-	Continuous variable which is the log of the consumer price adjusted personal income. There was a small number of respondents with negative 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. Household income was not adjusted to reflect family size/composition.
Health benefit	<ul style="list-style-type: none"> • No health related government benefit in reference period • Health-related government benefit in reference period 	This includes any ACC payments, sickness benefit, incapacity benefit and disability benefit.
Smoked	<ul style="list-style-type: none"> • Never smoked • Current or past smoker 	Estimated from whether a respondent currently smokes and, if not, whether they ever have.
Tenure	<ul style="list-style-type: none"> • Not owned • Owned with mortgage • Owned outright 	Derived from variable indicating ownership status of home.

Appendix B

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.^{37 38}

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. SoFIE interviews 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 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 back to 1990. For those respondents who agreed to the data linkage, and were successfully matched, it was possible to identify those respondents who are listed on the Cancer Register as having been diagnosed with cancer.³⁹ As the linked information only goes back to 1990 this is only a measure of recent cancer diagnosis. Where descriptive (prevalence) statistics are presented where only the linked sample is used, adjusted weights were used to realign the sample with the population (adjusted longitudinal weight) as opposed to the weights provided by Statistics New Zealand (standard longitudinal weights).⁴⁰

³⁷ 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 Māori population density and other socio-economic variables derived from the most recent census.

³⁸ The full SoFIE questionnaire can be found here:

<http://www2.stats.govt.nz/domino/external/quest/sddquest.nsf/12df43879eb9b25e4c256809001ee0fe/14d945bb95ab2bbbcc256fb70077b3bb?OpenDocument>.

³⁹ Around 80% of all SoFIE respondents agreed for their data to be linked. Of these, 97% were linked successfully.

⁴⁰ More information on the adjusted weight is available from the author.

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, 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 Waves 1, 2 or 3.^{41 42}

Figure B1 – 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

⁴¹ 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.

⁴² 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).

standard weights provided by Statistics New Zealand and adjusted 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.⁴³ 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 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 somewhat 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.

⁴³ More information on sample attrition and consent in SoFIE and the adjusted weights is available from the author.

Appendix C

Methods

Pooled logistic regressions

Initially, binomial logistic regression models were fitted to the data to quantify the relationship between the presence of different chronic diseases and labour force participation and between self-rated health and labour force participation, while holding all other variables constant. In the standard pooled regression models, responses in each wave were pooled together to form one large sample. Therefore each respondent had up to three responses in the sample. The fact that observations from the same person in different waves were not independent of each other, and therefore the error terms in the model were likely to be correlated, was accounted for by treating people as clusters.

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 are not participating (the latter was the reference category when a binomial logistic regression was carried out). The form of the equation can be seen in Figure C1. The unemployment rate at the time of the interview was included to reflect the possible differences in participation owing to the economic climate at the interview date. Maximum likelihood estimation was used to estimate the regression coefficients.⁴⁴

A multinomial logistic regression was then fitted to the data to quantify the impact of the presence of diseases on the chance of being in one of the four labour market outcomes while holding all other variables constant. This aimed to determine if the impact of the presence of each disease was consistent across each labour market outcome. As there are more than two response categories in the dependent variable there is now more than one logistic regression model. Each model is the same as that in Figure C1 with the L_i indicator replaced with indicators for full-time, part-time and unemployed (L_{FTi} , L_{PTi} and L_{Ui} respectively), with the reference category being those who are inactive. The formula for the probability of success in each case is similar to that for the binomial logistic regression but with the denominator being the sum of the odds of success across each of the three response categories (excluding the reference category).

The main limitation of standard binomial and multinomial logistic regressions is that they do not allow for endogeneity. In other words they assume that the explanatory variables are exogenous; that is, their values are not affected by labour force participation or by other unobserved characteristics. However, this assumption may not be strictly true for any generic health measure (H_i) and the failure to account for endogeneity means that any significant relationships that are established are associations and do not imply causality; for instance, the fact that the model may prove a relationship between the dependent and predictor variables does not mean that the predictor variables caused the outcome (Tabachnick and Fidell, 2001)

⁴⁴ Fitting models separately for each gender was considered. However, for all chronic diseases other than psychiatric conditions the relationship between chronic disease and participation was in the same direction and of the same magnitude irrespective of gender. Further, for each disease the confidence intervals for the coefficients overlapped for male and female. For this reason, and owing to the relatively small numbers with certain diseases such as cancer, it was decided to fit the model for combined genders with interactions included for parameter estimates that appeared to differ by gender. These were psychiatric conditions, social marital status and the presence of children. This approach was continued when considering self-rated health to aid comparability.

Figure C1 – Form of binomial logistic regression model

$$L_i = 1(\mathbf{CD}_i'\boldsymbol{\beta} + \mathbf{X}_i'\boldsymbol{\lambda} + u_i > 0) \quad i = 1, \dots, n$$

where:

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

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

$\boldsymbol{\beta}, \boldsymbol{\lambda}$ = a vector of regression coefficients

\mathbf{CD}_i = a vector of chronic disease indicators

\mathbf{X}_i = a vector of explanatory variables

u_i = error term associated with person i

$e^{\mathbf{CD}_i'\boldsymbol{\beta} + \mathbf{X}_i'\boldsymbol{\lambda}}$ = odds of success

$$P(L = 1 | \mathbf{X}, \mathbf{CD}) = \frac{e^{\mathbf{CD}'\boldsymbol{\beta} + \mathbf{X}'\boldsymbol{\lambda}}}{1 + e^{\mathbf{CD}'\boldsymbol{\beta} + \mathbf{X}'\boldsymbol{\lambda}}}$$

Note: The relationship between the responses for each person in the different waves (ie, time = 1, 2 or 3) is accounted for by identifying people as clusters.

Fixed and random effects panel logistic regression⁴⁵

While there were a number of control variables included in the standard pooled regressions, there may be some important individual characteristics that were not observed. The unobserved variables may significantly influence participation; they may influence (or be correlated) with ill health; or they may influence both of these. When the omitted variables are correlated with health, the estimates of the relationship between health and participation from the pooled regression model will be biased because the error term in the model will be correlated with the health variable (that is, health is endogenous, not exogenous, therefore violating an assumption of the logistic regression analysis).

One advantage of SoFIE is its panel aspect; that is, there are up to three observations per person. This opens up the prospect of fixed or random effects panel models to allow for time-constant unobserved heterogeneity. A fixed effects model exploits the panel nature of the data to determine how health shocks (changes in health) over time relate to changes in labour force participation allowing for time-invariant omitted variables that may be correlated with the explanatory variables (ie, the endogenous health). The fixed effects model is derived from the starting equation in Figure C2. The error term from the standard pooled regression model u_i now has a time dimension and is made up of two components. These are α_i , the time-constant unobserved variables for the i th person which may or may not be correlated with H_{it} , and the error term ε_{it} , which includes the true error and any unobserved variables that are time-varying. It is assumed that the time-variant unobserved variables are not correlated with the explanatory variables so that

⁴⁵ This section draws heavily on unpublished lecture notes by Dean Hyslop.

the error term, ε_{it} , is not correlated with L_{it} or H_{it} . Conditional logistic analysis differs from regular logistic regression in that data are grouped (with those who exhibit no changes in the outcome variable over the periods considered dropped) and the likelihood is calculated relative to each other group; that is, a conditional likelihood is used. The conditional likelihoods do not involve α_i , so they do not need to be estimated (Stata, 2007). The model compares changes in the covariates with a change in the dependent variable. The coefficients indicate the relationship between a change in that covariate and the chance of participating. One drawback of the fixed effects model is that it removes all explanatory variables from the model which are time-invariant; for example, gender.⁴⁶ It also drops all respondents for whom the dependent variable (labour force participation) did not change over time. This significantly reduced the sample available for analysis.

Figure C2 – Initial form of the fixed and standard random effects logistic panel model

$$L_{it} = 1(H'_{it}\beta + X'_{it}\lambda + \alpha_i + \varepsilon_{it} > 0) \quad i = 1, \dots, n; \quad t = 1, 2, 3$$

where:

L_{it} = a binary response variable for participation for the i th person at time t

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

β, λ = a vector of regression coefficients

H_{it} = a vector of variables to indicate self-rated health

X_{it} = a vector of explanatory variables

α_i = unobserved time-invariant variables

ε_{it} = idiosyncratic error representing unobserved factors that change over time and affect L_{it} (Note: $\alpha_i + \varepsilon_{it} = u_{it}$)

Fixed effects model: $\text{Cov}(H_{it}, \alpha_i) \neq 0 \quad i = 1, \dots, n; t = 1, 2, 3$

Random effects model: $\text{Cov}(H_{it}, \alpha_i) = 0 \quad i = 1, \dots, n; t = 1, 2, 3$

An alternative way to control for unobserved time-invariant variables is using a random effects model. The starting form of this model is the same as that presented in Figure C2, however, this time the assumption is that while the unobserved variables influence the dependent variable (labour force participation) they are not correlated with health. This means that the coefficient estimates from the standard pooled regression will not suffer from omitted variable bias, but that the error terms in the model will be serially correlated. The random effects model subtracts a fraction of that time averaged value, where the fraction depends on the variation of the unobserved variables, the variation of the idiosyncratic error and the number of time periods (for more explanation, see Wooldridge, 2006). The advantage of the method is that it includes explanatory and dependent variables that are constant over time. This means that the sample size available for analysis is not reduced as with the fixed effects model and that estimates of the effect of

⁴⁶ Further, it is considered best practice to remove from the model specification all variables that may change over time, but are more or less fixed in reality.

time constant variables are provided. However, the assumption that the omitted variables are not correlated with health is a disadvantage given that the unobserved variables that are correlated with health are of concern. One way to use the random effects model where some of the unobserved time constant variables are thought to be correlated with health is to make an assumption about the relationship between health and the unobserved time-invariant variables. This is the correlated random effects model. More specifically, as shown in Figure C3 it can be assumed that the expected value of the unobserved variables is equal to a linear function of the average time spent in each health state over the three waves together with a random term representing the unobserved time-invariant coefficients that are not correlated with health. Substituting this expected value into the starting equation for the fixed effects model results in the remaining unobserved time-variant coefficients being uncorrelated with health. A random effects model can therefore be used.

Figure C3 – Equations used in the correlated random effects logistic regression panel model

From Figure C2 the starting form of the fixed effects equation is:

$$L_{it} = 1(\mathbf{H}'_{it}\boldsymbol{\beta} + \mathbf{X}'_{it}\boldsymbol{\lambda} + \alpha_i + \varepsilon_{it} > 0) \quad (1)$$

Where:

i = person = 1, ..., n

t = time = 1, 2, 3

It is assumed that:

$$\begin{aligned} E(\alpha_i | \mathbf{H}_{it}) &= \sum_{j=2}^5 \sum_{t=1}^3 \frac{H_{jit}}{3} \gamma_{jt} + \eta_i = \sum_{j=2}^5 \bar{H}_{ji} \gamma_j + \eta_i \\ &= \bar{H}_{2i} \gamma_2 + \bar{H}_{3i} \gamma_3 + \bar{H}_{4i} \gamma_4 + \bar{H}_{5i} \gamma_5 + \eta_i \end{aligned} \quad (2)$$

where:

j = health state = 1(excellent), ..., 5(poor)

\mathbf{H}_{it} = a vector of variables to indicate self-rated health

For each health state j : \bar{H}_{ji} = Proportion of time in the health state

For each person i : $\sum_{j=1}^5 \bar{H}_{ji} = 1$

η_i = unobserved time-invariant variables

and $\text{Cov}(\mathbf{H}_{it}, \eta_i) = 0$

Combining equations (1) and (2) gives the standard form of the random effects model:

$$L_{it} = 1(\mathbf{H}'_{it}\boldsymbol{\beta} + \mathbf{X}'_{it}\boldsymbol{\lambda} + \bar{H}_{2i} \gamma_2 + \bar{H}_{3i} \gamma_3 + \bar{H}_{4i} \gamma_4 + \bar{H}_{5i} \gamma_5 + \eta_i + \varepsilon_{it} > 0) \quad (3)$$

Results for both the fixed and correlated random effects models are presented in this paper. While the fixed and correlated random effect panel model goes further than the standard pooled regression, there are drawbacks. Firstly, the model only accounts for omitted variables that are time-constant, so any time-variant unobserved effects are in the error term. The assumption is that these time-varying omitted variables are uncorrelated with participation or with any of the explanatory variables. Secondly, while using fixed or

correlated random effects models to look at how health changes are related to participation changes within respondents does control for the subjective nature of the self-rated health question (in the sense that some people will consistently be more optimistic in their health rating and some consistently more pessimistic) these models do not control for the other health measurement issues with self-rated health outlined in Section 4.2.2. Thirdly, these models do not allow the feedback effect to be estimated. Finally, an issue with the fixed effects model is that it only looks at how changes in health relate to changes in participation. It does not include estimates of the effect of poor health which possibly prevents a person working in the first place. This average health effect for the three waves is picked up in part in the correlated fixed effects model. However, if the assumption for the random effects model, that the expectation of the correlated unobserved time-invariant variables is a linear function of the average time in a health state, is incorrect, this model will be flawed.

Standard pooled, fixed and correlated random effects logistic regression with adjusted health measure

While self-reported health may be a more encompassing measure of current health than considering previous diagnosis of individual chronic diseases from which the respondent may no longer suffer symptoms, it is also a more subjective measure and open to bias. Despite the possibility that self-rated health may not completely reflect true health it is still widely used where no alternative measures exist.

Of the problems with self-reported health reported in Section 4.2.2 the one of main concern here is rationalisation bias. This is where individuals who are inactive may report worse-than-actual health to justify their inactive labour market state. Disney *et al* (2003) point out that this may be for self-esteem if nothing else. This bias, if it exists, will cause self-reported health to be correlated with the error term in the labour market participation regression models if unadjusted self-rated health is used as an explanatory variable and result in the relationship between health and participation being overestimated.

One approach to attempt to remove this problem (suggested by Bound *et al*, 1999 and used by Disney *et al*, 2003) is to construct an adjusted health measure using personal characteristics and more objective health measures. The relationships between true health and measured self-rated health are shown in Figure C4. This method aims to purge self-rated health of its rationalisation bias and better reflect true health. The adjusted health variable, which is a standardised index derived from equation 3 in Figure C4, is then included in a second model to assess the impact of adjusted health on participation. Using this adjusted health means that, unlike when unadjusted health is used, the error term should no longer be correlated with labour market participation as the rationalisation bias is included in the error term of equation 3 in Figure C4.

Figure C4 – Relationship between true health and measured health in each wave

Assume that at time t a person's true health, H_i^* , can be modelled using the following equation:

$$H_i^* = 1(\mathbf{Z}_i' \boldsymbol{\beta} + \mathbf{Y}_i' \boldsymbol{\lambda} + \varepsilon_i > 0) \quad i = 1, \dots, n \quad (1)$$

where:

$\boldsymbol{\beta}, \boldsymbol{\lambda}$ = a vector of regression coefficients

\mathbf{Z}_i = a vector of objective health indicators

\mathbf{Y}_i = a vector of explanatory personal characteristics that may affect health some of which overlap with the explanatory variables in X_i in the participation equation

ε_i = error term associated with person i

$$\text{Corr}(\mathbf{Z}_i, \varepsilon_i) = 0 \text{ and } \text{Corr}(\mathbf{Y}_i, \varepsilon_i) = 0$$

However, health may be measured with error:

$$H_i = 1(H_i^* + v_i > 0) \quad i = 1, \dots, n \quad (2)$$

where v_i = reporting errors

If H_i is subject to rationalisation bias then including this in the participation equation will result in biased estimates as v_i will not be random and $\text{Corr}(v_i, L_i) \neq 0$

Assuming that $\text{Corr}(v_i, \varepsilon_i) = 0$

Combining equations (1) and (2) gives:

$$H_i = 1(\mathbf{Z}_i' \boldsymbol{\beta} + \mathbf{Y}_i' \boldsymbol{\lambda} + u_i > 0) \quad i = 1, \dots, n \quad (3)$$

Where:

$$u_i = v_i + \varepsilon_i$$

Using a standardised form of the predicted value of H_i from (3) to estimate an adjusted health measure should purge health of any rationalisation bias as this bias should be contained in the error term for the model.

To construct the adjusted health measure, each wave of SoFIE was taken in turn and all adult longitudinal respondents considered (ie, even if respondents are over 64 or full-time students). An ordered logit model was used to predict self-rated health using a vector of personal characteristics (some of which overlap with the personal characteristics used in the participation equation) and a vector of objective health measures. The objective health measures were the presence of various chronic diseases; whether the respondent has ever been a regular smoker; and whether the respondent received any health related benefits in the reference period (which, if the benefit system is effective, should be an indicator of the severity of health problems). The form of this model is similar to that

described in Figure C1 but with self-reported health as the dependent variable. There are now numerous outcomes for each of the five self-reported health states which are ordered, so an ordered logit model is used to predict health.

The probability of being in poor health was then predicted for each person using the model results. As in the IFS paper (Disney *et al*, 2003) these probabilities were then standardised in relation to the average health for that year to form the adjusted health measure (so the mean for each year for all longitudinal respondents was zero and the standard deviation one). This process is conducted independently for each year and results in a health measure for each person relative to that year's average. This adjusted health measure is then included in the standard pooled logit regression and in the fixed and correlated random effects models in place of self-rated health to determine the relationship between this adjusted measure and labour force participation.⁴⁷

This method is similar to an instrumental variable or two-stage approach; however, the aim of it is just to purge self-rated health of potential bias rather than using the instruments to account for the unobserved heterogeneity of health. One drawback of using this adjusted health measure in the second model rather than unadjusted health is that interpreting what a unit change in the adjusted health measure equates to in the real world is less intuitive than, say, a change from excellent to poor health when the self-rated health measure is used. However, using this method is worthwhile to see if any relationships between health and participation remain when an adjusted health measure is used.

Instrumental variables/two-stage approach and simultaneous equations

An alternative method of controlling for unobserved heterogeneity is to use an instrumental variable approach (also called two-stage regression). This approach enables both time-variant and time-invariant unobserved variables that are correlated with health to be controlled for by instrumenting the endogenous variable, health. In the first equation, health is regressed against all the exogenous variables in the participation equation along with the instrument(s).⁴⁸ The second stage uses the predicted values of health in the labour force equation model. While this approach seems attractive, given that it controls for time-variant and invariant unobserved variables it is very difficult to find suitable instruments. For a variable to be a valid instrument it should be correlated with health, but should not affect participation other than through health (ie, it should not belong in the labour force participation equation once health is included) (Wooldridge, 2006). When a valid instrument(s) is found, an equation is said to be identified. A literature review by Currie and Madrian *et al* (1999) concluded that relatively little research has been devoted to assessing the empirical importance of potential endogeneity bias; however, for those studies that attempt to deal with endogeneity of health using instrumental variables, it is difficult to find compelling sources of identification. The majority of the studies they reviewed relied on arbitrary exclusion restrictions and the resulting estimates were very sensitive to these identification assumptions.⁴⁹

An effort was made to find instruments for self-rated health. Possible candidates available in all three waves were whether a respondent has ever smoked (making an assumption that very few people will have started smoking in the survey period) and whether they had

⁴⁷ The correlated fixed model included an average health stock measure for each person across all waves as discussed in the methodology section for the random effects model using unadjusted self-rated health.

⁴⁸ If there is only one instrument this is equivalent to regressing health against the instrument.

⁴⁹ In previous studies factors such as physical activity, whether a person has ever smoked, whether the respondent has a health condition or whether they are a heavy drinker have been used to instrument self-rated health.

any chronic condition. Both of these would be expected to affect participation only through health. Both of these variables are correlated with self-rated health. If self-rated health was a perfect measure of health then these variables may have been considered to be valid instruments. However, as there are problems with how self-rated health is measured, smoking and chronic disease presence could justifiably be associated with participation outside of self-rated health; perhaps as proxies for health-related aspects not measured accurately by self-rated health. While not a valid test of an instrument, basic models indicated that there appeared to be some correlation between both smoking and chronic disease presence and participation above self-rated health.⁵⁰ After much consideration it was decided that these were not feasible instruments.

Further to there being possible unobserved heterogeneity the link between health and participation is not necessarily one way. If working affects people's health then there is a feedback effect. This feedback effect could be positive or negative. Working long hours in a stressful environment may lead to poorer health or participation may lead to a higher sense of personal and economic security and thus better health (Laplagne *et al*, 2007). The leading method for solving simultaneous equations is by instrumental variables (Wooldridge, 2006). If it had been possible to identify the health equation using an instrument then this feedback effect could have been assessed using simultaneous equations. The first equation would be the identified health equation inclusive of participation, and the second the identified participation equation including the health variable.

There are numerous examples of research that has aimed to assess the feedback effect. For example, Stern (1989) used a simultaneous equations approach using a list of symptoms to instrument self-rated health or the presence of a health condition that limits work that can be undertaken. In this paper he used presence of different chronic conditions to identify his health equation. However, the indicator variable of whether any chronic disease was present was not significant in the model on top of self-reported health, so these chronic conditions identify the health equation. In the case of the SoFIE data the same is not true. In contrast to other literature, while Stern found that there was a significant feedback effect, he found this was not large (Currie *et al*, 1999). In any case, the impact of participation on health is not clearly in any one direction.

As a result of the fact that there does not appear to be any compelling instruments to identify the health equation for all three waves of SoFIE and that, in any case, conclusions can be very sensitive to the instruments chosen, this work does not attempt to adjust for unobserved heterogeneity that is time-variant or to assess the feedback effect.

General model information

For all of the models in this paper the logit model was used as opposed to the probit model; however, this choice is not critical as it has been proven that the two give very similar results (Freese and Scott Long *et al*, 2006).⁵¹ The models were fitted using Stata Version 9. Analysis was undertaken at the Statistics New Zealand Datalab. All variables and variable categories were included in the model even if they were found not to be

⁵⁰ Note these variables could have been included in the original participation equation; however, as they are correlated with self-rated health they were excluded. Further, even if they are proxies for time-invariant unobserved variables these will be removed in the panel models. In any case including them in the pooled regressions only marginally increases the R^2 and the coefficients for the health variables are largely unchanged.

⁵¹ The differences in the coefficients from the logit and probit models are owing to different assumptions about the distribution of errors. The magnitude of the coefficients from the logit and probit models is proportional and there is little or no difference in the predicted probabilities.

significant at the 95% level. This was done for completeness and to aid comparability between models. Residual plots of the models were examined but are not presented as Statistics New Zealand does not release them. Significance in this report is reported at the 95% level unless stated.

All descriptive figures presented in the report are based on weighted data; this is to ensure the figures are representative of the population. While Stata allows sampling weights to be accounted for in basic logit models it is not always possible to allow for these in the more advanced models. Further, while the survey command in Stata enables the sampling design to be taken into account, this survey command cannot be used with more advanced models. In any case, owing to confidentiality, not all the information on the sampling scheme is available which would allow full adjustment.⁵² For this reason all models were carried out unweighted and without adjusting for the sampling design. This is likely to make little difference to the magnitude of resulting estimates and lead to the same conclusions.⁵³

⁵² Statistics New Zealand provides information that allows the identification of the primary sampling units (PSUs) (geographical areas), secondary sampling units (SSUs) (households) and strata, however, as the total number of PSUs in each strata and the total number of SSUs in each PSU are not currently available to SoFIE users, the survey command in Stata would assume that the PSUs were sampled with replacement from the strata, therefore resulting in the secondary sampling stages being ignored. This means that in the pooled logistic model the fact that the responses for the same person are not independent could not be accounted for.

⁵³ The impact of not adjusting for the sampling weight or the survey design is likely to be small. Using the pooled logistic regression model, models were run with and without the weights and accounting and not accounting for the survey design (the SSU and the relationship between the responses of the same person in the different waves could not be accounted for as explained in footnote 39) to get an idea of the impact of not accounting for these factors. There was little difference in the conclusions reached using weighted or unweighted data or data adjusted or unadjusted for the sampling design. Therefore all models' results presented in this paper are based on unweighted data to aid comparability. Allowing for the sampling weights affects the estimated coefficient and the estimated standard errors (SEs). The weights result in coefficients that are slightly lower than those estimates that don't allow for the sampling weights. However, the differences are small and lead to the same conclusions being made about the variables that are and are not significant. Accounting for the survey design impacts on the SEs of the estimates rather than the estimates themselves. As would be expected, not accounting for the strata results in SEs that are higher than they otherwise would be. Reversely, not accounting for the PSU clusters results in the SEs being smaller than they otherwise would have been. Not accounting for the strata and the PSU clusters results in SEs that are only very slightly smaller than if these had been adjusted for.

Appendix D

Appendix Table D1 – Estimated coefficients for labour force participation – pooled logistic regression model – grouped chronic diseases: 2002/03 to 2004/05

	Coefficient	Standard error	P value	95% confidence intervals	
				Lower	Upper
Sex (base=male)					
Female	-0.247**	0.104	0.017	-0.452	-0.043
Region (base=Auckland)					
Waikato	0.134	0.088	0.128	-0.038	0.306
Wellington	0.004	0.076	0.960	-0.144	0.152
Rest of North Island	-0.024	0.067	0.723	-0.155	0.108
Canterbury	0.066	0.075	0.381	-0.081	0.213
Rest of South Island	-0.068	0.077	0.378	-0.220	0.083
Born in New Zealand (base=yes)					
No	-0.111*	0.067	0.096	-0.243	0.020
Ethnicity (base=NZ/European)					
Māori	-0.135**	0.066	0.042	-0.265	-0.005
Pacific Islander	-0.032	0.109	0.771	-0.244	0.181
Other	-0.170*	0.101	0.093	-0.368	0.028
Age at interview date					
	-0.108***	0.005	0.000	-0.118	-0.098
Aged 50 and over (base=15-49)					
Aged 50 and over	6.510***	0.670	0.000	5.198	7.823
Highest qualification					
(base=school qualification)					
Post-school vocational qualification	0.190***	0.057	0.001	0.079	0.301
Degree or higher	0.840***	0.078	0.000	0.687	0.992
No qualification	-0.414***	0.063	0.000	-0.537	-0.291
Chronic disease presence (base=no (or u/k) chronic diseases)					
One or more known chronic diseases	-0.378***	0.045	0.000	-0.467	-0.289
Studying (base=no studying)					
	-0.357***	0.056	0.000	-0.466	-0.248
Other household income					
	-0.007	0.006	0.241	-0.018	0.005
Partner (base=working partner)					
Non-working partner	-1.396***	0.102	0.000	-1.596	-1.196
No partner	-1.053***	0.102	0.000	-1.253	-0.854

	Coefficient	Standard error	P value	95% confidence intervals	
				Lower	Upper
Children (base=no children)					
Child(ren) minimum age 0-<5	0.634***	0.158	0.000	0.325	0.943
Child(ren) minimum age 5-17	-0.239**	0.107	0.025	-0.448	-0.029
Years paid employment	0.182***	0.009	0.000	0.165	0.199
Years paid employment squared	-0.001***	0.000	0.000	-0.001	-0.001
Unemployment rate	-0.124***	0.031	0.000	-0.185	-0.063
Interactions					
Female*Child(ren) minimum age 0-<5	-2.925***	0.166	0.000	-3.251	-2.599
Female*Child(ren) minimum age 5-17	-0.254**	0.123	0.038	-0.495	-0.014
Female*Non-working partner	0.084	0.147	0.567	-0.204	0.372
Female*No partner	0.369***	0.115	0.001	0.144	0.594
Aged 50 and over*Age	-0.130***	0.013	0.000	-0.155	-0.106
Constant	5.024***	0.224	0.000	4.586	5.462
Model summary statistics					
Number of observations			39,310		
Number of unique respondents (clusters)			13,940		
Chi-squared			3,401.23		
Log-likelihood			-13,178.71		
Pseudo R ²			0.2968		

Source: SoFIE Waves 1-3 Version 4, unweighted, Statistics New Zealand

Notes:

1. Based on original sample members with responses in all three waves 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. Responses in each wave are included in the model separately. The relationship between person responses in each wave was accounted for by defining the people as clusters. The number of observations in each wave is not equal owing to the small number of missing values for variables of interest in certain waves or owing to student/retirement status changing between waves. 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. Psychiatric conditions include depression, manic depression and schizophrenia.
4. The likelihood of labour market participation was modelled. Not participating is the base category.

Appendix Table D2 – Mean and standard deviations of variables – pooled regression models – individual chronic diseases: 2002/03 to 2004/05

	Mean	Standard deviation
Labour force participation (participation=1, not participating=0)	0.816	0.387
Labour market outcome (full-time=0, part-time=1, unemployed=2, inactive=3)	0.758	1.153
Gender (male=0, female=1)	0.538	0.499
Region (base=Auckland)		
Waikato (=1)	0.089	0.285
Wellington (=1)	0.135	0.342
Rest of North Island (=1)	0.217	0.412
Canterbury (=1)	0.162	0.368
Rest of South Island (=1)	0.143	0.350
Born in NZ (yes=1, no=0)	0.198	0.398
Ethnicity (base=NZ/European)		
Māori (=1)	0.117	0.321
Pacific Islander (=1)	0.046	0.209
Other (=1)	0.067	0.250
Age at interview date	42.250	12.242
Age 50 and over (15-49=0, 50 and over=1)	0.311	0.463
Highest Qualification (base=school qualification)		
Post-school vocational qualification (=1)	0.371	0.483
Degree or higher (=1)	0.161	0.367
No qualification (=1)	0.212	0.409
Asthma (asthma=1, no asthma=0)	0.186	0.389
High blood pressure (High blood pressure=1, no high blood pressure=0)	0.163	0.370
High cholesterol (High cholesterol=1, no high cholesterol=0)	0.140	0.347
Heart disease (Heart disease=1, no heart disease=0)	0.032	0.177
Diabetes (diabetes=1, no diabetes=0)	0.033	0.177
Stroke (stroke=1, no stroke=0)	0.011	0.105

	Mean	Standard deviation
Migraine (migraine=1, no migraine=0)	0.140	0.347
Psychiatric conditions (Psychiatric conditions=1, no psychiatric conditions=0)	0.103	0.304
Cancer (base=no cancer) Cancer (=1) Unknown (=1)	0.029 0.235	0.169 0.424
Studying (no studying in reference period=0, studying in reference period=1)	0.119	0.323
Other household income	8.398	4.083
Partner (base=working partner) Non-working partner (=1) No partner (=1)	0.113 0.308	0.317 0.462
Children (base=no children) Child(ren) minimum age 0-<5 Child(ren) minimum age 5-17	0.161 0.272	0.367 0.445
Years paid employment	22.116	12.399
Unemployment rate	4.174	0.531
Number of observations	39,310	

Source: SoFIE Waves 1-3 Version 4, unweighted, Statistics New Zealand

Notes:

1. Based on original sample members with responses in all three waves 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. Data for all three waves is pooled together to create an average rate.
2. Psychiatric conditions include depression, manic depression and schizophrenia.

Appendix Table D3 – Estimated coefficients for labour force participation – pooled logistic regression model – individual chronic diseases: 2002/03 to 2004/05

	Coefficient	Standard error	P value	95% confidence intervals	
				Lower	Upper
Sex (base=male)					
Female	-0.383***	0.106	0.000	-0.592	-0.174
Region (base=Auckland)					
Waikato	0.135	0.088	0.127	-0.038	0.308
Wellington	0.018	0.076	0.813	-0.130	0.166
Rest of North Island	-0.015	0.068	0.827	-0.147	0.118
Canterbury	0.103	0.076	0.173	-0.045	0.252
Rest of South Island	-0.077	0.078	0.323	-0.229	0.075
Born in New Zealand (base=yes)					
No	-0.118*	0.067	0.081	-0.250	0.014
Ethnicity (base=NZ/European)					
Māori	-0.139**	0.067	0.038	-0.270	-0.008
Pacific Islander	0.031	0.109	0.774	-0.182	0.245
Other	-0.149	0.102	0.142	-0.348	0.050
Age at interview date	-0.101***	0.005	0.000	-0.111	-0.091
Aged 50 and over (base=15-49)					
Aged 50 and over	6.669***	0.676	0.000	5.343	7.994
Highest qualification (base=school qualification)					
Post-school vocational qualification	0.194***	0.057	0.001	0.083	0.305
Degree or higher	0.812***	0.078	0.000	0.660	0.965
No qualification	-0.392***	0.063	0.000	-0.515	-0.268
Asthma (base=no asthma)	-0.090	0.055	0.102	-0.198	0.018
High blood pressure (base=no high blood pressure)	-0.169***	0.064	0.008	-0.294	-0.045
High cholesterol (base=no high cholesterol)	-0.080	0.071	0.257	-0.218	0.058
Heart disease (base=no heart disease)	-0.662***	0.120	0.000	-0.898	-0.426
Diabetes (base=no diabetes)	-0.553***	0.117	0.000	-0.782	-0.324
Stroke (base=no stroke)	-0.897***	0.181	0.000	-1.253	-0.541

	Coefficient	Standard error	P value	95% confidence intervals	
				Lower	Upper
Migraine (base=no migraine)	-0.043	0.064	0.501	-0.168	0.082
Psychiatric conditions					
(base=no psychiatric conditions)	-1.207***	0.115	0.000	-1.433	-0.981
Cancer (base=no cancer)					
Cancer	-0.068	0.129	0.598	-0.321	0.185
Unknown	-0.129**	0.053	0.016	-0.234	-0.024
Studying (base=no studying)	-0.355***	0.056	0.000	-0.464	-0.245
Other household income	-0.010*	0.006	0.090	-0.021	0.002
Partner (base=working partner)					
Non-working partner	-1.384***	0.105	0.000	-1.589	-1.179
No partner	-1.015***	0.103	0.000	-1.216	-0.813
Children (base=no children)					
Child(ren) minimum age 0-<5	0.589***	0.160	0.000	0.276	0.902
Child(ren) minimum age 5-17	-0.304***	0.107	0.004	-0.513	-0.095
Years paid employment	0.179***	0.009	0.000	0.162	0.196
Years paid employment squared	-0.001***	0.000	0.000	-0.001	-0.001
Unemployment rate	-0.129***	0.031	0.000	-0.190	-0.068
Interactions					
Female*Psychiatric conditions	0.701***	0.137	0.000	0.432	0.971
Female*Child(ren) minimum age 0-<5	-2.863***	0.169	0.000	-3.193	-2.533
Female*Child(ren) minimum age 5-17	-0.203	0.123	0.101	-0.444	0.039
Female*Non-working partner	0.085	0.149	0.567	-0.206	0.377
Female*No partner	0.367***	0.115	0.001	0.141	0.593
Aged 50 and over*Age	-0.134***	0.013	0.000	-0.159	-0.109
Constant	4.965***	0.225	0.000	4.523	5.406
Model summary statistics					
Number of observations				39,310	
Number of unique respondents (clusters)				13,940	
Chi-squared				3,543.77	
Log-likelihood				-12,949.58	
Pseudo R ²				0.309	

Source: SoFIE Waves 1-3 Version 4, unweighted, Statistics New Zealand

Note: See footnotes on Table D1.

Appendix Table D4 – Estimated coefficients for labour market outcome – pooled multinomial logistic regression model – grouped chronic diseases: 2002/03 to 2004/05

	Coefficients		
	Full-time	Part-time	Unemployed
Sex (base=male)			
Female	-0.503***	0.856***	-0.608***
Region (base=Auckland)			
Waikato	-0.128	0.116	0.347**
Wellington	-0.014	-0.008	0.314**
Rest of North Island	-0.098	0.065	0.389***
Canterbury	-0.015	0.214**	0.211
Rest of South Island	-0.122	0.064	-0.234
Born in New Zealand (base=yes)	-0.125*	-0.107	0.248*
Ethnicity (base=NZ/European)			
Māori	-0.105	-0.308***	0.441***
Pacific Islander	0.123	-0.365***	0.202
Other	-0.184*	-0.208*	0.376**
Age at interview date	-0.145***	-0.065***	-0.052***
Aged 50 and over (base=15-49)			
Aged 50 and over	7.442***	4.150***	7.693***
Highest qualification (base=school qualification)			
Post-school vocational qualification	0.205***	0.132**	0.306***
Degree or higher	1.052***	0.516***	0.298*
No qualification	-0.481***	-0.434***	0.304**
Chronic disease presence (base=no (or u/k) chronic diseases)	-0.417***	-0.310***	-0.130
One or more known chronic diseases			
Studying (base=no studying)	-0.423***	-0.243***	-0.113
Other household income	-0.014**	0.009	-0.018
Partner (base=working partner)			
Non-working partner	-1.417***	-1.120***	-1.102***
No partner	-1.227***	-0.534***	-0.264
Children (base=no children)			
Child(ren) minimum age 0-<5	0.505***	0.502**	0.415
Child(ren) minimum age 5-17	-0.362***	-0.201	0.114

	Coefficients		
	Full-time	Part-time	Unemployed
Years paid employment	0.235***	0.137***	0.048***
Years paid employment squared	-0.001***	-0.001***	0.000
Unemployment rate	-0.203***	-0.005	0.096
Interactions			
Female*Child(ren) minimum age 0-<5	-3.482***	-1.591***	-2.324***
Female*Child(ren) minimum age 5-17	-0.407***	0.360**	-0.388*
Female*Non-working partner	0.0569	-0.197	0.078
Female*No partner	0.566***	-0.232	0.333
Aged 50 and over*Age	-0.148***	-0.083***	-0.149***
Constant	6.059***	0.595**	-0.494
Model summary statistics			
Number of observations	39,310		
Number of unique respondents (clusters)	13,940		
Chi-squared	5,433.48		
Log-likelihood	-29,708.28		
Pseudo R ²	0.2301		

Source: SoFIE Waves 1-3 Version 4, unweighted, Statistics New Zealand

Notes:

1. See footnotes 1-3 Table D1.
2. The likelihood of different labour market outcomes was modelled. Inactive is the base category.

Appendix Table D5 – Estimated coefficients for labour market outcome – pooled multinomial logistic regression model – individual chronic diseases: 2002/03 to 2004/05

	Coefficients		
	Full-time	Part-time	Unemployed
Sex (base=male)			
Female	-0.641***	0.749***	-0.754***
Region (base=Auckland)			
Waikato	0.128	0.120	0.323*
Wellington	-0.003	0.011	0.320**
Rest of North Island	-0.094	0.077	0.392***
Canterbury	0.022	0.250***	0.217
Rest of South Island	-0.134	0.058	-0.237
Born in New Zealand (base=yes)	-0.129**	-0.115	0.230*
Ethnicity (base=NZ/European)			
Māori	-0.109	-0.311***	0.462***
Pacific Islander	0.201*	-0.322**	0.301
Other	-0.154	-0.206*	0.433**
Age at interview date	-0.138***	-0.06***	-0.048***
Aged 50 and over (base=15-49)			
Aged 50 and over	7.698***	4.226***	7.655***
Highest qualification			
(base=school qualification)			
Post-school vocational qualification	0.21***	0.136**	0.307***
Degree or higher	1.025***	0.498***	0.28*
No qualification	-0.451***	-0.421***	0.312**
Asthma (base=no asthma)	-0.080	-0.114*	-0.039
High blood pressure			
(base=no high blood pressure)	-0.164**	-0.183**	-0.172
High cholesterol			
(base=no high cholesterol)	-0.088	-0.080	-0.003
Heart disease			
(base=no heart disease)	-0.619***	-0.635***	-0.875***
Diabetes (base=no diabetes)	-0.700***	-0.361***	-0.015
Stroke (base=no stroke)	-1.119***	-0.492**	-0.808**
Migraine (base=no migraine)	-0.080	-0.011	0.229*

	Coefficients		
	Full-time	Part-time	Unemployed
Psychiatric conditions (base=no psychiatric conditions)	-1.328***	-0.751***	-0.597***
Cancer (base=no cancer)			
Cancer	-0.056	-0.068	-0.188
Unknown	-0.186***	-0.017	-0.237**
Studying (base=no studying)	-0.420***	-0.241***	-0.119
Other household income	-0.017***	0.007	-0.020*
Partner (base=working partner)			
Non-working partner	-1.411***	-1.106***	-1.082***
No partner	-1.196***	-0.513***	-0.243
Children (base=no children)			
Child(ren) minimum age 0-<5	0.451***	0.460**	0.367
Child(ren) minimum age 5-17	-0.434***	-0.253*	0.048
Years paid employment	0.232***	0.136***	0.046***
Years paid employment squared	-0.001***	-0.001***	0.000
Unemployment rate	-0.210***	-0.008	0.089
Interactions			
Female*Psychiatric conditions	0.694***	0.364**	0.554**
Female*Child(ren) minimum age 0-<5	-3.416***	-1.539***	-2.283***
Female*Child(ren) minimum age 5-17	-0.355***	0.399***	-0.331
Female*Non-working partner	0.065	-0.200	0.073
Female*No partner	0.580***	-0.226	0.318
Aged 50 and over*Age	-0.154***	-0.085***	-0.149***
Constant	6.017***	0.538**	-0.438
Model summary statistics			
Number of observations	39,310		
Number of unique respondents (clusters)	13,940		
Chi-squared	5,601.24		
Log-likelihood	-29,422.78		
Pseudo R ²	0.2375		

Source: SoFIE Waves 1-3 Version 4, unweighted, Statistics New Zealand

Notes:

1. See footnotes 1-3 Table D1.
2. The likelihood of different labour market outcomes was modelled. Inactive is the base category.

Appendix E

Appendix Table E1 – Mean and standard deviations – pooled regression models – self-rated health: 2002/03 to 2004/05

	Mean	Standard deviation
Self-rated health (base=excellent)		
Very good	0.341	0.474
Good	0.196	0.397
Fair	0.055	0.229
Poor	0.015	0.123
Number of observations		39,310

Source: SoFIE Waves 1-3 Version 4, unweighted, Statistics New Zealand

Notes:

1. Based on original sample members with responses in all three waves 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. Data for all three waves is pooled together to create an average rate. The sample is restricted so it is the same as that considered in the models with individual chronic diseases (ie, those with missing indicators of chronic diseases are excluded from this analysis).
2. The means and standard deviations for the non-health variables are the same as those from the individual disease models (Table D2).

Appendix Table E2 – Estimated coefficients for labour force participation – pooled logistic regression model – self-rated health: 2002/03 to 2004/05

	Coefficient	Standard error	P value	95% confidence intervals	
				Lower	Upper
Sex (base=male)					
Female	-0.407***	0.104	0.000	-0.611	-0.204
Region (base=Auckland)					
Waikato	0.163*	0.090	0.068	-0.012	0.339
Wellington	0.006	0.075	0.935	-0.141	0.153
Rest of North Island	0.013	0.068	0.843	-0.120	0.147
Canterbury	0.088	0.075	0.245	-0.060	0.236
Rest of South Island	0.007	0.078	0.927	-0.146	0.161
Born in New Zealand (base=yes)					
No	-0.082	0.068	0.224	-0.215	0.050
Ethnicity (base=NZ/European)					
Māori	-0.079	0.068	0.247	-0.212	0.054
Pacific Islander	0.064	0.108	0.555	-0.148	0.276
Other	-0.062	0.100	0.536	-0.259	0.135
Age at interview date	-0.097***	0.005	0.000	-0.108	-0.087
Aged 50 and over (base=15-49)					
Aged 50 and over	6.922***	0.671	0.000	5.607	8.238
Highest qualification (base=school qualification)					
Post-school vocational qualification	0.197***	0.057	0.001	0.085	0.309
Degree or higher	0.757***	0.078	0.000	0.605	0.910
No qualification	-0.312***	0.064	0.000	-0.437	-0.187
Self-rated health (base=excellent)					
Very good	-0.064	0.045	0.156	-0.152	0.024
Good	-0.578***	0.052	0.000	-0.681	-0.475
Fair	-1.440***	0.078	0.000	-1.594	-1.287
Poor	-2.545***	0.135	0.000	-2.809	-2.280
Studying (base=no studying)	-0.365***	0.057	0.000	-0.477	-0.253
Other household income	-0.012**	0.006	0.048	-0.023	0.000
Partner (base=working partner)					
Non-working partner	-1.337***	0.104	0.000	-1.540	-1.133
No partner	-1.016***	0.102	0.000	-1.217	-0.815

	Coefficient	Standard error	P value	95% confidence intervals	
				Lower	Upper
Children (base=no children)					
Child(ren) minimum age 0-<5	0.560***	0.161	0.000	0.245	0.875
Child(ren) minimum age 5-17	-0.284***	0.107	0.008	-0.494	-0.074
Years paid employment	0.178***	0.009	0.000	0.160	0.195
Years paid employment squared	-0.001***	0.000	0.000	-0.001	-0.001
Unemployment rate	-0.139***	0.032	0.000	-0.201	-0.078
Interactions					
Female*Child(ren) minimum age 0-<5	-2.858***	0.170	0.000	-3.191	-2.526
Female*Child(ren) minimum age 5-17	-0.209*	0.124	0.091	-0.451	0.033
Female*Non-working partner	0.082	0.149	0.581	-0.209	0.373
Female*No partner	0.434***	0.115	0.000	0.209	0.660
Aged 50 and over*Age	-0.139***	0.013	0.000	-0.164	-0.114
Constant	4.945***	0.227	0.000	4.500	5.389
Model summary statistics					
Number of observations			39,310		
Number of unique respondents (clusters)			13,940		
Chi-squared			3,749.79		
Log-likelihood			-12,691.30		
Pseudo R ²			0.3227		

Source: SoFIE Waves 1-3 Version 4, unweighted, Statistics New Zealand

Notes:

1. Based on original sample members with responses in all three waves 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. Responses in each wave are included in the model separately. The relationship between person responses in each wave was accounted for by defining the people as clusters. The number of observations in each wave is not equal owing to the small number of missing values for variables of interest in certain waves or owing to student/retirement status changing between waves. 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 sample is restricted so it is the same as that considered in the models with individual chronic diseases (ie, those with missing indicators of chronic diseases are excluded from this analysis).
4. The likelihood of labour market participation was modelled. Not participating is the base category.

Appendix Table E3 – Estimated coefficients for labour market outcome – pooled multinomial logistic regression model – self-rated health: 2002/03 to 2004/05

	Coefficients		
	Full-time	Part-time	Unemployed
Sex (base=male)			
Female	-0.698***	0.703***	-0.727***
Region (base=Auckland)			
Waikato	0.163*	0.136	0.339**
Wellington	-0.008	-0.007	0.301**
Rest of North Island	-0.055	0.093	0.389***
Canterbury	0.011	0.229***	0.206
Rest of South Island	-0.038	0.126	-0.214
Born in New Zealand (base=yes)	-0.096	-0.083	0.260*
Ethnicity (base=NZ/European)			
Māori	-0.042	-0.264***	0.456***
Pacific Islander	0.238*	-0.286**	0.251
Other	-0.061	-0.120	0.411**
Age at interview date	-0.133***	-0.057***	-0.048***
Aged 50 and over (base=15-49)			
Aged 50 and over	7.929***	4.588***	7.971***
Highest qualification			
(base=school qualification)			
Post-school vocational qualification	0.217***	0.137**	0.307***
Degree or higher	0.963***	0.453***	0.267
No qualification	-0.367***	-0.356***	0.338***
Self-rated health (base=excellent)			
Very good	-0.078	-0.026	0.036
Good	-0.665***	-0.468***	-0.035
Fair	-1.750***	-0.944***	-0.621***
Poor	-2.914***	-1.975***	-1.235***
Studying (base=no studying)	-0.436***	-0.250***	-0.115
Other household income	-0.019***	0.005	-0.021*
Partner (base=working partner)			
Non-working partner	-1.368***	-1.068***	-1.069***
No partner	-1.208***	-0.522***	-0.271
Children (base=no children)			
Child(ren) minimum age 0-<5	0.424**	0.428**	0.347
Child(ren) minimum age 5-17	-0.414***	-0.248*	0.063

	Coefficients		
	Full-time	Part-time	Unemployed
Years paid employment	0.23***	0.135***	0.049***
Years paid employment squared	-0.001***	-0.001***	0.000
Unemployment rate	-0.225***	-0.019	0.089
Interactions			
Female*Child(ren) minimum age 0-<5	-3.417***	-1.533***	-2.267***
Female*Child(ren) minimum age 5-17	-0.365***	0.399***	-0.349
Female*Non-working partner	0.070	-0.204	0.074
Female*No partner	0.656***	-0.163	0.398*
Aged 50 and over*Age	-0.158***	-0.092***	-0.155***
Constant	6.019***	0.577**	-0.503
Model Summary Statistics			
Number of observations	39,310		
Number of unique respondents (clusters)	13,940		
Chi-squared	5,869.14		
Log-likelihood	-29,136.51		
Pseudo R ²	0.2449		

Source: SoFIE Waves 1-3 Version 4, unweighted, Statistics New Zealand

Notes:

1. See footnotes 1-3 of Table E2.
2. The likelihood of different labour market outcomes was modelled. Inactive is the base category.

Appendix F

Appendix Table F1 – Estimated coefficients for labour force participation – fixed effects logistic regression model – self-rated health: 2002/03 to 2004/05

	Coefficient	Standard error	P value	95% confidence intervals	
				Lower	Upper
Region (base=Auckland)					
Waikato	-0.509	0.411	0.216	-1.314	0.297
Wellington	-0.627	0.408	0.125	-1.428	0.173
Rest of North Island	-0.118	0.353	0.738	-0.809	0.573
Canterbury	-1.023**	0.502	0.041	-2.006	-0.039
Rest of South Island	-1.075**	0.485	0.027	-2.026	-0.125
Age at interview date	0.193**	0.080	0.015	0.037	0.350
Aged 50 and over (base=15-49)					
Aged 50 and over	20.605***	3.405	0.000	13.931	27.279
Self-rated health (base=excellent)					
Very good	0.028	0.088	0.750	-0.144	0.200
Good	-0.078	0.105	0.459	-0.284	0.128
Fair	-0.563***	0.153	0.000	-0.863	-0.263
Poor	-1.503***	0.258	0.000	-2.008	-0.999
Other household income	-0.010	0.013	0.405	-0.035	0.014
Partner (base=working partner)					
Non-working partner	-1.479***	0.257	0.000	-1.983	-0.976
No partner	-0.373	0.310	0.228	-0.980	0.234
Children (base=no children)					
Child(ren) minimum age 0-<5	-0.163	0.382	0.670	-0.912	0.586
Child(ren) minimum age 5-17	-0.424	0.282	0.133	-0.977	0.129
Unemployment rate	-0.054	0.148	0.716	-0.344	0.236
Interactions					
Female*Child(ren) minimum age 0-<5	-1.925***	0.434	0.000	-2.775	-1.075
Female*Child(ren) minimum age 5-17	-0.409	0.353	0.246	-1.101	0.282
Female*Non-working partner	0.222	0.338	0.512	-0.441	0.885
Female*No partner	-0.055	0.353	0.877	-0.748	0.638
Aged 50 and over*Age	-0.427***	0.068	0.000	-0.560	-0.293

Model summary statistics

Number of observations	5,710
Number of unique respondents (clusters)	1,970
Chi-squared	329.44
Log-likelihood	-1,918.34

Source: SoFIE Waves 1-3 Version 4, unweighted, Statistics New Zealand

Notes:

1. Based on original sample members with responses in all three waves 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. Variables that do not change over time (ie, gender and place of birth), that are little or slow changing (eg, ethnicity and highest qualification) or that could be impacted on by health changes (ie, studying status and years in paid employment) are excluded from these models. Significant and insignificant variables or variable categories are kept in for completeness.
2. The relationship between changes in self-rated health and participation was modelled.
3. *Significant at the 90% level. **Significant at the 95% level. ***Significant at the 99% level.

Appendix Table F2 – Estimated coefficients for labour force participation – correlated random effects logistic regression model – self-rated health: 2002/03 to 2004/05

	Coefficient	Standard error	P value	95% confidence intervals	
				Lower	Upper
Sex (base=male)					
Female	-2.034***	0.119	0.000	-2.268	-1.800
Region (base=Auckland)					
Waikato	0.167	0.117	0.153	-0.062	0.395
Wellington	0.044	0.100	0.659	-0.152	0.240
Rest of North Island	-0.012	0.088	0.890	-0.184	0.160
Canterbury	0.198**	0.097	0.040	0.009	0.388
Rest of South Island	0.112	0.101	0.267	-0.086	0.311
Born in New Zealand (base=yes)	-0.422***	0.077	0.000	-0.573	-0.270
Age at interview date	0.051***	0.004	0.000	0.043	0.060
Aged 50 and over (base=15-49)					
Aged 50 and over	13.620***	0.695	0.000	12.257	14.983
Self-rated health (base=excellent)					
Very good	0.003	0.075	0.970	-0.143	0.149
Good	-0.069	0.090	0.446	-0.246	0.108
Fair	-0.419***	0.130	0.001	-0.673	-0.165
Poor	-1.058***	0.214	0.000	-1.477	-0.639
Average time in health state (base=excellent health)					
Very good	-0.157	0.126	0.211	-0.404	0.089
Good	-1.648***	0.139	0.000	-1.921	-1.375
Fair	-3.380***	0.209	0.000	-3.790	-2.970
Poor	-5.357***	0.365	0.000	-6.074	-4.641
Other household income	-0.022***	0.007	0.003	-0.037	-0.008
Partner (base=working partner)					
Non-working partner	-1.876***	0.130	0.000	-2.130	-1.621
No partner	-2.122***	0.130	0.000	-2.377	-1.866
Children (base=no children)					
Child(ren) minimum age 0-5	0.250	0.181	0.166	-0.104	0.604
Child(ren) minimum age 5-17	-0.744***	0.131	0.000	-1.000	-0.487
Unemployment rate	-0.203***	0.042	0.000	-0.285	-0.122

	Coefficient	Standard error	P value	95% confidence intervals	
				Lower	Upper
Interactions					
Female*Child(ren) minimum age 0-<5	-3.407***	0.197	0.000	-3.793	-3.021
Female*Child(ren) minimum age 5-17	-0.324**	0.151	0.032	-0.621	-0.028
Female*Non-working partner	-0.258	0.178	0.147	-0.607	0.091
Female*No partner	1.289***	0.144	0.000	1.008	1.570
Aged 50 and over*Age	-0.280***	0.013	0.000	-0.305	-0.255
Constant	5.728***	0.293	0.000	5.153	6.303
Model summary statistics					
$\ln(\sigma^2_\alpha)$	1.561	0.023		1.515	1.606
σ^2_α	2.182	0.025		2.133	2.232
ρ	0.591	0.006		0.580	0.602
Number of observations			39,310		
Number of unique respondents (clusters)			13,940		
Chi-squared			3,909.77		
Log-likelihood			-11,994.97		

Source: SoFIE Waves 1-3 Version 4, unweighted, Statistics New Zealand

Notes:

1. Based on original sample members with responses in all three waves 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. Variables that are little or slow changing (eg, ethnicity and highest qualification) or that could be impacted on by health changes (ie, studying status and years in paid employment) are excluded from these models. Significant and insignificant variables or variable categories are kept in for completeness.
2. The relationship between changes and stocks of self-rated health and participation was modelled.
3. *Significant at the 90% level. **Significant at the 95% level. ***Significant at the 99% level.

Appendix G

Appendix Table G1 – Estimated coefficients for self-rated health – ordered logistic regression model: 2002/03 to 2004/05

	Coefficients		
	Wave 1	Wave 2	Wave 3
Sex (base=male)			
Female	-0.133***	0.016	-0.124***
Region (base=Auckland)			
Waikato	0.271***	0.197***	0.124***
Wellington	0.078	-0.049	0.310***
Rest of North Island	0.193***	0.080	0.182***
Canterbury	0.084	-0.062	0.167***
Rest of South Island	0.279***	0.337***	0.055
Born in New Zealand (base=yes)	0.147***	0.121**	0.581***
Ethnicity (base=NZ/European)			
Māori	0.350***	0.344***	-0.090
Pacific Islander	0.162*	0.230**	0.234***
Other	0.312***	0.459***	0.135
Age at interview date	0.026***	-0.130***	0.021***
Aged 50 and over (base=15-49)			
Aged 50 and over	-0.294	-0.181	-0.008***
Highest qualification			
(base=school qualification)			
Post-school vocational qualification	-0.018	-0.090**	0.347***
Degree or higher	-0.342***	-0.363***	-0.070*
No qualification	0.317***	0.296***	-0.316***
Asthma (base=no asthma)	0.485***	0.340***	0.440***
High blood pressure			
(base=no high blood pressure)	0.500***	0.406***	0.522***
High cholesterol			
(base=no high cholesterol)	0.177***	0.507***	0.156***
Heart disease			
(base=no heart disease)	0.917***	0.154***	0.915***
Diabetes (base=no diabetes)	1.086***	0.884***	0.975***

	Coefficients		
	Wave 1	Wave 2	Wave 3
Stroke (base=no stroke)	0.736***	0.859***	0.519***
Migraine (base=no migraine)	0.368***	0.594***	0.386***
Psychiatric conditions (base=no psychiatric conditions)	0.953***	0.381***	0.986***
Cancer (base=no cancer)			
Cancer	0.418***	0.873***	0.582***
Unknown	0.095**	0.460***	0.043
Studying (base=no studying)	-0.170***	-0.129**	0.257***
Total household income	-0.059***	-0.109***	0.014
Partner (base=working partner)			
Non-working partner	0.158***	0.264***	-0.055
No partner	0.128***	0.100**	0.139***
Children (base=no children)			
Child(ren) minimum age 0-<5	-0.182***	-0.107*	0.133***
Child(ren) minimum age 5-17	-0.053	0.017	-0.042
Tenure (base=not owned)			
Owned with mortgage	-0.140***	-0.006	-0.086***
Owned outright	-0.123**	0.060	0.002
Years paid employment	-0.014***	-0.010***	0.052
Sickness benefit (base=no sickness benefit)	1.286***	0.004	0.868***
Smoked (base=never smoked)	0.381***	1.009***	0.344***
Interactions			
Female*Psychiatric conditions	-0.250**	-0.063	-0.269**
Aged 50 and over*Age	0.005	0.024***	0.002
Cut points			
Cut point 1	0.432	-0.013	0.015
Cut point 2	2.146	1.652	1.748
Cut point 3	4.025	3.515	3.601
Cut point 4	6.034	5.437	5.459

	Coefficients		
	Wave 1	Wave 2	Wave 3
Model summary statistics			
Number of respondents	17,190	17,195	17,355
Chi-squared	3,629.70	3,918.74	3,601.95
Log-likelihood	-19,648.90	-19,715.08	-20,633.86
Pseudo R ²	0.1084	0.1131	0.1052

Source: SoFIE Waves 1-3 Version 4, standard longitudinal weights, Statistics New Zealand

Notes:

1. Based on original sample members with responses in all three waves who are aged over 15 at the end of the reference period in Wave 1. Responses in each wave are included in the model separately. The number of observations in each wave is not equal owing to the small number of missing values for variables of interest in certain waves. 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. Psychiatric conditions include depression, manic depression and schizophrenia.
4. The likelihood of different self-rated health states was modelled using an ordinal logistic regression model.

Appendix Table G2 – Estimated coefficients for labour force participation – pooled logistic regression model – adjusted health stock: 2002/03 to 2004/05

	Coefficient	Standard error	P value	95% confidence intervals	
				Lower	Upper
Sex (base=male)					
Female	-0.421***	0.107	0.000	-0.630	-0.211
Region (base=Auckland)					
Waikato	0.196**	0.090	0.029	0.020	0.371
Wellington	0.041	0.075	0.581	-0.106	0.189
Rest of North Island	0.051	0.068	0.453	-0.082	0.183
Canterbury	0.122	0.075	0.107	-0.026	0.269
Rest of South Island	0.033	0.078	0.675	-0.121	0.186
Born in New Zealand (base=yes)					
No	-0.090	0.067	0.180	-0.222	0.042
Ethnicity (base=NZ/European)					
Māori	0.000	0.069	0.997	-0.135	0.135
Pacific Islander	0.101	0.108	0.348	-0.110	0.312
Other	-0.053	0.100	0.595	-0.249	0.143
Age at interview date	-0.092***	0.005	0.000	-0.103	-0.082
Aged 50 and over (base=15-49)					
Aged 50 and over	6.488***	0.692	0.000	5.132	7.845
Highest qualification (base=school qualification)					
Post-school vocational qualification	0.163***	0.056	0.004	0.053	0.274
Degree or higher	0.716***	0.077	0.000	0.565	0.868
No qualification	-0.249***	0.064	0.000	-0.375	-0.123
Health stock	-0.837***	0.079	0.000	-0.991	-0.682
Studying (base=no studying)	-0.381***	0.056	0.000	-0.490	-0.272
Other household income	-0.016***	0.006	0.006	-0.028	-0.005
Partner (base=working partner)					
Non-working partner	-1.266***	0.109	0.000	-1.479	-1.053
No partner	-0.978***	0.105	0.000	-1.184	-0.771
Children (base=no children)					
Child(ren) minimum age 0-<5	0.507***	0.163	0.002	0.188	0.826
Child(ren) minimum age 5-17	-0.357***	0.108	0.001	-0.569	-0.145

	Coefficient	Standard error	P value	95% confidence intervals	
				Lower	Upper
Years paid employment	0.178***	0.009	0.000	0.161	0.195
Years paid employment squared	-0.001***	0.000	0.000	-0.001	-0.001
Unemployment rate	-0.088***	0.032	0.005	-0.150	-0.026
Interactions					
Female*Child(ren) minimum age 0-<5	-2.782***	0.171	0.000	-3.118	-2.446
Female*Child(ren) minimum age 5-17	-0.192	0.125	0.126	-0.437	0.054
Female*Non-working partner	0.082	0.152	0.590	-0.216	0.380
Female*No partner	0.423***	0.117	0.000	0.193	0.653
Aged 50 and over*Age	-0.131***	0.013	0.000	-0.156	-0.105
Constant	4.147***	0.233	0.000	3.690	4.604
Model summary statistics					
Number of observations				39,270	
Number of unique respondents (clusters)				13,930	
Chi-squared				3,333.23	
Log-likelihood				-12,723.33	
Pseudo R ²				0.3201	

Source: SoFIE Waves 1-3 Version 4, unweighted, Statistics New Zealand

Notes:

1. See footnotes 1 and 2 of Table G1.
2. The likelihood of labour market participation was modelled. Not participating was the base category.
3. The number of observations the model is based on is lower than that for the self-rated health or individual disease models owing to small numbers of missing values for the objective health measures.

Appendix Table G3 – Estimated coefficients for labour force participation – fixed effects logistic regression model – adjusted health stock: 2002/03 to 2004/05

	Coefficient	Standard error	P value	95% confidence intervals	
				Lower	Upper
Region (base=Auckland)					
Waikato	-0.439	0.414	0.290	-1.250	0.373
Wellington	-0.613	0.407	0.132	-1.411	0.184
Rest of North Island	-0.169	0.359	0.638	-0.872	0.534
Canterbury	-1.006**	0.496	0.042	-1.978	-0.035
Rest of South Island	-1.126**	0.483	0.020	-2.073	-0.180
Age at interview date	0.186**	0.081	0.021	0.028	0.344
Aged 50 and over (base=15-49)					
Aged 50 and over	19.692***	3.449	0.000	12.931	26.452
Health stock	-0.572***	0.140	0.000	-0.846	-0.298
Other household income	-0.012	0.013	0.340	-0.037	0.013
Partner (base=working partner)					
Non-working partner	-1.462***	0.266	0.000	-1.982	-0.942
No partner	-0.347	0.316	0.272	-0.966	0.273
Children (base=no children)					
Child(ren) minimum age 0-<5	-0.036	0.382	0.925	-0.784	0.712
Child(ren) minimum age 5-17	-0.378	0.283	0.182	-0.934	0.177
Unemployment rate	-0.039	0.150	0.795	-0.332	0.254
Interactions					
Female*Child(ren) minimum age 0-<5	-2.029***	0.433	0.000	-2.877	-1.180
Female*Child(ren) minimum age 5-17	-0.419	0.354	0.236	-1.112	0.274
Female*Non-working partner	0.225	0.346	0.515	-0.453	0.904
Female*No partner	-0.049	0.359	0.891	-0.753	0.655
Aged 50 and over*Age	-0.409***	0.069	0.000	-0.544	-0.274

Model summary statistics

Number of observations	5,575
Number of unique respondents (clusters)	1,925
Chi-squared	302.90
Log-likelihood	-1,882.69

Source: SoFIE Waves 1-3 Version 4, unweighted, Statistics New Zealand

Notes:

1. Based on original sample members with responses in all three waves 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. Variables that do not change over time (ie, gender and place of birth), that are little or slow changing (eg, ethnicity and highest qualification) or that could be impacted on by health changes (ie, studying status and years in paid employment) are excluded from these models. Significant and insignificant variables or variable categories are kept in for completeness.
2. The relationship between changes in self-rated health and participation was modelled.
3. *Significant at the 90% level. **Significant at the 95% level. ***Significant at the 99% level.

Appendix Table G4 – Estimated coefficients for labour force participation – correlated random effects logistic regression model – adjusted health stock: 2002/03 to 2004/05

	Coefficient	Standard error	P value	95% confidence intervals	
				Lower	Upper
Sex (base=male)					
Female	-1.913***	0.118	0.000	-2.145	-1.681
Region (base=Auckland)					
Waikato	0.224*	0.116	0.054	-0.004	0.452
Wellington	0.113	0.100	0.257	-0.083	0.308
Rest of North Island	0.036	0.087	0.680	-0.135	0.208
Canterbury	0.230**	0.096	0.017	0.041	0.419
Rest of South Island	0.106	0.101	0.293	-0.091	0.303
Born in New Zealand (base=yes)	-0.482***	0.077	0.000	-0.633	-0.331
Age at interview date	0.053***	0.004	0.000	0.045	0.062
Aged 50 and over (base=15-49)					
Aged 50 and over	12.774***	0.698	0.000	11.406	14.142
Health stock	-0.560***	0.120	0.000	-0.797	-0.324
Average health stock	-1.148***	0.129	0.000	-1.400	-0.896
Other household income	-0.026***	0.007	0.000	-0.041	-0.011
Partner (base=working partner)					
Non-working partner	-1.710***	0.129	0.000	-1.964	-1.457
No partner	-2.011***	0.129	0.000	-2.265	-1.757
Children (base=no children)					
Child(ren) minimum age 0-<5	0.226	0.181	0.211	-0.128	0.580
Child(ren) minimum age 5-17	-0.794***	0.130	0.000	-1.049	-0.539
Unemployment rate	-0.162***	0.041	0.000	-0.243	-0.080
Interactions					
Female*Child(ren) minimum age 0-<5	-3.388***	0.197	0.000	-3.773	-3.002
Female*Child(ren) minimum age 5-17	-0.352**	0.150	0.019	-0.647	-0.058
Female*Non-working partner	-0.326*	0.177	0.066	-0.673	0.021
Female*No partner	1.191***	0.143	0.000	0.911	1.470
Aged 50 and over*Age	-0.263***	0.013	0.000	-0.288	-0.238
Constant	4.363	0.291	0.000	3.793	4.934

Model summary statistics

$\ln(\sigma_\alpha^2)$	1.569	0.023	1.523	1.614
σ_α^2	2.191	0.025	2.142	2.241
ρ	0.593	0.006	0.582	0.604
Number of observations		39,270		
Number of unique respondents (clusters)		13,925		
Chi-squared		3,642.92		
Log-likelihood		-12,079.28		

Source: SoFIE Waves 1-3 Version 4, unweighted, Statistics New Zealand

Notes:

1. Based on original sample members with responses in all three waves 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. Variables that are little or slow changing (eg, ethnicity and highest qualification) or that could be impacted on by health changes (ie, studying status and years in paid employment) are excluded from these models. Significant and insignificant variables or variable categories are kept in for completeness.
2. The relationship between changes and stocks of self-rated health and participation was modelled.
3. *Significant at the 90% level. **Significant at the 95% level. ***Significant at the 99% level.