Int. J. Epidemiol. Advance Access published May 28, 2009

Published by Oxford University Press on behalf of the International Epidemiological Association
International Journal of Epidemiology 2009;1–7
doi:10.1093/ije/dyp215

COHORT PROFILE

Cohort profile: Survey of Families, Income and Employment (SoFIE) and Health Extension (SoFIE-health)

Kristie N Carter,[1]* Matt Cronin,[2] Tony Blakely,[1] Michael Hayward[1] and Ken Richardson[1]

Accepted 21 April 2009

How did the study come about?

Panel studies in Western countries have transformed and greatly improved understanding of many social, economic and health trends, such as the British Panel Household Survey[1,2] and the Whitehall study in the UK,[3] and the Household, Income and Labour Dynamics in Australia (HILDA) survey.[4] In New Zealand (NZ) there are a number of birth cohort and population-specific longitudinal studies: the Dunedin multidisciplinary health and development study,[5] the Christchurch health and development study,[6] the Pacific Islands Family Study[7] and the Health Work and Retirement longitudinal study.[8] However, there was a need for a longitudinal study that covered all age ranges which could provide an understanding of the dynamics of the NZ economy and its interrelationship between the social and economic well-being of individuals, families and households and the factors affecting this well-being.

Statistics New Zealand was granted funding from the Foundation for Research, Science and Technology (Government organization) in 1997 to conduct a feasibility study for a longitudinal survey of income, employment and family dynamics. Following the feasibility study, the Survey of Families, Income and Employment (SoFIE) study was developed and first went into the field in October 2002.[9]

SoFIE is a single fixed panel longitudinal survey with duration of 8 years. Information is collected once a year from the same individuals on income levels, sources and changes; and on the major influences on income such as employment and education experiences, household and family status and changes, demographic factors and health status. Every 2 years (Waves 2, 4, 6 and 8) information on assets and liabilities is collected to monitor net worth and savings. A successful bid was made to the Health Research Council of NZ by health researchers to have a battery of health questions in Waves 3, 5 and 7—giving rise to the SoFIE-Health sub-study.

What does the study cover?

The overall objective of SoFIE is to provide information about changes over time in the economic well-being of individuals and their families, and about factors influencing those changes. The objectives are intended to provide information that will help design and evaluate government policy in the areas of income support, employment, education, training, retirement provision and family support.

The longitudinal objectives of SoFIE are to: identify the pattern of income level experience over time for individuals and their families; measure significant shifts in income levels for individuals and their families and explore the relationship of these shifts to labour market activity, receipt of government income support and family status; determine patterns over time of labour market activity, participation in education and training and receipt of government income support; identify transitions between spells of labour market activity, education and training participation and receipt of government income support, and examine the factors that influence transitions from one status to another; determine patterns of saving for retirement and relate these to income dynamics and life cycle stages; and determine patterns of change in family status.

The SoFIE-Health add-on is comprised of 20 minutes of questionnaire time in Waves 3 (2004–05), 5 (2006–07) and 7 (2008–09), in the following health-related

[1] Health Inequalities Research Program, School of Medicine and Health Sciences, University of Otago, Wellington, New Zealand.
[2] Statistics New Zealand, Wellington.
* Corresponding author. Health Inequalities Research Program, School of Medicine and Health Sciences, University of Otago, Wellington, PO Box 7343, Wellington South, Wellington 6242, New Zealand. E-mail: kristie.carter@otago.ac.nz

domains: health status (SF-36 and Kessler scale), perceived stress, chronic conditions (heart disease, diabetes and injury-related disability), tobacco smoking, alcohol consumption, health care utilization and access and continuity of primary health care, and an individual deprivation score. There are three major and ongoing goals within SoFIE-Health: (i) determining the impact of labour market factors, asset wealth, income and family dynamics on health; (ii) determining the impact of health status on labour market factors, income trajectories, asset wealth and family dynamics; and (iii) determining the contribution of access, continuity and co-ordination of primary health care to health status and to social inequalities in health.

Who is in the sample?

The target longitudinal population for SoFIE is:

> the usually resident population of New Zealand living in permanent, private dwellings on the main islands in the North and South Islands, including Waiheke Island as at the first wave of the panel.[10]

The survey excludes overseas visitors resident in NZ for <12 months and who intend to stay in NZ for <12 months; non-NZ diplomats and diplomatic staff and their dependants; members of non-NZ armed forces stationed in NZ and their dependants; and people living in institutions or in other non-private dwelling establishments such as boarding houses, hotels, motels and hostels, as well as people living on offshore islands (excluding Waiheke Island) at Wave 1. SoFIE respondents are tracked after Wave 1 if they have moved to a non-private dwelling but respondents who moved to an institute are not contacted.

Sampling frame

SoFIE used the standard Statistics New Zealand sampling frame that partitions the North Island, South Island and Waiheke Island into 19 000 Primary Sampling Units (PSUs). On average, a PSU contains around 70 dwellings, but each can range in size from 30 to 260 dwellings.

Sample selection

The selection of the random sample for SoFIE was a three-stage stratified cluster approach. First, PSUs were assigned to strata according to region, urban/rural, high/low Māori population density and other socioeconomic variables derived from the 1996 census. Then systematic sampling was used to select a sample of PSUs independently from each stratum.

The next stage of sampling involved taking a systematic random sample of permanent private dwellings within the PSUs selected. All of the eligible residents of each selected household who agreed to participate were then included in the sample. These people were designated as original sample members (OSMs) and will be followed and interviewed over time. The sample comprised 1500 PSUs, with an average 7.7 (full and partially) responding households obtained per PSU.

Sample size

At Wave 1, a total of 15 000 randomly-selected households were approached, of which 11 500 agreed to be interviewed (response rate of 77%), with data collected from 29 000 individuals, over 22 000 aged $\geqslant 15$. The longitudinal population represents 3 918 250 individuals. Children <15 years of age at Wave 1 are interviewed as adults at the wave following their 15th birthday. All OSMs are followed up and re-interviewed in subsequent years, regardless of changes in their place of residence. Attempts were made to track all OSMs. If an OSM refused follow up or could not be found and was not interviewed for $\geqslant 2$ or more consecutive years they were no longer tracked. Future members of an OSM's household in Waves 2 and beyond who are not OSMs ('non-OSMs') are interviewed from Wave 2 onwards as part of the cross-sectional population, while they remain living with an OSM. Non-OSMs are asked a reduced set of questions and are not followed up if they leave the OSM's household.

What has been measured?

SoFIE is conducted using computer-assisted interviewing (CAI). In the field, interviewers use laptop computers to administer an electronic questionnaire (EQ), face-to-face, in respondents' homes. The EQ is organized into modules.

There are two sets of questionnaires to SoFIE. For every household selected, a Household Questionnaire (HQ) is first administered to a single adult OSM, which includes questions on the characteristics of the entire household, and then a Personal Questionnaire (PQ) is administered to every person aged $\geqslant 15$. In general, each adult answers their own PQ. However, in some cases a proxy is allowed to fill in the PQ for someone else, for reasons such as disability or language difficulties. Children (aged $\leqslant 14$ at the household enumeration date) do not fill out a PQ. Instead, a nominated adult OSM in the household answers the child module for each child as part of their PQ. Once a child turns 15, they answer the full PQ.

A copy of the questionnaires can be found on the Statistics New Zealand website (www.stats.govt.nz).

The HQ is answered by one OSM in each household and contains two modules.

- Household (e.g. household type, family type).
- Standard of living (e.g. type of housing, appliances owned).

The PQ is administered to all OSMs and new respondents aged $\geqslant 15$, and consists of eight standard modules.

- Demographics (e.g. age, country of birth, self-defined ethnicity).
- Child [demographics about any child(ren) <15 years residing in the household—answered by a nominated adult].
- Labour market history (e.g. age at first paid job).
- Education (e.g. highest qualification).
- Family (e.g. existence of partners, children).
- Labour market (current activity—e.g. details of employment).
- Income (e.g. from paid work, superannuation, government payments).
- Contact (respondent's contact details for ease of follow-up).

Depending on the module, the SoFIE questionnaire collects both point-in-time data and spell data. Point-in-time data relate to a single date, usually the interview date (e.g. the respondent's educational qualifications as at the interview date). Spell data relate to a period of time in the last year with a defined start and end date reported by the respondent (e.g. the period of time a respondent lived with a family member or the length of time a person worked for a particular employer). For example, the start of a new household spell occurs when a person enters or leaves a household.

Information on assets and liabilities is collected every second wave (Waves 2, 4, 6 and 8) and provides detailed estimates of the type and level of assets and liabilities held by individuals, families and households at a point in time. Net worth (overall wealth) is calculated by subtracting the total value of all liabilities from the total value of all assets for individuals.

Every other year (Waves 3, 5 and 7) the health module is asked. The SF-36 (Version 2) is used to measure eight domains of health-related quality-of-life.[11] The Kessler-10 creates a score of non-specific psychological distress across 10 questions.[12] Four questions are used to measure a global level of perceived stress, dealing with the degree to which situations in one's life are appraised as stressful.[13] The prevalence of major chronic diseases (asthma, high blood pressure, high cholesterol, heart disease, diabetes, stroke, migraines, chronic depression, manic depression or schizophrenia) are reported. Information on current and past cigarette use and current alcohol consumption are asked to enable the investigation into changes in health behaviours. The New Zealand Individual Deprivation (NZiDep) index is a tool for measuring socio-economic position for individuals, a personal measure of deprivation/need, and is based on eight simple questions.[14] The health module also includes a number of questions related to first contact access to primary health care providers, first contact utilization and continuity of care, which were modified for application in NZ from the US Primary Care Assessment Tool (PCAT).[15,16] In the first health module (Wave 3) the SoFIE respondents were asked to consent to linking their information with cancer registration, hospital morbidity and mortality data within the NZ Health Information Service. Over 80% of respondents at Wave 3 consented to having their data linked, providing a rich dataset to explore the relationships between prior hospitalizations and cancer registrations with future changes and trajectories in social or economic factors.

How often have participants been followed up?

The first wave (or interview cycle) of SoFIE was conducted from 1 October 2002 to 30 September 2003 and asked the standard HQ and PQ. The second wave was conducted from 1 October 2003 to 30 September 2004 and asked the standard questionnaires as well as an additional asset and liability module. The third wave was conducted from 1 October 2004 to 30 September 2005 and asked the standard questionnaires as well as an additional health module. To date, the first four waves of the SoFIE data are available (on application to the Government Statistician) in microdata format for use in the Statistics New Zealand Data Laboratory environment in Wellington, NZ.

What is attrition like?

Figure 1 shows the flow and attrition of eligible respondents across the first three waves of the SoFIE study (data Version 3). Of the 15 000 households sampled by Statistics New Zealand, 11 500 (77%) agreed to participate in the study. In these households there were 29 685 OSMs (22 165 adults and 7520 children). In Wave 2 there were over 26 000 OSMs responding (89% of Wave 1) and in Wave 3 there were over 23 390 OSMs (88% of Wave 2 OSMs, 79% of Wave 1 OSMs). Attrition was greater in children and a child OSM who turns age 15 years before an interview is subsequently counted and interviewed as an adult OSM. The balanced panel across all three waves was 23 290 OSMs. The overall response rate of 79% of Wave 1 responders re-interviewed in Wave 3 (based on unweighted estimates), combined with the household response rate at Wave 1 of 77%, gives an estimated effective response rate of 61%.

What has been found?

SoFIE, and the SoFIE-Health sub-study, are in early stages of analysis. Here we describe some of the descriptive features of the SoFIE sample population,

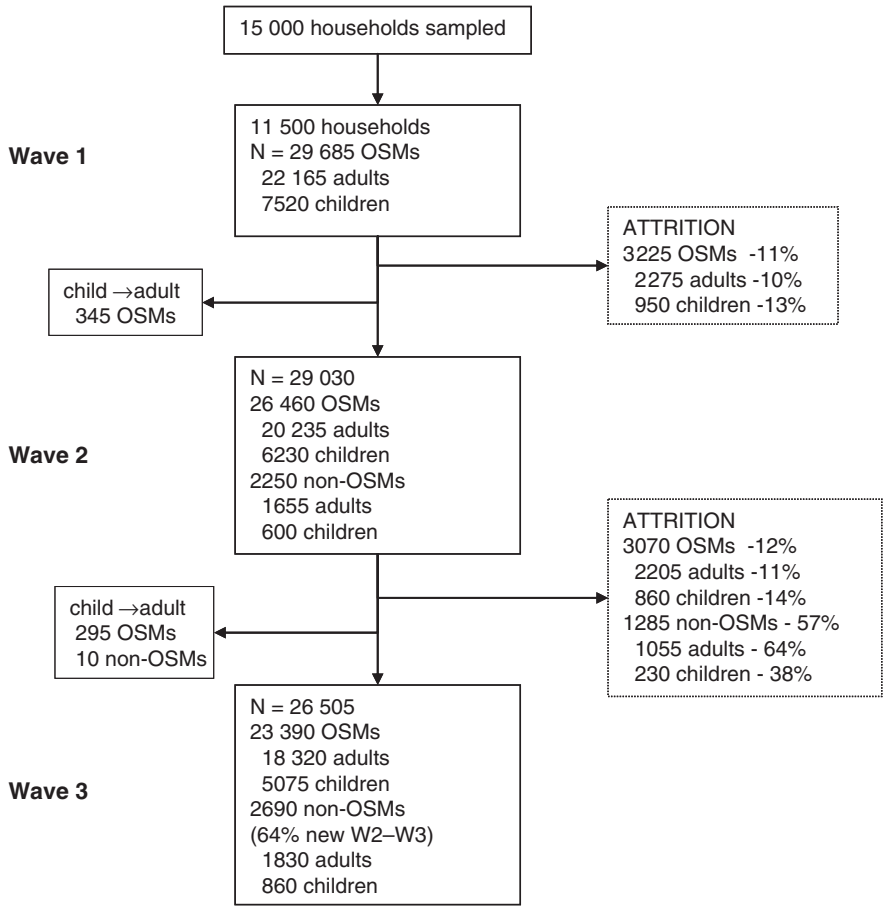

Figure 1 Flow of SoFIE respondents over Waves 1–3

and early analytical findings. Table 1 shows the distribution of respondents at Wave 1 across a number of demographic, social and economic characteristics. Although the sampling strategy in SoFIE did not over sample ethnic minority populations, the ethnic structure of the SoFIE population is relatively similar to the 2001 census population.[17] It seems that the original SoFIE population is relatively healthy, with 70% of adults rating their health as excellent or very good in Wave 1. This is higher than the 60% of adults in the most recent NZ Health Survey 2006/07 rating their health as excellent or very good.[18] There are a number of reasons for the better health status of the SoFIE population compared with the recent NZ Health Survey: (i) different sampling populations (the NZ Health Survey oversampled Māori, Pacific and Asian New Zealanders); (ii) different survey designs (longitudinal panel survey versus cross-sectional health survey); or (iii) attrition or overall response rates.

A recent publication on the first four waves of SoFIE has shown that over half (54%) of those people who received employee earnings in both their first and fourth years of the survey were in the same quintile in both years, nearly 30% moved up one or more quintiles, and 16% moved down one or more quintiles.[10] Over half of people employed, during the first four waves of SoFIE, were employed for the entire survey period (i.e. 4 years, and when unemployment rates were very low in NZ, 3–4%) and 78% of all people aged $\geqslant$15 years spent at least part of the time employed. It has also been shown that 81% of SoFIE participants were in the same type family at Waves 1 and 4.[10] This was supported by a report conducted by the Families Commission using SoFIE data, which found that only 1 in 10 New Zealanders changed family living arrangements between Waves 1 and 2.[19]

The SoFIE data has been used to show that considerable disparities in wealth occur in NZ by age, ethnic group and family status. The top 10% of wealthy individuals own over half of total net worth,[20] whereas, the bottom half of the population collectively owns a mere 5% of total net worth, although this takes into account the 6% of the population with negative net worth. A Gini coefficient, which is a summary measure of inequalities, for net worth was calculated at 0.693, a level comparable with other NZ sources, but lower than the USA and UK.[21]

Table 1 Demographic, social and economic characteristics of SoFIE respondents at Wave 1

Wave 1[a] ($N = 29685$)	*N*	%		*N*	%
Sex			**Personal income[b]**		
Male	14 155	47.7	q1: low–<7600	4100	18.5
Female	15 530	52.3	q2: 7600–<14 310	4085	18.5
Age at interview			q3: 14 310–<25 650	4955	22.4
0–14	7520	25.3	q4: 25 650–<41 480	4495	20.3
15–24	3800	12.8	q5: 41 480–high	4510	20.4
25–34	3730	12.6	**Household income**		
35–44	4530	15.3	q1: low–<21 070	5350	18.0
45–54	3760	12.7	q2: 21 070–<34 010	7280	24.5
55–64	2895	9.8	q3: 34 010–<49 380	6065	20.4
65+	3450	11.6	q4: 49 380–<72 280	5510	18.6
Marital status[b]			q5: 72 280–<high	5470	18.4
Never Married	7320	33.0	**Labour market activity[b]**		
Married	11 090	50.0	Working	13 590	61.3
Divorced	1465	6.6	Not employed, looking for work	605	2.7
Widowed	1410	6.4	Not employed, not looking for work	7965	35.9
Separated	875	3.9	Overseas	10	0.0
DK/REF[c]	10	0.1	**NZ deprivation**		
Ethnicity prioritized			NZDepQ1 (least)	5260	17.7
NZ/European	20 385	68.7	NZDepQ2	5540	18.7
Māori	4865	16.4	NZDepQ3	4910	16.5
Pacific	2130	7.2	NZDepQ4	6695	22.6
Asian	1750	5.9	NZDepQ5 (most)	7270	24.5
Other	550	1.9	**Major region**		
Self-rated health[b]			Auckland	8510	28.7
Excellent	8335	37.6	Waikato	2740	9.2
Very Good	7145	32.2	Wellington	3650	12.3
Good	4600	20.8	Rest of North Island	6760	22.9
Fair	1590	7.2	Canterbury	4235	14.3
Poor	490	2.2	Rest of South Island	3785	12.8
Highest educational qualification[b]					
No qualification	5980	26.9			
School qualification	6185	27.9			
Post-school vocational	7125	32.1			
Degree or higher	2880	13.0			

[a]All counts and values in the tables have been randomly rounded (up or down) to the nearest multiple of five and cells with counts less than 10 were imputed with the value 10. Therefore, table totals may differ from the sum of individual cells. Some column percentages in the tables may also sum to greater than 100, as the percentages were calculated according to the random rounded totals.
[b]Asked of adults only.
[c]Don't know/refused to answer.

The NZ Treasury used the SoFIE data to study housing wealth in household portfolios and explore the effect of some home equity withdrawal on the required saving rates.[22] In the SoFIE population, 60% of households are recorded as owning a home and over half of these do not have mortgage debt. Calculating the median ratio of net equity in property to total net worth indicates that property represents over half of total net worth (in a typical household). These data have been used to make assumptions on rates of saving needed to sustain consumption levels after retirement.

The SoFIE-Health website includes links to SoFIE-Health reports, peer-reviewed publications and other documents using the SoFIE data: http://www.wnmeds.ac.nz/academic/dph/research/HIRP/SoFIE/SofieIndex.html

What are the main strengths and weaknesses?

The main strength of the SoFIE study is the large representative sample size of the NZ population living in private residences. On the other hand, this is also a potential limitation of the study as people living in non-private residences, i.e. institutions are not included in the initial sample, inherently limiting the sample to a healthy population. Also, response rates are lower in people reporting fair or poor health status, leading to an overall healthier responding population compared with the general NZ population.

The attrition rates in SoFIE are similar to other (similarly representative) panel studies. In the HILDA survey 82% of the original sample was retained by Wave 3.[23] The original household response rate in SoFIE was 77%, which is better than shown in other household panel surveys. Therefore the overall response rate in HILDA, at Wave 3, taking into account the original 66% household response rate was 54%. The annual response rates are better than were projected at the outset of the study by Statistics New Zealand.

The SoFIE study collects detailed information on income, employment, education, family and household structure with good temporal resolution collecting detailed information about changes in these factors over the preceding year. This detailed information enables detailed investigations in these dynamics over time and temporal mapping of changes between variables, such as changes in employment linked to changes in income. SoFIE provides a unique platform for studying the impact of the current global economic recession on New Zealanders. However, the limited duration of the survey means that we will only be able to investigate initial impact of the recession and not the full extent of it.

The SoFIE-Health sub-study aims to bring these strengths in data and analysis to epidemiology and public health audiences. Much of the analyses and literature from panel studies has been conducted by economists and there is a need to close the gap between these disciplines. The linkage of the SoFIE population to hospitalization, cancer registration and mortality databases provides a unique opportunity to investigate the effects of previous and new events on social and economic factors. To our knowledge, no other panel study has linked through to hospitalization data.

The SoFIE study was initiated and is conducted through our National Statistical Agency (Statistics New Zealand), therefore the high household response rate may be due to the perception of the *Statistics Act 1975*, where responding is seen as 'compulsory'. However, as the study is conducted through Statistics New Zealand, there are strict confidentiality rules for using the data (upholding the public perception of all official statistics) and restricted access to the microdata (via Statistics New Zealand Data Laboratories in Auckland, Wellington and Christchurch, NZ) for researchers and external users. Unfortunately, the SoFIE study is also of limited duration of 8 years.

Due to limited capacity and time for each interview the questionnaires need to be confined to a manageable time. Therefore, regrettably, there are no data on physical activity and nutrition or other measures of psycho-social states. There is also limited information on children (aged <15 years) as they were not interviewed and information on them is collected from an adult OSM of the household. However, interviewing children adds much more complexity to any survey.

SoFIE and the SoFIE-Health sub-study are moving into exciting times, with more waves of data becoming available and collaboration between departments and government agencies leading to more opportunities for interdisciplinary research.

Statistics New Zealand Security Statement

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. The results in this study and any errors contained therein are those of the author, not Statistics New Zealand.

Funding

SoFIE-Health is primarily funded by the Health Research Council of New Zealand as part of the Health Inequalities Research Programme. Establishment funding was also received from the University of Otago, Accident Compensation Corporation of New Zealand (ACC) and the Alcohol Liquor Advisory Council (ALAC).

Conflict of interest: None declared.

References

[1] Taylor M (ed), Brice J, Buck N, Prentice-Lane E. *British Household Panel Survey User Manual Volume A: Introduction, Technical Report and Appendicies*. Colchester: University of Essex, 2007.

[2] Sacker A, Clarke P, Wiggins RD, Bartley M. Social dynamics of health inequalities: a growth curve analysis of aging and self assessed health in the British Household

Panel Survey 1991–2001. *J Epidemiol Community Health* 2005;**59:**495–501.

[3] Marmot M, Brunner E. Cohort Profile: The Whitehall II study. *Int J Epidemiol* 2005;**34:**251–56.

[4] Watson N, Wooden M. The HILDA survey four years on. *Aust Econ Rev* 2004;**37:**343–49.

[5] McGee R, Silva PA. *A Thousand New Zealand Children. Their Health and Development from Birth to Seven*. Dunedin: Medical Research Council of New Zealand, 1982.

[6] Fergusson D. The Christchurch Health and Development Study: An overview and some key findings. *Social Policy J New Zealand* 1998;**10:**154–76.

[7] Paterson J, Percival T, Schluter P *et al*. Cohort Profile: The Pacific Islands Families (PIF) Study. *Int J Epidemiol* 2008; **37:**273–79.

[8] Alpass F, Towers A, Stephens C, Fitzgerald E, Stevenson B, Davey J. Independence, well-being, and social participation in an aging population. *Ann NY Acad Sci* 2007;**1114:**241–50.

[9] Statistics New Zealand. *A Longitudinal Survey of Income, Employment and Family Dynamics. Feasibility Project Final Report*. Wellington: Statistics New Zealand, 2001.

[10] Statistics New Zealand. *Survey of Family, Income and Employment: Wave Four – September 2006*. Wellington: Statistics New Zealand, 2008.

[11] Ware JE, Snow KK, Kosinki MA, Gandek MS. *SF-36 Health Survey Manual and Interpretation Guide*. Lincoln RI: Quality Metric Incorporated, 2000.

[12] Kessler RC, Andrews G, Colpe LJ *et al*. Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychol Med* 2002;**32:** 959–76.

[13] Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. *J. Health Social Behav* 1983;**24:**385–96.

[14] Salmond C, Crampton P. *NZDep2001 Index of Deprivation*. Wellington: Department of Public Health: University of Otago, 2002.

[15] Shi L, Starfield B. The effect of primary care physician supply and income inequality on mortality among blacks and whites in US metropolitan areas. *Am J Public Health* 2001;**91:**1246–50.

[16] Starfield B. *Primary Care: Balancing Health Needs, Services and Technology*. New York: Oxford University Press, 1998.

[17] Statistics New Zealand. *2001 Census of Population and Dwellings*. Wellington: Statistics New Zealand, 2001.

[18] Ministry of Health. *A Portrait of Health. Key Results of the 2006/07 New Zealand Health Survey*. Wellington: Ministry of Health, 2008.

[19] Poland M, Cameron A, Wong K, Fletcher M. *Moving On. Changes in a year of family living arrangements*. Wellington: Families Commission, 2007.

[20] Cheung J. *Wealth Disparities in New Zealand*. Wellington: Statistics New Zealand, 2007.

[21] Schneider M. *The Distribution of Wealth*. Cheltenham, UK: Edward Elgar, 2004.

[22] Scobie G, Le T, Gibson J. *Housing in the Household Portfolio and Implications for Retirement Saving: Some Initial Findings from SOFIE*. Wellington: New Zealand Treasury, 2007.

[23] Headey B, Warren D. *Families, Incomes and Jobs, Volume 2: A Statistical Report on Waves 1 to 4 of the HILDA Survey*. Melbourne: Melbourne Institute of Applied Economic and Social Research, The University of Melbourne, 2007.