
Volume 74
Issue 7
April 2012
ISSN 0277-9536
ELSEVIER
SOCIAL SCIENCE & MEDICINE
an international journal
MOST CITED
Impact Factor
2.742
SOCIAL SCIENCE JOURNAL

Social Science & Medicine 74 (2012) 1117–1124

Contents lists available at SciVerse ScienceDirect

Social Science & Medicine

journal homepage: www.elsevier.com/locate/socscimed

Comparing self-rated health and self-assessed change in health in a longitudinal survey: Which is more valid?

Fiona Imlach Gunasekara*, Kristie Carter, Tony Blakely

Health Inequalities Research Programme, School of Medicine and Health Sciences, University of Otago, 23 Mein St, Newtown, Wellington, New Zealand

ARTICLE INFO

Article history:
Available online 2 February 2012

Keywords:
New Zealand
Self-rated health
Self-assessed change in health
Longitudinal survey
Longitudinal validity
Survey of family, income and employment

ABSTRACT

Self-rated health (SRH) is commonly used in longitudinal analyses as a repeated outcome measure. This assumes that computed changes in SRH over time truly represent within-individual changes in underlying health. The longitudinal validity of SRH, however, is threatened by ceiling effects (where people reporting the highest level of SRH cannot report subsequent improved health), insensitivity to small changes within SRH categories, reference group effects (where individuals assess their health changes relative to their peers) and stability in SRH even when change in underlying health is occurring. We assessed the longitudinal validity of SRH by comparing computed changes in SRH with a measure of self-assessed change in health (SACH). We used two waves of data (2003–2005) from the New Zealand longitudinal Survey of Family, Income and Employment (SoFIE). Computed change in SRH and SACH were compared directly and also in regression models using an objective measure of health outcome change (hospitalisations within the past year).

Computed change in SRH and SACH were not well correlated, consistent with ceiling and/or categorisation effects in SRH. In regression models, SACH was more strongly predictive of hospitalisation than computed change in SRH (worse SACH was associated with an increased odds of hospitalisation of 3.7 compared to 1.8 for decreased computed change in SRH). SACH may be affected by recall bias, but if SRH is used as a repeated outcome measure in longitudinal analyses, results may also be biased, if change in SRH does not occur in response to significant health events.

Background

Self-rated health (SRH), where individuals are asked to rate their health along a scale such as “excellent, very good, good, fair or poor”, is a commonly used health outcome in many disciplines. It is generally accepted as being valid, reliable, and predictive of mortality in a wide range of populations (Benyamini & Idler, 1999; Burstrom & Fredlund, 2001; DeSalvo, Bloser, Reynolds, He, & Muntner, 2006; Idler & Benyamini, 1997; Jylhä, 2009; Singh-Manoux, Gueguen, et al., 2007), with statements on its validity primarily based on single assessments of SRH compared to future mortality (Benítez-Silva & Ni, 2008; Idler & Benyamini, 1997; Sadana, Mathers, Lopez, Murray, & Iburg, 2000). Recent challenges to the reliability of SRH have suggested that socioeconomic position may modify the SRH-mortality relationship (Dowd & Zajacova, 2007; Huisman, van Lenthe, & Mackenbach, 2007; Quesnel Vallee, 2007; Singh-Manoux, Dugravot, et al., 2007). However, this effect is not always found (Burstrom & Fredlund, 2001; van Doorslaer & Gerdtham, 2003) and it has been argued that, overall, SRH can still be considered a reliable tool to assess social inequalities in health (Subramanian & Ertel, 2008a, 2008b).

However, the validity of SRH in a longitudinal context has not been so closely scrutinised. The longitudinal validity of SRH is not concerned with whether a cross-sectional measure of SRH predicts future mortality but whether computed changes in SRH, collected over several time periods in a longitudinal survey, measure true changes in health. By “computed changes” we mean either where the analyst computes changes in the SRH question (asked of the same people over time) by calculating the difference in responses to SRH from one study period to the next or where repeated SRH responses are analysed in a longitudinal model so that estimates are interpreted as a function of change in SRH. The main aim of first differenced (and fixed effects models, which subtract the individual mean from all time-varying variables in the model) is to eliminate the individual heterogeneity or unmeasured time-invariant confounding and give an estimate of change in the outcome (Imlach Gunasekara, Carter, & Blakely, 2008; Wooldridge, 2002). However, models that rely on change in the outcome are sensitive to measurement error, which is a well-recognised problem (Cronbach

* Corresponding author. Tel.: +64 43855541.
E-mail address: fiona.imlach-gunasekara@otago.ac.nz (F.I. Gunasekara).

0277-9536/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.socscimed.2011.11.038

& Furby, 1970; Hsiao, 2003), but one that is often overlooked when these methods are applied to the outcome of SRH. A recent systematic review (Imlach Gunasekara, Carter, & Blakely, 2011) which included thirteen studies investigating the longitudinal association between income and SRH, found that many used first differenced, fixed effects or related methods, but only two discussed measurement bias as a potential cause of attenuation bias.

The problem of measurement error in SRH in longitudinal analysis is also known as longitudinal validity (or responsiveness) (Husted, Cook, Farewell, & Gladman, 2000; Terwee, Dekker, Wiersinga, Prummel, & Bossuyt, 2003) and raises several issues. The first is whether SRH, as a highly skewed variable, can accurately measure change over time (Seymour, McNamee, Scott, & Tinelli, 2010). The most obvious example of this is the ceiling effect in SRH, where a large proportion of people report the highest level of health (e.g. excellent) (Bech, Olsen, Kjoller, & Rasmussen, 2003; Bowling & Windsor, 2008). In New Zealand, 20% of individuals typically report excellent health in general health surveys (Gerritsen, Stefanogiannis, & Galloway, 2008) but reporting excellent health can be as high as 38% (Carter, Cronin, Blakely, Hayward, & Richardson, 2010). However, individuals who rate their health as excellent but feel their health has improved at a later time cannot indicate this improvement using the SRH question. Also, the categorisation of SRH means that information about "true" changes in health may be lost, so that "no computed change in SRH" does not necessarily signify an absence of actual, and meaningful, change in health over time (Benítez-Silva & Ni, 2008). For example, a person may experience an improvement in their "true" underlying health over two time periods but still remain within the SRH category of "poor".

The second problem of longitudinal validity is whether changes in SRH accurately correspond to changes in "true" underlying health. This can be measured by testing how well changes in SRH correlate to changes in more objective or external measures of health (Husted et al., 2000). This begs the question of what is "true" health and how might this be measured. Although most SRH validation studies use mortality as the gold-standard comparison health measure, this limits the understanding of "true" health to an absence of life-threatening disease (Quesnel Vallee, 2007). However, SRH measures more than the absence of disease, and studies have shown that mental health, symptom scores and health behaviours are all important determinants of SRH (Benyamini, Idler, Leventhal, & Leventhal, 2000; Contoyannis & Jones, 2004; Singh-Manoux et al., 2006). For longitudinal analyses, the relationship between change in SRH and "true" health is of more relevance, but *changes* in SRH are also predictive of more than mortality (Han et al., 2005; Nielsen et al., 2009). Increased illness and comorbidity are associated with worsened SRH (Heller, Ahern, Pringle, & Brown, 2009; Manor, Matthews, & Power, 2001; Orfila, Ferrer, Lamarca, & Alonso, 2000; Rodin & McAvay, 1992) and studies in the elderly have found that declines in physical, mental, psychosocial, and/or cognitive functioning are predictive of decreases in SRH (Benyamini et al., 2000; Leinonen, Heikkinen, & Jylhä, 2001; Rodin & McAvay, 1992). These studies support the notion that changes in SRH reflect changes in "true" health that extend beyond life-threatening disease and mortality to reflect a host of risk factors and quality of life issues, at least in the elderly and less healthy populations. This still leaves the problem of accurately measuring "true" health, as many other supposedly objective measures may also rely on self-reports – of diseases, functional limitations and the Health Utility Index (Dowd & Zajacova, 2010).

Although change in SRH is often demonstrated in response to change in "true" health, this is not always the case. Studies have compared SRH to other assessments of health over time and found that although change in SRH does occur, there is also considerable stability in SRH even when change occurs in other assessments of health (Bailis, Segall, & Chipperfield, 2003; Boardman, 2006; Perruccio, Badley, Hogg-Johnson, & Davis, 2010). This stability has also been observed in the health economics literature (Contoyannis, Jones, & Rice, 2004), where considerable state dependence in SRH is seen even after controlling for observed and unobserved confounding. Even when changes in "true" health occur, previously reported SRH may be the strongest predictor of current SRH (Bailis et al., 2003; Perruccio et al., 2010).

Self-rated health may be affected by several other problems of longitudinal validity, where changes in SRH may not correspond to changes in "true" health. Some people may have difficulty distinguishing between categories of SRH, leading to "grey areas" between categories, so that a person with "true" fair health may report "fair" health at one time period, and "poor" health at another, introducing error (Benítez-Silva & Ni, 2008). Response category cut-point shifts are another problem, where different individuals have a different comprehension of how the SRH categories relate to true underlying health (Benítez-Silva & Ni, 2008; Lindeboom & van Doorslaer, 2004). However, as long as individuals have a constant understanding of how their own SRH and underlying health are related, this may not always affect longitudinal validity, for example, if data are analysed using fixed effects models, as these account for individual heterogeneity by utilising only changes within individuals over time (Allison, 2005). Cut-point shift is most problematic in cross-country comparisons, where calibration tests (with objective measures) (Lindeboom & van Doorslaer, 2004) or vignettes are useful to make subjective measures like SRH comparable (Murray, Tandon, Salomon, & Mathers, 2001; Salomon, Tandon, & Murray, 2004).

A final issue is that of reference group effects (or scale of reference bias), where individuals adjust their reported health status to what they perceive to be appropriate for their age, situation and stage of life (even if an explicit reference group is not given), compared to their peers (Groot, 2003). This leads to bias in subjective health questions as some may report good health when their "true" health is actually poor when measured more objectively, because they perceive they are doing better than others around them, or vice versa (Groot, 2003).

Given these limitations of SRH, it is surprising that lack of longitudinal validity in SRH is often overlooked as a source of bias in longitudinal analyses. An alternative way to assess change in health is to directly ask the interviewee whether they think their health has got better or worse over a period of time (for example, as asked using the health transition question from the SF-36: "How do you rate your health now compared to twelve months ago?" with possible responses of: "Much better, a little better, same, a little worse or much worse"). This question is hereafter referred to as "self-assessed change in health" (SACH) (Ware, Kosinki, & Gandek, 2005). By way of contrast, the SACH question may be less prone to this bias, potentially being less affected by ceiling effects, loss of information from categorisation and the issue of "grey areas". The SF-36 (including the SACH question) has been found to be a valid measure of change in health over time (Hemingway, Stafford, Stansfeld, Shipley, & Marmot, 1997). Changes in the SACH question have also been found to correspond to changes in health, although mostly in comparison to changes in other self-reported health questions (Frijters, Geishecker, Haisken-DeNew, & Shields, 2004); assessment of responsiveness is limited (Inwood & Roberts, 2010). Several studies have compared change in SRH and SACH, including a small longitudinal study of elderly people in Finland, where more participants reported deterioration in health from a SACH question than was evident from computed changes in SRH (at two points in time) (Leinonen, Heikkinen, & Jylhä, 1998; Leinonen et al., 2001). Participants appeared to adjust for "age-appropriate" health problems when answering the SRH question but not the SACH question (reference group effects). In an Australian study of older adults that directly compared SACH to change in SRH, using seven years of data, more people rated their health the same from the SACH question, even though computed SRH had worsened (Sargent-Cox, Anstey, & Luszcz, 2010). This was also theorised by the authors as due to reference group effects in SRH (as the same pattern of change occurred in SRH with and without including an age-comparison in the question) but also possibly ceiling effects in SACH (in that older adults perceived that their health cannot or does not get any worse and so rated their health the same). In other circumstances, the SACH question would be expected to be less subject to ceiling effects.

Another study compared changes in SRH to SACH using longitudinal fixed effects regressions on self-reported longevity in an older population from the Health and Retirement Study (Benítez-Silva & Ni, 2008). This found SACH a better reflection of "true" health dynamics and less affected by confounding bias than change in SRH.

Although these studies provide somewhat conflicting results on the usefulness of change in SRH and SACH as measures of health change, they have limitations as they apply primarily to older individuals and do not compare the change variables to an objective measure of health change to assess which might be the more valid measure of change over time ((Benítez-Silva & Ni, 2008) uses a subjective comparison measure). Our study investigates these variables using a survey including adults of all ages and comparing them against an objective measure of change in health. We investigate whether SACH and computed changes in SRH are comparable measures of health change by comparing the two variables in cross-tabulations. We also regress the two measures of change in SRH and SACH on the outcome of experiencing a hospitalisation in the previous year (in separate regressions and together in the same model). This was chosen because it is an objective measure of health change, so it is not affected by biases that may be present in subjective, self-reported data, such as self-reports of chronic conditions, which may be misreported and lead to attenuation bias (Baker, Stabile, & Deri, 2004; Carter, Barber, & Shaw, 2010) but has the advantage over mortality, which does not measure change in health. Hospitalisations also have the benefit of signalling a significant health event but include diseases that are not life-threatening and impact solely on quality of life. From the literature, we know that change in SRH also reflects change in this type of health state, making hospitalisation an appropriate outcome to use. As a sensitivity analysis, we also repeat the final model using cancer registration as the health outcome, as an alternative objective measure of health change. Hospitalisations (and cancer registrations) are used as proxy measures for "true" or unobserved change in health state, that subjective measures such as SRH seek to reflect. The major advantage of this analysis is having an objective measure of health change, from an independent data source (linked hospital records), to which we can compare the subjective self-reported health change variables.

Methods

Data

This study used two waves (wave two from October 2003 to September 2004; wave three from October 2004 to September 2005; data version 6) of the New Zealand (NZ) longitudinal household panel Survey of Families, Income and Employment (SoFIE). The population covered by SoFIE was the usually resident population of NZ living in private dwellings (excluding people living in institutions or in establishments such as boarding houses and rest homes). The initial SoFIE sample comprised of approximately 11 500 responding private households and 22 165 adults (aged 15 years and older) sampled within them (Fig. 1) (K.N. Carter et al., 2010). There were 20 005 adult original sample members in wave two and 18 950 in wave three. In wave three, a module of health questions (the SoFIE-Health sub-study) was asked of all adults. Data were collected via face-to-face computer-assisted interviews. In addition, as part of the SoFIE-Health sub-study, participants were asked for consent to link their survey data to health records, including the national minimum dataset, which is an electronic collection of public and private hospital discharge data. Approximately 80% of wave three participants consented to this linkage (Carter, Shaw, Hayward, & Blakely, 2010; K.N Carter et al., 2010). Ethical approval for the linkage was given by the Multi-region Ethics Committee, and administration of SoFIE was governed by the Statistics Act 1975, as part of the work programme of Statistics New Zealand (the national statistics agency).

Measures

Self-rated health was asked annually of all participants, using the question: "In general would you say your health is: excellent, very good, good, fair or poor?" To compute a change variable, the difference between SRH in wave two and three was calculated. All individuals whose SRH increased between waves two and three (e.g. moved from good to excellent, a positive score) were coded as "Increased health"; if SRH decreased (e.g. moved from excellent to good, a negative score) they were coded as "Decreased health"; or if SRH remained the same, this was coded as "No change". The SRH question was assumed to reflect the health status of individuals at the time of each interview so the change between the questions reflected the difference in health over the last 12 months (from wave two to three).

In wave three, as part of the Short Form-36 (SF-36) questionnaire (Ware et al., 2005), an alternative question on self-assessed change in health was asked: "How do you rate your health now compared to twelve months ago?" with responses: "Much better", "A little better", "About the same", "A little worse" and "Much worse". Since this question asks participants to estimate their change in health status from a year before the time of interview to the time of interview (at wave three), these responses were compared to the computed change in SRH from wave two to wave three. To make the two variables more comparable, the responses "Much better" or "A little better" were combined into one

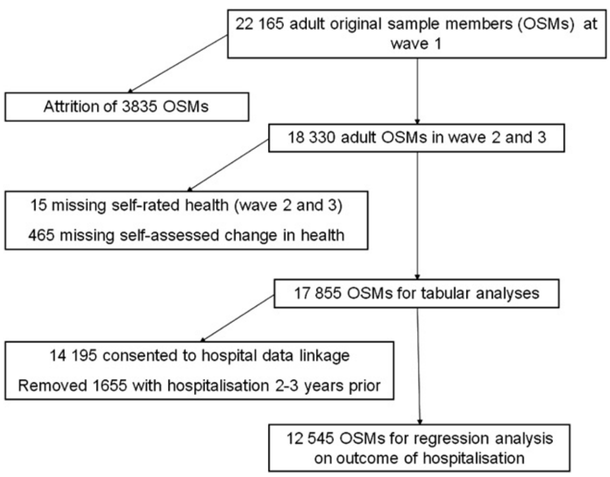

Fig. 1. Dataflow from wave one data collection to data analysis.

category: "Better". "A little worse" and "Much worse" were combined into the category: "Worse".

The experience of an overnight hospitalisation over the year prior to the interview in wave three was used as an objective measure of a change in health in regression models. Only individuals who consented to the linkage to health records were included in this analysis (see Fig. 1). To ensure that these hospitalisations reflected a relatively new or unexpected health change or health 'shock', individuals who had experienced a hospitalisation two or three years before the interview in wave three were excluded (to exclude those who may be experiencing multiple hospitalisations as part of a chronic disease process) as were hospitalisations due to childbirth, leaving 665 individuals with hospitalisation (overnight stay) events. Covariates included in the regression models were age, sex, ethnicity, education (highest qualification attained at wave three), annual household income, equivalised to adjust for household size and composition using the NZ-specific Jensen Index (Jensen, 1988), labour force status, marital status, family structure and NZ Deprivation Index 2001 (Salmond & Crampton, 2002) (a measure of small area deprivation, dividing areas into quintiles of deprivation). As a sensitivity test, an alternative 'health shock' outcome was also modelled – that of cancer registrations over the past three years (270 were recorded).

Statistical analyses

Tabular analyses were conducted on all eligible adults who participated in wave two and three of SoFIE and responded to the SRH and SACH questions ($N = 17\,855$) (see Fig. 1). To test whether the two variables related to a "true" change in health, we used logistic regression models with the outcome of hospitalisation in the year between interviews (or the presence of a cancer registration for the sensitivity analysis). All analyses were performed using SAS version 8.2.

Results

A descriptive summary of the sample by presence of a hospitalisation event is presented in Table 1. The mean age of those who had a hospitalisation was greater (54 years) compared to those who did not (45 years) and the mean household income lower ($43 100 compared to $54 900). Of note is that 13.1% of those who reported worse SACH experienced a hospitalisation compared to 7.5% of those with a decrease in SRH, suggesting that SACH may predict hospitalisation better than change in SRH. Additional demographic and socioeconomic characteristics by the two health change variables are given in Supplementary tables (Table 6 and Table 7).

Cross-tabulations of the two health change variables are in Table 2. Overall, more people had a decrease in computed SRH ($N = 4625$) than reported worse health compared to 12 months ago from the SACH question ($N = 2790$). The italicised diagonal highlights the imperfect concordance between the questions at each category. For example, it was expected that the majority of people with a computed increase in SRH would also report better health, but only 29.9% did so. Similarly, only 26.7% of those with a computed decrease in SRH reported worse health. The comparability between the two health change variables was poor (Spearman correlation coefficient $r = -0.16$, $p < 0.01$; kappa = 0.01).

Although approximately two thirds of those who did not change SRH also rated their health as the same on the SACH question (the shaded row in Table 2) of the remaining third, more people reported better SACH, which could be due to loss of information on "true" health change due to categorisation of SRH or ceiling effects in computed change in SRH. Ceiling effects were investigated further by exploring SACH within categories of SRH at wave two

Table 1
Characteristics of the analysis sample by hospitalisation.[a]

Characteristics	Hospitalisation		
	No (*N*/row %)[b]	Yes (*N*/row %)[b]	Total (*N*)
Self-assessed change in health (12 months prior to wave 3)			
Worse	1620 (86.9)	245 (13.1)	1865
Same	7700 (96.7)	260 (3.3)	7960
Better	2560 (93.9)	160 (5.9)	2725
Change in self-rated health (from wave 2–3)			
Decrease	2960 (92.5)	240 (7.5)	3200
No change	6830 (95.7)	310 (4.3)	7140
Increase	2090 (94.6)	115 (5.2)	2210
Sex			
Male	5440 (94.4)	325 (5.6)	5765
Female	6440 (94.9)	340 (5.0)	6785
Ethnicity			
European/other	9715 (94.5)	570 (5.5)	10 285
Māori	1235 (95.7)	55 (4.3)	1290
Pacific	410 (95.3)	20 (4.7)	430
Asian	525 (97.2)	15 (2.8)	540
Highest education level (at wave 3)			
No qualification	2510 (93.0)	190 (7.0)	2700
School qualification	3250 (95.2)	165 (4.8)	3415
Post school qualification	4235 (94.5)	240 (5.4)	4480
Degree or higher	1885 (96.7)	70 (3.6)	1950
Marital status (wave 2)			
Never married	3625 (96.5)	130 (3.5)	3755
Divorced, widowed, separated	1910 (92.9)	150 (7.3)	2055
Married	6345 (94.3)	385 (5.7)	6730
Family structure (wave 2)			
Couple only	3315 (93.4)	235 (6.6)	3550
Sole parent	1135 (95.0)	65 (5.4)	1195
Not in family	2395 (94.1)	150 (5.9)	2545
Couple with children	5035 (95.9)	215 (4.1)	5250
Labour force status (wave 2)			
Not employed, not looking for work	3540 (91.7)	320 (8.3)	3860
Not employed, looking for work	235 (95.9)	10 (4.1)	245
Employed	8100 (96.1)	330 (3.9)	8430
New Zealand deprivation index level (wave 2)			
NZDepQ1 (least deprived)	2520 (95.8)	105 (4.0)	2630
NZDepQ2	2480 (95.0)	130 (5.0)	2610
NZDepQ3	2165 (94.7)	120 (5.3)	2285
NZDepQ4	2560 (94.1)	160 (5.9)	2720
NZDepQ5 (most deprived)	2155 (93.5)	145 (6.3)	2305
Total	**11 880 (94.7)**	**665 (5.3)**	**12 545**

[a] All counts in tables were rounded to the nearest multiple of five and cells with counts less than ten were imputed with the value ten, as per Statistics New Zealand protocol.
[b] Row% is the row percentage, dividing each cell by the row total.

and three (Table 3). Of particular interest were the 6790 people who reported excellent health at wave two. Nearly two thirds (4300) also reported excellent health in wave three, but 13% (875) of these reported better health from the SACH question. A small percentage (2%) reported worse health over the past 12 months, but excellent health at both waves.

Lack of responsiveness in change in SRH could affect at least 18.7% of observations (2125 and 1215 individuals with no computed change in SRH but change in SACH). This may lead to dampened estimates of "true" health changes when using computed change in SRH as the outcome variable.

To test how the two health change variables estimated a "true" health change, both variables were compared as predictors in separate logistic regression models where the outcome was an objective measure of health change or a health shock – hospitalisation between waves two and three (Table 4). These models show that

Table 2
Comparison between computed change in self-rated health from wave two to three and self-assessed change in health (alternative health change question).

Computed change in self-rated health	Self-assessed change in health			
	Better (col %/row %)[a]	Same (col %/row %)	Worse (col %/row %)	Total (col%)
Increase	*980 (24.8/29.9)*	1950 (17.6/59.5)	345 (12.4/10.5)	3280 (18.4)
No change	2125 (53.7/21.4)	*6610 (59.5/66.4)*	1215 (43.5/12.2)	9950 (55.7)
Decrease	850 (21.5/18.4)	2545 (22.9/55.0)	*1235 (44.3/26.7)*	4625 (25.9)
Total (row%)	**3955 (22.2)**	**11 105 (62.2)**	**2790 (15.6)**	**17 855**

[a] Col% is the column percentage, dividing each cell by the column total; row% is the row percentage, dividing each cell by the row total.

reporting worse health (compared to reporting the same health) by the SACH variable was more strongly associated with hospitalisation than a decrease in computed SRH (compared to no change in computed SRH). In addition, reporting better health (compared to the same health) by the SACH question was positively associated with the probability of hospitalisation, but this effect was much less apparent (and not statistically significant) for increase in computed SRH. These differences persisted when age, sex, ethnicity and other characteristics were controlled for in the analysis (although the strength of association reduced). The model using SACH also had a statistically better fit (using the Akaike Information Criterion – AIC).

The final model included both health change variables in the same model to test which was most strongly associated with the outcome (Table 4). The association between decreased SRH and hospitalisation reduced by a third, but the SACH association was little changed.

To see whether the ceiling effect was affecting the estimates of change in SRH, we also re-ran the final fully-adjusted model (with both health change variables included) on a sample in which people in excellent health at wave two were excluded (Table 5). The estimates for change in SRH from this model were virtually the same as for the full analysis and SACH remained the stronger predictor of hospitalisation events.

As a sensitivity analysis, the final model, including both health change variables, was repeated for another health shock outcome – that of cancer registrations over the past three years. This found a similar pattern as for hospitalisations, although the stronger association with SACH was even more apparent. (Results in supplementary Table 8.)

Discussion

These analyses compared different health change variables, to test their validity in a longitudinal setting. In the SoFIE dataset,

Table 3
Three-way comparison between self-rated health at wave two, self-assessed change in health and self-rated health at wave three.

Self-rated health (wave 2)	Self-assessed change in health	Self-rated health (wave 3)					Total
		Excellent	Very good	Good	Fair	Poor	
Excellent	Better	875	380	105	15	10	1380
	Same	3275	1275	305	15	10	4870
	Worse	150	195	140	40	10	545
	Total	4300	1850	550	70	10	6790
Very good	Better	310	740	215	30	10	1300
	Same	850	2075	720	45	10	3695
	Worse	45	270	290	125	20	750
	Total	1205	3085	1230	200	30	5745
Good	Better	120	295	385	80	10	880
	Same	190	615	1025	140	10	1980
	Worse	15	115	420	245	60	855
	Total	325	1025	1830	465	75	3720
Fair	Better	15	45	125	105	10	305
	Same	10	45	190	200	25	480
	Worse	10	15	90	270	100	475
	Total	30	110	405	575	140	1260
Poor	Better	10	10	30	25	20	95
	Same	10	10	10	30	35	80
	Worse	0	10	10	50	105	165
	Total	10	15	50	105	160	340
Total	Better	1330	1470	860	255	60	3960
	Same	4335	4020	2250	430	90	11 105
	Worse	220	605	950	730	295	2790
Grand total		**5870**	**6085**	**4065**	**1415**	**415**	**17 855**

Table 4
Logistic regression models with outcome of hospitalisations in waves 2–3.

	Hospitalisation (event in past year = 1; no event = 0)	
	Baseline[a] $N = 12\,545$ Odds ratio (95% confidence interval)	Multivariate[b] $N = 12\,530$ Odds ratio (95% confidence interval)
Models including only the health change variable self-assessed change in health		
Self-assessed change in health (wave three)		
Same health	1	1
Worse health	4.5 (3.7–5.4)	3.7 (3.1–4.5)
Better health	1.9 (1.5–2.3)	2.1 (1.7–2.6)
	AIC = 4965	AIC = 4826
Models including only the health change variable computed change in self-rated health		
Computed change in self-rated health (wave two to wave three)		
No change in health	1	1
Decreased health	1.8 (1.5–2.1)	1.8 (1.5–2.1)
Increased health	1.2 (1.0–1.5)	1.1 (0.9–1.4)
	AIC = 5166	AIC = 4974
Models including simultaneously both health change variables self-assessed change in health and computed change in self-rated health		
	Baseline[c] $N = 12\,545$ Odds ratio (95% confidence interval)	Multivariate[b, c] $N = 12\,530$ Odds ratio (95% confidence interval)
Self-assessed change in health (wave three)		
	1	1
Worse health	4.2 (3.5–5.0)	3.5 (2.9–4.2)
Better health	1.9 (1.5–2.3)	2.1 (1.7–2.6)
Computed change in self-rated health		
No change in health	1	1
Decreased health	1.4 (1.2–1.7)	1.5 (1.2–1.8)
Increased health	1.2 (1.0–1.5)	1.1 (0.9–1.4)
	AIC = 4956	AIC = 4813

AIC = Akaike Information Criterion.
[a] Includes each health change variable separately and also includes wave (time).
[b] Also includes age, sex, ethnicity, marital status, family structure, labour force status, income, education and NZ deprivation index.
[c] Includes both health change variables and wave.

Table 5
Logistic regression of hospitalisations in waves 2–3 excluding individuals with excellent self-rated health at wave two.

	Multivariate[a] $N = 7530$ Odds ratio (95% confidence interval)
Self-assessed change in health (wave three)	
	1
Worse health	2.9 (2.3–3.7)
Better health	2.3 (1.8–3.0)
Computed change in self-rated health	
No change in health	1
Decreased health	1.6 (1.3–2.0)
Increased health	1.0 (0.8–1.3)
	AIC = 3366

AIC = Akaike Information Criterion.
[a] Includes age, sex, ethnicity, marital status, family structure, labour force status, income, education, NZ deprivation index, both health change variables and wave.

ceiling effects in SRH are a particular concern, as over a third of individuals report excellent health (K.N. Carter et al., 2010). In cross-tabulations of computed change in SRH with SACH, the finding that a decline in computed SRH was more common than an increase, but reporting better health from the SACH question was more common, could be due to ceiling effects in SRH. In the three-way cross-classification of SRH at wave two with SACH and SRH at wave three, a ceiling effect was clearly detected. The 2% of people who had discordant reports, reporting both worse health (on the SACH question) but excellent health at both waves (on the SRH question), could represent measurement error occurring in subjective health reports or may be an issue with the categorisation of SRH – that is, some people may feel their health has declined, but is still excellent.

We assumed that computed change in SRH and SACH would be comparable measures of a change in underlying health state and highly correlated. However, we found a relatively poor correlation between these two measures of changing health and different results when tested against an objective measure of health change (hospitalisation). The regression analyses attempted to circumvent the ceiling effect in computed change in SRH, since the main association was anticipated to be between decreasing SRH and hospitalisation (overnight stay in hospital), and few people report poor health (the lowest category of SRH), creating little impediment to decreasing SRH over time (i.e. there is not a major floor effect). The regression models did show an association between decreasing SRH and hospitalisation but the association was stronger between reporting worse SACH and hospitalisation, suggesting that SACH is a better predictor of health deterioration. Our research supports that of Benítez-Silva, which found that SACH may perform better than change in SRH in longitudinal analyses, when using a subjective health outcome as a comparison measure (Benítez-Silva & Ni, 2008).

However, the regression models also found that reporting better health (SACH question), but not an increase in computed SRH, was associated with a hospitalisation event. This is plausible, as hospitalisation can denote events that improve health as well as being a marker for health decline. For example, health can improve after hospitalisation for elective surgeries such as a hip replacement that relieves pain and restores mobility. The "same/no change" in health category emerged as the best health state, which has been found in studies of health dynamics using Markov models, where individuals with stable (good) health were the most healthy (McDonough, Worts, & Sacker, 2010; Sacker, Wiggins, Bartley, & McDonough, 2007). We did not attempt to categorise hospitalisations by severity or disease type as such categorisations can be arbitrary and would have reduced the power of our analyses to detect differences between groups.

Comparing changes in SRH to changes in objective health status has previously been investigated in analyses that used change in SRH and objective health measures as exposures, with the outcome being labour force decisions (Au, Crossley, & Schellhorn, 2005). These found that change in SRH was affected by significant measurement error when compared to analyses using health measures that were more objective or adjusted for measurement error, and this measurement error in SRH caused attenuation of the estimate of impact of health (SRH) change on labour market outcomes. Subjective quality of life measures have also been found to be less strongly associated with each other than expected, due to both unmeasured differences across individuals and ceiling and floor effects (Seymour et al., 2010).

The findings from the regression models in this study raise the possibility of significant measurement error in computed change in SRH due to poor longitudinal validity, either due to the limited ability of the SRH question to actually change (e.g. due to the ceiling effect) and/or to respond to "true" health change (e.g. due to state dependence in SRH) (Perruccio et al., 2010). However, the stronger association between hospitalisations and SACH may also be due to recall bias in SACH. This is not a problem with change in SRH, as the baseline SRH question is asked prior to any hospitalisation events, but SACH is asked after the event and may be directly affected. Different results from our models/cross-tabulations may be due to differences in the two health questions, which make them less comparable than they initially appear. Firstly, the health questions vary in how they refer to time – the SACH question specifically asks about health over the past 12 months whereas the SRH question does not include a time frame, leaving the question open for participants to interpret. If participants report their current SRH as being over the entire year of the survey, rather than the last few days or weeks, the comparability of change in SRH and the SACH question may be reduced. Secondly, the SACH question is asked later in the health module, so participants may be more stimulated to reflect on their health during the other questions. Self-rated health is asked earlier during the data collection on demographics. Thirdly, the SACH question, comparing health to twelve months ago, gives more freedom for people who want to report an improved health state whereas computed change in SRH has more potential for a ceiling effect. The final models, which included both health change variables, and found estimates that were fairly similar to the separate regressions, suggest that change in SRH and SACH may predict different components of health change.

The limitations of differencing are well known and the possibility of compounding measurement error with this method is always a concern. Both first differenced and fixed effects models, which rely on change in the outcome to produce estimates, are subject to this concern (and over two waves, a fixed effects model and first differenced model will give the same estimates – beyond this, results may vary depending on the presence of serial correlation of the errors)(Wooldridge, 2006). Alternative methods, which are less subject to measurement bias, are often used, such as residualised change analysis (Cronbach & Furby, 1970), multilevel and survival models (Singer & Willett, 2003), and structural equation models (Kline, 2005). But these do not all offer the rigorous control for time-invariant unmeasured confounding that is considered the main advantage of the fixed effects-type models. However, this advantage of fixed effects models, when the outcome of SRH is used, may be more than offset by the problem of measurement error, an issue which has been inadequately acknowledged in some current research (Imlach Gunasekara et al., 2011).

Our analyses may have been affected by selection bias if people who died (approximately 150 individuals from wave two to three) would have reported much worse SRH or SACH, and had a recent hospitalisation (or cancer registration). The loss of these individuals means the association of hospitalisation with deteriorations in both SRH and SACH may actually be stronger. In summary, our analyses cast some doubt over the longitudinal validity of SRH when change in SRH is compared against SACH using the outcome of an objective measure of health change. More investigation into the longitudinal validity of subjective health measures like SRH in a variety of contexts is needed. Researchers who use SRH should be aware of its potential limitations and weaknesses as a repeated health outcome measure, although it is convenient, flexible and widely available in panel data, and consider using alternative measures, where available, in sensitivity analyses.

Acknowledgements

This work, and the SoFIE-Health sub-study (reference 08/048), was supported by the Health Research Council of New Zealand, and was initiated as part of a PhD thesis within the Health Inequalities Research Programme, University of Otago. Publication was supported by the University of Otago Research Committee, by means of the University of Otago Postgraduate Publishing Bursary. 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 authors, not Statistics New Zealand.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.socscimed.2011.11.038.

References

Allison, P. D. (2005). *Fixed effects regression analysis for longitudinal data using SAS*. Cary, North Carolina: SAS Institute Inc.
Au, D. W. H., Crossley, T. F., & Schellhorn, M. (2005). The effect of health changes and long-term health on the work activity of older Canadians. *Health Economics, 14*(10), 999–1018.
Bailis, D. S., Segall, A., & Chipperfield, J. G. (2003). Two views of self-rated general health status. *Social Science & Medicine, 56*(2), 203–217.
Baker, M., Stabile, M., & Deri, C. (2004). What do self-reported, objective, measures of health measure. *Journal of Human Resources, 39*(4), 1067–1093.
Bech, P., Olsen, L. R., Kjoller, M., & Rasmussen, N. K. (2003). Measuring well-being rather than the absence of distress symptoms: a comparison of the SF-36 mental health subscale and the WHO-five well-being scale. *International Journal of Methods in Psychiatric Research, 12*(2), 85–91.
Benítez-Silva, H., & Ni, H. (2008). Health status and health dynamics in an empirical model of expected longevity. *Journal of Health Economics, 27*(3), 564–584.
Benyamini, Y., & Idler, E. L. (1999). Community studies reporting association between self-rated health and mortality: additional studies, 1995 to 1998. *Research on Aging, 21*(3), 392–401.
Benyamini, Y., Idler, E. L., Leventhal, H., & Leventhal, E. A. (2000). Positive affect and function as influences on self-assessments of health: expanding our view beyond illness and disability. *Journals of Gerontology Series B-Psychological Sciences & Social Sciences, 55*(2), 107–116.
Boardman, J. D. (2006). Self-rated health among U.S. adolescents. *The Journal of Adolescent Health: Official Publication of the Society for Adolescent Medicine, 38*(4), 401–408.
Bowling, A., & Windsor, J. (2008). The effects of question order and response-choice on self-rated health status in the English longitudinal study of ageing (ELSA). *Journal of Epidemiology & Community Health, 62*(1), 81–85.
Burstrom, B., & Fredlund, P. (2001). Self rated health: is it as good a predictor of subsequent mortality among adults in lower as well as in higher social classes? *Journal of Epidemiology and Community Health, 55*(11), 836–840.
Carter, K., Barber, A., & Shaw, C. (2010). How does self-reported history of stroke compare to hospitalization data in a population-based survey in New Zealand? *Stroke, 41*, 2678–2680.
Carter, K., Shaw, C., Hayward, M., & Blakely, T. (2010). Understanding the determinants of consent for linkage of administrative health data with a longitudinal survey. *Ko-tuitui: New Zealand Journal of Social Science Online, 5*(2), 53–60.
Carter, K. N., Cronin, M., Blakely, T., Hayward, M., & Richardson, K. (2010). Cohort profile: survey of families, income and employment (SoFIE) and health extension (SoFIE-health). *International Journal of Epidemiology, 39*(3), 653–659.
Contoyannis, P., Jones, A. M., & Rice, N. (2004). The dynamics of health in the British household panel survey. *Journal of Applied Econometrics, 19*(4), 473–503.
Contoyannis, P., & Jones, A. M. (2004). Socio-economic status, health and lifestyle. *Journal of Health Economics, 23*(5), 965–995.
Cronbach, L. J., & Furby, L. (July 1970). How we should measure "change": or should we? *Psychological Bulletin, 74*(1), 68–80.
DeSalvo, K. B., Bloser, N., Reynolds, K., He, J., & Muntner, P. (2006). Mortality prediction with a single general self-rated health question. *Journal of General Internal Medicine, 21*(3), 267–275.
Dowd, J. B., & Zajacova, A. (2007). Does the predictive power of self-rated health for subsequent mortality risk vary by socioeconomic status in the US? *International Journal of Epidemiology, 36*(6), 1214–1221.
Dowd, J. B., & Zajacova, A. (2010). Does self-rated health mean the same thing across socioeconomic groups? Evidence from biomarker data. *Annals of Epidemiology, 20*(10), 743–749.
Frijters, P., Geishecker, I., Haisken-DeNew, J. P., & Shields, M. A. (2004). *Income and life satisfaction in post-transition Russia: A new empirical methodology for panel data* (pp. 1–25). Working paper 186C. School of Economics and Finance, Queensland University of Technology.
Gerritsen, S., Stefanogiannis, N., & Galloway, Y. (2008). *A portrait of health: Key results from the 2006/07 New Zealand health survey*. Wellington: Ministry of Health.
Groot, W. (2003). Scale of reference bias and the evolution of health. *European Journal of Health Economics, 4*(3), 176–183.
Han, B., Phillips, C., Ferrucci, L., Bandeen-Roche, K., et al. (2005). Change in self-rated health and mortality among community-dwelling disabled older women. *The Gerontologist, 45*(2), 216–221.
Heller, D. A., Ahern, F. M., Pringle, K. E., & Brown, T. V. (2009). Among older adults, the responsiveness of self-rated health to changes in Charlson comorbidity was moderated by age and baseline comorbidity. *Journal of Clinical Epidemiology, 62*(2), 177–187.
Hemingway, H., Stafford, M., Stansfeld, S., Shipley, M., & Marmot, M. (1997). Is the SF-36 a valid measure of change in population health? Results from the Whitehall II study. *BMJ, 315*(7118), 1273–1279.
Hsiao, C. (2003). *Analysis of panel data*. Cambridge: Cambridge University Press.
Huisman, M., van Lenthe, F., & Mackenbach, J. (2007). The predictive ability of self-assessed health for mortality in different educational groups. *International Journal of Epidemiology, 36*(6), 1207–1213.
Husted, J. A., Cook, R. J., Farewell, V. T., & Gladman, D. D. (2000). Methods for assessing responsiveness: a critical review and recommendations. *Journal of Clinical Epidemiology, 53*(5), 459–468.
Idler, E. L., & Benyamini, Y. (1997). Self-rated health and mortality: a review of twenty-seven community studies. *Journal of Health & Social Behavior, 38*, 21–37.
Imlach Gunasekara, F., Carter, K., & Blakely, T. (2008). Glossary for econometrics and epidemiology. *Journal of Epidemiology and Community Health, 62*, 858–861.
Imlach Gunasekara, F., Carter, K., & Blakely, T. (2011). Change in income and change in self-rated health: systematic review of studies using repeated measures to control for confounding bias. *Social Science & Medicine, 72*(2), 193–201.
Inwood, K., & Roberts, E. (2010). *Longitudinal studies of human growth and health: A review of recent historical research* (40 pp.). University of Guelph, Department of Economics. Working papers: 1013.
Jensen, J. (1988). *Income equivalences and the estimation of family expenditures on children*. Wellington: Department of Social Welfare.
Jylhä, M. (2009). What is self-rated health and why does it predict mortality? Towards a unified conceptual model. *Social Science & Medicine, 69*(3), 307–316.
Kline, R. B. (2005). *Principles and practice of structural equation modelling*. New York: Guilford Press.
Leinonen, R., Heikkinen, E., & Jylhä, M. (1998). Self-rated health and self-assessed change in health in elderly men and women – a five-year longitudinal study. *Social Science & Medicine, 46*(4–5), 591–597.
Leinonen, R., Heikkinen, E., & Jylhä, M. (2001). Predictors of decline in self-assessments of health among older people – a 5-year longitudinal study. *Social Science & Medicine, 52*(9), 1329–1341.
Lindeboom, M., & van Doorslaer, E. (2004). Cut-point shift and index shift in self-reported health. *Journal of Health Economics, 23*(6), 1083–1099.
Manor, O., Matthews, S., & Power, C. (2001). Self-rated health and limiting long-standing illness: inter-relationships with morbidity in early adulthood. *International Journal of Epidemiology, 30*(3), 600–607.
McDonough, P., Worts, D., & Sacker, A. (2010). Socioeconomic inequalities in health dynamics: a comparison of Britain and the United States. *Social Science & Medicine, 70*(2), 251–260.
Murray, C. J., Tandon, A., Salomon, J., & Mathers, C. D. (2001). *Enhancing cross-population comparability of survey results*. GPE discussion paper no. 35 WHO/EIP. Geneva: World Health Organization.
Nielsen, A. B. S., Siersma, V., Kreiner, S., Hiort, L. C., Drivsholm, T., Eplov, L. F., et al. (2009). The impact of changes in self-rated general health on 28-year mortality among middle-aged Danes. *Scandinavian Journal of Primary Health Care, 27*(3), 160–166.
Orfila, F., Ferrer, M., Lamarca, R., & Alonso, J. (2000). Evolution of self-rated health status in the elderly: cross-sectional vs. longitudinal estimates. *Journal of Clinical Epidemiology, 53*(6), 563–570.
Perruccio, A. V., Badley, E. M., Hogg-Johnson, S., & Davis, A. M. (2010). Characterizing self-rated health during a period of changing health status. *Social Science & Medicine, 71*(9), 1636–1643.

Quesnel Vallee, A. (2007). Self-rated health: caught in the crossfire of the quest for 'true' health? *International Journal of Epidemiology, 36*(6), 1161–1164.

Rodin, J., & McAvay, G. (1992). Determinants of change in perceived health in a longitudinal study of older adults. *Journal of Gerontology, 47*(6), P373–P384.

Sacker, A., Wiggins, R., Bartley, M., & McDonough, P. (2007). Self-rated health trajectories in the United States and the United Kingdom: a comparative study. *American Journal of Public Health, 97*(5), 812–818.

Sadana, R., Mathers, C. D., Lopez, A. D., Murray, C. J., & Iburg, K. (2000). *Comparative analysis of more than 50 household surveys on health status*. GPE discussion paper no. 15. Geneva: World Health Organization.

Salmond, C., & Crampton, P. (2002). *NZDep2001 index of deprivation*. Wellington: Department of Public Health, University of Otago.

Salomon, J. A., Tandon, A., & Murray, C. J. L. (2004). Comparability of self rated health: cross sectional multi-country survey using anchoring vignettes. *BMJ, 328*(7434), 258.

Sargent-Cox, K. A., Anstey, K. J., & Luszcz, M. A. (2010). Patterns of longitudinal change in older adults' self-rated health: the effect of the point of reference. *Health Psychology March, 29*(2), 143–152.

Seymour, J., McNamee, P., Scott, A., & Tinelli, M. (2010). Shedding new light onto the ceiling and floor? A quantile regression approach to compare EQ-5D and SF-6D responses. *Health Economics, 19*(6), 683–696.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis*. Oxford: Oxford University Press.

Singh-Manoux, A., Dugravot, A., Shipley, M. J., Ferrie, J. E., Martikainen, P., Goldberg, M., et al. (2007). The association between self-rated health and mortality in different socioeconomic groups in the GAZEL cohort study. *International Journal of Epidemiology, 36*(6), 1222–1228.

Singh-Manoux, A., Martikainen, P., Ferrie, J., Zins, M., Marmot, M., & Goldberg, M. (2006). What does self rated health measure? Results from the British Whitehall II and French Gazel cohort studies. *Journal of Epidemiology & Community Health, 60*(4), 364–372.

Singh-Manoux, A. P., Gueguen, A. P., Martikainen, P. P., Ferrie, J. P., Marmot, M. F., & Shipley, M. M. (2007). Self-rated health and mortality: short- and long-term associations in the Whitehall II study. *Psychosomatic Medicine, 69*(2), 138–143.

Subramanian, S. V., & Ertel, K. (2008a). Is the use of self-rated health measures to assess health inequalities misleading? *International Journal of Epidemiology, 37*(6), 1436–1440.

Subramanian, S. V., & Ertel, K. (2008b). Self-rated health may be adequate for broad assessments of social inequalities in health. *International Journal of Epidemiology, 38*, 319–320.

Terwee, C. B., Dekker, F. W., Wiersinga, W. M., Prummel, M. F., & Bossuyt, P. M. M. (2003). On assessing responsiveness of health-related quality of life instruments: guidelines for instrument evaluation. *Quality of Life Research, 12*, 349–362.

van Doorslaer, E., & Gerdtham, U. G. (2003). Does inequality in self-assessed health predict inequality in survival by income? Evidence from Swedish data. *Social Science & Medicine, 57*(9), 1621–1629.

Ware, J., Kosinki, M., & Gandek, B. (2005). *SF-36 health survey: Manual and interpretation guide*. Lincoln, RI: Quality Metric Incorporated.

Wooldridge, J. M. (2002). *Econometric analysis of cross section and panel data*. Cambridge, Massachusetts: MIT Press.

Wooldridge, J. M. (2006). *Introductory econometrics: A modern approach*. Mason: Thomson South-Western.