EVALUATING THE EFFECTS OF SURVEY DESIGN AND ADMINISTRATION ON THE MEASUREMENT OF SUBJECTIVE PHENOMENA: THE CASE OF SELF-ASSESSED HEALTH STATUS

Kathleen M. Danchik and Thomas F. Drury, National Center for Health Statistics

Health survey folklore suggests that if you are going to ask one question about a person's health, you could do no better than to ask respondents to rate their own health in overall terms. This belief assumes, among other things, that responses to such a question are insensitive to variations in survey design and administration. This paper assesses the tenability of this assumption by examining the effects of selected variations in questionnaire design, interview process, and organizational contexts upon measurements of self-assessed health status.

The specific aspects of survey design which we evaluated are shown in figure 1. We assessed the effects of these aspects of survey design and administration through a series of quasi-experimental case studies using data available from 34 national surveys. Figure 2 lists the national surveys used in this study. Figure 3 shows the variations in question wording and answer categories by survey.

Figure 4 summarizes the data which were compared for each survey on the specific aspects of survey design studies. We used this figure to identify strategic comparisons for surveys for each aspect under study. A "strategic comparison" is one in which two surveys differed with respect to the specific aspect of survey design and administration being analyzed but were as similar as possible in all other design and administration characteristics. Only two comparisons met this definition exactly, and some comparisons more closely approximated it than others.

To evaluate whether a particular aspect of survey design and administration had an effect on the measurement of self-assessed health status, we looked at three types of effects for each comparison: (1) the effects on the whole distribution, (2) the effects on the percent assessing their health as excellent, and (3) the effects on the percent assessing their health as fair or poor. The effects on the whole distribution are shown by an index of dissimilarity for each comparison[1]. This is a descriptive measure which is calculated by taking half the sum of the absolute differences between the percentages in each category in the distribution. The index represents the percentage of cases that would have to change categories to make the two distributions identical. The effects on the proportions of adults assessing their health as "excellent," or as "fair or poor" are described simply by showing the percentage point difference between the two compared survey estimates.

In the paper from which this presentation is abstracted, each aspect is treated separately. In this condensed version, I have divided the findings into three groups: Variations whose effects on the self-assessed health status item are minimal, those whose effects are moderate, and those whose effects are major. For each type of effect, I will give one example.

We found the following aspects of survey design and administration to have minimal effects:
(a) The use of an age reference in the question stem. But when we looked at the response distributions by age we found that adults aged 50 and over were somewhat affected by this guided comparison. Using an age reference, they tend to increase self-assessments of excellent health.

(b) Certain context variations.
(1) Simple changes in context, such as preceding the self-assessed health status item with different lists of chronic conditions, do not affect response distributions. The NHIS did this from 1978 to 1981.
(2) Using relatively objective questions prior to the self-assessed health status item (such as types of disabilities or medical expenses) in surveys of similar content (such as NMCES and NMCUES) do not affect the response distributions.
(c) Time of year the interview is conducted. Self-assessments of personal happiness have been linked to seasonality, i.e., spring highs and winter lows, but we found nothing of this sort for self-assessed health status.
(d) The data collection organization. The house can affect the measurement of SAHS through methods of handling nonresponse, their interviewer procedures, training and staff differences, coding rules and sampling procedures. House effects on self-assessed health status were minimal for the surveys we looked at.

As an example of an aspect that has minimal effects on the measurement of self-assessed health status, figure 5 shows a strategic comparison of the time of year the survey is administered. We chose to combine three years of NHIS data and look at the response distributions by quarter, which is roughly equivalent to the seasons, starting with quarter 1 as the winter season, months January, February and March. We found very consistent distributions of response: A 1.1 index of dissimilarity for the distributions with the largest differences, a 0.8 percentage point difference in estimates of adults in excellent health and a 0.5 percentage point difference in estimates of adults in fair or poor health.

Aspects of survey design and administration which we found to have moderate effects are:
(a) Question placement.
(b) Contexts which involve other subjective evaluations, such as happiness ratings in the NORC General Social Surveys.
(c) General survey content. This aspect appears to have small to moderate effects on self-assessment of health, though it is difficult to tease out these effects from possible context effects. Respondents were found to rate their health somewhat higher in health-related surveys as opposed to more general surveys. (The relative deprivation theory)
(d) Proxy respondents.

As an example of an aspect of survey design and administration with moderate effects on self-assessed health status, figure 6 compares self and proxy responses. The 1976 NHIS had a health habit supplement which required self-respondents. Perceived health status was included in this supplement. Some sample adult family members answered this question for themselves after another family member had previously responded for them to the same question in the main questionnaire. Findings parallel the conclusions of Kovar and Wilson from the 1972 NHIS Respondent Experiment[2]. Compared to proxy respondents, self-respondents tend to give themselves

lower "excellent" ratings and higher "good" ratings. Estimates of assessed "fair and poor" health remain about the same. The index of dissimilarity between these distributions is 5.6. The estimated proportions assessed as in excellent health differ by 5.6 percentage points, while the estimated proportions in fair or poor health differ by only 0.2 percentage points.

The aspects of survey design and administration which have major effects on self-assessed health status are the number of answer categories and the choice of adjectives used.

As an example of an aspect with a major effect on self-assessed health status, let's look at the number of answer categories used. As you can see in figure 7, the proportion of adults rating their health as very good draws mainly from the "excellent" and "good" categories. The proportion of adults rating their health as "fair or poor" remains relatively constant.

The 1981 NHIS question used four answer categories and the 1982 NHIS, five. The proportions of self-respondents answering fair or poor in these two years only differed by 0.2 percentage points. The percentage point difference between estimates of excellent health was 9.8.

A comparison of distributions from the fourth quarter of the regular 1979 NHIS and the 1979 Fourth Quarter Evaluation Study paints a similar picture: the proportions of adults rating their health as "fair or poor" differed only by 0.9 percentage points. Estimates of excellent health differed by 12.2 percentage points.

A summary of our findings is shown in figure 8. Please keep in mind that we made several comparisons of survey aspects which were less than our strategic comparison ideal. In one case, we didn't make a comparison at all. Use of the global referent "in general" had no adequate comparisons available. Two other aspects studied resulted in conflicting findings. Comparisons made in reference to the mode of questionnaire administration showed moderate or major effects which we suspect are influenced by question context variations. The effects of survey sponsor were found to be very minimal in one case and moderate in two other comparisons. These effects are also suspected to be influenced by contextual differences. Thus the effects of these aspects are left open for further study.

I feel I must mention here that context was not controlled for in 7 of the 11 sets of comparisons made. The effects of certain other aspects besides mode of administration and survey sponsor, such as question location and survey content could possibly have been minimized had the effects of different contexts been eliminated.

We believe that our study has indicated that self-assessed health status questions are not greatly affected by variations in survey design and administration, either for the general population or the population grouped by age and sex. However, some questions for future study arise out of the limitations of this case-study approach.

(1) One very important implication of the case studies undertaken here is the need for laboratory approaches to measurement errors. Despite the large number of surveys available to us, it was difficult to identify "strategic comparisons" for many of the case studies. However, these case studies serve the purpose of directing attention to components of measurement error, such as context and question wording, which laboratory approaches can deal with more efficiently.

(2) One of the areas we do not address here is the effect of survey design and administration on intra-individual variations in responses to self-assessed health status questions. At first glance, it would appear that reinterview surveys would be an appropriate vehicle for looking at this topic. An implication of our findings is that reinterview survey findings for subjective measures may be systematically affected by changes in the context of questionnaire items, as most reinterview surveys are shorter than the original and may change sequence and context of questions.

(3) According to Turner, questions which measure subjective phenomena that are particularly vulnerable to response variations have certain characteristics in common[3,4]. They are generally vague in their meaning, are concerned with topics of low salience to respondents, have a seemingly arbitrary choice of response categories and have answers with no explicit behavioral implications in the everyday lives of respondents.

We submit that self-assessed health status measurements may be robust to variations in survey design and administration characteristics precisely because they ask about a highly salient topic through a question with clear meaning, use well known and widely used answer categories, and are sensitive to changes in health and illness behaviors. It may even be that single-item measures of subjective phenomena don't get any better than this. Additional studies of the effects of survey design and administration on other types of subjective measurements would further clarify the subject matter and question wording to avoid unnecessary measurement error.

In view of our findings, we would recommend the following question wording:

Compared to other persons your age, would you say your health is excellent, very good, good, fair or poor?

We prefer the reference to age because self-assessments of health by adults around 50 and over are affected by this guided comparison. We prefer five answer categories because it permits a finer classification of people at the positive end of the assessed health spectrum without affecting responses to the fair or poor categories. We recommend placing the question either at the beginning, whenever possible as the very first item, or at the end of the questionnaire after demographic items, but before a question on family income. We would avoid placing the question in the middle of the questionnaire, or after other health questions, particularly any which will be subsequently related to self-assessed health status.

We are aware that major national surveys are now using the question:

"Would you say your health in general is excellent, very good, good, fair or poor?"

Since our recommendation for question wording is different from what is now being used, we propose that these two variations in wording be looked at further through laboratory studies of measurement error. Another proposal is that a national survey such as the National Health Interview Survey allocate part of its sample to testing the effects of these two different forms of question wording on self-assessed health status.

References

1. Shryock HS, Siegel JS, and associates: The Methods and Materials of Demography. Vol. 1. U. S. Department of Commerce. Bureau of the Census. Issued May 1973.

2. Kovar MG and Wilson RW: Perceived health status--how good is proxy reporting? 1976 Social Statistics Section. Proceedings of the American Statistical Association.

3. Turner CF: Surveys of subjective phenomena: a working paper. In Measurement of Subjective Phenomena, ed. Denis F. Johnston. U. S. Department of Commerce. Bureau of the Census. Issued October 1981.

4. Turner CF and Martin E, eds.: Surveying Subjective Phenomena Vol. 1. Russell Sage Foundation. N.Y. 1983.

A longer version of this paper, including detailed references and tables, is available from Ms. Danchik, Room 2-27 Center Building, 3700 East-West Highway, Hyattsville, Maryland 20782.

FIGURE 1. SELECTED ASPECTS OF SURVEY DESIGN
AND ADMINISTRATION

Questionnaire design

Question wording
    Uses an age reference
    Number of answer categories
    Uses "in general"
    Adjectives used
Question placement
Specific question context
General survey content

Interview process

Time of year
Mode of administration
Self/proxy

Organizational contexts

Survey sponsor
Survey field organization (house)

FIGURE 2. SELECTED NATIONAL SURVEYS WHICH INCLUDE
A SELF-ASSESSED HEALTH STATUS ITEM

National Health Interview Survey (NHIS) 1972-82
NHIS Health Habits Supplement (NHIS-HHS) 1976
NHIS Fourth Quarter Evaluation Study (FQES) 1979
NHIS Survey Research Center Telephone Experiment
    (NTHIS-SRC) 1979
National Telephone Health Interview Survey Wave I (NTHIS I) 1980
National Telephone Health Interview Survey Wave II (NTHIS II) 1981
National Medical Care Expenditure Survey (NMCES) 1977
National Medical Care Utilization and Expenditure Survey
    (NMCUES) 1980
National Health and Nutrition Examination Survey I
    (NHANES I) 1971-75
National Health and Nutrition Examination Survey II
    (NHANES II) 1976-80
National Survey of Personal Health Practices and
    Consequences Wave I (NSPHPC I) 1979
National Survey of Personal Health Practices and
    Consequences Wave II (NSPHPC I) 1980
National Opinion Research Center - General Social Surveys
    (NORC-GSS) 1972-77, 1980
National Survey of Access to Medical Care (NORC-CHAS) 1975-76
1978 Survey of Disability and Work (SSA-SDW) 1978
National Longitudinal Survey of Mature Men (NLS)
    1966, 1969, 1978, 1980, 1981
Canada Fitness Survey (CFS) 1981

FIGURE 3. VARIATIONS IN QUESTION WORDING AND ANSWER CATEGORIES BY SURVEY

QUESTION	SURVEY
Compared to other persons/people your age, would you say your health is excellent, good, fair or poor?	NHIS 1972-1981 NTHIS-SRC 1979 NTHIS I 1980 NTHIS II 1981 NMCES 1977 NMCUES 1980 NHIS-HHS 1976 NSPHPC I 1979 NSPHPC II 1980
Would you say your own health, in general, is excellent, good, fair or poor?	NORC-GSS 1972-77, 80 NORC-CHAS 1975-76
Would you say your health in general is excellent, very good, good, fair or poor?	NHANES I 1971-75 NHANES II 1976-80 NHIS 1982
First, would you say your health is excellent, good, fair or poor?	NSPHPC I 1979 NSPHPC II 1980
Compared to other persons your age, would you say your health is excellent, very good, good, fair or poor?	FQES 1979
About your health now, would you say it is excellent, good, fair or poor?	SSA-SDW 1978
Would you rate your health, compared with other men of about your age, as excellent, good, fair or poor?	NLS 1966, 1969, 1978, 1980, 1981
In general, how would you describe your state of health -- very good, good, average, poor or very poor?	CFS 1981

FIGURE 4. SELECTED NATIONAL SURVEYS WHICH INCLUDE A SELF-ASSESSED HEALTH STATUS ITEM BY SELECTED ASPECTS OF SURVEY DESIGN AND ADMINISTRATION

Survey	QUESTIONNAIRE DESIGN							INTERVIEW PROCESS			ORGANIZATIONAL CONTEXTS	
	Uses "compared to other people your age"	Uses "in general"	Number of answer categories	Adjectives used[1]	Question placement	Specific question context	General survey content	Time of year	Questionnaire administration[2]	Self/ proxy[3]	Survey sponsor	Survey Field Organization (House)
NHIS (1972-81)	Yes	No	4	EGFP	middle	health conditions	health	Jan-Dec	P	Yes	NCHS	Census
NHIS (1982)	No	Yes	5	EVGFP	middle	doctor visits	health	Jan-Dec	P	Yes	NCHS	Census
NHIS-HHS (1976)	Yes	No	4	EGFP	last/middle	health habits	health	Jan-Dec	P	Yes	NCHS	Census
FQES (1979)	Yes	No	5	EVGFP	middle	conditions/utilization	health	Oct-Dec	P	Yes	NCHS	Census
NTHIS-SRC (1979)	Yes	No	4	EGFP	late	number of hospitalizations	health	Oct-Dec	T	Yes	NCHS	SRC
NTHIS I (1980)	Yes	No	4	EGFP	first	first	smoking	Jul-Dec	T	No	NCHS	NCHS
NTHIS II (1981)	Yes	No	4	EGFP	first	first	smoking	Jan-Jun	T	No	NCHS	NCHS
NMCES (1977)	Yes	No	4	EGFP	middle	disability	utilization & expenditures	Jan-Dec	P	Yes	NCHS/NCHSR	RTI/NORC
NMCUES (1980)	Yes	No	4	EGFP	middle	medical expenses	utilization & expenditures	Jan-Dec	P	Yes	NCHS/NCHSR	RTI/NORC
NHANES I (1971-74)	No	Yes	5	EGVFP	first or early	first or demographics	health	4/71 - 6/74	P	No	NCHS	NCHS
NHANES I (1974-75)	No	Yes	5	EGVFP	early	first or demographics	health	7/74 - 9/75	P	No	NCHS	Census
NHANES II (1976-80)	No	Yes	5	EGVFP	first or early	demographics	health	2/76 - 2/80	P	No	NCHS	Census
NSPHPC I (1979) Q1	No	No	4	EGFP	first	first	health	May	T	No	NCHS	Chilton
NSPHPC I (1979) Q53	Yes	No	4	EGFP	middle	control over future health	health	May	T	No	NCHS	Chilton
NSPHPC II (1980) Q1	No	No	4	EGFP	first	first	health	May	T	No	NCHS	Chilton
NSPHPC II (1980) Q50	Yes	No	4	EGFP	middle	control over future health	health	May	T	No	NCHS	Chilton
NORC-GSS (1972-77, 80)	No	Yes	4	EGFP	middle	happiness	general	Feb-Apr	P	No	NSF	NORC
NORC-CHAS (1975-76)	No	Yes	4	EGFP	middle	utilization	access to medical care	late '75-early '76	P	Yes	RWJF/NCHSR	NORC
SSA-SDW (1978)	No	No	4	EGFP	middle	number of doctor visits	disability & work experience	Jul-Sept	P	No	SSA	Census
NLS (1966, 1969, 1981)	Yes	No	4	EGFP	middle	1966, 1969 - limitation of activity	work experience	Jul-Sept	P	No	DOL/CHRR	Census
						1978, 1980 - comparative health						
NLS (1978, 1980)	Yes	No	4	EGFP	middle	1981 - help with ADL	work experience	Jul-Sept	T	No	DOL/CHRR	Census
CFS (1981)	No	Yes	5	VGAPV	middle	health habits	fitness	Feb-Jul	SA	No	Fitness Canada	CFS

[1] EGFP: Excellent, good, fair, poor
EVGFP: Excellent, very good, good, fair, poor
VGAPV: Very good, good, average, poor, very poor

[2] P: Personal interview
T: Telephone interview
SA: Self-administered

[3] Yes: Proxy respondents accepted
No : Proxy respondents not accepted

Figure 5.
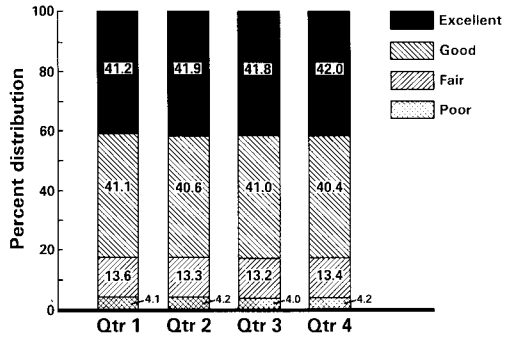### Effects of time of year of survey on self-assessed health status: 1975-77 NHIS

Percent distribution

	Qtr 1	Qtr 2	Qtr 3	Qtr 4
Excellent	41.2	41.9	41.8	42.0
Good	41.1	40.6	41.0	40.4
Fair	13.6	13.3	13.2	13.4
Poor	4.1	4.2	4.0	4.2

Figure 6.
Effects of proxy responses on self-assessed health status: 1976 NHIS

Percent distribution

	Self (from supplement)	Proxy (from main questionnaire)
Excellent	45.5	51.1
Good	43.7	38.3
Fair	8.9	8.8
Poor	1.9	1.8
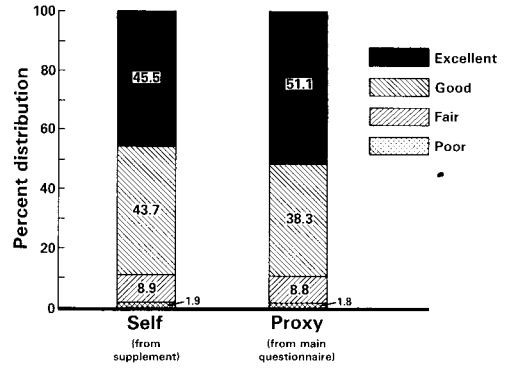
Figure 7.
Effects of the number of answer categories on self-assessed health status

Percent distribution

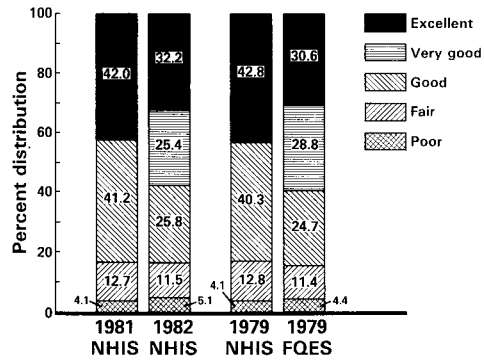|	1981 NHIS	1982 NHIS	1979 NHIS	1979 FQES
Excellent	42.0	32.2	42.8	30.6
Very good		25.4		28.8
Good	41.2	25.8	40.3	24.7
Fair	12.7	11.5	12.8	11.4
Poor	4.1	5.1	4.1	4.4

FIGURE 8. SUMMARY OF EFFECTS OF SELECTED ASPECTS OF SURVEY DESIGN AND ADMINISTRATION ON SELF-ASSESSED HEALTH STATUS

Design aspect	Surveys compared	Uncontrolled aspects	Age range	Unadjusted Index of Dissimilarity	Effect — Percentage point difference in estimates of Excellent health	Fair or poor health
Use of age reference	1980 NTHIS I 1980 NSPHPC II	Time of year House	20-64	2.6	1.9	0.1
	1978 NHIS (Qtr 3) 1978 SSA-SDW	Survey content Sponsor Specific context	20-64	0.9	0.1	0.8
	1979 FQES 1982 NHIS	Age reference Specific context Year of interview	20+	4.5	1.5	0.5
Use of "in general"	None	(5 or more aspects)	---	----	---	---
Number of answer categories	1981 NHIS 1982 NHIS	Age reference Specific context	20+	NA	9.8	0.2
	1979 NHIS (Qtr 4) 1979 FQES	Specific context	20+	NA	12.2	0.9
Adjectives used	1982 NHIS (Qtrs 1 & 2) 1981 CFS	(5 aspects)	20+	23.7[1]	NA	NA
	1976-80 NHANES II 1981 CFS	(6 aspects)	20-74	21.0[1]	NA	NA
Question placement	1976-80 NHANES II 1982 NHIS	(Year of interview) Specific context	20-74	6.2[1]	6.1	2.3
	1979 NSPHPC I (Q1) 1979 NSPHPC I (Q53)	Age reference Specific context	20-64	5.3	4.9	0.2
Specific question context	1979-80 NHIS by 6 condition lists	None	20+	2.3[2]	1.4[2]	1.4[2]
	1977 NMCES 1980 NMCUES	Co-sponsor (Year of interview)	20+	1.9	1.3	0.4
	1980 NORC-GSS by form A, form B	None	20+	5.6 (all) 6.8 (married)	1.7 2.7	1.2 0.7
General survey content	1975-76 NORC-GSS 1975-76 NORC-CHAS	Specific context Sponsor	20+	5.0	2.4	5.1
	1981 NLS 1981 NHIS	Specific context Sponsor	men 59-75	3.1	2.1	3.0
Time of year	1975-77 NHIS by quarter	None	20+	1.1[2]	0.8[2]	0.5[2]
	1980 NTHIS I 1981 NTHIS II	None	20+	1.5	0.2	1.7
Questionnaire administration	1979 NTHIS (Qtr 2) 1979 NSPHPC I	Specific context House	20-64	13.9	13.9	2.8
	1979 NHIS (Qtr 4) 1979 NTHIS (SRC)	Specific context House	25+	5.0	0.3	4.8
Self/proxy	1976 NHIS-HHS by respondent status	Specific context	20+	5.6	5.6	0.2
Survey sponsor	1978 SSA-SDW 1978 NHIS (Qtr 3)	Age reference Specific context Survey content	20-64	0.9	0.1	0.8
	1975-76 NORC-GSS 1975-76 NORC-CHAS	Specific context Sponsor	20+	5.0	2.4	5.1
	1981 NLS 1981 NHIS	Specific context Sponsor	men 59-75	3.1	2.1	3.0
Survey field organization (house)	1980 NTHIS I 1980 NSPHPC II	Age reference Time of year	20-64	2.6	1.9	0.1

[1]Five answer categories.
[2]Calculated for the greatest range difference.