CENTRE FOR ECONOMIC POLICY RESEARCH

Australian National University

DISCUSSION PAPERS

The Stability of Self Assessed Health Status*

Thomas F. Crossley[a], Steven Kennedy[b,**]

Economics Program
Research School of Social Sciences
Australian National University

DISCUSSION PAPER NO. 421

November 2000

ISSN: 1442-8636
ISBN: 0 7315 2285

---

[a] Department of Economics, York University, Toronto and Centre for Economic Policy Research, The Australian National University.

[b] Economics Program, Research School of Social Sciences, Australia National University

** Corresponding Author. Economics Program, Research School of Social Sciences, The Australia National University, Canberra, ACT, Australia 0200. skennedy@coombs.anu.edu.au

* The first author gratefully acknowledges the financial support of the Social Sciences and Humanities Research Council of Canada. We thank the Australian Bureau of Statistics and seminar participants at The Australian National University and particularly Jeff Borland, Deborah Cobb-Clark, Paul Grootendorst, Jeremiah Hurley, Michael Ornstein, Arthur Sweetman, Byron Spencer and Michael Veall for helpful comments. All remaining errors are our own.

CONTENTS

ABSTRACT

The use of self assessed health status as a measure of health is common in empirical research. We analyse a unique Australian survey in which a random sub-sample of respondents answer a standard self assessed health question twice – before and after an additional set of health related questions. 28% of respondents change their reported health status. Response instability is related to age, income and occupation. We also compare the responses of these individuals to other respondents who are queried only once. The distributions of responses to both questions by the former group are statistically different from the distribution of responses by the latter group.

Keywords: Self assessed health status

JEL: I10; C80

1. Introduction

Self assessed health status is an increasingly common measure of health in empirical research (e.g., Smith, 1999; Kennedy et al., 1998; Deaton and Paxson, 1998; Schofield, 1996; Ettner, 1996; Saunders, 1996). This is supported by a literature that shows that self assessed health status predicts mortality and morbidity (e.g., Idler and Kasl, 1995; McCallum et al., 1994; Connelly et al., 1989; Okun et al., 1984). Furthermore, Gerdtham et al., (1999) have demonstrated that a continuous health status measure constructed from a categorical response by the method of Wagstaff and van Doorslaer (1994) is highly correlated with other continuous measures of health.[1]

The 1995 Australian National Health Survey provides a unique opportunity to examine self assessed health status measures in a different way. The following standard self assessed health status question was asked of all respondents and *twice* of a random subsample:

In general, would you say that your health is:

Excellent?
Very good?
Good?
Fair?
Poor?

For the "treatment" group, this question is first asked at the beginning of a general health and well being questionnaire. The question is asked again after the respondent has completed the general health and well being questionnaire and answered some other non-health related questions. The distributions of responses to these two questions are statistically different. In addition, both distributions are statistically different from the distribution of responses by the group that was asked only once.

Among respondents who were asked the self reported health status question twice, approximately 28% change their response, though, only 3% change their response by more than one category. Some socio-economic groups are more likely than others to revise their self assessed health status on repeated questioning. For example, a higher proportion of older than younger persons change their self assessed health status.

These patterns of responses and changes in responses admit several interpretations and may have implications for empirical research employing self assessed health measures. Before turning to those interpretations and implications, we provide further detail on the data, and the patterns therein.

---

[1] Wagstaff and van Doorslaer (1994) constructed a continuous health measure from the categorical self assessed health status variable used in this study by assuming an underlying latent health status variable with a lognormal distribution.

2. The 1995 Australian National Health Survey

The Australian Bureau of Statistics (ABS) 1995 National Health Survey[2] was conducted over a 12 month period from January 1995 to January 1996 and based on a sample of private and non-private dwellings. Approximately 23,800 dwelling households were surveyed and the overall response rate for households was 91.5%. An important feature of the data is that they are representative of the entire Australian adult population which allows us to compare the stability of self assessed health status among different groups.

In addition to responding to the standard "face-to-face" interview, approximately half of the original sample was asked to complete a written supplement comprising the Short Form 36 health status questionnaire (SF-36). The SF-36 is a well-known measure of general health and well being; it produces scores for eight dimensions of health.[3] Selection into this "treatment" group was based on the random assignment of blocks within census districts.

In households selected to respond to the SF-36, persons aged 18 and over were asked to complete the SF-36 questionnaire *prior* to the standard interview. The first question on the SF-36 is the self assessed health status question presented in the introduction. This question is asked again (but by an interviewer rather than through a self completed form) in the standard interview. In the standard interview, the self assessed health question is the first question about the respondent's health, but is preceded by questions regarding gender, age, marital status, race, country of birth, year of arrival in Australia, language spoken at home and employment status. Thus, for the treatment group, the responses to the two self assessed health status questions are separated by the rest of the health questions on the SF-36 form and the socioeconomic questions listed above. 18436 persons responded to both self assessed health status questions.

Individuals in households that were not selected to respond to the SF-36 were asked the self assessed health question only once, in the standard (face-to-face) interview, after the socioeconomic questions listed above. We refer to these individuals as the "control" group. Figure 1 summarizes the self assessed health questions asked of the treatment and control groups. Appendix Table A.1 presents the distributions of socioeconomic characteristics and the results of tests for differences between the treatment and control groups. We find no evidence to suggest the randomization was inadequate.

Aside from the two self assessed health status variables and answers to the SF-36 questionnaire, other variables of interest include gender, age, employment status, equivalent income and occupation. The definitions of all variables are presented in Table 1.

---

[2] See Australian Bureau of Statistics (1995).

[3] See Ware et al. (1993)

Figure 1: Survey Process

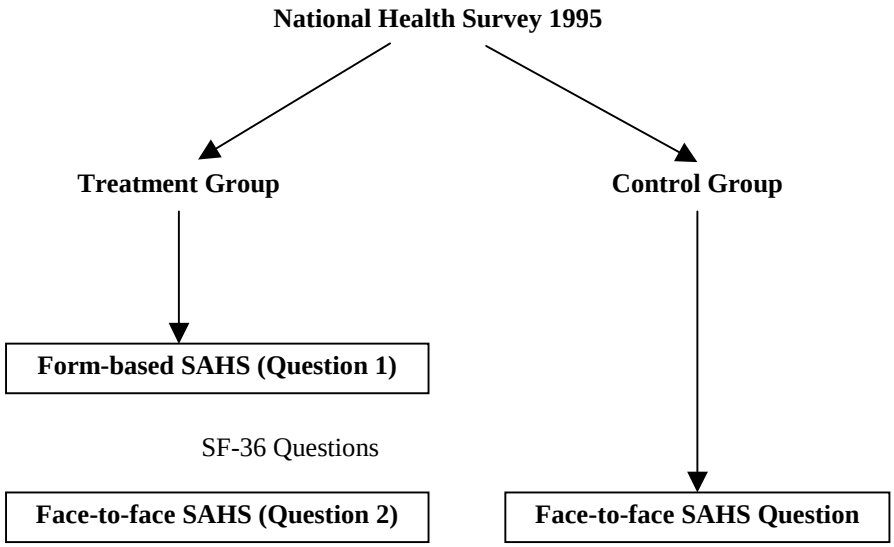

Table 1: Variable Definitions

Variable	Definition
Age	Age is defined in 4 groups, age 18 to 24 years, age 25 to 54 years, age 55 to 69 years, and age 70 years and over.
Gender	Males and females.
Self Assessed Health Status (SAHS)	Responses to the self assessed health status questions are scored as follows (1 = excellent, 2=very good, 3=good, 4=fair, 5=poor).
Occupation	There are 3 occupation groups: managers and professionals (white collar); para-professional, clerks, salespersons and personal service workers (other white collar); and labourers and related workers, tradespersons, and plant and machine operators and drivers (blue collar).
Employment Status	Employment status was represented by 3 groups; employed, unemployed, and not in the labour force.
Equivalent Income	Equivalent income was coded according to a person's equivalent income quintile. The equivalent income measure took into account different household types by applying Henderson simplified equivalence scales at the income unit level. See Australian Bureau Statistics (1995) for a more detailed description of this variable.

3. The Distribution of Responses by Question and Group

The distribution of responses to both self assessed health questions for the treatment group and to the single self assessed health question in the case of the control group are presented in Table 2. In Table 3, we present tests for differences in mean and distribution across questions within the treatment group, and also across the treatment and control groups.

Comparing the distributions of responses to the two questions for the treatment group, we cannot reject the null that the means are the same (at a 5 % level of significance) but the null that the distributions are the same is strongly rejected. This is due to a fairly symmetric thickening of the tails. When respondents were questioned a second time, by face-to-face interview rather than by written questionnaire, the middle category ("Good") was reported less frequently and *all* the other categories became more prevalent.

Table 2: Self Assessed Health Status Response Frequencies (percent)

	Treatment Group SAHS Question 1 Self completed, written response, asked first	Treatment Group SAHS Question 2 Personal interview, asked second	Control Group SAHS Personal Interview, asked first
Excellent	17.3	18.0	20.2
Very Good	37.2	37.5	34.5
Good	31.3	27.5	29.6
Fair	11.3	13.2	11.8
Poor	3.0	3.8	3.8

Table 3: Differences in Means and Distributions

	Difference in Means t-test (P value)	Difference in Multinomial Distributions Pearson Chi-square (P value)
SAHS Q1 vs SAHS Q2	1.58 (0.12)	93.48 (0.00)
SAHS Q1 vs SAHS Control group	0.98 (0.33)	92.41 (0.00)
SAHS Q2 vs SAHS Control group	2.53 (0.01)	77.88 (0.00)

Comparing the treatment and control groups, we find a statistically significant difference between the distribution of responses generated by the control group and that generated by either question asked of the treatment group. The differences cannot be simply characterised by either spread or location. We know from Appendix Table A.1 that the distribution of co-variates does not differ across the treatment and control groups, and we have just observed that the unconditional distribution of responses does differ across the groups; thus we can infer that conditional distributions of responses differ across the treatment and control groups. This is confirmed in Table 4 where we see that most conditional distributions of responses do in fact vary across the treatment and control groups.

Table 4: Differences in Multinomial Distributions - Pearson Chi-square (P value)

Variables	SAHS Q1 vs SAHS Q2	SAHS Q1 vs SAHS Control group	SAHS Q2 vs SAHS Control group
Age			
Age 18-24	29.18 (0.00)	35.03 (0.00)	30.63 (0.00)
Age 25-54	50.27 (0.00)	55.80 (0.00)	49.18 (0.00)
Age 55-69	24.27 (0.00)	18.20 (0.00)	8.96 (0.06)
Age 70 +	6.84 (0.15)	5.20 (0.27)	2.48 (0.65)
Gender			
Female	35.02 (0.00)	47.41 (0.00)	34.56 (0.00)
Male	61.91 (0.00)	48.71 (0.00)	50.52 (0.00)
Occupation			
White collar	29.38 (0.00)	27.13 (0.00)	21.67 (0.00)
Other white collar	24.85 (0.00)	26.35 (0.00)	32.32 (0.00)
Blue collar	21.28 (0.00)	12.45 (0.01)	23.64 (0.00)
Employment Status			
Not in Labour force	72.16 (0.00)	58.89 (0.00)	71.31 (0.00)
Unemployed	17.28 (0.00)	20.60 (0.00)	14.12 (0.01)
Employed	7.85 (0.10)	18.36 (0.00)	8.94 (0.06)
Income			
Income Quintile 1	13.23 (0.01)	25.91 (0.00)	21.00 (0.00)
Income Quintile 2	13.62 (0.01)	13.89 (0.01)	9.72 (0.05)
Income Quintile 3	28.34 (0.00)	12.23 (0.02)	12.08 (0.02)
Income Quintile 4	13.23 (0.01)	11.18 (0.03)	16.21 (0.00)
Income Quintile 5	19.67 (0.00)	20.99 (0.00)	13.84 (0.01)
All	93.49 (0.00)	92.42 (0.00)	77.89 (0.00)

Focusing on the treatment group, we find in Table 5 that for both questions, there were differences in mean self assessed health status across socio-economic groups. For example, older age groups tended to have higher mean self assessed health status, indicating that their reported health status was generally poorer than that of younger age groups (the categorical variable is coded numerically from 1 "excellent" to 5 "poor").[4] The self assessed health status means and variances (Table 6) for persons in blue collar occupations show that their health is poorer and more disperse than that of persons in white collar occupations.[5]

[4] As Deaton and Paxson (1998) point out, it is not surprising to find people reporting poorer health as they age. However, this means people probably assess their health by reference not only to their own age group but also younger age groups. Older age groups also showed larger variation in health status

[5] Ideally, we would have examined different groups defined by educational attainment. However, persons answering the SF-36 were not asked questions about their educational attainment.

Similarly, persons in low equivalent income quintiles seem to have poorer health and exhibit more dispersion in reported health outcomes compared to persons in high equivalent income quintiles.

Table 5: Self Assessed Health Status Means

Variables	Number	Treatment Group SAHS question 1 Self completed, written response, asked first	Treatment Group SAHS question 2 Personal interview, asked second	Number	Control Group SAHS Personal interview, asked first
Age					
Age 18-24	4445	2.320	2.308	4943	2.243
Age 25-54	9321	2.316	2.347	9904	2.319
Age 55-69	2986	2.742	2.765	3185	2.793
Age 70 +	1684	3.078	3.085	1833	3.066
Gender					
Female	9583	2.451	2.474	10231	2.434
Male	8853	2.461	2.471	9634	2.458
Occupation					
White collar	3156	2.165	2.166	3244	2.101
Other white collar	4584	2.234	2.238	4880	2.177
Blue collar	3934	2.374	2.378	4227	2.359
Employment Status					
Not in Labour force	3344	2.693	2.751	3770	2.759
Unemployed	792	2.528	2.587	901	2.532
Employed	11671	2.262	2.265	12350	2.219
Income					
Income Quintile 1	2876	2.651	2.683	3182	2.659
Income Quintile 2	2920	2.827	2.850	3265	2.832
Income Quintile 3	3059	2.435	2.469	3216	2.433
Income Quintile 4	3315	2.300	2.314	3349	2.259
Income Quintile 5	3708	2.199	2.184	3857	2.165
All	18436	2.456	2.473	19865	2.445

Table 6: Self Assessed Health Status Standard Deviations

Variables	Number	Treatment Group SAHS question 1 Self completed, written response, asked first	Treatment Group SAHS question 2 Personal interview, asked second	Number	Control Group SAHS Personal interview, asked first
Age					
Age 18-24	4445	0.887	0.931	4943	0.923
Age 25-54	9321	0.943	0.994	9904	1.004
Age 55-69	2986	1.061	1.131	3185	1.127
Age 70 +	1684	1.127	1.169	1833	1.174
Gender					
Female	9583	1.004	1.049	10231	1.056
Male	8853	0.996	1.051	9634	1.059
Occupation					
White collar	3156	0.866	0.919	3244	0.908
Other white collar	4584	0.867	0.915	4880	0.899
Blue collar	3934	0.905	0.942	4227	0.946
Employment Status					
Not in Labour force	3344	1.090	1.136	3770	1.150
Unemployed	792	0.944	1.021	901	1.066
Employed	11671	0.884	0.929	12350	0.923
Income					
Income Quintile 1	2876	1.077	1.130	3182	1.142
Income Quintile 2	2920	1.078	1.123	3265	1.138
Income Quintile 3	3059	0.959	1.018	3216	1.010
Income Quintile 4	3315	0.910	0.940	3349	0.942
Income Quintile 5	3708	0.872	0.917	3857	0.919
All	18436	1.000	1.050	19865	1.057

4. Who Revises their Self assessed Health Status?

As noted in the introduction, 28% of the treatment group change their response between the two self assessed health status questions. 13.6% reported a higher level of health whilst 14.8% reported a lower level of health. These *gross* flows are large relative to the *net* changes discussed above. (In Table 2, responses in the top two categories combined went up by 1 percentage point between the first and second question, while responses to the bottom to categories increased by a combined 2.7 percentage points). There is evidently considerable "churning". Table 7 shows in detail the transitions from self assessed health status question 1

categories to question 2 categories. For example, of those people reporting their health as "good" in question 1, 64.19% still report their health as "good" in question 2.

The amount of revision does vary by socioeconomic group. Table 8 demonstrates that the proportion of persons changing their response was significantly higher for older age groups compared to younger age groups. Variation in revisions across age groups illustrates the importance of examining both gross and net flows. The oldest age group was the only one with no significant variation in net flows even though it had the largest gross flows of all age groups. A smaller proportion of persons in the two white collar occupation groups changed their self assessed health status than those in blue collar occupations. The proportion of persons changing there self assessed health status in the two top quintiles of equivalent income was lower than in other equivalent income quintiles. To counter the effect that age might be having on occupation and family income groups, these groupings were age standardised. Standardising for age did not qualitatively alter the patterns by occupation and family income. We also present weighted and unweighted Kappa scores in the final two columns of Table 8; these measure the agreement between the two responses by socioeconomic group. They reveal the same patterns as the earlier columns.[6]

Table 7: Transition Matrix (percent)

	Self Assessed Health Status question 2					
Self Assessed Health Status question 1	Excellent	Very Good	Good	Fair	Poor	Total Proportion
Excellent	77.75	17.73	3.70	0.72	0.09	17.3
Very Good	10.32	74.27	13.65	1.53	0.23	37.2
Good	2.08	20.68	64.19	12.44	0.61	31.3
Fair	0.77	3.22	14.42	71.83	9.76	11.3
Poor	0.36	0.72	1.80	16.58	80.54	3.0
Total Proportion	18.0	37.5	27.5	13.2	3.8	100

[6] The Kappa statistic was developed by Cohen (1960, 1968). See Grootendorst et al. (1997) for an example of the use of the Kappa statistic.

Table 8: Proportion Changing between Self Assessed Health Status Questions

Variables	Number	Proportion changing between SAHS question 1 and SAHS question 2	F and t tests*	Kappa scores	Weighted Kappa scores
Age			17.338		
Age 18-24	4445	0.272	5.409	0.60	0.69
Age 25-54	9321	0.270	6.018	0.62	0.71
Age 55-69	2986	0.312	2.113	0.58	0.71
Age 70 +	1684	0.341		0.55	0.69
Gender			11.919		
Female	9583	0.273	3.452	0.62	0.73
Male	8853	0.296		0.59	0.70
Occupation			19.908		
White collar	3156	0.243	5.841	0.65	0.72
Other white collar	4584	0.257	4.940	0.62	0.70
Blue collar	3934	0.305		0.56	0.65
Employment Status			5.484		
Not in Labour force	3344	0.292	2.380	0.61	0.73
Unemployed	792	0.308	2.570	0.57	0.68
Employed	11671	0.269		0.61	0.69
Income			13.523		
Income Quintile 1	2876	0.306	5.580	0.59	0.71
Income Quintile 2	2920	0.311	6.081	0.58	0.70
Income Quintile 3	3059	0.301	5.231	0.58	0.68
Income Quintile 4	3315	0.272	2.681	0.60	0.70
Income Quintile 5	3708	0.244		0.64	0.72
All	18436	0.284		0.61	0.72

* All t and F tests are derived from OLS regressions where change in self assessed health status is regressed on each set of variables of interest. All F tests were highly significant.

The stability of self assessed responses among the older population may be of particular interest because of the growing empirical literature on the health and labour market activity of older persons. Table 8 shows the frequency of revisions by age group and initial response. Older persons exhibit higher revision propensities for every initial response except "Fair". Even more detail can be uncovered by examining the complete transition matrices by age group. These are presented in Appendix Table A.2. In younger age groups, most responses to self assessed health status were excellent, very good and good, and therefore most transitions were from these categories. In older age groups, responses were more evenly spread amongst self assessed health status categories and there tended to be slighter higher rates of transition between questions.

5. Interpretations

To summarise, we find that the distribution of self assessed health status responses differs between:

(1) A form-based questionnaire and a face-to-face question asked of the same individuals in quick succession;

(2) Form-based and face-to-face questions asked of two randomly allocated groups; and

(3) Face-to-face questions asked of a control group and a randomly selected treatment group which was "pre-treated" with form-based questions.

We also find that for the treatment group, changes in self assessed health between the initial form-based questionnaire and the subsequent face-to-face interview are numerous and related to age, occupation and income. How can we interpret these patterns?

Individuals in the treatment group might revise their self assessed health status for at least three reasons. First, it may of course be that they assess their health with some "error" and each response to a question represents a new draw from the measurement error distribution. If this were all that was going on, the frequency of revisions would seem to suggest a large degree of underlying uncertainty (or measurement error).

Second, it may be that there is an "instrument effect" as the respondents go from a written questionnaire to a face-to-face interview. There is a literature which suggests that people respond more candidly to sensitive questions when self completing a form as opposed to being personally interviewed (e.g., Tourangeau et al., 1996).

Finally, and perhaps most intriguingly, it may be that respondents in the treatment group "learn" about their health status between the first and second self assessment of health. Recall that after the first summary self assessed health question they respond to a number of detailed questions about various aspects of their health status. It may be that these detailed questions stimulate a process of introspection that leads to different responses to the second question. We use the term "learn" somewhat loosely: the later responses might or might not be more accurate. Another interpretation of the effect of preceding questions is framing, see for example, Tversky et al. (1998). In what follows we use the term "learning" to refer to any effect of the intervening questions on responses to the second self assessed health status question.

It might at first be tempting to ask if responses to the intervening specific health questions "predict" subsequent revision. However, such a correlation could simply reflect mean reversion, and thus would be just as consistent with our first explanation ("measurement error") as our third ("learning"). Suppose that each of the self assessed health questions and the SF-36 index represented "true" health status plus an independent draw from a measurement error distribution. Then it is easy to show that the best predictor of an individual's response to the second self assessed health status question is a (weighted) mean

of their response to the first self assessed health status question and the SF-36 index.[7] Thus if the SF-36 indicates better health status than the first self assessment, one would expect the second self assessment, on average, to indicate better health status as well, even though the SF-36 has no causal effect on the second assessment.

One way we might able to distinguish between learning and mean reversion is to appeal to temporal ordering. Learning should proceed in only one temporal direction, so that if learning is occurring, the intervening health questions should better predict responses to self assessed health status question 2 than to self assessed health status question 1. By contrast, mean reversion should work equally in either direction so that the intervening questions would predict self assessed health status question 1 and question 2 equally well. With this in mind, we estimated two ordered probit regressions in which the self assessed health status questions were regressed against the 35 intervening (SF-36) health questions. For self assessed health status question 1, the pseudo R squared was 0.31, for self assessed health status question 2 it was 0.34. Thus, there is some slight evidence of learning as the intervening questions better predict responses to self assessed health status question 2 than self assessed health status question 1.

The control group may provide a more promising way of untangling the three explanations listed at the beginning of this section. If we compare the distributions of responses across the randomly selected treatment and control groups, measurement error, which should have the same structure for the two randomly selected groups, does not explain any observed differences. Thus, we can focus on other explanations.

If we compare the *first* response of the treatment group to the responses of the control group, neither group has previously responded to any other health questions. However, the treatment group are self-completing a form, while the control group are responding verbally. Thus this comparison isolates (of the explanations we have considered) the instrument effect. As noted in Table 2, we find a statistically significant effect.

If we compare the *second* response of the treatment group to the responses of the control group, both groups are answering verbally. However, the treatment group has previously responded to the SF-36 questionnaire. Thus, this comparison isolates the "learning" effect. Again, as noted in Table 2, we find a statistically significant effect.

Thus we are forced to conclude that both the instrument (written or verbal) and the sequence of previous questions are determinants of individual's responses to a self assessed health question.

Turning now to the revision propensities of different socioeconomic groups, can we conclude, for example, that the aged are less accurate in their self assessments of health? While older persons are more likely to change categories, we know from Appendix Table A.2 that the elderly have a different initial distribution of responses than younger age groups. Response instability will reflect both the amount of uncertainty in responses and the underlying distribution of "true" health status, so that care must be taken with cross group comparisons.

---

[7] Where the weights would reflect the variances of the distributions of measurement error relative to "true" health status of the first self assessed health question and the SF-36 Index.

To see this simply, consider a model of responses that abstracts from the "learning" and instrument effects documented above. Suppose that the categorical self assessed health status is generated in the following way: There is an underlying continuous latent variable corresponding to "true" health. Individuals can assess the sum of this variable and a draw from some measurement error distribution. They then generate their categorical response by comparing the sum just defined to a set of fixed "cut points".[8] Each assessment leads the individual to make a new draw from the measurement error distribution. It is easy to show that, given the "cut points", the predicted number of revisions depends not only on the variance of the measurement error distribution but also on distribution of the latent health status.[9] Thus, one needs to be careful about interpreting the "accuracy" of self-health assessment of different socioeconomic groups. A greater propensity to revise assessments may not reflect a greater underlying uncertainty about "true" health, but rather a different distribution of health status.[10]

However, our results would still seem to suggest that older persons have greater difficulty in self assessing their health. Crucially, as we noted in Table 9, they have a higher propensity to revise for all but one category of initial response (and for that category their propensity to revise is essentially the same as younger persons). This observation would seem to argue against an interpretation that attributed their higher revision propensity entirely to a different distribution of underlying "true" health.

Table 9: Proportion changing between SAHS questions

	Proportion changing by response to SAHS question 1					
	Excellent	Very Good	Good	Fair	Poor	Total Proportion
Age 18-24	0.223	0.224	0.356	0.313	0.167	0.272
Age 25-54	0.211	0.248	0.346	0.264	0.180	0.270
Age 55-69	0.230	0.325	0.378	0.264	0.172	0.312
Age 70 +	0.331	0.368	0.390	0.307	0.236	0.341
Total Proportion	0.223	0.257	0.358	0.282	0.195	0.284

[8] We have in mind an ordered probit or ordered logit model.

[9] To see this in a trivial way, consider a subgroup of the population whose distribution "true" health has a single point of support far to the right of the rightmost "cut point". This group could have a very large measurement error variance and very few revisions.

[10] If we define accuracy without reference to some notion or model of underlying uncertainty, but rather in terms of revisions or response stability, then those who revise more frequently are by definition less accurate.

6. Implications for Empirical Research

Our results suggest that individuals' responses to a self assessed health question depend on both the nature of the survey (particularly whether responses are written or verbal) and the sequence of preceding questions. These are important factors to keep in mind when comparing the distribution of self assessed health status across different surveys. Many health surveys contain multiple measures of health status. Our results suggest that it is necessary to consider the order in which health status questions are asked when comparing the results of different surveys.

Our results also suggest that there is considerable measurement error or underlying uncertainty in individual's self assessment of health. This leads to response instability. Measurement error will lead to inconsistent estimation of models in which self assessed health appears as an explanatory variable (for example, a model of retirement). Furthermore, the degree of this uncertainty appears to be related to common observable characteristics. That is, this uncertainty is, at minimum, heteroscedastic. This means that maximum likelihood estimate of a health status model (an ordered probit or logit for example) will be inconsistent, if such estimation assumes homoscedastic error structures (as is usually the case). It seems likely (though it cannot be shown with our data) that individuals' uncertainty about their health status is also correlated *in mean* with socioeconomic characteristics. This would be problematic for studies that take self assessed health status as the dependent variable even if they are estimated by OLS.

How important might such considerations be? To provide a partial answer to that question, we consider a simple linear probability model of employment status, and examine the consequences of improving the estimates by combining both our measures of self assessed health. A simple model of measurement error in an explanatory variable (such as can be found in most econometrics textbooks) leads to a prediction of attenuation bias – the coefficient is biased towards zero, and the degree of attenuation decreases as the variance of the measurement error decreases. If we average two draws from the measurement error distribution, the variance of the average is less than the variance of a single draw, and thus a strategy which uses the average of two measures as the explanatory variable should result in reduced attenuation bias. Another possibility where two measures exist is to use one as an instrumental variable for the other. If the two measurement errors are uncorrelated, this leads to a consistent estimate. However, in our context – repeated self assessment of health – that seems unlikely to be true. Nevertheless, even if the measurement errors are correlated, under reasonable assumptions an instrumental variables (IV) estimate should lead to less attenuation bias than either a simple OLS estimate or an estimation strategy that averages the two responses. A simple model supporting this intuition is provided in Appendix A. The motivation behind our empirical example is that the actual attenuation bias must be at least as large as the difference between the OLS and IV estimates (if the measurement errors are uncorrelated) or even larger but in the direction indicated by that comparison (if the measurement errors are correlated).

In Table 10: Employment Regressions

Estimation Method:	OLS	OLS	OLS	IV
Independent Variable:	SAHS question 1	SAHS question 2	Mean of SAHS question 1 and 2	
Instrument:				SAHS question 2
	Coefficient (t-stat)	Coefficient (t-stat)	Coefficient (t-stat)	Coefficient (t-stat)
Excellent	0.123 (1.02)	0.133 (12.87)	0.135 (12.15)	0.155 (10.66)
Very Good	0.112 (13.54)	0.116 (13.52)	0.129 (13.42)	0.170 (10.58)
Fair	-0.230 (19.43)	-0.179 (15.65)	-0.227 (17.13)	-0.241 (10.43)
Poor	-0.423 (20.57)	-0.410 (22.04)	-0.454 (21.55)	-0.513 (16.98)

The dependant variable is employment (0=not employed, 1=employed). The omitted category for the independent variables is Good. There were 18436 observations, we report estimating linear probability models of employment on our treatment group, for

In Table 10, we report estimating linear probability models of employment on our treatment group, for whom the overall rate of employment is 74%. The explanatory variables are dummy variables indicating that the respondent reported their health as “excellent”, “very good”, “fair” or “poor” (“good” is the omitted category). This example is intended to be illustrative, and so we are abstracting from issues of reverse causation.

The first two columns of Table 10 report OLS estimates, where the explanatory variables are derived from the first and second self assessed health questions respectively. Health status is a statistically significant determinant of employment status. For example, the OLS estimates suggest that those who report their health as “fair” have an employment probability that is some twenty percentage points below those who reported “good” and “poor” health corresponds to an employment probability more than forty percentage points below that of the group reporting “good”.

When we average the two self assessed health measures (column 3) or use the second self assessed health as an instrumental variable for the first (column 4) we find even larger differences among the groups. This is consistent with the idea that the OLS estimates suffer from attenuation bias. Some of the differences are substantial. The OLS estimates suggest that “poor” health corresponds to an employment probability of 42 percentage points below those with “good” health, while the IV estimates suggest a differential of 51 percentage points – a 9 percentage point difference. If (as seems likely) the IV estimates also suffer from attenuation bias (because the measurement errors are correlated), the true degree of attenuation bias may be even larger. Thus, at least in some contexts the underlying uncertainty in self assessed health status can have important effects in empirical applications.

There are two other features of this empirical illustration to note. First, as our simple model predicts, a strategy of averaging the two measures leads to estimates that lie between the OLS and IV estimates. Second, the OLS estimates do not differ substantially when we

derive our regressors from the second rather than first self assessed health status measure. This suggests that "learning" that occurred between the first and second response did not substantial reduce the attenuation bias.

Empirical research must always proceed with the best available measures of quantities and concepts of interest. Nevertheless, it is important to bear in mind what evidence there is on the limitations of the best available measures, and this note has provided some evidence on the limitations of self assessed health measures.

Appendix A

We now present a simple model, which underlies the discussion of attenuation bias in the previous section. Suppose that

$$y_i = \beta hs_i + e_i \qquad (1)$$

Where y_i is an outcome of interest and hs_i is "true" health. For convenience, we express all variables as deviations from means. We have two error-ridden measures of health:

Question 1: $x_i = hs_i + u_i$

Question 2: $z_i = hs_i + v_i$

We define the following:

$$\sigma_{hs}^2 = plim\frac{1}{n}\sum_1^n hs_i^2 \quad \sigma_u^2 = plim\frac{1}{n}\sum_1^n u_i^2 \quad \sigma_v^2 = plim\frac{1}{n}\sum_1^n v_i^2$$

We assume that the measurement errors are "classical" – uncorrelated with "true" health status and with the disturbance in the structural equation,

$$plim\frac{1}{n}\sum_1^n e_i v_i = plim\frac{1}{n}\sum_1^n e_i u_i = plim\frac{1}{n}\sum_1^n hs_i v_i = plim\frac{1}{n}\sum_1^n hs_i u_i = 0$$

but allow the possibility that the measurement errors can be correlated (respondents can persistently over or underestimate their health),

$$plim\frac{1}{n}\sum_1^n u_i v_i = \sigma_{uv}$$

An OLS estimate of (1) solves:

$$x_i(y_i - b^{OLS}x_i) = 0\,.$$

With a little substitution and manipulation, it can be shown that:

$$plim\ b^{OLS} = \frac{\beta\sigma_{hs}^2}{\sigma_{hs}^2 + \sigma_u^2}\,.$$

This is the usual textbook result that measurement error leads to attenuation bias, with the attenuation bias disappearing as the variance of the measurement error goes to zero.

One way to use the second question to improve such an estimate would be to use the responses of the second question as an instrumental variable for the first question. The instrumental variable (IV) estimate of (1) solves:

$$z_i(y_i - b^{IV} x_i) = 0 .$$

For z_i to be a valid instrument it must be uncorrelated with the measurement error in x_i, in which case the IV estimate is consistent. However, this is not the case in our example as we have allowed the measurement errors in z_i and x_i to be correlated. It can be shown that:

$$plim\ b^{IV} = \frac{\beta\sigma_{hs}^2}{\sigma_{hs}^2 + \sigma_{uv}} .$$

The IV estimate, while not consistent, exhibits less attenuation bias than the OLS estimate as long as $\sigma_{uv} < \sigma_u^2$, or equivalently $\rho\sigma_v < \sigma_u$. (This seems quite likely both because the correlation ρ is less than one and because if respondents learn about their health via the intervening questions then $\sigma_v < \sigma_u$. That respondents learn about their health is suggested by the results of ordered probit regressions discussed in Section 5. However, our model is only intended to be illustrative, as it is not necessarily the case that more information will produce less uncertainty.)

Another way to use the information might be to combine the two self assessed health responses in an attempt to "average out" the measurement error. This might lead to an estimator that solves:

$$\frac{x_i + z_i}{2}(y_i - b^{avg} \frac{x_i + z_i}{2}) = 0 ,$$

and it can be shown that:

$$plim\ b^{avg} = \frac{\beta\sigma_{hs}^2}{\sigma_{hs}^2 + \frac{1}{4}(\sigma_u^2 + 2\sigma_{uv} + \sigma_v^2)}$$

This will typically lie between the OLS and IV estimates.

References

Australian Bureau of Statistics., 1995. National Health Survey Users' Guide. ABS Catalogue No. 4363.0. Australian Government Publishing Service.

Cohen, J., 1960. A coefficient of agreement for nominal scales. Educational and Psychological Measurement 20, 36-46.

Cohen, J., 1968. Weighted kappa: nominal scale agreement with provision for scales disagreement or partial credit. Psychological Bulletin 70, 213-220.

Connelly, J.E., Philbrick, J.T., Smith, R.G., Kaiser, D.L., Wymer, A., 1989. Health perceptions of primary care patients and the influence on health care utilisation. Medical Care 27, S99-S109.

Deaton, A.S., Paxson, C.H., 1998. Aging and inequality in income and health. American Economic Review, papers and proceedings 88, 248-253.

Ettner, S.L., 1996. New evidence on the relationship between income and health. Journal of Health Economics 15, 67-85.

Gerdtham, U.G., Johannesson, M., Lundberg, L., Isacson, D., 1999. A note on validating Wagstaff and van Dooslair's health measure in the analysis of inequality in health. Journal of Health Economics 18, 117-224.

Grootendorst, P., Feeny, D., Furlong, W., 1997. Does it matter whom and how you ask? Inter and intra-rater agreement in the Ontario health survey. Journal of Clinical Epidemiology 50, 127-136.

Idler, E.L., Kasl, S.V., 1995. Self-ratings of health: do they also predict change in functional ability? Journal of Gerontology 50B, S344-S353.

Kennedy, B.P., Kawachi, I., Glass, R., Prothrow-Stith, D., 1998. Income distribution, socioeconomic status, and self rated health in the United States: multilevel analysis. British Medical Journal 317, 917-921.

McCallum, J., Shadbolt, B., Wang, D., 1994. Self-rated health and survival: a 7-year follow-up study of Australian elderly. American Journal of Public Health 847, 1100-1105.

Okun, M.A., Stock, W.A., Haring, M.J., Witter, R.A., 1984. Health and subjective well-being: a meta-analysis. International Journal of Aging and Human Development 192, 111-132.

Saunders, P., 1996. Income, health and happiness. The Australian Economic Review, 353-366.

Schofield, D., 1996. The impact of employment and hours of work on health status and health service use. National Centre for Social and Economic Modelling Discussion Paper No. 11.

Smith, J.P., 1999. Health bodies and thick wallets: the dual relation between health and economic status. Journal of Economic Perspectives 13, 145-166.

Tourangeau, R., Smith, T.W., 1996. Asking sensitive questions the impact of data collection mode, question format, and question context. Public Opinion Quarterly 60, 275-304.

Tversky, A., Kahneman, D., 1998. Rational choice and the framing of decisions in: Bell, D., Raiffa, D., Tversky, A. (Eds.), Decision Making: Descriptive, Normative, and Prescriptive Interactions. Cambridge University Press, Great Britain.

Wagstaff, A., van Doorslair, E., 1994. Measuring inequalities in health in the presence of multiple category morbidity indicators. Health Economics 3, 281-291.

Ware, J., Snow, K., Gandek, B., 1993. SF-36 Health Survey Manual and Interpretation Guide. The Health Institute, New England Medical Centre. Boston, Massachusetts.

Appendix

Table A.1: Cell Counts of Variables of Interest

Variables	Treatment Group Number	Control Group Number	Difference in Multinomial Distributions Pearson Chi-square (P value)
Age			3.516 (0.31)
Age 18-24	4445	4943	
Age 25-54	9321	9904	
Age 55-69	2986	3185	
Age 70 +	1684	1833	
Gender			0.872 (0.35)
Female	9583	10231	
Male	8853	9634	
Occupation			1.912 (0.39)
White collar	3156	3244	
Other white collar	4584	4880	
Blue collar	3934	4227	
Employment Status			6.836 (0.03)
Not in Labour force	3344	3770	
Unemployed	792	901	
Employed	11671	12350	
Income			11.76 (0.02)
Income Quintile 1	2876	3182	
Income Quintile 2	2920	3265	
Income Quintile 3	3059	3216	
Income Quintile 4	3315	3349	
Income Quintile 5	3708	3857	
All	18436	19865	

There were a number of missing observations for occupation, employment status and income variables.

Table A.2a: Transition Matrix (percent) Age 18 to 24 years

	Self Assessed Health Status Question 2					
Self Assessed Health Status Question 1	Excellent	Very Good	Good	Fair	Poor	Total Proportion
Excellent	77.69	18.55	3.26	0.50	0	17.95
Very Good	10.50	77.61	10.66	1.02	0.22	41.80
Good	2.15	21.86	64.44	11.11	0.43	31.38
Fair	1.96	5.31	19.27	68.72	4.75	8.05
Poor	0	0	5.56	11.11	83.33	0.81
Total Proportion	19.17	43.06	26.86	9.63	1.28	100

Table A.2b: Transition Matrix (percent) Age 25 to 54 years

	Self Assessed Health Status Question 2					
Self Assessed Health Status Question 1	Excellent	Very Good	Good	Fair	Poor	Total Proportion
Excellent	78.88	17.23	3.30	0.49	0.11	19.86
Very Good	9.69	75.19	13.35	1.54	0.24	40.51
Good	1.92	20.56	65.45	11.60	0.47	29.69
Fair	0.13	3.47	14.42	73.56	8.41	8.04
Poor	1.12	1.12	0.56	15.17	82.02	1.91
Total Proportion	20.19	40.29	26.66	10.36	2.50	100

Table A.2c: Transition Matrix (percent) Age 55 to 69 years

	Self Assessed Health Status Question 2					
Self Assessed Health Status Question 1	Excellent	Very Good	Good	Fair	Poor	Total Proportion
Excellent	76.98	15.87	5.03	1.85	0.26	12.66
Very Good	12.12	67.55	18.48	1.73	0.12	29.00
Good	2.00	20.15	62.17	14.73	0.95	35.23
Fair	0.95	1.90	11.76	73.62	11.76	17.65
Poor	0	0.61	0.61	15.95	82.82	5.46
Total Proportion	14.13	29.07	30.01	19.79	7.00	100

Table A.2d: Transition Matrix (percent) Age 70 years and over

	Self Assessed Health Status Question 2					
Self Assessed Health Status Question 1	Excellent	Very Good	Good	Fair	Poor	Total Proportion
Excellent	66.88	23.75	7.50	1.88	0	9.50
Very Good	11.68	63.25	20.80	3.70	0.57	20.84
Good	2.91	19.31	61.02	15.66	1.09	32.60
Fair	0.67	2.69	13.68	69.28	13.68	26.48
Poor	0	0.56	3.37	19.66	76.40	10.57
Total Proportion	9.92	22.51	28.92	26.48	12.17	100